

Marc-Antoine Péguet

Natural Language Processing specialist

With a Master's degree in Automatic Language Processing (2024), I have worked on morphosemantics and lexical derivation through computational analysis of Russian suffixes. My background has given me expertise in derivational morphology, morphosyntactic analysis and distributional representations (word embeddings). I wish to continue this research at doctoral level, particularly in the computational modelling of word formation from a multilingual perspective.

EDUCATION

Master's Degree in Foreign and Regional Languages, Literatures and Civilisations, specialising in Automatic Language Processing (with distinction)

Université de Franche-Comté, Centre de Recherches Interdisciplinaires et Transculturelle (CRIT)
2022–2024 | Besançon - France

Skills: Machine Learning, Deep Learning, NLP, Python, Transformers, Spacy, Sickitlearn, NLTK, Linguistics, SQL, dependency grammar, lexicography

First year Master's degree in French as a Foreign Language, specialising in Language and Education Policy and Digital Environments

Université de Franche-Comté, Centre de Télé-enseignement Universitaire
2021–2022 | Besançon - France

Skills: The French-speaking world, Linguistics, Didactics, language policy, creation of teaching units.

Bachelor's Degree in Applied Foreign Languages (English – Russian)

Université de Franche-Comté, Sciences du Langage, de l'Homme et de la Société
2018–2021 | Besançon - France

Erasmus experience: University of Latvia (Riga, Latvia) from February to July 2021

Skills: English, Russian, translation, Economics, management, British civilisation, American civilisation, Russian civilisation.

RESEARCH EXPERIENCE

Data Science & NLP Intern: Automatic Classification of Knowledge Domains in Lexicographical Entries

Laboratories ERIC (Université Lyon 2) & LIRIS (INSA Lyon)
Feb. 2024–Aug. 2024 | Bron - France
Supervised by Julien Velcin (ERIC) and Ludovic Moncla (LIRIS)

- Development and testing of automatic classification models (BERT) to identify areas of knowledge in the 'Dictionnaire Universel François-Latin de Trévoux' (18th-century corpus).
- Construction and enrichment of an annotated dataset based on explicit/implicit markers, using natural language processing and semantic similarity.
- Implementation of supervised and semi-supervised approaches (constraint clustering, multi-level embeddings: word, sentence, article).
- Diachronic analyses comparing the 1743 and 1771 editions to study thematic and stylistic evolution.

Skills: Machine Learning, Deep Learning, NLP, Digital Humanities, Python, Transformers, word embeddings, BERT, BERTopic.

Master's thesis : The morphosemantics of Russian suffixes: developing a digital resource from a paper dictionary

Université de Franche-Comté, Centre de Recherches Interdisciplinaires et Transculturelle (CRIT)
2022–2024 | Besançon - France
Supervised by Izabella Thomas and Yağmur Öztürk

- Linguistic analysis of nominal suffixation in Russian and its morphosemantic uses.
- Design and structuring of a digital dictionary listing Russian noun suffixes.
- Structuring lexicographical data to facilitate its digital exploitation in standard XML format.
- Participation at the Fête de la Science (2024) in Besançon as a presenter at the CRIT stand.

Skills: Morphosemantics, Linguistics, suffixation, Russian language, NLP, XML, Word-Formation



+33 7 70 25 80 84



peguet.marcantoine@gmail.com



51 Rue de Lure, Athesans Etroitefontaine (70110)



www.linkedin.com/in/marc-antoine-peguet-a749061b8



<https://github.com/MAPEguet>

LANGUAGES

- French : Native language
- English : C1
- Russian : B1
- German : A2

HARD SKILLS

- Programming language : **Python**
- Libraries : **NLTK, Spacy, Pandas, Sickitlearn**
- WEB : **HTML, CSS**
- IDE : **Visual Studio Code, Jupyter, Ollama**
- Data base : **SQL (MySQL)**
- Markup : **XML, LATEX, TEI**
- Others : **TXM (corpus analyses), Protege (ontology), BERTopic**

OTHER WORK EXPERIENCE

Cashier

Supermarchés MATCH

Lure - France

Oct. 2024–Aug. 2025

July 2021–Jan. 2024

Apr. 2018–Jan. 2021