

# Mesurer l'impact de Jean-Martin Charcot *via* l'extraction des termes médicaux : quelle approche adopter ?

Ljudmila PETKOVIC<sup>1,2,3,4</sup>

prenom.nom@sorbonne-universite.fr

<sup>1</sup> Sorbonne Université, Faculté des Lettres, UFR Littératures françaises et comparée, ED III (ED019)  
<sup>2</sup> Sorbonne Université, Centre d'étude de la langue et des littératures françaises (CELLF), UMR 8599  
<sup>3</sup> Sorbonne Université, Observatoire des textes, des idées et des corpus (ObTIC)  
<sup>4</sup> Sorbonne Université, UFR Sociologie et Informatique pour les Sciences Humaines

Séminaire doctoral ObTIC  
SCAI, salle 1  
Paris, le 6 février 2025



- 1 Projet Charcot
- 2 Problématique
- 3 Méthodologie
- 4 Expériences
- 5 Conclusion

# Valorisation des archives de Jean-Martin Charcot

*Dans les petits papiers de Charcot : de l'expérimentation aux prémisses de la neurologie moderne*<sup>1</sup>

Thèse en cours  
initiative OPUS<sup>a</sup>

- dir. : Prof. Dr Glenn ROE
- co-enc. : Dr Motasem ALRAHABI

<sup>a</sup> <https://institut-opus.sorbonne-universite.fr/node/478>

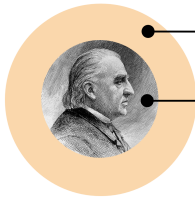


Fig. 1 – J.-M. Charcot (1825-1893), Wikipédia.

- père de la neurologie moderne
- hystérie, « Parkinson », SLA...
- Freud, de la Tourette, Babinski...

<sup>1</sup> <https://theses.fr/s382733>

- 1 Projet Charcot
- 2 Problématique
- 3 Méthodologie
- 4 Expériences
- 5 Conclusion



- Formaliser une approche numérique pour tracer l'évolution des concepts médicaux en général
- Pister numériquement la circulation du discours médical de Charcot

Est-il possible de mesurer l'impact de Charcot sur son réseau\* scientifique en s'appuyant sur les termes scientifiques qu'il a employés?

\* élèves, collègues et successeurs  
 → collaborateurs

Le texte comme point d'entrée pour étudier les tendances de la circulation des idées à l'aide des caractéristiques structurelles spécifiques (MILIA, 2023)

Certains **termes médicaux** dont Charcot a été l'inventeur (SLA) ou le transmetteur (*hystérie*) ont été repris de manière significative dans les écrits de son réseau scientifique.

Tendances de circulation linguistique dans les fronts de recherche

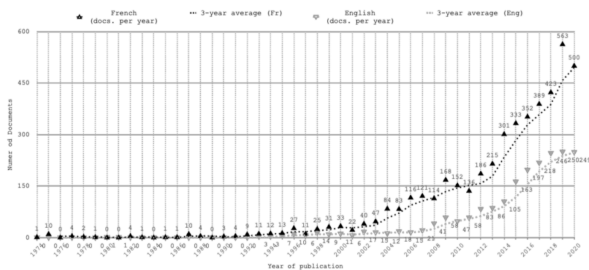


Fig. 2 – Nombre de documents représentant les intérêts thématiques en anglais et en français (MILIA, 2023).

Tendances de circulation linguistique dans les fronts de recherche

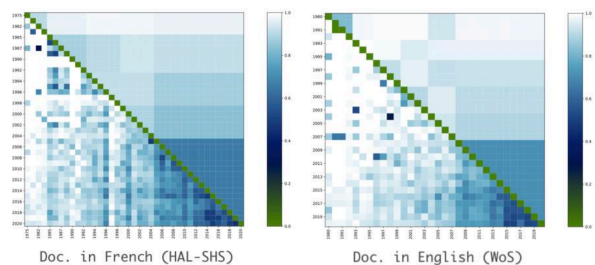


Fig. 3 – Détection de période à l'aide d'une corrélation significative des termes au fil des années dans deux sources et langues différentes (MILIA, 2023).

- réception de la pensée scientifique de C. Bernard (RIGUET, 2018)
- détection des réemplois textuels (FEDCHENKO et al., 2024)
- Rankingdom – mesurer l'importance d'une entité (SOULET, 2024)

[https://github.com/ljpetkovic/Charcot\\_circulations/tree/main/concepts](https://github.com/ljpetkovic/Charcot_circulations/tree/main/concepts)

<p>  
<s>Hystérie (V. Epidémie, Hémianes-thésie, Hyperesthésie ovarienne,  
<p>  
<s>Hystéro-épilepsie, 332, 367. — Si-gnification de ce mot, 368 ; —  
<p>

**Fig. 7**—Concepts médicaux, document XML.

Fig. 6 – Index des termes (CHARCOT, 1892).

hystérie(s)?
hystérie(s)? épileptiforme(s)?
hystérie(s)? ovarienne(s)?
hystérie(s)? grave(s)?
hystérie(s)? locale(s)?
hystérie(s)? infantile(s)?
hystérie(s)? locale(s)? traumatique(s)?
hystéro-épilepsie(s)?

- 1 entre <s> et , - (regex)
- 2 sans termes génériques (*os*, *peau*)
- 3 prise en compte des sg. / pl. (regex)

Fig. 8 – Liste finale des concepts médicaux.

- système à base de règles + mesures statistiques
  - TF-IDF, spécificité, fréquences...
  - 261 termes en commun

Corpus Charcot				Corpus Autres			
Motif POS	Effectif	Freq. relat. (%)	Exemple	Effectif	Freq. relat. (%)	Exemple	
N	261	52.10	<i>hystérie</i>	271	54.20	<i>somnambule</i>	
A	151	30.14	<i>cérébral</i>	149	29.80	<i>hypnotique</i>	
N A	73	14.57	<i>système nerveux</i>	73	14.60	<i>lame médullaire</i>	
N P N	12	2.40	<i>cas de folie</i>	6	1.20	<i>scissure de sylvius</i>	
N A A	3	0.60	<i>système nerveux central</i>	0	0.00	—	
R	1	0.20	<i>[d']emblée</i>	1	0.20	<i>obliquement</i>	
Total	501	100.00		500	100.00		

Tab. 2 – Répartition des parties du discours constituant les termes médicaux dans les deux corpus.

3. <https://termsuite.github.io/>

4. [https://github.com/ljpetkovic/Charcot\\_TermSuite/tree/main](https://github.com/ljpetkovic/Charcot_TermSuite/tree/main)

- unigrammes de noms [N] et des adjectifs [A] : les plus fréquents
- trigrammes : les séquences les plus longues extraites
  - aucune occurrence du trigramme  *sclérose latérale amyotrophique*
  - *seule trace* : adjectif [A]  *latérale*
  - *idem* pour le quadrigramme  *état de mal hystéro-épileptique*

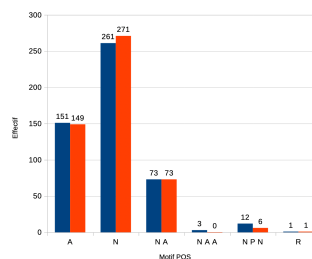


Fig. 9—Analyse comparative des séquences syntaxiques.

Premiers 12 termes communs, à partir du TF-IDF  $\approx 0,15$

- prévalence des unigrammes

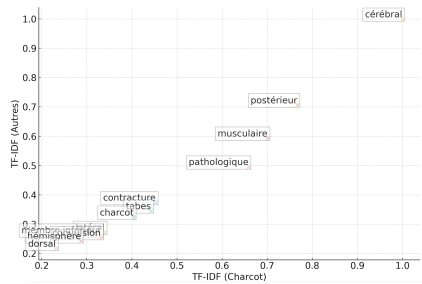


Fig. 10 – Termes communs avec leurs valeurs TF-IDF respectives.

- 1 Projet Charcot
- 2 Problématique
- 3 Méthodologie
- 4 Expériences
- 5 Conclusion

Il est possible de quantifier la pertinence des termes scientifiques partagés entre Charcot et son réseau

⚠ Limitations de l'approche linguistique de l'extraction des termes  
→ les termes plus pointus sont pénalisés

À faire :

- rendre compte du contexte d'énonciation des termes
- déterminer l'approche à adopter (plongements dynamiques?)
- trouver un corpus « externe » : prouver la pertinence des termes détectés pour la communauté scientifique

CHARCOT, J. M. (1892). *Œuvres complètes de J. M. Charcot. Leçons sur les maladies du système nerveux. T. 1*. <https://patrimoine.sorbonne-universite.fr/viewer/3468/?offset=1#page=2&viewer=picture&o=&n=0&q=> Bureaux du progrès medical (voir p. 13).

FEDCHENKO, V., D. M. NICOLSI et G. ROE (2024). À la recherche des réseaux intertextuels : défis de la recherche littéraire à grande échelle. In : *Humanités numériques* 9. <https://journals.openedition.org/revuehn/3940> (voir p. 9).

MILIA, M. (2023). Using Digital Text-Based Approaches to Study Knowledge Circulation. In : *Routledge Handbook of Academic Knowledge Circulation*. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003290650-52/using-digital-text-based-approaches-study-knowledge-circulation-mat%C3%ADas-milia>. Routledge, p. 556-574 (voir pp. 6-8).

RICUET, M. (2018). L'impact de la physiologie dans la critique littéraire de la fin du XIX<sup>ème</sup> siècle : l'exemple de Claude Bernard. In : *Epistémocritique : Littérature et savoirs*. <https://hal.science/hal-01903871> (voir p. 9).

SOULET, A. (2024). *Vers l'analyse à la demande des connaissances de Wikidata*. [https://afia.asso.fr/wp-content/uploads/2024/05/SOULET\\_IA-et-HN-2024-05-03.pdf](https://afia.asso.fr/wp-content/uploads/2024/05/SOULET_IA-et-HN-2024-05-03.pdf). Consulté le 2 mars 2025 (voir p. 9).