

Mesurer l'influence de Jean-Martin Charcot sur ses contemporains à l'aide de l'extraction des phrases-clés

Ljudmila PETKOVIC^{1,2,3}

`prenom.nom@sorbonne-universite.fr`

¹ Sorbonne Université, Faculté des Lettres, UFR Littératures françaises et comparée, ED 3

² Centre d'étude de la langue et des littératures françaises (CELLF), UMR 8599

³ Observatoire des textes, des idées et des corpus (ObTIC)

Journée « IA et Humanités Numériques »

BNF, salle 70

Paris, le 3 mai 2024



Plan

1. Contexte de recherche
2. Problématique et objectif
3. Approche supervisée
4. Approche non supervisée
5. Conclusion et recherches futures

1. Contexte de recherche

2. Problématique et objectif

3. Approche supervisée

4. Approche non supervisée

5. Conclusion et recherches futures

« Napoléon des névroses » ou « Paganini de l'hystérie » (MARMION, 2015)



Source : [Wikipedia](#).

JEAN-MARTIN CHARCOT (1825-1893)

- père de la neurologie moderne en France au XIX^e s.
- leçons cliniques du mardi à l'hôpital de la Salpêtrière à Paris
« Mecque de la neurologie »

● Contributions majeures :

hystérie

hypnose

SEP

SLA

maladie de Parkinson

← lésion dynamique des circuits cérébraux

analyse et traitement des symptômes hystériques

description de la *sclérose en plaques* disséminée¹

description de la *sclérose latérale amyotrophique*²

concepteur du terme (avec Alfred Vulpian)

(CAMARGO et al., 2024)

1. ou *sclérose multiple*.

2. *maladie de Charcot* ou *maladie Lou-Gehrig*.

Impact de Charcot sur sa discipline et au-delà

(Quelques) collaborateurs et élèves

« réseau scientifique »

Sigmund FREUD (1856-1939)

théorie psychanalytique

Gilles DE LA TOURETTE (1857-1932)

syndrome de Tourette

Joseph BABINSKI (1857-1904)

pithiatisme, signe de Babinski

(BROUSSOLLE et al., 2012)

(Quelques) écrivains naturalistes français et européens

- références à Charcot et aux descriptions de crises hystériques

Émile ZOLA (1840–1902)

Lourdes

Léon TOLSTOÏ (1828–1910)

La Sonate à Kreutzer

Luigi CAPUANA (1839–1915)

La Torture

(KOEHLER, 2013)

1. Contexte de recherche

2. Problématique et objectif

3. Approche supervisée

4. Approche non supervisée

5. Conclusion et recherches futures

Circulation du discours médical au prisme du numérique

Objectif : aborder computationnellement la question des circulations des phénomènes textuels complexes.

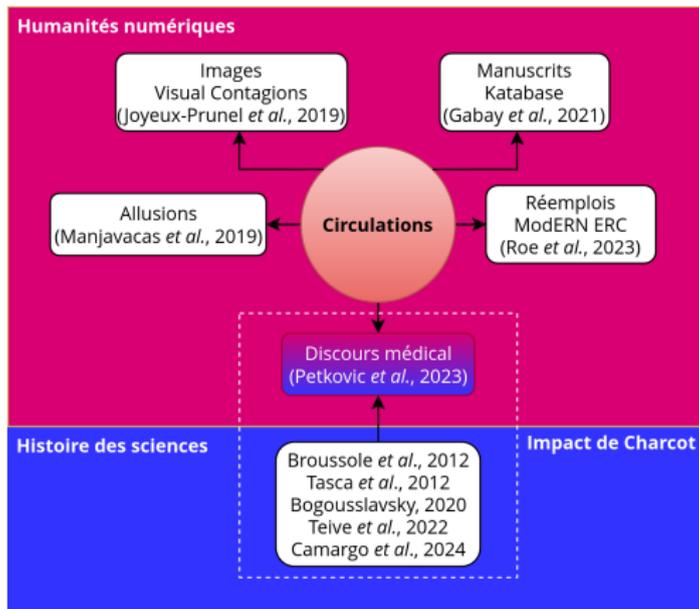


Fig. 1 – Études (numériques) des circulations des savoirs.

Question de recherche

Comment mesurer le degré d'intertextualité entre le discours de Charcot et celui de son réseau scientifique au prisme du numérique?

1. Contexte de recherche

2. Problématique et objectif

3. Approche supervisée

4. Approche non supervisée

5. Conclusion et recherches futures

Fonds Charcot en ligne

SorbonNum

Bibliothèque de Sorbonne Université (BSU)

201 documents XML OCRisés (sans post-correction)

Corpus	Nb de docs	Nb de tokens
Charcot textes rédigés par Charcot	68	12 190 649 (38,12%)
Autres textes rédigés par les membres de son réseau scientifique	133	19 788 830 (61,88%)
Total	201	31 979 479 (100%)

Tab. 1 – Répartition du fonds Charcot selon les auteurs.

Corpus Charcot en ligne

Corpus Charcot accessible sur la plateforme OBVIE (ALRAHABI, 2022)

- fouille avancée des corpus en XML-TEI
- textes similaires : mots fréquents / en commun, noms cités

The screenshot displays the OBVIE platform interface for comparing two corpora. At the top, there is a search bar and navigation tabs: CORPUS, NUAGE, RÉSEAU, FRÉQUENCES, EXTRAITS, CONCORDANCE, COMPARER, and AIDE. The selected view is 'COMPARER', showing two corpora side-by-side:

- Left Corpus:** CHARCOT, Jean-Martin : BOURNEVILLE, Désiré Magloire. (1881) *Archives de*
- Right Corpus:** GILLES DE LA TOURETTE, Georges. (1901) *Nouvelle iconographie de la Salpêtrière*

Under each corpus, there are three sections:

- Mots fréquents:** Lists common words and their occurrences in each text. For Charcot, words include 'revue, métal, température, gaine, cérébral, pathologie, nerf, accès, myéline, délire, cylindre, circonvolution, auteur, segment, névrite, asile, anesthésie, paralysie, if, observation, dégénération, application, nerveux, muscle, mental, anatomie, sensibilité, restauration, épileptique, portion, hémisphère, lésion, planche, mentir, pathologique, frontal, concours, démence, autopsie, axe, thermomètre, côté, épilepsie, dé, mince, contracture, amidon, sain, ecchymose, action'.
- Mots en commun:** Lists words shared by both corpora, such as 'malade, cas, partie, gauche, côté, lésion, membre, nerveux, moelle, présenter, observation, état, droit, inférieur, cérébral, trouver, fibre, paralysie, main, nerf, normal, postérieur, muscle, cellule, trouble, antérieur, corps, travail, face, donner, sembler, auteur, maladie, droite, forme, mouvement, région, supérieur, substance, constater, revue, cordon, surtout, exister, enfant, tumeur, sujet, mentir, interne'.
- Noms cités:** Lists authors mentioned in the texts. For Charcot, names include 'J, T, B, Burq, H. d' O., T. R., W, PL, P, Schwann, H. de B., T., U, M. Blaise, Chauvet, H, PÉRI-AXILE, MÉTALLOSCOPE, N, J., R, PATHOGÉNIE un tremblement, Maragliano, Arnozan, Seguin, Wrissberg, J, Landouzy, JI, JI, Laffont, MÉTALLOTHÉRAPIE, Charcot, Bogdanow, Schiff, Burman, Vulpian, Despine, Treub, Archiv, F, M. Russell, Beard, J9, Ir, Ranvier, Journ, ÆSTHÉSIOGÈNES, PERI-AXILE, Lasègue'.

At the bottom left, there is an 'ARCHIVES' button.

Fig 2 – Points similaires entre un ouvrage de Charcot et celui de de la Tourette.

Mesurer le degré d'intertextualité

Mesurer informatiquement l'impact de Charcot sur son réseau
→ intertextualité uni-directionnelle

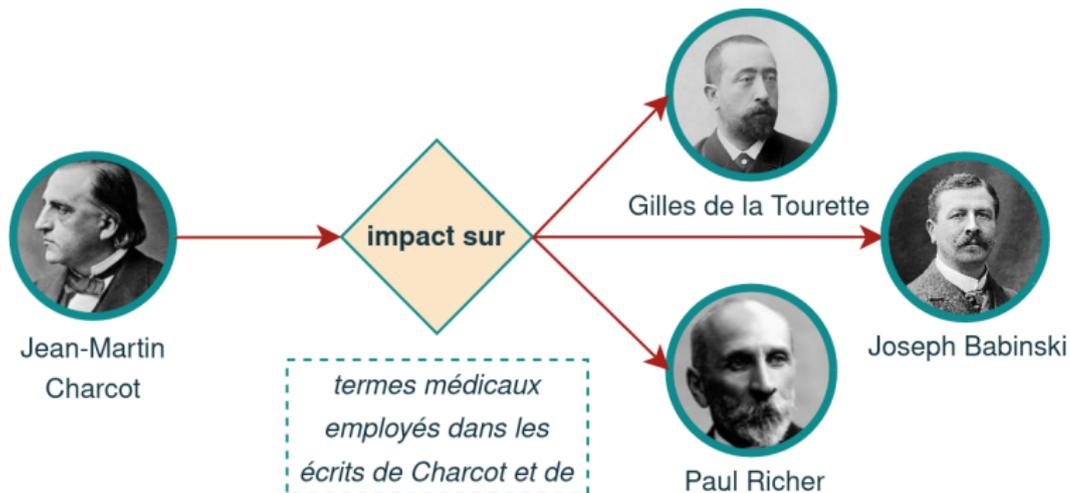


Fig. 3 – Opérationnalisation de l'impact de Charcot sur ses élèves.

Liste des concepts médicaux

Extraction semi-automatique des termes en lien avec Charcot.

HYSTÉRIE (V. ÉPIDÉMIE, HÉMIANESTHÉSIE, HYPERESTHÉSIE OVARIENNE, ISCHURIE, SECOURS); — *épileptiforme*, 369; — *ovarienne*, 302; — *grave*, 306, 383; — *locale*, 320. — *infantile*, 451. — *locale traumatique*, 450.

HYSTÉRO-ÉPILEPSIE, 332, 367. — Signification de ce mot, 368; — à crises distinctes, 371. — Variétés de l' —, 370. — Nature de l' —, 373. — Température dans l' —,

```
<p>
<s>Hystérie (V. Epidémie, Hémianesthésie, Hyperesthésie ovarienne,
</p>
<p>
<s>Hystéro-épilepsie, 332, 367. — Signification de ce mot, 368; —
</s>
```

Fig. 5 – Concepts médicaux, document XML.

Fig. 4 – Index des termes (CHARCOT, 1892).

<u>hystérie(s)?</u>
<u>hystérie(s)? épileptiforme(s)?</u>
<u>hystérie(s)? ovarienne(s)?</u>
<u>hystérie(s)? grave(s)?</u>
<u>hystérie(s)? locale(s)?</u>
<u>hystérie(s)? infantile(s)?</u>
<u>hystérie(s)? locale(s)? traumatique(s)?</u>
<u>hystéro-épilepsie(s)?</u>

Fig. 6 – Liste finale des concepts médicaux.

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

Calcul de pertinence des concepts

Trois mesures de pondération : TF-IDF, BM25 et BERT.

Terme	Corpus «Autres»			
	Fréquence	TF-IDF	BM25	BERT
Arthrite déformante	24	0,02	0,50	0,40
Ataxie locomotrice	169	0,08	0,25	0,39
Atrophie musculaire	1465	0,43	0,15	0,42
Atrophie progressive	22	0,02	0,53	0,39
Catalepsie	975	0,28	0,15	0,39
Épilepsie	577	0,12	0,10	0,41
Hystérie	4934	0,45	0,05	0,41
Langue	3591	0,11	0,02	0,41
Maladie de Parkinson	130	0,09	0,35	0,37
Paralysie bulbaire	93	0,09	0,52	0,40
Paralysie rhumatismale	14	0,02	0,68	0,44
Sclérose latérale	127	0,09	0,37	0,41
Sclérose en plaque disséminées	12	0,02	0,83	0,40
Somnambulisme	3410	1	0,15	0,43

Tab. 2 – Pertinence des concepts sous forme des scores TF-IDF, BM25 et BERT, corpus «Autres».

Intensification du discours de Charcot dans le corpus Autres

Le terme le plus impactant dans le réseau de Charcot selon BM25 :
 sclérose en plaque disséminées ? (pertinence : 83%)

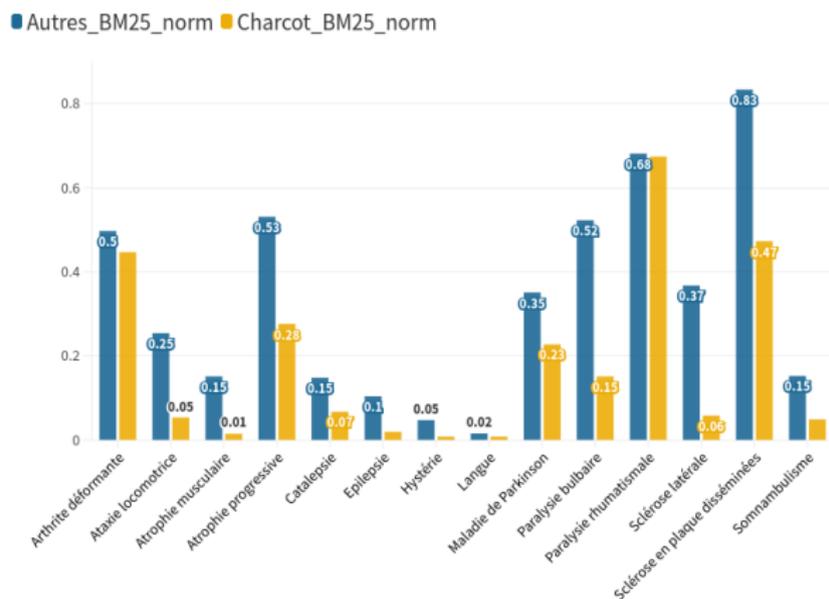


Fig. 7 – Pertinence des concepts dans les deux corpus (BM25).

BERT

VASWANI et al., 2017

- plongements lexicaux et des mécanismes d'attention
- modèle bert-base-multilingual-cased

Corpus «Charcot»		Corpus «Autres»	
diplopie	0,92	préambule	0,47
myélite partielle	0,91	délire	0,47
état de mal épileptique	0,91	miracle	0,47
paralysie labio-glosso-laryngée	0,91	cicatrices vicieuses	0,46
PATHOLOGIES		NOTIONS ABSTRAITES	

1. Contexte de recherche
2. Problématique et objectif
3. Approche supervisée
- 4. Approche non supervisée**
5. Conclusion et recherches futures

Extraction des phrases-clés : méthode keybert

- 1 entrée : un document
- 2 tokénisation du document en phrases-clés candidates (PCC)
- 3 génération des plongements du doc. et des PCC par un modèle de langage
- 4 calcul de la similarité cosinus entre le document et les PC

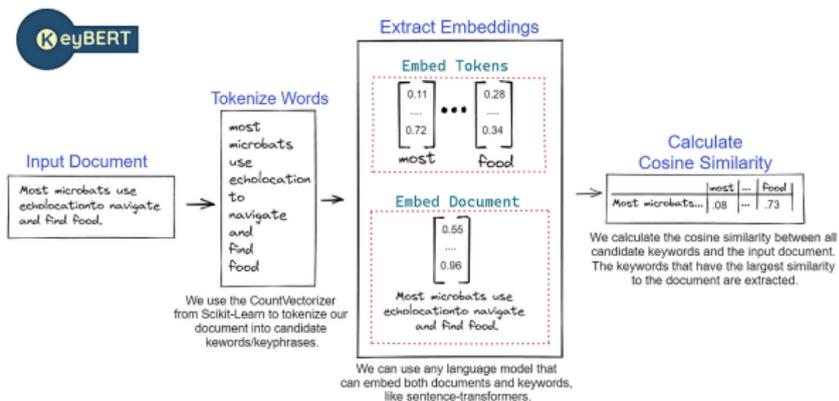


Fig. 8 – Pipeline de la librairie keybert (GROOTENDORST, 2020).

Limitations de keybert

⚠ manque de diversification des résultats + (non-)grammaticalité

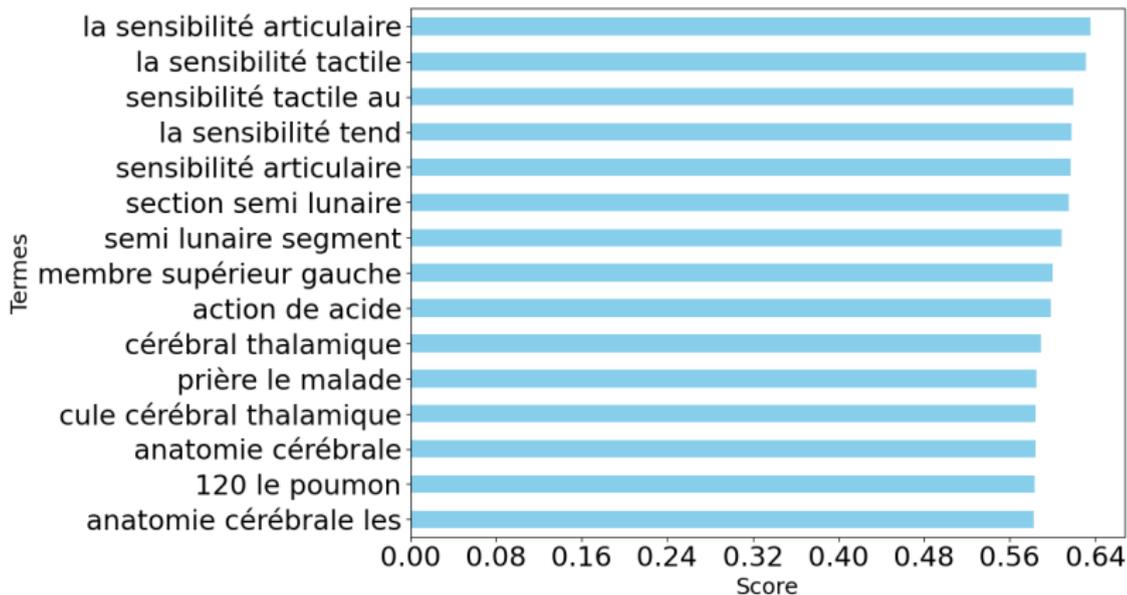


Fig. 9 – Répartition des 15 termes les plus pertinents dans le corpus « Autres » selon keybert.

Phrases-clés *hapax* partagés dans les deux corpus selon *keybert*

Les seuls termes partagés avec le corpus Charcot :

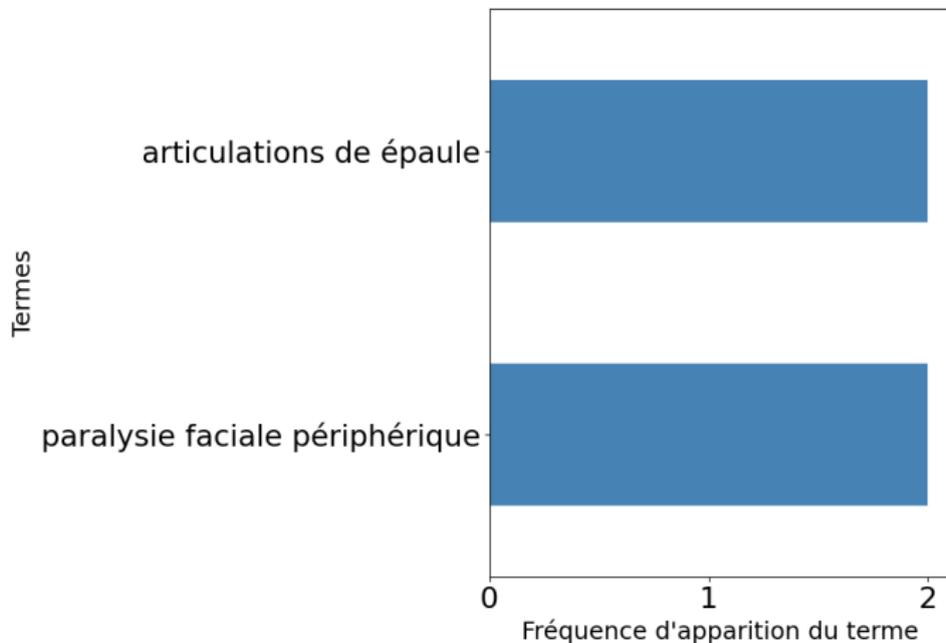


Fig. 10 – Répartition des termes les plus pertinents dans les deux corpus selon *keybert*.

Extraction des phrases-clés : méthode *PatternRank*

Librairie *keyphrase-vectorizers*

- 1 entrée : un seul document texte tokenisé
- 2 étiquetage des tokens avec les balises du partie du discours (POS)
- 3 sélection des tokens selon le motif POS → phrases-clés candidates (PCC)
- 4 génération des plongements du doc. et des PCC par un modèle de langue
- 5 calcul des similarités cosinus entre ces deux types de plongements + classement des PCC par ordre décroissant
- 6 extraction des *N* PC les plus représentatives

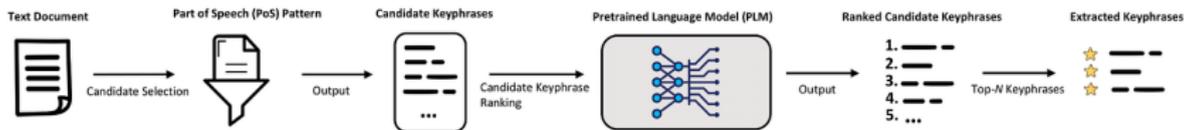


Fig. 11 – Workflow de la méthode *PatternRank* (SCHOPF et al., 2022).

Les termes partagés les plus fréquents | keyphrase-vectorizers

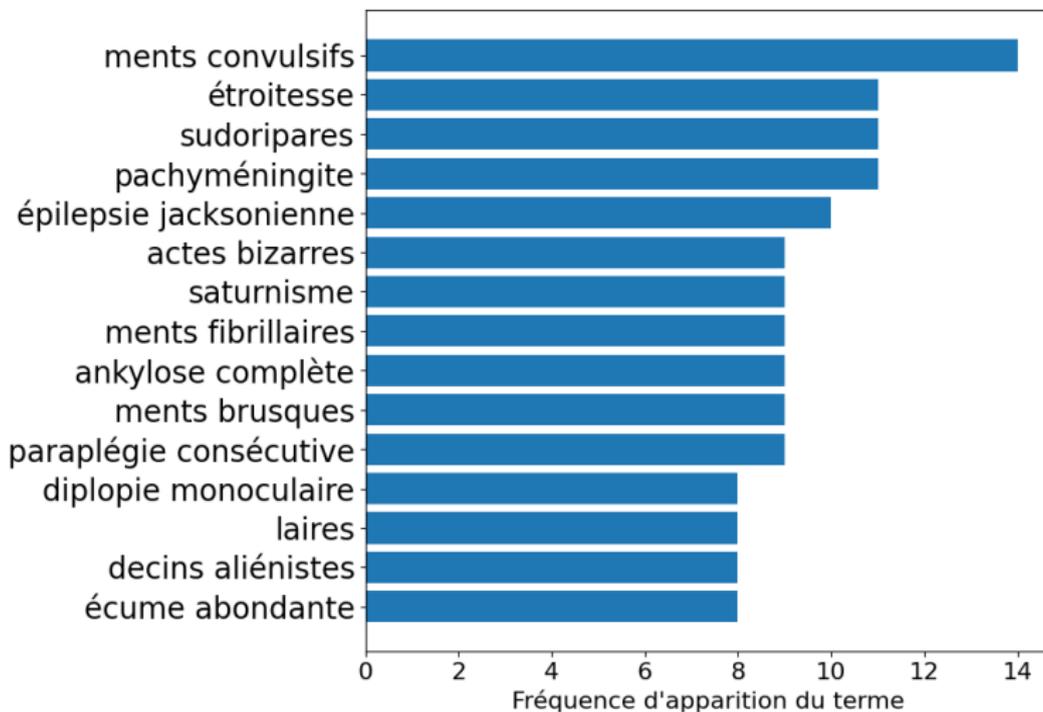


Fig. 12 – Les 15 termes les plus fréquents dans les deux corpus selon keyphrase-vectorizers.

1. Contexte de recherche
2. Problématique et objectif
3. Approche supervisée
4. Approche non supervisée
5. Conclusion et recherches futures

Conclusion et recherches futures

Contributions :

- étendue de l'impact de Charcot sur la neurologie
- étude numérique de son héritage scientifique

Prochaines étapes :

- 1 diversification des phrases-clés extraites par keybert
- 2 définir les notions des CIRCULATIONS NUMÉRIQUES et du CONCEPT du point de vue du TAL / linguistique
- 3 évaluation des phrases-clés extraites (semi-)automatique + retour d'un spécialiste de Charcot
- 4 contexte des réemplois textuels (affirmation, contestation ...)

Remerciements

Un grand merci à :

- **Valentina Fedchenko**
ingénieure de recherche, équipe-projet OB TIC
- **Motasem Alrahabi**
ingénieur de recherche, équipe-projet OB TIC
- **Glenn Roe**
professeur des universités, équipe-projet OB TIC
- **Simon Gabay**
maître-assistant, Chaire des humanités numériques, univ. de Genève
- **unité de service SACADO**
hôte de la plateforme MESU de Sorbonne Université, sur laquelle les expériences ont été réalisées

Dépôts GitHub

Les données et les scripts utilisés dans le cadre de cette étude sont disponibles dans les dépôts GitHub suivants :

- Mesurer l'influence de Charcot sur ses contemporains à l'aide de l'extraction de phrases-clés
- *Tracking the circulation of Jean-Martin Charcot's medical discourse : first observations.*

Références I



ALRAHABI, M. (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 durant la conférence francophone sur l'Extraction et la Gestion des Connaissances (egc2022)* (voir p. 11).



BOGOUSLAVSKY, J. (2020). The mysteries of hysteria : a historical perspective. In : *International Review of Psychiatry* 32.5-6, p. 437-450 (voir p. 7).



BROUSSOLLE, E., J. POIRIER, F. CLARAC et J.-G. BARBARA (2012). Figures and institutions of the neurological sciences in Paris from 1800 to 1950. Part III : Neurology. In : *Revue Neurologique* 168.4, p. 301-320 (voir p. 5).



CAMARGO, C. H. F., L. COUTINHO, Y. CORREA NETO, E. ENGELHARDT, P. MARANHÃO FILHO, O. WALUSINSKI et H. A. G. TEIVE (2024). Jean-Martin Charcot : the polymath. In : *Arquivos de Neuro-psiquiatria* 81, p. 1098-1111 (voir p. 4).



CHARCOT, J.-M. (1892). *Œuvres complètes de J.-M. Charcot : Leçons sur les maladies du système nerveux*. T. 1. Paris : Bureaux du Progrès médical (voir p. 13).



GABAY, S., L. PETKOVIC, A. BARTZ, M. G. LEVENSON et L. R. DU NOYER (2021). Katabase : À la recherche des manuscrits vendus. In : *Humanistica 2021* (voir p. 7).



GROOTENDORST, M. (2020). *KeyBERT : Minimal keyword extraction with BERT*. Version v0.3.0 (voir p. 18).

Références II



JOYEUX-PRUNEL, B. (2019). Visual Contagions, the Art Historian, and the Digital Strategies to Work on Them. In : *Artl@s Bulletin* 8.3, p. 8 ([voir p. 7](#)).



KOEHLER, P. J. (2013). Charcot, La Salpêtrière, and Hysteria as Represented in European Literature. In : *Progress in Brain Research* 206, p. 93-122 ([voir p. 5](#)).



MANJAVACAS, E., B. LONG et M. 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*. Minneapolis, USA : Association for Computational Linguistics, p. 104-114 ([voir p. 7](#)).



MARMION, J.-F. (2015). *Freud et la psychanalyse*. Sciences Humaines ([voir p. 4](#)).



SCHOPF, T., S. KLIMEK et F. MATTHES (2022). PatternRank : Leveraging Pretrained Language Models and Part of Speech for Unsupervised Keyphrase Extraction. In : *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022) – KDIR*. INSTICC. SciTePress, p. 243-248 ([voir p. 21](#)).



TASCA, C., M. RAPETTI, M. G. CARTA et B. FADDA (2012). Women and hysteria in the history of mental health. In : *Clinical practice and epidemiology in mental health : CP & EMH* 8, p. 110 ([voir p. 7](#)).



TEIVE, H. A. G., L. COUTINHO, C. H. F. CAMARGO, R. P. MUNHOZ et O. WALUSINSKI (2022). Thomas Willis' legacy on the 400th anniversary of his birth. In : *Arquivos de Neuro-Psiquiatria* 80, p. 759-762 ([voir p. 7](#)).

Références III



VASWANI, A., N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES, A. N. GOMEZ, L. KAISER et I. POLOSUKHIN (2017). **Attention Is All You Need**. In : *CoRR abs/1706.03762*. arXiv : 1706.03762 (*voir p. 16*).