NAACL/HLT 2013

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

June 13, 2013
Atlanta, Georgia

Listening. Learning. Leading.®

©2013 The Association for Computational Linguistics

209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-47-3

Preface

Research focusing on 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. NLP-based educational applications continue to develop in order to serve the learning and assessment needs of students, teachers, schools, and testing organizations, often guided by educational policy and learner needs.

The practical need for language-analysis capabilities has been further motivated by increased requirements for state and national assessments, and a growing population of foreign and second language learners. In the United States, the need for applications for language analysis is emphasized by the Common Core State Standards Initiative (Standards), now adopted by 46 States: (http://www.corestandards.org/). The Standards describe what K-12 students should be learning with regard to Reading, Writing, Speaking, Listening, Language, and Media and Technology, and have clear alignments with NLP research and potential applications. Motivated by the Common Core State Standards Initiative, the use of NLP in educational contexts took two major steps forward. First, outside of the computational linguistics community, the Hewlett Foundation reached out to both the public and private sectors and sponsored two competitions: one on automated essay scoring (Automated Student Assessment Prize: ASAP, Phase 1), and a second on short-answer scoring (Phase 2). The motivation driving these competitions was to engage the larger scientific community to harness the collective knowledge toward the development of new ideas and methods. In April 2013, a New York Times article by John Markoff discussed automated essay scoring use by EdX, one of the two competing Massive Online Educational Course (MOOC) companies. Within the computational linguistics community, a breakthrough for educational applications is a new Shared Task co-located with the BEA workshop, NLI-2013, in which the task involves identifying the native language (L1) of a writer based solely on a sample of their writing. Independent of the BEA workshop, there were two additional shared task competitions: the CoNLL Shared Task on Grammatical Error Correction, and a SemEval Shared Task on Student Response Analysis. NAACL and ACL each hosted other education-centered workshops, including the Workshop on Using NLP to Improve Text Accessibility at NAACL, and the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations at ACL. Further, a new book, The Handbook of Automated Essay Evaluation (2013) (Eds., Mark Shermis and Jill Burstein) reports on the state-of-the-art in the field, and a Special Issue of the International Journal of Applied Linguistics, Current research in readability and text simplification (forthcoming) (Eds. Thomas François and Delphine Bernhard) calls for new work. The competitions, the recent deployment of automated essay grading in MOOCs, the education-related workshops, and are evidence of the high visibility of Educational Applications in NLP.

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 one specific subfield. In this workshop, we present papers from these subfields: tools for automated scoring of text and speech, dialogue and intelligent tutoring, use of corpora, grammatical error detection, and native language identification. Consistent with 2012, the workshop made an attempt to focus on contributions that could be described in core educational problem spaces, including: 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). This workshop is the eighth in a series, specifically related to "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), ACL/HLT 2011 (Portland), NAACL/HLT 2012 (Montreal), and now, NAACL/HLT 2013 (Atlanta). This year, the workshop is co-located with the NLI-2013 (Native Language Identification Shared Task) – another indication of how this field is developing.

We received 25 submissions and accepted nine papers as oral presentations and six as poster presentation plus an oral presentation of the summary report for the NLI Shared Task. All of the papers appear in these proceedings. Each paper was reviewed by three members of the Program Committee who were most appropriate for each paper. We continue to have a very strong policy to deal with conflicts of interest. First, we made a concerted effort to not assign papers to reviewers to evaluate if the paper had an author from their institution. Second, with respect to the organizing committee, authors of papers where there was a conflict of interest recused themselves from the discussion.

This workshop offers an opportunity to present and publish work that is highly relevant to NAACL/HLT, but is also highly specialized, and so this workshop is often a more appropriate venue for such work. The Poster session offers more breadth in terms of topics related to NLP and education, and maintains the original concept of a workshop. We believe that the workshop framework designed to introduce work 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 for this workshop. The total number of acceptances represents a 60% acceptance rate across oral and poster presentations.

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 wider variety of topics and institutions. 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 accepted papers fall under several main themes:

Automatic Writing Assessment Measures: Four papers focus on writing assessment and feedback. Östling et al. describe work into automatic scoring of Swedish essays and Andersen et al. describe a system which provides automatic on English learners' writing. Vajjala and Loo describe work into proficiency classification of Estonian language learners, and Madnani et al. describe work into the automatic scoring of a summarization task designed to measure reading comphrension in young students.

Assessing Speech: Four papers focus on different methods of assessing spoken the language of different populations of non-native speakers of English (Xie and Chen; Evanini et al.; Zechner and Wang; Chen).

Grammatical Error Correction: Two papers describe work into the creation of an error-annotated corpus of learner English (Dahlmeier et al.) and the automatic detection of hyphens in learner English (Cahill et al.).

Other Learning Assistance Research: Finally, we have several papers on other topics which use NLP to develop educational applications. Topics include intelligent tutoring (Dzikovska et al.), use of machine translation metrics to rate student translations (Michaud and McCoy), semantic analysis of

interactive learner sentences (Levi and Dickinson), dependency annotation in learner writing (Ragheb and Dickinson) and the use of linguistic error codes for identifying neurodevelopmental disorders (Morley et al.).

This year, we are excited to host the first Shared Task in Native Language Identification (http://www.nlisharedtask2013.org/). The task involves automatically predicting the native language of a English language learner based solely on their essay. 29 teams competed and 24 teams submitted descriptions of their submitted systems. These papers are found in these proceedings and are presented as posters in conjunction with the BEA7 poster session. A summary report of the shared task (Tetreault et al.) is also found in the proceedings.

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. The eighth edition of the BEA workshop is notable one as this is the first year that the workshop has sponsors. We would like to thank our four sponsors: Appen Butler-Hill, CTB/McGraw-Hill, Educational Testing Service, and PacificMetrics, whose contributions allowed us to subsidize students at the workshop dinner, and make workshop t-shirts! In addition, we would like to thank Joya Tetreault for creating the t-shirt design.

Joel Tetreault, Nuance Communications, Inc.
Jill Burstein, Educational Testing Service
Claudia Leacock, CTB/McGraw-Hill

Organizers:

Joel Tetreault, Nuance Communications, Inc.
Jill Burstein, Educational Testing Service
Claudia Leacock, CTB McGraw-Hill

Program Committee:

Andrea Abel, EURAC, Italy
Sumit Basu, Microsoft Research, USA
Lee Becker, Avaya Labs, USA
Beata Beigman Klebanov, Educational Testing Service, USA
Delphine Bernhard, Universite de Strasbourg, France
Jared Bernstein, Pearson, USA
Kristy Boyer, North Carolina State University, USA
Chris Brew, Educational Testing Service, USA
Ted Briscoe, University of Cambridge, UK
Chris Brockett, Microsoft, USA
Aoife Cahill, Educational Testing Service, USA
Martin Chodorow, Hunter College, CUNY, USA
Mark Core, USC Institute for Creative Technologies, USA
Daniel Dahlmeier, National University of Singapore, Singapore
Markus Dickinson, Indiana University, USA
Bill Dolan, Microsoft, USA
Myrosia Dzikovska, University of Edinburgh, UK
Keelan Evanini, Educational Testing Service, USA
Michael Flor, Educational Testing Service, USA
Peter Foltz, Pearson Knowledge Technologies, USA
Jennifer Foster, Dublin City University, Ireland
Horacio Franco, SRI, USA
Michael Gamon, Microsoft, USA
Caroline Gasperin, SwiftKey, UK
Kallirroi Georgila, USC Institute for Creative Technologies, USA
Iryna Gurevych, University of Darmstadt, Germany
Kadri Hacioglu, Rosetta Stone, USA
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
Ross Israel, Indiana University, USA
Heng Ji, Queens College, USA
Pamela Jordan, University of Pittsburgh, USA
Ola Knutsson, Stockholm University, Sweden

Mamoru Komachi, Nara Institute of Science and Technology, Japan
John Lee, City University of Hong Kong, China
Jackson Liscombe, Nuance Communications Inc., USA
Diane Litman, University of Pittsburgh, USA
Annie Louis, University of Pennsylvania, USA
Xiaofei Lu, Penn State University, USA
Nitin Madnani, Educational Testing Service, USA
Montse Maritxalar, University of the Basque Country, Spain
James Martin, University of Colorado, USA
Aurélien Max, LIMSI-CNRS, France
Detmar Meurers, University of Tübingen, Germany
Lisa Michaud, Merrimack College, USA
Michael Mohler, University of North Texas, USA
Smaranda Muresan, Rutgers University, USA
Ani Nenkova, University of Pennsylvania, USA
Hwee Tou Ng, National University of Singapore, Singapore
Rodney Nielsen, University of North Texas, USA
Ted Pedersen, University of Minnesota, USA
Bryan Pellom, Rosetta Stone, USA
Heather Pon-Barry, Arizona State University, USA
Patti Price, PPRICE Speech and Language Technology, USA
Andrew Rosenberg, Queens College, CUNY, USA
Mihai Rotaru, TextKernel, The Netherlands
Dan Roth, UIUC, USA
Alla Rozovskaya, UIUC, USA
Izhak Shafran, Oregon Health and Science University, USA
Serge Sharoff, University of Leeds, UK
Richard Sproat, Google, USA
Svetlana Stenchikova, Columbia University, USA
Helmer Strik, Radboud University Nijmegen, The Netherlands
Joseph Tepperman, Rosetta Stone, USA
Nai-Lung Tsao, National Central University, Taiwan
Elena Volodina, University of Gothenburg, Sweden
Monica Ward, Dublin City University, Ireland
Pete Whitelock, Oxford University Press, UK
David Wible, National Central University, Taiwan
Peter Wood, University of Saskatchewan in Saskatoon, Canada
Klaus Zechner, Educational Testing Service, USA
Torsten Zesch, University of Darmstadt, Germany

Table of Contents

The Utility of Manual and Automatic Linguistic Error Codes for Identifying Neurodevelopmental Disorders
Eric Morley, Brian Roark and Jan van Santen 1

Shallow Semantic Analysis of Interactive Learner Sentences
Levi King and Markus Dickinson 11

Building a Large Annotated Corpus of Learner English: The NUS Corpus of Learner English
Daniel Dahlmeier, Hwee Tou Ng and Siew Mei Wu 22

Developing and testing a self-assessment and tutoring system
Øistein E. Andersen, Helen Yannakoudakis, Fiona Barker and Tim Parish 32

Automated Essay Scoring for Swedish
Robert Östling, André Smolentzov, Björn Tyrefors Hinnerich and Erik Höglin 42

A Report on the First Native Language Identification Shared Task
Joel Tetreault, Daniel Blanchard and Aoife Cahill 48

Applying Unsupervised Learning To Support Vector Space Model Based Speaking Assessment
Lei Chen 58

Role of Morpho-Syntactic Features in Estonian Proficiency Classification
Sowmya Vajjala and Kaidi Loo 63

Automated Content Scoring of Spoken Responses in an Assessment for Teachers of English
Klaus Zechner and Xinhao Wang 73

Experimental Results on the Native Language Identification Shared Task
Amjad Abu-Jbara, Rahul Jha, Eric Morley and Dragomir Radev 82

VTEX System Description for the NLI 2013 Shared Task
Vidas Daudaravicius 89

Feature Space Selection and Combination for Native Language Identification
Cyril Goutte, Serge Léger and Marine Carpuat 96

Discriminating Non-Native English with 350 Words
John Henderson, Guido Zarrella, Craig Pfeifer and John D. Burger 101

Maximizing Classification Accuracy in Native Language Identification
Scott Jarvis, Yves Bestgen and Steve Pepper 111

Recognizing English Learners' Native Language from Their Writings
Baoli LI 119

NLI Shared Task 2013: MQ Submission
Shervin Malmasi, Sze-Meng Jojo Wong and Mark Dras 124

NAIST at the NLI 2013 Shared Task
Tomoya Mizumoto, Yuta Hayashibe, Keisuke Sakaguchi, Mamoru Komachi and Yuji Matsumoto 134

Cognate and Misspelling Features for Natural Language Identification
Garrett Nicolai, Bradley Hauer, Mohammad Salameh, Lei Yao and Grzegorz Kondrak 140

Exploring Syntactic Representations for Native Language Identification
Ben Swanson 146

Simple Yet Powerful Native Language Identification on TOEFL11
Ching-Yi Wu, Po-Hsiang Lai, Yang Liu and Vincent Ng 152

Prompt-based Content Scoring for Automated Spoken Language Assessment
Keelan Evanini, Shasha Xie and Klaus Zechner 157

Automated Scoring of a Summary-Writing Task Designed to Measure Reading Comprehension
Nitin Madnani, Jill Burstein, John Sabatini and Tenaha O'Reilly 163

Inter-annotator Agreement for Dependency Annotation of Learner Language
Marwa Ragheb and Markus Dickinson 169

Native Language Identification with PPM
Victoria Bobicev 180

Using Other Learner Corpora in the 2013 NLI Shared Task
Julian Brooke and Graeme Hirst 188

Combining Shallow and Linguistically Motivated Features in Native Language Identification
Serhiy Bykh, Sowmya Vajjala, Julia Krivanek and Detmar Meurers 197

Linguistic Profiling based on General–purpose Features and Native Language Identification
Andrea Cimino, Felice Dell'Orletta, Giulia Venturi and Simonetta Montemagni 207

Improving Native Language Identification with TF-IDF Weighting
Binyam Gebrekidan Gebre, Marcos Zampieri, Peter Wittenburg and Tom Heskes 216

Native Language Identification: a Simple n-gram Based Approach
Binod Gyawali, Gabriela Ramirez and Thamar Solorio 224

Feature Engineering in the NLI Shared Task 2013: Charles University Submission Report
Barbora Hladka, Martin Holub and Vincent Kriz 232

Native Language Identification: A Key N-gram Category Approach
Kristopher Kyle, Scott Crossley, Jianmin Dai and Danielle McNamara 242

Using N-gram and Word Network Features for Native Language Identification
Shibamouli Lahiri and Rada Mihalcea 251

LIMSI's participation to the 2013 shared task on Native Language Identification
Thomas Lavergne, Gabriel Illouz, Aurélien Max and Ryo Nagata 260

Native Language Identification using large scale lexical features
André Lynum 266

The Story of the Characters, the DNA and the Native Language
Marius Popescu and Radu Tudor Ionescu 270

Identifying the L1 of non-native writers: the CMU-Haifa system
Yulia Tsvetkov, Naama Twitto, Nathan Schneider, Noam Ordan, Manaal Faruqui, Victor Chahuneau, Shuly Wintner and Chris Dyer 279

Evaluating Unsupervised Language Model Adaptation Methods for Speaking Assessment
Shasha Xie and Lei Chen 288

Improving interpretation robustness in a tutorial dialogue system
Myroslava Dzikovska, Elaine Farrow and Johanna Moore 293

Detecting Missing Hyphens in Learner Text
Aoife Cahill, Martin Chodorow, Susanne Wolff and Nitin Madnani 300

Applying Machine Translation Metrics to Student-Written Translations
Lisa Michaud and Patricia Ann McCoy 306

Conference Program

Thursday, June 13, 2013

8:45–9:00 Load Presentations

9:00–9:15 Opening Remarks

9:15–9:40 *The Utility of Manual and Automatic Linguistic Error Codes for Identifying Neurodevelopmental Disorders*
Eric Morley, Brian Roark and Jan van Santen

9:40–10:05 *Shallow Semantic Analysis of Interactive Learner Sentences*
Levi King and Markus Dickinson

10:05–10:30 *Building a Large Annotated Corpus of Learner English: The NUS Corpus of Learner English*
Daniel Dahlmeier, Hwee Tou Ng and Siew Mei Wu

10:30–11:00 Break

11:00–11:25 *Developing and testing a self-assessment and tutoring system*
Øistein E. Andersen, Helen Yannakoudakis, Fiona Barker and Tim Parish

11:25–11:45 *Automated Essay Scoring for Swedish*
Robert Östling, André Smolentzov, Björn Tyrefors Hinnerich and Erik Höglin

11:45–12:10 *A Report on the First Native Language Identification Shared Task*
Joel Tetreault, Daniel Blanchard and Aoife Cahill

12:10–1:50 Lunch

1:50–2:40 BEA8 Poster Session A

Applying Unsupervised Learning To Support Vector Space Model Based Speaking Assessment
Lei Chen

Role of Morpho-Syntactic Features in Estonian Proficiency Classification
Sowmya Vajjala and Kaidi Loo

Thursday, June 13, 2013 (continued)

Automated Content Scoring of Spoken Responses in an Assessment for Teachers of English
Klaus Zechner and Xinhao Wang

1:50–2:40 NLI 2013 Poster Session A

Experimental Results on the Native Language Identification Shared Task
Amjad Abu-Jbara, Rahul Jha, Eric Morley and Dragomir Radev

VTEX System Description for the NLI 2013 Shared Task
Vidas Daudaravicius

Feature Space Selection and Combination for Native Language Identification
Cyril Goutte, Serge Léger and Marine Carpuat

Discriminating Non-Native English with 350 Words
John Henderson, Guido Zarrella, Craig Pfeifer and John D. Burger

Maximizing Classification Accuracy in Native Language Identification
Scott Jarvis, Yves Bestgen and Steve Pepper

Recognizing English Learners' Native Language from Their Writings
Baoli LI

NLI Shared Task 2013: MQ Submission
Shervin Malmasi, Sze-Meng Jojo Wong and Mark Dras

NAIST at the NLI 2013 Shared Task
Tomoya Mizumoto, Yuta Hayashibe, Keisuke Sakaguchi, Mamoru Komachi and Yuji Matsumoto

Cognate and Misspelling Features for Natural Language Identification
Garrett Nicolai, Bradley Hauer, Mohammad Salameh, Lei Yao and Grzegorz Kondrak

Exploring Syntactic Representations for Native Language Identification
Ben Swanson

Simple Yet Powerful Native Language Identification on TOEFL11
Ching-Yi Wu, Po-Hsiang Lai, Yang Liu and Vincent Ng

Thursday, June 13, 2013 (continued)

2:40–3:30 BEA8 Poster Session B

Prompt-based Content Scoring for Automated Spoken Language Assessment
Keelan Evanini, Shasha Xie and Klaus Zechner

Automated Scoring of a Summary-Writing Task Designed to Measure Reading Comprehension
Nitin Madnani, Jill Burstein, John Sabatini and Tenaha O'Reilly

Inter-annotator Agreement for Dependency Annotation of Learner Language
Marwa Ragheb and Markus Dickinson

2:40–3:30 NLI 2013 Poster Session B

Native Language Identification with PPM
Victoria Bobicev

Using Other Learner Corpora in the 2013 NLI Shared Task
Julian Brooke and Graeme Hirst

Combining Shallow and Linguistically Motivated Features in Native Language Identification
Serhiy Bykh, Sowmya Vajjala, Julia Krivanek and Detmar Meurers

Linguistic Profiling based on General–purpose Features and Native Language Identification
Andrea Cimino, Felice Dell'Orletta, Giulia Venturi and Simonetta Montemagni

Improving Native Language Identification with TF-IDF Weighting
Binyam Gebrekidan Gebre, Marcos Zampieri, Peter Wittenburg and Tom Heskes

Native Language Identification: a Simple n-gram Based Approach
Binod Gyawali, Gabriela Ramirez and Thamar Solorio

Feature Engineering in the NLI Shared Task 2013: Charles University Submission Report
Barbora Hladka, Martin Holub and Vincent Kriz

Native Language Identification: A Key N-gram Category Approach
Kristopher Kyle, Scott Crossley, Jianmin Dai and Danielle McNamara

Thursday, June 13, 2013 (continued)

Using N-gram and Word Network Features for Native Language Identification
Shibamouli Lahiri and Rada Mihalcea

LIMSI's participation to the 2013 shared task on Native Language Identification
Thomas Lavergne, Gabriel Illouz, Aurélien Max and Ryo Nagata

Native Language Identification using large scale lexical features
André Lynum

The Story of the Characters, the DNA and the Native Language
Marius Popescu and Radu Tudor Ionescu

Identifying the L1 of non-native writers: the CMU-Haifa system
Yulia Tsvetkov, Naama Twitto, Nathan Schneider, Noam Ordan, Manaal Faruqui, Victor Chahuneau, Shuly Wintner and Chris Dyer

3:30–4:00 Break

4:00–4:20 *Evaluating Unsupervised Language Model Adaptation Methods for Speaking Assessment*
Shasha Xie and Lei Chen

4:20–4:40 *Improving interpretation robustness in a tutorial dialogue system*
Myroslava Dzikovska, Elaine Farrow and Johanna Moore

4:40–5:00 *Detecting Missing Hyphens in Learner Text*
Aoife Cahill, Martin Chodorow, Susanne Wolff and Nitin Madnani

5:00–5:20 *Applying Machine Translation Metrics to Student-Written Translations*
Lisa Michaud and Patricia Ann McCoy

5:20–5:30 Closing Remarks