



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

Observatoire des Patrimoines (OPUS)
École doctorale III, Faculté des Lettres

Sorbonne Université, Paris (France)

Candidate :
Ljudmila PETKOVIĆ

n° de matricule :
19-337-757

Directeur de thèse :
Glenn ROE
Professeur des Universités

Co-encadrant :
Motasem ALRAHABI
Ingénieur de recherche, doctorat

Mots clés : humanités numériques, valorisation du patrimoine, fonds de Jean-Martin Charcot, traitement automatique des langues, fouille de textes, alignement, XML-TEI.

1. Contexte

Ce projet de thèse, à la jonction des domaines médico-scientifiques et des sciences humaines et sociales, propose une étude pluridisciplinaire dont l'objectif est la valorisation des fonds patrimoniaux dans la perspective des humanités numériques (HN). Notre objet d'étude porte sur les archives de Jean-Martin Charcot, précurseur de la psychopathologie et fondateur de la neurologie moderne au XIX^e siècle en France. Ces archives volumineuses et d'une grande importance scientifique (sous forme physique et numérisée) sont situées à la Bibliothèque de Neurosciences Jean-Martin Charcot / Bibliothèque numérique patrimoniale de la Faculté des Sciences de Sorbonne Université (Jubilothèque), ainsi que sur Gallica, bibliothèque numérique de la Bibliothèque nationale de France (BNF).

Figure emblématique et directeur de L'École de la Salpêtrière, Charcot a laissé une marque indélébile dans le domaine de la neurologie grâce à sa réhabilitation de l'hypnose et de l'hystérie comme objets d'étude scientifique. Après sa mort, son travail a été poursuivi par ses étudiants, parmi lesquels se distinguaient Sigmund Freud, un de ses premiers traducteurs en allemand (Charcot, 2002), dont les études psychanalytiques sur l'hystérie et l'hypnose puisent leurs racines dans l'œuvre de Charcot (Oliveira *et al.*, 2018) ; Joseph Babinski, deuxième chef de clinique à la Salpêtrière (Massie, 2004), qui approfondit les recherches dans le domaine de l'hystérie associée aux psychonévroses de la I^{re} Guerre mondiale (Bogousslavsky, 2011) ; Alfred Binet, contribuant, en partenariat avec Théodore Simon, au développement d'échelles de test d'intelligence (Teive *et al.*, 2017). Pierre Janet, développeur des bases de la théorie « traumatico-dissociative » des névroses (Perez, 2011) ; Gilles de la Tourette, qui publie son traité¹ en trois volumes sur l'hystérie ; Paul Richer, dont les études cliniques² résument les concepts de Charcot élaborés dans ses premières études sur l'hystérie (Bogousslavsky, 2014).

Le fonds en question est fortement transdisciplinaire grâce aux nombreuses collaborations de Charcot avec des savants appartenant à différents domaines scientifiques. Cette double valeur est exprimée sous forme discursive et iconographique : les implications dans le domaine de l'esthétique découlent des travaux entrepris par Charcot avec le médecin et sculpteur Paul Richer ;³ d'autre part, nous pouvons citer ses travaux avec le psychologue Alfred Binet et le philosophe Théodule Ribot (Schultheiss, 2021), ainsi que ceux du neurologue et d'un des héritiers de Charcot, Désiré Magloire

¹ *Traité clinique et thérapeutique de l'hystérie d'après l'enseignement de la Salpêtrière.*

² *Études cliniques sur l'hystéro-épilepsie ou grande hystérie (Studies on Hystero-Epilepsy).*

³ cf. les images dans *Les démoniaques dans l'art* de Charcot et Richer – <https://gallica.bnf.fr/view3if/ga/ark:/12148/bpt6k6492840j/f24> (consulté le 2 mai 2021).

Bourneville, travaux situés au carrefour de la psychiatrie et de la philosophie (Bourneville, 1864). La place prise par la religion au sein de la médecine clinique est évidente dans l'œuvre *La foi qui guérit* (Charcot, 2015), dont le point central est celui de la guérison miraculeuse. Ce transfert disciplinaire des pratiques médicales de Charcot aux disciplines des lettres et sciences humaines se reflète également dans la pièce théâtrale *Une leçon à la Salpêtrière* de Binet, conçue en collaboration avec le dramaturge André de Lorde (Lorde, 1908).

Afin de mettre en valeur un tel corpus, cette thèse englobera les disciplines distinctes et transversales au prisme des HN : la linguistique computationnelle (la fouille et l'alignement de textes) ; l'histoire de la pensée médicale dans le domaine de la neurologie du XIX^e siècle, les transferts de la neurologie à la psychologie et de la philosophie des sciences ; l'esthétique et la pensée critique en art et en littérature.

2. Objectifs de l'étude

Bien qu'il existe une pléthore de publications portant sur le fonds Charcot du point de vue théorique, il n'existe pas à notre connaissance d'initiatives en humanités numériques visant la constitution et l'exploitation de ce fonds. L'objectif de la thèse est donc de valoriser les archives de Charcot déjà numérisées, en les complétant par certaines œuvres importantes en collaboration avec la Bibliothèque de Sorbonne Université (BSU), grâce à laquelle nous serons en mesure d'identifier un échantillon représentatif des documents (imprimés, manuscrits, illustrés, etc.) pour les numériser et les fouiller. Cette valorisation sera concrétisée par la mise en place d'une chaîne de traitement automatique qui va de l'océrisation à la fouille sémantique et la comparaison de textes, en passant par la correction et la structuration de ceux-ci au format XML-TEI. Ces traitements nous permettront déjà de produire une édition numérique savante et exploitable dans notre cadre de recherche. À ce titre, nous nous intéressons tout particulièrement à l'analyse de la genèse et la migration du discours médical de pathologie anatomique, de neurologie et psychologie de Charcot dans les écrits co-réalisés en collaboration et dans les écrits de ses continuateurs et disciples.

3. Méthodologie

Notre démarche s'articulera en deux temps : l'analyse du discours neuropathologique de Charcot, et l'influence de celui-ci sur sa propre discipline et sur les autres domaines.

Tout d'abord, la fouille de textes, notamment les approches textométriques et les visualisations appropriées, générées par exemple par les plongements des mots,

faciliteront l'identification des champs lexicaux et les isotopies, ainsi que la modélisation des sujets ou des marqueurs spécifiques (iconographiques, religieux, théâtraux...) de la circulation des idées de Charcot. Les analyses préliminaires seront effectuées dans le logiciel textométrique TXM⁴. Dans la perspective de la modélisation des sujets, nous pourrions exploiter le logiciel LDAvis, utilisé dans l'étude de Huber (2016), alors que pour l'extraction des concepts médicaux nous allons tester les CRF⁵ et les systèmes d'apprentissage profond tels que BiLSTM⁶, MTL⁷ ou BioBERT⁸, proposés par Schulz *et al.* (2020). Pour générer les plongements de mots, les outils comme ELMo ou BERT semblent les outils les plus appropriés (Si *et al.* 2019).⁹

Ensuite, le projet va évaluer l'impact des approches neurologiques innovantes de Charcot aux autres disciplines en détectant les reprises textuelles, les filiations et les concepts communs. Ce choix méthodologique est justifié étant donné les travaux des scientifiques (Freud, Binet, Janet etc.) en lien avec les pratiques de Charcot, mentionnés dans la 1^{re} section, § 2 ; L'objectif ne réside pas uniquement dans le repérage de ces transferts : cette intertextualité, qu'il faudra mesurer dans sa précision, est observée de manière critique, en sorte que nous cherchions à mesurer si le transfert d'un champ disciplinaire à l'autre est littéral ou au contraire s'accompagne de modifications sémantiques et opératoires. La réponse à cette question nécessite d'aligner les textes de Charcot avec d'autres corpus, en utilisant les outils mis à disposition par le laboratoire d'excellence OBVIL. Les alignements de textes s'effectueront avec les outils MEDITE¹⁰, DeSeRT¹¹ et TextPAIR¹². La plateforme OBVIE¹³ fournira l'accès aux documents nécessaires pour la comparaison.

L'effet des pratiques de Charcot est également observable dans les textes littéraires, tels que *Lourdes* de Zola, *L'Hystérique de Lemmonier* ou *Le Horla* de Maupassant (Marquer, 2008). Sous cet angle, nous chercherons à identifier la manière dont les écrivains ont utilisé l'image de l'hystérie féminine telle qu'elle a été mise en scène par Charcot afin d'exorciser les angoisses concernant le déclin national, l'influence corruptrice de la vie urbaine et la menace toujours présente de la folie et de la fragmentation subjective. Nous incluons donc la dimension intertextuelle de la théorie riffaterrienne (Michael, 1980) au sein des discussions du rapport entre les travaux de Charcot et ses successeurs, en adoptant l'approche globale de la lecture distante, dans la lignée des travaux de Moretti et autres (2013; Underwood, 2017).

⁴ <http://textometrie.ens-lyon.fr/?lang=en> (consulté le 2 mai 2021).

⁵ Angl. *Conditional Random Fields*, i.e. les champs aléatoires conditionnels.

⁶ Angl. *Bidirectional Long/Short-Term Memory*.

⁷ Angl. *Multi-Task Learning*, i.e. l'apprentissage multi-tâche.

⁸ Angl. *Bidirectional Encoder Representations from Transformers*.

⁹ En comparaison avec les méthodes traditionnelles word2Vec, GloVe et fasttext.

¹⁰ <https://obvil.sorbonne-universite.fr/developpements/medite> (consulté le 2 mai 2021).

¹¹ <https://obvil.sorbonne-universite.fr/developpements/desert> (consulté le 2 mai 2021).

¹² <https://artfl-project.uchicago.edu/text-pair> (consulté le 2 mai 2021).

¹³ <https://github.com/OBVIL/obvie> (consulté le 2 mai 2021).

Nous souhaitons également croiser les données lexicales avec une analyse sémantique fine ; croisement qui est rendu possible grâce à la plateforme ARIANE¹⁴ développée au sein du LabEx pour analyser et comparer quel est le traitement du corps, par exemple, dans l'association de la névrose et du génie.

Il sera envisageable de visualiser les données, en repérant la circulation des pratiques de Charcot *via* des analyses de réseaux dans le logiciel Gephi¹⁵ (méthodologie notamment appliquée par Ji *et al.* (2018),¹⁶ ou dans le langage de programmation R. À part l'utilisation des plongements de mots avec ELMo ou BERT, nous essaierons de cartographier des termes relevant du discours de Charcot, en s'inspirant du travail de Ewing (2017), qui expose les connexions des termes à la fois à la grippe et à l'épidémie, ainsi que ceux qui sont liés soit à la grippe, soit à l'épidémie.

La numérisation exhaustive du corpus n'est pas donc l'objectif primordial de ce projet, en raison du temps limité pour mener le projet à bien. En revanche, il est prévu d'entraîner et d'expérimenter des outils d'OCR (p. ex. tesseract¹⁷ ou kraken¹⁸) ou d'HTR comme Transkribus¹⁹ (dans la mesure du possible), pour la numérisation des documents / objets manuscrits (p. ex. notes de cours de Charcot) qui restent à être numérisés dans l'optique de notre projet. L'établissement d'un tel corpus de recherche représentera l'un des apports majeurs du projet pour la communauté patrimoniale.

4. Corpus

Le fonds en question fait l'objet de différents travaux théoriques. Nous comptons ainsi l'histoire complète de l'hypnose et de la psychologie clinique (Nicolas, 2004) ; les réflexions philosophiques sur l'œuvre scientifique de Charcot (Micale, 2019) ; la démystification du phénomène de l'hystérie (Andrieu, 1999) ; l'alliance entre le théâtre et les modèles diagnostiques et pédagogiques de Charcot (Marshall, 2016) ; les descriptions des pratiques cliniques et expérimentales autour de l'hystérie (Didi-Huberman, 1982) ; un compte-rendu de la vie et des contributions scientifiques de Charcot (Tréanton, 1997 ; Goetz *et al.*, 1995).

L'OBVIL-ObTIC dispose déjà en format XML-TEI des neuf volumes des *Œuvres complètes* de Charcot qui demandent à être intégrés dans la chaîne de traitement. Le fonds est constitué de la bibliothèque personnelle de Jean-Martin Charcot, *i.e.* des

¹⁴ <https://obvil.huma-num.fr/ariane/> (consulté le 2 mai 2021).

¹⁵ <https://gephi.org> (consulté le 2 mai 2021).

¹⁶ Voir également une analyse en réseau basée sur la fréquence des artistes à des expositions prestigieuses, au sein du projet *Visual Contagions* (Joyeux-Prunel, 2019), dont le point fondamental représente la circulation du style et des œuvres de l'art au prisme des HN.

¹⁷ <https://github.com/tesseract-ocr/tesseract> (consulté le 2 mai 2021).

¹⁸ <http://kraken.re> (consulté le 2 mai 2021).

¹⁹ <https://readcoop.eu/transkribus/?sc=Transkribus> (consulté le 2 mai 2021).

ouvrages de lui-même et de ses élèves, périodiques, collections de thèses et de tirés à part, manuscrits, observations, collection neurologique couvrant la seconde part du XIX^e siècle, augmenté de la bibliothèque des Internes de la Salpêtrière (ouvrages, périodiques, thèses en neurologie et psychiatrie pour la période 1900-1950).

5. Résultats attendus

Cette thèse va permettre de mettre à disposition un corpus numérique inédit et exploitable de la part de la communauté scientifique et des participants dans la présente recherche. En confrontant ces différentes données et les divers outils numériques, nous allons fabriquer et évaluer la chaîne de traitement comprenant la numérisation des textes (l'OCR, la correction automatique de sortie OCR), la structuration des données au format XML-TEI, la fouille sémantique et l'alignement des textes, à l'aide de laquelle nous allons rendre compte de la circulation transdisciplinaire des emprunts, des influences et des reprises chez les contemporains, savants, artistes et intellectuels autour de Charcot. Ce projet va contribuer à mettre à jour l'état de l'art, et donner lieu à une réflexion sur la valorisation des archives patrimoniales en général, ainsi que sur celles hébergées par Sorbonne Université.

6. Conclusion

Cette recherche apportera deux contributions principales. Premièrement, la valorisation du patrimoine scientifique de Jean-Martin Charcot en constituant un corpus numérique à partir de ses archives déjà numérisées ainsi que la numérisation ponctuelle de nouveaux ouvrages, lequel sera mis à disposition à la communauté scientifique et utilisé dans notre cadre d'étude. L'établissement de ce corpus va donc ouvrir de nombreuses pistes de recherche pluridisciplinaires (historiques, médicales, lexicographiques, littéraires, etc.). Comme deuxième contribution, nous proposons la mise en place d'une chaîne de traitement automatique consistant en l'océrisation, la correction automatique de sortie OCR, la structuration des données au format XML-TEI, la fouille sémantique et l'alignement des textes pour étudier le transfert du discours de Charcot aux différentes disciplines. De cette manière, nous établirons un nouveau modèle de l'édition numérique savante et exploitable, grâce à laquelle nous pourrons établir et explorer l'influence interdisciplinaire de Charcot envers ses contemporains et continuateurs. Au-delà des finalités de ce projet de thèse, ce modèle généralisable sera aussi applicable à d'autres projets de numérisation et de valorisation des fonds patrimoniaux.

7. Calendrier de recherche / Plan de travail

| Semestre | Axes d'approches | Démarche | Production écrite |
|----------|-------------------------|---|---|
| 1 | Travail préliminaire | <ul style="list-style-type: none"> Découverte du corpus Utilisation d'une interface simple (comme TXM) pour effectuer les premières observations | <ul style="list-style-type: none"> Premières observations sur le corpus, sa définition, analyses statistiques sur sa composition Méthodologie de travail de structuration XML-TEI |
| 2 | Préparation des données | <ul style="list-style-type: none"> Océrisation selon les besoins établis pour l'étude Structuration des données | <ul style="list-style-type: none"> Organisation d'un plan provisoire (amené à évoluer en fonction des résultats obtenus ultérieurement) Notes sur la bibliographie étudiée en parallèle |
| 3 | Analyses préliminaires | <ul style="list-style-type: none"> Fouille globale des textes pour identifier les éléments principaux et établir les outils d'analyse précis qui devront être utilisés Chaîne de traitement basée sur le traitement automatique des langues | <ul style="list-style-type: none"> Méthodologie des traitements à effectuer sur les textes et des outils à utiliser pour l'analyse approfondie |
| 4 | Analyses finales | <ul style="list-style-type: none"> Utilisation des outils de fouille de texte : modélisation de sujets, plongements des mots, analyses | <ul style="list-style-type: none"> Élaboration d'un plan détaillé Début de la rédaction |

| | | | |
|---|-------------|--|--|
| | | textométriques et statistiques sur les champs lexicaux | |
| 5 | Publication | Rédaction | |
| 6 | | Relectures de la thèse rédigée | |

Bibliographie indicative

Andrieu, B. (1999). *L'homme naturel: la fin promise des sciences humaines*. Presses Universitaires Lyon.

Bourneville, D. M. (1864). Socrate était-il fou? In *Journal de Médecine Mentale, T. IV, Masson & Fils, Paris, 1864.*, p. 209-22.

Bogousslavsky, J. (2011). Hysteria after Charcot: back to the future. In *Following charcot: A forgotten history of neurology and psychiatry* (Vol. 29, pp. 137-161). Karger Publishers.

Bogousslavsky, J. (2014). Jean-Martin Charcot and his legacy. In *Hysteria: The rise of an enigma* (Vol. 35, pp. 44-55). Karger Publishers.

Charcot, J.-M. (2015). *La foi qui guérit*. FV Éditions.

Charcot, J.-M. (2002). *Leçons du mardi à la Salpêtrière*. Avant-propos par J. Sédat. Bibliothèque des Introuvables, Paris.

Didi-Huberman, G. (1982). *Invention de l'hystérie*. Editions Macula, Paris.

Ewing, E. T. (2017). Will It Come Here? Using Digital Humanities Tools to Explore Medical Understanding during the Russian Flu Epidemic, 1889–90. *Medical history*, 61(3), 474-477.

Goetz, C. G., Bonduelle, M., & Gelfand, T. (1995). *Charcot: constructing neurology*. Oxford University Press on Demand.

Huber, L. (2016, April). A web tool for medical topic modeling: student research abstract. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing* (pp. 165-166).

Ji, Y. A., Nam, S. J., Kim, H. G., Lee, J., & Lee, S. K. (2018). Research topics and trends in medical education by social network analysis. *BMC medical education*, 18(1), 1-10.

Joyeux-Prunel, B. (2019). Visual Contagions, the Art Historian, and the Digital Strategies to Work on Them. *Artl@s Bulletin*, 8(3), 8.

de Lorde, A. (1908). *Une leçon à la Salpêtrière: tableau dramatique en deux actes*. Charpentier et Fasquelle.

Marquer, B. (2008). *Les romans de la Salpêtrière: réception d'une scénographie clinique: Jean-Martin Charcot dans l'imaginaire fin-de-siècle* (Vol. 438). Librairie Droz.

Marshall, J. W. (2016). *Performing Neurology: The Dramaturgy of Dr Jean-Martin Charcot*. Springer.

Massie, R. (2004). Charcot et Babinski: au-delà de la simple relation professeur-élève. *Canadian journal of neurological sciences*, 31(3), 422-426.

Micale, M. S. (2019). *Approaching hysteria: Disease and its interpretations*. Princeton University Press.

Moretti, F. (2013). *Distant reading*. Verso Books.

Nicolas, S. (2004). *L'hypnose: Charcot face à Bernheim-L'école de la Salpêtrière face à l'école de Nancy*. Editions L'Harmattan.

Oliveira, L., Teive, H. A., Marques, P., Germiniani, F., & de Paola, L. (2018). Jean-Martin Charcot's influence on Sigmund Freud's career. *Neurology* Apr 2018, 90 (15 Supplement) P5.309.

Pérez-Rincón, H. (2011). Pierre Janet, Sigmund Freud and Charcot's psychological and psychiatric legacy. In *Following Charcot: A forgotten history of neurology and psychiatry* (Vol. 29, pp. 115-124). Karger Publishers.

Riffaterre, M. (1980). *La trace de l'intertexte*. La pensée, 215, 4-18.

Schultheiss, K. (2021). The Internal Image: Mind and Brain in the Age of Charcot. *Modern Intellectual History*, 18(1), 23-46.

Schulz, S., Ševa, J., Rodriguez, S., Ostendorff, M., & Rehm, G. (2020). Named Entities in Medical Case Reports: Corpus and Experiments. *arXiv preprint arXiv:2003.13032*.

Si, Y., Wang, J., Xu, H., & Roberts, K. (2019). Enhancing clinical concept extraction with contextual embeddings. *Journal of the American Medical Informatics Association*, 26(11), 1297-1304.

Teive, H. A., Teive, G. M., Dallabrida, N., & Gutierrez, L. (2017). Alfred Binet: Charcot's pupil, a neuropsychologist and a pioneer in intelligence testing. *Arquivos de neuro-psiquiatria*, 75(9), 673-675.

Tréanton, J. R. (1997). Bonduelle M., Gelfand T., Goetz C.G. Charcot, un grand médecin dans son siècle. *Revue française de sociologie*, 38(2).

Underwood, T. (2017). A Genealogy of Distant Reading. *Digital Humanities Quarterly*, 11(2).