

**53rd Annual Meeting of the Association for
Computational Linguistics and 7th
International Joint Conference on Natural
Language Processing of the Asian
Federation of Natural Language Processing
(ACL-IJCNLP 2015)**

Short Papers

**Beijing, China
26 – 31 July 2015**

ISBN: 978-1-5108-0846-1

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© (2015) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2015)

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

<i>A Framework for the Construction of Monolingual and Cross-lingual Word Similarity Datasets</i> José Camacho-Collados, Mohammad Taher Pilehvar and Roberto Navigli	1
<i>On metric embedding for boosting semantic similarity computations</i> Julien Subercaze, Christophe Gravier and Frédérique Laforest	8
<i>Improving Distributed Representation of Word Sense via WordNet Gloss Composition and Context Clustering</i> Tao Chen, Ruifeng Xu, Yulan He and Xuan Wang	15
<i>A Multitask Objective to Inject Lexical Contrast into Distributional Semantics</i> Nghia The Pham, Angeliki Lazaridou and Marco Baroni	21
<i>Semi-Stacking for Semi-supervised Sentiment Classification</i> Shoushan Li, Lei Huang, Jingjing Wang and Guodong Zhou	27
<i>Deep Markov Neural Network for Sequential Data Classification</i> Min Yang	32
<i>Semantic Analysis and Helpfulness Prediction of Text for Online Product Reviews</i> Yinfei Yang, Yaowei Yan, Minghui Qiu and Forrest Bao	38
<i>Document Classification by Inversion of Distributed Language Representations</i> Matt Taddy	45
<i>Using Tweets to Help Sentence Compression for News Highlights Generation</i> Zhongyu Wei, Yang Liu, Chen Li and Wei Gao	50
<i>Domain-Specific Paraphrase Extraction</i> Ellie Pavlick, Juri Ganitkevitch, Tsz Ping Chan, Xuchen Yao, Benjamin Van Durme and Chris Callison-Burch	57
<i>Simplifying Lexical Simplification: Do We Need Simplified Corpora?</i> Goran Glavaš and Sanja Štajner	63
<i>Zoom: a corpus of natural language descriptions of map locations</i> Romina Altamirano, Thiago Ferreira, Ivandré Paraboni and Luciana Benotti	69
<i>Generating overspecified referring expressions: the role of discrimination</i> Ivandré Paraboni, Michelle Galindo and Douglas Iacovelli	76
<i>Using prosodic annotations to improve coreference resolution of spoken text</i> Ina Roesiger and Arndt Riester	83
<i>Spectral Semi-Supervised Discourse Relation Classification</i> Robert Fisher and Reid Simmons	89
<i>I do not disagree: leveraging monolingual alignment to detect disagreement in dialogue</i> Ajda Gokcen and Marie-Catherine de Marneffe	94
<i>Language Models for Image Captioning: The Quirks and What Works</i> Jacob Devlin, Hao Cheng, Hao Fang, Saurabh Gupta, Li Deng, Xiaodong He, Geoffrey Zweig and Margaret Mitchell	100

<i>A Distributed Representation Based Query Expansion Approach for Image Captioning</i> Semih Yagcioglu, Erkut Erdem, Aykut Erdem and Ruket Cakici	106
<i>Learning language through pictures</i> Grzegorz Chrupała, Ákos Kádár and Afra Alishahi	112
<i>Exploiting Image Generality for Lexical Entailment Detection</i> Douwe Kiela, Laura Rimell, Ivan Vulić and Stephen Clark	119
<i>Lexicon Stratification for Translating Out-of-Vocabulary Words</i> Yulia Tsvetkov and Chris Dyer	125
<i>Recurrent Neural Network based Rule Sequence Model for Statistical Machine Translation</i> Heng Yu and Xuan Zhu	132
<i>Discriminative Preordering Meets Kendall's τ Maximization</i> Sho Hoshino, Yusuke Miyao, Katsuhito Sudoh, Katsuhiko Hayashi and Masaaki Nagata	139
<i>Evaluating Machine Translation Systems with Second Language Proficiency Tests</i> Takuya Matsuzaki, Akira Fujita, Naoya Todo and Noriko H. Arai	145
<i>Representation Based Translation Evaluation Metrics</i> Boxing Chen and Hongyu Guo	150
<i>Exploring the Planet of the APes: a Comparative Study of State-of-the-art Methods for MT Automatic Post-Editing</i> Rajen Chatterjee, Marion Weller, Matteo Negri and Marco Turchi	156
<i>Efficient Learning for Undirected Topic Models</i> Jiatao Gu and Victor O.K. Li	162
<i>A Hassle-Free Unsupervised Domain Adaptation Method Using Instance Similarity Features</i> Jianfei Yu and Jing Jiang	168
<i>Dependency-based Convolutional Neural Networks for Sentence Embedding</i> Mingbo Ma, Liang Huang, Bowen Zhou and Bing Xiang	174
<i>Non-Linear Text Regression with a Deep Convolutional Neural Network</i> Zsolt Bitvai and Trevor Cohn	180
<i>A Unified Learning Framework of Skip-Grams and Global Vectors</i> Jun Suzuki and Masaaki Nagata	186
<i>Pre-training of Hidden-Unit CRFs</i> Young-Bum Kim, Karl Stratos and Ruhi Sarikaya	192
<i>Distributional Neural Networks for Automatic Resolution of Crossword Puzzles</i> Aliaksei Severyn, Massimo Nicosia, Gianni Barlacchi and Alessandro Moschitti	199
<i>Word Order Typology through Multilingual Word Alignment</i> Robert Östling	205
<i>Measuring idiosyncratic interests in children with autism</i> Masoud Rouhizadeh, Emily Prud'hommeaux, Jan van Santen and Richard Sproat	212

<i>Frame-Semantic Role Labeling with Heterogeneous Annotations</i>	
Meghana Kshirsagar, Sam Thomson, Nathan Schneider, Jaime Carbonell, Noah A. Smith and Chris Dyer	218
<i>Semantic Interpretation of Superlative Expressions via Structured Knowledge Bases</i>	
Sheng Zhang, Yansong Feng, Songfang Huang, Kun Xu, Zhe Han and Dongyan Zhao	225
<i>Grounding Semantics in Olfactory Perception</i>	
Douwe Kiela, Luana Bulat and Stephen Clark	231
<i>Word-based Japanese typed dependency parsing with grammatical function analysis</i>	
Takaaki Tanaka and Masaaki Nagata	237
<i>KLcpo3 - a Language Similarity Measure for Delexicalized Parser Transfer</i>	
Rudolf Rosa and Zdenek Zabokrtsky	243
<i>CCG Supertagging with a Recurrent Neural Network</i>	
Wenduan Xu, Michael Auli and Stephen Clark	250
<i>An Efficient Dynamic Oracle for Unrestricted Non-Projective Parsing</i>	
Carlos Gómez-Rodríguez and Daniel Fernández-González	256
<i>Synthetic Word Parsing Improves Chinese Word Segmentation</i>	
Fei Cheng, Kevin Duh and Yuji Matsumoto	262
<i>If all you have is a bit of the Bible: Learning POS taggers for truly low-resource languages</i>	
Željko Agić, Dirk Hovy and Anders Søgaard	268
<i>Improving distant supervision using inference learning</i>	
Roland Roller, Eneko Agirre, Aitor Soroa and Mark Stevenson	273
<i>A Lexicalized Tree Kernel for Open Information Extraction</i>	
Ying Xu, Christoph Ringlstetter, Mi-young Kim, Grzegorz Kondrak, Randy Goebel and Yusuke Miyao	279
<i>A Dependency-Based Neural Network for Relation Classification</i>	
Yang Liu, Furu Wei, Sujian Li, Heng Ji, Ming Zhou and Houfeng WANG	285
<i>Embedding Methods for Fine Grained Entity Type Classification</i>	
Dani Yogatama, Daniel Gillick and Nevena Lazic	291
<i>Sieve-Based Entity Linking for the Biomedical Domain</i>	
Jennifer D’Souza and Vincent Ng	297
<i>Open IE as an Intermediate Structure for Semantic Tasks</i>	
Gabriel Stanovsky and Ido Dagan	303
<i>Recovering dropped pronouns from Chinese text messages</i>	
Yaqin Yang, Yalin Liu and Nianwen Xue	309
<i>The Users Who Say ‘Ni’: Audience Identification in Chinese-language Restaurant Reviews</i>	
Rob Voigt and Dan Jurafsky	314
<i>Chinese Zero Pronoun Resolution: A Joint Unsupervised Discourse-Aware Model Rivaling State-of-the-Art Resolvers</i>	
Chen Chen and Vincent Ng	320

<i>Co-Simmate: Quick Retrieving All Pairwise Co-Simrank Scores</i>	
Yu Weiren and Julie McCann	327
<i>Retrieval of Research-level Mathematical Information Needs: A Test Collection and Technical Terminology Experiment</i>	
Yiannos Stathopoulos and Simone Teufel	334
<i>Learning to Mine Query Subtopics from Query Log</i>	
Zhenzhong Zhang, Le Sun and Xianpei Han	341
<i>Learning Topic Hierarchies for Wikipedia Categories</i>	
Linmei Hu, Xuzhong Wang, Mengdi Zhang, Juanzi Li, Xiaoli Li, Chao Shao, Jie Tang and Yongbin Liu	346
<i>Semantic Clustering and Convolutional Neural Network for Short Text Categorization</i>	
Peng Wang, Jiaming Xu, Bo Xu, Chenglin Liu, Heng Zhang, Fangyuan Wang and Hongwei Hao	352
<i>Document Level Time-anchoring for TimeLine Extraction</i>	
Egoitz Laparra, Itziar Aldabe and German Rigau	358
<i>Event Detection and Domain Adaptation with Convolutional Neural Networks</i>	
Thien Huu Nguyen and Ralph Grishman	365
<i>Seed-Based Event Trigger Labeling: How far can event descriptions get us?</i>	
Ofer Bronstein, Ido Dagan, Qi Li, Heng Ji and Anette Frank	372
<i>An Empirical Study of Chinese Name Matching and Applications</i>	
Nanyun Peng, Mo Yu and Mark Dredze	377
<i>Language Identification and Modeling in Specialized Hardware</i>	
Kenneth Heafield, Rohan Kshirsagar and Santiago Barona	384
<i>Cross-lingual Transfer of Named Entity Recognizers without Parallel Corpora</i>	
Ayah Zirikly	390
<i>Robust Multi-Relational Clustering via ℓ_1-Norm Symmetric Nonnegative Matrix Factorization</i>	
Kai Liu and Hua Wang	397
<i>Painless Labeling with Application to Text Mining</i>	
Sajib Dasgupta	402
<i>FrameNet+: Fast Paraphrastic Tripling of FrameNet</i>	
Ellie Pavlick, Travis Wolfe, Pushpendre Rastogi, Chris Callison-Burch, Mark Dredze and Benjamin Van Durme	408
<i>IWNLP: Inverse Wiktionary for Natural Language Processing</i>	
Matthias Liebeck and Stefan Conrad	414
<i>TR9856: A Multi-word Term Relatedness Benchmark</i>	
Ran Levy, Liat Ein-Dor, Shay Hummel, Ruty Rinott and Noam Slonim	419
<i>PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification</i>	
Ellie Pavlick, Pushpendre Rastogi, Juri Ganitkevitch, Benjamin Van Durme and Chris Callison-Burch	425

<i>Automatic Discrimination between Cognates and Borrowings</i>	
Alina Maria Ciobanu and Liviu P. Dinu	431
<i>The Media Frames Corpus: Annotations of Frames Across Issues</i>	
Dallas Card, Amber E. Boydston, Justin H. Gross, Philip Resnik and Noah A. Smith	438
<i>deltaBLEU: A Discriminative Metric for Generation Tasks with Intrinsically Diverse Targets</i>	
Michel Galley, Chris Brockett, Alessandro Sordoni, Yangfeng Ji, Michael Auli, Chris Quirk, Margaret Mitchell, Jianfeng Gao and Bill Dolan	445
<i>Tibetan Unknown Word Identification from News Corpora for Supporting Lexicon-based Tibetan Word Segmentation</i>	
Minghua Nuo, Huidan Liu, Congjun Long and Jian Wu	451
<i>Learning Lexical Embeddings with Syntactic and Lexicographic Knowledge</i>	
Tong Wang, Abdelrahman Mohamed and Graeme Hirst	458
<i>Non-distributional Word Vector Representations</i>	
Manaal Faruqui and Chris Dyer	464
<i>Early and Late Combinations of Criteria for Reranking Distributional Thesauri</i>	
Olivier Ferret	470
<i>Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases</i>	
Kristina Gulordava, Paola Merlo and Benoit Crabbé	477
<i>Tagging Performance Correlates with Author Age</i>	
Dirk Hovy and Anders Søgaard	483
<i>User Based Aggregation for Biterm Topic Model</i>	
Weizheng Chen, Jinpeng Wang, Yan Zhang, Hongfei Yan and Xiaoming Li	489
<i>The Fixed-Size Ordinally-Forgetting Encoding Method for Neural Network Language Models</i>	
ShiLiang Zhang, Hui Jiang, MingBin Xu, JunFeng Hou and LiRong Dai	495
<i>Unsupervised Decomposition of a Multi-Author Document Based on Naive-Bayesian Model</i>	
Khaled Aldebei, Xiangjian He and Jie Yang	501
<i>Extended Topic Model for Word Dependency</i>	
Tong Wang, Vish Viswanath and Ping Chen	506
<i>Dependency Recurrent Neural Language Models for Sentence Completion</i>	
Piotr Mirowski and Andreas Vlachos	511
<i>Point Process Modelling of Rumour Dynamics in Social Media</i>	
Michal Lukasik, Trevor Cohn and Kalina Bontcheva	518
<i>Learning Hidden Markov Models with Distributed State Representations for Domain Adaptation</i>	
Min Xiao and Yuhong Guo	524
<i>MT Quality Estimation for Computer-assisted Translation: Does it Really Help?</i>	
Marco Turchi, Matteo Negri and Marcello Federico	530
<i>Context-Dependent Translation Selection Using Convolutional Neural Network</i>	
Baotian Hu, Zhaopeng Tu, Zhengdong Lu, Hang Li and Qingcai Chen	536

<i>Learning Word Reorderings for Hierarchical Phrase-based Statistical Machine Translation</i> jingyi zhang, Masao Utiyama, Eiichiro Sumita and Hai Zhao	542
<i>UNRAVEL—A Decipherment Toolkit</i> Malte Nuhn, Julian Schamper and Hermann Ney	549
<i>Multi-Pass Decoding With Complex Feature Guidance for Statistical Machine Translation</i> Benjamin Marie and Aurélien Max	554
<i>What’s in a Domain? Analyzing Genre and Topic Differences in Statistical Machine Translation</i> Marlies van der Wees, Arianna Bisazza, Wouter Weerkamp and Christof Monz	560
<i>Learning Cross-lingual Word Embeddings via Matrix Co-factorization</i> Tianze Shi, Zhiyuan Liu, Yang Liu and Maosong Sun	567
<i>Improving Pivot Translation by Remembering the Pivot</i> Akiva Miura, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura	573
<i>BrailleSUM: A News Summarization System for the Blind and Visually Impaired People</i> Xiaojun Wan and Yue Hu	578
<i>Automatic Identification of Age-Appropriate Ratings of Song Lyrics</i> Anggi Maulidyani and Ruli Manurung	583
<i>Ground Truth for Grammaticality Correction Metrics</i> Courtney Napoles, Keisuke Sakaguchi, Matt Post and Joel Tetreault	588
<i>Radical Embedding: Delving Deeper to Chinese Radicals</i> Xinlei Shi, Junjie Zhai, Xudong Yang, Zehua Xie and Chao Liu	594
<i>Automatic Detection of Sentence Fragments</i> Chak Yan Yeung and John Lee	599
<i>A Computational Approach to Automatic Prediction of Drunk-Texting</i> Aditya Joshi, Abhijit Mishra, Balamurali AR, Pushpak Bhattacharyya and Mark J Carman	604
<i>Reducing infrequent-token perplexity via variational corpora</i> Yusheng Xie, Pranjal Daga, Yu Cheng, Kunpeng Zhang, Ankit Agrawal and Alok Choudhary .	609
<i>A Hierarchical Knowledge Representation for Expert Finding on Social Media</i> Yanran Li, Wenjie Li and Sujian Li	616
<i>Tackling Sparsity, the Achilles Heel of Social Networks: Language Model Smoothing via Social Regularization</i> Rui Yan, Xiang Li, Mengwen Liu and Xiaohua Hu	623
<i>Twitter User Geolocation Using a Unified Text and Network Prediction Model</i> Afshin Rahimi, Trevor Cohn and Timothy Baldwin	630
<i>Automatic Keyword Extraction on Twitter</i> Luis Marujo, Wang Ling, Isabel Trancoso, Chris Dyer, Alan W Black, Anatole Gershman, David Martins de Matos, João Neto and Jaime Carbonell	637
<i>Towards a Contextual Pragmatic Model to Detect Irony in Tweets</i> Jihen Karoui, Benamara Farah, Véronique MORICEAU, Nathalie Aussenac-Gilles and Lamia Hadrich-Belguith	644

<i>Annotation and Classification of an Email Importance Corpus</i> Fan Zhang and Kui Xu	651
<i>Lexical Comparison Between Wikipedia and Twitter Corpora by Using Word Embeddings</i> Luchen Tan, Haotian Zhang, Charles Clarke and Mark Smucker	657
<i>The Discovery of Natural Typing Annotations: User-produced Potential Chinese Word Delimiters</i> Dakui Zhang, Yu Mao, Yang Liu, Hanshi Wang, Chuyuan Wei and Shiping Tang	662
<i>One Tense per Scene: Predicting Tense in Chinese Conversations</i> Tao Ge, Heng Ji, Baobao Chang and Zhifang Sui	668
<i>A Language-Independent Feature Schema for Inflectional Morphology</i> John Sylak-Glassman, Christo Kirov, David Yarowsky and Roger Que	674
<i>Rhetoric Map of an Answer to Compound Queries</i> Boris Galitsky, Dmitry Ilvovsky and Sergey O. Kuznetsov	681
<i>Thread-Level Information for Comment Classification in Community Question Answering</i> Alberto Barrón-Cedeño, Simone Filice, Giovanni Da San Martino, Shafiq Joty, Lluís Màrquez, Preslav Nakov and Alessandro Moschitti	687
<i>Learning Hybrid Representations to Retrieve Semantically Equivalent Questions</i> Cicero dos Santos, Luciano Barbosa, Dasha Bogdanova and Bianca Zadrozny	694
<i>Machine Comprehension with Syntax, Frames, and Semantics</i> Hai Wang, Mohit Bansal, Kevin Gimpel and David McAllester	700
<i>A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering</i> Di Wang and Eric Nyberg	707
<i>Answer Sequence Learning with Neural Networks for Answer Selection in Community Question Answering</i> Xiaoqiang Zhou, Baotian Hu, Qingcai Chen, Buzhou Tang and Xiaolong Wang	713
<i>Bilingual Word Embeddings from Non-Parallel Document-Aligned Data Applied to Bilingual Lexicon Induction</i> Ivan Vulić and Marie-Francine Moens	719
<i>How Well Do Distributional Models Capture Different Types of Semantic Knowledge?</i> Dana Rubinstein, Effi Levi, Roy Schwartz and Ari Rappoport	726
<i>Low-Rank Tensors for Verbs in Compositional Distributional Semantics</i> Daniel Fried, Tamara Polajnar and Stephen Clark	731
<i>Constrained Semantic Forests for Improved Discriminative Semantic Parsing</i> Wei Lu	737
<i>Automatic Identification of Rhetorical Questions</i> Shohini Bhattachali, Jeremy Cytryn, Elana Feldman and Joonsuk Park	743
<i>Lifelong Learning for Sentiment Classification</i> Zhiyuan Chen, Nianzu Ma and Bing Liu	750
<i>Harnessing Context Incongruity for Sarcasm Detection</i> Aditya Joshi, Vinita Sharma and Pushpak Bhattacharyya	757

<i>Emotion Detection in Code-switching Texts via Bilingual and Sentimental Information</i>	
Zhongqing Wang, Sophia Lee, Shoushan Li and Guodong Zhou	763
<i>Model Adaptation for Personalized Opinion Analysis</i>	
Mohammad Al Boni, Keira Zhou, Hongning Wang and Matthew S. Gerber	769
<i>Linguistic Template Extraction for Recognizing Reader-Emotion and Emotional Resonance Writing Assistance</i>	
Yung-Chun Chang, Cen-Chieh Chen, Yu-lun Hsieh, Chien Chin Chen and Wen-Lian Hsu	775
<i>Aspect-Level Cross-lingual Sentiment Classification with Constrained SMT</i>	
Patrik Lambert	781
<i>Predicting Valence-Arousal Ratings of Words Using a Weighted Graph Method</i>	
Liang-Chih Yu, Jin Wang, K. Robert Lai and Xue-jie Zhang	788
<i>Multi-domain Dialog State Tracking using Recurrent Neural Networks</i>	
Nikola Mrkšić, Diarmuid Ó Séaghdha, Blaise Thomson, Milica Gasic, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen and Steve Young	794
<i>Dialogue Management based on Sentence Clustering</i>	
Wendong Ge and Bo Xu	800
<i>Compact Lexicon Selection with Spectral Methods</i>	
Young-Bum Kim, Karl Stratos, Xiaohu Liu and Ruhi Sarikaya	806
<i>The Impact of Listener Gaze on Predicting Reference Resolution</i>	
Nikolina Koleva, Martin Villalba, Maria Staudte and Alexander Koller	812
<i>A Simultaneous Recognition Framework for the Spoken Language Understanding Module of Intelligent Personal Assistant Software on Smart Phones</i>	
Changsu Lee, Youngjoong Ko and Jungyun Seo	818
<i>A Deeper Exploration of the Standard PB-SMT Approach to Text Simplification and its Evaluation</i>	
Sanja Štajner, Hannah Bechara and Horacio Saggion	823
<i>Learning Summary Prior Representation for Extractive Summarization</i>	
Ziqiang Cao, Furu Wei, Sujian Li, Wenjie Li, Ming Zhou and Houfeng WANG	829
<i>A Methodology for Evaluating Timeline Generation Algorithms based on Deep Semantic Units</i>	
Sandro Bauer and Simone Teufel	834
<i>Unsupervised extractive summarization via coverage maximization with syntactic and semantic concepts</i>	
Natalie Schluter and Anders Søgaard	840
<i>Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser</i>	
Long Duong, Trevor Cohn, Steven Bird and Paul Cook	845
<i>Semantic Structure Analysis of Noun Phrases using Abstract Meaning Representation</i>	
Yuichiro Sawai, Hiroyuki Shindo and Yuji Matsumoto	851
<i>Boosting Transition-based AMR Parsing with Refined Actions and Auxiliary Analyzers</i>	
Chuan Wang, Nianwen Xue and Sameer Pradhan	857
<i>Generative Incremental Dependency Parsing with Neural Networks</i>	
Jan Buys and Phil Blunsom	863

<i>Labeled Grammar Induction with Minimal Supervision</i>	
Yonatan Bisk, Christos Christodoulopoulos and Julia Hockenmaier	870
<i>On the Importance of Ezafe Construction in Persian Parsing</i>	
Alireza Nourian, Mohammad Sadegh Rasooli, Mohsen Imany and Heshaam Faili	877