

Courtney Napoles

Center for Language and Speech Processing
Johns Hopkins University
3400 N. Charles St., Hackerman 225
Baltimore, MD 21218

a.k.a Courtney Napoles Cohen
napoles@cs.jhu.edu
www.cs.jhu.edu/~napoles
(917) 319-7534

RESEARCH INTERESTS

Applying machine learning and natural language processing techniques to problems involving user-generated dialogue, social media, noisy text, education, crowdsourcing, language resource creation, and evaluation.

EDUCATION

Johns Hopkins University, Baltimore, MD

- Ph.D., Computer Science expected September 2017
Thesis: *Generation and Evaluation of Sentential Paraphrases*
- M.S.E., Computer Science December 2012
Thesis: *Computational Approaches to Shortening and Simplifying Text*
- Advisors: Chris Callison-Burch and Benjamin Van Durme

Columbia University, New York, NY

- Post-baccalaureate program in Computer Science 2008–2009

Princeton University, Princeton, NJ

- A.B., Psychology August 2005
Certificate in Linguistics
Thesis: *Conceptual Combination of Nominal Compounds by French and English Speakers*

RESEARCH EXPERIENCE

Yahoo Research, New York, NY

Summer 2016

- Intern Research Scientist
Mentors: Aasish Pappu and Joel Tetreault
Analyzed user-generated comment threads to identify constructive sub-dialogues using neural nets and graphical and linear models, and compiled a novel corpus of annotated comment threads.

Educational Testing Services, Princeton, NJ

Summer 2015

- Research Intern
Mentors: Aoife Cahill and Nitin Madnani
Investigated computer generated feedback for student writing, quantified the impact of grammatical errors on the NLP pipeline, and investigated methods for comparing unaligned trees.

Johns Hopkins University, Baltimore, MD

2010–2013, 2014–present

- Research Assistant
Mentors: Chris Callison-Burch and Benjamin Van Durme
Projects include exploring sentential paraphrasing using techniques from machine translation, machine learning, linguistic theory, and crowdsourcing; best practices for evaluating automatically generated text; and educational applications of NLP, including automatic essay scoring and error correction.

Princeton University Cognitive Science Laboratory, Princeton, NJ

Summer 2004

- Research Assistant
Mentor: Christiane Fellbaum
 Assisted in data preparation and corpus annotation for a word-sense disambiguation task.

AWARDS

National Science Foundation Graduate Research Fellowship, 2011

PROFESSIONAL EXPERIENCE

Freelance Editor, New York, NY & Baltimore, MD 2006–2009

- Indexing, layout, book packaging, and developmental, line, and copy editing.

Da Capo Press (formerly **Marlowe Books**), New York, NY

- *Assistant Editor* 2007–2008
 - *Editorial Assistant* 2006–2007
- Acquired and edited non-fiction books through all stages of production; performed market research and shaped manuscripts to target specific markets; recruited and managed college interns.

Strong Literary Agency, New York, NY 2005–2006

- *Assistant Agent*
 Evaluated, packaged, and marketed manuscripts to book publishers and negotiated contracts.

Applause Books, New York, NY 2005–2006

- *Editorial and Production Assistant*
 Prepared manuscripts for publication, including copy editing, proofreading, indexing, and layout.

REFEREED PUBLICATIONS

Journal Articles

1. K. Sakaguchi, **C. Napoles**, M. Post, and J. Tetreault. *Reassessing the Goals of Grammatical Error Correction: Fluency Instead of Grammaticality*. Transactions of the Association for Computational Linguistics (TACL), 2016.
 * First two authors made equal contributions.
2. W. Xu, **C. Napoles**, E. Pavlick, Q. Chen, and C. Callison-Burch. *Optimizing Statistical Machine Translation for Text Simplification*. TACL, 2016.
3. W. Xu, C. Callison-Burch, and **C. Napoles**. *Problems in Current Text Simplification Research: New Data Can Help*. TACL, 2015.

Main Conference Proceedings

4. **C. Napoles**, A. Pappu, and J. Tetreault. *Automatically Identifying Good Conversations Online (Yes, They Do Exist!)*. 11th International AAAI Conference on Web and Social Media (ICWSM), 2017.
5. **C. Napoles**, K. Sakaguchi, and J. Tetreault. *JFLEG: A Fluency Corpus and Benchmark for Grammatical Error Correction*. 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2017.

6. **C. Napoles**, K. Sakaguchi, and J. Tetreault. *There's No Comparison: Reference-less Evaluation Metrics in Grammatical Error Correction*. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016.
7. **C. Napoles**, K. Sakaguchi, M. Post, and J. Tetreault. *Ground Truth for Grammatical Error Correction Metrics*. 53rd Annual Meeting of the Association for Computational Linguistics (ACL), 2015.
8. J. Ganitkevitch, C. Callison-Burch, **C. Napoles**, and B. Van Durme. *Learning Sentential Paraphrases from Bilingual Parallel Corpora for Text-to-Text Generation*. EMNLP, 2011.

Demonstrations

9. **C. Napoles**, C. Callison-Burch, M. Post. *Sentential Paraphrasing as Black-Box Machine Translation*. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2016.

Workshop Proceedings

10. **C. Napoles** and C. Callison-Burch. *Systematically Adapting Machine Translation for Grammatical Error Correction*. Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2017.
11. K. Sakaguchi, **C. Napoles**, and J. Tetreault. *GEC into the Future: Where Are We Going and How Do We Get There?* BEA, 2017.
12. **C. Napoles**, J. Tetreault, E. Rosato, B. Provenzale, and A. Pappu. *Finding Good Conversations Online: The Yahoo News Annotated Corpus*. 11th Linguistic Annotation Workshop (LAW-XI), 2017.
13. **C. Napoles**, A. Cahill, and N. Madnani. *The Effect of Multiple Grammatical Errors on Processing Non-Native Writing*. BEA, 2016.
14. V. Daudaravicius, R.E. Banchs, E. Volodina, and **C. Napoles**. *A report on the automatic evaluation of scientific writing shared task*. BEA, 2016.
15. **C. Napoles** and C. Callison-Burch. *Automatically Scoring Freshman Writing: A Preliminary Investigation*. BEA, 2015.
16. **C. Napoles**, M. Gormley, and B. Van Durme. *Annotated Gigaword*. Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, 2012.
17. **C. Napoles**, C. Callison-Burch, and B. Van Durme. *Evaluating Sentence Compression: Pitfalls and Suggested Remedies*. Workshop on Monolingual Text-To-Text Generation, 2011.
18. **C. Napoles**, C. Callison-Burch, J. Ganitkevitch, and B. Van Durme. *Paraphrastic Sentence Compression with a Character-based Metric: Tightening without Deletion*. Workshop on Monolingual Text-To-Text Generation, 2011.
19. **C. Napoles** and M. Dredze. *Learning Simple Wikipedia: A Cogitation in Ascertaining Abecedarian Language*. Workshop on Computational Linguistics and Writing, 2010.

Abstracts Presented

20. W. Xu, **C. Napoles**, E. Pavlick, Q. Chen, and C. Callison-Burch. *Optimizing Statistical Machine Translation for Simplification*. Mid-Atlantic Student Colloquium (MASC), 2016.
21. **C. Napoles**. *Predicting College Essay Scores in the Presence of Grader Bias*. 10th Women in Machine Learning Workshop (WiML), 2015.
22. **C. Napoles**, C. Callison-Burch, J. Ganitkevitch, and B. Van Durme. *Paraphrastic Sentence Compression*. MASC, 2011.

SOFTWARE & RESOURCES

Datasets: Collected and released 3 annotated datasets

- *Annotated Gigaword*: 4 billion words of news-wire text with layers of NLP annotations [16]
- *Yahoo News Annotated Comments Corpus*: 9 thousand comments on news articles annotated with argumentation and sentiment labels [12]
- *JFLEG*: 1.5 thousand sentences from English Language Learners, corrected for grammatical errors [5]

Software: Developed and released several NLP tools in Java and Python (<https://github.com/cnap>)

RELATED ACTIVITIES

Shared Task Co-organizer

- First Automatic Evaluation of Scientific Writing Shared Task (AESW), 2016

Reviewer

- *Journals*: Computer Speech and Language (2017), Language Resources and Evaluation (2017), Computational Linguistics (co-reviewer: 2016)
- *Conferences*: ACL (2016, 2017), COLING (2016), EACL (2017), EMNLP (2016, 2017, secondary reviewer: 2011), AAAI (secondary reviewer: 2011, 2015, 2016)
- *Workshops*: *CL Student Research Workshop (2014–2017), Abusive Language Workshop (2017), Workshop on Innovative Use of NLP for Building Educational Applications (BEA; 2016, 2017), Workshop on Natural Language Processing NLP Techniques for Educational Applications (NLP-TEA; 2016)

Teaching Assistant

- *Information Retrieval*
Johns Hopkins University, Prof. David Yarowsky, Spring 2010
Johns Hopkins CLSP Summer School, Prof. Doug Oard, Summer 2011
Columbia University, Prof. Dragomir Radev, Spring 2009
- *Data Structures*
Johns Hopkins University, Prof. Joanne Selinski, Fall 2009

Co-chair, Half-minute Madness @ Conference on Empirical Methods in NLP (EMNLP) 2016

Outreach

- Mentor, STEM Achievement in Baltimore City Schools (SABES), 2015–2016
- Mentor, Thread (formally Incentive Mentoring Program), 2012–2013
- Site Organizer/Recruiter, North American Computational Linguistics Olympiad (NACLO), 2009–2012
- School visits and STEM presentations with JHU Center for Educational Outreach, 2009–2012

SKILLS

Programming: Python, Java, Perl, R, awk, bash

Machine learning: Scikit-Learn, CRFSuite; familiar with Gensim/Word2vec, Tensor Flow, Theano

Other: Git, Matplotlib, NumPy, Pandas, SciPy, Mechanical Turk; familiar with AWS, Hadoop, SQL

Related coursework: Machine learning, parallel programming, probability, statistics, syntax, semantics, artificial intelligence

Languages: English (native), French and Spanish (intermediate)