

L'extraction terminologique : un levier pour analyser la circulation des savoirs ? Le cas du corpus Charcot.

Ljudmila PETKOVIC^{1,2,3,4}

`prenom.nom@sorbonne-universite.fr`

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

Journée de travail du consortium ARIANE (GT4)
SCAI, salle du Conseil
Paris, le 17 juin 2025

Sommaire

- 1 Projet Charcot
- 2 Problématique et hypothèses
- 3 Extraction de la terminologie
- 4 Résultats
- 5 Conclusion et perspectives

1 Projet Charcot

1.1. Présentation générale du sujet

1.2. Contexte et importance de l'étude

1.3. Objectifs de la recherche

2 Problématique et hypothèses

3 Extraction de la terminologie

4 Résultats

5 Conclusion et perspectives

1 Projet Charcot

1.1. Présentation générale du sujet

1.2. Contexte et importance de l'étude

1.3. Objectifs de la recherche

2 Problématique et hypothèses

3 Extraction de la terminologie

4 Résultats

5 Conclusion et perspectives

Entre l'histoire des sciences et des humanités numériques

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

Valorisation numérique des archives de Jean-Martin Charcot

Circulation des savoirs et intertextualité



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

Thèse en cours (2021–)

initiative OPUS ^a

- dir. : Prof. D^r Glenn ROE
- co-enc. : D^r Motasem ALRAHABI

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

- père de la neurologie moderne
- contributions et influences :
 - hystérie, « Parkinson », SLA...
 - Freud, de la Tourette, Babinski...
- héritage scientifique vivant

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

1 Projet Charcot

1.1. Présentation générale du sujet

1.2. Contexte et importance de l'étude

1.3. Objectifs de la recherche

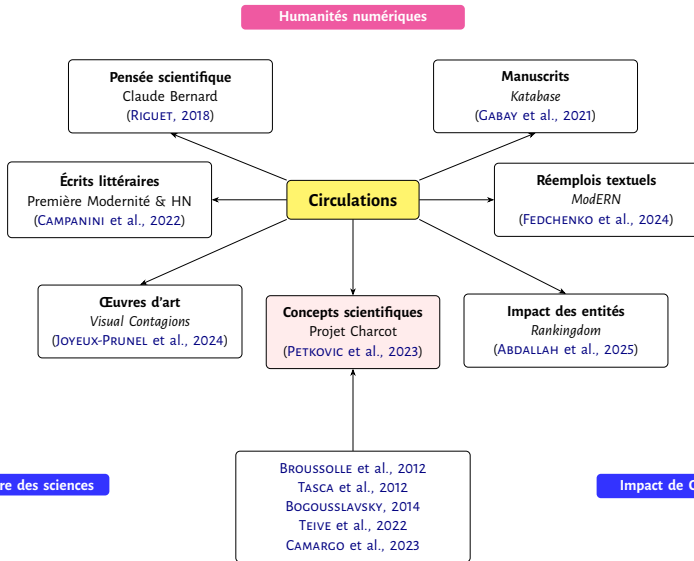
2 Problématique et hypothèses

3 Extraction de la terminologie

4 Résultats

5 Conclusion et perspectives

Études des circulations culturelles et scientifiques



1 Projet Charcot

1.1. Présentation générale du sujet

1.2. Contexte et importance de l'étude

1.3. Objectifs de la recherche

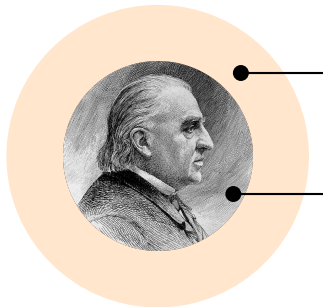
2 Problématique et hypothèses

3 Extraction de la terminologie

4 Résultats

5 Conclusion et perspectives

Double objectif



Formaliser une approche numérique pour tracer l'évolution des concepts médicaux en général

Pister numériquement la circulation des concepts médicaux associés à Charcot

- 1 définir le terme *concept scientifique* du point de vue de TAL
 - concept? idée? terme? mot? mot-clé? entité nommée?
- 2 comment des concepts circulent d'une discipline à une autre

(LANDAIS, 2014, p. 331)

5 Conclusion et perspectives

Problématique principale

Évaluer l'influence de Charcot *via* les termes repris dans son réseau

Quels concepts médicaux associés à Charcot ont eu un impact computationnellement mesurable sur son réseau scientifique ?

Concepts scientifiques en TAL :

- termes candidats ou entités nommées (EN) propres à un domaine ;
- détectables par des patrons linguistiques (morpho-syntaxiques) ;
- ancrage référentiel : valeur sémantique particulière à leur contexte ;
- pertinence élevée selon les métriques de pondération des termes.

(OMRANE et al., 2011, pp. 1-2) :

5 Conclusion et perspectives

Hypothèses formulées

- 1 Le texte comme point d'entrée pour étudier les tendances de la circulation des idées à l'aide des caractéristiques structurales.

(MILIA, 2023, p. 2)

- 2 Certains termes médicaux associés à Charcot ont été repris de manière significative dans les écrits de son réseau scientifique.
- 3 Chronologie d'une locution : indice de croissance de l'impact.
 - évolution de la fréquence des termes au sein des deux corpus ²
 - ex. : convergence entre des termes : fin XIX^e, début XX^e s.

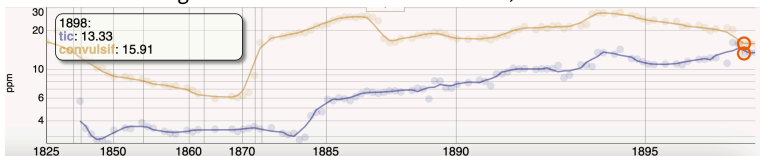


Fig. 2 – Chronologie de la fréquence du terme *tic convulsif*.

- 1 Projet Charcot
- 2 Problématique et hypothèses
- 3 Extraction de la terminologie**
 - 3.1. Méthodes de collecte de données
 - 3.2. *Design* de la recherche
 - 3.3. Outils et techniques utilisées
- 4 Résultats
- 5 Conclusion et perspectives

Fonds Charcot⁴

SorbonNum

Bibliothèque de Sorbonne Université (BSU)

201 documents XML OCRisés (sans post-correction)

Corpus	Docs	Tokens ³	%	Types	Lemmes	Diversité	Mémoire (Mo)
Charcot*	68	15 025 612	38,27	1 147 371	809 611	7,64	130,9
Autres**	133	24 232 207	61,73	1 773 538	1 218 074	7,32	179,6
Total	201	39 257 819	100	2 920 909	2 027 685	14,96	310,5

Tab. 1 – Description du corpus d'étude.

* textes rédigés par Charcot

** textes rédigés par son réseau scientifique

3. https://spacy.io/models/fr#fr_core_news_lg

4. <https://patrimoine.sorbonne-universite.fr/collection>

Distribution des ouvrages du fonds Charcot

Entre 1825 et 1934.

- corpus Charcot : fin XIX^e-début XX^e s. (1881-1907)
- corpus Autres : période plus longue (1825-1934)

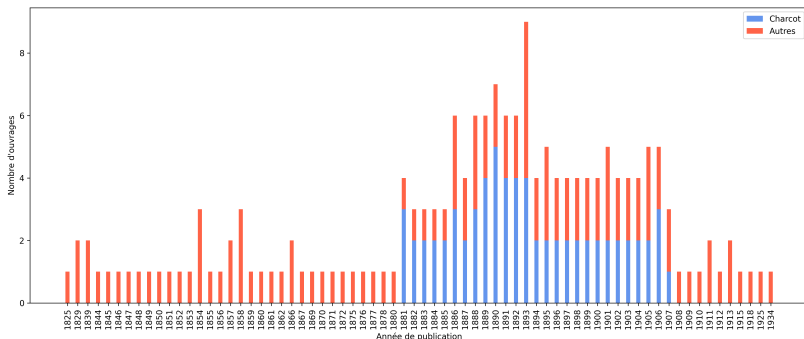


Fig. 3 – Répartition des ouvrages constituant les corpus «Charcot» et «Autres» par année.

5 Conclusion et perspectives

Formalisation de l'approche

- comparaison des résultats avec la liste des concepts (vérité terrain)
- recenser le score le plus élevé sur le terme ou sur son synonyme

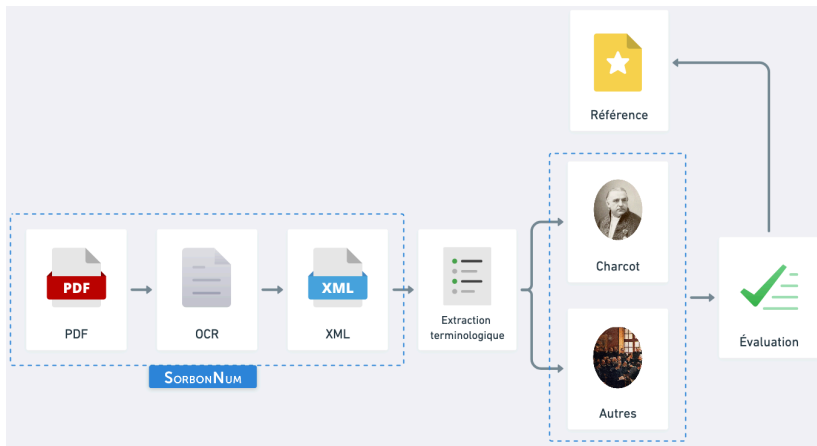


Fig. 4 – Pipeline pour pister la circulation des termes médicaux associés à Charcot.

Liste des concepts médicaux – vérité terrain

Extraction semi-automatique des termes en lien avec Charcot.

https://github.com/ljpetkovic/Charcot_circulations/tree/main/concepts

HYSTÉRIE (V. ÉPIDÉMIE, HÉMIANES-THÉSIE, 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, Hémianes-thésie, Hyperesthésie ovarienne,  
</p>  
<p>  
<s>Hystéro-épilepsie, 332, 367. — Signification de ce mot, 368 ; —  
</p>
```

Fig. 6 – Concepts médicaux, document XML.

Fig. 5 – 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>

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

Fig. 7 – Liste finale des concepts médicaux.

Enrichissement de la liste des concepts

Termes inventés par Charcot :

nom traditionnel	nom moderne / synonyme
paralysie agitante	maladie de Parkinson
ataxie locomotrice progressive	<i>tabes dorsalis</i>
arthropathies tabétiques	arthropathie de Charcot
sclérose latérale amyotrophique	maladie de Charcot / Lou Gehrig
idée(s) fixe(s), maladie des tics	syndrome de Tourette
...	

≠ Termes transmis par Charcot :

athétose	mouvements involontaires
hystérie	névrose
épilepsie	attaques convulsives
hypnose	transe
sclérose en plaques disséminées	sclérose multiple

(WALUSINSKI, 2025; CAMARGO et al., 2023)

5 Conclusion et perspectives

Approches comparées

- ❶ **TermSuite**⁵ (CRAM et DAILLE, 2016)
 - linguistique, à base de règles → TD-IDF
- ❷ **TF-IDF, BM25**⁶ (ROBERTSON et JONES, 1976)
 - statistique
- ❸ **PatternRank** (SCHOPF et al., 2022)⁷
 - apprentissage profond
 - keybert + keyphrase-vectorizers
 - utilisation des étiquettes POS

Traitements effectués en local (1,2) et *via* la plateforme MeSU^a (3).

- appliqués à tout le corpus

a. <https://sacado.sorbonne-universite.fr/fr/plateforme-mesu/>

5. https://github.com/ljpetkovic/Charcot_TermSuite

6. https://github.com/ljpetkovic/Charcot_circulations

7. https://github.com/ljpetkovic/Seminaire_doctoral_0bTIC_130325/blob/main/0_main.pdf ◀ ▶ ☰ ☷ 🔍 ↺

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

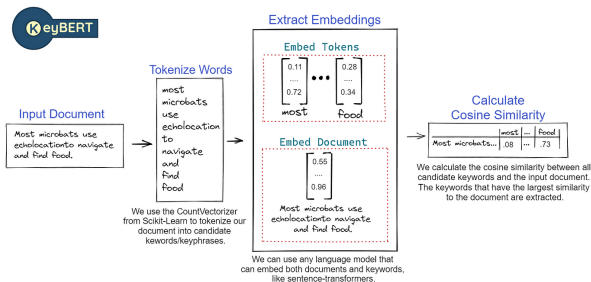


Fig. 8 – Pipeline de la librairie keybert (grootendorst2020keybert).

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

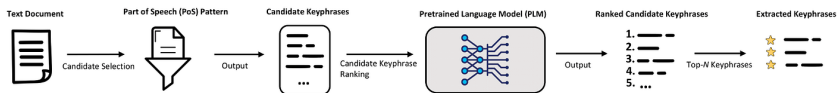


Fig. 9 – Workflow de la méthode *PatternRank* (SCHOPF et al., 2022).

- 1 Projet Charcot
- 2 Problématique et hypothèses
- 3 Extraction de la terminologie
- 4 Résultats**
 - 4.1. Présentation des résultats principaux
 - 4.2. Analyse et interprétation des résultats
- 5 Conclusion et perspectives

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

Domaine impactant dans les écrits de Charcot : **hystérie**

Terme	TF-IDF (<i>TermSuite</i>)	TF-IDF	BM25	PatternRank	Moyenne
<i>maladie de Parkinson</i>	NA	0,2478	0,2397	0,7926	0,3200
<i>ataxie locomotrice progressive</i>	0,1981	0,4114	0,2313	0,7912	0,408
<i>arthropathies tabétiques</i>	0,4424	0,1655	0,337	0,8050	0,4375
<i>trépidation épileptoïde du pied</i>	0,0379	0,2581	0,053	0,7581	0,2768
<i>sclérose en plaques disséminées</i>	NA	0,2935	0,4812	0,7611	0,3840
<i>tremblement</i>	NA	0,2712	0,0213	0,7834	0,2690
<i>nystagmus</i>	NA	0,2142	0,0488	0,7683	0,2578
<i>embarras parole</i>	NA	0,0724	0,9143	0,8159	0,4507
<i>sclérose latérale amyotrophique</i>	NA	0,3287	0,1152	0,7514	0,2988
<i>tics convulsifs</i>	0,0670	0,2273	0,1696	0,8073	0,3178
<i>atrophie musculaire progressive</i>	0,1161	0,2321	0,0797	0,7874	0,3038
<i>aphasie</i>	0,1722	0,345	0,0289	0,7824	0,3321
<i>astasia-abasie</i>	0,1281	0,7022	0,1912	0,7891	0,4527
<i>athétose</i>	NA	0,226	0,0797	0,7910	0,2742
<i>chorées</i>	0,1593	0,1933	0,0213	0,8030	0,2942
<i>hystérie</i>	0,6892	0,5407	0,0213	0,8194	0,5177
<i>épilepsie</i>	0,0062	0,534	0,0213	0,8170	0,3446
<i>hypnose</i>	0,0311	0,4294	0,0994	0,7955	0,3389
<i>systématisation de l'organisation de la moëlle épinière</i>	NA	0	0	NA	0
<i>localisations cérébrales</i>	NA	0,27	0,0943	0,7493	0,2784

Tab. 2 – Les scores de pertinence pour les termes de référence à partir du corpus « Charcot ».

Moyenne pour tous les termes combinés : 0,3280

Domaine impactant dans les écrits des Autres : syndrome de Tourette

Terme	TF-IDF (TermSuite)	TF-IDF	BM25	PatternRank	Moyenne
<i>maladie de Parkinson</i>	0,05	0,0775	0,333	0,7936	0,3135
<i>ataxie locomotrice progressive</i>	0,32	0,0386	0,4877	0,7431	0,3974
<i>arthropathies tabétiques</i>	0,33	0,0934	0,4928	0,7506	0,4167
<i>trépidation épileptoïde du pied</i>	0,0198	0,1227	0,2919	0,7597	0,2985
<i>sclérose en plaques disséminées</i>	NA	0,178	0,8089	NA	0,2467
<i>tremblement</i>	NA	0,1686	0,0362	0,7683	0,2432
<i>nystagmus</i>	0,0243	0,1326	0,146	0,7474	0,2626
<i>embarras parole</i>	NA	NA	0,0018	0,9347	0,2341
<i>sclérose latérale amyotrophique</i>	NA	0,044	0,6586	NA	0,1757
tics convulsifs	NA	0,1293	0,8385	0,8331	0,4502⁸
<i>atrophie musculaire progressive</i>	0,40	0,1118	0,3489	0,8053	0,4165
<i>aphasie</i>	0,0587	0,2245	0,1334	0,7960	0,3031
<i>astasia-abasie</i>	NA	0,0478	0,3565	0,7375	0,2855
<i>athétose</i>	NA	0,2029	0,274	0,8068	0,3209
<i>chorées</i>	NA	0,1336	0,0701	0,8047	0,2521
<i>hystérie</i>	0,2724	0,3711	0,0442	0,8018	0,3723
<i>épilepsie</i>	NA	0,164	0,0247	0,8199	0,2521
<i>hypnose</i>	0,3543	1	0,2922	0,7738	0,6050
<i>systématisation de l'organisation de la moëlle épinière</i>	NA	NA	NA	0,7550	0,1888
<i>localisations cérébrales</i>	0,43	0,034	0,3017	0,8090	0,3937

Tab. 3 – Les scores de pertinence pour les termes de référence à partir du corpus « Autres ».

Moyenne pour tous les termes combinés : 0,3214

8. Le score le plus élevé parmi les termes **inventés** par Charcot ≠ *hypnose*.

Limitations de keybert

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

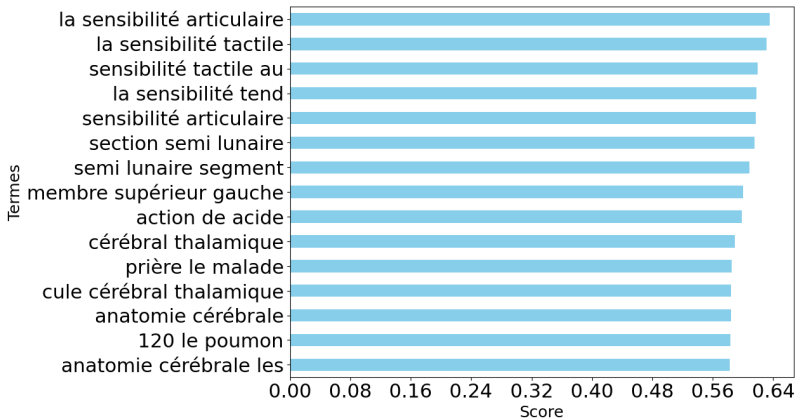


Fig. 10 – Répartition des 15 termes les plus pertinents dans le corpus « Autres » selon keybert.

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

Les seuls termes partagés avec le corpus Charcot :

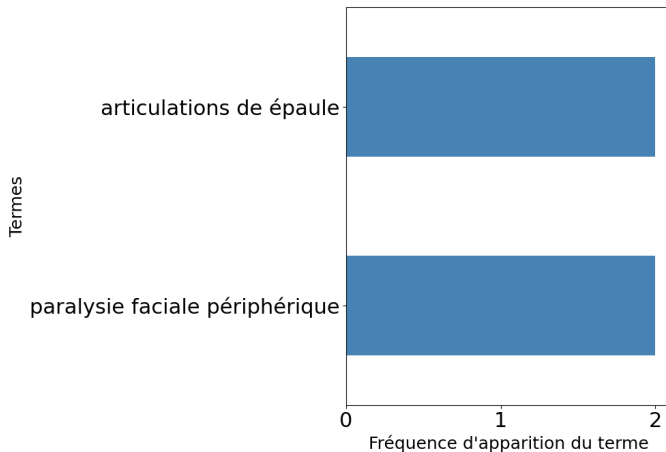


Fig. 11 – Répartition des termes les plus pertinents dans les deux corpus selon keybert.

Les termes partagés les plus fréquents | keyphrase-vectorizers

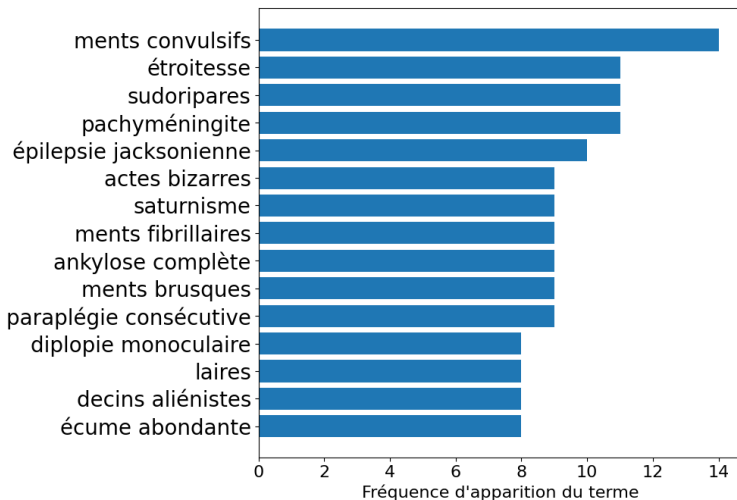


Fig. 12 – Les 15 termes les plus fréquents dans les deux corpus selon keyphrase-vectorizers.

Analyse comparative des approches employées

- *PatternRank* valorise systématiquement les termes
- pas de consensus entre les métriques
 - l'écart le plus petit entre eux : *hypnose*

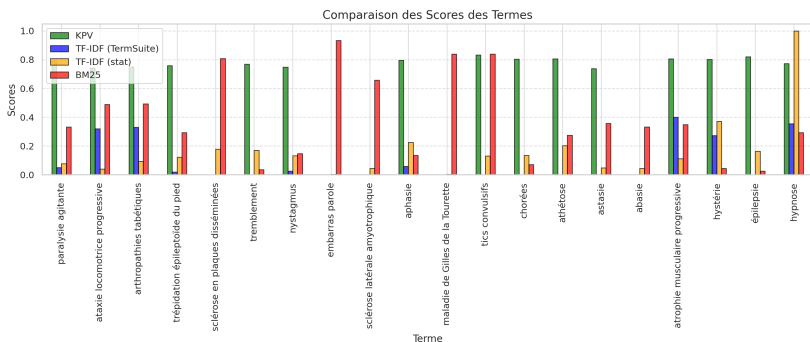


Fig. 13 – Visualisation des scores de pertinences pour chaque terme de référence.

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ↺ 🔍 ↻

Analyse des concordances des termes médicaux

Analyses effectuées dans TXM.

- on utilise un terme dans le contexte où on cite Charcot
 - terme traditionnel (ex. *paralysie agitante*)
 - synonyme ou terme relevant du champ conceptuel en question
 - *pied tabétique* → forme particulière d'*arthropathies tabétiques*

Exemple

```
([word = "paralysie"] [word = "agitante"] []* [word = "Charcot"] | [word = "Charcot"] []* [word = "paralysie"] [word = "agitante"])
```

- repérer toutes les occurrences dans un paragraphe où *paralysie agitante* et *Charcot* apparaissent dans n'importe quel ordre, séparés par 0 ou plusieurs mots

Références à Charcot

Terme	Contexte
<i>maladie de Parkinson</i>	[...] <u>paralysie agitante</u> que Charcot a eu raison de dénommer <u>maladie de Parkinson</u>
<i>ataxie locomotrice progressive</i>	[...] arthropathies [...] de l' <u>ataxie locomotrice</u> , [...] signalées, pour la première fois, par M. Charcot . [...] <u>ataxie locomotrice progressive</u> , constatée par douze médecins, parmi lesquels, [...] MM. Charcot [...]
<i>arthropathies tabétiques</i>	[...] mais personne n'avait encore décrit des cas d' <u>arthropathie tabétique</u> , lorsque, en 1808, M. le professeur Charcot publia la première observation d'arthropathie chez un ataxique.
<i>trépidation épileptoïde du pied</i>	[...] trépidation, qui se propage parfois à tous les <u>membres</u> . Ce spasme, [...] peut entraîner à sa suite des rétractions fibro-tendineuses analogues à celles que Charcot a décrites chez l'homme, [...]
<i>sclérose en plaques disséminées</i>	[...] une combinaison de la <u>sclérose en plaques</u> , bien décrite déjà par Charcot et Vulpian [...]

Tab. 4 – Concordance des termes médicaux faisant référence à Charcot – corpus Autres.

Références à Charcot

Terme	Contexte
<i>tremblement</i>	[...] Charcot présentait, dans son amphithéâtre, pour démontrer les caractères oscillatoires des diverses variétés de <u>tremblements</u> .
<i>nystagmus</i>	SCLÉROSE EN PLAQUES [...] [...], <u>nystagmus</u> , [...]. La difficulté de la résoudre est d'autant plus grande qu'à côté des foyers de sclérose en plaques avec tous les caractères histologiques classiques décrits depuis Charcot , il y a des foyers avec une destruction plus ou moins complète des cylindraxes [...]
<i>embarras parole</i>	Lorsqu'on se trouve en présence d'un malade ayant de l' <u>embarras de la parole</u> [...] A. La réponse à la première proposition n'est nullement embarrassante, si l'on veut se rappeler ces paroles de M. le professeur Charcot : [...]
<i>sclérose latérale amyotrophique</i>	[...] <u>sclérose latérale amyotrophique</u> , maladie découverte par mon illustre maître Charcot .
<i>tics convulsifs</i>	désignée par M. Charcot sous le nom de maladie des <u>tics convulsifs</u>
<i>atrophie musculaire progressive</i>	Charcot et Marie ont décrit la « forme particulière d' <u>atrophie musculaire progressive</u> »
<i>aphasie</i>	Lorsqu'il y a, dit M. le professeur Charcot (1), suppression de la mémoire pour l'articulation des mots, c'est l' <u>aphasie</u> motrice d'articulation ou <u>aphasie</u> de Broca qui se présente.

Tab. 5 – Concordance des termes médicaux faisant référence à Charcot – corpus Autres (suite).

Références à Charcot

Terme	Contexte
<i>astasie-abasie</i>	[...] forme particulière d'impuissance motrice dont M. P. Blocq a donné la définition suivante « [...], et qu'il a désigné sous le nom expressif d' astasie et d' abasie . C'est là un état morbide sur lequel M. le professeur Charcot est fréquemment revenu dans ses Leçons du mardi [...]
<i>athétose</i>	symptôme désigné par M. W. Hammond sous le nom d' <u>athétose</u> [...] M. Charcot a fait remarquer que cette définition était imparfaite pour les motifs suivants : [...]
<i>chorées</i>	<u>Chorée</u> hystérique ou rythmique. C'est à M. le professeur Charcot que nous devons une exacte description de cet état pathologique.
<i>hystérie</i>	C'est encore à lui [Charcot] que nous devons la connaissance de l' <u>hystérie</u> traumatique [...]
<i>épilepsie</i>	M. Charcot a décrit avec le plus grand soin l' <u>épilepsie</u> partielle d'origine syphilitique [...]
<i>hypnose</i>	Les trois états de l' <u>hypnose</u> décrits par M. Charcot sont devenus classiques, [...]
<i>systématisation de l'organisation de la moëlle épinière</i>	systématisation de la moelle, synthèses [...]. Mais Charcot , on l'a vu, est, par nature, enclin à la synthèse.
<i>localisations cérébrales</i>	Je vous ai montré Charcot , concourant pour la plus grosse part, à l'édification de la doctrine des <u>localisations cérébrales</u> , qui est devenue quelque chose comme la préface d'une psychologie nouvelle.

Tab. 6 – Concordance des termes médicaux faisant référence à Charcot – corpus Autres (fin).

Analyse des cooccurrences

- quels cooccurrents avec les termes médicaux ciblés?
 - Charcot, Babinski, Necker...
- recensement des résultats pour le cooccurrent : Charcot
- sinon, autre cooccurrent (médecin) avec l'indice le plus élevé

Exemple

[word = "athétose"]

- liste des cooccurrents pour le terme *athétose*

Analyse des cooccurrences

Terme	Cooccurent	Fréquence	Co-fréquence	Indice	Distance moyenne
maladie de Parkinson	Charcot	2 968	3	2	3,3
ataxie locomotrice progressive	Charcot	2 968	3	2	5,3
arthropathies tabétiques	Charcot	2 968	4	5	1,5
trépidation épileptoïde du pied	Babinski · Charcot	1 134	8	12	4,8
sclérose en plaques disséminées	Charcot	2 968	7	3	6,3
tremblement	Achard	137	3	2	2,3
nystagmus	Barany	11	3	7	3,3
embarras parole	Chervin	41	8	17	4,5
sclérose latérale amyotrophique	Charcot	2 968	4	3	3,5
tics convulsifs	Charcot	2 968	6	8	5,2
atrophie musculaire progressive	Charcot	2 968	4	3	6,0
aphasie	Charcot	2 968	7	2	4,0
astasie-abasie	Charcot	2 968	2	2	1,5
athétose	Hammond	34	2	4	1,5
chorées	Sydenham	129	63	163	1,1
hystérie	Charcot	2 968	52	19	5,4
épilepsie	Jackson	52	34	78	0,2
hypnose	Braid	567	14	12	4,7
systematisation de l'organisation de la moëlle épinière	NA	NA	NA	NA	NA
localisations cérébrales	Charcot	2 968	9	7	5,3

Tab. 7 – Analyse des cooccurrences des termes médicaux à partir du corpus Autres.

Analyse des cooccurrences

Absence occasionnelle de cooccurrent Charcot explicable :

- *épilepsie* : terme créé par J. H. Jackson
- *hypnose* : terme créé par J. Braid
- *athétose* : terme créé par W. A. Hammond
- *chorées* : définition moderne par T. Sydenham
- *trépidation épileptoïde du pied* : Babinski ? Vulpian ? Charcot ?

Analyse des autres cooccurents

Terme	Contexte
<i>trépidation épileptoïde du pied (clonus)</i>	On [Babinski] le désigne alors sous la dénomination de « clonus du pied », « <u>trépidation épileptoïde du pied</u> »
<i>tremblement</i>	D'après Charcot et surtout d'après Achard , ce <u>tremblement</u> aurait de certaines analogies avec le tremblement sénile, [...]
<i>nystagmus</i>	[...] les recherches inspirées par les travaux de Barany sur le <u>nystagmus</u> provoqué indiqueraient une certaine fréquence de troubles labyrinthiques [...]
<i>embarras parole</i>	M. le Dc Chervin , [...] vient de rédiger un nouveau résumé des notions cliniques fondamentales indispensables à connaître sur quelques <u>troubles fonctionnels de la parole</u> et notamment sur le bégaiement.
<i>athétose</i>	[...] nous avons affaire au symptôme désigné par M. W. Hammond sous le nom d' <u>athétose</u> .
<i>chorées</i>	Cependant, la <u>chorée</u> de Sydenham présente quelques particularités sémiologiques que nous allons passer en revue.
<i>épilepsie</i>	Depuis les remarquables travaux de M. Hughlings Jackson sur la forme d' <u>épilepsie</u> à laquelle il a attaché son nom, [...]
<i>hypnose</i>	Selon Braid , l' <u>hypnose</u> est caractérisée par des phénomènes mentaux et physiques, particuliers à cette condition.

Tab. 8 – Concordance des termes médicaux faisant référence à d'autres médecins – corpus Autres.

- 1 Projet Charcot
- 2 Problématique et hypothèses
- 3 Extraction de la terminologie
- 4 Résultats
- 5 Conclusion et perspectives**

Conclusion et perspectives

- 1 *PatternRank* : la méthode la plus robuste
 - capture la sémantique jusqu'aux pentagrammes
 - *méningite syphilitique hémorragique fibrineuse aiguë*
 - produit des scores de pertinence plus élevés
 - exception : scores BM25 (SLA, *embarras parole*) et TF-IDF (*hypnose*)
- 2 les termes les plus impactants dans les corpus :
 - Charcot : *hystérie, astasie-abasie, embarras parole*
 - Autres : *hypnose**, *syndrome de Tourette*, *arthropathies tabétiques*

Les résultats sont alignés avec les faits historiques.

Recherches futures : tester les *LLM* ou les *LCM* (angl. *Large Concept Models*) ?

Références I



ABDALLAH, H., B. MARKHOFF, M. OVIDE, L. PARKIN et A. SOULET (2025). Rankingdom : une architecture coopérative pour l'analyse à la demande de Wikidata. In : *Extraction et Gestion des Connaissances 2025*.
<https://hal-hceres.archives-ouvertes.fr/LIBDTLN/hal-04977222v1> (voir p. 7).



BOGOUSLAVSKY, J. (2014). The Mysteries of Hysteria. In : *Neurosciences and History 2.2*.
https://nah.sen.es/vmfiles/abstract/NAHV2N2201454_73EN.pdf, p. 54-73 (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. <https://doi.org/10.1016/j.neurol.2011.10.006>, p. 301-320 (voir p. 7).



CAMARGO, C. H. F., L. COUTINHO, Y. CORREA NETO, E. ENGELHARDT, P. MARANHÃO FILHO, O. WALUSINSKI et H. A. G. TEIVE (2023). Jean-Martin Charcot : the polymath. In : *Arquivos de Neuro-psiquiatria* 81. <https://www.thieme-connect.de/products/ejournals/pdf/10.1055/s-0043-1775984.pdf>, p. 1098-1111 (voir pp. 7, 22).

Références II



CAMPANINI, M., M. PARRA, C. PRIMOT, A. RÉACH-NGÔ, C. SAIGNOL, M. SPEYER, S. VERTENT-GIRAUD, R. WALTER et al. (2022). Circulation des écrits littéraires de la Première Modernité & Humanités numériques. In : *LE VERGER*.
<http://cornucopia16.com/blog/2022/05/01/bouquet-xxiii-circulation-des-ecrits-litteraires-de-la-premiere-modernite-humanites-numeriques-2/> (voir p. 7).



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. 21).



CRAM, D. et B. DAILLE (2016). Terminology Extraction with Term Variant Detection. In : *Proceedings of ACL-2016 system demonstrations*.
<https://aclanthology.org/P16-4003.pdf>, p. 13-18 (voir p. 24).



FEDCHENKO, V., D. M. NICOLASI 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. 7).



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*.
<https://hal.science/hal-03066108>, p. 1-7 (voir p. 7).

Références III



JOYEUX-PRUNEL, B., M. BARRAS et N. CARBONI (2024). Greater Aura? The 19th and 20th Century Circulation of Art Reproductions in the Press through a Digital and Historical Perspective. In : *Život umjetnosti : časopis o modernoj i suvremenoj umjetnosti i arhitekturi* 114.1. <https://doi.org/10.31664/zu.2024.114.02>, p. 26-49 (voir p. 7).



LANDAIS, É. (2014). Frédéric Darbellay, éd., *La circulation des savoirs. Interdisciplinarité, concepts nomades, analogies, métaphores* : Berne, P. Lang, 2012, 245 pages. In : *Questions de communication* 26. <https://doi.org/10.4000/questionsdecommunication.9367>, p. 331-333 (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 p. 14).



OMRANE, N., A. NAZARENKO et S. SZULMAN (2011). Le poids des entités nommées dans le filtrage des termes d'un domaine. In : *9^e conférence internationale de Terminologie et Intelligence Artificielle*. <https://hal.science/hal-00626843v4>, p. 80-86 (voir p. 12).

Références IV



PETKOVIC, L., M. ALRAHABI et G. ROE (2023). Circulation du discours médical de Jean-Martin Charcot. In : *Humanistica 2023*.

<https://hal.science/HUMANISTICA-2023/hal-04107099v1> (voir p. 7).



RIGUET, M. (2018). L'impact de la physiologie dans la critique littéraire de la fin du XIX^{ème} siècle : l'exemple de Claude Bernard. In : *Epistémocritique : Littérature et savoirs*.

<https://hal.science/hal-01903871> (voir p. 7).



ROBERTSON, S. E. et K. S. JONES (1976). Relevance Weighting of Search Terms. In : *Journal of the American Society for Information science* 27.3.

[https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.4630270302?casa_token=TfyVkMGkDQsAAAAA:](https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.4630270302?casa_token=TfyVkMGkDQsAAAAA:TCuXWzGHjo31RdxGR9jECRG2rZzqvOK3G0zHF7yAa2NfxtdFqxe-MmSHMC6e80FiFxI4sLj2aW60yDk)

[TCuXWzGHjo31RdxGR9jECRG2rZzqvOK3G0zHF7yAa2NfxtdFqxe-MmSHMC6e80FiFxI4sLj2aW60yDk](https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.4630270302?casa_token=TfyVkMGkDQsAAAAA:TCuXWzGHjo31RdxGR9jECRG2rZzqvOK3G0zHF7yAa2NfxtdFqxe-MmSHMC6e80FiFxI4sLj2aW60yDk), p. 129-146 (voir p. 24).



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*. <http://dx.doi.org/10.5220/0011546600003335>. SCITEPRESS – Science et Technology Publications. DOI : 10.5220/0011546600003335. URL : <http://dx.doi.org/10.5220/0011546600003335> (voir pp. 24, 26).

Références V



TASCA, C., M. RAPETTI, M. G. CARTA et B. FADDA (2012). Women And Hysteria In The History Of Mental Health. In : *Clinical Practice & Epidemiology in Mental Health : CP & EMH* 8. <https://doi.org/10.2174/1745017901208010110>, p. 110-119 (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. <https://doi.org/10.1055/s-0042-1755278>, p. 759-762 (voir p. 7).



WALUSINSKI, O. (2025). *Jean-Martin Charcot's Birth Bicentennia*. <http://ishn.org/>. Consulté le 17 juin 2025 (voir p. 22).