2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

(NAACL HLT 2013)

Atlanta, Georgia, USA
9-14 June 2013

Volume 1 of 2

ISBN: 978-1-62748-838-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2013) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Model With Minimal Translation Units, But Decode With Phrases
Nadir Durrani, Alexander Fraser and Helmut Schmid 1

Beyond Left-to-Right: Multiple Decomposition Structures for SMT
Hui Zhang, Kristina Toutanova, Chris Quirk and Jianfeng Gao 12

Improved Reordering for Phrase-Based Translation using Sparse Features
Colin Cherry 22

Simultaneous Word-Morpheme Alignment for Statistical Machine Translation
Elif Eyigöz, Daniel Gildea and Kemal Oflazer 32

Multi-faceted Event Recognition with Bootstrapped Dictionaries
Ruihong Huang and Ellen Riloff 41

Named Entity Recognition with Bilingual Constraints
Wanxiang Che, Mengqiu Wang, Christopher D. Manning and Ting Liu 52

Minimally Supervised Method for Multilingual Paraphrase Extraction from Definition Sentences on the Web
Yulan Yan, Chikara Hashimoto, Kentaro Torisawa, Takao Kawai, Jun'ichi Kazama and Stijn De Saeger 63

Relation Extraction with Matrix Factorization and Universal Schemas
Sebastian Riedel, Limin Yao, Andrew McCallum and Benjamin M. Marlin 74

Extracting the Native Language Signal for Second Language Acquisition
Ben Swanson and Eugene Charniak 85

An Analysis of Frequency- and Memory-Based Processing Costs
Marten van Schijndel and William Schuler 95

Cross-Lingual Semantic Similarity of Words as the Similarity of Their Semantic Word Responses
Ivan Vulić and Marie-Francine Moens 106

Combining multiple information types in Bayesian word segmentation
Gabriel Doyle and Roger Levy 117

Training Parsers on Incompatible Treebanks
Richard Johansson 127

Learning a Part-of-Speech Tagger from Two Hours of Annotation
Dan Garrette and Jason Baldridge 138

Experiments with Spectral Learning of Latent-Variable PCFGs
Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster and Lyle Ungar 148

Representing Topics Using Images
Nikolaos Aletras and Mark Stevenson 158

Drug Extraction from the Web: Summarizing Drug Experiences with Multi-Dimensional Topic Models
Michael J. Paul and Mark Dredze 168

Towards Topic Labeling with Phrase Entailment and Aggregation
Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng and Shafiq Joty 179

Topic Segmentation with a Structured Topic Model
Lan Du, Wray Buntine and Mark Johnson 190

Text Alignment for Real-Time Crowd Captioning
Iftekhar Naim, Daniel Gildea, Walter Lasecki and Jeffrey P. Bigham 201

Discriminative Joint Modeling of Lexical Variation and Acoustic Confusion for Automated Narrative Retelling Assessment
Maider Lehr, Izhak Shafran, Emily Prud'hommeaux and Brian Roark 211

Using Out-of-Domain Data for Lexical Addressee Detection in Human-Human-Computer Dialog
Heeyoung Lee, Andreas Stolcke and Elizabeth Shriberg 221

Segmentation Strategies for Streaming Speech Translation
Vivek Kumar Rangarajan Sridhar, John Chen, Srinivas Bangalore, Andrej Ljolje and Rathinavelu Chengalvarayan 230

Enforcing Subcategorization Constraints in a Parser Using Sub-parses Recombining
Seyed Abolghasem Mirroshandel, Alexis Nasr and Benoît Sagot 239

Large-Scale Discriminative Training for Statistical Machine Translation Using Held-Out Line Search
Jeffrey Flanigan, Chris Dyer and Jaime Carbonell 248

Measuring Term Informativeness in Context
Zhaohui Wu and C. Lee Giles 259

Unsupervised Learning Summarization Templates from Concise Summaries
Horacio Saggion 270

Classification of South African languages using text and acoustic based methods: A case of six selected languages
Peleira Nicholas Zulu 280

Improving Syntax-Augmented Machine Translation by Coarsening the Label Set
Greg Hanneman and Alon Lavie 288

Keyphrase Extraction for N-best Reranking in Multi-Sentence Compression
Florian Boudin and Emmanuel Morin 298

Development of a Persian Syntactic Dependency Treebank
Mohammad Sadegh Rasooli, Manouchehr Kouhestani and Amirsaeid Moloodi 306

Improving reordering performance using higher order and structural features
Mitesh M. Khapra, Ananthakrishnan Ramanathan and Karthik Visweswariah 315

Massively Parallel Suffix Array Queries and On-Demand Phrase Extraction for Statistical Machine Translation Using GPUs
Hua He, Jimmy Lin and Adam Lopez 325

Discriminative Training of 150 Million Translation Parameters and Its Application to Pruning
Hendra Setiawan and Bowen Zhou 335

Applying Pairwise Ranked Optimisation to Improve the Interpolation of Translation Models
Barry Haddow 342

Dialectal Arabic to English Machine Translation: Pivoting through Modern Standard Arabic
Wael Salloum and Nizar Habash 348

What to do about bad language on the internet
Jacob Eisenstein 359

Minibatch and Parallelization for Online Large Margin Structured Learning
Kai Zhao and Liang Huang 370

Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters
Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider and Noah A. Smith 380

Parser lexicalisation through self-learning
Marek Rei and Ted Briscoe 391

Mining User Relations from Online Discussions using Sentiment Analysis and Probabilistic Matrix Factorization
Minghui Qiu, Liu Yang and Jing Jiang 401

Focused training sets to reduce noise in NER feature models
Amber McKenzie 411

Learning to Relate Literal and Sentimental Descriptions of Visual Properties
Mark Yatskar, Svitlana Volkova, Asli Celikyilmaz, Bill Dolan and Luke Zettlemoyer 416

Morphological Analysis and Disambiguation for Dialectal Arabic
Nizar Habash, Ryan Roth, Owen Rambow, Ramy Eskander and Nadi Tomeh 426

Using a Supertagged Dependency Language Model to Select a Good Translation in System Combination
Wei-Yun Ma and Kathleen McKeown 433

Dudley North visits North London: Learning When to Transliterate to Arabic
Mahmoud Azab, Houda Bouamor, Behrang Mohit and Kemal Oflazer 439

Better Twitter Summaries?
Joel Judd and Jugal Kalita 445

Training MRF-Based Phrase Translation Models using Gradient Ascent
Jianfeng Gao and Xiaodong He 450

Automatic Morphological Enrichment of a Morphologically Underspecified Treebank
Sarah Alkuhlani, Nizar Habash and Ryan Roth 460

A Beam-Search Decoder for Normalization of Social Media Text with Application to Machine Translation
Pidong Wang and Hwee Tou Ng 471

Parameter Estimation for LDA-Frames
Jiří Materna 482

Approximate PCFG Parsing Using Tensor Decomposition
Shay B. Cohen, Giorgio Satta and Michael Collins 487

Negative Deceptive Opinion Spam
Myle Ott, Claire Cardie and Jeffrey T. Hancock 497

Improving speech synthesis quality by reducing pitch peaks in the source recordings
Luisina Violante, Pablo Rodríguez Zivic and Agustín Gravano 502

Robust Systems for Preposition Error Correction Using Wikipedia Revisions
Aoife Cahill, Nitin Madnani, Joel Tetreault and Diane Napolitano 507

Supervised Bilingual Lexicon Induction with Multiple Monolingual Signals
Ann Irvine and Chris Callison-Burch 518

Creating Reverse Bilingual Dictionaries
Khang Nhut Lam and Jugal Kalita 524

Identification of Temporal Event Relationships in Biographical Accounts
Lucian Silcox and Emmett Tomai 529

Predicative Adjectives: An Unsupervised Criterion to Extract Subjective Adjectives
Michael Wiegand, Josef Ruppenhofer and Dietrich Klakow 534

Modeling Syntactic and Semantic Structures in Hierarchical Phrase-based Translation
Junhui Li, Philip Resnik and Hal Daumé III 540

Using Derivation Trees for Informative Treebank Inter-Annotator Agreement Evaluation
Seth Kulick, Ann Bies, Justin Mott, Mohamed Maamouri, Beatrice Santorini and Anthony Kroch 550

Embracing Ambiguity: A Comparison of Annotation Methodologies for Crowdsourcing Word Sense Labels
David Jurgens 556

Compound Embedding Features for Semi-supervised Learning
Mo Yu, Tiejun Zhao, Daxiang Dong, Hao Tian and Dianhai Yu 563

On Quality Ratings for Spoken Dialogue Systems – Experts vs. Users
Stefan Ultes, Alexander Schmitt and Wolfgang Minker 569

Overcoming the Memory Bottleneck in Distributed Training of Latent Variable Models of Text
Yi Yang, Alexander Yates and Doug Downey 579

Processing Spontaneous Orthography
Ramy Eskander, Nizar Habash, Owen Rambow and Nadi Tomeh 585

Purpose and Polarity of Citation: Towards NLP-based Bibliometrics
Amjad Abu-Jbara, Jefferson Ezra and Dragomir Radev 596

Estimating effect size across datasets
Anders Søgaard 607

Systematic Comparison of Professional and Crowdsourced Reference Translations for Machine Translation
Rabih Zbib, Gretchen Markiewicz, Spyros Matsoukas, Richard Schwartz and John Makhoul . 612

Down-stream effects of tree-to-dependency conversions
Jakob Elming, Anders Johannsen, Sigrid Klerke, Emanuele Lapponi, Hector Martinez Alonso and Anders Søgaard 617

The Life and Death of Discourse Entities: Identifying Singleton Mentions
Marta Recasens, Marie-Catherine de Marneffe and Christopher Potts 627

Automatic Generation of English Respellings
Bradley Hauer and Grzegorz Kondrak 634

A Simple, Fast, and Effective Reparameterization of IBM Model 2
Chris Dyer, Victor Chahuneau and Noah A. Smith 644

Phrase Training Based Adaptation for Statistical Machine Translation
Saab Mansour and Hermann Ney 649

Translation Acquisition Using Synonym Sets
Daniel Andrade, Masaki Tsuchida, Takashi Onishi and Kai Ishikawa 655

Supersense Tagging for Arabic: the MT-in-the-Middle Attack
Nathan Schneider, Behrang Mohit, Chris Dyer, Kemal Oflazer and Noah A. Smith 661

Zipfian corruptions for robust POS tagging
Anders Søgaard 668

A Multi-Dimensional Bayesian Approach to Lexical Style
Julian Brooke and Graeme Hirst 673

Unsupervised Domain Tuning to Improve Word Sense Disambiguation
Judita Preiss and Mark Stevenson 680

What's in a Domain? Multi-Domain Learning for Multi-Attribute Data
Mahesh Joshi, Mark Dredze, William W. Cohen and Carolyn P. Rosé 685

An opinion about opinions about opinions: subjectivity and the aggregate reader
Asad Sayeed 691

An Examination of Regret in Bullying Tweets
Jun-Ming Xu, Benjamin Burchfiel, Xiaojin Zhu and Amy Bellmore 697

A Cross-language Study on Automatic Speech Disfluency Detection
Wen Wang, Andreas Stolcke, Jiahong Yuan and Mark Liberman 703

Distributional semantic models for the evaluation of disordered language
Masoud Rouhizadeh, Emily Prud'hommeaux, Brian Roark and Jan van Santen 709

Atypical Prosodic Structure as an Indicator of Reading Level and Text Difficulty
Julie Medero and Mari Ostendorf 715

Using Document Summarization Techniques for Speech Data Subset Selection
Kai Wei, Yuzong Liu, Katrin Kirchhoff and Jeff Bilmes 721

Semi-Supervised Discriminative Language Modeling with Out-of-Domain Text Data
Arda Çelebi and Murat Saraçlar 727

More than meets the eye: Study of Human Cognition in Sense Annotation
Salil Joshi, Diptesh Kanojia and Pushpak Bhattacharyya 733

Improving Lexical Semantics for Sentential Semantics: Modeling Selectional Preference and Similar Words in a Latent Variable Model
Weiwei Guo and Mona Diab 739

Linguistic Regularities in Continuous Space Word Representations
Tomas Mikolov, Wen-tau Yih and Geoffrey Zweig 746

TruthTeller: Annotating Predicate Truth
Amnon Lotan, Asher Stern and Ido Dagan 752

PPDB: The Paraphrase Database
Juri Ganitkevitch, Benjamin Van Durme and Chris Callison-Burch 758

Exploiting the Scope of Negations and Heterogeneous Features for Relation Extraction: A Case Study for Drug-Drug Interaction Extraction
Md. Faisal Mahbub Chowdhury and Alberto Lavelli 765

Graph-Based Seed Set Expansion for Relation Extraction Using Random Walk Hitting Times
Joel Lang and James Henderson 772

Distant Supervision for Relation Extraction with an Incomplete Knowledge Base
Bonan Min, Ralph Grishman, Li Wan, Chang Wang and David Gondek 777

Measuring the Structural Importance through Rhetorical Structure Index
Narine Kokhlikyan, Alex Waibel, Yuqi Zhang and Joy Ying Zhang 783

Separating Fact from Fear: Tracking Flu Infections on Twitter
Alex Lamb, Michael J. Paul and Mark Dredze 789

Differences in User Responses to a Wizard-of-Oz versus Automated System
Jesse Thomason and Diane Litman 796

Improving the Quality of Minority Class Identification in Dialog Act Tagging
Adinoyi Omuya, Vinodkumar Prabhakaran and Owen Rambow 802

Discourse Connectors for Latent Subjectivity in Sentiment Analysis
Rakshit Trivedi and Jacob Eisenstein 808

Coherence Modeling for the Automated Assessment of Spontaneous Spoken Responses
Xinhao Wang, Keelan Evanini and Klaus Zechner 814

Disfluency Detection Using Multi-step Stacked Learning
Xian Qian and Yang Liu 820

Using Semantic Unification to Generate Regular Expressions from Natural Language
Nate Kushman and Regina Barzilay 826

Probabilistic Frame Induction
Jackie Chi Kit Cheung, Hoifung Poon and Lucy Vanderwende 837

A Quantum-Theoretic Approach to Distributional Semantics
William Blacoe, Elham Kashefi and Mirella Lapata 847

Answer Extraction as Sequence Tagging with Tree Edit Distance
Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch and Peter Clark 858

Open Information Extraction with Tree Kernels
Ying Xu, Mi-Young Kim, Kevin Quinn, Randy Goebel and Denilson Barbosa 868

Finding What Matters in Questions
Xiaoqiang Luo, Hema Raghavan, Vittorio Castelli, Sameer Maskey and Radu Florian 878

A Just-In-Time Keyword Extraction from Meeting Transcripts
Hyun-Je Song, Junho Go, Seong-Bae Park and Se-Young Park 888

Same Referent, Different Words: Unsupervised Mining of Opaque Coreferent Mentions
Marta Recasens, Matthew Can and Daniel Jurafsky 897

Global Inference for Bridging Anaphora Resolution
Yufang Hou, Katja Markert and Michael Strube 907

Classifying Temporal Relations with Rich Linguistic Knowledge
Jennifer D'Souza and Vincent Ng 918

Improved Information Structure Analysis of Scientific Documents Through Discourse and Lexical Constraints
Yufan Guo, Roi Reichart and Anna Korhonen 928

Adaptation of Reordering Models for Statistical Machine Translation
Boxing Chen, George Foster and Roland Kuhn 938

Multi-Metric Optimization Using Ensemble Tuning
Baskaran Sankaran, Anoop Sarkar and Kevin Duh 947

Grouping Language Model Boundary Words to Speed K–Best Extraction from Hypergraphs
Kenneth Heafield, Philipp Koehn and Alon Lavie 958

A Systematic Bayesian Treatment of the IBM Alignment Models
Yarin Gal and Phil Blunsom 969

Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering
Ekaterina Shutova and Lin Sun 978

Three Knowledge-Free Methods for Automatic Lexical Chain Extraction
Steffen Remus and Chris Biemann 989

Combining Heterogeneous Models for Measuring Relational Similarity
Alisa Zhila, Wen-tau Yih, Christopher Meek, Geoffrey Zweig and Tomas Mikolov 1000

Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter
Shane Bergsma, Mark Dredze, Benjamin Van Durme, Theresa Wilson and David Yarowsky 1010

To Link or Not to Link? A Study on End-to-End Tweet Entity Linking
Stephen Guo, Ming-Wei Chang and Emre Kiciman 1020

A Latent Variable Model for Viewpoint Discovery from Threaded Forum Posts
Minghui Qiu and Jing Jiang 1031

Identifying Intention Posts in Discussion Forums
Zhiyuan Chen, Bing Liu, Meichun Hsu, Malu Castellanos and Riddhiman Ghosh 1041

Dependency-based empty category detection via phrase structure trees
Nianwen Xue and Yaqin Yang 1051

Target Language Adaptation of Discriminative Transfer Parsers
Oscar Täckström, Ryan McDonald and Joakim Nivre 1061

Emergence of Gricean Maxims from Multi-Agent Decision Theory
Adam Vogel, Max Bodoia, Christopher Potts and Daniel Jurafsky 1072

Open Dialogue Management for Relational Databases
Ben Hixon and Rebecca J. Passonneau 1082

A method for the approximation of incremental understanding of explicit utterance meaning using predictive models in finite domains
David DeVault and David Traum 1092

Paving the Way to a Large-scale Pseudosense-annotated Dataset
Mohammad Taher Pilehvar and Roberto Navigli 1100

Labeling the Languages of Words in Mixed-Language Documents using Weakly Supervised Methods
Ben King and Steven Abney 1110

Learning Whom to Trust with MACE
Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani and Eduard Hovy 1120

Supervised All-Words Lexical Substitution using Delexicalized Features
György Szarvas, Chris Biemann and Iryna Gurevych 1131

A Tensor-based Factorization Model of Semantic Compositionality
Tim Van de Cruys, Thierry Poibeau and Anna Korhonen 1142

A Participant-based Approach for Event Summarization Using Twitter Streams
Chao Shen, Fei Liu, Fuliang Weng and Tao Li 1152

Towards Coherent Multi-Document Summarization
Janara Christensen, Mausam, Stephen Soderland and Oren Etzioni 1163

Generating Expressions that Refer to Visible Objects
Margaret Mitchell, Kees van Deemter and Ehud Reiter 1174

Supervised Learning of Complete Morphological Paradigms
Greg Durrett and John DeNero 1185

Optimal Data Set Selection: An Application to Grapheme-to-Phoneme Conversion
Young-Bum Kim and Benjamin Snyder 1196

Knowledge-Rich Morphological Priors for Bayesian Language Models
Victor Chahuneau, Noah A. Smith and Chris Dyer 1206