

2012 International Conference on Asian Language Processing

(IALP 2012)

**Hanoi, Vietnam
13 – 15 November 2012**



**IEEE Catalog Number: CFP1244I-PRT
ISBN: 978-1-4673-6113-2**

2012 International Conference on Asian Language Processing

IALP 2012

Table of Contents

Message from General Chair.....	X
Message from Program Chairs.....	xi
Message from the Local Organizing Chair.....	xii
Organizing Committee.....	xiii
Program Committee.....	xiv
Reviewers.....	xix

Session 1A: Computational Linguistics and Language Studies 1

Classification of Emotional Onomatopoeias Based on Questionnaire Surveys	1
<i>Yuzu Uchida, Kenji Araki, and Jun Yoneyama</i>	
An Acoustic Investigation of Chinese Tone 4 Produced by German Learners	5
<i>Hongwei Ding, Rüdiger Hoffmann, Hussein Hussein, and Oliver Jokisch</i>	
Computer Identification of the Ambiguity Structure a Shuo Shenmeyer B	9
<i>Lin He and Shuzhen Shi</i>	
Multi-view Learning for Semi-supervised Sentiment Classification	13
<i>Yan Su, Shoushan Li, Shengfeng Ju, Guodong Zhou, and Xiaojun Li</i>	
Linguistic Features for Subjectivity Classification	17
<i>Huong Nguyen Thi Xuan, Anh Cuong Le, and Le Minh Nguyen</i>	

Session 1B: Language Information Processing 1

From Raw Text to Morphological Rules for Iban Morphological Analyser	21
<i>Suhaila Saeed, Lay-Ki Soon, Tek-Yonglim, Bali Ranaivo-Malançon, and Enya Kong Tang</i>	
A Hierarchical Graph Method Using in A* Algorithm for Vietnamese Parsing Technique	25
<i>Le Quang Thang and Đỗ Đạt Tran</i>	

Study on the Influencing Factors of Chinese Word Segmentation	29
<i>Chi Xiu and Rou Song</i>	
NLP Oriented Japanese Pun Classification	33
<i>Pawel Dybala, Rafal Rzepka, Kenji Araki, and Kohichi Sayama</i>	
Exploring Self-Training and Co-training for Hindi Dependency Parsing Using Partial Parses	37
<i>Rahul Goutam</i>	

Session 2A: Computational Linguistics and Language Studies 2

Word Sense Induction Using Correlated Topic Model	41
<i>Thanh Tung Hoang and Phuong Thai Nguyen</i>	
Analyzing and Classifying the Yami Emotion Phrases Using Ontological Structure and Computation	45
<i>Meng-Chien Yang, D. Victoria Rau, and Yi-Hsin Wu</i>	
Anaphora Resolution in Bangla Using Global Discourse Knowledge	49
<i>Apurbalal Senapati and Utpal Garain</i>	
Improve Chinese Semantic Dependency Parsing via Syntactic Dependency Parsing	53
<i>Meishan Zhang, Wanxiang Che, Yanqiu Shao, and Ting Liu</i>	
Feature-Based Subjectivity Classification of Filipino Text	57
<i>Ralph Vincent J. Regalado and Charibeth K. Cheng</i>	

Session 2B: Language Information Processing 2

Event Information Extraction Approach Based on Complex Chinese Texts	61
<i>Bao Jiana, Li Tingyu, and Yao Tianfang</i>	
A Research of Chinese Sentence Constituent Labeling Based on History Hidden Markov Model	65
<i>Zhijian Xu, Siming Luo, Dan Li, and Hongxia Li</i>	
Weirdness Coefficient as a Feature Selection Method for Arabic Special Domain Text Classification	69
<i>Abdalmohsen Al-Thubaity, Albandari Alanazi, Itisam Hazzaa, and Haya Al-Tuwajri</i>	
Development and Evaluation of a Model for Japanese Honorific Expressions Using Assembly Minutes	73
<i>Hokuto Ototake and Kenji Yoshimura</i>	
Quality Assurance for Segmentation and Tagging of Chinese Novels in the Ming and Qing Dynasties	77
<i>Dan Xiong, Qin Lu, Fengju Lo, Dingxu Shi, and Tin-Shing Chiu</i>	

Session 3A: Language Technology

Initial Experiments with Tamil LVCSR	81
<i>Melvin Jose J., Ngoc Thang Vu, and Tanja Schultz</i>	
Named Entity Recognition for Vietnamese Documents Using Semi-supervised Learning Method of CRFs with Generalized Expectation Criteria	85
<i>Thi-Ngan Pham, Le Minh Nguyen, and Quang-Thuy Ha</i>	
Mining Space-Time Elements of Opinion	89
<i>Tianfang Yao, Jun Liu, and Wei Qiu</i>	
A Novel Approach Towards Building a Portable NLIDB System Using the Computational Paninian Grammar Framework	93
<i>Abhijeet Gupta, Arjun Akula, Deepak Malladi, Puneeth Kukkadapu, Vinay Ainavolu, and Rajeev Sangal</i>	
Automatically Mining Parallel Corpora for Minority Languages from Web Pages	97
<i>Zede Zhu, Miao Li, Lei Chen, and Weihui Zeng</i>	

Session 3B: Machine Translation 1

A Dictionary Based Urdu Word Segmentation Using Maximum Matching Algorithm for Space Omission Problem	101
<i>Rabiya Rashid and Seemab Latif</i>	
An Enhanced Model for Lexical Gap Processing in English-Vietnamese Machine Translation	105
<i>Tuoi Phan Thi and Hai Le Manh</i>	
A Malay Dialect Translation and Synthesis System: Proposal and Preliminary System	109
<i>Tien-Ping Tan, Sang-Seong Goh, and Yen-Min Khaw</i>	
Improving Word Alignment for Statistical Machine Translation Based on Constraints	113
<i>Le Quang-Hung and Le Anh-Cuong</i>	

Session 4A: Machine Translation 2

Mongolian Morphological Segmentation with Hidden Markov Model	117
<i>Miantao He, Miao Li, and Lei Chen</i>	
Improving the Quality of Word Alignment by Integrating Pearson's Chi-Square Test Information	121
<i>Cuong Hoang, Cuong-Anh Le, and Son-Bao Pham</i>	
A Systematic Comparison of SVM and Maximum Entropy Classifiers for Translation Error Detection	125
<i>Jinhua Du and Sha Wang</i>	

Modality-Preserving Phrase-Based Statistical Machine Translation	129
<i>Masamichi Ideue, Kazuhide Yamamoto, Masao Utiyama, and Eiichiro Sumita</i>	
Dependency Parsing on Source Language with Reordering Information in SMT	133
<i>Lei Chen, Miao Li, Miantao He, and Hui Liu</i>	

Session 4B: Spoken Language Processing 1

Automatic Speech Recognition of Code Switching Speech Using 1-Best Rescoring	137
<i>Basem H.A. Ahmed and Tien-Ping Tan</i>	
YAST: A Scalable ASR Toolkit Especially Designed for Under-Resourced Languages	141
<i>Emmanuel Ferreira, Pascal Nocera, Maria Goudi, and Ngoc Diep Do Thi</i>	
Combining Evidences from Mel Cepstral Features and Cepstral Mean Subtracted Features for Singer Identification	145
<i>Hemant A. Patil, Purushotam G. Radadia, and T.K. Basu</i>	
Person Recognition Using Humming, Singing and Speech	149
<i>Hemant A. Patil, Maulik C. Madhavi, and Nirav H. Chhayani</i>	
Pronunciation Modeling for Malaysian English	153
<i>Yen Min Khaw and Tien Ping Tan</i>	

Session 5: Poster Session

Building an English-Vietnamese Bilingual Corpus for Machine Translation	157
<i>Quoc Hung Ngo and Werner Winiwarter</i>	
Applying Smartphone Technology to Compile Innovative Arabic Learner's Dictionaries	161
<i>Jack Halpern</i>	
A Corpus-Based Analysis of Mixed Code in Hong Kong Speech	165
<i>John Lee</i>	
Combining Documentation and Research: Ongoing Work on an Endangered Language	169
<i>Alexis Michaud, Andrew Hardie, Séverine Guillaume, and Martine Toda</i>	
Automatic Generation with Manual Post-Editing of Shallow-Transfer Rule for English-to-Indonesian	173
<i>Evi Yulianti, Mirna Adriani, Hisar M. Manurung, Indra Budi, and Achmad N. Hidayanto</i>	
Phonetic Transcription of Fricatives and Plosives for Gujarati and Marathi Languages	177
<i>Hemant A. Patil, Maulik C. Madhavi, Kewal D. Malde, and Bhavik B. Vachhani</i>	
A New Model of Time Expressions Detection and Annotation in Vietnamese: The hôm Case	181
<i>Philippe Lambert, Sylviane R. Schwer, and Nicolas Boffo</i>	

A Hybrid Approach for Word Alignment in English-Hindi Parallel Corpora with Scarce Resources	185
<i>Jyoti Srivastava and Sudip Sanyal</i>	
A Pointwise Approach for Vietnamese Diacritics Restoration	189
<i>Tuan Anh Luu and Kazuhide Yamamoto</i>	
Source-Side Suffix Stripping for Bengali-to-English SMT	193
<i>Rejwanul Haque, Sergio Penkale, Jie Jiang, and Andy Way</i>	
Perception of the Vietnamese Short Vowels / , , ə/ in /bVk/ Produced by Female Voice	197
<i>Tran Thi Anh Xuan, René Carré, Eric Castelli, and Nathalie Vallée</i>	
An Approach to N-Gram Language Model Evaluation in Phrase-Based Statistical Machine Translation	201
<i>Jinsong Su, Qun Liu, Huailin Dong, Yidong Chen, and Xiaodong Shi</i>	
 Session 6A: Machine Translation 3	
Machine Transliteration Based on Error-Driven Learning	205
<i>Ying Qin</i>	
Morpheme Segmentation Using Bilingual Features	209
<i>Hui Liu, Miao Li, Jian Zhang, and Lei Chen</i>	
How Good is Web Based MT: A Case Study on Paper Title Translation	213
<i>Jia Yao, Muyun Yang, Junmao Meng, Dongqing Xiao, Tiejun Zhao, and Sheng Li</i>	
Identification and Transformation of Comparative Sentences in Patent Chinese-English Machine Translation	217
<i>Runxiang Zhang and Yaohong Jin</i>	
 Session 6B: Spoken Language Processing 2	
Tonal Co-articulation on Particles in Vietnamese Language	221
<i>Thị Lan Nguyễn and Đỗ Đạt Trần</i>	
N-gram Language Model Based on Multi-Word Expressions in Web Documents for Speech Recognition and Closed-Captioning	225
<i>Shinya Takahashi and Tsuyoshi Morimoto</i>	
Analysis of Malay Speech Recognition for Different Speaker Origins	229
<i>Sarah Samson Juan, Laurent Besacier, and Tien-Ping Tan</i>	
A Phone Mapping Technique for Acoustic Modeling of Under-Resourced Languages	233
<i>Van Hai Do, Xiong Xiao, Eng Siong Chng, and Haizhou Li</i>	
Culturomics on a Bengali Newspaper Corpus	237
<i>Shanta Phani, Shibamouli Lahiri, and Arindam Biswas</i>	
 Author Index	 241