

CoNLL-2014

**Eighteenth Conference on  
Computational Natural Language Learning**

**Proceedings of the Shared Task**

June 26-27, 2014  
Baltimore, Maryland, USA

©2014 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](mailto:acl@aclweb.org)

ISBN 978-1-941643-19-8

## Introduction

This volume contains papers describing the CoNLL-2014 Shared Task and the participating systems. This year, we continue the tradition of the Conference on Computational Natural Language Learning (CoNLL) of having a high profile shared task in natural language processing, centered on automatic grammatical error correction of English essays. The grammatical error correction task is impactful since it is estimated that hundreds of millions of people in the world are learning English as a second language, and they benefit directly from an automated grammar checker.

This task is a continuation of the CoNLL shared task in 2013. We have only one track in which shared task participants are provided with an annotated training corpus, but are allowed to use additional resources as long as they are publicly available. The training corpus, NUCLE (NUS Corpus of Learner English), is a large collection of English essays written by students at the National University of Singapore (NUS) who are non-native speakers of English. The essays were annotated by professional English instructors at the NUS. As in other shared tasks, we provide a common test set with gold-standard annotations, and a scorer to evaluate the submitted system output.

This year's shared task requires a participating system to correct all error types present in an essay, instead of only the five error types in the CoNLL-2013 shared task. Also, the evaluation metric has been changed to  $F_{0.5}$ , weighting precision twice as much as recall.

A total of 13 participating teams submitted system output and 12 of them submitted system description papers. Many different approaches were adopted to perform grammatical error correction. We hope that these approaches help to advance the state of the art in grammatical error correction, and that the test set and scorer, which are freely available after the shared task, can be useful resources for those interested in grammatical error correction.

Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant  
Organizers of the CoNLL-2014 Shared Task  
May 2014



**Organizers:**

Hwee Tou Ng, National University of Singapore  
Siew Mei Wu, National University of Singapore  
Ted Briscoe, University of Cambridge  
Christian Hadiwinoto, National University of Singapore  
Raymond Hendy Susanto, National University of Singapore  
Christopher Bryant, National University of Singapore

**Program Committee:**

Pushpak Bhattacharyya, Indian Institute of Technology Bombay  
Francis Bond, Nanyang Technological University  
Aoife Cahill, Educational Testing Service  
Martin Chodorow, City University of New York  
Daniel Dahlmeier, SAP Singapore  
Dan Flickinger, Stanford University  
Michael Heilman, Educational Testing Service  
Gary Geunbae Lee, Pohang University of Science and Technology  
Yuji Matsumoto, Nara Institute of Science and Technology  
Detmar Meurers, Universität Tübingen  
Alla Rozovskaya, Columbia University  
Joel Tetreault, Yahoo! Labs  
Antal van den Bosch, Radboud University Nijmegen  
Torsten Zesch, University of Duisburg-Essen

**Additional Reviewer:**

Peter Berck, Tilburg University



## Table of Contents

<i>The CoNLL-2014 Shared Task on Grammatical Error Correction</i>	
Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Henty Susanto and Christopher Bryant .....	1
<i>Grammatical error correction using hybrid systems and type filtering</i>	
Mariano Felice, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis and Ekaterina Kochmar	15
<i>The AMU System in the CoNLL-2014 Shared Task: Grammatical Error Correction by Data-Intensive and Feature-Rich Statistical Machine Translation</i>	
Marcin Junczys-Dowmunt and Roman Grundkiewicz.....	25
<i>The Illinois-Columbia System in the CoNLL-2014 Shared Task</i>	
Alla Rozovskaya, Kai-Wei Chang, Mark Sammons, Dan Roth and Nizar Habash.....	34
<i>RACAI GEC – A hybrid approach to Grammatical Error Correction</i>	
Tiberiu Boroş, Stefan Daniel Dumitrescu, Adrian Zafiu, Verginica Barbu Mititelu and Ionut Paul Vaduva .....	43
<i>Grammatical Error Detection Using Tagger Disagreement</i>	
Anubhav Gupta.....	49
<i>CoNLL 2014 Shared Task: Grammatical Error Correction with a Syntactic N-gram Language Model from a Big Corpora</i>	
S. David Hdez. and Hiram Calvo.....	53
<i>Tuning a Grammar Correction System for Increased Precision</i>	
Anoop Kunchukuttan, Sriram Chaudhury and Pushpak Bhattacharyya.....	60
<i>POSTECH Grammatical Error Correction System in the CoNLL-2014 Shared Task</i>	
Kyusong Lee and Gary Geunbae Lee .....	65
<i>Grammatical Error Detection and Correction using a Single Maximum Entropy Model</i>	
Peilu Wang, Zhongye Jia and Hai Zhao .....	74
<i>Factored Statistical Machine Translation for Grammatical Error Correction</i>	
Yiming Wang, Longyue Wang, Xiaodong Zeng, Derek F. Wong, Lidia S. Chao and Yi Lu.....	83
<i>NTHU at the CoNLL-2014 Shared Task</i>	
Jian-Cheng Wu, Tzu-Hsi Yen, Jim Chang, Guan-Cheng Huang, Jimmy Chang, Hsiang-Ling Hsu, Yu-Wei Chang and Jason S. Chang .....	91
<i>A Unified Framework for Grammar Error Correction</i>	
Longkai Zhang and Houfeng WANG .....	96





# Conference Program

**Thursday, June 26, 2014**

## **Session 1: Oral Presentation**

- 11:00–11:30 *The CoNLL-2014 Shared Task on Grammatical Error Correction*  
Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto and Christopher Bryant
- 11:30–11:50 *Grammatical error correction using hybrid systems and type filtering*  
Mariano Felice, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis and Ekaterina Kochmar
- 11:50–12:10 *The AMU System in the CoNLL-2014 Shared Task: Grammatical Error Correction by Data-Intensive and Feature-Rich Statistical Machine Translation*  
Marcin Junczys-Dowmunt and Roman Grundkiewicz
- 12:10–12:30 *The Illinois-Columbia System in the CoNLL-2014 Shared Task*  
Alla Rozovskaya, Kai-Wei Chang, Mark Sammons, Dan Roth and Nizar Habash

Thursday, June 26, 2014 (continued)

**Session 2: Poster Presentation**

- 15:30–17:00 *RACAI GEC – A hybrid approach to Grammatical Error Correction*  
Tiberiu Boros, Stefan Daniel Dumitrescu, Adrian Zafiu, Verginica Barbu Mititelu and Ionut Paul Vaduva
- Grammatical error correction using hybrid systems and type filtering*  
Mariano Felice, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis and Ekaterina Kochmar
- Grammatical Error Detection Using Tagger Disagreement*  
Anubhav Gupta
- CoNLL 2014 Shared Task: Grammatical Error Correction with a Syntactic N-gram Language Model from a Big Corpora*  
S. David Hdez. and Hiram Calvo
- The AMU System in the CoNLL-2014 Shared Task: Grammatical Error Correction by Data-Intensive and Feature-Rich Statistical Machine Translation*  
Marcin Junczys-Dowmunt and Roman Grundkiewicz
- Tuning a Grammar Correction System for Increased Precision*  
Anoop Kunchukuttan, Sriram Chaudhury and Pushpak Bhattacharyya
- POSTECH Grammatical Error Correction System in the CoNLL-2014 Shared Task*  
Kyungho Lee and Gary Geunbae Lee
- The Illinois-Columbia System in the CoNLL-2014 Shared Task*  
Alla Rozovskaya, Kai-Wei Chang, Mark Sammons, Dan Roth and Nizar Habash
- Grammatical Error Detection and Correction using a Single Maximum Entropy Model*  
Peilu Wang, Zhongye Jia and Hai Zhao
- Factored Statistical Machine Translation for Grammatical Error Correction*  
Yiming Wang, Longyue Wang, Xiaodong Zeng, Derek F. Wong, Lidia S. Chao and Yi Lu
- NTHU at the CoNLL-2014 Shared Task*  
Jian-Cheng Wu, Tzu-Hsi Yen, Jim Chang, Guan-Cheng Huang, Jimmy Chang, Hsiang-Ling Hsu, Yu-Wei Chang and Jason S. Chang
- A Unified Framework for Grammar Error Correction*  
Longkai Zhang and Houfeng WANG