ACL 2019

Innovative Use of NLP for Building Educational Applications

Proceedings of the 14th Workshop

August 2, 2019 Florence, Italy

Gold Sponsors









Silver Sponsors



Bronze Sponsors



©2019 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 978-1-950737-34-5

Organizers:

Helen Yannakoudakis, University of Cambridge Ekaterina Kochmar, University of Cambridge Claudia Leacock, Grammarly Nitin Madnani, Educational Testing Service Ildikó Pilán, Develop Diverse Torsten Zesch, University of Duisburg-Essen

Program Committee:

Tazin Afrin, University of Pittsburgh David Alfter, University of Gothenburg Dimitrios Alikaniotis, Grammarly

Rajendra Banjade, Audible Inc.

Timo Baumann, Carnegie Mellon University

Lee Becker, Pearson

Beata Beigman Klebanov, Educational Testing Service

Kay Berkling, Cooperation State university Karlsruhe Germany

Suma Bhat, University of Illinois at Urbana-Champaign

Sameer Bhatnagar, Polytechnique Montreal

Joachim Bingel, University of Copenhagen

Karim Bouzoubaa, Mohammed V University in Rabat

Chris Brew, Facebook

Ted Briscoe, University of Cambridge

Julian Brooke, University of British Columbia

Dominique Brunato, Institute for Computational Linguistics, ILC-CNR, Pisa, Italy

James Bruno, Educational Testing Service

Christopher Bryant, University of Cambridge

Paula Buttery, University of Cambridge

Aoife Cahill, Educational Testing Service

Andrew Caines, University of Cambridge

Mei-Hua Chen, Department of Foreign Languages and Literature, Tunghai University

Martin Chodorow, ETS & City University of New York

Shamil Chollampatt, National University of Singapore

Mark Core, University of Southern California

Vidas Daudaravicius, UAB VTEX

Kordula De Kuthy, University of Tübingen

Carrie Demmans Epp, University of Alberta

Yo Ehara, Faculty of Informatics, Shizuoka Institute of Science and Technology

Keelan Evanini, Educational Testing Service

Mariano Felice, University of Cambridge

Michael Flor, Educational Testing Service

Thomas François, Université catholique de Louvain

Yoko Futagi, Educational Testing Service

Michael Gamon, Microsoft Research

Dipesh Gautam, The University of Memphis

Christian Gold, University of Bergen

Sian Gooding, University of Cambridge

Jonathan Gordon, Vassar College

Cyril Goutte, National Research Council Canada

Iryna Gurevych, UKP Lab, TU Darmstadt

Binod Gyawali, Educational Testing Service

Na-Rae Han, University of Pittsburgh

Jiangang Hao, Educational Testing Service

Homa Hashemi, Microsoft

Trude Heift, Simon Fraser University

Derrick Higgins, American Family Insurance

Heiko Holz, LEAD Graduate School & Research Network at the University of Tübingen

Andrea Horbach, University Duisburg-Essen

Chung-Chi Huang, Frostburg State University

Yi-Ting Huang, Academia Sinica

Radu Tudor Ionescu, University of Bucharest

Lifeng Jin, The Ohio State University

Pamela Jordan, University of Pittsburgh

Taraka Kasicheyanula, University of Oslo

Elma Kerz, RWTH Aachen

Mamoru Komachi, Tokyo Metropolitan University

Lun-Wei Ku, Academia Sinica

Chong Min Lee, Educational Testing Service

Ji-Ung Lee, UKP Lab, Technische Universität Darmstadt

John Lee, City University of Hong Kong

Lung-Hao Lee, National Central University

Ben Leong, Educational Testing Service

James Lester, North Carolina State University

Chen Liang, Facebook

Diane Litman, University of Pittsburgh

Yang Liu, Laix

Peter Ljunglöf, University of Gothenburg

Anastassia Loukina, Educational Testing Service

Xiaofei Lu, Pennsylvania State University

Luca Lugini, University of Pittsburgh

Nabin Maharjan, University of Memphis

Jean Maillard, University of Cambridge

Shervin Malmasi, Harvard Medical School

Montse Maritxalar, University of the Basque Country

Ditty Mathew, IIT Madras

Julie Medero, Harvey Mudd College

Beata Megyesi, Uppsala University

Detmar Meurers, Universität Tübingen

Margot Mieskes, University of Applied Sciences, Darmstadt

Elham Mohammadi, CLaC Laboratory, Concordia University

Maria Moritz, German Research Center for Artificial Intelligence

William Murray, Pearson

Courtney Napoles, Grammarly

Diane Napolitano, LCSR, Rutgers University

Hwee Tou Ng, National University of Singapore

Huy Nguyen, LingoChamp

Rodney Nielsen, University of North Texas

Nobal Niraula, Boeing Research and Technology

Yoo Rhee Oh, Electronics and Telecommunications Research Institute (ETRI)

Constantin Orasan, University of Wolverhampton

Ulrike Pado, HFT Stuttgart

Alexis Palmer, University of North Texas

Martí Quixal, Universität Tübingen

Vipul Raheja, Grammarly

Zahra Rahimi, University of Pittsburgh

Lakshmi Ramachandran, Amazon Search

Vikram Ramanarayanan, Educational Testing Service

Hanumant Redkar, Indian Institute of Technology Bombay

Marek Rei, University of Cambridge

Robert Reynolds, Brigham Young University

Brian Riordan, Educational Testing Service

Andrew Rosenberg, Google

Mark Rosenstein, Pearson

Alla Rozovskaya, City University of New York

C. Anton Rytting, University of Maryland

Keisuke Sakaguchi, Allen Institute for Artificial Intelligence

Allen Schmaltz, Harvard University

Serge Sharoff, University of Leeds

Swapna Somasundaran, Educational Testing Service

Richard Sproat, Google

Helmer Strik, Radboud University Nijmegen

Jan Švec, NTIS, University of West Bohemia

Anaïs Tack, UCLouvain & KU Leuven

Joel Tetreault, Grammarly

Yuen-Hsien Tseng, National Taiwan Normal University

Giulia Venturi, Institute for Computational Linguistics "A. Zampolli", Italy

Aline Villavicencio, Federal University of Rio Grande do Sul and University of Essex

Carl Vogel, Trinity College Dublin

Elena Volodina, University of Gothenburg

Shuting Wang, Facebook Inc

Xinhao Wang, Educational Testing Service

Michael White, The Ohio State University

Michael Wojatzki, LTL, University of Duisburg-Essen

Magdalena Wolska, Eberhard Karls Universität Tübingen

Huichao Xue, LinkedIn

Victoria Yaneva, National Board of Medical Examiners / University of Wolverhampton

Zheng Yuan, University of Cambridge

Marcos Zampieri, University of Wolverhampton (UK)

Klaus Zechner, Educational Testing Service

Fan Zhang, Google

Haoran Zhang, University of Pittsburgh

Ramon Ziai, University of Tübingen

Table of Contents

The many dimensions of algorithmic fairness in educational applications Anastassia Loukina, Nitin Madnani and Klaus Zechner
Predicting the Difficulty of Multiple Choice Questions in a High-stakes Medical Exam Le An Ha, Victoria Yaneva, Peter Baldwin and Janet Mee
An Intelligent Testing Strategy for Vocabulary Assessment of Chinese Second Language Learners Wei Zhou, Renfen Hu, Feipeng Sun and Ronghuai Huang
Computationally Modeling the Impact of Task-Appropriate Language Complexity and Accuracy on Human Grading of German Essays Zarah Weiss, Anja Riemenschneider, Pauline Schröter and Detmar Meurers
Analysing Rhetorical Structure as a Key Feature of Summary Coherence Jan Šnajder, Tamara Sladoljev-Agejev and Svjetlana Kolić Vehovec
The BEA-2019 Shared Task on Grammatical Error Correction Christopher Bryant, Mariano Felice, Øistein E. Andersen and Ted Briscoe
A Benchmark Corpus of English Misspellings and a Minimally-supervised Model for Spelling Correction Michael Flor, Michael Fried and Alla Rozovskaya
Artificial Error Generation with Fluency Filtering Mengyang Qiu and Jungyeul Park
Regression or classification? Automated Essay Scoring for Norwegian Stig Johan Berggren, Taraka Rama and Lilja Øvrelid
Context is Key: Grammatical Error Detection with Contextual Word Representations Samuel Bell, Helen Yannakoudakis and Marek Rei
How to account for mispellings: Quantifying the benefit of character representations in neural content scoring models Brian Riordan, Michael Flor and Robert Pugh
The Unreasonable Effectiveness of Transformer Language Models in Grammatical Error Correction Dimitris Alikaniotis, Vipul Raheja and Joel Tetreault
(Almost) Unsupervised Grammatical Error Correction using Synthetic Comparable Corpus Satoru Katsumata and Mamoru Komachi
Learning to combine Grammatical Error Corrections Yoav Kantor, Yoav Katz, Leshem Choshen, Edo Cohen-Karlik, Naftali Liberman, Assaf Toledo, Amir Menczel and Noam Slonim
Erroneous data generation for Grammatical Error Correction Shuyao Xu, Jiehao Zhang, Jin Chen and Long Qin
The LAIX Systems in the BEA-2019 GEC Shared Task Ruobing Li, Chuan Wang, Yefei Zha, Yonghong Yu, Shiman Guo, Qiang Wang, Yang Liu and Hui
Lin

The CUED's Grammatical Error Correction Systems for BEA-2019 Felix Stahlberg and Bill Byrne
The AIP-Tohoku System at the BEA-2019 Shared Task Hiroki Asano, Masato Mita, Tomoya Mizumoto and Jun Suzuki
CUNI System for the Building Educational Applications 2019 Shared Task: Grammatical Error Correction
Jakub Náplava and Milan Straka
Noisy Channel for Low Resource Grammatical Error Correction Ophélie Lacroix, Simon Flachs and Anders Søgaard
The BLCU System in the BEA 2019 Shared Task Liner Yang and Chencheng Wang
TMU Transformer System Using BERT for Re-ranking at BEA 2019 Grammatical Error Correction or Restricted Track
Masahiro Kaneko, Kengo Hotate, Satoru Katsumata and Mamoru Komachi
A Neural Grammatical Error Correction System Built On Better Pre-training and Sequential Transfer Learning Yo Joong Choe, Jiyeon Ham, Kyubyong Park and Yeoil Yoon
Neural and FST-based approaches to grammatical error correction Zheng Yuan, Felix Stahlberg, Marek Rei, Bill Byrne and Helen Yannakoudakis
Improving Precision of Grammatical Error Correction with a Cheat Sheet Mengyang Qiu, Xuejiao Chen, Maggie Liu, Krishna Parvathala, Apurva Patil and Jungyeul Park 240
Multi-headed Architecture Based on BERT for Grammatical Errors Correction Bohdan Didenko and Julia Shaptala
Neural Grammatical Error Correction Systems with Unsupervised Pre-training on Synthetic Data Roman Grundkiewicz, Marcin Junczys-Dowmunt and Kenneth Heafield
Evaluation of automatic collocation extraction methods for language learning Vishal Bhalla and Klara Klimcikova
Anglicized Words and Misspelled Cognates in Native Language Identification Ilia Markov, Vivi Nastase and Carlo Strapparava
Linguistically-Driven Strategy for Concept Prerequisites Learning on Italian Alessio Miaschi, Chiara Alzetta, Franco Alberto Cardillo and Felice Dell'Orletta
Grammatical-Error-Aware Incorrect Example Retrieval System for Learners of Japanese as a Second Language Mio Arai, Masahiro Kaneko and Mamoru Komachi
Toward Automated Content Feedback Generation for Non-native Spontaneous Speech Su-Youn Yoon, Ching-Ni Hsieh, Klaus Zechner, Matthew Mulholland, Yuan Wang and Nitin Madnani

Analytic Score Prediction and Justification Identification in Automated Short Answer Scoring Tomoya Mizumoto, Hiroki Ouchi, Yoriko Isobe, Paul Reisert, Ryo Nagata, Satoshi Sekine and Kentaro Inui
Content Customization for Micro Learning using Human Augmented AI Techniques Ayush Shah, Tamer Abuelsaad, Jae-Wook Ahn, Prasenjit Dey, Ravi Kokku, Ruhi Sharma Mittal Aditya Vempaty and Mourvi Sharma
Curio SmartChat: A system for Natural Language Question Answering for Self-Paced K-12 Learning Srikrishna Raamadhurai, Ryan Baker and Vikraman Poduval
Supporting content evaluation of student summaries by Idea Unit embedding Marcello Gecchele, Hiroaki Yamada, Takenobu Tokunaga and Yasuyo Sawaki
On Understanding the Relation between Expert Annotations of Text Readability and Target Reader Comprehension
Sowmya Vajjala and Ivana Lucic
Measuring Text Complexity for Italian as a Second Language Learning Purposes Luciana Forti, Alfredo Milani, Luisa Piersanti, Filippo Santarelli, Valentino Santucci and Stefania Spina
Simple Construction of Mixed-Language Texts for Vocabulary Learning Adithya Renduchintala, Philipp Koehn and Jason Eisner
Analyzing Linguistic Complexity and Accuracy in Academic Language Development of German across Elementary and Secondary School Zarah Weiss and Detmar Meurers
Content Modeling for Automated Oral Proficiency Scoring System Su-Youn Yoon and Chong Min Lee
Learning Outcomes and Their Relatedness in a Medical Curriculum Sneha Mondal, Tejas Dhamecha, Shantanu Godbole, Smriti Pathak, Red Mendoza, K Gayathr. Wijayarathna, Nabil Zary, Swarnadeep Saha and Malolan Chetlur
Measuring text readability with machine comprehension: a pilot study
Marc Benzahra and François Yvon
Metaphors in Text Simplification: To change or not to change, that is the question Yulia Clausen and Vivi Nastase
Application of an Automatic Plagiarism Detection System in a Large-scale Assessment of English Speaking Proficiency Xinhao Wang, Keelan Evanini, Matthew Mulholland, Yao Qian and James V. Bruno435
Equity Beyond Bias in Language Technologies for Education Elijah Mayfield, Michael Madaio, Shrimai Prabhumoye, David Gerritsen, Brittany McLaughlin Ezekiel Dixon-Román and Alan W Black
From Receptive to Productive: Learning to Use Confusing Words through Automatically Selected Example Sentences Chieh-Yang Huang, Yi-Ting Huang, MeiHua Chen and Lun-Wei Ku

Equipping Educational Applications with Domain Knowledge Tarek Sakakini, Hongyu Gong, Jong Yoon Lee, Robert Schloss, JinJun Xiong and Suma Bhat. 472
The Unbearable Weight of Generating Artificial Errors for Grammatical Error Correction Phu Mon Htut and Joel Tetreault
Automated Essay Scoring with Discourse-Aware Neural Models Farah Nadeem, Huy Nguyen, Yang Liu and Mari Ostendorf
Modeling language learning using specialized Elo rating Jue Hou, Koppatz Maximilian, José María Hoya Quecedo, Nataliya Stoyanova and Roman Yangarber
Rubric Reliability and Annotation of Content and Argument in Source-Based Argument Essays Yanjun Gao, Alex Driban, Brennan Xavier McManus, Elena Musi, Patricia Davies, Smaranda Muresan and Rebecca J. Passonneau

Workshop Program

Friday, August 2, 2019

8:30-9:00	Loading of Oral Presentations
9:00-9:15	Opening Remarks
9:15-10:30	Oral Presentations 1
9:15–9:40	The many dimensions of algorithmic fairness in educational applications Anastassia Loukina, Nitin Madnani and Klaus Zechner
9:40–10:05	Predicting the Difficulty of Multiple Choice Questions in a High-stakes Medical Exam
	Le An Ha, Victoria Yaneva, Peter Baldwin and Janet Mee
10:05-10:30	Ambassador paper: Effects of the self-view window during video-mediated survey interviews: An eye-tracking study, Shelley Feuer
10:30-11:00	Coffee Breaks
11:00-12:30	Oral Presentations 2
11:00–11:25	An Intelligent Testing Strategy for Vocabulary Assessment of Chinese Second Language Learners Wei Zhou, Renfen Hu, Feipeng Sun and Ronghuai Huang
11:25–11:50	Computationally Modeling the Impact of Task-Appropriate Language Complexity and Accuracy on Human Grading of German Essays Zarah Weiss, Anja Riemenschneider, Pauline Schröter and Detmar Meurers
11:50–12:10	Analysing Rhetorical Structure as a Key Feature of Summary Coherence Jan Šnajder, Tamara Sladoljev-Agejev and Svjetlana Kolić Vehovec

12:10–12:30 Shared Task Report

*The BEA-2019 Shared Task on Grammatical Error Correction*Christopher Bryant, Mariano Felice, Øistein E. Andersen and Ted Briscoe

12:30-14:00 Lunch

14:00-14:45 **Poster Session A**

A Benchmark Corpus of English Misspellings and a Minimally-supervised Model for Spelling Correction

Michael Flor, Michael Fried and Alla Rozovskaya

Artificial Error Generation with Fluency Filtering

Mengyang Qiu and Jungyeul Park

Regression or classification? Automated Essay Scoring for Norwegian

Stig Johan Berggren, Taraka Rama and Lilja Øvrelid

Context is Key: Grammatical Error Detection with Contextual Word Representations

Samuel Bell, Helen Yannakoudakis and Marek Rei

How to account for mispellings: Quantifying the benefit of character representations in neural content scoring models

Brian Riordan, Michael Flor and Robert Pugh

The Unreasonable Effectiveness of Transformer Language Models in Grammatical Error Correction

Dimitris Alikaniotis, Vipul Raheja and Joel Tetreault

(Almost) Unsupervised Grammatical Error Correction using Synthetic Comparable Corpus

Satoru Katsumata and Mamoru Komachi

Learning to combine Grammatical Error Corrections

Yoav Kantor, Yoav Katz, Leshem Choshen, Edo Cohen-Karlik, Naftali Liberman, Assaf Toledo, Amir Menczel and Noam Slonim

Erroneous data generation for Grammatical Error Correction

Shuyao Xu, Jiehao Zhang, Jin Chen and Long Qin

The LAIX Systems in the BEA-2019 GEC Shared Task

Ruobing Li, Chuan Wang, Yefei Zha, Yonghong Yu, Shiman Guo, Qiang Wang, Yang Liu and Hui Lin

The CUED's Grammatical Error Correction Systems for BEA-2019

Felix Stahlberg and Bill Byrne

The AIP-Tohoku System at the BEA-2019 Shared Task

Hiroki Asano, Masato Mita, Tomoya Mizumoto and Jun Suzuki

CUNI System for the Building Educational Applications 2019 Shared Task: Grammatical Error Correction

Jakub Náplava and Milan Straka

Noisy Channel for Low Resource Grammatical Error Correction

Ophélie Lacroix, Simon Flachs and Anders Søgaard

The BLCU System in the BEA 2019 Shared Task

Liner Yang and Chencheng Wang

TMU Transformer System Using BERT for Re-ranking at BEA 2019 Grammatical Error Correction on Restricted Track

Masahiro Kaneko, Kengo Hotate, Satoru Katsumata and Mamoru Komachi

A Neural Grammatical Error Correction System Built On Better Pre-training and Sequential Transfer Learning

Yo Joong Choe, Jiyeon Ham, Kyubyong Park and Yeoil Yoon

Neural and FST-based approaches to grammatical error correction

Zheng Yuan, Felix Stahlberg, Marek Rei, Bill Byrne and Helen Yannakoudakis

Improving Precision of Grammatical Error Correction with a Cheat Sheet

Mengyang Qiu, Xuejiao Chen, Maggie Liu, Krishna Parvathala, Apurva Patil and Jungyeul Park

Multi-headed Architecture Based on BERT for Grammatical Errors Correction

Bohdan Didenko and Julia Shaptala

Neural Grammatical Error Correction Systems with Unsupervised Pre-training on Synthetic Data

Roman Grundkiewicz, Marcin Junczys-Dowmunt and Kenneth Heafield

14:45–15:30 **Poster Session B**

Evaluation of automatic collocation extraction methods for language learning Vishal Bhalla and Klara Klimcikova

Anglicized Words and Misspelled Cognates in Native Language Identification Ilia Markov, Vivi Nastase and Carlo Strapparava

Linguistically-Driven Strategy for Concept Prerequisites Learning on Italian
Alessio Miaschi, Chiara Alzetta, Franco Alberto Cardillo and Felice Dell'Orletta

Grammatical-Error-Aware Incorrect Example Retrieval System for Learners of Japanese as a Second Language

Mio Arai, Masahiro Kaneko and Mamoru Komachi

Toward Automated Content Feedback Generation for Non-native Spontaneous Speech

Su-Youn Yoon, Ching-Ni Hsieh, Klaus Zechner, Matthew Mulholland, Yuan Wang and Nitin Madnani

Analytic Score Prediction and Justification Identification in Automated Short Answer Scoring

Tomoya Mizumoto, Hiroki Ouchi, Yoriko Isobe, Paul Reisert, Ryo Nagata, Satoshi Sekine and Kentaro Inui

Content Customization for Micro Learning using Human Augmented AI Techniques Ayush Shah, Tamer Abuelsaad, Jae-Wook Ahn, Prasenjit Dey, Ravi Kokku, Ruhi Sharma Mittal, Aditya Vempaty and Mourvi Sharma

Curio SmartChat: A system for Natural Language Question Answering for Self-Paced K-12 Learning

Srikrishna Raamadhurai, Ryan Baker and Vikraman Poduval

Supporting content evaluation of student summaries by Idea Unit embedding Marcello Gecchele, Hiroaki Yamada, Takenobu Tokunaga and Yasuyo Sawaki

On Understanding the Relation between Expert Annotations of Text Readability and Target Reader Comprehension

Sowmya Vajjala and Ivana Lucic

Measuring Text Complexity for Italian as a Second Language Learning Purposes Luciana Forti, Alfredo Milani, Luisa Piersanti, Filippo Santarelli, Valentino Santucci and Stefania Spina

Simple Construction of Mixed-Language Texts for Vocabulary Learning Adithya Renduchintala, Philipp Koehn and Jason Eisner

Analyzing Linguistic Complexity and Accuracy in Academic Language Development of German across Elementary and Secondary School Zarah Weiss and Detmar Meurers

Content Modeling for Automated Oral Proficiency Scoring System Su-Youn Yoon and Chong Min Lee

Learning Outcomes and Their Relatedness in a Medical Curriculum

Sneha Mondal, Tejas Dhamecha, Shantanu Godbole, Smriti Pathak, Red Mendoza, K Gayathri Wijayarathna, Nabil Zary, Swarnadeep Saha and Malolan Chetlur

Measuring text readability with machine comprehension: a pilot study Marc Benzahra and François Yvon

Metaphors in Text Simplification: To change or not to change, that is the question Yulia Clausen and Vivi Nastase

Application of an Automatic Plagiarism Detection System in a Large-scale Assessment of English Speaking Proficiency

Xinhao Wang, Keelan Evanini, Matthew Mulholland, Yao Qian and James V. Bruno

Equity Beyond Bias in Language Technologies for Education

Elijah Mayfield, Michael Madaio, Shrimai Prabhumoye, David Gerritsen, Brittany McLaughlin, Ezekiel Dixon-Román and Alan W Black

From Receptive to Productive: Learning to Use Confusing Words through Automatically Selected Example Sentences

Chieh-Yang Huang, Yi-Ting Huang, MeiHua Chen and Lun-Wei Ku

Equipping Educational Applications with Domain Knowledge

Tarek Sakakini, Hongyu Gong, Jong Yoon Lee, Robert Schloss, JinJun Xiong and Suma Bhat

The Unbearable Weight of Generating Artificial Errors for Grammatical Error Correction

Phu Mon Htut and Joel Tetreault

15:30-16:00 Mid-Afternoon Snacks

16:00–17:15 Oral Presentations **3**

16:00–16:25 Automated Essay Scoring with Discourse-Aware Neural Models
Farah Nadeem, Huy Nguyen, Yang Liu and Mari Ostendorf

Smaranda Muresan and Rebecca J. Passonneau

- 16:25–16:50 *Modeling language learning using specialized Elo rating*Jue Hou, Koppatz Maximilian, José María Hoya Quecedo, Nataliya Stoyanova and Roman Yangarber
- 16:50–17:15 Rubric Reliability and Annotation of Content and Argument in Source-Based Argument Essays

 Yanjun Gao, Alex Driban, Brennan Xavier McManus, Elena Musi, Patricia Davies,

17:15-17:30 Closing Remarks