

The 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)

WORKSHOP DESCRIPTION

The BEA Workshop, one of the largest one-day workshops in the ACL community, is a leading venue for NLP innovation for educational applications. The workshop's continuous growth illustrates an alignment between societal need and technological advances. NLP capabilities now support an array of learning domains, including writing, speaking, reading, science, and mathematics, as well as the related intra- (e.g., self-confidence) and inter-personal (e.g., collaboration) domains that support achievement in those domains. Automated writing evaluation (AWE) and speech scoring applications are commercially deployed in high-stakes assessment and instructional contexts, including primary and secondary educational settings. Commercially-deployed plagiarism detection is also prevalent. The current educational and assessment landscape in K-12, higher education, and adult learning (in academic and workplace settings) fosters a strong interest in technologies that yield analytics to support proficiency measures for complex constructs. For writing, there is a focus on innovation that supports writing tasks requiring source use, argumentative discourse, and factual content accuracy. For speech, there is an interest in advancing automated scoring to include the evaluation of discourse and content features in responses to spoken assessments. General advances in speech technology have promoted a renewed interest in spoken dialog and multimodal systems for instruction and assessment in, for example, workplace interviews and simulated teaching environments. The explosive growth of mobile applications for game-based and simulation applications for instruction and assessment is another place where NLP has begun to play a large role, especially in language learning.

NLP for educational applications has gained visibility outside of the NLP community. First, the Hewlett Foundation reached out to public and private sectors and sponsored two competitions – one for automated essay scoring, and another for scoring of short response items – to engage the larger scientific community in this enterprise. Subsequently, EdX, a non-profit enterprise founded by Harvard and MIT, announced the release of software designed to automatically grade text using AI technologies. Learning@Scale, a relatively new venue for NLP research in education, promotes interdisciplinary research on learning and teaching. Massive Open Online Courses (MOOCs) now incorporate AWE systems to manage the thousands of assignments that may be received during a single MOOC course. [MOOCs for Refugees](#) have more recently popped up in response to current social situations. Courses include language learning, and AWE and other NLP capabilities could support the coursework.

Research into education-related problems has also continued to grow within the NLP community. The last five years saw eight shared tasks (including five on grammatical error correction), two workshops on NLP and Computer-Assisted Language Learning ([TEA series](#)), and two sessions of the ACL devoted exclusively to the role of NLP in assisting language learners. Three of the shared tasks were sponsored by BEA. At ACL 2016, there were over 12 papers devoted to NLP in education, the highest number ever at a conference of this caliber. All of this has served to increase the visibility of, and interest in, the field of education applications in NLP.

The 12th workshop will have oral presentation sessions and a large poster session in order to maximize the amount of original work presented (the workshop normally has 50+ attendees). We will invite both full papers and short papers on topics including: automated scoring of textual and spoken responses, intelligent tutoring, peer review, grammatical error detection/correction, learner cognition, spoken dialog, multimodal applications, tools for teachers and test developers, and use of corpora. Research that incorporates NLP for use with mobile and game-based platforms will be of special interest. Specific topics include:

Automated scoring/evaluation for written student responses

- Content analysis for scoring/assessment
- Analysis of the structure of argumentation
- Grammatical error detection and correction
- Discourse and stylistic analysis
- Plagiarism detection
- Machine translation for assessment, instruction and curriculum development
- Detection of non-literal language (e.g., metaphor) and sentiment analysis
- Non-traditional genres (beyond essays)

Intelligent Tutoring (IT), Collaborative Learning Environments

- Game-based learning
- Dialogue systems in education
- Hypothesis formation and testing
- Multi-modal communication between students and computers
- Generation of tutorial responses
- Knowledge representation in learning systems
- Concept visualization in learning systems

Learner cognition

- Assessment of learners' language and cognitive skill levels
- Systems that detect and adapt to learners' cognitive or emotional states
- Tools for learners with special needs

Use of corpora in educational tools

- Data mining of learner and other corpora for tool building
- Annotation standards and schemas / annotator agreement

Tools and applications for classroom teachers and/or test developers

- NLP tools for second and foreign language learners
- Semantic-based access to instructional materials to identify appropriate texts
- Tools that automatically generate test questions
- Processing of and access to lecture materials across topics and genres
- Adaptation of instructional text to individual learners' grade levels
- Tools for text-based curriculum development
- E-learning tools for personalized course content
- Language-based educational games

SHARED TASK

The workshop will also host the 2nd Shared Task on Native Language Identification. The [first edition](#), also held with the BEA workshop in 2013, was a total success with 29 teams competing, making it one of the largest shared tasks that year. Native Language Identification (NLI) is the goal of automatically classifying the native language (L1) of a writer based solely on his or her essay writing in another language. For the second edition, we are partnering with speech researchers and are modifying the task to make the prediction based not only an essay sample but also a speech sample where the same writer is asked to orally produce a response to a short answer essay question. There will be three tracks for the task: NLI on the essay only, NLI on the speech response only, and NLI using both responses. We feel this will make for a more challenging shared task while building on the methods and results from 2013.

ESTIMATED NUMBER OF ATTENDEES

We expect 65 attendees to the workshop. This is based on previous years where we typically have 50 attendees. With the inclusion of a shared task, we expect at least 15 extra attendees.

WORKSHOP LENGTH

We expect the workshop to be one full day, starting at (approximately) 9:00 and ending at (approximately) 5:00 pm.

SPECIAL TECHNICAL REQUIREMENTS

We have no special requirements besides poster stands and internet access for any demos.

VENUE PREFERENCE

Our venue preference is ACL in Vancouver.

WORKSHOP ORGANIZERS

Joel Tetreault (Primary Contact) is Director of Research at Grammarly. His research focus is Natural Language Processing with specific interests in anaphora, dialogue and discourse processing, machine learning, and applying these techniques to the analysis of English language learning, automated essay scoring among others. Currently he works on the research and development of NLP tools and components for the next generation of intelligent writing assistance systems. Prior to joining Grammarly, he was a Senior Research Scientist at Yahoo Labs, Senior Principal Manager of the Core Natural Language group at Nuance Communications, Inc., and worked at Educational Testing Service for six years as a managing

research scientist where he researched automated methods for detecting grammatical errors by non-native speakers, plagiarism detection, and content scoring. Tetreault received his B.A. in Computer Science from Harvard University and his M.S. and Ph.D. in Computer Science from the University of Rochester. He was also a postdoctoral research scientist at the University of Pittsburgh's Learning Research and Development Center, where he worked on developing spoken dialogue tutoring systems. In addition, he has co-organized the Building Educational Application workshop series for 8 years, several shared tasks, and is currently NAACL Treasurer.

Contact Information:

Grammarly

548 Market Street

San Francisco, CA, 94104

USA

bea.nlp.workshop@gmail.com (primary workshop contact email address)

tetreaul@gmail.com

phone: +1 310-612-0315

Jill Burstein is the Research Director of the NLP Group at Educational Testing Service. Her background and expertise is in computational linguistics with a focus on education technology for writing, reading, and teacher professional development. She holds more than several patents for educational technology inventions. Her inventions include e-rater[®], an automated essay scoring system and an essay-based discourse analysis system – both used in *Criterion*[®], ETS' online essay evaluation system. Her current research uses NLP to develop capabilities to support English learners. She is the principal investigator on a grant funded by the Institute of Education Sciences (U.S. Department of Education) entitled: *A Technology-Rich Teacher Professional Development Intervention that Supports Content-based Curriculum Development for English Language Learners. The Language Muse[™] Activity Palette*, developed from this grant, automatically generates language-based activities to support reading and language skills for English learners. Dr. Burstein received her B.A. in Linguistics and Spanish from New York University and her M.A. and Ph.D. in Linguistics from the Graduate Center, City University of New York.

Contact Information:

Educational Testing Service

MS 12R

Princeton, NJ 08541 USA

jburstein@ets.org

phone: 609-734-5823

<https://sites.google.com/site/jbursteinets/>

Claudia Leacock is a consultant who has been working on using NLP in educational applications for over 20 years, focusing on automated scoring and grammatical error detection.

She was previously a consultant for Microsoft Research where she collaborated on the development of *ESL Assistant*: a web-based prototype tool for detecting and correcting grammatical errors of English language learners. As a Distinguished Member of Technical Staff at Pearson Knowledge Technologies, and previously as a Principal Development Scientist at Educational Testing Service, she pioneered the development of tools for automated assessment of short-response content-based questions, and for grammatical error detection still used in e-rater, ETS' essay scoring engine. As a member of the WordNet group at Princeton University's Cognitive Science Lab, her research focused on word sense disambiguation. Dr. Leacock received a B.A. in English from NYU, a Ph.D. in Linguistics from the City University of New York, Graduate Center, and was a post-doctoral fellow at IBM, T.J. Watson Research Center.

Contact Information:

80 LaSalle Street, Apt. 10F
New York, NY 10027
Claudia.leacock@gmail.com
phone: 212-866-1638

Helen Yannakoudakis is a Senior Research Associate at the Computer Laboratory of the University of Cambridge, working on Automated Language Teaching and Assessment (ALTA). She is also a Teaching Fellow at Girton college Cambridge, and a committee member of Women@CL. Previously, she was a postdoctoral research fellow at Cambridge English Language Assessment. She holds a PhD in Natural Language and Information Processing from the University of Cambridge, during which she also worked on the English Profile Programme (EPP) in collaboration with Cambridge Assessment; an MPhil in Computer Speech, Text and Internet Technology; and a BSc in Computer Science. Her research interests lie at the intersection of Computational Linguistics, Machine Learning and Visualisation, and particularly revolve around the areas of automated language teaching and assessment, self-assessment and tutoring systems, error detection and correction, and second language acquisition.

Contact Information:

University of Cambridge
Computer Laboratory
William Gates Building
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK
helen.yannakoudakis@cl.cam.ac.uk
phone: +44 (0)1223 763536

PROGRAM COMMITTEE

The following **139 scientists** are all confirmed reviewers. This PC is a new record for us.

David Alfter, University of Gothenburg, Sweden
Dimitrios Alikaniotis, University of Cambridge, UK
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Rajendra Banjade, University of Memphis, USA
Timo Baumann, Universität Hamburg, Germany
Lee Becker, Hapara, USA
Beata Beigman Klebanov, Educational Testing Service, USA
Lisa Beinborn, Technische Universität Darmstadt, Germany
Kay Berkling, Cooperative State University, Karlsruhe, Germany
Delphine Bernhard, LiLPa, Université de Strasbourg, France
Yevgeni Berzak, MIT, USA
Sameer Bhatnagar, Polytechnique Montreal, Canada
Serge Bibauw, KU Leuven & Université Catholique de Louvain, Belgium
Kristy Boyer, University of Florida, USA
Chris Brew, Digital Operatives LLC, USA
Ted Briscoe, University of Cambridge, UK
Chris Brockett, Microsoft Research, USA
Julian Brooke, University of Melbourne, Australia
Dominique Brunato, Istituto di Linguistica Computazionale (ILC-CNR), National Council of Research, Pisa (Italy), Italy
Christopher Bryant, University of Cambridge, UK
Andrew Caines, University of Cambridge, UK
Xiaoabin Chen, Tübingen University, Germany
Wei-Fan Chen, Academia Sinica, Taiwan
Lei Chen, Educational Testing Service, USA
Yeonsuk Cho, Educational Testing Service, USA
Martin Chodorow, City University of New York, USA
Shamil Chollampatt, National University of Singapore, Singapore
Mark Core, University of Southern California, USA
Scott Crossley, Georgia State University, USA
Ronan Cummins, University of Cambridge, UK
Luis Fernando D'Haro, Institute for Infocomm Research, A*STAR, Singapore
Vidas Daudaravicius, VTEX Research, Lithuania

Markus Dickinson, Indiana University, USA
Yo Ehara, National Institute of Advanced Industrial Science and Technology, Japan
Keelan Evanini, Educational Testing Service, USA
Mariano Felice, University of Cambridge, UK
Michael Flor, Educational Testing Service, USA
Peter Foltz, Pearson, USA
Hector-Hugo Franco-Penya, DIT, Ireland
Thomas François, Université Catholique de Louvain, Belgium
Michael Gamon, Microsoft Research, USA
Dipesh Gautam, University of Memphis, Nepal
Binyam Gebrekidan Gebre, Philips, Netherlands
Kallirroï Georgila, University of Southern California, USA
Cyril Goutte, National Research Council Canada, Canada
Iryna Gurevych, TU Darmstadt, Germany
Na-Rae Han, University of Pittsburgh, USA
Jiangang Hao, Educational Testing Service, USA
Rachel Harsley, University of Illinois at Chicago, USA
Homa B. Hashemi, University of Pittsburgh, USA
Trude Heift, Simon Fraser University, Canada
Derrick Higgins, American Family Insurance, USA
Andrea Horbach, Saarland University, Germany
Chung-Chi Huang, Frostburg State University, USA
Radu Tudor Ionescu, University of Bucharest, Romania
Ross Israel, Factual, USA
Pamela Jordan, University of Pittsburgh, USA
Fazel Keshtkar, St. John's University Professor, USA
Levi King, Indiana University, USA
Sigrid Klerke, University of Copenhagen, Denmark
Ekaterina Kochmar, University of Cambridge, UK
Mamoru Komachi, Tokyo Metropolitan University, Japan
Robert Krovetz, Lexical Research, USA
Girish Kumar, Stanford University, Singapore
Lun-Wei Ku, IIS, Academia Sinica, Taiwan
Kristopher Kyle, University of Hawaii, USA
Lung-Hao Lee, National Taiwan Normal University, Taiwan
John Lee, City University of Hong Kong, Hong Kong
James Lester, North Carolina State University, USA
Chen Wee Leong, Educational Testing Service, USA
Baoli Li, Henan University of Technology, China
Diane Litman, University of Pittsburgh, USA
Annie Louis, University of Essex, UK
Anastassia Loukina, Educational Testing Service, USA
Xiaofei Lu, Pennsylvania State University, USA
Fabiana MacMillan, Rosetta Stone, USA
Nitin Madnani, Educational Testing Service, USA

Shervin Malmasi, Harvard Medical School, USA
Liliana Mamani Sanchez, University College Dublin, Ireland
Montse Maritxalar, University of the Basque Country, Spain
Ditty Matthew, Indian Institute of Technology Madras, India
Julie Medero, Harvey Mudd College, USA
Beata Megyesi, Uppsala University, Sweden
Mohsen Mesgar, Heidelberg Institute for Theoretical Studies, Germany
Angeliki Metallinou, Amazon, USA
Detmar Meurers, University of Tübingen, Germany
Christian M. Meyer, Technische Universität Darmstadt, Germany
Lisa Michaud, Aspect Software, USA
Michael Mohler, Language Computer Corporation, USA
Maria Moritz, Faculty of Mathematics and Computer Science, Germany
Smaranda Muresan, Columbia University, USA
William Murray, Pearson, USA
Courtney Napoles, Johns Hopkins, USA
Hwee Tou Ng, National University of Singapore, Singapore
Huy Nguyen, University of Pittsburgh, USA
Nobal Niraula, Boeing Research & Technology, USA
Simon Ostermann, Saarland University, Germany
Alexis Palmer, University of North Texas, USA
Ted Pedersen, University of Minnesota, Duluth, USA
Isaac Persing, University of Texas at Dallas, USA
Patti Price, PPRICE Speech and Language Technology Consulting, USA
Martí Quixal, Universitat Oberta de Catalunya, Spain
Zahra Rahimi, University of Pittsburgh, USA
Lakshmi Ramachandran, A9.com, USA
David Randolph, University of Illinois at Chicago, USA
Livy Real, IBM Research, Brazil
Marek Rei, University of Cambridge, UK
Robert Reynolds, UiT The Arctic University of Norway, Norway
Brian Riordan, Educational Testing Service, USA
Andrew Rosenberg, IBM, USA
Alla Rozovskaya, Queens College (CUNY), USA
Alexander Rush, Harvard, USA
Anton Rytting, University of Maryland, USA
Elizabeth Salesky, MIT LL, USA
Allen Schmaltz, Harvard University, USA
Mathias Schulze, University of Waterloo, Canada
Swapna Somasundaran, Educational Testing Service, USA
Helmer Strik, Radboud University Nijmegen, Netherlands
David Suendermann-Oeft, Educational Testing Service, USA
Kaveh Taghipour, National University of Singapore, Singapore
Fateme Torabi Asr, Indiana University, USA
Yuen-Hsien Tseng, National Taiwan Normal University, Taiwan

Sowmya Vajjala, Iowa State University, USA
Giulia Venturi, Istituto di Linguistica Computazionale (ILC-CNR), Italy
Aline Villavicencio, Federal University of Rio Grande do Sul, Brazil
Carl Vogel, Computational Linguistics Lab, O'Reilly Institute, Ireland
Elena Volodina, University of Gothenburg, Sweden
Michael White, The Ohio State University, USA
Denise Whitelock, The Open University, UK
David Wible, National Central University, Taiwan
Alistair Willis, Open University, UK
Magdalena Wolska, Eberhard Karls Universität Tübingen, Germany
Michael Wojatzki, University of Duisburg-Essen, Germany
Huichao Xue, Google, USA
Zheng Yuan, University of Cambridge, UK
Marcos Zampieri, Saarland University, Germany
Torsten Zesch, University of Duisburg-Essen, Germany
Fan Zhang, University of Pittsburgh, USA
Xiaodan Zhu, National Research Council Canada, Canada