

# ANALYSE DES CHAMPS LEXICAUX DANS LE FONDS PATRIMONIAL DE JEAN-MARTIN CHARCOT

**Ljudmila PETKOVIĆ<sup>1</sup>**

```
prenom.nom@etu.unige.ch
```

<sup>1</sup> Université de Genève, Faculté des Lettres, Département de linguistique

Soutenance du mémoire du certificat de spécialisation en linguistique  
Université de Genève, bâtiment Landolt, salle L208  
Genève, le 29 octobre 2024

UNIVERSITÉ  
DE GENÈVE

FACULTÉ DES LETTRES

# Plan

1 À propos du projet

2 Problématique

3 Méthodologie

4 Résultats

5 Conclusion

# 1 À propos du projet

## 2 Problématique

## 3 Méthodologie

## 4 Résultats

## 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>

## CS en linguistique (2019-)

- directeur : Prof. D<sup>r</sup> Christopher LAENZLINGER
- co-directeur : Luka NERIMA

## Thèse en cours (2021-)

- directeur : Prof. D<sup>r</sup> Glenn ROE
- co-encadrant : D<sup>r</sup> Motasem ALRAHABI



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

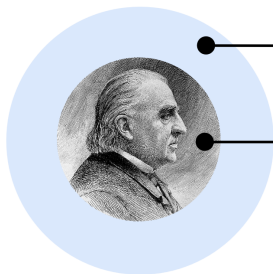
- Père de la neurologie moderne
- Impact sur son « réseau »  
Freud, de la Tourette, Babinski...
- hystérie, SLA, Parkinson...
- Fonds Charcot sur SorbonNum<sup>a</sup>

a. <https://patrimoine.sorbonne-universite.fr>

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

## 5 Conclusion

# Double objectif



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

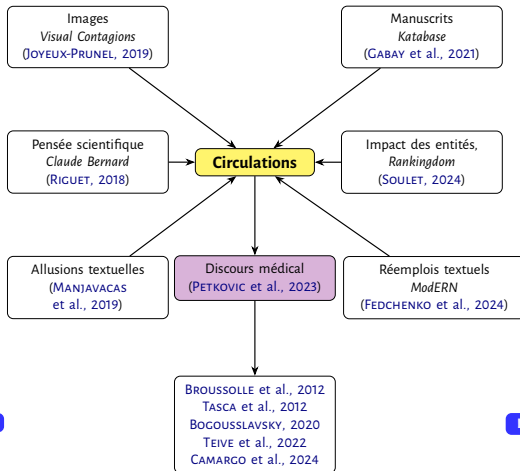
Pister numériquement la circulation du discours médical de Charcot

Quels sont les champs lexicaux\* dominants dans le discours de Charcot?

ensembles de mots liés par leur sémantique, traitant d'un domaine commun, p. ex. neurologie.

# Études numériques des circulations culturelles et scientifiques

## Humanités numériques



Histoire des sciences

Impact de Charcot

# Analyse de l'impact de l'entité Charcot *via* Rankingdom

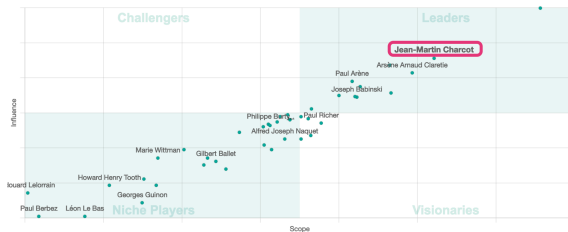


Fig. 2 – Analyse de quadrant : positionnement de l'entité Charcot au sein de son domaine.

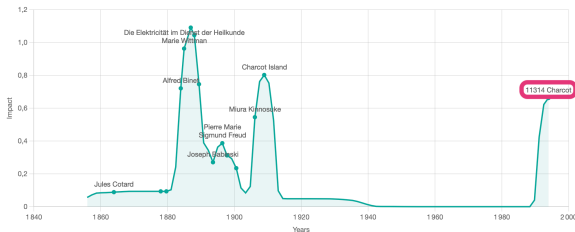


Fig. 3 – Analyse temporelle de l'impact de l'entité Charcot.



- 1 À propos du projet
- 2 Problématique
- 3 Méthodologie**
- 4 Résultats
- 5 Conclusion

# Mesurer le degré d'intertextualité

Mesurer informatiquement l'impact de Charcot sur son « réseau »

→ intertextualité uni-directionnelle

- rapports entre une œuvre et d'autres qui l'ont précédée ou suivie

(RIFATERRE, 1980)

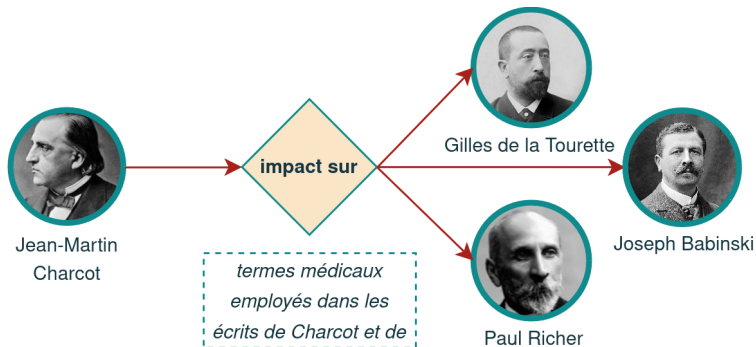


Fig. 4 – Opérationnalisation de l'impact de Charcot sur ses élèves.

# Corpus Charcot

SorbonNum<sup>a</sup> | Bibliothèque de Sorbonne Université (BSU)

a. <https://patrimoine.sorbonne-universite.fr/>

201 documents XML OCRisés (sans post-correction)

- Charcot : textes rédigés par Charcot
- Autres : textes rédigés par son réseau scientifique

Corpus	Nb de docs	Nb de tokens
Charcot	68	12 190 649 (38,12 %)
Autres	133	19 788 830 (61,88 %)
<b>Total</b>	<b>201</b>	<b>31 979 479 (100%)</b>

Tab. 1 – Répartition du corpus issu du fonds Charcot<sup>2</sup>.

2. <https://patrimoine.sorbonne-universite.fr/collection/Fonds-Charcot>

# OBVIE – recherche textuelle, corpus Charcot<sup>3</sup>

OBVIE : moteur de recherche avancée (ALRAHABI, 2022)

⚠ quantification de l'importance des « collocations » (NERIMA et al., 2006)

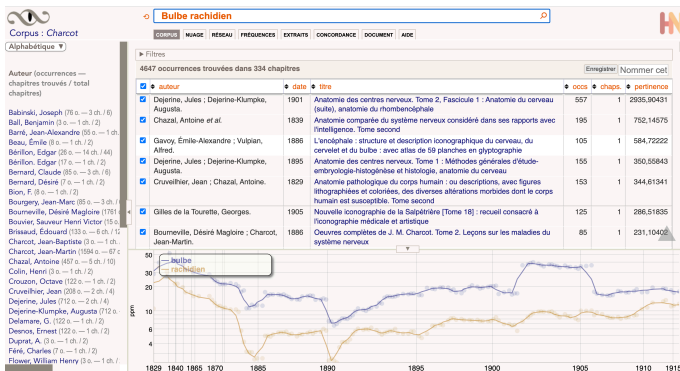


Fig. 5 – Distribution des fréquences des tokens avec la frise chronologique pour ceux constituant l'expression *bulbe rachidien* (issus des corpus « Charcot » et « Autres »).

# OBVIE – comparaison des documents similaires

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

**Charcot**

CORPUS | NUAGE | RÉSEAU | FRÉQUENCES | EXTRAITS | CONCORDANCE | COMPARER | AIDE

CHARCOT, Jean-Martin ; BOURNEVILLE, Désiré Magloire. (1881) Archives de

**Mots fréquents :** 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 .

**Noms cités :** J, T, B, Burg, H, d' O, T, R, W, Pl, P, Schwann, H, de B, T, U, M, Blaise, Chauvet, H, PÉRI-AXILE, MÉTALLOSCOPIE, N, J, R, PATHOGÉNIE un tremblement, Maragliano, Arnozan, Seguin, Wrisberg, J, Landouzy, J, Laffont, MÉTALLOTHÉRAPIE, Charcot, Bogdanow, Schiff, Burman, Vulpian, Despine, Treub, Archiv, F, M. Russell, Beard, J, Ir, Ranvier, Journ, ÆSTHÉSIOGÈNES, PERI-AXILE, Lasègue .

**Mots en commun :** ', 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 .

**Mots fréquents :** moelle, allonger, cordon, partie, kyste, membre, gauche, côté, os, byzantin, spina, tumeur, phot, cellule, doigt, cas, saignée, médius, main, achondroplasie, hypertrichose, sembler, lésion, pi, normal, postérieur, tissu, avant-bras, rachitisme, radiographie, inférieur, neurone, substance, bras, difformité, mosaïque, art, absence, nain, vaisseau, droit, dorsal, hypertrophie, travail, gélatineux, cellulaire, trouver, volume, coupe, phalange .

**Noms cités :** Cohausen, Hermippus, Exp, DUPRÉ et DEVAUX, T. XIV, W., T. 14, Campeon, FÉRÉ, Fig, Masson, Monreale, Menzel, Jésus-Christ, Pl, HEITZ, E. rapin, SOLOVTZOFF, BIFIDA, Goli, SWITALSKI, BEAUVOIS, Potel, MEIGE, F, VASCHIDE et VURPAS, Trophoedème, Zeiss, Spillmann, Salerne, ECTROMÉLIEN HÉMIMÈLE, NOEVUS veineux et hystérie, MONSIEUX,

ARCHIVES

Fig. 6 – Points similaires entre un ouvrage de Charcot et celui de de la Tourette.

# TextPair – alignement de textes, corpus Charcot<sup>4</sup>

⚠ nombre de résultats assez conséquent – filtrage nécessaire

1

Source

Charcot, Jean-Martin • Archives de neurologie [Tome 26, n° 77-82] : revue des maladies nerveuses et mentales •

nouveaux cas de sclérosé 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

View passage in context

Hide differences

View passage in context

Target

Gilles de la Tourette, Georges • *Nouvelle iconographie de la Salpêtrière [Tome 23] : iconographie médicale et artistique* •

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. Lennmalm. Bidrag till Kannedomen om den amyotrofiska laleralisklerosen., Upsala lékareförebund, 1887, n° 7. Analysé in Neurol. Centralbl, 1881, p. 550.

Browse by Metadata Counts

Source
Passage
Author
Title
Year
Passage Length
Target
Passage
Author
Title

Fig. 7 – 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*.

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

# 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 ; —
</o>
```

Fig. 9 – Concepts médicaux, document XML.

Fig. 8 – 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)?

Fig. 10 – Liste finale des concepts médicaux.

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

- 1 À propos du projet
- 2 Problématique
- 3 Méthodologie
- 4 Résultats**
- 5 Conclusion



# 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	<b>0,44</b>
Sclérose latérale	127	0,09	0,37	0,41
Sclérose en plaque disséminées	12	0,02	<b>0,83</b>	0,40
Somnambulisme	3410	<b>1</b>	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%)

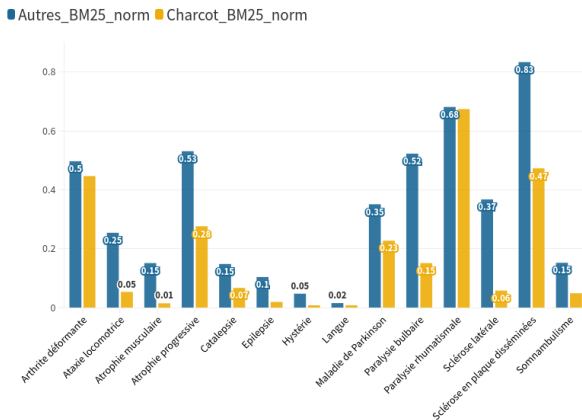


Fig. 11 – Pertinence des concepts dans les deux corpus (BM25).

# Types de concepts extraits avec BERT

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

(VASWANI et al., 2017)

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	

# 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

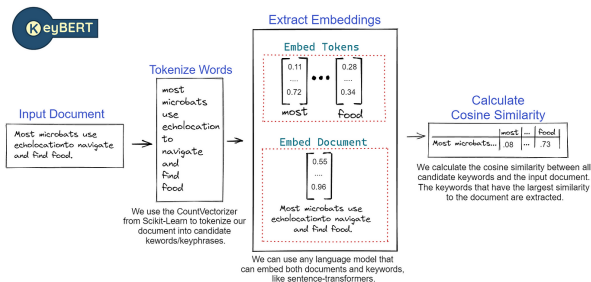


Fig. 12 – Pipeline de la librairie keybert (GROOTENDORST, 2020).

# Limitations de keybert

⚠ manque de diversification des résultats + (non-)grammaticalité  
2 termes communs : « articulations de épaule », « paralysie faciale périphérique »

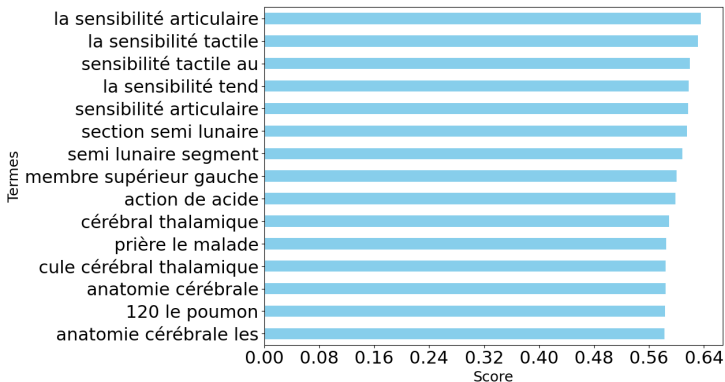


Fig. 13 – Répartition des 15 termes les plus pertinents dans le corpus « Autres » selon keybert.

# 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

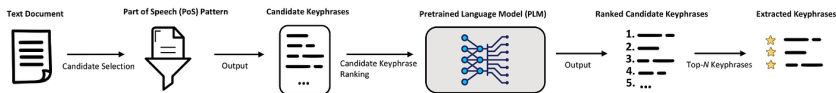


Fig. 14 – Workflow de la méthode *PatternRank* (SCHOPF et al., 2022).

# Les termes partagés les plus fréquents | keyphrase-vectorizers

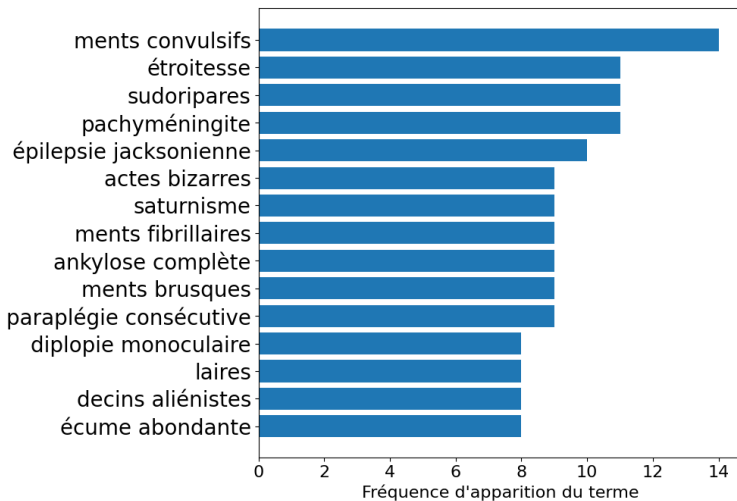


Fig. 15 – Les 15 termes les plus fréquents dans les deux corpus selon keyphrase-vectorizers.

## 5 Conclusion



# Évolution du projet

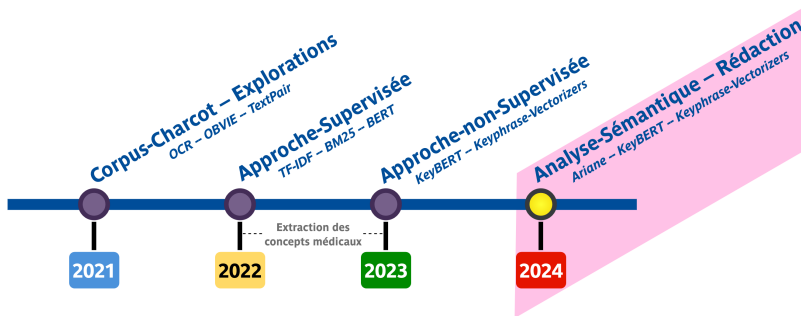


Fig. 16 – Méthodes computationnelles déjà expérimentées et à expérimenter.

→ histoire des concepts (alem. *Begriffsgeschichte*) (KOSELLECK et RICHTER, 2011)

# 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)*. <https://hal.science/hal-03543362/> (voir p. 12).



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. 7).



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. 7).



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. 15).



FEDCHENKO, V., D. M. NICOLosi et G. ROE (2024). À la recherche des réseaux intertextuels : défis de la recherche littéraire à grande échelle. In : *Humanités numériques* 9 (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, p. 1-7 (voir p. 7).

## Références II



GROOTENDORST, M. (2020). *KeyBERT : Minimal keyword extraction with BERT*. Version v0.3.0 (voir p. 20).



JOYEUX-PRUNEL, B. (2019). Visual Contagions, the Art Historian, and the Digital Strategies to Work on Them. In : *Artl@s Bulletin* 8.3, p. 128-144 (voir p. 7).



KOSELLECK, R. et M. RICHTER (2011). Introduction and Prefaces to the *Geschichtliche Grundbegriffe* : (Basic Concepts in History : A Historical Dictionary of Political and Social Language in Germany). In : *Contributions to the History of Concepts* 6.1, p. 1-37 (voir p. 25).



MANJAVACAS, E., B. LONG et M. KESTEMONT (2019). On the Feasibility of Automated Detection of Allusive Text Reuse. In : *Proceedings of the 3<sup>rd</sup> 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).



NERIMA, L., V. SERETAN et E. WEHRLI (jan. 2006). Le problème des collocations en TAL. In : *Nouveaux cahiers de linguistique française* 27, p. 95-115 (voir p. 12).



PETKOVIC, L., M. ALRAHABI et G. ROE (2023). Circulation du discours médical de Jean-Martin Charcot. In : *Humanistica* 2023 (voir p. 7).



RIFFATERRE, M. (1980). La trace de l'intertexte. In : *Pensée (La) Paris* 215, p. 4-18 (voir p. 10).



RIGUET, 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* (voir p. 7).

# Références III



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 14<sup>th</sup> International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. SCITEPRESS – Science et Technology Publications (voir p. 22).



SOULET, A. (2024). *Vers l'analyse à la demande des connaissances de Wikidata*. Journée Intelligence Artificielle et Humanités Numériques, DataLab, BnF (voir p. 7).



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, 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 400<sup>th</sup> anniversary of his birth. In : *Arquivos de Neuro-Psiquiatria* 80, p. 759-762 (voir p. 7).



VASWANI, A., N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES, A. N. GOMEZ, L. KAISER et I. POLOSUKHIN (2017). *Attention Is All You Need*. arXiv : 1706 . 03762 [cs . CL] (voir p. 19).