

Marc-Antoine Péguet

Spécialiste en Traitement Automatique des Langues

Diplômé d'un Master en Traitement Automatique des Langues (2024), j'ai travaillé sur la morphosémantique et la dérivation lexicale à travers l'analyse computationnelle des suffixes russes. Mon parcours m'a donné une expertise en morphologie dérivationnelle, analyse morphosyntaxique et représentations distributionnelles (word embeddings). Je souhaite poursuivre ces recherches en doctorat, notamment sur la modélisation computationnelle de la formation des mots dans une perspective multilingue.

FORMATIONS

Master Langues, Littératures et Civilisations Étrangères et Régionales, parcours Traitement Automatique des Langues (mention : Bien)

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

Mots-clés : Machine Learning, Deep Learning, NLP, Python, Transformers, Spacy, Sickitlearn, NLTK, linguistique, SQL, grammaire de dépendance, lexicographie

Master 1ère année Français Langue Étrangère, parcours Politiques Linguistiques - Éducatives et Environnements Numériques

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

Mots-clés : Francophonie, linguistique, didactique, politique linguistique, conception d'unité d'enseignement.

Licence en Langues Étrangères Appliquées (Anglais – Russe)

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

Stage Erasmus : University of Latvia (Riga - Lettonie) de février à juillet 2021

Mots-clés : Langue anglaise, langue russe, traduction, économie, management, civilisation britannique, civilisation américaine, civilisation russe.

EXPÉRIENCES DE RECHERCHES

Stagiaire en Data Science & TAL : Classification automatique des domaines de connaissance d'entrées lexicographiques

Laboratoires ERIC (Université Lyon 2) & LIRIS (INSA Lyon)
Fev. 2024 - Août 2024 | Bron - France
Encadré par Julien Velcin (ERIC) et Ludovic Moncla (LIRIS)

- Développement et expérimentation de modèles de classification automatique (BERT) pour identifier les domaines de connaissance dans le Dictionnaire Universel François-Latin de Trévoux (corpus XVIII^e). Étude, annotations et nettoyages de corpus.
- Construction et enrichissement d'un jeu de données annoté à partir de marqueurs explicites/implicites, via TAL et similarité sémantique.
- Mise en œuvre d'approches supervisées et semi-supervisées (clustering par contrainte, embeddings multi-niveaux : mot, phrase, article).
- Analyses diachroniques comparant les éditions 1743 et 1771 pour étudier l'évolution thématique et stylistique.

Mots-clés : Machine Learning, Deep Learning, NLP, humanités numériques, Python, Transformers, word embeddings, BERT, BERTopic.

Mémoire : La morphosémantique des suffixes russes : élaboration d'une ressource numérique à partir d'un dictionnaire papier

Université de Franche-Comté, Centre de Recherches Interdisciplinaires et Transculturelle (CRIT)
2022 - 2024 | Besançon - France
Encadré par Izabella Thomas et Yağmur Öztürk

- Analyse morphosémantique de suffixes russes (segmentation, catégorisation, annotations).
- Conception et structuration d'un dictionnaire numérique recensant les suffixes des substantifs russes.
- Structuration de données lexicographiques pour faciliter leur exploitation numérique au format standard TEI/XML.
- Participation à la Fête de la Science (2024) à Besançon en tant qu'animateur sur le stand du CRIT.

Mots-clés : Morphosémantique, linguistique, suffixation, langue russe, TAL, XML, formation de mots



+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

LANGUES

- Français : Langue maternelle
- Anglais : C1
- Russe : B1
- Allemand : A2

COMPÉTENCES

- Langage de programmation : **Python**
- Librairies : **NLTK, Spacy, Pandas, Sickitlearn**
- WEB : **HTML, CSS**
- IDE : **Visual Studio Code, Jupyter, Ollama**
- Bases de données : **SQL (MySQL)**
- Balises : **XML, LATEX, TEI**
- Autres : **TXM (analyse de corpus), Protege (ontologie), BERTopic**

EXPÉRIENCES PROFESSIONNELLES ANNEXES

Hôte de caisse

Supermarchés MATCH

Lure - France

Oct. 2024 - Août 2025

Juil. 2021 - Jan. 2024

Avr. 2018 - Jan. 2021