ACL 2018

# Natural Language Processing Techniques for Educational Applications

## Proceedings of the Fifth Workshop

July 19, 2018
Melbourne, Australia

©2018 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-948087-35-3

# Preface

Welcome to the 5th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2018), with a Shared Task on Chinese Grammatical Error Diagnosis.

The development of Natural Language Processing (NLP) has advanced to a level that affects the research landscape of many academic domains and has practical applications in many industrial sectors. On the other hand, educational environment has also been improved to impact the world society, such as the emergence of MOOCs (Massive Open Online Courses). With these trends, this workshop focuses on the NLP techniques applied to the educational environment. Research issues in this direction have gained more and more attention, examples including the activities like the workshops on Innovative Use of NLP for Building Educational Applications since 2005 and educational data mining conferences since 2008.

This is the fifth workshop held in the Asian area, with the first one NLPTEA 2014 workshop being held in conjunction with the 22nd International Conference on Computer in Education (ICCE 2014) from Nov. 30 to Dec. 4, 2014 in Japan. The second edition NLPTEA 2015 workshop was held in conjunction with the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015) from July 26- 31 in Beijing, China. The third version NLPTEA 2016 workshop was held in conjunction with the 26th International Conference on Computational Linguistics (COLING 2016) from December 11- 16 in Osaka, Japan. The fourth edition NLPTEA 2017 workshop was held in conjunction with the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017) from November 27- December 1 in Taipei, Taiwan. This year, we continue to promote this research line by holding the workshop in conjunction with the ACL 2018 conference and also holding the fourth shared task on Chinese Grammatical Error Diagnosis. We receive 33 valid submissions for research issues, each of which was reviewed by at least two experts, and have 12 teams participating in the shared task and submitting their task reports. In total, there are 10 oral papers and 20 posters accepted. We also organize a keynote speech in this workshop. The invited speaker Professor Yuji Matsumoto is expected to deliver a great talk entitled as "Multi-word Expressions in Second Language Learning".

We would like to thank the program committee members for their hard work in completing the review tasks. Their collective efforts achieved quality reviews of the submissions within a few weeks. Great thanks should also go to the speaker, authors, and participants for the tremendous supports in making the workshop a success.

Welcome you to the Melbourne city, and wish you enjoy the city as well as the workshop.

NLPTEA 2018 Workshop Chairs
Yuen-Hsien Tseng, National Taiwan Normal University
Hsin-Hsi Chen, National Taiwan University
Vincent Ng, The University of Texas at Dallas
Mamoru Komachi, Tokyo Metropolitan University

# Organization

**Workshop Organizers:**

Yuen-Hsien Tseng, National Taiwan Normal University
Hsin-Hsi Chen, National Taiwan University
Vincent Ng, The University of Texas at Dallas
Mamoru Komachi, Tokyo Metropolitan University

**Shared Task Organizers:**

Gaoqi Rao, Beijing Language and Culture University
Qi Gong, Beijing Language and Culture University
Baolin Zhang, Beijing Language and Culture University
Endong Xun, Beijing Language and Culture University

**Program Committee:**

David Alfter, University of Gothenburg
Chris Brockett, Microsoft Research
Christopher Bryant, Cambridge University
Tao Chen, John Hokins University
Vidas Daudaravicius, VTeX Solutions for Science Publishing
Mariano Felice, Cambridge University
Cyril Goutte, National Research Council Canada
Homa B. Hashemi, University of Pittsburgh
Trude Heift, Simon Fraser University
Tomoyuki Kajiwara, Tokyo Metropolitan University
Herbert Lange, University of Gothenburg
John Lee, The City University of Hong Kong
Lung-Hao Lee, National Taiwan University
Chen Li, Microsoft, USA
Chuan-Jie Lin, National Taiwan Ocean University
Shervin Malmasi, Harvard University
Tomoya Mizumoto, Tohoku University
Courtney Napoles, John Hopkins University
Gustavo Paetzold, University of Sheffield
Livy Real, IBM Research, Brazil
Elizabeth Salesky, Massachusetts Institute of Technology
Yukio Tono, Tokyo University of Foreign Studies
Elena Volodina, University of Gothenburg
Thuy Vu, University of California, Los Angles
Mats Wiren, Stockholm University
Shih-Hung Wu, Chaoyang University of Technology
Huichao Xue, Google, USA
Jui-Feng Yeh, National Chiayi University
Dong Yu, Beijing Language and Culture University
Liang-Chih Yu, Yuan Ze University
Zheng Yuan, Cambridge University
Marcos Zampieri, University of Wolverhampton

# Invited Speaker

Yuji Matsumoto, Professor of Information Science, Nara Institute of Science and Technology

**Title:**

Multi-word Expressions in Second Language Learning

**Abstract:**

Multi-word Expressions (MWEs) pose difficult problems to the learners of a second language. Effective learning of MWEs is important for them to become fluent speakers or writers. In this talk, I will discuss what kinds of resource and functionality are useful in computational assistance to language learners, and present our experiences on construction of MWE resources, MWE usage classification, MWE-aware error correction and proper usage suggestion.

**Biography:**

Yuji Matsumoto is currently a Professor of Information Science, Nara Institute of Science and Technology, and a Team Leader of the Knowledge Acquisition Team at Riken AIP. He received his M.S. and Ph.D. degrees in information science from Kyoto University in 1979 and in 1989. He joined Machine Inference Section of Electrotechnical Laboratory in 1979. He has then experienced an academic visitor at Imperial College of Science and Technology, a deputy chief of First Laboratory at ICOT, and an associate professor at Kyoto University. His main research interests are natural language understanding and machine learning. He is an ACL fellow and a fellow of Information Processing Society of Japan.

# Table of Contents

*Generating Questions for Reading Comprehension using Coherence Relations*
Takshak Desai, Parag Dakle and Dan Moldovan . . . . . 1

*Syntactic and Lexical Approaches to Reading Comprehension*
Henry Lin . . . . . 11

*Feature Optimization for Predicting Readability of Arabic L1 and L2*
Hind Saddiki, Nizar Habash, Violetta Cavalli-Sforza and Muhamed Al Khalil . . . . . 20

*A Tutorial Markov Analysis of Effective Human Tutorial Sessions*
Nabin Maharjan and Vasile Rus . . . . . 30

*Thank "Goodness"! A Way to Measure Style in Student Essays*
Sandeep Mathias and Pushpak Bhattacharyya . . . . . 35

*Overview of NLPTEA-2018 Share Task Chinese Grammatical Error Diagnosis*
Gaoqi RAO, Qi Gong, Baolin Zhang and Endong Xun . . . . . 42

*Chinese Grammatical Error Diagnosis using Statistical and Prior Knowledge driven Features with Probabilistic Ensemble Enhancement*
Ruiji Fu, Zhengqi Pei, Jiefu Gong, Wei Song, Dechuan Teng, Wanxiang Che, Shijin Wang, Guoping Hu and Ting Liu . . . . . 52

*A Hybrid System for Chinese Grammatical Error Diagnosis and Correction*
Chen Li, Junpei Zhou, Zuyi Bao, Hengyou Liu, Guangwei Xu and Linlin Li . . . . . 60

*Ling@CASS Solution to the NLP-TEA CGED Shared Task 2018*
Qinan Hu, Yongwei Zhang, Fang Liu and Yueguo Gu . . . . . 70

*Chinese Grammatical Error Diagnosis Based on Policy Gradient LSTM Model*
Changliang Li and Ji Qi . . . . . 77

*The Importance of Recommender and Feedback Features in a Pronunciation Learning Aid*
Dzikri Fudholi and Hanna Suominen . . . . . 83

*Selecting NLP Techniques to Evaluate Learning Design Objectives in Collaborative Multi-perspective Elaboration Activities*
Aneesha Bakharia . . . . . 88

*Augmenting Textual Qualitative Features in Deep Convolution Recurrent Neural Network for Automatic Essay Scoring*
Tirthankar Dasgupta, Abir Naskar, Lipika Dey and Rupsa Saha . . . . . 93

*Joint learning of frequency and word embeddings for multilingual readability assessment*
Dieu-Thu Le, Cam-Tu Nguyen and Xiaoliang Wang . . . . . 103

*MULLE: A grammar-based Latin language learning tool to supplement the classroom setting*
Herbert Lange and Peter Ljunglöf . . . . . 108

*Textual Features Indicative of Writing Proficiency in Elementary School Spanish Documents*
Gemma Bel-Enguix, Diana Dueñas Chavez and Arturo Curiel Díaz . . . . . 113

*Assessment of an Index for Measuring Pronunciation Difficulty*
Katsunori Kotani and Takehiko Yoshimi . . . . . 119

*A Short Answer Grading System in Chinese by Support Vector Approach*
Shih-Hung Wu and Wen-Feng Shih . . . . . 125

*From Fidelity to Fluency: Natural Language Processing for Translator Training*
Oi Yee Kwong . . . . . 130

*Countering Position Bias in Instructor Interventions in MOOC Discussion Forums*
Muthu Kumar Chandrasekaran and Min-Yen Kan . . . . . 135

*Measuring Beginner Friendliness of Japanese Web Pages explaining Academic Concepts by Integrating Neural Image Feature and Text Features*
Hayato Shiokawa, Kota Kawaguchi, Bingcai Han, Takehito Utsuro, Yasuhide Kawada, Masaharu Yoshioka and Noriko Kando . . . . . 143

*Learning to Automatically Generate Fill-In-The-Blank Quizzes*
Edison Marrese-Taylor, Ai Nakajima, Yutaka Matsuo and Ono Yuichi . . . . . 152

*Multilingual Short Text Responses Clustering for Mobile Educational Activities: a Preliminary Exploration*
Yuen-Hsien Tseng, Lung-Hao Lee, Yu-Ta Chien, Chun-Yen Chang and Tsung-Yen Li . . . . . 157

*Chinese Grammatical Error Diagnosis Based on CRF and LSTM-CRF model*
Yujie Zhou, Yinan Shao and Yong Zhou . . . . . 165

*Contextualized Character Representation for Chinese Grammatical Error Diagnosis*
Jianbo Zhao, Si Li and Zhiqing Lin . . . . . 172

*CMMC-BDRC Solution to the NLP-TEA-2018 Chinese Grammatical Error Diagnosis Task*
Zhang Yongwei, Hu Qinan, Liu Fang and Gu Yueguo . . . . . 180

*Detecting Simultaneously Chinese Grammar Errors Based on a BiLSTM-CRF Model*
Yajun Liu, Hongying Zan, Mengjie Zhong and Hongchao Ma . . . . . 188

*A Hybrid Approach Combining Statistical Knowledge with Conditional Random Fields for Chinese Grammatical Error Detection*
Yiyi Wang and Chilin Shih . . . . . 194

*CYUT-III Team Chinese Grammatical Error Diagnosis System Report in NLPTEA-2018 CGED Shared Task*
Shih-Hung Wu, JUN-WEI WANG, Liang-Pu Chen and Ping-Che Yang . . . . . 199

*Detecting Grammatical Errors in the NTOU CGED System by Identifying Frequent Subsentences*
Chuan-Jie Lin and Shao-Heng Chen . . . . . 203

# Workshop Program

**Thursday, July 19, 2018**

**09:20–09:30 Opening Remarks**

**09:30–10:30 Invited Talk**

09:30–10:30 *Multi-word Expressions in Second Language Learning*
Yuji Matsumoto

**10:30–11:00 Coffee Break**

**11:00–12:40 Regular Paper Session**

11:00–11:20 *Generating Questions for Reading Comprehension using Coherence Relations*
Takshak Desai, Parag Dakle and Dan Moldovan

11:20–11:40 *Syntactic and Lexical Approaches to Reading Comprehension*
Henry Lin

11:40–12:00 *Feature Optimization for Predicting Readability of Arabic L1 and L2*
Hind Saddiki, Nizar Habash, Violetta Cavalli-Sforza and Muhamed Al Khalil

12:00–12:20 *A Tutorial Markov Analysis of Effective Human Tutorial Sessions*
Nabin Maharjan and Vasile Rus

12:20–12:40 *Thank "Goodness"! A Way to Measure Style in Student Essays*
Sandeep Mathias and Pushpak Bhattacharyya

**Thursday, July 19, 2018 (continued)**

**12:40–14:10 Lunch**

**14:10–15:30 Shared Task Session**

14:10–14:30 *Overview of NLPTEA-2018 Share Task Chinese Grammatical Error Diagnosis*
Gaoqi RAO, Qi Gong, Baolin Zhang and Endong Xun

14:30–14:45 *Chinese Grammatical Error Diagnosis using Statistical and Prior Knowledge driven Features with Probabilistic Ensemble Enhancement*
Ruiji Fu, Zhengqi Pei, Jiefu Gong, Wei Song, Dechuan Teng, Wanxiang Che, Shijin Wang, Guoping Hu and Ting Liu

14:45–15:00 *A Hybrid System for Chinese Grammatical Error Diagnosis and Correction*
Chen Li, Junpei Zhou, Zuyi Bao, Hengyou Liu, Guangwei Xu and Linlin Li

15:00–15:15 *Ling@CASS Solution to the NLP-TEA CGED Shared Task 2018*
Qinan Hu, Yongwei Zhang, Fang Liu and Yueguo Gu

15:15–15:30 *Chinese Grammatical Error Diagnosis Based on Policy Gradient LSTM Model*
Changliang Li and Ji Qi

**15:30–16:00 Coffee Break**

**16:00–17:00 Poster Session**

*The Importance of Recommender and Feedback Features in a Pronunciation Learning Aid*
Dzikri Fudholi and Hanna Suominen

*Selecting NLP Techniques to Evaluate Learning Design Objectives in Collaborative Multi-perspective Elaboration Activities*
Aneesha Bakharia

*Augmenting Textual Qualitative Features in Deep Convolution Recurrent Neural Network for Automatic Essay Scoring*
Tirthankar Dasgupta, Abir Naskar, Lipika Dey and Rupsa Saha

**Thursday, July 19, 2018 (continued)**

*Joint learning of frequency and word embeddings for multilingual readability assessment*
Dieu-Thu Le, Cam-Tu Nguyen and Xiaoliang Wang

*MULLE: A grammar-based Latin language learning tool to supplement the classroom setting*
Herbert Lange and Peter Ljunglöf

*Textual Features Indicative of Writing Proficiency in Elementary School Spanish Documents*
Gemma Bel-Enguix, Diana Dueñas Chavez and Arturo Curiel Díaz

*Assessment of an Index for Measuring Pronunciation Difficulty*
Katsunori Kotani and Takehiko Yoshimi

*A Short Answer Grading System in Chinese by Support Vector Approach*
Shih-Hung Wu and Wen-Feng Shih

*From Fidelity to Fluency: Natural Language Processing for Translator Training*
Oi Yee Kwong

*Countering Position Bias in Instructor Interventions in MOOC Discussion Forums*
Muthu Kumar Chandrasekaran and Min-Yen Kan

*Measuring Beginner Friendliness of Japanese Web Pages explaining Academic Concepts by Integrating Neural Image Feature and Text Features*
Hayato Shiokawa, Kota Kawaguchi, Bingcai Han, Takehito Utsuro, Yasuhide Kawada, Masaharu Yoshioka and Noriko Kando

*Learning to Automatically Generate Fill-In-The-Blank Quizzes*
Edison Marrese-Taylor, Ai Nakajima, Yutaka Matsuo and Ono Yuichi

*Multilingual Short Text Responses Clustering for Mobile Educational Activities: a Preliminary Exploration*
Yuen-Hsien Tseng, Lung-Hao Lee, Yu-Ta Chien, Chun-Yen Chang and Tsung-Yen Li

*Chinese Grammatical Error Diagnosis Based on CRF and LSTM-CRF model*
Yujie Zhou, Yinan Shao and Yong Zhou

*Contextualized Character Representation for Chinese Grammatical Error Diagnosis*
Jianbo Zhao, Si Li and Zhiqing Lin

**Thursday, July 19, 2018 (continued)**

*CMMC-BDRC Solution to the NLP-TEA-2018 Chinese Grammatical Error Diagnosis Task*
Zhang Yongwei, Hu Qinan, Liu Fang and Gu Yueguo

*Detecting Simultaneously Chinese Grammar Errors Based on a BiLSTM-CRF Model*
Yajun Liu, Hongying Zan, Mengjie Zhong and Hongchao Ma

*A Hybrid Approach Combining Statistical Knowledge with Conditional Random Fields for Chinese Grammatical Error Detection*
Yiyi Wang and Chilin Shih

*CYUT-III Team Chinese Grammatical Error Diagnosis System Report in NLPTEA-2018 CGED Shared Task*
Shih-Hung Wu, JUN-WEI WANG, Liang-Pu Chen and Ping-Che Yang

*Detecting Grammatical Errors in the NTOU CGED System by Identifying Frequent Subsentences*
Chuan-Jie Lin and Shao-Heng Chen

**17:00–17:10 Closing Remarks**