

Circulation du discours médical de Jean-Martin Charcot : premières observations

Ljudmila Petkovic, Motasem Alrahabi, Glenn Roe

Sorbonne Université, équipe-projet ObTIC

{prenom.nom}@sorbonne-universite.fr

Résumé

Cette proposition de communication se veut une première réflexion liée à la mesure du degré d'intertextualité entre Jean-Martin Charcot (1825-1893) et son réseau scientifique et artistique au prisme des humanités numériques. L'approche consiste à comparer les textes écrits par Charcot avec ceux de ses collaborateurs et successeurs, *via* les concepts-clés liés à son discours scientifique. Les expériences menées sur la circulation des concepts d'un corpus à l'autre se base sur les mesures TF-IDF, BM-25 et BERT, et les résultats obtenus semblent probants. Au-delà du cas de Charcot, ce travail vise à établir un protocole permettant d'appréhender la circulation de concepts de manière automatisée.

1 Introduction

Depuis les premières descriptions du cerveau humain faites de manière rigoureuse par Constanzo Varolio (1543-1575) en 1573 (Kneib, 2011)¹, suivies par la création du terme *neurologia* par Thomas Willis (1621-1675)², l'histoire de la neurologie trouve son ancrage au XIX^e siècle dans les travaux de Jean-Martin Charcot (1825-1893), considéré comme le père de la neurologie française et moderne (Teive et al., 2022; Broussolle et al., 2012).

Figure emblématique du célèbre hôpital de la Salpêtrière à Paris, Charcot est essentiellement connu pour ses études sur les troubles névrotiques, notamment sur l'hystérie, l'hypnose, la double personnalité, la catalepsie et le somnambulisme³. Ses idées novatrices ont eu un grand retentissement parmi ses collaborateurs : Pierre Janet (1859-1947), Joseph Babinski (1857-1932), Gilles de La Tourette (1857-1904), Paul Richer (1849-1933), Dé-

1. Il s'agit des descriptions de la structure cérébrale, appelée *pont* (lat. *pons*) par Varolio (1573), puis *pont de Varole* en l'honneur du célèbre anatomiste.

2. Cf. Willis (1664).

3. Ces explorations des abîmes de l'esprit humain lui ont valu l'appellation de « Napoléon des névroses » (Ellenberger, 1970).

siré Magloire Bourneville (1840-1909), pour n'en nommer que quelques-uns (Bogousslavsky, 2011), mais aussi parmi ceux qui lui étaient radicalement opposés, comme Hippolyte Bernheim (1840-1919) et l'école de Nancy pendant les années 1880-1890⁴. L'impact colossal de Charcot sur sa propre discipline est bien connu : mentionnons seulement le changement d'intérêt radical de Sigmund Freud (1856-1939), passant de la neurologie générale à l'hystérie, l'hypnose et d'autres troubles psychologiques, ce qui a donné lieu au développement de la théorie psychanalytique, à la suite de son séjour dans le service de neurologie de Charcot pendant l'hiver 1885-1886 (Camargo et al., 2018).

Nous visons à mesurer informatiquement l'impact de Charcot sur son réseau scientifique et artistique⁵. Cette mesure se fonde sur l'analyse de ses concepts-clés en matière de discours scientifique, et plus particulièrement sur l'opérationnalisation du terme « influence », définie ici comme une intertextualité uni-directionnelle, allant des écrits de Charcot vers ceux de ses collaborateurs et successeurs (ci-après corpus « Autres »). Il s'agit donc *in fine* d'aborder computationnellement la question des circulations, non pas des artefacts matériels comme les manuscrits (Gabay et al., 2021) et les images (Joyeux-Prunel, 2019), mais des phénomènes textuels complexes (Manjavacas et al., 2019) ayant une dimension théorique forte.

2 Méthodologie

2.1 Outils existants

Le corpus de travail est constitué de 201 documents OCRisés (sans post-correction), fournis

4. Contrairement à la définition de Charcot, qui voyait en l'hypnose un état pathologique propre aux hystériques, Bernheim (1891) définit l'hypnose comme un simple sommeil produit par la suggestion et susceptible d'applications thérapeutiques.

5. Par ailleurs, le présent travail fait partie du projet doctoral en cours <https://obtic.sorbonne-universite.fr/projet/charcot>.



FIGURE 1 – Distribution des fréquences des tokens avec la frise chronologique pour ceux constituant l’expression *bulbe rachidien* (issus du corpus « Charcot » et du corpus « Autres ») dans l’application OBVIE.

au format XML par la Bibliothèque de Sorbonne Université (BSU)⁶. Nous avons procédé, dans un premier temps, à une restructuration des textes en XML-TEI, afin de permettre la fouille avancée du corpus Charcot à travers des outils développés au sein de notre équipe. D’une part, notre moteur de recherche OBVIE permet de repérer des textes similaires par ordre de pertinence à partir des termes en commun⁷. D’autre part, l’algorithme TextPair génère une liste de passages similaires, c’est-à-dire les séquences de mots qui se chevauchent (n-grammes de mots) pour chaque texte, en comparant ensuite ces résultats avec ceux de séquences dans d’autres textes⁸.

Afin de mesurer l’impact de Charcot sur son entourage et d’analyser la circulation de concepts véhiculés dans le corpus, nous avons commencé par séparer les textes rédigés par Charcot de ceux rédigés par ses co-auteurs (p. ex. Bourneville) ou les auteurs thématiquement proches de lui (p. ex. de la Tourette). Nous avons obtenu respectivement

68 (corpus « Charcot ») et 133 (corpus « Autres ») textes⁹. Une première exploration du corpus Charcot à travers l’application OBVIE nous a permis d’identifier les substantifs les plus importants de chaque corpus en utilisant les fréquences brutes ou des méthodes plus fines comme TF-IDF, BM25, χ^2 ou le Test Gamma. Cependant, l’application ne permet pas de quantifier la pertinence des expressions polylexicales (i.e. les n-grammes de mots), très fréquentes dans les deux corpus et dont la décomposition entraînerait une perte d’information (p. ex. le terme polysémique « bulbe » qui a une valeur spécifique dans l’expression figée *bulbe rachidien*, cf. fig. 1).

Concernant l’alignement des séquences similaires aux deux corpus, TextPair nous a permis, par une lecture attentive, de faire des comparaisons entre les textes et de rechercher des termes au sein des passages similaires, malgré le nombre de résultats assez conséquent (cf. fig. 2).

Les deux corpus issus du fonds Charcot sont librement disponibles et interrogeables sur les deux plateformes OBVIE¹⁰ et TextPair¹¹.

6. Ces archives sont disponibles sur le portail numérique SorbonNum (anc. Bibliothèque numérique patrimoniale de l’université Pierre et Marie Curie – UPMC) qui représente le point d’entrée unique vers les collections scientifiques patrimoniales et numériques de Sorbonne Université : <https://patrimoine.sorbonne-universite.fr/>.

7. <https://obtic.huma-num.fr/obvie/>. Pour d’amples informations sur le fonctionnement de cet outil, cf. Alrahabi (2022).

8. <https://artfl-project.uchicago.edu/text-pair/>.

9. Le corpus contient 31 979 479 tokens au total, dont 12 190 649 (38,12%) dans le corpus « Charcot » et 19 788 830 (61,88%) dans le corpus « Autres ».

10. <https://obtic.huma-num.fr/obvie/charcot/?view=corpus>

11. <https://anomander.uchicago.edu/text-pair/charcot2autres/>

Source	Target	Browse by Metadata Counts
Charcot, Jean-Martin • Archives de neurologie [Tome 26, n° 77-82] : revue des maladies nerveuses et mentales •	Gilles de la Tourette, Georges • Nouvelle iconographie de la Salpêtrière [Tome 23] : iconographie médicale et artistique •	Source
nouveaux cas de sclérose latérale amyotrophique suivis d'autopsie (en collaboration avec Marie), 1885 ; De l'Ozzonatomazie (en collaboration avec Magnan), 188 ? - Deux nouveaux cas de sclérose latérale amyotrophique suivis d'autopsie (en collaboration avec Marie), 1885 ; - Rapport médico-légal sur Annette G... (en collaboration avec Brouardel et Mottet), 1880 ; - Rapport présenté à M. le Ministre de	rale amyotrophique, dans lesquels ils ont noté l'atrophie et la disparition des cellules de Betz ; ils s'en ont servi pour délimiter la zone (1) CHARCOT et Marie. Deux nouveaux cas de sclérose latérale amyotrophique suivis d'autopsie . Arch. de Neurologie, 1885, nos 28-29. (2) F. Leninmalm. Bidrag till Kannedomen om den amyotrofiska lateralsklerosen., Upsala läkareförening, 1887, n° 7. Analysé in Neurol. Centralbl, 1881, p. 550.	Passage
		Author
		Title
		Year
		Passage Length
		Target
		Passage
		Author
		Title

FIGURE 2 – Alignement et comparaison des textes de Charcot à celui de Georges Gilles de la Tourette (le seul résultat) en lançant la requête *sclérose latérale amyotrophique*.

2.2 Une nouvelle approche

Afin de surmonter les limites rencontrées avec ces deux outils, nous proposons une nouvelle méthode pour identifier des concepts dans les deux corpus en nous basant sur le poids de leur apparition, calculé selon trois différentes mesures de pondération¹² :

- TF-IDF est une méthode qui permet d'évaluer l'importance d'un terme contenu dans un document relativement à un corpus plus large en récompensant la fréquence des termes et en pénalisant la fréquence des documents.
- BM25 est une tentative d'amélioration de TF-IDF, notamment pour traiter les documents de grande longueur et les problèmes engendrés par la possible saturation des termes (Robertson et Sparck Jones, 1988).
- BERT (Devlin et al., 2018) est un modèle pré-entraîné qui utilise l'apprentissage non-supervisé sur de grandes quantités de données textuelles pour apprendre des représentations de mots et de phrases, et comprendre le contexte et la sémantique. Il est basé sur l'architecture des transformeurs, qui est un type de réseau de neurones utilisé pour le traitement du langage naturel.

La liste des concepts retenus pour l'étude est composée de termes ou expressions popularisés par Charcot (Walusinski et Bogousslavsky, s.d.), comme *hystérie*, *sclérose latérale* etc.¹³. Elle provient du croisement entre la liste des termes obtenus avec OBVIE et l'index d'une édition des œuvres

12. Le code est disponible en ligne : https://github.com/ljpetkovic/Charcot_circulations.

13. Pour chaque entrée, nous avons pris en compte les formes du singulier et du pluriel obtenues grâce à des expressions régulières.

complètes de Charcot (1892), dont nous avons retiré les termes génériques (*os*, *cerveau*, etc.).

3 Résultats

Comme nous pouvons l'observer sur la fig. 3, la mesure BM25 révèle une intensification du lexique de Charcot dans le corpus « Autres », les scores étant plus élevés pour 14 termes (sur 14 évalués) utilisés par le réseau de Charcot. D'ailleurs, d'après le tableau 1, c'est la seule mesure dont les valeurs témoignent clairement d'un lexique partagé entre Charcot et ses successeurs et collaborateurs, *a contrario* des deux autres mesures, où le rapport en question est inversé (la grande majorité des termes étant plus pertinente dans le discours de Charcot, et son impact étant donc moins accentué). Concrètement, les termes les plus pertinents semblent être *sclérose en plaque disséminées* (score 0,83), *paralysie rhumatismale* (0,68), *atrophie progressive* (0,53) et *arthrite déformante* (0,50).

D'autre part, nous avons utilisé BERT pour mesurer le poids des termes dans les deux corpus. Bien que ce type de modèle ne fournisse pas directement de poids pour les mots, nous pourrions cependant en extraire des informations utiles pour estimer l'importance ou le poids des mots dans les textes. Différentes approches sont généralement utilisées pour obtenir une représentation de l'importance des mots, en exploitant les informations des plongements lexicaux et des mécanismes d'attention (Vaswani et al., 2017). Pour ce travail en cours, nous avons utilisé le modèle `bert-base-multilingual-cased`. Les premiers résultats obtenus se trouvent dans le tableau 1 et restent à améliorer. Cependant, nous avons observé que les termes les plus pertinents pour le discours de Charcot étaient ceux qui désignent les noms des différentes pathologies (*diplopie*, *myélite partielle*, *état de mal épileptique*, *paralysie labio-glosso-laryngée* etc.), contrairement à d'autres notions plus abstraites (*préambule*, *délire*, *miracle*) qui sont prédominantes dans le corpus « Autres » (termes non renseignés dans le tableau en question).

4 Discussion et perspectives

Ce travail constitue la première phase d'exploration du corpus de Charcot. Les deux outils, OBVIE et TextPair, nous offrent des fonctionnalités avancées de recherche et de comparaison de textes dans le cadre d'une analyse de textes assistée par ordina-

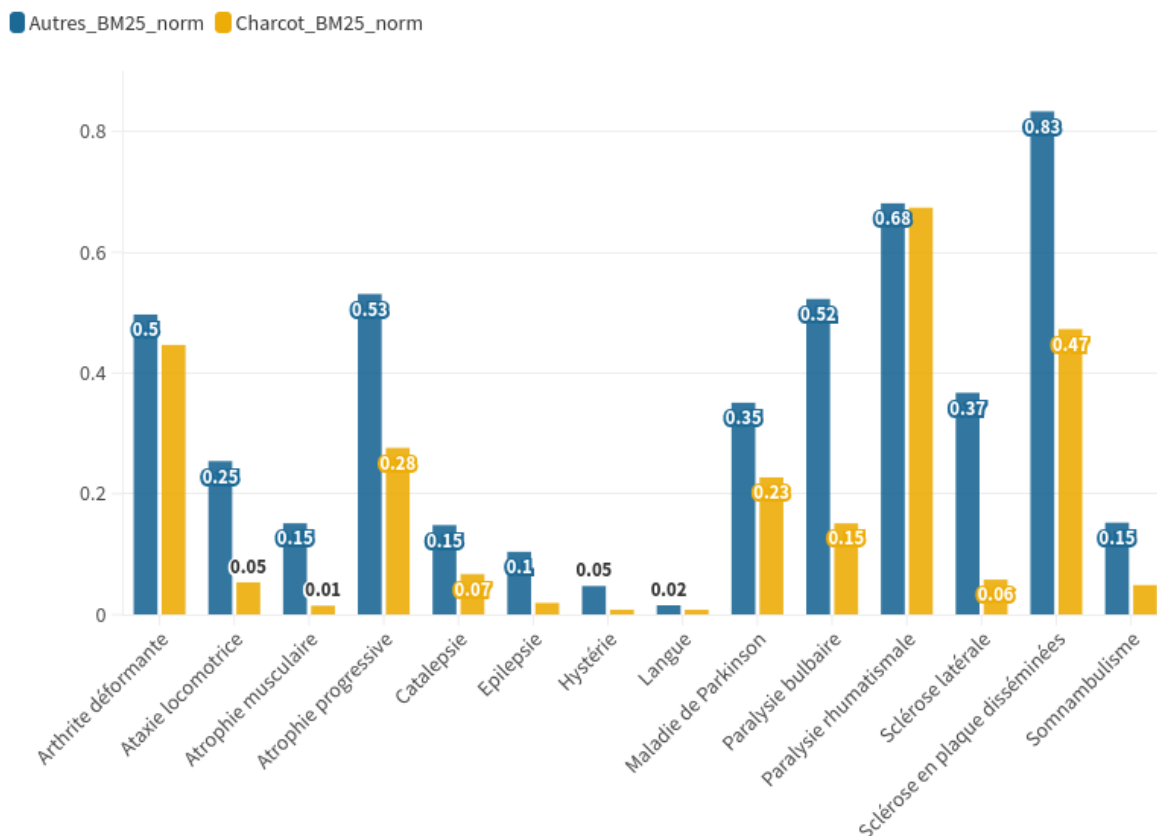


FIGURE 3 – Visualisation de pertinence des concepts dans les deux corpus suivant la mesure BM25. Les valeurs des concepts associées au corpus « Charcot » sont représentées en bleu, alors que celles du corpus « Autres » en jaune.

teur. Cependant, ils ne proposent pas de fonctionnalité de lecture distante permettant de rendre compte de l'impact de Charcot sur son réseau scientifique et artistique à travers les concepts principaux de ses travaux. L'analyse effectuée à l'aide d'un nouveau script nous a alors permis de quantifier les concepts polylexicaux dans les deux corpus, selon trois différentes métriques de pondération. La visualisation des résultats nous a permis d'observer des phénomènes qu'il serait nécessaire de valider auprès de spécialistes de Charcot.

Pour la suite, trois pistes de recherche devraient être suivies : 1. comparer l'œuvre de Charcot avec la production qui le précède afin de vérifier s'il est vraiment l'initiateur de certains termes ou bien s'il en est simplement le transmetteur ; 2. opérer une analyse sémantique des passages qui contiendraient nos concepts, afin d'étudier les différentes modalités de prise en charge énonciative : opinions, accords, désaccords, définitions, etc. (Alexandre et al., 2022). En effet, reprendre un terme ne veut pas dire y adhérer : on peut citer pour dire que l'on n'est pas d'accord ; 3. améliorer le texte issu de

l'OCR et évaluer l'impact de la correction orthographique de notre corpus sur ces résultats, à l'aide d'une approche à base d'apprentissage profond. (Petkovic et al., 2023).

5 Remerciements

Nous remercions vivement Simon Gabay et Valentina Fedchenko pour leur relecture et leurs conseils précieux.

Bibliographie

- Didier Alexandre, Motasem Alrahabi, Marine Riguet, Alexandre Duguet, Marie-Christine Renaud, et Frédéric Gay. 2022. *Le médical et le social. Analyse sémantique des rapports de l'immersion d'étudiants en médecine dans le Samu social. Cultures et pratiques savantes du numérique*, pages 817–846.
- Motasem Alrahabi. 2022. *Obvie: interface web pour la fouille et la comparaison de textes*. In *Atelier Digital Humanities and cultural heritage: data and knowledge management and analysis during the conference francophone sur l'Extraction et la Gestion des Connaissances (egc2022)*, Paris, France.
- Hippolyte Bernheim. 1891. *De la suggestion: et de ses*

- applications à la thérapeutique : (3e édition corrigée et augmentée)*. Octave Doin, Paris.
- Julien Bogousslavsky. 2011. *Hysteria after Charcot: Back to the Future*. In *Following Charcot : A Forgotten History of Neurology and Psychiatry*, volume 29, pages 137–161. Karger Publishers.
- Emmanuel Broussolle, Jacques Poirier, François Clarrac, et Jean-Gaël Barbara. 2012. *Figures and institutions of the neurological sciences in Paris from 1800 to 1950. Part III: Neurology*. *Revue Neurologique*, 168(4) :301–320.
- Carlos Henrique Camargo, Paula Teixeira Marques, Lívia Oliveira, Francisco Germiniani, Luciano de Paola, et Helio Teive. 2018. *Jean-Martin Charcot’s influence on Career of Sigmund Freud, and the Influence of this Meeting for the Brazilian Medicine*. *Revista Brasileira de Neurologia*, 54(2) :40–46.
- Jean-Martin Charcot. 1892. *Œuvres complètes de J.-M. Charcot : Leçons sur les maladies du système nerveux*, volume 1. Bureaux du Progrès médical, Paris.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, et Kristina Toutanova. 2018. *BERT: pre-training of deep bidirectional transformers for language understanding*. *CoRR*, abs/1810.04805.
- Henri F. Ellenberger. 1970. *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry*, volume 1. Basic books, New York.
- Simon Gabay, Ljudmila Petkovic, Alexandre Bartz, Matthias Gille Levenson, et Lucie Rondeau Du Noyer. 2021. *Katabase: À la recherche des manuscrits vendus*. In *Humanistica 2021*, Rennes, France.
- Béatrice Joyeux-Prunel. 2019. *Visual Contagions, the Art Historian, and the Digital Strategies to Work on Them*. *Artl@s Bulletin*, 8(3) :8.
- Marie Kneib. 2011. *Étude fonctionnelle d’un circuit inhibiteur du cortex cérébelleux de la souris: Importance pour la physiopathologie des retards mentaux*. Thèse de doctorat, Strasbourg.
- Enrique Manjavacas, Brian Long, et Mike Kestemont. 2019. *On the feasibility of automated detection of allusive text reuse*. In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 104–114, Minneapolis, USA. Association for Computational Linguistics.
- Ljudmila Petkovic, Motasem Alrahabi, et Glenn Roe. 2023. *Impact de la correction automatique de l’OCR/HTR sur la reconnaissance d’entités nommées dans un corpus bruité*. *Journal of Information Sciences*, 21(2).
- Stephen E. Robertson et Karen Sparck Jones. 1988. *Relevance Weighting of Search Terms*, page 143–160. Taylor Graham Publishing, GBR.
- Hélio Afonso Ghizoni Teive, Léo Coutinho, Carlos Henrique Ferreira Camargo, Renato Puppi Munhoz, et Olivier Walusinski. 2022. *Thomas Willis’ legacy on the 400th anniversary of his birth*. *Arquivos de Neuro-Psiquiatria*, 80 :759–762.
- Constanzo Varolio. 1573. *De nervis opticis nonnullisq; aliis praeter communem opinionem in humano capite obseruatis*. Patavii : apud P. et A. Meiettos fratres.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, et Illia Polosukhin. 2017. *Attention Is All You Need*. *CoRR*, abs/1706.03762.
- Olivier Walusinski et Julien Bogousslavsky. s.d. *Jean-Martin Charcot, Pierre Janet et leurs conceptions psychopathologiques*. Consulté sur le site <http://baillement.com/>.
- Thomas Willis. 1664. *Cerebri anatome: cui accessit nervorum descriptio et usus*. Londini : Typis Ja. Fleisher, impensis Jo. Martyn & Ja. Allestry, apud insigne Campanæ in Cœmeterio, D. Pauli.

Terme	Charcot				Autres			
	Fréquence	TF-IDF	BM25	BERT	Fréquence	TF-IDF	BM25	BERT
Arthrite déformante	30	0,16	0,45	0,80	24	0,02	0,50	0,40
Ataxie locomotrice	559	0,35	0,05	0,83	169	0,08	0,25	0,39
Atrophie musculaire	1105	0,20	0,02	0,84	1465	0,43	0,15	0,42
Atrophie progressive	40	0,14	0,27	0,72	22	0,02	0,53	0,39
Catalepsie	681	0,54	0,07	0,88	975	0,28	0,15	0,39
Épilepsie	414	0,09	0,02	0,78	577	0,12	0,10	0,41
Hystérie	5775	0,51	0,01	0,74	4934	0,45	0,05	0,41
Langue	2695	0,24	0,01	0,72	3591	0,11	0,02	0,41
Maladie de Parkinson	75	0,21	0,23	0,81	130	0,09	0,35	0,37
Paralysie bulbaire	149	0,27	0,15	0,89	93	0,09	0,52	0,40
Paralysie rhumatismale	8	0,07	0,67	0,86	14	0,02	0,68	0,44
Sclérose latérale	445	0,30	0,06	0,88	127	0,09	0,37	0,41
Sclérose en plaque disséminées	45	0,25	0,47	0,87	12	0,02	0,83	0,40
Somnambulisme	847	0,49	0,05	0,89	3410	1	0,15	0,43

TABLEAU 1 – Calcul de pertinence des concepts selon les mesures TF-IDF, BM25 et BERT dans les corpus « Charcot » et « Autres » (les meilleurs scores sont affichés en gras).