

Aaron M. Mueller

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

Courriel: amueller@jhu.edu
Site web: aaronmueller.github.io
Code: github.com/aaronmueller

INTÉRÊTS DE RECHERCHE

- TALN multilingue
- Psycholinguistique informatique
- Traduction automatique

FORMATION **Johns Hopkins University**, Baltimore, MD, É-U
Doctorant en informatique.
Maîtrise en informatique.
GPA: 3.9/4.0.
Directeurs de thèse: Mark Dredze, David Yarowsky.

Mai 2023 (attendu)
Mai 2020

New York University, New York, NY, É-U
Chercheur visiteur, Center for Data Science.
Conseiller: Tal Linzen.

Août 2021 – mai 2023

University of Kentucky, Lexington, KY, É-U
Baccalauréat en informatique (*avec mention*).
Baccalauréat en linguistique (*avec mention*).
GPA: 4.0/4.0. *Summa cum laude*.
Thèse: Neural Machine Translation for Canadian French.

Mai 2018
Mai 2018

EXPÉRIENCE DE L'INDUSTRIE **Meta**
Stagiaire en recherche, AI Integrity
Manager: Kanika Narang

Menlo Park, CA, É-U
Mai - septembre 2022

- Recherche en méta-apprentissage augmenté de récupération pour les systèmes de questions-réponses sur quelques exemples.
- Réalisation des F1 améliorés sur plusieurs tâches questions-réponses et classifications en utilisant moins de paramètres que les systèmes pré-entraînés de pointe.

Amazon Web Services (AWS)
Stagiaire en recherche appliqué, Lex
Manager: Saab Mansour

Santa Clara, CA, É-U
Mai - août 2021

- Recherche en améliorer les systèmes de dialogue orientés tâche.
- Réalisation des précisions de pointe sur la classification d'intention sur quelques exemples (>30% sur 1 exemple) et une publication à l'ACL.

Raytheon BBN Technologies, Cambridge, MA, É-U
Analytics & Machine Intelligence

Stagiaire en recherche, Analytics & Machine Intelligence
Superviseurs: Roger Bock, Ilana Heintz

Mai - août 2019

- Implementation des systèmes de traduction automatique convolutionnelle qui rivalise notre système seq2seq précédent, mais avec l'apprentissage 20% plus rapide et l'inference 50% plus rapide.
- Recherche en alignement des mots et annotation sémantique multilingue à faibles ressources en russe et ukrainien.

PUBLICATIONS COMPTES RENDUS DE CONFÉRENCES

1. **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." *Findings of the Association for Computational Linguistics (ACL)*, 2022.
2. **Aaron Mueller**, Jason Krone, Salvatore Romeo, Saab Mansour, Elman Mansimov, Yi Zhang, Dan Roth. "Label Semantic Aware Pre-training for Few-shot Text Classification." *Proceedings of the Association for Computational Linguistics (ACL)*, 2022.

3. **Aaron Mueller**, Yu Xia, Tal Linzen. “Causal Analysis of Syntactic Agreement Neurons in Multilingual Language Models.” *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*, 2022.
4. **Aaron Mueller**, Mark Dredze. “Fine-tuning Encoders for Improved Monolingual and Zero-shot Polylingual Neural Topic Modeling.” *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021.
5. **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.” *Proceedings of the Association for Computing Machinery (ACM) on Human-Computer Interaction (HCI)*, vol. CSCWI, 2021.
6. Matthew Finlayson*, **Aaron Mueller***, Sebastian Gehrmann, Stuart Shieber, Tal Linzen, Yonatan Belinkov. “Causal Analysis of Syntactic Agreement Mechanisms in Neural Language Models.” *Proceedings of the Association for Computational Linguistics (ACL)*, 2021. [*Contribution égale]
7. Alexandra DeLucia*, **Aaron Mueller***, Xiang Lisa Li, João Sedoc. “Decoding Methods for Neural Narrative Generation.” *Proceedings of the Workshop on Generation Evaluation and Metrics (GEM) at Association for Computational Linguistics (ACL)*, 2021. [*Contribution égale]
8. **Aaron Mueller**, Garrett Nicolai, Panayiota Petrou-Zeniou, Natalia Talmina, Tal Linzen. “Cross-linguistic Syntactic Evaluation of Word Prediction Models.” *Proceedings of the Association for Computational Linguistics (ACL)*, 2020.
9. **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.” *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2020.
10. 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.” *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2020.
11. 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.” *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2020.
12. Marten van Schijndel, **Aaron Mueller**, Tal Linzen. “Quantity Doesn’t Buy Quality Syntax with Neural Language Models.” *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.
13. Arya D. McCarthy, Winston Wu, **Aaron Mueller**, Bill Watson, David Yarowsky. “Modeling Color Terminology Across Thousands of Languages.” *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.
14. **Aaron Mueller***, Yash Kumar Lal*. “Sentence-Level Adaptation for Low-Resource Neural Machine Translation.” *Proceedings of the Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT) at Machine Translation Summit (MTSummit)*, 2019. [*Contribution égale]

NON-REVUES ET AUTRES

15. Julian Michael, Ari Holtzman, Alicia Parrish, **Aaron Mueller**, Alex Wang, Angelica Chen, Divyam Madaan, Nikita Nangia, Richard Yuanzhe Pang, Jason Phang, Samuel R. Bowman. “What do NLP Researchers Believe? Results of the NLP Community Metasurvey.” *Prépublication arXiv*, 2022.
16. **Aaron Mueller**, Mark Richard Lauersdorf, Kevin McGowan, Ramakanth Kavuluru. “Neural Machine Translation for Canadian French.” *Thèse de premier cycle*, 2018.

PRÉSENTATIONS INVITÉES

Aaron Mueller. “What Generalizations do Sequence-to-sequence Models Learn from Multilingual Text? Insights from Translation and Syntactic Transformations.” Au Conseil national de recherches Canada (Ottawa, ON, CA). 4 mars 2022. Présentation virtuelle.

Aaron Mueller. “Syntactic Agreement in Neural Language Models: How Well and Where Do They Perform Subject-Verb Agreement?” Au Mila – Institut de recherche, groupe Language & Understanding (Montréal, QC, CA). 22 mars 2021. Présentation virtuelle.

Aaron Mueller*, Sebastian Gehrmann*. “Causal Mediation Analysis for Analyzing Neural Networks.” À Google, groupe Fairness & Interpretability (New York, NY, É-U). 17 mars 2021. Présentation virtuelle. [*Contribution égale]

Aaron Mueller. “Causal Analysis of Syntactic Agreement Mechanisms in Neural Language Mod-

els.” À la Johns Hopkins University, séminaire de Center for Language & Speech Processing (Baltimore, MD, É-U). 12 février 2021. Virtual talk.

PRÉSENTATIONS
DE RECHERCHE

Aaron Mueller. Québec French and Language Technology. *Cérémonie de Boursiers Betts & Rowland 2018*, Lexington, KY, É-U, 4 avril 2018.

Aaron Mueller, Katherine Keith, Abe Handler, Su Lin Blodgett, et Brendan O’Connor. The Identification of Civilians Killed by Police with Supervised Entity-Event Extraction. *2017 UMass Amherst Research Experience for Undergraduates (REU) Showcase*, Amherst, MA, É-U, 9 août 2017.

Aaron Mueller, Raphael Finkel, et Hilaria Cruz. Documenting and Promoting the Chatino Language and Orthography. *Présentation de jury dans la satisfaction des besoins de la bourse Gaines*, Lexington, KY, É-U, 21 février 2017.

Aaron Mueller, Huda Khayrallah, Winston Wu, et David Yarowsky. A Lemma-Based Approach for English-Uyghur Statistical Machine Translation. *9th Annual Conference of the Illinois Language and Linguistics Society (ILLS9)*, Urbana, IL, É-U, 31 mars–1 avril 2017.

Aaron Mueller et Mark Richard Lauersdorf. Lexical and Semantic Shifts in the Linguistic Construction of Social Gender: A Corpus-Based Analysis of Written U.S. English. Poster. *9th Annual Toronto Undergraduate Linguistics Conference (TULCON9)*, Toronto, ON, Canada, 4–6 mars 2016.

EXPÉRIENCE

Johns Hopkins University, Baltimore, MD, É-U

ACADÉMIQUE

Assistant de recherche, Center for Language & Speech Processing

Août 2018 - présent

EN RECHERCHE

Superviseur: Mark Dredze, David Yarowsky

- Recherche en traduction automatique neuronale, génération du langage naturel, et le TALN multilingue.

Assistant de recherche, Center for Language & Speech Processing

Mai - août 2016

Superviseur: David Yarowsky

- Recherche en traduction automatique statistique pour les langues à faibles ressources.
- Mettre en œuvre un modèle de traduction anglais-ouïghour basé sur les lemmes, un générateur morphologique pour le tatar de Crimée, et des analyseurs syntaxiques pour soutenir les tables de traduction des éditions étrangères de Wiktionnaire.

New York University, New York, NY, É-U

Chercheur visiteur, Center for Data Science

Août 2021 - présent Superviseur: Tal Linzen

- Création des probes linguistiques pour comprendre la connaissance syntaxique et morphologique des modèles pré-entraînés.

University of Kentucky, Lexington, KY, É-U

Assistant de recherche, Institutional Research & Advanced Analytics

Mai - août 2018

Superviseurs: Nathan Jacobs, Craig Rudick

- Mettre en œuvre les LSTMs profonds pour apprendre les profils étudiants et cours.
- Dessiner des fonctions objectifs ordinales.

Projet de jury, Gaines Center

Août 2016 - mai 2017

Jurés: Hilaria Cruz, Raphael Finkel, Phil Harling

- Projet de préservation et revitalisation pour une langue autochtone à faible ressources—Chatino—parlée par 40 000 personnes.
- Entraîner un système de reconnaissance automatique de la parole avec Sphinx.
- Promouvoir et créer des ressources pour l’apprentissage des langues chatino, dont les données source-ouvert de parole, un corpus de parole et un site web.

University of Massachusetts Amherst, Amherst, MA, É-U

Statistical Social Language Analysis Lab

Assistant de recherche

Mai - août 2017

Superviseur: Brendan O’Connor

- Recherche en extraction des entités et événements des textes.
- Intégrer les systèmes d’extraction des relations des entités dans un modèle préexistant au niveau des phrases complètes. Intégrer l’information au niveau du document (p. ex., les coréférences de noms).

BOURSES ET PRIX	Boursier de recherche de la National Science Foundation	2018 - 2023
	Boursier Gaines	2016 - 2018
	Bourse de recherche de deux ans décernée aux étudiants de premier cycle pour leur rendement scolaire, capacité de mener des recherches, intérêt pour les affaires publiques et désir de faire avancer les connaissances de la condition humaine à travers les sciences humaines. Elle nécessite l'achèvement d'un projet de jury, un thèse et un cours séminaire en sciences humaines. (\$5 000 USD)	
	Boursier Patterson	2014 - 2018
	Décernée aux étudiants à l'Université du Kentucky qui ont gagné le Bourse national du mérite. (\$80 000+ USD)	
	Boursier Raymond F. Betts	2017
	Décernée à un étudiant par an pour aider ses travaux de thèse. Utilisé les fonds pour étudier le français québécois à Montréal et à Québec pendant l'hiver de 2017–2018. (\$2 500 USD)	
	Bourse Goldwater (Mention honorable)	2017
	Phi Beta Kappa	2017
	Prix de recherche en linguistique	2016
EXPÉRIENCE EN ENSEIGNEMENT	Attribué à un étudiant à l'Université du Kentucky pour faciliter un projet de recherche d'une année en linguistique. Utilisé les fonds pour un projet de sociolinguistique quantitative examinant les glissements sémantiques des mots sexospécifiques dans les médias américains sur 200 ans. (\$500 USD)	
	Bourse nationale du mérite	2014
	Johns Hopkins University , Baltimore, MD, É-U	
COMPÉTENCES INFORMATIQUES	<i>Assistant</i>	
	– Machine Learning: AI System Design and Development, Mathias Unberath. Printemps 2020.	
	Programmentation:	
LANGUES	– Langages (expert): Python	
	– Langages (compétent): C++, HTML, CSS, Javascript, Bash.	
	– Toolkits d'apprentissage automatique: PyTorch (dont huggingface, fairseq, sockeye), Keras, TensorFlow.	
INTÉRÊTS ET LOISIRS	– Contrôle de version: DVCS (Git, Bitbucket).	
	Outils linguistiques:	
	Praat, AntConc, QGIS, Audacity.	
INTÉRÊTS ET LOISIRS	Français (B2, canadien), anglais (langue maternelle). Expérience en allemand et finlandais par la recherche.	
	– Musique: guitare (classique, électrique), basse, banjo, synthétiseur, MaxMSP, Ableton Live.	
	– Athlétisme: randonnée d'hiver, danse disco et swing.	
INTÉRÊTS ET LOISIRS	– Loisirs: apprentissage du langage naturel, lecture (philosophie, politique), exploration urbaine.	
	– Compétences culinaires: thaïlandais, français, allemand, sud des États-Unis.	