



Lydia Nishimwe

PHD STUDENT IN MACHINE TRANSLATION

My thesis focuses on the **robustness of neural machine translation** models against the **non-standard text** found in **user-generated content** on social media.

Research Fields

- Machine Translation
- Lexical Normalisation
- Language Models
- Sentence Embeddings

- ✉ lydia.nishimwe@inria.fr
- 🌐 lydianishimwe.info
- 🌐 [linkedin.com/in/lydia-nishimwe](https://www.linkedin.com/in/lydia-nishimwe)
- 🐦 twitter.com/LydiaNishimwe
- 🐙 github.com/lydianish

Education

■ Ph.D. in Computer Science

Centre Inria de Paris / Sorbonne Université

October 2021 – Present

Paris, France

- Thesis topic: **Robust Neural Machine Translation**
- PhD supervisors: **Benoît Sagot** and **Rachel Bawden**

■ M.Eng. in Mathematics and Computer Science

École Centrale de Nantes

September 2017 – May 2021

Nantes, France

Specialisations:

- 3rd Year: **Applied Mathematics**
 - Track: **Statistics and Data Science**
 - Professional Track: **R&D**
- 2nd Year: **Computer Science**
 - Exchange semestre at the Polytechnic University of Madrid : **Organisational Engineering**

■ B.Sc. in Mathematics and Computer Science

Université Grenoble Alpes

September 2014 – July 2017

Grenoble, France

Work Experience

■ Internship – Research and Development

Orange Labs

June 2020 – December 2020

Lannion, France

- Title: **Inference of masked sequences**
- Environment: LaTeX, Python, TensorFlow, Keras
- Keywords: **(non)-autoregressive and (non)-monotonic seq2seq models, unmasking logs**

■ Internship – Software Development

Mean-In-Full

May 2017 – July 2017

Meylan, France

- Title: **Integration of the learning management system Opencast with the RoCamRoll software**
- Environment: Linux, LaTeX, Erlang, HTTP

■ Internship – Assembly Programming

Laboratoire TIMA

Grenoble, France

May 2016 – June 2016

- Title: **Functional verification of an ARM7 microprocessor**
- Environment: Linux, LaTeX, VHDL, C, ARM

■ Internship – Functional Programming

Laboratoire VERIMAG

May 2015

Grenoble, France

- Title: **Simulation of GPS tracks**
- Environment: Linux, LaTeX, Lutin, Lustre

Languages

English
Bilingual

French
Bilingual

Spanish
Advanced

German
Intermediate

Swahili
Intermediate

Kinyarwanda
Elementary

Programming Languages

Python

(PyTorch, TensorFlow, Scikit-Learn)

Matlab

R

C

Java

JavaScript

SQL

Soft Skills

Leadership

Team spirit

Autonomy

Resilience

Rigour

Writing

Public speaking

Interests

Choir

Guitar

K-Dramas

Traditional dances

Organisations

■ Groupe Sainte Marie

English–French interpreter

March 2016 – March 2019

Paris, France

■ Chrétiens en Grandes Écoles (CGE)

Head of the CGE network for the city of Nantes

April 2018 – January 2019

Nantes, France

■ Entraide Scolaire Amicale (ESA)

Tutoring high school students from underprivileged neighbourhoods

February 2018 – January 2019

Nantes, France

■ Chorale Jeunes de Tous Pays (JTP)

Director of an international students' choir

September 2016 – July 2017

Grenoble, France

References

Benoît Sagot

Leader of the ALMAnaCH research team, Inria Paris

benoit.sagot@inria.fr

Rachel Bawden

Researcher, ALMAnaCH team, Inria Paris

rachel.bawden@inria.fr

Publications

Lydia Nishimwe, Benoît Sagot, Rachel Bawden (2024). *Making Sentence Embeddings Robust to User-Generated Content*. Accepted at LREC-COLING. [↗](#)

Lydia Nishimwe (2023). *Normalisation lexicale de contenus générés par les utilisateurs sur les réseaux sociaux*. TALN-RÉCITAL. [↗](#) 🏆 Best Paper Award 🏆

Omer Goldman, Francesco Tinner, Hila Gonen, Benjamin Muller, Victoria Basmov, Shadrack Kirimi, **Lydia Nishimwe**, Benoît Sagot, Djamé Seddah, Reut Tsarfaty, Duygu Ataman (2022). *The MRL 2022 Shared Task on Multilingual Clause-level Morphology*. MRL. [↗](#)

Jesujoba O Alabi, **Lydia Nishimwe**, Benjamin Muller, Camille Rey, Benoît Sagot, Rachel Bawden (2022). *Inria-ALMAnaCH at the WMT 2022 shared task: Does Transcription Help Cross-Script Machine Translation?* WMT. [↗](#)