



# Lucía Ormaechea

PH.D. CANDIDATE · NLP RESEARCHER · COMPUTATIONAL LINGUIST

☎ (+41) 78 220 49 79 | ✉ lormaechea@hotmail.com | 🏠 luciaormaechea.com | 📱 lormaechea

| 💻 lucía-ormaechea | 🎓 Lucía Ormaechea

## Summary

Currently finishing a **joint Ph.D. in Natural Language Processing** at University of Geneva and University of Grenoble-Alpes. As highlighted in the list of Selected Publications, my research work to date has focused on: *i) Speech-to-Text Technologies* (with particular emphasis on **Speech Simplification** and **Speech Recognition**); *ii) Machine Translation*; and *iii) Large Language Models*.

**Actively seeking new job opportunities** where I can contribute with my expertise in the fields of **Natural Language Processing** and **Artificial Intelligence**. Open to learning new tools and technologies if needed.

## Experience

### Research Scientist & Teaching Assistant

Geneva, Switzerland

UNIVERSITÉ DE GENÈVE

Dec. 2020 - Present

- Participated in **PROPICTO** [↗](#):
  - Developed and fine-tuned **cascade and end-to-end models** for **simplification of spontaneous French**.
  - Authored numerous publications** in **high-impact conferences and journals**.
- Contributed to **BabelDr** [↗](#): deployed a **containerized Docker application** to perform ASR inside the **BabelDr medical translation device**.
- Collaborated in **Localization** [↗](#) (M.A. course): preparation of course materials, assignment grading, assistance to students in practical sessions.

### NLP Research Engineer

Grenoble, France

LABORATOIRE D'INFORMATIQUE DE GRENOBLE (LIG)

Feb. 2020 - Sep. 2020

- Created a complete pipeline for **injecting grammar-based language models** into the Kaldi speech recognition toolkit.
- Trained **HMM-DNN-based acoustic models** using open-source corpora; conducted **prototype testing and evaluation**.
- Developed **Kaldi-based ASR systems for specialized medical translation devices**.

### Translation & Localization Intern

Remote Working

NEW YORK HABITAT

Jul. 2018 - Sep. 2018

- Translated **ENG > ESP commercial texts**: video transcriptions – travel articles – apartment reviews – client testimonials.
- Selected publications**:
  - Bajo la sombra del cine: una guía del cine en Nueva York (parte 1: Manhattan)* [↗](#), (2018).
  - El cine de los distritos: un recorrido por el cine de Nueva York (parte 2)* [↗](#), (2018).

### Undergraduate Research Assistant

Pamplona, Spain

UNIVERSIDAD DE NAVARRA

Sep. 2016 - Jun. 2017

- Collaborated as a research assistant at the **Department of Philology** [↗](#)
- Development of **educational materials** aiming to automatically evaluate knowledge of Spanish language in L2 learners.
- One-on-one tutorial assessment** on academic writing; **proofreading** of academic papers.

## Education

### Ph.D. (Joint Degree) in Computer Science & Multilingual Information Processing

Geneva, Switzerland

UNIVERSITÉ DE GENÈVE – UNIVERSITÉ GRENOBLE-ALPES

Jun. 2022 - present

- Currently finishing a **joint Ph.D.** between the **Département of Translation Technology** [↗](#) at UNIGE, and the **GETALP team** [↗](#), at UGA.
- Ph.D. Project**: *Towards Simpler Transcripts: Investigating Automatic Simplification of French Spontaneous Speech* [↗](#)

### M.Sc. in Natural Language Processing

Paris, France

INSTITUT NATIONAL DES LANGUES ET CIVILISATIONS ORIENTALES (INALCO) – UNIVERSITÉ SORBONNE-NOUVELLE (PARIS 3)

Sep. 2018 - Sep. 2020

- Graduated with High Honors**.
- Coursework**: Scripting (Bash, Python, Perl) – Object-Oriented Programming (C++, Java) – Statistical methods for corpus exploitation – Text Mining – Convolutional Neural Networks (Keras) – Corpus Linguistics – Mark-up Languages (XML, XSLT) – Databases (SQL, Neo4J).
- Master's Thesis**: *Mise en place d'un système robuste de reconnaissance automatique de la parole appliqué au domaine médical* [↗](#) **GPA: 19/20**

### B.A. in Hispanic Philology

Pamplona, Spain

UNIVERSIDAD DE NAVARRA

Sep. 2014 - Jun. 2018

- Extraordinary End-of-Degree Award Nominee**.
- Coursework**: Lexicology – Semantics – Phonetics – Phonology – Morphology – Syntax – Discourse Analysis.

## Teaching & Assistantships

2021-2024	<b>M.A. in Translation, M.A. in Multilingual Communication Technology</b> – <i>Teaching Assistant</i> , Localization, with L. Morado (2021, 2022, 2024), and S. Girletti (2023), Université de Genève	
2016-2017	<b>Workshops on Spanish Language</b> – <i>Instructor</i> , Grammar courses for non-native undergraduate students, Universidad de Navarra	

## Languages

<b>Spanish</b>	<i>Native speaker</i>	
<b>English</b>	<i>Proficient</i>	<i>Advanced Certificate in English (CAE)</i>
<b>French</b>	<i>Proficient</i>	<i>Diplôme Approfondi de Langue Française (DALF)</i>
<b>Italian</b>	<i>Conversational</i>	

## Software Skills

<b>Programming</b>	Python, Bash ( <i>proficient</i> ) – Perl, SQL ( <i>comfortable</i> ) – C++, Java ( <i>some experience</i> )
<b>AI Models &amp; Techniques</b>	Supervised learning – LLMs – Pre-training – Fine-tuning (full and parameter-efficient) – Prompt engineering
<b>Libraries</b>	Transformers – Speechbrain – Langchain – Flask – Keras – Fairseq – Pandas – NLTK – SpaCy – Scrapy
<b>Tools</b>	Kaldi – Slurm – Git – Vim – LaTeX – Docker – Praat – Pip – Nginx – Sublime Text – Hugging Face Spaces
<b>Operating Systems</b>	GNU/Linux (Ubuntu, Debian, Linux Mint) – Windows (10, 11) – Android
<b>Web Development</b>	HTML – CSS – Jekyll – XML – XSLT – JavaScript – Gradio – MkDocs – GitHub Pages
<b>Software</b>	Microsoft Office – Gate – Trados – Passolo – Adobe Photoshop – GIMP

## Selected Publications

<b>Interspeech 2025</b>	<b>Lucía Ormaechea</b> , Nikos Tsourakis, Pierrette Bouillon, Benjamin Lecouteux and Didier Schwab. <i>Towards High-Quality LLM-Based Data for French Spontaneous Speech Simplification: an Exo-Refinement Approach</i> <a href="#">↗</a> In: Proc. Interspeech 2025 <a href="#">↗</a> , Rotterdam (Netherlands).
<b>IJST 2024</b>	<b>Lucía Ormaechea</b> and Nikos Tsourakis. <i>Automatic Text Simplification for French: Model Fine-Tuning for Simplicity Assessment and Simpler Text Generation</i> <a href="#">↗</a> In: International Journal of Speech Technology <a href="#">↗</a> (IJST), Springer, vol. 27.
<b>WMT 2024</b>	Jonathan Mutal and <b>Lucía Ormaechea</b> . <i>TIM-UNIGE Translation into Low-Resource Languages of Spain for WMT24</i> <a href="#">↗</a> In: Proceedings of the Ninth Conference on Machine Translation <a href="#">↗</a> (WMT), Miami (USA).
<b>LREC 2024</b>	<b>Lucía Ormaechea</b> , Nikos Tsourakis, Didier Schwab, Pierrette Bouillon and Benjamin Lecouteux. <i>Simplification Strategies in French Spontaneous Speech</i> <a href="#">↗</a> In: Proceedings of the Workshop on DeTermIt! Evaluating Text Difficulty in a Multilingual Context <a href="#">↗</a> , within the Joint International Conference LREC-COLING <a href="#">↗</a> , Torino (Italy).
<b>ICNSLP 2023</b>	<b>Lucía Ormaechea</b> , Nikos Tsourakis, Didier Schwab, Pierrette Bouillon and Benjamin Lecouteux. <i>Simple, Simpler and Beyond: A Fine-Tuning BERT-Based Approach to Enhance Sentence Complexity Assessment for Text Simplification</i> <a href="#">↗</a> In: Proceedings of the 6 <sup>th</sup> International Conference on Natural Language and Speech Processing (ICNSLP), Trento (Italy).
<b>SwissText 2023</b>	<b>Lucía Ormaechea</b> and Nikos Tsourakis. <i>Extracting Sentence Simplification Pairs from French Comparable Corpora Using a Two-Step Filtering Method</i> <a href="#">↗</a> In: Proceedings of the 8 <sup>th</sup> Swiss Text Analytics Conference <a href="#">↗</a> , Neuchâtel (Switzerland).
<b>EAMT 2023</b>	<b>Lucía Ormaechea</b> , Pierrette Bouillon, Maximin Coavoux, Emmanuelle Esperança-Rodier, Johanna Gerlach et al. <i>PROPICTO: Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility</i> <a href="#">↗</a> In: Proceedings of the 24 <sup>th</sup> Annual Conference of The European Association for Machine Translation <a href="#">↗</a> (EAMT), Tampere (Finland).

## Academic Service

2025	<b>Tenth Conference on Machine Translation (WMT)</b> <a href="#">↗</a> – Held at Nanjing University – <b>Member of Program Committee</b>
2025	<b>Machine Translation Summit (MT Summit)</b> <a href="#">↗</a> – Held at University of Geneva – <b>Member of Local Organizing Team</b>
2024	<b>Corpus Journal No. 26</b> <a href="#">↗</a> – Managing noise during computer processing of linguistic corpora – <b>Peer Reviewer</b>
2023	<b>Bruit de fond ou valeur ajoutée ? Gérer le bruit lors des traitements informatiques des corpus linguistiques</b> <a href="#">↗</a> – Seminar jointly organized by Université Grenoble-Alpes and Sapienza Università di Roma – <b>Member of Scientific Committee</b>
2018	<b>5th Forum of the Association for Discourse Linguistics</b> <a href="#">↗</a> – Held at Instituto de Cultura y Sociedad, Universidad de Navarra – <b>Local Organizer</b>

## Scholarships & Awards

- 2020 **Candoc Research Fellowship** – Granted by the Swiss National Science Foundation (SNSF)
- 2013 **Runner-up at the 4th Inter-University Debate Tournament** [↗](#) – Universidad Pública de Navarra

## Certifications

### Getting Started With Mistral [↗](#)

DEEPLARNING.AI

Short course on Mistral's API features to enhance structured LLM responses (JSON mode, function calling and RAG).

*Online Course*

*September 2024*

### 3<sup>rd</sup> Advanced Language Processing School (ALPS) [↗](#)

UNIVERSITÉ GRENOBLE-ALPES – NAVER LABS EUROPE

Winter school featuring advanced lectures, social and poster sessions covering a wide range of hot topics in NLP.

*Winter School*

*January 2023*

### Research Methods in Translation & Interpreting Studies (ReMeTIS) [↗](#)

UNIVERSITÉ DE GENÈVE

- Foundations & Data Analysis: qualitative and mixed methods, research ethics.
- Specific Research & Scientific Communication Skills: corpus construction, quantitative data analysis, publication strategies.

*Online Course*

*December 2022*

### Sequence Models [↗](#)

COURSERA, DEEPLARNING.AI

Deep Learning Specialization (V): Recurrent Neural Networks – Long Short Term Memory – Attention Models.

*Online Course*

*November 2022*

### Structuring Machine Learning Projects [↗](#)

COURSERA, DEEPLARNING.AI

Deep Learning Specialization (III): Error Analysis in Machine Learning Systems – Performance Improvement Strategies.

*Online Course*

*October 2022*

### Improving Deep Neural Networks: Hyperparameter Tuning, Regularization & Optimization [↗](#)

COURSERA, DEEPLARNING.AI

Deep Learning Specialization (II): Optimization Algorithms – Hyperparameter Tuning.

*Online Course*

*October 2022*

### Neural Networks & Deep Learning [↗](#)

COURSERA, DEEPLARNING.AI

Deep Learning Specialization (I): Deep Learning Foundations – Neural Network Architectures.

*Online Course*

*October 2022*

### 33<sup>rd</sup> European Summer School in Logic, Language & Information (ESSLI) [↗](#)

NATIONAL UNIVERSITY OF IRELAND GALWAY

Interdisciplinary summer school involving courses and workshops in logic, linguistics and computer science.

*Summer School*

*August 2022*

### 2<sup>nd</sup> Advanced Language Processing School (ALPS) [↗](#)

UNIVERSITÉ GRENOBLE-ALPES – NAVER LABS EUROPE

Winter school featuring advanced lectures, social and poster sessions covering a wide range of hot topics in NLP.

*Winter School*

*January 2022*

### Personal Data Protection [↗](#)

INSTITUT NATIONAL DE RECHERCHE EN SCIENCES ET TECHNOLOGIES DU NUMÉRIQUE (INRIA)

MOOC Course introducing data confidentiality regulations and privacy protection.

*MOOC Course*

*May 2020*

### Introduction to Corpus Linguistics [↗](#)

UNIVERSITÉ GRENOBLE-ALPES (UGA)

MOOC Course introducing corpus linguistics and its many implementations in lexicology, discourse analysis and NLP.

*MOOC Course*

*June 2019*