



Pedro J. Ortiz Suárez

DOCTORANT

✉ pedro.ortiz@inria.fr | 🌐 <https://pjortiz.fr/> | 📱 [pjox](#)

Intérêts

Language Modeling, Sequence Tagging, Reconnaissance d'Entités Nommées, Parsing, Linguistique de Corpus, Text Mining.

Formation

Sorbonne Université

DOCTORAT EN INFORMATIQUE

Contrat doctoral Inria

[Paris, France](#)

Oct. 2018 - Présent

Université de Paris 8, Vincennes-Saint-Denis

LICENCE MATHÉMATIQUES, INFORMATIQUE APPLIQUÉES AUX SCIENCES HUMAINES ET SOCIALES

Option sciences du langage

[Saint-Denis, France](#)

Nov. 2017 - Sep. 2018

Université d'Aix Marseille

MASTER EN MATHÉMATIQUES ET APPLICATIONS, PARCOURS MATHÉMATIQUES FONDAMENTALES

Bourse d'excellence académique LabEx Archimède

[Marseille, France](#)

Sep. 2016 - Jun. 2017

Universidad Nacional de Colombia

PREMIER CYCLE EN MATHÉMATIQUES FONDAMENTALES (BAC+4)

Admission honorable, troisième place au test d'admission

[Medellín, Colombie](#)

Jan. 2012 - Jun. 2016

Expérience

Inria - ALMAAnaCH Team

DOCTORANT

- Recherche en intelligence artificielle appliqué au traitement des langues naturelles.

[Paris, France](#)

Nov. 2018 - Présent

Sorbonne Université

CHARGÉ DE TRAVAUX DIRIGÉS

- Mathématiques discrètes, TD 40h.
- Éléments de programmation 2, TME 30h.

[Paris, France](#)

Sep. 2019 - Présent

Inria - ALMAAnaCH Team

RELAIS THÈSE

- Recherche en intelligence artificielle appliqué au traitement des langues naturelles.

[Paris, France](#)

Oct. 2018

Inria - ALMAAnaCH Team

STAGIAIRE

- Recherche en text mining et intelligence artificielle appliqué aux accords d'entreprise.

[Paris, France](#)

Apr. 2018 - Sep. 2018

Institut de Mathématiques de Marseille, I2M

STAGIAIRE

- Recherche en géométrie algébrique complexe, en surfaces canoniquement plongées de grands degrés.

[Marseille, France](#)

Avr. 2017 - Jun. 2017

Universidad Nacional de Colombia

CHARGÉ DE TRAVAUX DIRIGÉS

- Mathématiques discrètes, TD 128h.

[Medellín, Colombie](#)

Aoû. 2015 - Jun. 2016

Publications

[1] Pedro Javier Ortiz Suárez, Benoît Sagot, Laurent Romary. *A Monolingual Approach to Contextualized Word Embeddings for Mid-Resource Languages*, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Jul 2020, Association for Computational Linguistics, Online.

[2] Louis Martin*, Benjamin Muller*, Pedro Javier Ortiz Suárez*, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, Benoît Sagot. *CamemBERT : a Tasty French Language Model*, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Jul 2020, Association for Computational Linguistics, Online.

[3] Djamé Seddah, Farah Essaidi, Amal Fethi, Matthieu Futral, Benjamin Muller, Pedro Javier Ortiz Suárez, Benoît Sagot, Abhishek Srivastava. *Building a User-Generated Content North-African Arabizi Treebank : Tackling Hell*, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Jul 2020, Association for Computational Linguistics, Online.

[4] Louis Martin*, Benjamin Muller*, Pedro Javier Ortiz Suárez*, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, Benoît Sagot. *Les modèles de langue contextuels Camembert pour le Français : impact de la taille et de l'hétérogénéité des données d'entraînement*, 27e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2020), Jun 2020, Nancy, France.

[5] Pedro Javier Ortiz Suárez, Yoann Dupont, Benjamin Muller, Laurent Romary, Benoît Sagot. *Establishing a New State-of-the-Art for French Named Entity Recognition*, 12th International Conference on Language Resources and Evaluation (LREC 2020), May 2020, Marseille, France.

[6] Murielle Popa-Fabre, Pedro Javier Ortiz Suárez, Benoît Sagot, Éric de la Clergerie. *French Contextualized Word-Embeddings with a sip of CaBeRnet : a New French Balanced Reference Corpus*, 8th Workshop on the Challenges in the Management of Large Corpora (CMLC-8), May 2020, Marseille, France.

[7] Mohamed Khemakhem, Ioana Galleron, Geoffrey Williams, Laurent Romary, Pedro Javier Ortiz Suárez. *How OCR Performance can Impact on the Automatic Extraction of Dictionary Content Structures*, 19th annual Conference and Members' Meeting of the Text Encoding Initiative Consortium (TEI) -What is text, really? TEI and beyond, Sep 2019, Graz, Austria.

[8] Pedro Javier Ortiz Suárez, Benoît Sagot, Laurent Romary. *Asynchronous Pipelines for Processing Huge Corpora on Medium to Low Resource Infrastructures*, 7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7), Jul 2019, Cardiff, United Kingdom.

* Contribué à parts égales.

Talks

Séminaires du Lattice

DES MÉTHODES DE TAL MODERNES POUR L'ENRICHISSEMENT DE DOCUMENTS

École normale supérieure / Laboratoire Lattice

[Paris, France](#)

22/09/2020

58th Annual Meeting of the Association for Computational Linguistics

A MONOLINGUAL APPROACH TO CONTEXTUALIZED WORD EMBEDDINGS FOR MID-RESOURCE LANGUAGES

Association for Computational Linguistics

[En Ligne](#)

06/07/2020

7th Workshop on the Challenges in the Management of Large Corpora

ASYNCHRONOUS PIPELINE FOR PROCESSING HUGE CORPORA ON MEDIUM TO LOW RESOURCE INFRASTRUCTURES

Cardiff University

[Cardiff, Royaume-Uni](#)

22/07/2019

10th International Conference on Historical Lexicography and Lexicology

PREPARING THE DICTIONNAIRE UNIVERSEL FOR AUTOMATIC ENRICHMENT

Fryske Akademy

[Leeuwarden, Pays-Bas](#)

13/06/2019

GIG #3 : bring the cool back in the cloud

REDUCING COMPUTATION TIME BY MONTHS BY REWRITING BASH SCRIPTS IN GO

Golang Paris

[Paris, France](#)

24/03/2019

Distinctions

2016-2017 **Attribution bourse d'excellence Master 2**, Laboratoire d'excellence LabEx Archimède.

Marseille, France

Jan. 2012 **Bourse au mérite**, Universidad Nacional de Colombia.

Medellín, Colombie

Aoû. 2012 **Bourse au mérite**, Universidad Nacional de Colombia.

Medellín, Colombie

Langues

Espagnol - Langue maternelle, **Anglais** - Courant, **Français** - Courant, **Allemand** - Élémentaire.

Compétences informatiques

- **Courant** : C/C++, Python, Go, TensorFlow, Keras, \LaTeX , VS Code, GNU/Linux, OS X, Git, SQL, MySQL.
- **Niveau avancé** : Raspberry Pi, Matlab, Wolfram Mathematica, JavaScript, PHP, HTML5, Java, Windows, MongoDB, OpenSSH, GnuPG, Apache Server, NumPy, GIMP.