Aaron M. Mueller

CONTACT
INFORMATION

Center for Language & Speech Processing
Johns Hopkins University
3400 N. Charles St., Hackerman 319
Baltimore, MD 21218-2608 (U.S.A.)

E-mail: amueller@jhu.edu
Website: aaronmueller.github.io
Code: github.com/aaronmueller

RESEARCH
INTERESTS

- Multilingual natural language processing
- Computational psycholinguistics (syntax & morphology)
- Machine translation

EDUCATION

Johns Hopkins University   Baltimore, MD
   Ph.D., Computer Science.   August 2018 – May 2023 (expected)
   M.S.E., Computer Science.   May 2020
   GPA: 3.9/4.0.
   Advisors: Mark Dredze, David Yarowsky.

New York University   New York, NY
   Visiting academic, Center for Data Science.   August 2021 – May 2023
   Advisor: Tal Linzen.

University of Kentucky   Lexington, KY
   B.S., Computer Science. *Honors*.   May 2018
   B.S., Linguistics. *Honors*.   May 2018
   GPA: 4.0/4.0. *Summa cum laude*.

INDUSTRY
EXPERIENCE

Amazon Web Services (AWS)   Santa Clara, CA
Applied Scientist Intern, Lex Team   May 2021 – August 2021
Manager: Saab Mansour
- Research in pre-training methods for improving goal-oriented dialogue agents.
- Resulted in state-of-the-art few-shot intent classification accuracy (>30% 1-shot gains) and an ACL submission (under review).

Raytheon BBN Technologies   Cambridge, MA
Research Intern, Analytics & Machine Intelligence Team   May 2019 – August 2019
Manager: Ilana Heintz
- Implemented convolutional neural machine translation models rivaling our prior seq2seq model's BLEU with over 20% faster training and over 50% faster inference.
- Research in low-resource Ukrainian/Russian-English word alignment and entity linking.

PUBLICATIONS

REFEREED

1. **Aaron Mueller**, Mark Dredze. "Fine-tuning Encoders for Improved Monolingual and Zero-shot Polylingual Neural Topic Modeling." In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021.
2. **Aaron Mueller**, Zach Wood-Doughty, Silvio Amir, Mark Dredze, Alicia L. Nobles. "Demographic Representation and Collective Storytelling in the Me Too Twitter Hashtag Activism Movement." In *Proceedings of the Association for Computing Machinery (ACM) on Human-Computer Interaction (HCI), vol. CSCW1*, 2021.
3. Matthew Finlayson*, **Aaron Mueller***, Sebastian Gehrmann, Stuart Shieber, Tal Linzen, Yonatan Belinkov. "Causal Analysis of Syntactic Agreement Mechanisms in Neural Language Models." In *Proceedings of the Association for Computational Linguistics (ACL)*, 2021. [*Equal contribution]
4. Alexandra DeLucia*, **Aaron Mueller***, Xiang Lisa Li, João Sedoc. "Decoding Methods for Neural Narrative Generation." In *Proceedings of the Workshop on Generation Evaluation and Metrics (GEM) at Association for Computational Linguistics (ACL)*, 2021. [*Equal contribution]
5. **Aaron Mueller**, Garrett Nicolai, Panayiota Petrou-Zeniou, Natalia Talmina, Tal Linzen. "Cross-linguistic Syntactic Evaluation of Word Prediction Models." In *Proceedings of the Association for Computational Linguistics (ACL)*, 2020.

6. **Aaron Mueller**, Garrett Nicolai, Arya D. McCarthy, Dylan Lewis, Winston Wu, David Yarowsky. "An Analysis of Massively Multilingual Neural Machine Translation for Low-Resource Languages." In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2020.

7. Arya D. McCarthy, Rachel Wicks, Dylan Lewis, **Aaron Mueller**, Winston Wu, Oliver Adams, Garrett Nicolai, Matt Post, David Yarowsky. "The Johns Hopkins University Bible Corpus: 1600+ Tongues for Typological Exploration." In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2020.

8. Garrett Nicolai, Dylan Lewis, Arya D. McCarthy, **Aaron Mueller**, Winston Wu, David Yarowsky. "Fine-grained Morphosyntactic Analysis and Generation Tools for More Than One Thousand Languages." In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2020.

9. Marten van Schijndel, **Aaron Mueller**, Tal Linzen. "Quantity Doesn't Buy Quality Syntax with Neural Language Models." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

10. Arya D. McCarthy, Winston Wu, **Aaron Mueller**, Bill Watson, David Yarowsky. "Modeling Color Terminology Across Thousands of Languages." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

11. **Aaron Mueller***, Yash Kumar Lal*. "Sentence-Level Adaptation for Low-Resource Neural Machine Translation." In *Proceedings of the Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT) at Machine Translation Summit (MTSummit)*, 2019. [*Equal contribution]

UNDER REVIEW

12. **Aaron Mueller**, Jason Krone, Salvatore Romeo, Saab Mansour, Elman Mansimov, Yi Zhang, Dan Roth. "Label Semantic Aware Pre-training for Few-shot Text Classification." In submission to *Association for Computational Linguistics (ACL)*, 2022.

13. **Aaron Mueller**, Robert Frank, Tal Linzen, Luheng Wang, Sebastian Schuster. "Coloring the Blank Slate: Pre-training Imparts a Hierarchical Inductive Bias to Sequence-to-sequence Models." In submission to *Association for Computational Linguistics (ACL)*, 2022.

NON-REFEREED

14. **Aaron Mueller**, Mark Richard Lauersdorf, Kevin McGowan, Ramakanth Kavuluru. "Neural Machine Translation for Canadian French." *Undergraduate thesis*, 2018.

INVITED TALKS
– **Aaron Mueller**. "Syntactic Agreement in Neural Language Models: How Well and Where Do They Perform Subject-Verb Agreement?" At Mila – Québec Artificial Intelligence Institute, Language & Understanding Group (Montréal, QC). March 22, 2021. Virtual talk.

– **Aaron Mueller***, Sebastian Gehrmann*. "Causal Mediation Analysis for Analyzing Neural Networks." At Google Fairness & Interpretability Research Group (New York, NY). March 17, 2021. Virtual talk. [*Equal contribution]

– **Aaron Mueller**. "Causal Analysis of Syntactic Agreement Mechanisms in Neural Language Models." At Johns Hopkins University, Center for Language & Speech Processing Seminar (Baltimore, MD). February 12, 2021. Virtual talk.

RESEARCH PRESENTATIONS
– **Aaron Mueller**. "Québec French and Language Technology." 2018 Betts & Rowland Scholar Awards Ceremony, Lexington, KY, April 4, 2018.

– **Aaron Mueller**, Katherine Keith, Abram Handler, Su Lin Blodgett, Brendan O'Connor. "The Identification of Civilians Killed by Police with Supervised Entity-Event Extraction." UMass Amherst Research Experience for Undergraduates (REU) Showcase. Amherst, MA. August 9, 2017.

– **Aaron Mueller**, Raphael Finkel, Hilaria Cruz. "Documenting and Promoting the Chatino Language and Orthography." Juried Presentation in Satisfaction of the Requirements of the Gaines Fellowship. Lexington, KY. February 21, 2017.

– **Aaron Mueller**, Huda Khayrallah, Winston Wu, David Yarowsky. "A Lemma-Based Approach for English-Uyghur Statistical Machine Translation." Refereed Abstract at the *9th Annual Conference of the Illinois Language and Linguistics Society (ILLS9)*. Urbana, IL. March 31, 2017.

– **Aaron Mueller**, Mark Richard Lauersdorf. "Lexical and Semantic Shifts in the Linguistic Construction of Social Gender: A Corpus-Based Analysis of Written U.S. English." Poster. Refereed Abstract at the *9th Annual Toronto Undergraduate Linguistics Conference (TULCON9)*. Toronto, ON. March 5, 2016.

ACADEMIC RESEARCH EXPERIENCE	**Johns Hopkins University**	Baltimore, MD

ACADEMIC
RESEARCH
EXPERIENCE

Johns Hopkins University Baltimore, MD
Research Assistant, Center for Language & Speech Processing August 2018 – Present
Advisors: David Yarowsky, Mark Dredze
– Research in neural machine translation, natural language generation, and massively multilingual NLP.

Research Assistant, Center for Language & Speech Processing May 2016 – August 2016
Advisor: David Yarowsky
– Research in statistical machine translation for low-resource languages.
– Implemented a lemma-based English-Uyghur translation model, built a morphological generator for Crimean Tatar, and created parsers to extract translation tables from foreign editions of Wiktionary.

New York University New York, NY
Visiting Academic, Center for Data Science August 2021 – Present
Advisor: Tal Linzen
– Behavioral and causal probing into the (morpho)syntactic representations and abilities of neural language models and sequence-to-sequence models.

University of Kentucky Lexington, KY
Research Assistant, Institutional Research & Advanced Analytics Team May 2018 – August 2018
Advisors: Nathan Jacobs, Craig Rudick
– Implemented deep LSTMs to learn student and course profiles for grade prediction.
– Designed various ordinal loss functions.

Juried Project, Gaines Center August 2016 – May 2017
Jurors: Hilaria Cruz, Raphael Finkel, Phil Harling
– Preservation and revitalization project for a low-resource indigenous language—Chatino—spoken natively by approximately 40,000 individuals.
– Trained ASR systems with Sphinx.
– Created and publicized Chatino language-learning resources, including open-source speech data, a speech corpus, and a website.

University of Massachusetts Amherst Amherst, MA
Research Assistant, Statistical Social Language Analysis Lab May 2017 – August 2017
Advisor: Brendan O'Connor
– Research in entity-event extraction.
– Integrated entity mention and relation systems into pre-existing sentence-level entity-event extraction model. Also integrated document-level information (e.g., noun coreferences) as features.

FELLOWSHIPS
AND AWARDS

National Science Foundation Graduate Research Fellow 2018 - 2023
Gaines Fellow 2016 - 2018
Two-year fellowship awarded to undergraduates based on academic performance, ability to conduct research, an interest in public issues, and a desire to enhance understanding of the human condition through the humanities. Requires the completion of a juried project, a thesis project, and a seminar in the humanities. ($5,000)
Patterson Scholar 2014 - 2018
Awarded to University of Kentucky students who have earned National Merit semifinalist standing or better. ($80,000+)
Raymond F. Betts Scholar 2017
Awarded to rising seniors conducting thesis research. Used funds to study Québec French in Montréal and Québec City during winter of 2017–2018. ($2,500)
Goldwater Scholarship (Honorable Mention) 2017
Phi Beta Kappa 2017
Linguistics Research Award 2016
Awarded to a University of Kentucky undergraduate to facilitate a year-long research project in linguistics. Used funds for a quantitative sociolinguistics project examining semantic shifts in gendered words in U.S. media over 200 years. ($500)
National Merit Semifinalist 2014

TEACHING	**Johns Hopkins University** Baltimore, MD

Teaching Assistant
– Machine Learning: AI System Design & Development. Spring 2020

SKILLS   Programming:
– Languages (expert): Python
– Languages (proficient): C++, HTML, CSS, Javascript, Bash
– Machine Learning Toolkits: PyTorch (incl. HuggingFace, fairseq, sockeye), NLTK, Scikit-learn, numpy
– Version Control: DVCS (Git, Bitbucket)

Linguistic Tools:
Praat, AntConc, QGIS, Audacity

LANGUAGES   English (native language), French (B2, Canadian). Experience with German and Finnish through research.

INTERESTS AND
HOBBIES
– Music: guitar (classical, electric), bass, banjo, synthesizer, MaxMSP, Ableton Live.
– Athletics: cold-weather/snow hiking, swing dance, contra.
– Hobbies: natural language learning, reading (philosophy, politics), city exploration.
– Cooking: Thai, French, German, Southern.