Kenton Murray

kenton@jhu.edu
www.kentonmurray.com

Johns Hopkins University
Center for Language and Speech Processing

EDUCATION:

- **University of Notre Dame** — South Bend, IN
  Ph.D. in Computer Science — August 2014 – May 2020
  Dissertation: "Learning Hyperparameters for Neural Machine Translation"
  Advisor: David Chiang
  Graduate GPA: 4.0

- **Carnegie Mellon University School of Computer Science** — Pittsburgh, PA
  M.S. Language Technologies — August 2011 – August 2013
  Master's Thesis: "A Semantic Scan Statistic for Novel Disease Outbreak Detection"
  Advisor: Daniel Neill
  * Course Work Includes: Machine Learning, Machine Learning for Policy, Probabilistic Graphical Models, Language and Statistics, Social Media Analysis, Software Engineering, Algorithms for NLP, Grammars and Lexicons, Seminar on Endangered Languages

- **Princeton University** — Princeton, NJ
  Major: B.S.E. in Computer Science — September 2005 - June 2009
  Certificates (Minors): Robotics and Intelligent Systems, Finance
  Senior Thesis: **"**Summarization by Latent Dirichlet Allocation"
  Advisor: David Blei
  ***** Course Work Includes: Algorithms and Data Structures, Advanced Programming Techniques, Logic Design, Human Computer Interface, Robotics and Intelligent Systems, Corporate Finance, Micro and Macroeconomics, Programming Languages, Calculus through Multivariable, Linear Algebra, and Statistics

Awards and Honors:

- Outstanding Teaching Assistant - University of Notre Dame's Computer Science and Engineering Department. Awarded to One TA per semester out of approximately 30
- Yelp Dataset Winner. Awarded by Yelp as the most innovative use of their data.
- Best Poster Presentation – CMU Language Technologies Institute's Annual Student Research Symposium. Selected from a committee of Faculty and Students with expertise ranging across Machine Learning, NLP, Computer Science, and Language Technologies
- Best performing systems in spoken and written translation in both directions for Arabic/English Language Pair at IWSLT 2013.
- Masters Student Representative – Elected to be the sole representative for 50 Masters Students'. Voiced student concerns to the faculty. Held regular meetings regarding student's issues with classes.
- Sigma Xi Scientific Research Society Inductee for Senior Thesis on using Statistical NLP to Automatically Summarize Text Documents. Required Faculty Nomination. Currently one of the leading algorithms in the field.
- Won the Princeton Junior Orator Award for a Speech on Revolutionary America

RESEARCH:

Johns Hopkins University — January 2020 - Present
Center for Language and Speech Processing (CLSP) — P.I. Ben Van Durme
Postdoctoral Fellow

Working on Cross-Lingual Computational Semantics with a particular focus on Decompositional Semantics. IN particular, focusing on modeling challenges while also creating new datasets. Helping to run JHU's IARPA BETTER and DARPA AIDA/KAIROS projects. Mentoring PhD, Masters, and Undergrad students.

Notre Dame NLP Lab August 2014 – December 2019
PhD Student P.I. David Chiang
Worked on Neural and Statistical Machine Translation with a particular emphasis on morphologically rich languages and languages in a low-resource setting. Focusing on developing more robust and easy to use neural models for research in Machine Translation and broader NLP. In particular, addressing issues with understandability and ability to use these methods without exhaustive parameter tuning and grid-search approaches. Involved in entire DARPA LORELEI Project over 4 years.

Aristo Project June 2016 – September 2016
Allen Institute for Artificial Intelligence (AI2) Mentor Jayant Krishnamurthy
Research Intern
Worked on Automatic Question Answering for standardized science tests given to grade school students. In particular, questions that relied on referencing diagrams to answer. Replaced state-of-the-art log-linear model with neural architecture. Resulted in 50% performance gain. Combined probabilistic programming with neural networks in novel way.

Arabic Language Technologies Lab September 2013 – June 2014
Qatar Computing Research Institute P.I. Stephan Vogel
Research Associate
Working on Statistical Machine Translation. In particular, investigating statistical methods for phrase alignment and extraction. Specific focus on morophologically complex languages, such as Arabic. Developing open source phrase alignment toolkit for phrase based MT.

Event and Pattern Detection Lab August 2011 – August 2013
CMU School of Computer Science P.I. Daniel Neill
Graduate Research Assistant
Developed unsupervised ML techniques on free text in Hospital Emergency Department Records for Anomalous Pattern Detection. Expanded state-of-the-art pattern detection methods from reliance on expertly labeled data to natural language. Focused on robustness in graphical models and capturing low-frequency terms. Experiments include various Variational Inference Methods, Gibbs Samplers, EM, and corpus preprocessing steps. Received Full Graduate Fellowship for tuition and stipend equivalent to PhD candidates'.

CMU Machine Translation Lab Spring and Summer 2013
CMU's group focusing on MT. Submitted WMT 2013 paper. Part of reading and discussion group for Machine Translation. Discussed and presented work from recent publications.

Automatic Text Summarization using Topic Models September 2008 – May 2009
Independent Research Advised by David Blei
Investigated the use of Probabilistic Topic Models for the use in Summarizing Textual Documents. Investigated a non-standard summarization corpus. Evaluated using Crowdsourcing which had no prior applications in automatic summarization. Work resulted in induction into Sigma Xi.

PUBLICATIONS:

2020:
"The JHU Submission to the 2020 Duolingo Shared Task on Simultaneous Translation and Paraphrase for Language Education" Huda Khayrallah, Jacob Bremerman, Arya D. McCarthy, Kenton Murray, Winston Wu, and Matt Post, Workshop on Neural Generation and Translation, July 2020, Seattle, Wa

"Collecting Verified COVID-19 Question Answer Pairs" Adam Poliak et al., NLP COVID-19 Workshop, November 2020, Dominican Republic

"Universal Decompositional Parsing at the Syntax-Semantics Interface" Elias Stengel-Eskin, Kenton Murray, Sheng Zhang, Aaron White, Benjamin Van Durme, Under Review

"Gradual Fine-Tuning for Low-Resource Domain Adaptation" Haoran Xu, Seth Ebner, Mahsa Yarmohammadi, Aaron White, Benjamin Van Durme, Kenton Murray, Under Review

"An Expectation-Maximization Approach to BPE Segmentation" Kenton Murray and David Chiang, Under Review

"Towards Model Diagnosis of Probabilistic Sequence Models with Likelihood-Automated Evaluation Metrics Scatter Plots" Chu-Cheng Lin, Benjamin Van Durme, and Kenton Murray, Under Review

2019:
"Auto-Sizing the Transformer Network: Improving Speed, Efficiency, and Performance for Low-Resource Machine Translation" Kenton Murray, Jeffery Kinnison, Toan Q. Nguyen, Walter Scheirer, and David Chiang, Workshop on Neural Generation and Translation, November 2019, Hong Kong

"Efficiency through Auto-Sizing: Notre Dame NLP's Submission to the WNGT 2019 Efficiency Task" Kenton Murray, Brian DuSell, and David Chiang, Workshop on Neural Generation and Translation, November 2019, Hong Kong

2018:
"Correcting Length Bias in Neural Machine Translation" Kenton Murray and David Chiang, Conference on Machine Translation, November 2018, Brussels, Belgium.

2017:
"Incident-Driven Machine Translation and Name Tagging for Low-Resource Languages" Ulf Hermjakob et al. Journal of Machine Translation, October 2017.

2016:
"Probabilistic Neural Programs" Kenton W. Murray and Jayant Krishnamurthy. Neural Abstract Machines and Program Induction, December 2016, Barcelona, Spain.

2015:
"Auto-Sizing Neural Networks: With Applications to n-gram Language Models" Kenton Murray and David Chiang. Proceedings of the International Conference on Empirical Methods in Natural Language Processing, September 2015, Lisbon, Portugal.

2014:
"Collaborative Bayesian Filtering" Alex Beutel, Kenton Murray, Alex Smola, and Christos Faloutsos. Proceedings of the International World Wide Web Conference, April 2014, Seoul, Korea.

2013:
"QCRI at IWSLT 2013: Experiments in Arabic-English and English-Arabic Spoken Language Translation". Hassan Sajjad, Fransico Guzman, Preslav Nakov, Ahmed Abdelali, Kenton Murray, Fahad Al Obaidli, Stephan Vogel. Proceedings of IWSLT. December 2013, Heidelberg, Germany. (Best Arabic-English and English-Arabic Systems)

"The CMU machine translation systems at WMT 2013: Syntax, synthetic translation options, and pseudo-references" Waleed Ammar, Victor Chahuneau, Michael Denkowski, Greg Hanneman, Wang Ling, Austin Matthews, Kenton Murray, Nicola Segall, Yulia Tsvetkov, Alon Lavie, Chris Dyer. Proceedings of WMT. August 2013, Sofia, Bulgaria.

"A Semantic Scan Statistic for Novel Disease Outbreak Detection". Kenton Murray and Daniel Neill. Carnegie Mellon University Master's Thesis (available online). August 2013, Pittsburgh, PA.

2009:
"Summarization by Latent Dirichlet Allocation". Kenton Murray and David Blei. Princeton University Senior Thesis (available online and through Princeton University Libraries). May 2009, Princeton, NJ.

Invited Talks:

University of Sothern California Information Sciences Institute *Marina del Rey, CA* April 18, 2019
Airforce Research Lab *Dayton, OH* March 11, 2019
Johns Hopkins University *Baltimore, MD* February 25, 2019
Army Research Lab *Adelphi, MD* February 5, 2019
University of Sothern California Information Sciences Institute *Marina del Rey, CA* January 6, 2017

PROFESSIONAL EXPERIENCE:

NLP Start-up IP Consultant *Pittsburgh, PA* Summer 2014
Hired by a law firm to compare proprietary and open-source code bases for potential IP overlaps. Evaluated NLP pipelines in Java and Python for potential derivative works. Wrote a report as the technical expert for the due diligence of a start-up acquisition.

Safaba Translation Services *Software Developer (Part-time) | Pittsburgh, PA* July 2012 – July 2013
Developed Machine Translation Service for Fortune 500 companies using Moses providing translations for dozens of languages. Worked on proprietary MT software that improves state of the art translation for domain specific applications. Developed service monitoring framework.

Lighter Capital *Software Development Engineer Intern | Seattle, WA* Summer 2011
Implemented Build Automation System, Started Proper QA protocols. Sole developer of a proprietary asset pricing software for mitigating risk in firm's investment portfolio and to determine loan conditions to small businesses.

Microsoft *Software Development Engineer in Test | Bellevue, WA* Summer Intern: 2008, Full-Time: 2009 - 2010
Dynamics CRM Online. Part of team developing the backend of a cloud service. Sole tester on the Backup Service for all customer data running every 10 minutes to maintain SLA. Authored determining report for the GM that led to expanding product to 41 languages and 40 countries after evaluating feasibility affecting business strategy for the following year. Online Test Expert in Upgrade and Patching Scenarios for the product. In charge of deploying new releases to Asian and European Data Centers. Programmed Automated Test Suites in C# and the .Net used daily on the latest build (BVTs). Focused on preventing security flaws and viruses, consumer ease of use, and stress testing.

LECTURES and TEACHING:

University of Notre Dame Computer Science NLP Class *TA/Guest Lecturer | South Bend, IN* Fall 2016
TA for an upper-level undergraduate/graduate class on Natural Language Processing. Held weekly office hours, graded and helped write assignments, as well as helped develop with course structure.

University of Notre Dame Computer Science NLP Class *TA/Guest Lecturer | South Bend, IN* Spring 2015
Won Department's TA Award
Sole TA for an upper-level undergraduate/graduate class on Natural Language Processing. Held weekly office hours, graded and helped write assignments and tests, as well as helped develop with course structure. Wrote required background readings, and gave lectures for the classes focusing on CRFs and Linear Regression.

University of Notre Dame Philosophy Department *Guest Lecturer | South Bend, IN* October 11, 2017

Guest Lecturer in Digital Technology, Society, and Ethics. Presented instances where Machine Learning and Artificial Intelligence can and have been used in unethical ways. Focused on how choosing a dataset matters.

Westville Education Initiative (Holy Cross College) *General TA | Westville, IN* Spring 2017
General TA for all college level courses at the Westville Correctional Facility. Helping Incarcerated Felons to get University Degrees.

Carnegie Mellon University Materials Science Department *Guest Lecturer | Pittsburgh, PA* April 1, 2013
Guest Lecturer on Introduction to Text Mining. Course on basic machine learning techniques on impacts of material for advanced undergraduates (27-566). Covered basic challenges and methods for NLP and the impacts of statistics within corpora.

Carnegie Mellon University Heinz College *Guest Lecturer | Pittsburgh, PA* December 2011 and 2012
Gave guest lectures to a Master's Level Course for Public Policy Students (Large Scale Data Analysis). Presented current state-of-the-art methods for using ML techniques on Natural Language Data for Public Health Surveillance.

Princeton Computer Science Department *Laboratory TA, Grader | Princeton, NJ* Spring 2008 - Spring 2009
Held weekly lab hours for students in all 100 and 200 level computer science classes to help with programming assignments. Taught programming techniques, debugging, and general subject knowledge. Graded assignments.

PROFESSIONAL SERVICE:

EMNLP Reviewer 2017, 2018, 2019, 2020
ACL Reviewer 2018, 2019, 2020
NAACL Reviewer 2019
COLING Reviewer 2018
CoNLL Reviewer 2018
WMT Reviewer 2018, 2019, 2020
MT Summit Reviewer 2019
NAACL SRW Reviewer 2019
ACL SRW Reviewer/Mentor 2020

AMTA 2020
Program Committee. Running Student Research Workshop with Matt Post

Midwest Speech and Language Days May 2018
Program Chair and Local Organization Committee for the MSLD meeting at the University of Notre Dame

Machine Translation Marathon in the Americas: May 2016
Local Organization Committee for the 2nd annual MTMA meeting at the University of Notre Dame.

SOFTWARE:

NPLM – Contribute to the NPLM toolkit for continuous space language models. Added functionality for group regularizers for determining hidden-layer sizes.

MISCELLANEOUS:

- Programming Languages: Java, C#, C, C++, AWK, Python, .NET, PHP, SQL, Haskell
- Princeton Varsity Men's Crew Team (2 Letters) 2005 – 2007
- Princeton Charter Eating Club Officer (Managed $30,000 Social Budget) 2007 - 2009