



# Lucía Ormaechea

RESEARCHER IN NLP · SPEECH-TO-TEXT ENGINEER · PH.D.

|(+41) 78 220 49 79 | [lormaechea@hotmail.com](mailto:lormaechea@hotmail.com) | [luciaormaechea.com](http://luciaormaechea.com) | [lormaechea](https://lormaechea.com)  
| [lucía-ormaechea](https://luciaormaechea.com) | [Lucía Ormaechea](https://lormaechea.com)

## Summary

Currently working as an **ML Software Engineer** for WIPO (United Nations) at the ATAC center, for the **development of Speech-to-Text solutions**. Recently finished a joint Ph.D. in Natural Language Processing at University of Geneva and University of Grenoble-Alpes. Research interests: **Speech Recognition – Automatic Simplification – Large Language Models – Low-Resource NLP**.

## Experience

### Machine Learning Software Engineer

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO), UNITED NATIONS

Geneva, Switzerland

Oct. 2025 – Present

- Collaborating with the **Advanced Technology Applications Center (ATAC)**:
  - Participation in the development of the **WIPO Speech-to-Text platform**: **evaluation** of models, **optimization** of the existing pipeline.
  - Maintenance of the **WIPO Speech-to-Text deployment infrastructure**.

### Research Scientist & Teaching Assistant

UNIVERSITÉ DE GENÈVE

Geneva, Switzerland

Dec. 2020 – Dec. 2025

- Participated in **PROPICTO**:
  - Developed and fine-tuned **cascade** and **end-to-end models** for the **automatic 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

LABORATOIRE D'INFORMATIQUE DE GRENOBLE (LIG)

Grenoble, France

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

NEW YORK HABITAT

Remote Working

Jul. 2018 – Sep. 2018

- Translated **ENG > ESP commercial texts**: video transcriptions – travel articles – apartment reviews – client testimonials.

### Undergraduate Research Assistant

DEPARTMENT OF PHILOLOGY, UNIVERSIDAD DE NAVARRA

Pamplona, Spain

Sep. 2016 – Jun. 2017

- 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

UNIVERSITÉ DE GENÈVE – UNIVERSITÉ GRENOBLE-ALPES

Geneva, Switzerland

Jun. 2022 – Dec. 2025

- Completed a **joint Ph.D.** between the **Departement of Translation Technology** at UNIGE, and the **GETALP team** at UGA.
- Ph.D. Thesis:** *Towards Simpler Transcripts: Investigating Automatic Simplification of French Spontaneous Speech*.
- Grade:** *Très honorable avec les félicitations du jury (highest distinction)*.

### M.Sc. in Natural Language Processing

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

Paris, France

Sep. 2018 – Sep. 2020

- Coursework:** Scripting (Bash, Python, Perl) – Object-oriented programming (C++, Java) – Statistical methods for corpus exploitation – Text mining – 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 appliquée au domaine médical* **GPA: 19/20**
- Grade:** *mention très bien (highest distinction)*.

### B.A. in Hispanic Philology

UNIVERSIDAD DE NAVARRA

Pamplona, Spain

Sep. 2014 – Jun. 2018

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

# Teaching & Assistantships

---

2021-2025	<b>M.A. in Translation, M.A. in Multilingual Communication Technology</b> – <i>Teaching Assistant, Localization, with L. Morado (2021, 2022, 2024, 2025), and S. Girletti (2023)</i> , Université de Genève
2016-2017	<b>Workshops on Spanish Language</b> – <i>Instructor, Grammar courses for non-native undergraduate students, Universidad de Navarra</i>

# Languages

---

<b>Spanish</b>	Native speaker
<b>English</b>	Proficient
<b>French</b>	Proficient
<b>Italian</b>	Conversational

# 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> ↗ In: Proc. Interspeech 2025 ↗, 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> ↗ In: International Journal of Speech Technology ↗ (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> ↗ In: Proceedings of the Ninth Conference on Machine Translation ↗ (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> ↗ In: Proceedings of the Workshop on DeTermit! Evaluating Text Difficulty in a Multilingual Context ↗, within the Joint International Conference LREC-COLING ↗, 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> ↗ 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> ↗ In: Proceedings of the 8 <sup>th</sup> Swiss Text Analytics Conference ↗, 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> ↗ In: Proceedings of the 24 <sup>th</sup> Annual Conference of The European Association for Machine Translation ↗ (EAMT), Tampere (Finland).

# Academic Service

---

2025	<b>Tenth Conference on Machine Translation (WMT)</b> ↗ – Held in Suzhou (China) – <b>Member of Program Committee</b>
2025	<b>Machine Translation Summit</b> ↗ – Held at University of Geneva (Switzerland) – <b>Member of Local Organizing Team</b>
2024	<b>Corpus Journal No. 26</b> ↗ – 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> ↗ – 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> ↗ – Held at Instituto de Cultura y Sociedad, Universidad de Navarra (Spain) – <b>Member of Local Organizing Team</b>