

# Michael Heilman

<http://www.cs.cmu.edu/~mheilman>

## Professional Experience

Civis Analytics, Chicago, Illinois

- Data Scientist, 2015-present

Educational Testing Service, Princeton, New Jersey

- Research Scientist, NLP, 2014-2015
- Associate Research Scientist, NLP, 2011-2014

## Education

Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania

- Ph.D. in Language and Information Technologies, 2007-2011
- Advisor: Noah A. Smith

Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania

- M.S. in Language Technologies, 2005-2007
- Advisor: Maxine Eskenazi

University of Notre Dame, Notre Dame, Indiana

- B.S., Computer Science, with Honors, 2001-2005
- B.A. in Japanese Language, with Honors, 2001-2005

## Awards and Honors

- Educational Testing Service Presidential Award, 2013
- Program in Interdisciplinary Education Research (<http://www.cmu.edu/pier>), 2007-2011
- National Science Foundation Graduate Research Fellowship, 2007-2010
- Siebel Scholar (<http://www.siebelscholars.com>), 2007
- Tau Beta Pi, University of Notre Dame, 2005
- Upsilon Pi Epsilon, University of Notre Dame, 2005

## Languages

Computer Programming: Python (primary), Java, Javascript

Natural: English (native), Japanese (intermediate), Spanish (beginner)

## Conference and Workshop Reviewing

- ACL-HLT
- EACL
- CoNLL
- Workshop on Innovative Use of NLP for Building Educational Applications
- Semantic Evaluation Exercises (SemEval)
- FLAIRS, Applied Natural Language Processing Track
- Workshop on Question Generation
- Conference on Educational Data Mining
- Speech and Language Technology in Education (SLaTE)

## Journal Reviewing

- Journal of Artificial Intelligence Research
- Transactions of the Association for Computational Linguistics
- Computational Linguistics

- ACM Transactions on Intelligent Systems and Technology
- Language Resources and Evaluation

### Other Professional Service

- Organizer, CMU Program for Interdisciplinary Education Research speaker series, 2008-2010
- Organizing Committee Member, Speech and Language Technology in Education (SLaTE), ISCA Tutorial and Research Workshop, 2007
- Co-organizer and presenter, CMU, Pitt, Pittsburgh Company Demo Session, 9<sup>th</sup> International Conference on Spoken Language Processing, 2006

### Teaching Experience

Teaching Assistant, Grammars and Lexicons (11-721), fall 2009

- Course Instructors: Lori Levin and Teruko Mitamura
- Graded assignments, exams, and final projects. For this graduate-level course on linguistics for computer scientists

Teaching Assistant, Natural Language Processing (11-411), spring 2008

- Course Instructor: Noah A. Smith
- Presented 3 lectures on text classification, lexical semantics, and discourse processing
- Co-developed final project on question generation and answering for this undergraduate-level course

Undergraduate Mentoring

- Mentored Rishav Bhowmick, a student from CMU-Qatar participating in undergraduate research, 2009-2010
- Mentored Christine Feeney, an intern in the Pittsburgh Science of Learning Center Summer Internship program, 2007

### Other Educational Experiences

Research Experience for Undergraduates, University of Notre Dame, 2005

- Worked under Dr. Matthias Scheutz on natural language understanding for a social robot, which was presented at the AAAI 2005 Robot Competition

Research Experience for Undergraduates, University of Notre Dame, summer 2004

- Part of SPANIDS project team under Dr. Lambert Schaelicke

Princeton in Ishikawa, Kanazawa, Japan, summer 2003

- Participated in eight-week home-stay and intensive language training program

Hokkaido International Foundation, Hakodate, Japan, summer 2002

- Participated in eight-week home-stay and intensive language training program

University of Notre Dame London Program, London, England, fall 2003

### Doctoral Dissertation

M. Heilman. 2011. Automatic Factual Question Generation from Text. Ph.D. Dissertation, Carnegie Mellon University. CMU-LTI-11-004. (Committee Chair: Noah A. Smith)

### Publications in Refereed Conference Proceedings and Journals

K. Sakaguchi, M. Heilman and N. Madnani. 2015. Effective Feature Integration for Automated Short Answer Scoring. In Proc. NAACL (short papers).

D. Higgins and M. Heilman. 2014. Managing What We Can Measure: Quantifying the

- Susceptibility of Automated Scoring Systems to Gaming Behavior. *Educational Measurement: Issues and Practice*, 33 (3), pp. 36-46.
- M. Heilman, A. Cahill, N. Madnani, M. Lopez, M. Mulholland, and J. Tetreault. 2014. Predicting Grammaticality on an Ordinal Scale. In *Proc. of ACL (short papers)*.
- N. Madnani, M. Heilman, J. Tetreault and M. Chodorow. 2012. Identifying High Level Organizational Elements in Argumentative Discourse. In *Proc. of NAACL-HLT*.
- K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. 2011. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. In *Proc. of ACL, companion volume*.
- D. Yogatama, M. Heilman, B. O'Connor, C. Dyer, B. R. Routledge, and N. A. Smith. 2011. Predicting a Scientific Community's Response to an Article. In *Proc. of EMNLP*.
- M. Heilman and N. A. Smith. 2010. Tree Edit Models for Recognizing Textual Entailments, Paraphrases, and Answers to Questions. In *Proc. of NAACL/HLT*.
- M. Heilman and N. A. Smith. 2010. Good Question! Statistical Ranking for Question Generation. In *Proc. of NAACL/HLT*.
- M. Heilman, K. Collins-Thompson, M. Eskenazi, A. Juffs, L. Wilson. 2010. Personalization of Reading Passages Improves Vocabulary Acquisition. *International Journal of Artificial Intelligence in Education*, Vol. 20 (1).
- A. Kulkarni, M. Heilman, M. Eskenazi, and J. Callan. 2008. Word Sense Disambiguation for Vocabulary Learning. In *Proc. of the 9th international conference on Intelligent Tutoring Systems*.
- M. Heilman, K. Collins-Thompson, J. Callan, and M. Eskenazi. 2007. Combining lexical and grammatical features to improve readability measures for first and second language texts. In *Proc. NAACL-HLT*.
- M. Heilman, A. Juffs, and M. Eskenazi. 2007. Choosing reading passages for vocabulary learning by topic to increase intrinsic motivation. In *Proc. of AIED*.
- M. Heilman, K. Collins-Thompson, J. Callan, and M. Eskenazi. 2006. Classroom success of an intelligent tutoring system for lexical practice and reading comprehension. In *Proc. of the 9th International Conference on Spoken Language Processing*.
- K. M. Eberhard, M. Scheutz, and M. Heilman. 2005. An empirical and computational test of linguistic relativity. In *Proc. of the Cognitive Science Society Conference*.

### Publications in Workshop and Shared Task Proceedings

- M. Heilman and N. Madnani. 2015. The Impact of Training Data on Automated Short Answer Scoring Performance. In *Proc. 10th Workshop on Innovative Use of Natural Language Processing for Building Educational Applications (BEA)*.
- A. Loukina, K. Zechner, L. Chen and M. Heilman. 2015. Feature selection for automated speech scoring. In *Proc. 10th Workshop on Innovative Use of Natural Language Processing for Building Educational Applications (BEA)*.
- T. Zesch, M. Heilman and A. Cahill. 2015. Reducing Annotation Efforts in Supervised Short Answer Scoring. In *Proc. 10th Workshop on Innovative Use of Natural Language Processing for Building Educational Applications (BEA)*.
- B. Beigman Klebanov, B. Leong, M. Heilman, and M. Flor. 2014. Different Texts, Same Metaphors: Unigrams and Beyond. In *Proc. of the 2nd Workshop on Metaphor in NLP (ACL-Meta4NLP)*.
- Y. Song, M. Heilman, B. Beigman Klebanov and P. Deane. 2014. Applying Argumentation Schemes for Essay Scoring. In *Proc. of the First Workshop on Argumentation Mining*.
- M. Heilman and N. Madnani. 2013. ETS: Domain Adaptation and Stacking for Short Answer Scoring. *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on*

Semantic Evaluation (SemEval).

- M. Heilman and N. Madnani. 2013. HENRY-CORE: Domain Adaptation and Stacking for Text Similarity. Proceedings of the Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity.
- M. Heilman and N. Madnani. 2012. ETS: Discriminative Edit Models for Paraphrase Scoring. Proc. of the 6th International Workshop on Semantic Evaluation (SemEval).
- M. Heilman, A. Cahill, and J. Tetreault. 2012. Precision Isn't Everything: A Hybrid Approach to Grammatical Error Detection. Proc. of the Seventh Workshop on Innovative Use of NLP for Building Educational Applications.
- M. Heilman and N. A. Smith. 2010. Extracting Simplified Statements for Factual Question Generation. In Proc. of the 3rd Workshop on Question Generation.
- M. Heilman and N. A. Smith. 2010. Rating Computer-Generated Questions with Mechanical Turk. In Proc. of the NAACL/HLT workshop on Creating Speech and Language Data With Amazon's Mechanical Turk.
- M. Heilman and N. A. Smith. 2009. Ranking automatically generated questions as a shared task. In Proc. of the 2nd Workshop of Question Generation.
- N. A. Smith, M. Heilman, and R. Hwa. 2008. Question generation as a competitive undergraduate course project. In Proc. of the Workshop on the Question Generation Shared Task and Evaluation Challenge.
- M. Heilman and M. Eskenazi. 2008. Self-assessment in vocabulary tutoring. In Proc. of the 9th International Conference on Intelligent Tutoring Systems.
- C. Feeney and M. Heilman. 2008. Automatically generating and validating reading-check questions. In Proc. of the 9th International Conference on Intelligent Tutoring Systems.
- J. Pino, M. Heilman, and M. Eskenazi. 2008. A selection strategy to improve cloze question quality. In Proc. of the Workshop of Intelligent Tutoring Systems for Ill-Defined Domains. 9th International Conference on Intelligent Tutoring Systems.
- M. Heilman, K. Collins-Thompson, and M. Eskenazi. 2008. An analysis of statistical models and features for reading difficulty prediction. In Proc. of The 3rd Workshop on Innovative Use of NLP for Building Educational Applications.
- M. Heilman, L. Zhao, J. Pino, and M. Eskenazi. 2008. Retrieval of reading materials for vocabulary and reading practice. In Proc. of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications.
- M. Heilman and M. Eskenazi. 2007. Application of automatic thesaurus extraction for computer generation of vocabulary questions. In Proc. of the SLaTE Workshop on Speech and Language Technology in Education.
- M. Heilman and M. Eskenazi. 2006. Language learning: Challenges for intelligent tutoring systems. In Proc. of the Workshop of Intelligent Tutoring Systems for Ill-Defined Tutoring Systems. Eight International Conference on Intelligent Tutoring Systems.

## Technical Reports

- M. Heilman and K. Sagae. 2015. Fast Rhetorical Structure Theory Discourse Parsing. arXiv:1505.02425 [cs.CL].
- D. Higgins, C. Brew, M. Heilman, R. Ziai, L. Chen, A. Cahill, M. Flor, N. Madnani, J. Tetreault, D. Blanchard, D. Napolitano, C. M. Lee, J. Blackmore. 2014. Is getting the right answer just about choosing the right words? The role of syntactically-informed features in short answer scoring. arXiv:1403.0801 [cs.CL].
- B. O'Connor and M. Heilman. 2013. ARKref: a rule-based coreference resolution system. arXiv:1310.1975 [cs.CL].
- M. Heilman and N. Madnani. 2012. Topical Trends in a Corpus of Persuasive Writing. ETS

Research Report Series, RR-12-19.

- D. Yogatama, M. Heilman, B. O'Connor, C. Dyer, B. R. Routledge, and N. A. Smith. 2011. Predicting Responses and Discovering Social Factors in Scientific Literature. Language Technologies Institute, Carnegie Mellon University Technical Report CMU-LTI-11-015.
- N. Schneider, R. Hwa, P. Gianfortoni, D. Das, M. Heilman, A. W. Black, F. L. Crabbe, and N. A. Smith. 2010. Visualizing Topical Quotations Over Time to Understand News Discourse. Language Technologies Institute, Carnegie Mellon University Technical Report CMU-LTI-10-013.
- M. Heilman and N. A. Smith. 2009. Question Generation via Overgenerating Transformations and Ranking. Language Technologies Institute, Carnegie Mellon University Technical Report CMU-LTI-09-013.

### Unpublished Presentations

- M. Heilman and J. Tetreault. 2012. Using automated essay scoring to analyze the GURT data set. Georgetown Roundtable on Languages and Linguistics.
- M. Heilman and N. A. Smith. 2009. Tree Edit Models for Sentence Pair Tasks. Language Technologies Institute Student Research Symposium.
- M. Heilman and M. Eskenazi. 2006. Authentic, Individualized Practice for English as a Second Language Vocabulary. Presented at Interfaces of Intelligent Computer-Assisted Language Learning Workshop at the Ohio State University, Columbus, OH, 12/06. Unpublished.
- A. Juffs, M. Eskenazi, J. Callan, L. Wilson, M. Heilman, K. Collins-Thompson, and T. Pelletreau. 2006. Promoting robust learning of vocabulary through computer assisted language learning. Meeting of the American Association for Applied Linguistics, Montreal.
- A. Juffs, L. Wilson, M. Eskenazi, J. Callan, J. Brown, K. Collins-Thompson, M. Heilman, T. Pelletreau, and J. Sanders. 2006. Robust learning of vocabulary: investigating the relationship between learner behaviour and the acquisition of vocabulary (poster). At The 40th Annual TESOL Convention and Exhibit (TESOL 2006).

### Invited Talks

- M. Heilman. 2010. Using Natural Language Processing to Develop Instructional Content. Presented at Darmstadt University of Technology, Darmstadt, Germany.
- M. Heilman. 2010. Automatic Factual Question Generation. Presented at Darmstadt University of Technology, Darmstadt, Germany.