|  |  |
| --- | --- |

SORBONNE UNIVERSITÉ

ÉCOLE DOCTORALE III « Littérature française et comparée » (ED019)

Centre d’étude de la langue et des littératures françaises (CELLF), UMR 8599

Équipe-projet « Observatoire des Textes, des Idées et des Corpus » (ObTIC)

THÈSE

pour obtenir le grade de

DOCTEUR DE SORBONNE UNIVERSITÉ

Discipline : Humanités numériques

Présentée et soutenue par :

Ljudmila PETKOVIĆ

**Dans les petits papiers de Charcot : de l’expérimentation aux prémisses de la neurologie moderne**

Directeur : Prof. Dr Glenn ROE, Sorbonne Université

Co-encadrant : Dr Motasem ALRAHABI, Sorbonne Université

Feuille de route

Corpus et OCR

Correction

Exploration du corpus: circulation des concepts et alignement.

# 

# Résumé

# 

# 

# Avant-propos

Ce travail faisant l’objet de la présente thèse de doctorat s’inscrit dans le cadre du programme *Instituts et Initiative* de l’Observatoire des Patrimoines (OPUS) de l’Alliance Sorbonne Université (ASU). Une campagne d’attribution de contrats doctoraux financés par l’Initiative d’Excellence (IdEx), nommée « Campagne 2021 Contrats Doctoraux Instituts/Initiatives » est lancée en janvier 2021 ; grâce à cette initiative, le projet doctoral intitulé « Dans les petits papiers de Charcot : de l’expérimentation aux prémisses de la neurologie moderne » a été sélectionné et retenu fin mai 2021 dans le cadre de l’appel à candidatures.[[1]](#footnote-0)

Le projet en question est conçu par Prof. Dr Glenn Roe, directeur de l’équipe-projet ObTIC (Observatoire des Textes, des Idées et des Corpus)[[2]](#footnote-1) de Sorbonne Université (SU), en concertation avec Dr Motasem Alrahabi, coordinateur scientifique de l’ObTIC, ainsi qu’avec l’autrice de la présente thèse. La dernière avait été recrutée en qualité de doctorante contractuelle au sein de la même équipe et rattachée à l’École doctorale III « Littérature française et comparée » (ED019), ainsi qu’au Centre d’étude de la langue et des littératures françaises (CELLF), UMR 8599 de la Faculté des Lettres de SU.

Ce travail s’intègre dans quatre axes de recherche transversaux en humanités numériques, hérités de l’ancien laboratoire d’excellence Observatoire de la Vie Littéraire (LabEx OBVIL) :

1. l’édition numérique des textes ;
2. la fouille de données et l’herméneutique des textes dans une perspective interdisciplinaire ;
3. le développement d’applications d’analyse textuelle ;
4. la réflexion théorique et critique menée sur cette méthodologie.

Le même projet est le fruit d'une collaboration étroite de l’ObTIC avec plusieurs institutions universitaires et patrimoniales, notamment la Bibliothèque de Sorbonne Université (BSU), la Bibliothèque de Neurosciences Jean-Martin Charcot, ainsi que l’Université de Genève (Unige), Faculté des Lettres, Chaire des humanités numériques. Étant donné les implications médicales induites par ce projet, il a été considéré au départ comme susceptible d’intéresser l’Initiative Humanités Médicales (HuMed) de SU.

# 

# Remerciements

BSU… transcriptions

Bib Charcot… les ouvrages restants  
SU, OPUS…

Unige… outils, infrastructure (eScriptorium)

# 

# Organisation de ce présent livre

# Introduction

# I Contexte et problématique

# II État de l’art

# III Constitution du corpus Charcot

## 1. Définition du corpus

## 2. Récolte des ouvrages

### 2.1 BSU

### 2.2 HTR (eScriptorium)

### 2.3 Bibliothèque Charcot

## 4. Correction automatique des transcriptions

## 5. Outils de génération du corpus

### 5.1 OBVIE

* tour « du propriétaire » de la BDD
* question d’interface, d’outils et de visualisations
* ouverture du corpus (HumaNum)
* bilan provisoire

# IV Exploration du corpus (OBVIE)

## 1. Circulation des concepts

## 2. Comparaison des textes (alignement)

# 

# Résultats

# Discussion

# Conclusion et perspectives

# Bibliographie

# Annexes

# Liste des abréviations, des sigles et des acronymes

**ASU** Alliance Sorbonne Université

**BSU** Bibliothèque de Sorbonne Université

**HN** Humanités numériques

**HTR** angl. *Handwritten Character Recognition*

(reconnaissance d’écriture manuscrite)

**HuMed** Humanités médicales

**IdEx** Initiative d’excellence

**LabEx** Laboratoire d’excellence

**ObTIC** Observatoire des textes, des idées et des corpus

**OBVIL** Observatoire de la vie littéraire

**OCR** angl. *Optical Character Recognition*

(reconnaissance optique de caractères)

**OPUS** Observatoire des patrimoines de l’Alliance Sorbonne Université

**SU** Sorbonne Université

**UMR** Unite de mixte recherche

**Unige** Université de Genève

# Glossaire

# Liste des tableaux

# Table des figures

# Table des matières

# 

1. Plus de détails concernant le déroulement de la campagne en question : <https://institut-opus.sorbonne-universite.fr/actualites-opus/campagne-dattribution-de-contrats-doctoraux> (consulté le 10 mars 2022). [↑](#footnote-ref-0)
2. Présentation de l’ObTIC : <https://obtic.sorbonne-universite.fr/presentation/> (consulté le 11 mars 2022). [↑](#footnote-ref-1)