# Àlex R. Atrio

Chemin de Sous-Bois, 10
CH-1400 Yverdon-les-Bains
Vaud, Switzerland
✆ +41 (0)76 22 99 508
✉ alexratrio@gmail.com
🖰 https://alexratrio.github.io/

## Research Interests

Natural Language Processing, Machine Translation, Langauge Modeling, Language Generation, Machine Learning.

## Education

**2019 – 2023**  **Ph.D. in Machine Translation and NLP.**
**École Polytechnique Fédérale de Lausanne (EPFL) & Haute École d'Ingénierie et de Gestion du Canton de Vaud (HEIG-VD).**
*With a focus on low resource, machine translation, language generation, and language models.*
- *Thesis*: Regularization Techniques for Low-Resource Machine Translation.
- *Supervisors*: Andrei Popescu-Belis & Jean-Marc Odobez.

**2017 – 2018**  **Master in Cognitive Science and Language.**
**University Pompeu Fabra (UPF) & University of Barcelona (UB).** *GPA: 8.2/10*
*With a focus on cognitive science, sentiment analysis, and natural language processing.*
- *Thesis*: Machine Translation Inspired Word Reordering as Preprocessing for Cross-Lingual Sentiment Analysis.
- *Supervisors*: Toni Badia & Jeremy Barnes.

**2011 – 2016**  **Bachelor in Philosophy.**
**University of Barcelona (UB).** *GPA: 7.6/10*
*With focus on formal logic, philosophy of language and philosophy of science.*

## Publications

### Papers in Conference Proceedings

**2023**  **Àlex R. Atrio**, Alexis Allemann, Ljiljana Dolamic, and Andrei Popescu-Belis. A Simplified Training Pipeline for Low-Resource and Unsupervised Machine Translation. In *Proceedings of the Sixth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2023)*, pages 47–58, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.

**2023**  Andrei Popescu-Belis, **Àlex R. Atrio**, Bastien Bernath, Etienne Boisson, Xavier Theimer-Liemard, Teo Ferrari, and Giorgos Vernikos. GPoeT: a Language Model Trained for Rhyme Generation on Synthetic Data. In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 10–20, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.

**2022**  **Àlex R. Atrio** and Andrei Popescu-Belis. On the Interaction of Regularization Factors in Low-resource Neural Machine Translation. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pages 111–120, Ghent, Belgium, June 2022.

**2022**  Andrei Popescu-Belis, **Àlex R. Atrio**, Valentin Minder, Aris Xanthos, Gabriel Luthier, Simon Mattei, and Antonio Rodriguez. Constrained Language Models for Interactive Poem Generation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3519–3529, Marseille, France, June 2022.

2021 **Àlex R. Atrio** and Andrei Popescu-Belis. Small Batch Sizes Improve Training of Low-Resource Neural MT. In *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*, pages 18–24, National Institute of Technology Silchar, Silchar, India, December 2021.

2021 **Àlex R. Atrio**, Gabriel Luthier, Axel Fahy, Giorgos Vernikos, Andrei Popescu-Belis, and Ljiljana Dolamic. The IICT-Yverdon System for the WMT 2021 Unsupervised MT and Very Low Resource Supervised MT Task. In *Proceedings of the Sixth Conference on Machine Translation*, pages 973–981, Online, November 2021. Association for Computational Linguistics.

2019 **Àlex R. Atrio**, Toni Badia, and Jeremy Barnes. On the Effect of Word Order on Cross-lingual Sentiment Analysis. *Procesamiento del Lenguaje Natural (SEPLN)*, volume 63, pages 23–30, 2019.

### Presentations at Reviewed Conferences without Proceedings

2020 Andrei Popescu-Belis, Aris Xanthos, Valentin Minder, **Àlex R. Atrio**, Gabriel Luthier, and Antonio Rodriguez. Interactive poem generation: when language models support human creativity. *Presentation at the 5th Swiss Text Analytics and 16th KONVENS Conference (SwissText-KONVENS 2020)*, 2020.

### Presentations

2020 **Àlex R. Atrio**, Valentin Minder, Gabriel Luthier, and Andrei Popescu-Belis. Création Poétique Assistée par Ordinateur (CPAO). *Presentation at the Data Science Seminar, IICT / HEIG-VD*, 2020.

## Participation in Research Projects

2019 – 2023 **DOMAT - On-demand Knowledge for Document-level Machine Translation**
DOMAT (n. 175693) is a Swiss National Science Foundation project which aims at designing a novel approach for providing on-demand linguistic knowledge to neural machine translation systems.

2022 – 2023 **UNISUB - Unsupervised NMT with Innovative Multilingual Subword Models**
UNISUB is a project funded by Armasuisse. Its main goal is to improve unsupervised neural machine translation through an adaptive scheduling of the training tasks and through improved tokenization models, either using subword aligment across the source and target languages, or more advanced subword construction methods.

2020 – 2021 **FamilyMT - Multilingual Neural MT for Families of Languages and Dialects**
FamilyMT is an Armasuisse-funded project which aims at designing systems for closely related source languages of the same family, and show that this relatedness can be leveraged to improve translation of low-resource languages in multilingual neural machine translation.

2019 – 2020 **Digital Lyric - Création Poétique Assistée par Ordinateur**
Digital Lyric (n. 184330) is a project supported by the SNSF Agora program (main applicant: University of Lausanne). Its main goals are to design a system for interactive poem generation, which combines language models with explicit constraints that can be set by users on form, topic, emotion, and rhyming scheme. The system was showcased at a public exhibition at the Château de Morges (Vaud, Switzerland), from February 14 to May 10, 2020.

## Awards and Distinctions

2019 **Best Paper Award.** *"On the Effect of Word Order on Cross-lingual Sentiment Analysis",* SEPLN (Conference of the Spanish Society for Natural Language Processing), Bilbao.

## Professional Experience

2019 – 2023 **Research Assistant.** Haute École d'Ingénierie et de Gestion du Canton de Vaud, Switzerland.

## Teaching

### Teaching assistant for courses

2020 & 2022 Traitement Automatique des Langues (TAL) – BSc in computer science (3rd year).

## Languages (Proficiency)

| | |
|---|---|
| **Catalan** | Native |
| **Spanish** | Native |
| **English** | Fluent (C2 - *Proficiency Exam in English*) |
| **French** | Professional proficiency |
| **Galician** | Professional proficiency |

## Computer skills

| | |
|---|---|
| Programming | Python, Prolog |
| Visualizing | Matplotlib, Pandas |
| OS | Linux, Windows |
| Toolkits | PyTorch, TensorFlow |
| Misc | LaTeX, Git |