

KRISTY HOLLINGSHEAD SEITZ, PH.D.

www.umiacs.umd.edu/ hollingk

hollingk@gmail.com

OBJECTIVE Research scientist/programmer position working on interesting, cutting-edge problems in natural language processing, with particular interest and experience in parsing, machine translation, and language modeling.

EDUCATION

- Ph.D.** Computer Science Engineering Oregon Health & Science University, 2010
Dissertation: *Formalizing the use and characteristics of constraints in pipeline systems* GPA: 3.89/4.0
Committee: Brian Roark (chair), Peter Heeman, Deniz Erdogmus, Mark Johnson
- M.S.** Computer Science Engineering Oregon Health & Science University, 2004
Concentration: *Spoken Language Systems* GPA: 3.93/4.0
- B.A.** English–Creative Writing University of Colorado at Boulder, 2000
Minor: Computer Science GPA: 3.91/4.0

SKILLS

Languages:	Hadoop/Cloud computing	Quick learner
Proficient in: C, Python, Unix shell scripting	Machine learning techniques	Experienced instructor
Experienced in: C++, Java, Spanish	Large-scale data	Highly skilled presenter
Familiar with: MATLAB, JavaScript	Statistical analysis	Team player
Operating Systems: UNIX, Linux, Mac, Windows	System optimization	

WORK EXPERIENCE

Post-doctoral Research Associate, 9/11–present *University of Maryland — College Park, MD*

- Led the UMD machine translation team on (1) Chinese-English translation of newswire and web genre data, for the DARPA BOLT program; (2) Arabic-English translation of newswire, web, and broadcast conversation genre data, for the DARPA GALE program; and (3) Haitian-English SMS translation. Implemented a transformation technique for “single-language” translation, as a method to clean noisy data given a small amount of cleaned data.
- Designed and implemented graph-based representation of machine translation output to analyze the density and diversity of an MT n -best list. Applied novel technique of “pipeline iteration” to machine translation, producing a substantial improvement in BLEU. Reduced computational complexity of machine translation, in collaboration with the Center of Excellence (COE) at Johns Hopkins University.
- Taught Computational Linguistics I, Fall 2011

Graduate Research Assistant, 9/03–8/11 *Oregon Health & Science University — Beaverton, OR*

- Researched pipeline systems, wherein data is processed by a sequence of stages such that the output from one stage provides the input to the next. Projects included:
 1. Reducing run-time complexity of context-free parsing from $O(n^3)$ to $O(n^2)$ in the worst case, with observably linear ($O(n)$) performance;
 2. Applying graph theory metrics to analyze the density, area, diversity, and uniformity of the search space explored at each stage of a pipeline system;
 3. Developing the novel, generalizable technique of “pipeline iteration,” and applying to a parsing pipeline, resulting in an improvement upon state-of-the-art parsing accuracy;
 4. Designing techniques to automatically extract and calculate syntactic complexity measures for automated diagnosis of cognitive impairment;
 5. Exploring various language modeling techniques to improve assistive technology for “locked-in” patients.
- Developed a finite-state parser which performs competitively on several shallow parsing tasks. Discovered and corrected several flaws in shallow parsing evaluation metrics. Achieved new state-of-the-art levels of accuracy on several shallow parsing tasks.
- Taught Computational Linguistics II, Winter 2010

Visiting Student/Summer Workshop Participant, 6/08–8/08 *Johns Hopkins University — Baltimore, MD*

- Participated in the ‘Multilingual Spoken Term Detection: Finding and Testing New Pronunciations’ workshop team. Developed various methods to extract English, Chinese, and Arabic transliteration candidates from comparable (Gigaword) using Hadoop cloud cluster; the only workshop participant to use the cloud cluster.

Research Assistant, 9/02–9/03 *Oregon Health & Science University — Beaverton, OR*

- Developed dynamic grammar-based language models to automatically detect and diagnose stuttered speech using a statistical speech recognition system.
- Built models to represent the underlying structure of human-human dialogues and the roles of different speech acts within a dialogue.
- Analyzed human-subject data on how users adapted to computer systems in changing environments.

Technical Editor, 7/00–11/02 *IBM — Boulder, CO*

SELECT PUBLICATIONS

JOURNAL ARTICLES

- Brian Roark, Kristy Hollingshead, Nathan Bodenstab. Submitted 2011. Finite-state chart constraints for reduced complexity context-free parsing pipelines. Submitted to *Computational Linguistics*.
- Brian Roark, Margaret Mitchell, John-Paul Hosom, Kristy Hollingshead and Jeffrey A. Kaye. 2011. Spoken language derived measures for detecting Mild Cognitive Impairment. *IEEE Transactions on Audio, Speech and Language Processing*, 19(7).
- Fan Yang, Peter A. Heeman, Kristy Hollingshead, and Susan E. Strayer. 2008. DialogueView: Annotating dialogues in multiple views with abstraction. In *Natural Language Engineering*, 14(1).

REFEREED CONFERENCE PROCEEDINGS

- Nathan Bodenstab, Kristy Hollingshead, and Brian Roark. 2011. Unary constraints for context-free parsing. In *Proceedings of ACL*.
- Ting Qian, Kristy Hollingshead, Su-youn Yoon, Kyoung-young Kim, and Richard Sproat. 2010. A Python toolkit for universal transliteration. In *Proceedings of LREC*.
- Brian Roark and Kristy Hollingshead. 2009. Linear complexity context-free parsing pipelines via chart constraints. In *Proceedings of HLT-NAACL*.
- Brian Roark and Kristy Hollingshead. 2008. Classifying chart cells for quadratic complexity context-free inference. In *Proceedings of COLING*.
- Kristy Hollingshead and Brian Roark. 2007. Pipeline Iteration. In *Proceedings of ACL*.
- Kristy Hollingshead, Seeger Fisher, and Brian Roark. 2005. Comparing and combining finite-state and context-free parsers. In *Proceedings of HLT/EMNLP*.

REFEREED WORKSHOP PROCEEDINGS

- Vlad Eidelman, Kristy Hollingshead, and Philip Resnik. 2011. Noisy SMS machine translation in low-density languages. In *Proceedings of the EMNLP Sixth Workshop on Statistical Machine Translation (WMT)*.
- Christian Monson, Kristy Hollingshead, and Brian Roark. 2010. Simulating morphological analyzers with stochastic taggers for confidence estimation. In *Multilingual Information Access Evaluation, Vol. I, 10th Workshop of the Cross-Language Evaluation Forum (CLEF 2009)*. Lecture Notes in Computer Science, Springer.
- Brian Roark, Margaret Mitchell, and Kristy Hollingshead. 2007. Syntactic complexity measures for detecting Mild Cognitive Impairment. In *Proceedings of the Workshop on Biomedical Natural Language Processing (BioNLP)*.

PROFESSIONAL AFFILIATIONS AND AWARDS

Reviewer for several top-tier conferences, 2006-present.

Member of Association for Computational Linguistics (ACL), 2006-present.

National Science Foundation Graduate Research Fellow (NSF GRF), 2004-2007.

U.S. Citizen.

REFERENCES Available upon request.