

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

**Ljudmila PETKOVIC<sup>1,2,3</sup>**, **Motasem ALRAHABI<sup>1,3</sup>**, **Glenn ROE<sup>1,2,3</sup>**

`prenom.nom@sorbonne-universite.fr`

<sup>1</sup> Sorbonne Université, Faculté des Lettres, UFR Littératures françaises et comparée

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

<sup>3</sup> Observatoire des textes, des idées et des corpus (ObTIC)

Séminaire doctoral du CERES  
Maison de la Recherche, Sorbonne Université  
Paris, le 27 mars 2024



# Plan

1. Contexte de recherche
2. Problématique et objectifs
3. Méthodologie de recherche
4. Résultats
5. Conclusion et recherches futures

## 1. Contexte de recherche

## 2. Problématique et objectifs

## 3. Méthodologie de recherche

## 4. Résultats

## 5. Conclusion et recherches futures

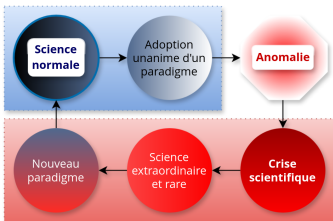
# Le progrès de la science comme un processus discontinu

## Rupture épistémologique

Passage radical d'un paradigme <sup>a</sup> à un autre, dans la façon dont les scientifiques abordent un domaine donné.

a. découverte scientifique universellement reconnue.

- « rupture entre observation et expérimentation » (BACHELARD, 1934)
- « révolution scientifique » (KOYRÉ, 1957)
- « changement de paradigme » (KUHN, 1962)



Conception kuhnienne du progrès scientifique, adaptée de AMIRI (2012).

Le nouveau paradigme est incompatible avec le précédent, p. ex. géo- vs. héliocentrisme.

# Évolution du terme *hystérie*

## Exemple du changement de paradigme : le terme d'HYSTÉRIE

- gr. ὑστέρα, lat. *hystera* : « utérus », « matrice »

Période	Sexe	Étiologie
Antiquité	♀	déplacement de l'utérus, selon Hippocrate (TASCA et al., 2012) <i>hystérique</i> : (femme) malade de l'utérus
Moyen Âge	♀	possession démoniaque, superstition religieuse de l'Église → chasses, tortures, exorcismes (TASCA et al., 2012)
Renaissance	♀/♂	localisation dans le cerveau, <i>sensorium commune</i> (LE POIS, 1618) « siège commun de la sensibilité » <sup>1</sup> , ensemble des perceptions
Lumières	♀/♂	explosion des « esprits animaux » dans le cerveau maladie convulsive (WILLIS, 1667/1681) <sup>2</sup>
XIX <sup>e</sup> s.	♀/♂	dégénérescence héréditaire du système nerveux (CHARCOT, 1870) maladie systématiquement traitée comme un trouble neurologique

1. KANT, 1863.

2. créateur du terme *neurologia* en 1664.

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



Source : [Wikipedia](#).

## JEAN-MARTIN CHARCOT (1825-1893)

- père de la neurologie moderne et française au XIX<sup>e</sup> s.
- leçons cliniques du mardi à l'hôpital de la Salpêtrière à Paris  
« Mecque de la neurologie »

### ● Contributions majeures :

- hystérie : résultat d'une lésion dynamique des circuits cérébraux
- hypnose : analyse des symptômes hystériques et outil thérapeutique
- SEP<sup>3</sup> disséminée (description) → sclérose multiple
- SLA<sup>4</sup> (description) → maladie de Charcot / Lou Gehrig
- maladie de Parkinson : concepteur du terme (avec A. Vulpian)

(CAMARGO et al., 2024)

---

3. abbr. *sclérose en plaques*.

4. abbr. *sclérose latérale amyotrophique*.

# Impact de Charcot sur sa discipline et au-delà

## Collaborateurs et élèves

Sigmund FREUD (1856-1939)	théorie psychanalytique
Gilles DE LA TOURETTE (1857-1932)	syndrome de Tourette
Joseph BABINSKI (1857-1904)	pithiatisme, signe de Babinski
Pierre JANET (1859-1947)	psychopathologie

(BROUSSOLLE et al., 2012)

## Littéraires

- références à Charcot et descriptions de crises hystériques dans la littérature naturaliste française et européenne

Émile ZOLA (1840–1902)	<i>Lourdes</i>
Léon TOLSTOÏ (1828–1910)	<i>La Sonate à Kreutzer</i>
Luigi CAPUANA (1839–1915)	<i>La Torture</i>
Bjørnstjerne BJØRNSON (1832–1910)	<i>Over Ævne</i>

(KOEHLER, 2013)

1. Contexte de recherche

2. Problématique et objectifs

3. Méthodologie de recherche

4. Résultats

5. Conclusion et recherches futures



# Circulation du discours médical au prisme du numérique

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

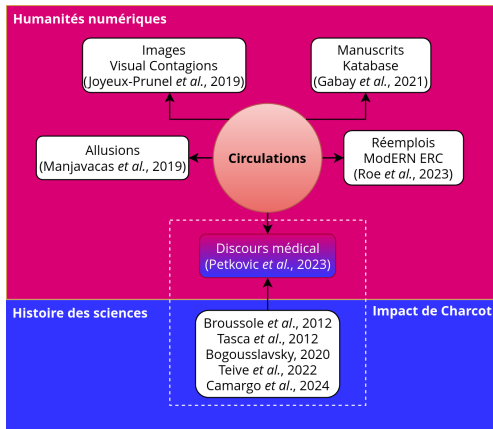


Fig. 1 – Études (numériques) des circulations des savoirs.

# Question de recherche

Comment mesurer le degré d'intertextualité entre Charcot et son réseau scientifique et/ou artistique au prisme du numérique ?

1. Contexte de recherche
2. Problématique et objectifs
3. Méthodologie de recherche
4. Résultats
5. Conclusion et recherches futures

# Mesurer le degré d'intertextualité

(Traductologie) Mesurer l'influence d'un écrivain sur le style de son traducteur  
(OSEKI-DÉPRÉ, 2007).

→ mesurer informatiquement l'impact de Charcot sur son réseau

- analyse des concepts-clés médicaux
- intertextualité uni-directionnelle CHARCOT → AUTRES

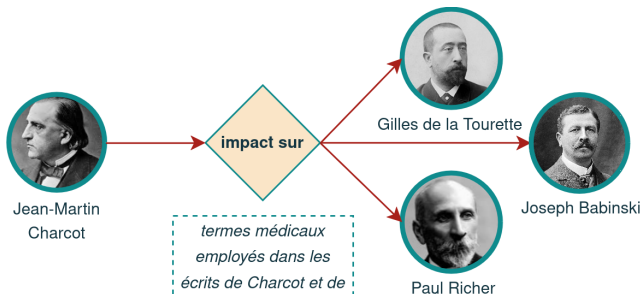


Fig. 2 – Opérationnalisation de l'impact de Charcot sur ses élèves.

# Fonds Charcot<sup>5</sup>

## SorbonNum – Bibliothèque de Sorbonne Université (BSU)

201 documents XML OCRisés (sans post-correction)

Répartition du corpus :

- Charcot : textes rédigés par Charcot et ses collègues
- Autres : textes rédigés par ses collègues

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>

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

# Première analyse du corpus Charcot

## OBVIE<sup>6</sup>

- moteur de recherche pour la fouille avancée des corpus en XML-TEI
- identification des substantifs les plus importants
  - fréquences brutes, mesures TF-IDF, BM25,  $\chi^2$ , Test Gamma
- repérage des textes similaires par ordre de pertinence
  - à partir des termes en commun et termes fréquents

---

6. <https://obtic.huma-num.fr/obvie/>

## Deuxième analyse du corpus Charcot

### TextPair<sup>7</sup>

- aligne des passages similaires dans une collection de textes
  - passages incluant des citations, plagiats, emprunts, réemplois...
- génère une liste de séquences similaires pour chaque texte
  - séquences de mots qui se chevauchent (trigrammes de mots)
- compare les séquences générées du texte *source* au texte *cible*

---

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

# keybert

Librairie Python qui exploite les plongements de mots BERT pour générer des mots/phrases-clés les plus similaires à un document.

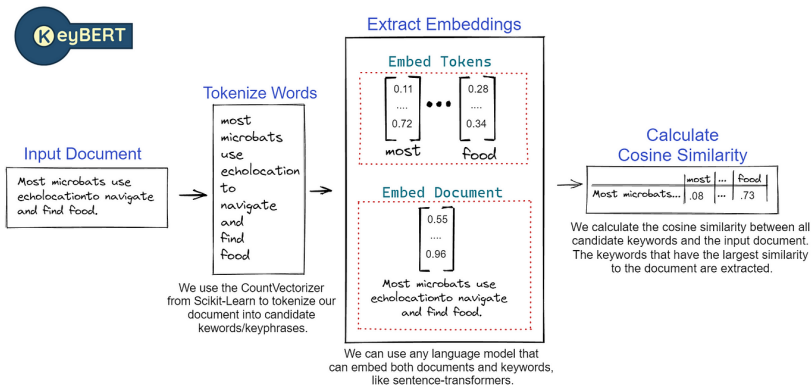


Fig. 3 – Pipeline de la librairie keybert (GROOTENDORST, 2020).



# Améliorer les sorties de keybert

## Limitations de keybert :

- on doit spécifier la longueur (optimale) des n-grammes à extraire
- la grammaticalité des phrases n'est pas prise en compte  
p. ex. *spinal les muscles*
- mots très similaires dans la liste des mots-clés  
*réflexes tendineux* sont, 0.468  
*les réflexes tendineux*, 0.4615

## Alternative : réglage fin des sorties

- diversification des résultats : *Maximal Marginal Relevance*
  - « pertinence marginale maximale » : combinaison linéaire de la pertinence par rapport à la requête et la nouveauté
  - un document a une PMM s'il répond à la requête et contient peu de similarité avec les documents précédents ([BOUTIN, 2006](#))
- préservation de la grammaticalité de la phrase (motifs POS)  
→ keyphrase-vectorizers ([SCHOPF, KLIMEK et MATTHES, 2022](#))

1. Contexte de recherche

2. Problématique et objectifs

3. Méthodologie de recherche

4. Résultats

5. Conclusion et recherches futures

OBVIE – corpus Charcot<sup>8</sup>

⚠ impossible de quantifier l'importance des phrasèmes

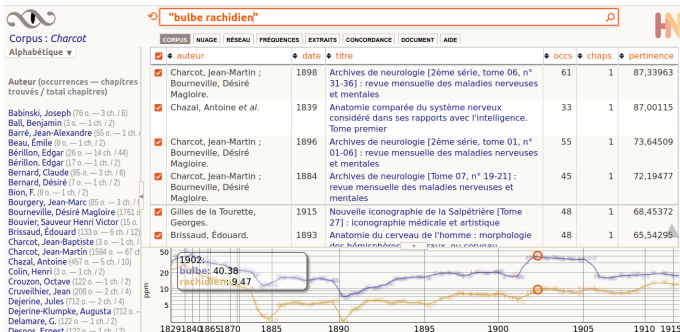


Fig. 4 – Distribution d'occurrences des tokens avec la frise chronologique pour ceux constituant l'expression *bulbe rachidien*.

# TextPair – corpus Charcot<sup>9</sup>

⚠ nombre de résultats parfois assez conséquent → filtrage

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'Ozzonatzomaazie (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 lékarefbreu för, 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. 5 – Réemploi du terme *sclérose latérale amyotrophique* dans les textes de Charcot et de de la Tourette (le seul résultat).

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

# Liste des concepts médicaux

Extraction semi-automatique des termes en lien avec Charcot (*hystérie, sclérose latérale amyotrophique* etc.)

- index d'une édition des œuvres complètes de Charcot<sup>10</sup>
- extraction des termes à l'intérieur des balises XML <s> jusqu'à la virgule ou à un tiret avec des regex
- post-traitement de la liste : élimination des termes génériques (*os, cerveau, peau, etc.*)
- prise en compte des formes singulier / pluriel avec des regex

---

10. CHARCOT, 1892

# TF-IDF<sup>11</sup>

Mesure pour quantifier l'importance ou la pertinence des représentations lexicales (mots, phrases, lemmes...) dans un document parmi une collection de documents (corpus).

- **TF** : fréquence d'un terme particulier par rapport au document
- **IDF** : calcule à quel point un terme est courant (ou rare) dans le corpus
  - pénalise des termes fréquents et récompense les termes peu fréquents (considérés comme plus discriminants)

---

11. angl. *term frequency-inverse document frequency*.

BM-25<sup>12</sup>

Modèle de classement basé sur des termes qui vise à fournir des résultats de recherche précis et pertinents en classant les documents en fonction de la fréquence de leurs termes et de leur longueur.

- 1 TF
- 2 IDF
- 3 normalisation de la longueur de plusieurs documents
- 4 saturation des termes de requête

TF-IDF et BM25 sont les mesures couramment utilisées dans le domaine de recherche d'information (angl. *information retrieval*)  
→ mathématiquement assez proches, surtout si appliquées sur un seul document.

---

12. angl. *best match 25*.

# Intensification du lexique de Charcot dans le corpus «Autres»

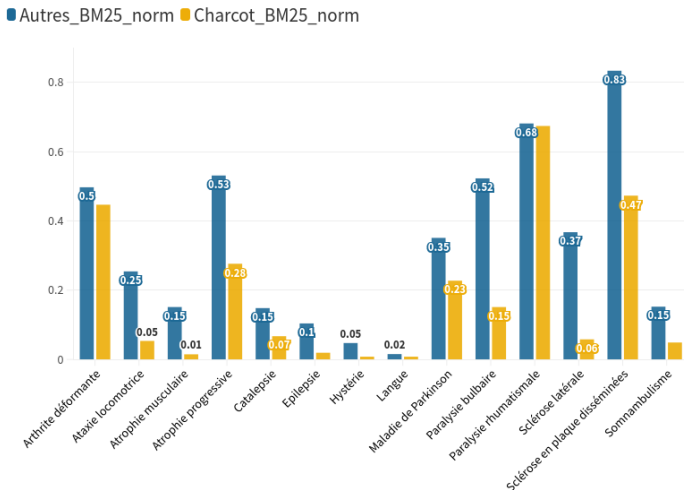


Fig. 6 – Pertinence des concepts dans les deux corpus (BM25).



# BERT

## VASWANI et al., 2017

- plongements lexicaux et des mécanismes d'attention
- modèle bert-base-multilingual-cased
- censé bien capturer la sémantique

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

## Calcul de pertinence des concepts : corpus «Charcot»

Terme	Corpus «Charcot»			
	Fréquence	TF-IDF	BM25	BERT
Arthrite déformante	30	0.16	0.45	0.80
Ataxie locomotrice	559	0.35	0.05	0.83
Atrophie musculaire	1 105	0.20	0.02	0.84
Atrophie progressive	40	0.14	0.27	0.72
Catalepsie	681	<b>0.54</b>	0.07	0.88
Épilepsie	414	0.09	0.02	0.78
Hystérie	5 775	0.51	0.01	0.74
Langue	2 695	0.24	0.01	0.72
Maladie de Parkinson	75	0.21	0.23	0.81
Paralysie bulbaire	149	0.27	0.15	0.89
Paralysie rhumatismale	8	0.07	0.67	0.86
Sclérose latérale	445	0.30	0.06	0.88
Sclérose en plaque disséminées	45	0.25	0.47	0.87
Somnambulisme	847	0.49	0.05	0.89

Table 1 – Calcul de pertinence des concepts selon TF-IDF, BM25 et BERT, corpus «Charcot».

## Calcul de pertinence des concepts : corpus « Autres »

Terme	Corpus « Autres »			
	Fréquence	TF-IDF	BM25	BERT
Arthrite déformante	24	0.02	<b>0.50</b>	0.40
Ataxie locomotrice	169	0.08	0.25	0.39
Atrophie musculaire	1 465	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	4 934	0.45	0.05	0.41
Langue	3 591	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	<b>0.68</b>	0.44
Sclérose latérale	127	0.09	0.37	0.41
Sclérose en plaque disséminées	12	0.02	<b>0.83</b>	0.40
Somnambulisme	3 410	1	0.15	0.43

Table 2 – Calcul de pertinence des concepts selon TF-IDF, BM25 et BERT, corpus « Autres ».

# keybert

bla

1. Contexte de recherche
2. Problématique et objectifs
3. Méthodologie de recherche
4. Résultats
5. Conclusion et recherches futures

# Vers une lecture plus distante du corpus Charcot

## Premières explorations du corpus Charcot

- recherche avancée et alignement de textes (OBVIE, TextPair)  
analyse de textes assistée par ordinateur
- manque de fonctionnalités pour mesurer l'impact de Charcot sur son réseau *via* les concepts de ses travaux médicaux
  - [→] recherche d'un outil de « lecture distante »

## Une nouvelle approche

- quantification de la pertinence des concepts polylexicaux dans les corpus, selon trois différentes métriques de pondération
- repérage des phénomènes lexicaux grâce aux visualisations (validation auprès de spécialistes de Charcot nécessaire)

# Perspectives

- ① Charcot vs. Autres : initiateur ou transmetteur de certains termes?
- ② analyse sémantique des passages contenant ces concepts → modalités de prise en charge énonciative
  - opinions, accords, désaccords, définitions, etc.
- ③ post-correction automatique d'OCR avec la librairie `neuspe11`  
(JAYANTHI, PRUTHI et NEUBIG, 2020)
  - évaluation de son impact sur des tâches en aval
- ④ modélisation de sujets assistée par l'apprentissage automatique  
(GROOTENDORST, 2022)
- ⑤ OCRiser les manuscrits (« leçons ») de Charcot avec eScriptorium<sup>13</sup>

---

13. <https://escriptorium.isir.upmc.fr/>

# Données et scripts

Dépôts GitHub :

- Tracking the circulation of Jean-Martin Charcot's medical discours...
- Extraction des mots-clés à partir des textes



# Remerciements

Un grand merci à Valentina Fedchenko (ingénieure de recherche de l'équipe-projet ObTIC) et à Simon Gabay (maître-assistant de la Chaire des humanités numériques à l'Unige) pour leurs conseils précieux.

# Références I



AMIRI, V. V. (2012). *T.S. Kuhn*. (Voir p. 4).



BACHELARD, G. (1934). *La formation de l'esprit scientifique : contribution à une psychanalyse de la connaissance*. Vrin (voir p. 4).



BOGOUSLAVSKY, Julien (2020). The mysteries of hysteria : a historical perspective. In : *International Review of Psychiatry* 32.5-6, p. 437-450 (voir p. 9).



BOUTIN, E. (2006). Biais cognitifs et recherche d'information sur internet. Quelles perspectives pour les indicateurs de pertinence des moteurs de recherche. In : *VSST 2006*, p. 1-7 (voir p. 17).



BROUSSOLLE, E. et al. (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. et al. (2024). Jean-Martin Charcot : the polymath. In : *Arquivos de Neuro-psiquiatria* 81, p. 1098-1111 (voir p. 6).



CHARCOT, J.-M. (1870). *Manuscrit n° 3 : leçons (Manuscrits des leçons et observations de J.M. Charcot)*. Bibliothèque de Sorbonne Université (BSU) – Bibliothèque Charcot (voir p. 5).



— (1892). *Œuvres complètes de J.-M. Charcot : Leçons sur les maladies du système nerveux*. T. 1. Paris : Bureaux du Progrès médical (voir p. 21).



GABAY, S. et al. (2021). Katabase : À la recherche des manuscrits vendus. In : *Humanistica* 2021 (voir p. 9).

## Références II



GROOTENDORST, M. (2022). BERTopic : Neural topic modeling with a class-based TF-IDF procedure. In : *arXiv preprint arXiv :2203.05794* ([voir p. 31](#)).



— (2020). *KeyBERT : Minimal keyword extraction with BERT*. Version v0.3.0 ([voir p. 16](#)).



JAYANTHI, S. M., D. PRUTHI et G. NEUBIG (2020). Neuspell : A Neural Spelling Correction Toolkit. In : *arXiv preprint arXiv :2010.11085* ([voir p. 31](#)).



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. 9](#)).



KANT, É. (1863). *Anthropologie d'un point de vue pragmatique*, (trad. J. Tissot). Librairie Ladrangé (originellement publié en 1798) ([voir p. 5](#)).



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. 7](#)).



KOYRÉ, A. (1957). *From the closed world to the infinite universe*. T. 1. Baltimore, Johns Hopkins Press ([voir p. 4](#)).



KUHN, T. S. (1962). *La structure des révolutions scientifiques*. Traduit par Laure Meyer. Flammarion Paris ([voir p. 4](#)).



LE POIS, C. (1618). *Selectiorum observationum et consiliorum de praetervis hactenus morbis affectibusque praeter naturum, ab aqua seu serosa colluvie et diluvie ortis, liber singularis*. Authore Carolo Pisone, Ponte ad Monticulum, apud Carolum Mercatorem ([voir p. 5](#)).

## Références III



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. 9).



MARMION, J.-F. (2015). *Freud et la psychanalyse*. Sciences Humaines (voir p. 6).



OSEKI-DÉPRÉ, I. (2007). *De Walter Benjamin à nos jours : essais de traductologie*. Paris : Honoré Champion (voir p. 12).



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 - KDIR*. INSTICC. SciTePress, p. 243-248 (voir p. 17).



TASCA, C. et al. (2012). Women and hysteria in the history of mental health. In : *Clinical practice and epidemiology in mental health : CP & EMH* 8, p. 110 (voir pp. 5, 9).



VASWANI, A. et al. (2017). Attention Is All You Need. In : CoRR abs/1706.03762. arXiv : 1706.03762 (voir p. 25).



WILLIS, T. (1667/1681). *Pathologiae cerebri et nervosi generis specimen*. An essay of the pathology of the brain and nervous stock in which convulsive diseases are treated of (voir p. 5).