

29th Pacific Asia Conference on Language, Information and Computation (PACLIC 2015)

Shanghai, China

30 October – 1 November 2015

ISBN: 978-1-5108-1756-2

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

Two-level Word Class Categorization Model in Analytic Languages and Its Implications for POS Tagging in Modern Chinese Corpora <i>Renqiang Wang and Changning Huang</i>	1
A Review of Corpus-based Statistical Models of Language Variation <i>Yao Yao</i>	11
Translation of Unseen Bigrams by Analogy Using an SVM Classifier <i>Hao Wang, Lu Lyu and Yves Lepage</i>	16
Machine Translation Experiments on PADIC: A Parallel Arabic DIAlect Corpus <i>Karima Meftouh, Salima Harrat, Salma Jamoussi, Mourad Abbas and Kamel Smail</i>	26
Surrounding Word Sense Model for Japanese All-words Word Sense Disambiguation <i>Kanako Komiya, Yuto Sasaki, Hajime Morita, Minoru Sasaki, Hiroyuki Shinnou and Yoshiyuki Kotani</i>	35
Computing Semantic Text Similarity Using Rich Features <i>Yang Liu, Chengjie Sun, Lei Lin, Xiaolong Wang and Yuming Zhao</i>	44
Mechanical Turk-based Experiment vs Laboratory-based Experiment: A Case Study on the Comparison of Semantic Transparency Rating Data <i>Shichang Wang, Chu-Ren Huang, Yao Yao and Angel Chan</i>	53
Discourse Relation Recognition by Comparing Various Units of Sentence Expression with Recursive Neural Network <i>Atsushi Otsuka, Toru Hirano, Chiaki Miyazaki, Ryo Masumura, Ryuichiro Higashinaka, Toshiro Makino and Yoshihiro Matsuo</i>	63
Bidirectional Long Short-Term Memory Networks for Relation Classification <i>Shu Zhang, Dequan Zheng, Xinchun Hu and Ming Yang</i>	73
Distant Supervision for Entity Linking <i>Miao Fan, Qiang Zhou and Thomas Fang Zheng</i>	79
Toward Algorithmic Discovery of Biographical Information in Local Gazetteers of Ancient China <i>Chao-Lin Liu, Chih-Kai Huang, Hongsu Wang and Peter K. Bol</i>	87

Fast and Large-scale Unsupervised Relation Extraction <i>Sho Takase, Naoaki Okazaki and Kentaro Inui</i>	96
Reducing Lexical Features in Parsing by Word Embeddings <i>Hiroya Komatsu, Ran Tian, Naoaki Okazaki and Kentaro Inui</i>	106
High-order Graph-based Neural Dependency Parsing <i>Zhisong Zhang and Hai Zhao</i>	114
A Dynamic Syntax Modelling of Postposing in Japanese Narratives <i>Tohru Seraku</i>	124
Unsupervised and Lightly Supervised Part-of-Speech Tagging Using Recurrent Neural Networks <i>Othman Zennaki, Nasredine Semmar and Laurent Besacier</i>	133
Identifying Prepositional Phrases in Chinese Patent Texts with Rule-based and CRF Methods <i>Hongzheng Li and Yaohong Jin</i>	143
Japanese Sentiment Classification with Stacked Denoising Auto-Encoder using Distributed Word Representation <i>Peinan Zhang and Mamoru Komachi</i>	150
Is Wikipedia Really Neutral? A Sentiment Perspective Study of War-related Wikipedia Articles since 1945 <i>Yiwei Zhou, Alexandra Cristea and Zachary Roberts</i>	160
A Comprehensive Filter Feature Selection for Improving Document Classification <i>Nguyen Hoai Nam Le and Bao Quoc Ho</i>	169
Sentiment Analyzer with Rich Features for Ironic and Sarcastic Tweets <i>Piyoros Tunthamthiti, Enrico Santus, Hongzhi Xu, Chu-Ren Huang and Shirai Kiyooki</i>	178
Thai Stock News Sentiment Classification using Wordpair Features <i>Ponrudee Netisopakul and Apinan Chattupan</i>	188
Sentiment Classification of Arabic Documents: Experiments with multi-type features and ensemble algorithms <i>Amine Bayoudhi, Hatem Ghorbel and Lamia Hadrich Belguith</i>	196
The Invertible Construction in Chinese <i>Cong Yan, Lian-Hee Wee and Chu-Ren Huang</i>	206

Pan's (2001) puzzle revisited <i>Hyunjun Park</i>	212
English Right Dislocation <i>Kohji Kamada</i>	221
A Comparative Study on Mandarin and Cantonese Resultative Verb Compounds <i>Helena Yan Ping Lau and Sophia Yat Mei Lee</i>	231
Complex-NP Islands in Korean: An Experimental Approach <i>Yong-Hun Lee and Yeonkyung Park</i>	240
Two Types of Multiple Subject Constructions (MSCs) in Korean <i>Ji-Hye Kim, Eunah Kim and James Yoon</i>	250
A Large-scale Study of Statistical Machine Translation Methods for Khmer Language <i>Ye Kyaw Thu, Vichet Chea, Andrew Finch, Masao Utiyama and Eiichiro Sumita</i>	259
English to Chinese Translation: How Chinese Character Matters <i>Rui Wang, Hai Zhao and Bao-Liang Lu</i>	270
Well-Formed Dependency to String translation with BTG Grammar <i>Xiaoqing Li, Kun Wang, Dakun Zhang and Jie Hao</i>	281
Large-scale Dictionary Construction via Pivot-based Statistical Machine Translation with Significance Pruning and Neural Network Features <i>Raj Dabre, Chenhui Chu, Fabien Cromieres, Toshiaki Nakazawa and Sadao Kuroha- shi</i>	289
Annotation and Classification of French Feedback Communicative Functions <i>Laurent Prévot, Jan Gorisch and Sankar Mukherjee</i>	298
Automatic conversion of sentence-end expressions for utterance characterization of d- ialogue systems <i>Chiaki Miyazaki, Toru Hirano, Ryuichiro Higashinaka, Toshiro Makino and Yoshihiro Matsuo</i>	307
Auditory Synaesthesia and Near Synonyms: A Corpus-Based Analysis of sheng1 and yin1 in Mandarin Chinese <i>Qingqing Zhao, Chu-Ren Huang and Hongzhi Xu</i>	315
System Utterance Generation by Label Propagation over Association Graph of Words and Utterance Patterns for Open-Domain Dialogue Systems	

<i>Hiroshi Tsukahara and Kei Uchiumi</i>	323
The Cross - modal Representation of Metaphors <i>Yutung Chang and Kawai Chui</i>	332
Writing to Read: the Case of Chinese <i>Qi Zhang and Ronan Reilly</i>	341
Design of a Learner Corpus for Listening and Speaking Performance <i>Katsunori Kotani and Takehiko Yoshimi</i>	351
Understanding Infants' Language Development in Relation to Levels of Consciousness: An Approach in Building up an Agent-based Model <i>Helena Hong Gao and Can Guo</i>	359
Pivot-Based Topic Models for Low-Resource Lexicon Extraction <i>John Richardson, Toshiaki Nakazawa and Sadao Kurohashi</i>	369
A Corpus-Based Study of <i>zunshou</i> and Its English Equivalents <i>Ying Liu</i>	378
Self Syntactico-Semantic Enrichment of LMF Normalized Dictionaries <i>Imen Elleuch, Bilel Gargouri and Abdelmajid Ben Hamadou</i>	387
When Embodiment Meets Generative Lexicon: The Human Body Part Metaphors in Sinica Corpus <i>Ren-Feng Duann and Chu-Ren Huang</i>	396
Degree Variables by Choose Degree in Izyooni 'than'-Clauses <i>Toshiko Oda</i>	404
Not Voice but Case Identity in VP Ellipsis of English <i>Myungkwan Park and Sunjoo Choi</i>	413
A Statistical Modeling of the Correlation between Island Effects and Working-memory Capacity for L2 Learners <i>Euhee Kim and Myungkwan Park</i>	422
De-verbalization and Nominal Categories in Mandarin Chinese: A corpus-driven study in both Mainland Mandarin and Taiwan Mandarin <i>Jiajuan Xiong and Chu-Ren Huang</i>	431
Zero Object Resolution in Korean <i>Arum Park, Seunghee Lim and Munpyo Hong</i>	439

An Improved Hierarchical Word Sequence Language Model Using Directional Information <i>Xiaoyi Wu and Yuji Matsumoto</i>	449
Neural Network Language Model for Chinese Pinyin Input Method Engine <i>Shenyuan Chen, Hai Zhao and Rui Wang</i>	455
Real-time Detection and Sorting of News on Microblogging Platforms <i>Wenting Tu, David Cheung, Nikos Mamoulis, Min Yang and Ziyu Lu</i>	462
Trouble information extraction based on a bootstrap approach from Twitter <i>Kohei Kurihara and Kazutaka Shimada</i>	471
Using Twitter Data to Infer Personal Values of Japanese Consumers <i>Yinjun Hu and Yasuo Tanida</i>	480
Distant-supervised Language Model for Detecting Emotional Upsurge on Twitter <i>Yoshinari Fujinuma, Hikaru Yokono, Pascual Martínez-Gómez and Akiko Aizawa</i> ..	488
Hybrid Method of Semi-supervised Learning and Feature Weighted Learning for Domain Adaptation of Document Classification <i>Hiroyuki Shinnou, Liying Xiao, Minoru Sasaki and Kanako Komiya</i>	496
Paraphrase Detection Based on Identical Phrase and Similar Word Matching <i>Hoang-Quoc Nguyen-Son, Yusuke Miyao and Isao Echizen</i>	504
Multi-aspects Rating Prediction Using Aspect Words and Sentences <i>Takuto Nakamuta and Kazutaka Shimada</i>	513
Understanding Rating Behaviour and Predicting Ratings by Identifying Representative Users <i>Rahul Kamath, Masanao Ochi and Yutaka Matsuo</i>	522
Cross-lingual Pseudo Relevance Feedback Based on Weak Relevant Topic Alignment <i>Xuwen Wang, Qiang Zhang, Xiaojie Wang and Junlian Li</i>	529
Corpus annotation with a linguistic analysis of the associations between event mentions and spatial expressions <i>Jin-Woo Chung, Jinseon You and Jong C. Park</i>	535
Recognizing Complex Negation on Twitter <i>Junta Mizuno, Canasai Kruengkrai, Kiyonori Ohtake, Chikara Hashimoto, Kentaro Torisawa, Julien Kloetzer and Kentaro Inui</i>	544

Topic Model for Identifying Suicidal Ideation in Chinese Microblog	
<i>Xiaolei Huang, Xin Li, Tianli Liu, David Chiu, Tingshao Zhu and Lei Zhang.....</i>	<i>553</i>
Predicting Sector Index Movement with Microblogging Public Mood Time Series on Social Issues	
<i>Yujie Lu, Jinlong Guo, Kotaro Sakamoto, Hideyuki Shibuki and Tatsunori Mori.....</i>	<i>563</i>