

# Tolúlopé Ògúnrẹmí | Curriculum Vitae

✉ tolulope@cs.stanford.edu • 📄 tolulope.carrd.co

## Education

---

- **PhD Computer Science**

*Stanford University*

*2021–Present*

Low-resource natural language processing and speech processing. Advised by Professor Dan Jurafsky and Professor Chris Manning.

- **MSc Speech and Language Processing**

*University of Edinburgh*

*2020–2021*

**Modules:** Accelerated Natural Language Processing, Natural Language Understanding, Generation, and Machine Translation, Speech Processing, Automatic Speech Recognition, Speech Synthesis, Machine Learning and Pattern Recognition, Reinforcement Learning (audited).

- **Double licence en mathématiques et informatique**

*Université Paris Diderot [Erasmus Semester during undergraduate degree]*

*2019*

**Modules:** Programmation réseaux, Introduction au TAL (top of the class), Projet Master (Théorie de l'information), Projet math-info, Expression et Communication.

- **BSc Mathematics with Computer Science**

*University of Southampton*

*2016–2019*

**Modules:** Programming III, Intelligent Systems, Statistical Distribution Theory, Analysis, Partial Differential Equations, Stochastic Processes, Statistical Methods I.

**Computer Science projects:** Lambda Calculus Compiler in Haskell, Artificially intelligent agent navigating a grid.

## Publications

---

- **Multilingual self-supervised speech representations improve the speech recognition of low-resource African languages with codeswitching** Tolúlopé Ògúnrẹmí, Christopher D. Manning, Dan Jurafsky – Computational Approaches to Linguistic Code-Switching, CALCS 2023 co-located with EMLP 2023
- **ÌròyìnSpeech: A multi-purpose Yorùbá Speech Corpus** Tolúlopé Ògúnrẹmí, Kòlá Túbòsún, Anuoluwapo Aremu, Iroko Orife, David Ifeoluwa Adelani – under submission but on arXiv
- **Mini But Mighty: Efficient Multilingual Pretraining with Linguistically-Informed Data Selection** Tolúlopé Ògúnrẹmí, Dan Jurafsky, Christopher D. Manning – EACL Findings 2023
- **Leveraging Bias in Pre-trained Word Embeddings for Unsupervised Microaggression Detection [Best Paper Award]** Tolúlopé Ògúnrẹmí, Nazanin Sabri, Valerio Basile and Tommaso Caselli – CLiC-it 2021

- **Automated speech tools for helping communities process restricted-access corpora for language revival efforts** *Nay San, Martijn Bartelds, Tolúlopé Ògúnremí, Alison Mount, Ruben Thompson, Michael Higgins, Roy Barker, Jane Simpson and Dan Jurafsky*, – CompuTEL-5 (2022)
- **An evaluation and analysis fine-tuned representations for code-switched speech recognition in a zero-resource scenario** *Tolúlopé Ògúnremí* – In Preparation

## Awards, Grants and Fellowships

---

### Grants.....

- **Imminent Research Grant (Linguistic Data)**  
*Creation of Yorùbá speech data which will be open-sourced on Common Voice*

### Awards and Fellowships.....

- **University of Southampton Outstanding Alumni Award: Rising Star 2023**
- **Stanford EDGE Fellowship 2021-2026**
- **Stanford School of Engineering Fellowship 2021-2024**
- **Dean's List Award BSc Mathematics with Computer Science 2016-2019**  
 I was on the Dean's List for each individual year of my undergraduate study and received a Dean's List Award for the entire degree.

## Invited Talks

---

- **Grandes architectures, peu de ressources: l'application du traitement automatique de langage naturel et de la parole sur les langues Africaines peu dotées:** Laboratoire LLACAN (CNRS), Paris (virtually), November 2022
- **Computer Aided Pronunciation Training of Yorùbá using Bayesian Item Response Theory:** NLP@Deep Learning Indaba, August 2022
- **An evaluation and analysis fine-tuned representations for code-switched speech recognition in a zero-resource scenario:** Natural Language Processing Group at Università di Torino, July 2022
- **State-of-the-Art Automatic Speech Recognition:** Adesfa Natural Language Processing Boot Camp, Ibadan (virtually), November 2021

## Technical and Personal skills

---

- **Programming Languages:** Python, Java, JavaScript (Vanilla, React), R, C, Haskell
- **Spoken Languages:** English (fluent), Yoruba (mother-tongue), French (extremely proficient) and Italian. Also currently learning Spanish, Arabic, Kinyarwanda and Kiswahili. I can speak, read and write all languages listed.

## Technical and Personal skills

---

- **Programming Languages:** Python, Java, JavaScript (Vanilla, React), R, C, Haskell
- **Spoken Languages:** English (fluent), Yoruba (mother-tongue), French (extremely proficient) and Italian. Also currently learning Spanish, Arabic, Kinyarwanda and Kiswahili. I can speak, read and write all languages listed.

## Service

---

- **Reviewing:** EMLP 2023, ACL 2023, AfricaNLP 2023.

## Interests and extra-curricular activity

---

- I am passionate about democratising access to education. I have set up Coders of Colour to educate young, underrepresented people of colour learn to code via free workshops and programmes. To tackle adult illiteracy in Africa, I have set up Project Hezi, providing educational resources to young Africans, easing the burden caused by attending school.
- I am a member of a number of research groups at Stanford and beyond. I am part of Masakhane, the a grassroots organisation whose mission is to strengthen and spur NLP research in African languages, for Africans, by Africans. I am part of the NLP for Low-resource languages special interest group within Stanford NLP. I am a member of the Niger-Volta Language Technology Institute, YorùbáNames research team and African Languages Technology Institute (Alt-i).