Lucía **Ormaechea**

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

(+41) 78 220 49 79 | l.ormaechea@hotmail.com | luciaormaechea.com | lormaechea

| lucía-ormaechea | Lucía Ormaechea

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
Geneva, Switzerland

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO), UNITED NATIONS
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
Geneva, Switzerland

UNIVERSITÉ DE GENÈVE
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
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.

Undergraduate Research Assistant
Pamplona, Spain

DEPARTMENT OF PHILOLOGY, UNIVERSIDAD DE NAVARRA
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
Geneva, Switzerland

UNIVERSITÉ DE GENÈVE – UNIVERSITÉ GRENOBLE-ALPES
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
Paris, France

INSTITUT NATIONAL DES LANGUES ET CIVILISATIONS ORIENTALES (INALCO) – UNIVERSITÉ SORBONNE-NOUVELLE (PARIS 3)
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 appliqué au domaine médical* **GPA: 19/20**
- **Grade**: *Mention très bien* (**highest distinction**).

B.A. in Hispanic Philology
Pamplona, Spain

UNIVERSIDAD DE NAVARRA
Sep. 2014 – Jun. 2018

- **Coursework**: Lexicology – Semantics – Phonetics – Phonology – Morphology – Syntax – Discourse analysis – Sociolinguistics.
- **Grade**: Extraordinary end-of-degree award nominee.

Teaching & Assistantships

2021-2025 **M.A. in Translation, M.A. in Multilingual Communication Technology** – *Teaching Assistant*, Localization, with L. Morado (2021, 2022, 2024, 2025), and S. Girletti (2023), Université de Genève

2016-2017 **Workshops on Spanish Language** – *Instructor*, Grammar courses for non-native undergraduate students, Universidad de Navarra

Languages

Spanish *Native speaker*

English *Proficient* *Advanced Certificate in English (CAE)*

French *Proficient* *Diplôme Approfondi de Langue Française (DALF)*

Italian *Conversational*

Software Skills

Programming Python, Bash (*proficient*) – Perl, SQL (*comfortable*) – C++, Java (*some experience*)

AI Models & Techniques Supervised learning – LLMs – Pre-training – Fine-tuning (full and parameter-efficient) – Prompt engineering

Libraries Transformers – Speechbrain – Langchain – Flask – Keras – Fairseq – Pandas – NLTK – SpaCy – Scrapy

Tools Kaldi – Slurm – Git – Vim – LaTeX – Docker – Praat – Pip – Nginx – Sublime Text – Hugging Face Spaces

Operating Systems GNU/Linux (Ubuntu, Debian, Linux Mint) – Windows (10, 11) – Android

Web Development HTML – CSS – Jekyll – XML – XSLT – JavaScript – Gradio – MkDocs – GitHub Pages

Software Microsoft Office – Gate – Trados – Passolo – Adobe Photoshop – GIMP

Selected Publications

Interspeech 2025 **Lucía Ormaechea**, Nikos Tsourakis, Pierrette Bouillon, Benjamin Lecouteux and Didier Schwab. *Towards High-Quality LLM-Based Data for French Spontaneous Speech Simplification: an Exo-Refinement Approach* ⧉ In: Proc. Interspeech 2025 ⧉, Rotterdam (Netherlands).

IJST 2024 **Lucía Ormaechea** and Nikos Tsourakis. *Automatic Text Simplification for French: Model Fine-Tuning for Simplicity Assessment and Simpler Text Generation* ⧉ In: International Journal of Speech Technology ⧉ (IJST), Springer, vol. 27.

WMT 2024 Jonathan Mutal and **Lucía Ormaechea**. *TIM-UNIGE Translation into Low-Resource Languages of Spain for WMT24* ⧉ In: Proceedings of the Ninth Conference on Machine Translation ⧉ (WMT), Miami (USA).

LREC 2024 **Lucía Ormaechea**, Nikos Tsourakis, Didier Schwab, Pierrette Bouillon and Benjamin Lecouteux. *Simplification Strategies in French Spontaneous Speech* ⧉ In: Proceedings of the Workshop on DeTermIt! Evaluating Text Difficulty in a Multilingual Context ⧉, within the Joint International Conference LREC-COLING ⧉, Torino (Italy).

ICNSLP 2023 **Lucía Ormaechea**, Nikos Tsourakis, Didier Schwab, Pierrette Bouillon and Benjamin Lecouteux. *Simple, Simpler and Beyond: A Fine-Tuning BERT-Based Approach to Enhance Sentence Complexity Assessment for Text Simplification* ⧉ In: Proceedings of the 6th International Conference on Natural Language and Speech Processing ⧉ (ICNSLP), Trento (Italy).

SwissText 2023 **Lucía Ormaechea** and Nikos Tsourakis. *Extracting Sentence Simplification Pairs from French Comparable Corpora Using a Two-Step Filtering Method* ⧉ In: Proceedings of the 8th Swiss Text Analytics Conference ⧉, Neuchâtel (Switzerland).

EAMT 2023 **Lucía Ormaechea**, Pierrette Bouillon, Maximin Coavoux, Emmanuelle Esperança-Rodier, Johanna Gerlach *et al.* *PROPICTO: Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility* ⧉ In: Proceedings of the 24th Annual Conference of The European Association for Machine Translation ⧉ (EAMT), Tampere (Finland).

Academic Service

2025 **Tenth Conference on Machine Translation (WMT)** ⧉ – Held in Suzhou (China) – **Member of Program Committee**

2025 **Machine Translation Summit** ⧉ – Held at University of Geneva (Switzerland) – **Member of Local Organizing Team**

2024 *Corpus* **Journal No. 26** ⧉ – Managing noise during computer processing of linguistic corpora – **Peer Reviewer**

2023 **Bruit de fond ou valeur ajoutée ?** *Gérer le bruit lors des traitements informatiques des corpus linguistiques* ⧉ – Seminar jointly organized by Université Grenoble-Alpes and Sapienza Università di Roma – **Member of Scientific Committee**

2018 **5th Forum of the Association for Discourse Linguistics** ⧉ – Held at Instituto de Cultura y Sociedad, Universidad de Navarra (Spain) – **Member of Local Organizing Team**