Introduction

Research in NLP applications for education continues to progress using innovative NLP techniques. New technologies have made it possible to include speech in both assessment and in Intelligent Tutoring Systems (ITS). NLP techniques are also being used to generate assessments and tools for curriculum development of reading materials, as well as tools to support assessment and test development. As a community, we continue to improve existing capabilities and to identify and generate innovative and creative ways to use NLP in applications for writing, reading, speaking, critical thinking, and assessment.

In this workshop, we focus on contributions to 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 scoring). The need for, and the rapid development of, language-based capabilities have been driven by increased requirements for state and national assessments and a growing population of foreign and second language learners.

This is the sixth in a series of workshops on Building NLP Applications for Education that began at NAACL/HLT 2003 (Edmonton), and continued at ACL 2005 (Ann Arbor), ACL/HLT 2008 (Columbus), NAACL/HLT 2009 (Boulder), NAACL/HLT 2010 (Los Angeles), and now ACL/HLT 2011 (Portland). Research in this area continues to grow, and there is ever-increasing interest and practical application that was evidenced this year, again, by the large number of submissions.

We received a record 35 submissions and accepted 8 full papers as oral presentations and 14 papers as poster presentations. Each paper was carefully reviewed by at least three members of the Program Committee. We selected reviewers most appropriate for each paper so as to give more helpful feedback and comments. This workshop offers an opportunity to present and publish work that is highly relevant to ACL, but is also specialized, so the workshop is often a more appropriate venue. The decision to have a poster session this year was made so as to offer more breadth in terms of topics related to NLP and education and to reinstate the original concept of a workshop as a venue for fully developed work as well as work in progress. Also, we continue to have a strong policy with respect to conflicts of interest and made a concerted effort to not assign papers to reviewers if the paper had an author from their institution.

The papers accepted to this workshop were selected on the basis of several factors: 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 papers fall under several main themes.

Assessing Speech – Five papers focus on assessing spoken language of non-native speakers of English (Chen and Yoon; Cook, et al; Downey, et al; Yoon and Higgins; and Yoon, et al).

Grammatical Error Detection – Five papers deal with grammatical error detection for non-native speakers, ranging from new paradigms and methodologies (Gamon; Dickinson, et al; West, et al), to CALL applications (Huang, et al), to using grammar checking to measure language development (Hassanali and Liu).

Generation – Five papers address different aspects of generating questions, exercises and examples for
students (Agarawal, et al; Agarawal and Mannem; Mostow and Duan; Olney, et al; and Theune, et al).


Finally, we also have four papers on other topics. Xiong and Litman use NLP techniques to determine the effectiveness of peer-review. Van Oosten and Hoste investigate the efficacy of using experts and crowdsourcing for readability assessment. Dela Rosa and Eskenazi investigate the effect of word complexity on vocabulary learning for language learners. Yang and Heines apply NLP techniques to the novel task of determining the best transfer course equivalencies.

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 attends this workshop. All of these factors contribute to a truly rich and successful event. And, the informal post-workshop dinner is getting more crowded every year!

Joel Tetreault, Educational Testing Service
Jill Burstein, Educational Testing Service
Claudia Leacock, Butler Hill Group
Organizers:

Joel Tetreault, Educational Testing Service
Jill Burstein, Educational Testing Service
Claudia Leacock, Butler Hill Group

Program Committee:

Delphine Bernhard, LIMSI-CNRS, France
Jared Bernstein, Pearson, USA
Chris Brockett, MSR, USA
Lei Chen, ETS, USA
Martin Chodorow, Hunter College, CUNY, USA
Mark Core, Institute for Creative Technologies
Barbara Di Eugenio, University of Illinois at Chicago, USA
Markus Dickinson, Indiana University, USA
Bill Dolan, Microsoft, USA
Maxine Eskenazi, Carnegie Mellon University, USA
Keelan Evanini, ETS, USA
Peter Foltz, Pearson Knowledge Technologies, USA
Jennifer Foster, Dublin City University, Ireland
Horacio Franco, SRI, USA
Michael Gamon, Microsoft, USA
Caroline Gasperin, TouchType Ltd., UK
Kallirroi Georgila, Institute for Creative Technologies
Iryna Gurevych, University of Darmstadt, Germany
Na-Rae Han, University of Pittsburgh, USA
Trude Heift, Simon Frasier University, Canada
Derrick Higgins, ETS, USA
Emi Izumi, Kyoto University of Foreign Studies, Japan
Heng Ji, Queens College, USA
Pamela Jordan, University of Pittsburgh, USA
Ola Knutsson, KTH Nada, Sweden
John Lee, City University of Kong, USA
Jackson Liscombe, SpeechCycle, Inc., USA
Diane Litman, University of Pittsburgh, USA
Annie Louis, University of Pennsylvania
Nitin Madnani, ETS, USA
Montse Maritxalar, University of the Basque Country, Spain
James Martin, University of Colorado
Aurelien Max, LIMSI-CNRS, France
Detmar Meurers, University of Tübingen, Germany
Lisa Michaud, Merrimack College, USA
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Question Generation using Discourse Cues</td>
<td>1</td>
</tr>
<tr>
<td>Manish Agarwal, Rakshit Shah and Prashanth Mannem</td>
<td></td>
</tr>
<tr>
<td>Understanding Differences in Perceived Peer-Review Helpfulness using Natural Language Processing</td>
<td>10</td>
</tr>
<tr>
<td>Wenting Xiong and Diane Litman</td>
<td></td>
</tr>
<tr>
<td>Generating Varied Narrative Probability Exercises</td>
<td>20</td>
</tr>
<tr>
<td>Mariët Theune, Roan Boer Rookhuiszen, Rieks op den Akker and Hanneke Geerlings</td>
<td></td>
</tr>
<tr>
<td>Elicited Imitation for Prediction of OPI Test Scores</td>
<td>30</td>
</tr>
<tr>
<td>Kevin Cook, Jeremiah McGhee and Deryle Lonsdale</td>
<td></td>
</tr>
<tr>
<td>Detecting Structural Events for Assessing Non-Native Speech</td>
<td>38</td>
</tr>
<tr>
<td>Lei Chen and Su-Youn Yoon</td>
<td></td>
</tr>
<tr>
<td>Performance of Automated Scoring for Children’s Oral Reading</td>
<td>46</td>
</tr>
<tr>
<td>Ryan Downey, David Rubin, Jian Cheng and Jared Bernstein</td>
<td></td>
</tr>
<tr>
<td>Automatic Gap-fill Question Generation from Text Books</td>
<td>56</td>
</tr>
<tr>
<td>Manish Agarwal and Prashanth Mannem</td>
<td></td>
</tr>
<tr>
<td>Exploring Effective Dialogue Act Sequences in One-on-one Computer Science Tutoring Dialogues</td>
<td>65</td>
</tr>
<tr>
<td>Lin Chen, Barbara Di Eugenio, Davide Fossati, Stellan Ohlsson and David Cosejo</td>
<td></td>
</tr>
<tr>
<td>Effect of Word Complexity on L2 Vocabulary Learning</td>
<td>76</td>
</tr>
<tr>
<td>Kevin Dela Rosa and Maxine Eskenazi</td>
<td></td>
</tr>
<tr>
<td>Developing Methodology for Korean Particle Error Detection</td>
<td>81</td>
</tr>
<tr>
<td>Markus Dickinson, Ross Israel and Sun-Hee Lee</td>
<td></td>
</tr>
<tr>
<td>Measuring Language Development in Early Childhood Education: A Case Study of Grammar Checking in Child Language Transcripts</td>
<td>87</td>
</tr>
<tr>
<td>Khairun-nisa Hassanali and Yang Liu</td>
<td></td>
</tr>
<tr>
<td>GRASP: Grammar- and Syntax-based Pattern-Finder in CALL</td>
<td>96</td>
</tr>
<tr>
<td>Chung-chi Huang, Mei-hua Chen, Shih-ting Huang, Hsien-chin Liou and Jason S. Chang</td>
<td></td>
</tr>
<tr>
<td>Generating Example Contexts to Illustrate a Target Word Sense</td>
<td>105</td>
</tr>
<tr>
<td>Jack Mostow and Weisi Duan</td>
<td></td>
</tr>
<tr>
<td>Generating Concept Map Exercises from Textbooks</td>
<td>111</td>
</tr>
<tr>
<td>Andrew Olney, Whitney Cade and Claire Williams</td>
<td></td>
</tr>
<tr>
<td>Readability Annotation: Replacing the Expert by the Crowd</td>
<td>120</td>
</tr>
<tr>
<td>Philip van Oosten and Véronique Hoste</td>
<td></td>
</tr>
</tbody>
</table>
Story Assembly in a Dyslexia Fluency Tutor  
Arthur Ward and Rebecca Crowley .......................................................... 130

Predicting Change in Student Motivation by Measuring Cohesion between Tutor and Student  
Arthur Ward, Diane Litman and Maxine Eskenazi ........................................ 136

Using Semantic Distance to Automatically Suggest Transfer Course Equivalencies  
Beibei Yang and Jesse M. Heines ................................................................. 142

Non-scorable Response Detection for Automated Speaking Proficiency Assessment  
Su-Youn Yoon, Keelan Evanini and Klaus Zechner ...................................... 152

Non-English Response Detection Method for Automated Proficiency Scoring System  
Su-Youn Yoon and Derrick Higgins ............................................................ 161

Bilingual Random Walk Models for Automated Grammar Correction of ESL Author-Produced Text  
Randy West, Y. Albert Park and Roger Levy ............................................... 170

High-Order Sequence Modeling for Language Learner Error Detection  
Michael Gamon .............................................................................................. 180
Conference Program

Friday, June 24, 2011

8:45–9:00 Load Presentations

9:00–9:15 Opening Remarks

9:15–9:40 Automatic Question Generation using Discourse Cues
Manish Agarwal, Rakshit Shah and Prashanth Mannem

9:40–10:05 Understanding Differences in Perceived Peer-Review Helpfulness using Natural Language Processing
Wenting Xiong and Diane Litman

10:05–10:30 Generating Varied Narrative Probability Exercises
Mariët Theune, Roan Boer Rookhuizen, Rieks op den Akker and Hanneke Geerlings

10:30–11:00 Break

11:00–11:25 Elicited Imitation for Prediction of OPI Test Scores
Kevin Cook, Jeremiah McGhee and Deryle Lonsdale

11:25–11:50 Detecting Structural Events for Assessing Non-Native Speech
Lei Chen and Su-Youn Yoon

11:50–12:15 Performance of Automated Scoring for Children’s Oral Reading
Ryan Downey, David Rubin, Jian Cheng and Jared Bernstein

12:15–1:45 Lunch

1:45–3:30 Poster Session

Automatic Gap-fill Question Generation from Text Books
Manish Agarwal and Prashanth Mannem

Exploring Effective Dialogue Act Sequences in One-on-one Computer Science Tutoring Dialogues
Lin Chen, Barbara Di Eugenio, Davide Fossati, Stellan Ohlsson and David Cosejo
Effect of Word Complexity on L2 Vocabulary Learning
Kevin Dela Rosa and Maxine Eskenazi

Developing Methodology for Korean Particle Error Detection
Markus Dickinson, Ross Israel and Sun-Hee Lee

Measuring Language Development in Early Childhood Education: A Case Study of Grammar Checking in Child Language Transcripts
Khairun-nisa Hassanali and Yang Liu

GRASP: Grammar- and Syntax-based Pattern-Finder in CALL
Chung-chi Huang, Mei-hua Chen, Shih-ting Huang, Hsien-chin Liou and Jason S. Chang

Generating Example Contexts to Illustrate a Target Word Sense
Jack Mostow and Weisi Duan

Generating Concept Map Exercises from Textbooks
Andrew Olney, Whitney Cade and Claire Williams

Readability Annotation: Replacing the Expert by the Crowd
Philip van Oosten and Véronique Hoste

Story Assembly in a Dyslexia Fluency Tutor
Arthur Ward and Rebecca Crowley

Predicting Change in Student Motivation by Measuring Cohesion between Tutor and Student
Arthur Ward, Diane Litman and Maxine Eskenazi

Using Semantic Distance to Automatically Suggest Transfer Course Equivalencies
Beibei Yang and Jesse M. Heines

Non-scorable Response Detection for Automated Speaking Proficiency Assessment
Su-Youn Yoon, Keelan Evanini and Klaus Zechner

Non-English Response Detection Method for Automated Proficiency Scoring System
Su-Youn Yoon and Derrick Higgins
Friday, June 24, 2011 (continued)

3:30–4:00 Break

4:00–4:25 Bilingual Random Walk Models for Automated Grammar Correction of ESL Author-Produced Text
Randy West, Y. Albert Park and Roger Levy

4:25–4:50 High-Order Sequence Modeling for Language Learner Error Detection
Michael Gamon

4:50–5:00 Closing Remarks