49th Annual Meeting of the Association for Computational Linguistics 2011

(ACL 2011)

Portland, Oregon, USA
19 – 24 June 2011

Volume 1 of 2

ISBN: 978-1-61839-111-7

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© (2011) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2011)

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

A Word-Class Approach to Labeling PSCFG Rules for Machine Translation
Andreas Zollmann and Stephan Vogel 1

Deciphering Foreign Language
Sujith Ravi and Kevin Knight 12

Effective Use of Function Words for Rule Generalization in Forest-Based Translation
Xianchao Wu, Takuya Matsuzaki and Jun'ichi Tsujii 22

Combining Morpheme-based Machine Translation with Post-processing Morpheme Prediction
Ann Clifton and Anoop Sarkar 32

Evaluating the Impact of Coder Errors on Active Learning
Ines Rehbein and Josef Ruppenhofer 43

A Fast and Accurate Method for Approximate String Search
Ziqi Wang, Gu Xu, Hang Li and Ming Zhang 52

Domain Adaptation by Constraining Inter-Domain Variability of Latent Feature Representation
Ivan Titov 62

Exact Decoding of Syntactic Translation Models through Lagrangian Relaxation
Alexander M. Rush and Michael Collins 72

Jigs and Lures: Associating Web Queries with Structured Entities
Patrick Pantel and Ariel Fuxman 83

Semi-Supervised SimHash for Efficient Document Similarity Search
Qixia Jiang and Maosong Sun 93

Joint Annotation of Search Queries
Michael Bendersky, W. Bruce Croft and David A. Smith 102

Query Weighting for Ranking Model Adaptation
Peng Cai, Wei Gao, Aoying Zhou and Kam-Fai Wong 112

Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification
Yulan He, Chenghua Lin and Harith Alani 123

Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification
Danushka Bollegala, David Weir and John Carroll 132

Learning Word Vectors for Sentiment Analysis
Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts 142

Target-dependent Twitter Sentiment Classification
Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu and Tiejun Zhao 151

A Comprehensive Dictionary of Multiword Expressions
Kosho Shudo, Akira Kurahone and Toshifumi Tanabe 161

Multi-Modal Annotation of Quest Games in Second Life
Sharon Gower Small, Jennifer Strommer-Galley and Tomek Strzalkowski 171

A New Dataset and Method for Automatically Grading ESOL Texts
Helen Yannakoudakis, Ted Briscoe and Ben Medlock 180

Collecting Highly Parallel Data for Paraphrase Evaluation
David Chen and William Dolan 190

A Large Scale Distributed Syntactic, Semantic and Lexical Language Model for Machine Translation
Ming Tan, Wenli Zhou, Lei Zheng and Shaojun Wang 201

Goodness: A Method for Measuring Machine Translation Confidence
Nguyen Bach, Fei Huang and Yaser Al-Onaizan 211

MEANT: An inexpensive, high-accuracy, semi-automatic metric for evaluating translation utility based on semantic roles
Chi-kiu Lo and Dekai Wu 220

An exponential translation model for target language morphology
Michael Subotin 230

Bayesian Inference for Zodiac and Other Homophonic Ciphers
Sujith Ravi and Kevin Knight 239

Interactive Topic Modeling
Yuening Hu, Jordan Boyd-Graber and Brianna Satinoff 248

Faster and Smaller N-Gram Language Models
Adam Pauls and Dan Klein 258

Learning to Win by Reading Manuals in a Monte-Carlo Framework
S.R.K Branavan, David Silver and Regina Barzilay 268

Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity
Tony Veale 278

Local Histograms of Character N-grams for Authorship Attribution
Hugo Jair Escalante, Thamar Solorio and Manuel Montes-y-Gomez 288

Word Maturity: Computational Modeling of Word Knowledge
Kirill Kireyev and Thomas K Landauer 299

Finding Deceptive Opinion Spam by Any Stretch of the Imagination
Myle Ott, Yejin Choi, Claire Cardie and Jeffrey T. Hancock 309

Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora
Bin Lu, Chenhao Tan, Claire Cardie and Benjamin K. Tsou 320

A Pilot Study of Opinion Summarization in Conversations
Dong Wang and Yang Liu 331

Contrasting Opposing Views of News Articles on Contentious Issues
Souneil Park, Kyung Soon Lee and Junehwa Song 340

Content Models with Attitude
Christina Sauper, Aria Haghighi and Regina Barzilay 350

Recognizing Named Entities in Tweets
Xiaohua LIU, Shaodian ZHANG, Furu WEI and Ming ZHOU 359

Lexical Normalisation of Short Text Messages: Makn Sens a #twitter
Bo Han and Timothy Baldwin 368

Topical Keyphrase Extraction from Twitter 379
Xin Zhao, Jing Jiang, Jing He, Yang Song, Palakorn Achanauparp, Ee-Peng Lim and Xiaoming Li

Event Discovery in Social Media Feeds
Edward Benson, Aria Haghighi and Regina Barzilay 389

How do you pronounce your name? Improving G2P with transliterations
Aditya Bhargava and Grzegorz Kondrak 399

Unsupervised Word Alignment with Arbitrary Features
Chris Dyer, Jonathan H. Clark, Alon Lavie and Noah A. Smith 409

Model-Based Aligner Combination Using Dual Decomposition
John DeNero and Klaus Macherey 420

An Algorithm for Unsupervised Transliteration Mining with an Application to Word Alignment
Hassan Sajjad, Alexander Fraser and Helmut Schmid 430

Beam-Width Prediction for Efficient Context-Free Parsing
Nathan Bodenstab, Aaron Dunlop, Keith Hall and Brian Roark 440

Optimal Head-Driven Parsing Complexity for Linear Context-Free Rewriting Systems
Pierluigi Crescenzi, Daniel Gildea, Andrea Marino, Gianluca Rossi and Giorgio Satta 450

Prefix Probability for Probabilistic Synchronous Context-Free Grammars
Mark-Jan Nederhof and Giorgio Satta 460

A Comparison of Loopy Belief Propagation and Dual Decomposition for Integrated CCG Supertagging and Parsing
Michael Auli and Adam Lopez 470

Jointly Learning to Extract and Compress
Taylor Berg-Kirkpatrick, Dan Gillick and Dan Klein 481

Discovery of Topically Coherent Sentences for Extractive Summarization
Asli Celikyilmaz and Dilek Hakkani-Tur 491

Coherent Citation-Based Summarization of Scientific Papers
Amjad Abu-Jbara and Dragomir Radev 500

A Class of Submodular Functions for Document Summarization
Hui Lin and Jeff Bilmes 510

Semi-supervised Relation Extraction with Large-scale Word Clustering
Ang Sun, Ralph Grishman and Satoshi Sekine 521

In-domain Relation Discovery with Meta-constraints via Posterior Regularization
Harr Chen, Edward Benson, Tahira Naseem and Regina Barzilay 530

Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations
Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer and Daniel S. Weld 541

Exploiting Syntactico-Semantic Structures for Relation Extraction
Yee Seng Chan and Dan Roth 551

Together We Can: Bilingual Bootstrapping for WSD
Mitesh M. Khapra, Salil Joshi, Arindam Chatterjee and Pushpak Bhattacharyya 561

Which Noun Phrases Denote Which Concepts?
Jayant Krishnamurthy and Tom Mitchell 570

Semantic Representation of Negation Using Focus Detection
Eduardo Blanco and Dan Moldovan 581

Learning Dependency-Based Compositional Semantics
Percy Liang, Michael Jordan and Dan Klein 590

Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections
Dipanjan Das and Slav Petrov 600

Global Learning of Typed Entailment Rules
Jonathan Berant, Ido Dagan and Jacob Goldberger 610

Incremental Syntactic Language Models for Phrase-based Translation
Lane Schwartz, Chris Callison-Burch, William Schuler and Stephen Wu 620

An Unsupervised Model for Joint Phrase Alignment and Extraction
Graham Neubig, Taro Watanabe, Eiichiro Sumita, Shinsuke Mori and Tatsuya Kawahara 632

Learning Hierarchical Translation Structure with Linguistic Annotations
Markos Mylonakis and Khalil Sima'an 642

Phrase-Based Translation Model for Question Retrieval in Community Question Answer Archives
Guangyou Zhou, Li Cai, Jun Zhao and Kang Liu 653

Neutralizing Linguistically Problematic Annotations in Unsupervised Dependency Parsing Evaluation
Roy Schwartz, Omri Abend, Roi Reichart and Ari Rappoport 663

Dynamic Programming Algorithms for Transition-Based Dependency Parsers
Marco Kuhlmann, Carlos Gómez-Rodríguez and Giorgio Satta 673

Shift-Reduce CCG Parsing
Yue Zhang and Stephen Clark 683

Web-Scale Features for Full-Scale Parsing
Mohit Bansal and Dan Klein 693

The impact of language models and loss functions on repair disfluency detection
Simon Zwarts and Mark Johnson 703

Learning Sub-Word Units for Open Vocabulary Speech Recognition
Carolina Parada, Mark Dredze, Abhinav Sethy and Ariya Rastrow 712

Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech
Miao Chen and Klaus Zechner 722

N-Best Rescoring Based on Pitch-accent Patterns
Je Hun Jeon, Wen Wang and Yang Liu 732

Lexically-Triggered Hidden Markov Models for Clinical Document Coding
Svetlana Kiritchenko and Colin Cherry 742

Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments
Michael Mohler, Razvan Bunescu and Rada Mihalcea 752

Age Prediction in Blogs: A Study of Style, Content, and Online Behavior in Pre- and Post-Social Media Generations
Sara Rosenthal and Kathleen McKeown 763

Extracting Social Power Relationships from Natural Language
Philip Bramsen, Martha Escobar-Molano, Ami Patel and Rafael Alonso 773

Bootstrapping coreference resolution using word associations
Hamidreza Kobdani, Hinrich Schuetze, Michael Schiehlen and Hans Kamp 783

Large-Scale Cross-Document Coreference Using Distributed Inference and Hierarchical Models
Sameer Singh, Amarnag Subramanya, Fernando Pereira and Andrew McCallum 793

A Cross-Lingual ILP Solution to Zero Anaphora Resolution
Ryu Iida and Massimo Poesio 804

Coreference Resolution with World Knowledge
Altaf Rahman and Vincent Ng 814

How to train your multi bottom-up tree transducer
Andreas Maletti 825

Binarized Forest to String Translation
Hao Zhang, Licheng Fang, Peng Xu and Xiaoyun Wu 835

Learning to Transform and Select Elementary Trees for Improved Syntax-based Machine Translations
Bing Zhao, Young-Suk Lee, Xiaoqiang Luo and Liu Li 846

Rule Markov Models for Fast Tree-to-String Translation
Ashish Vaswani, Haitao Mi, Liang Huang and David Chiang 856

A Hierarchical Pitman-Yor Process HMM for Unsupervised Part of Speech Induction
Phil Blunsom and Trevor Cohn 865

Using Deep Morphology to Improve Automatic Error Detection in Arabic Handwriting Recognition
Nizar Habash and Ryan Roth 875

A Discriminative Model for Joint Morphological Disambiguation and Dependency Parsing
John Lee, Jason Naradowsky and David A. Smith 885

Unsupervised Bilingual Morpheme Segmentation and Alignment with Context-rich Hidden Semi-Markov Models
Jason Naradowsky and Kristina Toutanova 895

A Graph Approach to Spelling Correction in Domain-Centric Search
Zhuowei Bao, Benny Kimelfeld and Yunyao Li 905

Grammatical Error Correction with Alternating Structure Optimization
Daniel Dahlmeier and Hwee Tou Ng 915

Algorithm Selection and Model Adaptation for ESL Correction Tasks
Alla Rozovskaya and Dan Roth 924

Automated Whole Sentence Grammar Correction Using a Noisy Channel Model
Y. Albert Park and Roger Levy 934

A Generative Entity-Mention Model for Linking Entities with Knowledge Base
Xianpei Han and Le Sun 945

Simple supervised document geolocation with geodesic grids
Benjamin Wing and Jason Baldridge 955

Piggyback: Using Search Engines for Robust Cross-Domain Named Entity Recognition
Stefan Rüd, Massimiliano Ciaramita, Jens Müller and Hinrich Schütze 965

Template-Based Information Extraction without the Templates
Nathanael Chambers and Dan Jurafsky 976

Classifying arguments by scheme
Vanessa Wei Feng and Graeme Hirst 987

Automatically Evaluating Text Coherence Using Discourse Relations
Ziheng Lin, Hwee Tou Ng and Min-Yen Kan 997

Underspecifying and Predicting Voice for Surface Realisation Ranking
Sina Zarrieß, Aoife Cahill and Jonas Kuhn 1007

Recognizing Authority in Dialogue with an Integer Linear Programming Constrained Model
Elijah Mayfield and Carolyn Penstein Rosé 1018

Reordering Metrics for MT
Alexandra Birch and Miles Osborne 1027

Reordering with Source Language Collocations
Zhanyi Liu, Haifeng Wang, Hua Wu, Ting Liu and Sheng Li 1036

A Joint Sequence Translation Model with Integrated Reordering
Nadir Durrani, Helmut Schmid and Alexander Fraser 1045

Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results
Roger Levy 1055

Metagrammar engineering: Towards systematic exploration of implemented grammars
Antske Fokkens 1066

Simple Unsupervised Grammar Induction from Raw Text with Cascaded Finite State Models
Elias Ponvert, Jason Baldridge and Katrin Erk 1077

Extracting Paraphrases from Definition Sentences on the Web 1087
Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jun'ichi Kazama and Sadao Kurohashi

Learning From Collective Human Behavior to Introduce Diversity in Lexical Choice
Vahed Qazvinian and Dragomir R. Radev 1098

Ordering Prenominal Modifiers with a Reranking Approach
Jenny Liu and Aria Haghighi 1109

Unsupervised Semantic Role Induction via Split-Merge Clustering
Joel Lang and Mirella Lapata 1117

Using Cross-Entity Inference to Improve Event Extraction
Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou and Qiaoming Zhu 1127

Peeling Back the Layers: Detecting Event Role Fillers in Secondary Contexts
Ruihong Huang and Ellen Riloff 1137

Knowledge Base Population: Successful Approaches and Challenges
Heng Ji and Ralph Grishman 1148

Nonlinear Evidence Fusion and Propagation for Hyponymy Relation Mining
Fan Zhang, Shuming Shi, Jing Liu, Shuqi Sun and Chin-Yew Lin 1159

A Pronoun Anaphora Resolution System based on Factorial Hidden Markov Models
Dingcheng Li, Tim Miller and William Schuler 1169

Disentangling Chat with Local Coherence Models
Micha Elsner and Eugene Charniak 1179

An Affect-Enriched Dialogue Act Classification Model for Task-Oriented Dialogue
Kristy Boyer, Joseph Grafsgaard, Eun Young Ha, Robert Phillips and James Lester 1190

Fine-Grained Class Label Markup of Search Queries
Joseph Reisinger and Marius Pasca 1200

Creating a manually error-tagged and shallow-parsed learner corpus
Ryo Nagata, Edward Whittaker and Vera Sheinman 1210

Crowdsourcing Translation: Professional Quality from Non-Professionals
Omar F. Zaidan and Chris Callison-Burch 1220

A Statistical Tree Annotator and Its Applications
Xiaoqiang Luo and Bing Zhao 1230

Consistent Translation using Discriminative Learning - A Translation Memory-inspired Approach
Yanjun Ma, Yifan He, Andy Way and Josef van Genabith 1239

Machine Translation System Combination by Confusion Forest
Taro Watanabe and Eiichiro Sumita 1249

Hypothesis Mixture Decoding for Statistical Machine Translation
Nan Duan, Mu Li and Ming Zhou 1258

Minimum Bayes-risk System Combination
Jesús González-Rubio, Alfons Juan and Francisco Casacuberta 1268

Adjoining Tree-to-String Translation
Yang Liu, Qun Liu and Yajuan Lü 1278

Enhancing Language Models in Statistical Machine Translation with Backward N-grams and Mutual Information Triggers
Deyi Xiong, Min Zhang and Haizhou Li 1288

Translating from Morphologically Complex Languages: A Paraphrase-Based Approach
Preslav Nakov and Hwee Tou Ng 1298

Gappy Phrasal Alignment By Agreement
Mohit Bansal, Chris Quirk and Robert Moore 1308

Translationese and Its Dialects
Moshe Koppel and Noam Ordan 1318

Rare Word Translation Extraction from Aligned Comparable Documents
Emmanuel Prochasson and Pascale Fung 1327

Using Bilingual Parallel Corpora for Cross-Lingual Textual Entailment
Yashar Mehdad, Matteo Negri and Marcello Federico 1336

Using Large Monolingual and Bilingual Corpora to Improve Coordination Disambiguation
Shane Bergsma, David Yarowsky and Kenneth Church 1346

Unsupervised Decomposition of a Document into Authorial Components
Moshe Koppel, Navot Akiva, Idan Dershowitz and Nachum Dershowitz 1356

Discovering Sociolinguistic Associations with Structured Sparsity
Jacob Eisenstein, Noah A. Smith and Eric P. Xing 1365

Local and Global Algorithms for Disambiguation to Wikipedia
Lev Ratinov, Dan Roth, Doug Downey and Mike Anderson 1375

A Stacked Sub-Word Model for Joint Chinese Word Segmentation and Part-of-Speech Tagging
Weiwei Sun 1385

Language-independent compound splitting with morphological operations
Klaus Macherey, Andrew Dai, David Talbot, Ashok Popat and Franz Och 1395

Parsing the Internal Structure of Words: A New Paradigm for Chinese Word Segmentation
Zhongguo Li 1405

A Simple Measure to Assess Non-response
Anselmo Peñas and Alvaro Rodrigo 1415

Improving Question Recommendation by Exploiting Information Need
Shuguang Li and Suresh Manandhar 1425

Semi-Supervised Frame-Semantic Parsing for Unknown Predicates
Dipanjan Das and Noah A. Smith 1435

A Bayesian Model for Unsupervised Semantic Parsing
Ivan Titov and Alexandre Klementiev 1445

Unsupervised Learning of Semantic Relation Composition
Eduardo Blanco and Dan Moldovan 1456

Unsupervised Discovery of Domain-Specific Knowledge from Text
Dirk Hovy, Chunliang Zhang, Eduard Hovy and Anselmo Peñas 1466

Latent Semantic Word Sense Induction and Disambiguation
Tim Van de Cruys and Marianna Apidianaki 1476

Confidence Driven Unsupervised Semantic Parsing
Dan Goldwasser, Roi Reichart, James Clarke and Dan Roth 1486

Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews
Jianxing Yu, Zheng-Jun Zha, Meng Wang and Tat-Seng Chua 1496

Collective Classification of Congressional Floor-Debate Transcripts
Clinton Burfoot, Steven Bird and Timothy Baldwin 1506

Integrating history-length interpolation and classes in language modeling
Hinrich Schütze 1516

Structural Topic Model for Latent Topical Structure Analysis
Hongning Wang, Duo Zhang and ChengXiang Zhai 1526

Automatic Labelling of Topic Models
Jey Han Lau, Karl Grieser, David Newman and Timothy Baldwin 1536

Using Bilingual Information for Cross-Language Document Summarization
Xiaojun Wan 1546

Exploiting Web-Derived Selectional Preference to Improve Statistical Dependency Parsing
Guangyou Zhou, Jun Zhao, Kang Liu and Li Cai 1556

Effective Measures of Domain Similarity for Parsing
Barbara Plank and Gertjan van Noord 1566

Efficient CCG Parsing: A versus Adaptive Supertagging*
Michael Auli and Adam Lopez 1577

Improving Arabic Dependency Parsing with Form-based and Functional Morphological Features
Yuval Marton, Nizar Habash and Owen Rambow 1586

Partial Parsing from Bitext Projections
Prashanth Mannem and Aswarth Dara 1597

Ranking Class Labels Using Query Sessions
Marius Pasca 1607

Insights from Network Structure for Text Mining
Zornitsa Kozareva and Eduard Hovy 1616

Event Extraction as Dependency Parsing
David McClosky, Mihai Surdeanu and Christopher Manning 1626

Extracting Comparative Entities and Predicates from Texts Using Comparative Type Classification
Seon Yang and Youngjoong Ko 1636