NAACL-HLT 2012

Proceedings of the Seventh Workshop on Innovative Use of NLP for Building Educational Applications

June 7, 2012
Montréal, Canada

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2012 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

ISBN13: 978-1-937284-20-6
ISBN10: 1-937284-20-4

Preface

Research in natural language processing (NLP) applications for education has continued to progress using innovative statistical and rule-based NLP methods, or most commonly, a combination of the two. As a community, we continue to improve existing capabilities and to identify and generate innovative ways to use NLP in applications for writing, reading, speaking, critical thinking, curriculum development, and assessment. Steady growth in the development of NLP-based applications for education has prompted an increased number of workshops, typically focusing on a single subfield. In this workshop, researchers present papers from many subfields: tools for automated scoring of text and speech, intelligent tutoring, readability measures, use of corpora, grammatical error detection, and tools for teachers and test developers. These focus on contributions to the three core educational problem spaces: development of curriculum and assessment (e.g., applications that help teachers develop reading materials), delivery of curriculum and assessments (e.g., applications where the student receives instruction and interacts with the system), and reporting of assessment outcomes (e.g., automated essay and other constructed response scoring).

NLP-based educational applications continue to develop in order to serve the learning and assessment needs of students, teachers, schools, and assessment organizations. The practical need for language-analysis capabilities has been motivated even further by increased requirements for state and national assessments, and a growing population of foreign and second language learners. There are currently a number of commercial systems that handle automated scoring of free-text and speech as well as systems that address linguistic complexity in text – commonly referred to as readability measures. More recently, the need for language analysis tools is, in part, driven by a new influence in the educational landscape in the United States: the Common Core State Standards initiative (http://www.corestandards.org/). The initiative has been adopted by 46 states for use in Kindergarten through 12th grade (K-12) classrooms and is likely to have a strong influence on teaching standards, as well as how NLP research and applications are applied in the classroom.

This workshop is the seventh in a series related to Building NLP Applications for Education. The series began at NAACL/HLT (2003), and continued at ACL 2005 (Ann Arbor), ACL/HLT 2008 (Columbus), NAACL/HLT 2009 (Boulder), NAACL/HLT 2010 (Los Angeles), ACL/HLT 2011 (Portland), and now NAACL/HLT 2012 (Montréal). This year, we received a record 42 submissions and accepted 8 full papers as oral presentations and 16 papers as poster presentations, as well as an invited talk by Robert Dale describing the *HOO2012 Shared Task*. The acceptance rate is 57%. All of the papers are published in these proceedings. Each paper was carefully reviewed by at least three members of the Program Committee. We carefully selected reviewers most appropriate for each paper so as to get knowledgeable reviews. This workshop offers an opportunity to present and publish work that is highly relevant to NAACL/HLT, but is also specialized. Thus, the BEA workshop is often a more appropriate venue for such work. We believe that the workshop framework designed to introduce works in progress and new ideas needs to be revived, and we hope that we have achieved this with the breadth and variety of research accepted presented here.

While the field is growing, we do recognize that there is a core group of institutions and researchers who work in this area. With a higher acceptance rate, we were able to include papers from a broad range of topics and institutions. We continue to have a strong policy to avoid conflicts of interest. We did not assign papers to reviewers if the paper had an author from the same institution. Second, with respect to the organizing committee, authors of papers where there was a conflict of interest did not participate in the discussion.

The papers accepted to this workshop were selected on the basis of several factors, including the relevance to a core educational problem space, the novelty of the approach or domain, and the strength of the research. The final set of 24 papers fall under several main themes:

Assessing Speech: Four papers focus on assessing spoken language of non-native speakers of English (Chen; Chen and Zechner; Huant et al, and Yoon et al.).

Automated Scoring Tools: Six papers focus on aspects of scoring textual responses, such as short answer scoring (Hahn and Meurers; Rus and Lintean; and Ziai et al.), measuring coherence in learner essays (Yannakoudakis and Briscoe), measuring the use of factual information (Beigman-Klebanov and Higgins), and automatically grading responses to science questions (Sil et al.).

Generation: Two papers (both Perez-Beltrachini et al.) present work into generation cloze questions and grammar exercises.

Grammatical Error Detection: Three papers target grammatical error detection. Madnina et al. discuss novel techniques for error correction and Ferraro et al. judge grammatically. The third paper (Flor and Futagi) focuses on automatic spell correction in student essays.

Intelligent Tutoring: Two papers discuss issues related to intelligent tutoring systems (Becker et al.; and Bethard et al.).

Readability and Reading Assistance Tools: Four papers investigate aspects of readability ranging from developing tools for student reading assistance to detecting a document's reading level (Talukdar and Cohen; Eom et al.; Maamouri et al.; and Vajjala and Meurers).

Other Learning Assistance Research: Finally, we have three papers on other topics. Xiong et al., present a tool for peer-review exploration. Dickinson et al. present a method for predicting which college level Hebrew class a student should place into. And Chen et al. present an approach to generating paraphrases for language learning.

This year, we are pleased to host the *Helping Our Own* (HOO-2012) shared task on grammatical

error detection (http://www.correcttext.org/hoo2012), organized by Robert Dale et al. In its second year, this instantiation of the shared task focuses on the detection and correction of determiner and preposition errors in texts written by non-native speakers of English. These error types are two of the most frequent, and nettlesome, ones for English learners. 14 teams took part in the shared task and descriptions of their submitted systems are found in these proceedings and are presented as posters in conjunction with the BEA7 poster session.

We wish to thank everyone who showed interest and submitted a paper, all of the authors for their contributions, the members of the Program Committee for their thoughtful reviews, and everyone who attended this workshop. All of these factors contribute to a truly enriching event!

Joel Tetreault, Educational Testing Service
Jill Burstein, Educational Testing Service
Claudia Leacock, CTB McGraw-Hill

Organizers:

Joel Tetreault, Educational Testing Service
Jill Burstein, Educational Testing Service
Claudia Leacock, CTB McGraw-Hill

Program Committee:

Andrea Abel, EURAC, Italy
Shane Bergsma, Johns Hopkins University, USA
Delphine Bernhard, Université de Strasbourg, France
Jared Bernstein, Pearson, USA
Daniel Blanchard, Educational Testing Service, USA
Kristy Boyer, North Carolina State University, USA
Chris Brew, Educational Testing Service, USA
Chris Brockett, Microsoft Research, USA
Aoife Cahill, Educational Testing Service, USA
Martin Chodorow, Hunter College of CUNY, USA
Mark Core, USC Institute for Creative Technologies, USA
Daniel Dahlmeier, National University of Singapore, Singapore
Markus Dickinson, Indiana University, USA
Robert Dale, Macquarie University, Australia
Bill Dolan, Microsoft Research, USA
Maxine Eskenazi, Carnegie Mellon University, USA
Keelan Evanini, Educational Testing Service, USA
Jennifer Foster, Dublin City University, Ireland
Annette Frank, University of Heidelberg, Germany
Michael Gamon, Microsoft Research, USA
Caroline Gasperin, TouchType, Brazil
Kallirroi Georgila, USC Institute for Creative Technologies, USA
Iryna Gurevych, University of Darmstadt, Germany
Na-Rae Han, University of Pittsburgh, USA
Trude Heift, Simon Frasier University, Canada
Michael Heilman, Educational Testing Service, USA
Derrick Higgins, Educational Testing Service, USA
Heng Ji, Queens College of CUNY, USA
Pamela Jordan, University of Pittsburgh, USA
Ola Knutsson, Stockholm University, Sweden
John Lee, City University of Hong Kong, China
Xiaofei Lu, Penn State University, USA
Roger Levy, University of California San Diego, USA
Jackson Liscombe, SpeechCycle, USA
Diane Litman, University of Pittsburgh, USA

Annie Louis, University of Pennsylvania, USA
Nitin Madnani, Educational Testing Service, USA
Montse Maritxalar, University of the Basque Country, Spain
Aurélien Max, LIMSI-CNRS, France
Detmar Meurers, University of Tübingen, Germany
Lisa Michaud, Merrimack College, USA
Rada Mihalcea, University of North Texas, USA
Michael Mohler, University of North Texas, USA
Jack Mostow, Carnegie Mellon University, USA
Smaranda Muresan, Rutgers University, USA
Ani Nenkova, University of Pennsylvania, USA
Rodney Nielsen, University of Colorado, USA
Hwee Tou Ng, National University of Singapore, USA
Matt Post, Johns Hopkins University, USA
Patti Price, PPRICE Speech and Language Technology, USA
Andrew Rosenberg, Queens College of CUNY, USA
Mihai Rotaru, TextKernel, the Netherlands
Dan Roth, University of Illinois Urbana-Champagne, USA
Alla Rozovskaya, University of Illinois Urbana-Champagne, USA
Mathias Schulze, University of Waterloo, Canada
Stephanie Seneff, Massachusetts Institute of Technology, USA
Izhak Shafran, Oregon Health and Science University, USA
Serge Sharoff, University of Leeds, UK
Svetlana Stenchikova, Open University, UK
Helmer Strik, Radboud University Nijmegen, the Netherlands
Joseph Tepperman, Rosetta Stone, USA
Nai-Lung Tsao, National Central University, Taiwan
Benjamin Van Durme, Johns Hopkins University, USA
Arthur Ward, University of Pittsburgh, USA
Monica Ward, Dublin City University, Ireland
David Wible, National Central University, Taiwan
Sze Wong, Macquarie University, Australia
Peter Wood, University of Saskatchewan in Saskatoon, Canada
Klaus Zechner, Educational Testing Service, USA

Table of Contents

Question Ranking and Selection in Tutorial Dialogues
Lee Becker, Martha Palmer, Sarel van Vuuren and Wayne Ward 1

Identifying science concepts and student misconceptions in an interactive essay writing tutor
Steven Bethard, Ifeyinwa Okoye, Md. Arafat Sultan, Haojie Hang, James H. Martin and Tamara Sumner 12

Automatic Grading of Scientific Inquiry
Avirup Sil, Angela Shelton, Diane Jass Ketelhut and Alexander Yates 22

Modeling coherence in ESOL learner texts
Helen Yannakoudakis and Ted Briscoe 33

Exploring Grammatical Error Correction with Not-So-Crummy Machine Translation
Nitin Madnani, Joel Tetreault and Martin Chodorow 44

HOO 2012: A Report on the Preposition and Determiner Error Correction Shared Task
Robert Dale, Ilya Anisimoff and George Narroway 54

Measuring the Use of Factual Information in Test-Taker Essays
Beata Beigman Klebanov and Derrick Higgins 63

Utilizing Cumulative Logit Model and Human Computation on Automated Speech Assessment
Lei Chen 73

PREFER: Using a Graph-Based Approach to Generate Paraphrases for Language Learning
Mei-Hua Chen, Shi-Ting Huang, Chung-Chi Huang, Hsien-Chin Liou and Jason S. Chang 80

Using an Ontology for Improved Automated Content Scoring of Spontaneous Non-Native Speech
Miao Chen and Klaus Zechner 86

Predicting Learner Levels for Online Exercises of Hebrew
Markus Dickinson, Sandra Kübler and Anthony Meyer 95

On using context for automatic correction of non-word misspellings in student essays
Michael Flor and Yoko Futagi 105

Judging Grammaticality with Count-Induced Tree Substitution Grammars
Francis Ferraro, Matt Post and Benjamin Van Durme 116

Scoring Spoken Responses Based on Content Accuracy
Fei Huang, Lei Chen and Jana Sukkarieh 122

Developing ARET: An NLP-based Educational Tool Set for Arabic Reading Enhancement
Mohammed Maamouri, Wajdi Zaghouani, Violetta Cavalli-Sforza, Dave Graff and Mike Ciul 127

Generating Diagnostic Multiple Choice Comprehension Cloze Questions
Jack Mostow and Hyeju Jang 136

Generating Grammar Exercises
Laura Perez-Beltrachini, Claire Gardent and German Kruszewski 147

A Comparison of Greedy and Optimal Assessment of Natural Language Student Input Using Word-to-Word Similarity Metrics
Vasile Rus and Mihai Lintean 157

On Improving the Accuracy of Readability Classification using Insights from Second Language Acquisition
Sowmya Vajjala and Detmar Meurers 163

An Interactive Analytic Tool for Peer-Review Exploration
Wenting Xiong, Diane Litman, Jingtao Wang and Christian Schunn 174

Vocabulary Profile as a Measure of Vocabulary Sophistication
Su-Youn Yoon, Suma Bhat and Klaus Zechner 180

Short Answer Assessment: Establishing Links Between Research Strands
Ramon Ziai, Niels Ott and Detmar Meurers 190

Detection and Correction of Preposition and Determiner Errors in English: HOO 2012
Pinaki Bhaskar, Aniruddha Ghosh, Santanu Pal and Sivaji Bandyopadhyay 201

Informing Determiner and Preposition Error Correction with Hierarchical Word Clustering
Adriane Boyd, Marion Zepf and Detmar Meurers 208

NUS at the HOO 2012 Shared Task
Daniel Dahlmeier, Hwee Tou Ng and Eric Jun Feng Ng 216

VTEX Determiner and Preposition Correction System for the HOO 2012 Shared Task
Vidas Daudaravicius 225

Precision Isn't Everything: A Hybrid Approach to Grammatical Error Detection
Michael Heilman, Aoife Cahill and Joel Tetreault 233

HOO 2012 Error Recognition and Correction Shared Task: Cambridge University Submission Report
Ekaterina Kochmar, Øistein Andersen and Ted Briscoe 242

Korea University System in the HOO 2012 Shared Task
Jieun Lee, Jung-Tae Lee and Hae-Chang Rim 251

A Naive Bayes classifier for automatic correction of preposition and determiner errors in ESL text
Gerard Lynch, Erwan Moreau and Carl Vogel 257

KU Leuven at HOO-2012: A Hybrid Approach to Detection and Correction of Determiner and Preposition Errors in Non-native English Text
Li Quan, Oleksandr Kolomiyets and Marie-Francine Moens 263

The UI System in the HOO 2012 Shared Task on Error Correction
Alla Rozovskaya, Mark Sammons and Dan Roth 272

NAIST at the HOO 2012 Shared Task
Keisuke Sakaguchi, Yuta Hayashibe, Shuhei Kondo, Lis Kanashiro, Tomoya Mizumoto, Mamoru Komachi and Yuji Matsumoto 281

Memory-based text correction for preposition and determiner errors
Antal van den Bosch and Peter Berck 289

Helping Our Own: NTHU NLPLAB System Description
Jian-Cheng Wu, Joseph Chang, Yi-Chun Chen, Shih-Ting Huang, Mei-Hua Chen and Jason S. Chang 295

HOO 2012 Shared Task: UKP Lab System Description
Torsten Zesch and Jens Haase 302

Crowdsourced Comprehension: Predicting Prerequisite Structure in Wikipedia
Partha Talukdar and William Cohen 307

Sense-Specific Lexical Information for Reading Assistance
Soojeong Eom, Markus Dickinson and Rebecca Sachs 316

Evaluating the Meaning of Answers to Reading Comprehension Questions: A Semantics-Based Approach
Michael Hahn and Detmar Meurers 326

Conference Program

Thursday, June 7, 2012

8:45–9:00 Load Presentations

9:00–9:15 Opening Remarks

\+ 9:15–9:40

Question Ranking and Selection in Tutorial Dialogues
Lee Becker, Martha Palmer, Sarel van Vuuren and Wayne Ward

\+ 9:40–10:05

Identifying science concepts and student misconceptions in an interactive essay writing tutor
Steven Bethard, Ifeyinwa Okoye, Md. Arafat Sultan, Haojie Hang, James H. Martin and Tamara Sumner

\+ 10:05–10:30

Automatic Grading of Scientific Inquiry
Avirup Sil, Angela Shelton, Diane Jass Ketelhut and Alexander Yates

10:30–11:00 **Break**

\+ 11:00–11:25

Modeling coherence in ESOL learner texts
Helen Yannakoudakis and Ted Briscoe

\+ 11:25–11:50

Exploring Grammatical Error Correction with Not-So-Crummy Machine Translation
Nitin Madnani, Joel Tetreault and Martin Chodorow

\+ 11:50–12:15

HOO 2012: A Report on the Preposition and Determiner Error Correction Shared Task
Robert Dale, Ilya Anisimoff and George Narroway

Thursday, June 7, 2012 (continued)

12:15–1:45 **Lunch**

1:45–3:30 **Poster Session**

\+ BEA7 Posters

Measuring the Use of Factual Information in Test-Taker Essays
Beata Beigman Klebanov and Derrick Higgins

Utilizing Cumulative Logit Model and Human Computation on Automated Speech Assessment
Lei Chen

PREFER: Using a Graph-Based Approach to Generate Paraphrases for Language Learning
Mei-Hua Chen, Shi-Ting Huang, Chung-Chi Huang, Hsien-Chin Liou and Jason S. Chang

Using an Ontology for Improved Automated Content Scoring of Spontaneous Non-Native Speech
Miao Chen and Klaus Zechner

Predicting Learner Levels for Online Exercises of Hebrew
Markus Dickinson, Sandra Kübler and Anthony Meyer

On using context for automatic correction of non-word misspellings in student essays
Michael Flor and Yoko Futagi

Judging Grammaticality with Count-Induced Tree Substitution Grammars
Francis Ferraro, Matt Post and Benjamin Van Durme

Scoring Spoken Responses Based on Content Accuracy
Fei Huang, Lei Chen and Jana Sukkarieh

Developing ARET: An NLP-based Educational Tool Set for Arabic Reading Enhancement
Mohammed Maamouri, Wajdi Zaghouani, Violetta Cavalli-Sforza, Dave Graff and Mike Ciul

Generating Diagnostic Multiple Choice Comprehension Cloze Questions
Jack Mostow and Hyeju Jang

Thursday, June 7, 2012 (continued)

Generating Grammar Exercises
Laura Perez-Beltrachini, Claire Gardent and German Kruszewski

A Comparison of Greedy and Optimal Assessment of Natural Language Student Input Using Word-to-Word Similarity Metrics
Vasile Rus and Mihai Lintean

On Improving the Accuracy of Readability Classification using Insights from Second Language Acquisition
Sowmya Vajjala and Detmar Meurers

An Interactive Analytic Tool for Peer-Review Exploration
Wenting Xiong, Diane Litman, Jingtao Wang and Christian Schunn

Vocabulary Profile as a Measure of Vocabulary Sophistication
Su-Youn Yoon, Suma Bhat and Klaus Zechner

Short Answer Assessment: Establishing Links Between Research Strands
Ramon Ziai, Niels Ott and Detmar Meurers

\+ HOO2012 Posters

Detection and Correction of Preposition and Determiner Errors in English: HOO 2012
Pinaki Bhaskar, Aniruddha Ghosh, Santanu Pal and Sivaji Bandyopadhyay

Informing Determiner and Preposition Error Correction with Hierarchical Word Clustering
Adriane Boyd, Marion Zepf and Detmar Meurers

NUS at the HOO 2012 Shared Task
Daniel Dahlmeier, Hwee Tou Ng and Eric Jun Feng Ng

VTEX Determiner and Preposition Correction System for the HOO 2012 Shared Task
Vidas Daudaravicius

Precision Isn't Everything: A Hybrid Approach to Grammatical Error Detection
Michael Heilman, Aoife Cahill and Joel Tetreault

HOO 2012 Error Recognition and Correction Shared Task: Cambridge University Submission Report
Ekaterina Kochmar, Øistein Andersen and Ted Briscoe

Korea University System in the HOO 2012 Shared Task
Jieun Lee, Jung-Tae Lee and Hae-Chang Rim

A Naive Bayes classifier for automatic correction of preposition and determiner errors in ESL text
Gerard Lynch, Erwan Moreau and Carl Vogel

KU Leuven at HOO-2012: A Hybrid Approach to Detection and Correction of Determiner and Preposition Errors in Non-native English Text
Li Quan, Oleksandr Kolomiyets and Marie-Francine Moens

The UI System in the HOO 2012 Shared Task on Error Correction
Alla Rozovskaya, Mark Sammons and Dan Roth

NAIST at the HOO 2012 Shared Task
Keisuke Sakaguchi, Yuta Hayashibe, Shuhei Kondo, Lis Kanashiro, Tomoya Mizumoto, Mamoru Komachi and Yuji Matsumoto

Memory-based text correction for preposition and determiner errors
Antal van den Bosch and Peter Berck

Helping Our Own: NTHU NLPLAB System Description
Jian-Cheng Wu, Joseph Chang, Yi-Chun Chen, Shih-Ting Huang, Mei-Hua Chen and Jason S. Chang

HOO 2012 Shared Task: UKP Lab System Description
Torsten Zesch and Jens Haase

3:30–4:00 **Break**

\+ 4:00–4:25

Crowdsourced Comprehension: Predicting Prerequisite Structure in Wikipedia
Partha Talukdar and William Cohen

\+ 4:25–4:50

Sense-Specific Lexical Information for Reading Assistance
Soojeong Eom, Markus Dickinson and Rebecca Sachs

Thursday, June 7, 2012 (continued)

+ 4:50–5:15

Evaluating the Meaning of Answers to Reading Comprehension Questions: A Semantics-Based Approach
Michael Hahn and Detmar Meurers

5:15–5:30 Closing Remarks