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
# Table of Contents

**A Framework for the Construction of Monolingual and Cross-lingual Word Similarity Datasets**
José Camacho-Collados, Mohammad Taher Pilehvar and Roberto Navigli .......................... 1

**On metric embedding for boosting semantic similarity computations**
Julien Subercaze, Christophe Gravier and Frédérique Laforest ................................. 8

**Improving Distributed Representation of Word Sense via WordNet Gloss Composition and Context Clustering**
Tao Chen, Ruifeng Xu, Yulan He and Xuan Wang ..................................................... 15

**A Multitask Objective to Inject Lexical Contrast into Distributional Semantics**
Ngia The Pham, Angeliki Lazaridou and Marco Baroni ............................................. 21

**Semi-Stacking for Semi-supervised Sentiment Classification**
Shoushan Li, Lei Huang, Jingjing Wang and Guodong Zhou ..................................... 27

**Deep Markov Neural Network for Sequential Data Classification**
Min Yang ..................................................................................................................... 32

**Semantic Analysis and Helpfulness Prediction of Text for Online Product Reviews**
Yinfei Yang, Yaowei Yan, Minghui Qiu and Forrest Bao ............................................ 38

**Document Classification by Inversion of Distributed Language Representations**
Matt Taddy .................................................................................................................. 45

**Using Tweets to Help Sentence Compression for News Highlights Generation**
Zhongyu Wei, Yang Liu, Chen Li and Wei Gao ......................................................... 50

**Domain-Specific Paraphrase Extraction**
Ellie Pavlick, Juri Ganitkevitch, Tsz Ping Chan, Xuchen Yao, Benjamin Van Durme and Chris Callison-Burch .................................................................................. 57

**Simplifying Lexical Simplification: Do We Need Simplified Corpora?**
Goran Glavaš and Sanja Štajner ................................................................................. 63

**Zoom: a corpus of natural language descriptions of map locations**
Romina Altamirano, Thiago Ferreira, Ivandré Paraboni and Luciana Benotti ............. 69

**Generating overspecified referring expressions: the role of discrimination**
Ivandré Paraboni, Michelle Galindo and Douglas Iacobelli ..................................... 76

**Using prosodic annotations to improve coreference resolution of spoken text**
Ina Roesiger and Arndt Riester .................................................................................. 83

**Spectral Semi-Supervised Discourse Relation Classification**
Robert Fisher and Reid Simmons ............................................................................... 89

**I do not disagree: leveraging monolingual alignment to detect disagreement in dialogue**
Ajda Gokcen and Marie-Catherine de Marneffe ....................................................... 94

**Language Models for Image Captioning: The Quirks and What Works**
Jacob Devlin, Hao Cheng, Hao Fang, Saurabh Gupta, Li Deng, Xiaodong He, Geoffrey Zweig and Margaret Mitchell ................................................................. 100
A Distributed Representation Based Query Expansion Approach for Image Captioning
Semih Yagcioglu, Erkut Erdem, Aykut Erdem and Ruket Cakici .......................... 106

Learning language through pictures
Grzegorz Chrupała, Ákos Kádár and Afra Alishahi ........................................... 112

Exploiting Image Generality for Lexical Entailment Detection
Douwe Kiela, Laura Rimell, Ivan Vulić and Stephen Clark ................................. 119

Lexicon Stratification for Translating Out-of-Vocabulary Words
Yulia Tsvetkov and Chris Dyer ............................................................................ 125

Recurrent Neural Network based Rule Sequence Model for Statistical Machine Translation
Heng Yu and Xuan Zhu ....................................................................................... 132

Discriminative Preordering Meets Kendall’s $\tau$ Maximization
Sho Hoshino, Yusuke Miyao, Katsuhito Sudoh, Katsuhiko Hayashi and Masaaki Nagata .... 139

Evaluating Machine Translation Systems with Second Language Proficiency Tests
Takuya Matsuzaki, Akira Fujita, Naoya Todo and Noriko H. Arai .......................... 145

Representation Based Translation Evaluation Metrics
Boxing Chen and Hongyu Guo ........................................................................... 150

Exploring the Planet of the APEs: a Comparative Study of State-of-the-art Methods for MT Automatic Post-Editing
Rajen Chatterjee, Marion Weller, Matteo Negri and Marco Turchi ......................... 156

Efficient Learning for Undirected Topic Models
Jiatao Gu and Victor O.K. Li .............................................................................. 162

A Hassle-Free Unsupervised Domain Adaptation Method Using Instance Similarity Features
Jianfei Yu and Jing Jiang ................................................................................... 168

Dependency-based Convolutional Neural Networks for Sentence Embedding
Mingbo Ma, Liang Huang, Bowen Zhou and Bing Xiang ..................................... 174

Non-Linear Text Regression with a Deep Convolutional Neural Network
Zsolt Bitvai and Trevor Cohn ............................................................................ 180

A Unified Learning Framework of Skip-Grams and Global Vectors
Jun Suzuki and Masaaki Nagata ....................................................................... 186

Pre-training of Hidden-Unit CRFs
Young-Bum Kim, Karl Stratos and Ruhi Sarikaya ............................................. 192

Distributional Neural Networks for Automatic Resolution of Crossword Puzzles
Aliaksei Severyn, Massimo Nicosia, Gianni Barlacchi and Alessandro Moschitti ....... 199

Word Order Typology through Multilingual Word Alignment
Robert Östling .................................................................................................... 205

Measuring idiosyncratic interests in children with autism
Masoud Rouhizadeh, Emily Prud’hommeaux, Jan van Santen and Richard Sproat .... 212
Frame-Semantic Role Labeling with Heterogeneous Annotations  
Meghana Kshirsagar, Sam Thomson, Nathan Schneider, Jaime Carbonell, Noah A. Smith and Chris Dyer ................................................................. 218

Semantic Interpretation of Superlative Expressions via Structured Knowledge Bases  
Sheng Zhang, Yansong Feng, Songfang Huang, Kun Xu, Zhe Han and Dongyan Zhao .... 225

Grounding Semantics in Olfactory Perception  
Douwe Kiela, Luana Bulat and Stephen Clark ......................................................... 231

Word-based Japanese typed dependency parsing with grammatical function analysis  
Takaaki Tanaka and Masaaki Nagata ................................................................. 237

KLcpos3 - a Language Similarity Measure for Delexicalized Parser Transfer  
Rudolf Rosa and Zdenek Zabokrtsky ................................................................. 243

CCG Supertagging with a Recurrent Neural Network  
Wenduan Xu, Michael Auli and Stephen Clark ......................................................... 250

An Efficient Dynamic Oracle for Unrestricted Non-Projective Parsing  
Carlos Gómez-Rodríguez and Daniel Fernández-González ........................................ 256

Synthetic Word Parsing Improves Chinese Word Segmentation  
Fei Cheng, Kevin Duh and Yuji Matsumoto .............................................................. 262

If all you have is a bit of the Bible: Learning POS taggers for truly low-resource languages  
Željko Agić, Dirk Hovy and Anders Søgaard ............................................................... 268

Improving distant supervision using inference learning  
Roland Roller, Eneko Agirre, Aitor Soroa and Mark Stevenson ................................... 273

A Lexicalized Tree Kernel for Open Information Extraction  
Ying Xu, Christoph Ringlstetter, Mi-young Kim, Grzegorz Kondrak, Randy Goebel and Yusuke Miyao ........................................................................................................... 279

A Dependency-Based Neural Network for Relation Classification  
Yang Liu, Furu Wei, Sujian Li, Heng Ji, Ming Zhou and Houfeng WANG ...................... 285

Embedding Methods for Fine Grained Entity Type Classification  
Dani Yogatama, Daniel Gillick and Nevena Lazic .................................................... 291

Sieve-Based Entity Linking for the Biomedical Domain  
Jennifer D’Souza and Vincent Ng .............................................................................. 297

Open IE as an Intermediate Structure for Semantic Tasks  
Gabriel Stanovsky and Ido Dagan ............................................................................. 303

Recovering dropped pronouns from Chinese text messages  
Yaqin Yang, Yalin Liu and Nianwen Xue ................................................................. 309

The Users Who Say ‘Ni’: Audience Identification in Chinese-language Restaurant Reviews  
Rob Voigt and Dan Jurafsky ..................................................................................... 314

Chen Chen and Vincent Ng ..................................................................................... 320
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Discrimination between Cognates and Borrowings</td>
<td>Alina Maria Ciobanu and Liviu P. Dinu</td>
<td>431</td>
</tr>
<tr>
<td>The Media Frames Corpus: Annotations of Frames Across Issues</td>
<td>Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik and Noah A. Smith</td>
<td>438</td>
</tr>
<tr>
<td>deltaBLEU: A Discriminative Metric for Generation Tasks with Intrinsically Diverse Targets</td>
<td>Michel Galley, Chris Brockett, Alessandro Sordoni, Yangfeng Ji, Michael Auli, Chris Quirk, Margaret Mitchell, Jianfeng Gao and Bill Dolan</td>
<td>445</td>
</tr>
<tr>
<td>Tibetan Unknown Word Identification from News Corpora for Supporting Lexicon-based Tibetan Word Segmentation</td>
<td>Minghua Nuo, Huidan Liu, Congjun Long and Jian Wu</td>
<td>451</td>
</tr>
<tr>
<td>Learning Lexical Embeddings with Syntactic and Lexicographic Knowledge</td>
<td>Tong Wang, Abdelrahman Mohamed and Graeme Hirst</td>
<td>458</td>
</tr>
<tr>
<td>Non-distributional Word Vector Representations</td>
<td>Manaal Faruqui and Chris Dyer</td>
<td>464</td>
</tr>
<tr>
<td>Early and Late Combinations of Criteria for Reranking Distributional Thesauri</td>
<td>Olivier Ferret</td>
<td>470</td>
</tr>
<tr>
<td>Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases</td>
<td>Kristina Gulordava, Paola Merlo and Benoit Crabbé</td>
<td>477</td>
</tr>
<tr>
<td>Tagging Performance Correlates with Author Age</td>
<td>Dirk Hovy and Anders Søgaard</td>
<td>483</td>
</tr>
<tr>
<td>User Based Aggregation for Biterm Topic Model</td>
<td>Weizheng Chen, Jinpeng Wang, Yan Zhang, Hongfei Yan and Xiaoming Li</td>
<td>489</td>
</tr>
<tr>
<td>The Fixed-Size Ordinally-Forgetting Encoding Method for Neural Network Language Models</td>
<td>ShiLiang Zhang, Hui Jiang, MingBin Xu, JunFeng Hou and LiRong Dai</td>
<td>495</td>
</tr>
<tr>
<td>Unsupervised Decomposition of a Multi-Author Document Based on Naïve-Bayesian Model</td>
<td>Khaled Aldebei, Xiangjian He and Jie Yang</td>
<td>501</td>
</tr>
<tr>
<td>Extended Topic Model for Word Dependency</td>
<td>Tong Wang, Vish Viswanath and Ping Chen</td>
<td>506</td>
</tr>
<tr>
<td>Dependency Recurrent Neural Language Models for Sentence Completion</td>
<td>Piotr Mirowski and Andreas Vlachos</td>
<td>511</td>
</tr>
<tr>
<td>Point Process Modelling of Rumour Dynamics in Social Media</td>
<td>Michal Lukasik, Trevor Cohn and Kalina Bontcheva</td>
<td>518</td>
</tr>
<tr>
<td>Learning Hidden Markov Models with Distributed State Representations for Domain Adaptation</td>
<td>Min Xiao and Yuhong Guo</td>
<td>524</td>
</tr>
<tr>
<td>MT Quality Estimation for Computer-assisted Translation: Does it Really Help?</td>
<td>Marco Turchi, Matteo Negri and Marcello Federico</td>
<td>530</td>
</tr>
<tr>
<td>Context-Dependent Translation Selection Using Convolutional Neural Network</td>
<td>Baotian Hu, Zhaopeng Tu, Zhengdong Lu, Hang Li and Qingcai Chen</td>
<td>536</td>
</tr>
</tbody>
</table>
Learning Word Reorderings for Hierarchical Phrase-based Statistical Machine Translation
jingyi zhang, Masao Utiyama, Eiichro Sumita and Hai Zhao .......................... 542

UNRAVEL—A Decipherment Toolkit
Malte Nuhn, Julian Schamper and Hermann Ney ........................................ 549

Multi-Pass Decoding With Complex Feature Guidance for Statistical Machine Translation
Benjamin Marie and Aurélien Max ............................................................... 554

What’s in a Domain? Analyzing Genre and Topic Differences in Statistical Machine Translation
Marlies van der Wees, Arianna Bisazza, Wouter Weerkamp and Christof Monz ........ 560

Learning Cross-lingual Word Embeddings via Matrix Co-factorization
Tianze Shi, Zhiyuan Liu, Yang Liu and Maosong Sun .................................. 567

Improving Pivot Translation by Remembering the Pivot
Akiva Miura, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura .... 573

BrailleSUM: A News Summarization System for the Blind and Visually Impaired People
Xiaojun Wan and Yue Hu .............................................................................. 578

Automatic Identification of Age-Appropriate Ratings of Song Lyrics
Anggi Maulidyani and Ruli Manurung .......................................................... 583

Ground Truth for Grammaticality Correction Metrics
Courtney Napoles, Keisuke Sakaguchi, Matt Post and Joel Tetreault .................. 588

Radical Embedding: Delving Deeper to Chinese Radicals
Xinlei Shi, Junjie Zhai, Xudong Yang, Zehua Xie and Chao Liu .......................... 594

Automatic Detection of Sentence Fragments
Chak Yan Yeung and John Lee ....................................................................... 599

A Computational Approach to Automatic Prediction of Drunk-Texting
Aditya Joshi, Abhijit Mishra, Balamurali AR, Pushpak Bhattacharyya and Mark J Carman ... 604

Reducing infrequent-token perplexity via variational corpora
Yusheng Xie, Pranjal Daga, Yu Cheng, Kunpeng Zhang, Ankit Agrawal and Alok Choudhary . 609

A Hierarchical Knowledge Representation for Expert Finding on Social Media
Yanran Li, Wenjie Li and Sujian Li ................................................................. 616

Tackling Sparsity, the Achilles Heel of Social Networks: Language Model Smoothing via Social Regularization
Rui Yan, Xiang Li, Mengwen Liu and Xiaohua Hu ............................................ 623

Twitter User Geolocation Using a Unified Text and Network Prediction Model
Afshin Rahimi, Trevor Cohn and Timothy Baldwin ....................................... 630

Automatic Keyword Extraction on Twitter
Luis Marujo, Wang Ling, Isabel Trancoso, Chris Dyer, Alan W Black, Anatole Gershman, David Martins de Matos, João Neto and Jaime Carbonell ................................. 637

Towards a Contextual Pragmatic Model to Detect Irony in Tweets
Jihen Karoui, Benamara Farah, Véronique MORICEAU, Nathalie Aussenac-Gilles and Lamia Hadrich-Belguith ......................................................... 644
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation and Classification of an Email Importance Corpus</td>
<td>Fan Zhang and Kui Xu</td>
<td>651</td>
</tr>
<tr>
<td>Lexical Comparison Between Wikipedia and Twitter Corpora by Using Word Embeddings</td>
<td>Luchen Tan, Haotian Zhang, Charles Clarke and Mark Smucker</td>
<td>657</td>
</tr>
<tr>
<td>The Discovery of Natural Typing Annotations: User-produced Potential Chinese Word Delimiters</td>
<td>Dakui Zhang, Yu Mao, Yang Liu, Hanshi Wang, Chuyuan Wei and Shiping Tang</td>
<td>662</td>
</tr>
<tr>
<td>One Tense per Scene: Predicting Tense in Chinese Conversations</td>
<td>Tao Ge, Heng Ji, Baobao Chang and Zhifang Sui</td>
<td>668</td>
</tr>
<tr>
<td>A Language-Independent Feature Schema for Inflectional Morphology</td>
<td>John Sylak-Glassman, Christo Kirov, David Yarowsky and Roger Que</td>
<td>674</td>
</tr>
<tr>
<td>Rhetoric Map of an Answer to Compound Queries</td>
<td>Boris Galitsky, Dmitry Ilvovsky and Sergey O. Kuznetsov</td>
<td>681</td>
</tr>
<tr>
<td>Thread-Level Information for Comment Classification in Community Question Answering</td>
<td>Alberto Barrón-Cedeño, Simone Filice, Giovanni Da San Martino, Shafiq Joty, Lluís Márquez, Preslav Nakov and Alessandro Moschitti</td>
<td>687</td>
</tr>
<tr>
<td>Learning Hybrid Representations to Retrieve Semantically Equivalent Questions</td>
<td>Cicero dos Santos, Luciano Barbosa, Dasha Bogdanova and Bianca Zadrozny</td>
<td>694</td>
</tr>
<tr>
<td>Machine Comprehension with Syntax, Frames, and Semantics</td>
<td>Hai Wang, Mohit Bansal, Kevin Gimpel and David McAllester</td>
<td>700</td>
</tr>
<tr>
<td>A Long Short-Term Memory Model for Answer Sentence Selection in Question Answering</td>
<td>Di Wang and Eric Nyberg</td>
<td>707</td>
</tr>
<tr>
<td>Answer Sequence Learning with Neural Networks for Answer Selection in Community Question Answering</td>
<td>Xiaoqiang Zhou, Baotian Hu, Qingcai Chen, Buzhou Tang and Xiaolong Wang</td>
<td>713</td>
</tr>
<tr>
<td>Bilingual Word Embeddings from Non-Parallel Document-Aligned Data Applied to Bilingual Lexicon Induction</td>
<td>Ivan Vulić and Marie-Francine Moens</td>
<td>719</td>
</tr>
<tr>
<td>Low-Rank Tensors for Verbs in Compositional Distributional Semantics</td>
<td>Daniel Fried, Tamara Polajnar and Stephen Clark</td>
<td>731</td>
</tr>
<tr>
<td>Constrained Semantic Forests for Improved Discriminative Semantic Parsing</td>
<td>Wei Lu</td>
<td>737</td>
</tr>
<tr>
<td>Automatic Identification of Rhetorical Questions</td>
<td>Shohini Bhattachar, Jeremy Cytryn, Elana Feldman and Joonsuk Park</td>
<td>743</td>
</tr>
<tr>
<td>Lifelong Learning for Sentiment Classification</td>
<td>Zhiyuan Chen, Nianzu Ma and Bing Liu</td>
<td>750</td>
</tr>
<tr>
<td>Harnessing Context Incongruity for Sarcasm Detection</td>
<td>Aditya Joshi, Vinita Sharma and Pushpak Bhattacharyya</td>
<td>757</td>
</tr>
</tbody>
</table>
Emotion Detection in Code-switching Texts via Bilingual and Sentimental Information
Zhongqing Wang, Sophia Lee, Shoushan Li and Guodong Zhou ......................... 763

Model Adaptation for Personalized Opinion Analysis
Mohammad Al Boni, Keira Zhou, Hongning Wang and Matthew S. Gerber ................. 769

Linguistic Template Extraction for Recognizing Reader-Emotion and Emotional Resonance Writing Assistance
Yung-Chun Chang, Cen-Chieh Chen, Yu-lun Hsieh, Chien Chin Chen and Wen-Lian Hsu .... 775

Aspect-Level Cross-lingual Sentiment Classification with Constrained SMT
Patrik Lambert ................................................................. 781

Predicting Valence-Arousal Ratings of Words Using a Weighted Graph Method
Liang-Chih Yu, Jin Wang, K. Robert Lai and Xue-jie Zhang ................................. 788

Multi-domain Dialog State Tracking using Recurrent Neural Networks
Nikola Mrkšić, Diarmuid Ó Séaghdha, Blaise Thomson, Milica Gasic, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen and Steve Young ......................................................... 794

Dialogue Management based on Sentence Clustering
Wendong Ge and Bo Xu ........................................................................ 800

Compact Lexicon Selection with Spectral Methods
Young-Bum Kim, Karl Stratos, Xiaohu Liu and Ruhi Sarikaya ................................. 806

The Impact of Listener Gaze on Predicting Reference Resolution
Nikolina Koleva, Martin Villalba, Maria Staudte and Alexander Koller ..................... 812

A Simultaneous Recognition Framework for the Spoken Language Understanding Module of Intelligent Personal Assistant Software on Smart Phones
Changsu Lee, Youngjoong Ko and Jungyun Seo .................................................. 818

A Deeper Exploration of the Standard PB-SMT Approach to Text Simplification and its Evaluation
Sanja Štajner, Hannah Bechara and Horacio Saggion .......................................... 823

Learning Summary Prior Representation for Extractive Summarization
Ziqiang Cao, Furu Wei, Sujian Li, Wenjie Li, Ming Zhou and Houfeng WANG .............. 829

A Methodology for Evaluating Timeline Generation Algorithms based on Deep Semantic Units
Sandro Bauer and Simone Teufel ....................................................................... 834

Unsupervised extractive summarization via coverage maximization with syntactic and semantic concepts
Natalie Schluter and Anders Søgaard ................................................................. 840

Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser
Long Duong, Trevor Cohn, Steven Bird and Paul Cook ......................................... 845

Semantic Structure Analysis of Noun Phrases using Abstract Meaning Representation
Yuichiro Sawai, Hiroyuki Shindo and Yuji Matsumoto .......................................... 851

Boosting Transition-based AMR Parsing with Refined Actions and Auxiliary Analyzers
Chuan Wang, Nianwen Xue and Sameer Pradhan ................................................ 857

Generative Incremental Dependency Parsing with Neural Networks
Jan Buys and Phil Blunsom .............................................................................. 863
Labeled Grammar Induction with Minimal Supervision
Yonatan Bisk, Christos Christodoulopoulos and Julia Hockenmaier .......................... 870

On the Importance of Ezafe Construction in Persian Parsing
Alireza Nourian, Mohammad Sadegh Rasooli, Mohsen Imany and Heshaam Faili .............. 877