

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.

Depuis août 2018

Mai 2020

Directeurs de thèse: Mark Dredze, David Yarowsky.

University of Kentucky, Lexington, KY, É-U

Baccalauréat en informatique (avec mention).

Mai 2018

Baccalauréat en linguistique (avec mention).

Mai 2018

GPA: 4.0/4.0. *Summa cum laude*.

Thèse: Neural Machine Translation for Canadian French.

EXPÉRIENCE **BBN Technologies**, Cambridge, MA, É-U

EN RECHERCHE **Analytics & Machine Intelligence**

Stagiaire en recherche linguistique et multimédia

Mai - août 2019

Superviseurs: Roger Bock, Ilana Heintz

- Recherche en améliorer le compromis entre vitesse et précision en traduction automatique neuronale, alignement des mots et annotation sémantique.

Université du Kentucky, Lexington, KY, É-U

Institutional Research & Advanced Analytics

Assistant de recherche en apprentissage automatique

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.

University du Massachusetts à Amherst, Amherst, MA, É-U

Statistical Social Language Analysis Lab

Assistant de recherche en TALN

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).

University of Kentucky, Lexington, KY, É-U

Département de linguistique

Projet de jury

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.

Johns Hopkins University, Baltimore, MD, É-U
Center for Language & Speech Processing

Assistant de recherche en TALN

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.

PAPERS

COMPTES RENDUS DE CONFÉRENCES

1. **Aaron Mueller**, Garrett Nicolai, Panayiota Petrou-Zeniou, Natalia Talmina, et Tal Linzen. *Cross-Linguistic Syntactic Evaluation of Word Prediction Models*. Association for Computational Linguistics (ACL), 2020.
2. **Aaron Mueller**, Garrett Nicolai, Arya D. McCarthy, Dylan Lewis, Winston Wu, et David Yarowsky. *An Analysis of Massively Multilingual Neural Machine Translation for Low-Resource Languages*. Language Resources and Evaluation Conference (LREC), 2020.
3. Arya D. McCarthy, Rachel Wicks, Dylan Lewis, **Aaron Mueller**, Winston Wu, Oliver Adams, Garrett Nicolai, Matt Post, et David Yarowsky. *The Johns Hopkins University Bible Corpus: 1600+ Tongues for Typological Exploration*. Language Resources and Evaluation Conference (LREC), 2020.
4. Garrett Nicolai, Dylan Lewis, Arya D. McCarthy, **Aaron Mueller**, Winston Wu, et David Yarowsky. *Fine-Grained Morphosyntactic Analysis and Generation Tools for More Than One Thousand Languages*. Language Resources and Evaluation Conference (LREC), 2020.
5. Marten van Schijndel, **Aaron Mueller**, et Tal Linzen. *Quantity Doesn't Buy Quality Syntax with Neural Language Models*. Empirical Methods in Natural Language Processing (EMNLP), 2019.
6. Arya D. McCarthy, Winston Wu, **Aaron Mueller**, Bill Watson, et David Yarowsky. *Modeling Color Terminology Across Thousands of Languages*. Empirical Methods in Natural Language Processing (EMNLP), 2019.
7. **Aaron Mueller*** et Yash Kumar Lal*. *Sentence-Level Adaptation for Low-Resource Neural Machine Translation*. Machine Translation Summit (MTSummit) Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT), 2019. [***Contribution égale**]

NON-REVUES ET AUTRES

8. **Aaron Mueller**, Mark Richard Lauersdorf, Kevin McGowan et Ramakanth Kavuluru. *Neural Machine Translation for Canadian French*. Thèse de baccalauréat, 2018.

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
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

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

REPRÉSENTATIVES

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 EN **Johns Hopkins University**, Baltimore, MD, É-U

ENSEIGNEMENT *Assistant*

– Machine Learning: AI System Design and Development, Mathias Unberath. Printemps 2020.

EXPÉRIENCE **Chellgren Center for Undergraduate Excellence**, Lexington, KY, É-U

SUPPLÉMENTAIRE

Équipe de sensibilisation des étudiants

Août 2017 - mai 2018

- Organiser les tables rondes, ateliers et lectures sur l'écriture des résumés, le dessin des posters de recherches et les présentations académiques.
- Diriger le comité d'examen pour la Showcase of Undergraduate Scholars, une conférence universitaire annuelle pour les étudiants de l'University of Kentucky.

UK Special Collections Research Center, Lexington, KY, É-U

Stagiaire en archivage

Septembre 2015 - mai 2016

- Créer un corpus diachronique de documentation architecturale de Lexington, dont les ordonnances de zonage et les procès-verbaux du comité organisateur de Lexington.
- Coopérer sur un projet multimédia, dont les logiciels SIG, données de municipalité, photos d'archives, et outils d'humanités numériques.

COMPÉTENCES **Programmation:**

INFORMATIQUES

- Languages: Python, C++, C, HTML, CSS, Javascript, Java, script shell UNIX.
- Toolkits d'apprentissage automatique: PyTorch (dont huggingface, fairseq, sockeye), Keras, TensorFlow.
- Contrôle de version: DVCS (Git, Bitbucket).
- Environnements de développement: Vim, Emacs, VSCode.

Outils linguistiques:

Praat, AntConc, QGIS, Audacity.

LANGUES

Français (B2, canadien), anglais (langue maternelle). Expérience en chatino et ouïghour par la recherche.

INTÉRÊTS ET LOISIRS

- Musique: guitare (classique, électrique), basse, banjo, synthétiseur, MaxMSP, Ableton Live.
- Athlétisme: randonnée d'hiver, danse disco et swing.
- Loisirs: apprentissage du langage naturel, lecture (philosophie, politique), exploration urbaine.
- Compétences culinaires: thaïlandais, français, allemand, sud des États-Unis.