

# Caroline KOUDORO-PARFAIT

DOCTEURE (PHD) · HUMANITÉS NUMÉRIQUES, LITTÉRATURE COMPARÉE

Sorbonne Université, Paris, France

✉ caroline.parfait@sorbonne-universite.fr | 🌐 <https://carolinekoudoroparfait.github.io/> | 📄 <https://github.com/These-SCAI2023> |

📞 caroline-koudoro-parfait-668245159

## Parcours Universitaire

### Sorbonne Université, France

Paris, France

#### THÈSE DE DOCTORAT, HUMANITÉS NUMÉRIQUES ET LITTÉRATURE COMPARÉE

2020 - 2025

- Directeurs de Thèse : Glenn Roe et Gaël Lejeune ; co-encadrant de Thèse : Motasem Alrahabi
- Titre de la thèse de doctorat : *Des IA au service de l'espace littéraire du XIX<sup>e</sup> siècle : évaluation et analyse des outils de reconnaissance d'entités nommées spatiales.*
- Obtic - Labex Obvil, Observatoire des textes, des idées et des corpus ; Sorbonne Center for Artificial Intelligence (SCAI) ; Laboratoire Sens Texte Informatique Histoire (STIH)

### Centre d'études supérieures de la Renaissance de Tours (CESR)

Tours, France

#### MASTER INTELLIGENCE DES DONNÉES DE LA CULTURE ET DES PATRIMOINES

2018 - 2020

- Directrice de mémoire : Elena Pierazzo
- Titre du mémoire : *Inventorier, Modéliser, Concevoir, Évaluer : enjeux et pratiques des Humanités Numériques au sein d'une entreprise de développement informatique.*

### École Nationale Supérieure d'Arts de Paris-Cergy (ENSAPC)

Cergy, France

#### DIPLÔME NATIONAL SUPÉRIEUR D'EXPRESSION PLASTIQUE GRADE MASTER

2011 - 2017

- Directeur de mémoire : François Bon
- Titre du mémoire : *L'inertie, le mouvant*
- Pratiques plastiques : Sérigraphie, vidéo, creative writing et écritures numériques

### Université, Paris 4, Paris-Sorbonne

Paris, France

#### MAÎTRISE DE PHILOSOPHIE DE L'ART ET ESTHÉTIQUE

2007 - 2011

- Enseignements mineurs : Philosophie et histoire des sciences, langages logiques, communication et philosophie politique et éthique.

## Expériences professionnelles

- |              |   |
|--------------|---|
| 2023-Présent | Attachée temporaire d'enseignement et de recherche (A.T.E.R.), Sorbonne Université  |
| 2025         | Co-encadrement du stage de Master 1, <i>Reconnaissance d'Entités Nommées sur un corpus de poèmes du XIX<sup>e</sup> siècle</i> , LOPES M. Licence sciences du langage, Sorbonne Université  |
| 2025         | Co-encadrement du stage de Master 1, <i>Entraînement et évaluation d'un modèle de Reconnaissance d'Entités Nommées sur données bruitées</i> , FIEF, C., CANONGE Y. Licence sciences du langage, Sorbonne Université   |
| 2024         | Co-encadrement du stage de Licence 3, <i>Évaluation des problèmes d'étiquetage de la REN.</i> , FIEF, C. Licence sciences du langage, Sorbonne Université   |
| 2021-2023    | Chargée d'enseignement vacataire, Programmation Python, Institut National des Langues et Civilisations Orientales - INALCO  |
| 2023         | Co-encadrement du stage de Master 2, <i>Évaluation et développement de méthodes pour la désambiguïsation sémantique d'entités nommées spatiales</i> , DORLÉANS, C. Master Intelligences des données de la culture et des patrimoines, Centre d'études supérieures de la Renaissance, Tours et Sorbonne Université |
| 2022         | Co-encadrement du Mémoire de Master 1, <i>Résolution de problèmes d'Entity Linking sur des sorties OCR bruitées</i> , BUTH, P-R., Sciences du Langage, Langue et informatique, Sorbonne Université  |
| 2010-2022    | Répétitrice de Littérature française, cours collectifs et individuels - Lycée, Acadomia   |
| 2020         | Stage d'ingénierie en Humanités numériques : Conception et développement d'une bibliothèque numérique sous Omeka S et état de l'art des outils numériques des SHS, Acatius Informatique, Orléans  |

## Détails des expériences d'enseignement

2024 - Présent	<b>Responsable, L5SOPRG - Programmation Python, CM et TD</b> , LICENCE – Niveau L3	Sorbonne Université
2023 - Présent	<b>L6SOPRG - Programmation Python, CM et TD</b> , LICENCE – Niveau L3	Sorbonne Université
2023 - Présent	<b>PIX, CM et TD</b> , LICENCE	Sorbonne Université
2023 - Présent	<b>M2SOL023 - Analyse Lexicale en TAL, CM et TD</b> , MASTER	Sorbonne Université
2023 - Présent	<b>M4LFHUNU - Humanités Numériques, CM et TD</b> , MASTER	Sorbonne Université
2023 - 2024	<b>L5SOPRG - Programmation Python, CM et TD</b> , LICENCE – Niveau L3	Sorbonne Université
2023 - 2024	<b>L4SOINGE - Ingénierie des Langues, TD</b> , LICENCE	Sorbonne Université
2021 - 2023	<b>Algorithmique et structuration des données</b> , LICENCE LLCER parcours Traitement automatique des langues – Niveau L3	INALCO
2021 - 2023	<b>Programmation python</b> , LICENCE LLCER parcours Traitement automatique des langues – Niveau L3	INALCO
2012-2023	<b>Préparation des épreuves anticipées écrites et orales du baccalauréat de français</b> , suivi hebdomadaire et formations intensives en groupe, Oraux blancs	Academia
2021-2022	<b>Préparation de l'épreuve du grand oral du cycle terminal STD2A (Sciences et technologies du design et des arts appliqués)</b> , cours particuliers et oraux blancs	Academia
2010 - 2021	<b>Répétitrice de français en cours particuliers</b> , suivi hebdomadaire et formations intensives	Academia
2012 - 2018	<b>Méthodologie : Apprendre à apprendre en 2nde</b> , formations intensives en groupe	Academia
2012 - 2016	<b>Stage de français niveau 2nde</b> , suivi hebdomadaire et formations intensives en groupe	Academia
2012 - 2016	<b>Préparation de l'épreuve de français du Brevet des collèges</b> , formations intensives en groupe	Academia

## Compétences informatiques

LANGAGES : Python, HTML5 / CSS3

**TRAITEMENT DONNÉES TEXTUELLES, FOUILLE DE TEXTES** : Spacy, Stanza, CasEn, SEM, Unitex, TXM, Voyant Tools

**DONNÉES STRUCTURÉES** : XML-TEI/MEI/EAD, Dublin Core, RDF, RDFs, SKOS, CIDOC-CRM

**REPRÉSENTATION STATISTIQUE :**

Matplotlib, Seaborn, Pandas, OpenCv, Rstudio, Gephy

**LOGICIELS DE TRAITEMENT MULTIMÉDIA :**

QGIS, Unity, inDesign, Photoshop, illustrator

## Publications

Quelle solution pour améliorer les performances de la reconnaissance d'entités nommées sur des données bruitées, corriger l'entrée ou filtrer la sortie ? Petkovic, L., Koudoro-Parfait, C., Lejeune, G. Article accepté par la revue Corpus, à paraître en janvier 2025.

Analyse multilingue de l'impact de la correction automatique de l'OCR sur la reconnaissance d'entités nommées spatiales dans des corpus littéraires. Koudoro-Parfait, C., Petkovic, L., ROE, G. Article accepté par la revue TAL, à paraître en janvier 2025.

Reconnaissance des Entités Nommées spatiales sur un corpus littéraire bruité : des entités à la carte. Koudoro-Parfait, C., Lejeune G. Chapitre dans l'ouvrage "Des sources aux systèmes d'information géographique : des outils pour la cartogra-

phie dans les humanités numériques”, Laboratoire DYPAC UVSQ Paris-Saclay, Plateforme Géomatique de EHESS. Editions Matériologiques dans la collection Modélisations, simulations, systèmes complexes. Article accepté, à paraître.

Mapping spatial named entities from noisy OCR output: Épiméthée from OCR to map. Koudoro-Parfait, C., Alrahabi, M., Dupont, Y., Lejeune, G., Roe, G. 2023. Scholger, W., Vogeler, G., Tasovac, T., Baillot, A., Helling, P., ADHO Digital Humanities Conference 2023 (DH2023) Graz. 10.5281/zenodo.7961822 (DOI)

SPQR@Deft2023: Résolution automatique de QCM médicaux à partir de corpus de domaine et de mesures de similarité. Bezançon J., Boubehziz T., Chutaux C., Zine O., Acensio, L., Koudoro-Parfait, C., Briglia A., Lejeune G. Actes du Défi Fouille de Textes@TALN2023, ATALA, Jun 2023, Paris, France. [https://coria-taln-2023.sciencesconf.org/data/proceedings\\_DEFT.pdf](https://coria-taln-2023.sciencesconf.org/data/proceedings_DEFT.pdf) hal-04131847

Évaluation de la tâche de clustering pour l’alignement de formes contaminées d’entités nommées issues d’un corpus OCR bruité. Koudoro-Parfait, C. 2022. Corro, C., Lejeune, G. Actes de la journée d’étude sur la robustesse des systèmes de TAL, p. 5.

Reconnaissance d’entités nommées sur des sorties OCR bruitées: des pistes pour la désambiguïsation morphologique automatique. Koudoro-Parfait, C., Lejeune G., Buth, R. 2022. Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Atelier TAL et Humanités Numériques (TAL-HN), p. 45-55.

Stylo@ DEFT2022: Notation automatique de copies d’étudiant·e·s par combinaisons de méthodes de similarité. Ben Ltaifa, I., Boubehziz, T., Briglia, A., Chutaux, C., Dupont, Y., González-Gallardo, C., Koudoro-Parfait, C., Lejeune, G. 2022. Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Atelier Défi Fouille de Textes (DEFT), ATALA, p. 11-22.

Spatial Named Entity Recognition in Literary Texts: What is the Influence of OCR Noise ?. Koudoro-Parfait, C., Lejeune, G., Roe, G. 2021. Moncla, L., Brando, C., McDonough, K. GeoHumanities@SIGSPATIAL 2021: Proceedings of the 5th ACM, SIGSPATIAL International Workshop on Geospatial Humanities, Beijing, China, 2 - 5 Novembre, p. 13-21.

## Présentations

---

L’impact de l’IA sur les pratiques documentaires : regards croisés entre chercheurs et bibliothécaires. Koudoro-Parfait, C., Bouchard, A., Solnon, C. 2024. Bouju, P. ENSSIB.

S’appropriier sa recherche. La posture des doctorantes et doctorants ..., axe : « ... face aux Intelligences artificielles (IA) ». Koudoro-Parfait, C., Nascimento, L., Courtois, M. 2024. Collectif doctoral Sorbonne Actuelle.

De l’OCR à la visualisation de données. Koudoro-Parfait, C., Hernandez, M. 2024. Atelier ObTIC, BnF.

Garder trace de son travail et le partager en traitement automatique des langues. Koudoro-Parfait, C. 2023. Atelier Digit\_Hum 2023, Manipuler des données en SHS : R, Python et les carnets interactifs. Massot M.-L., Tricoche, A., Schultz, E. École normale supérieure.

OCR impact on Named Entity Recognition on historical documents : Problem Presentation ...and Epistemological consequences. Koudoro-Parfait, C., Lejeune, G. 2023. Summer School on Digital Humanities, UNAM/SCAI, Université nationale autonome du Mexique.

Mapping spatial named entities from noisy OCR output: Épiméthée from OCR to map. Koudoro-Parfait, C., Alrahabi, M., Dupont, Y., Lejeune, G., Roe, G. 2023. ADHO Digital Humanities Conference 2023 (DH2023) Graz.

Correction automatique des interférences OCR dans la reconnaissance d’entités nommées spatiales : réel gain ou perte de l’information ?. Koudoro-Parfait, C., Petkovic, L., Lejeune, G. 2023. Journée d’étude à Grenoble organisée par l’Université Grenoble Alpes et Roma La Sapienza.

Reconnaissance des Entités Nommées spatiales dans un corpus littéraire bruité : des entités à la carte. Koudoro-Parfait, C. 2022. Journée d’étude organisée par ObTIC : Analyse spatiale des textes littéraires, Alrahabi, M., Roe, G.

Des formats de données pour mes sources : un casse-tête ?!. Koudoro-Parfait, C. 2022. Séminaire doctoral CERES. Grison, T., Bordier, C.

Annotation et REN sur des OCR bruités : données, utilisateurs, outils. Koudoro-Parfait, C. 2022. Séminaire doctoral, Natural Language Processing meets Digital Humanities, Baillot, A., Magué, J.-P.

De la création des corpus numériques à leur utilisation : les interférences dans la reconnaissance d’entités nommées spatiales. Koudoro-Parfait, C., Lejeune, G. 2022. Séminaire “Des sources aux systèmes d’information géographique : des outils pour la cartographie dans les humanités numériques”, Laboratoire DYPAC UVSQ Paris-Saclay, Plateforme Géomatique de EHESS.

Spatial Named Entity Recognition in Literary Texts: What is the Influence of OCR Noise ?. Koudoro-Parfait, C., Lejeune, G., Roe, G. Novembre 2021. Moncla, L., Brando, C., McDonough, K. GeoHumanities@SIGSPATIAL 2021: Proceedings of the 5th ACM, SIGSPATIAL International Workshop on Geospatial Humanities.

Réutiliser les corpus numériques constitués par OCR : bruit et silence dans les data papers. Koudoro-Parfait, C., Lejeune, G., Tangy, J.-B. Novembre 2021. DHNord2021.

Reconnaissance des Entités Nommées (REN) spatiales dans un corpus littéraire : robustesse des systèmes existants ?. Koudoro-Parfait, C. Novembre 2021. Séminaire Humanités Numériques Littéraires, Roe, G., ObTIC.

Workshop ROC & NER. Parfait, C., Tangy, J.-B. 2021. Maison de la recherche, Sorbonne Université, Paris.

Reconnaissance d'entités nommées spatiales dans des transcriptions OCR bruitées de textes littéraires. Koudoro-Parfait, C., Lejeune, G. 2021. Atelier Humanités Numériques Spatialisées SAGEO 2021.

Discovering Spatial Relations in Literature : What is the influence of OCR noise?. Koudoro-Parfait, C., Lejeune, G. 2021. NewsEye International Conference, Programme de l'Union Européenne.