

bbb

**Ljudmila PETKOVIC<sup>1,2,3,4</sup>**

`prenom.nom@sorbonne-universite.fr`

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

Séminaire doctoral ObTIC  
SCAI, salle du Conseil  
Paris, le 13 mars 2025



# 1 Contributions principales de Charcot

## 2 Approches comparées

## 3 Conclusion

## Termes inventés par Charcot : référence

paralysie agitante  
 ataxie locomotrice progressive  
 arthropathies tabétiques  
 trépidation épileptoïde du pied  
 sclérose en plaques disséminées  
 sclérose latérale amyotrophique  
 idée(s) fixe(s), maladie des tics  
 mouvements involontaires  
 incapacité d'être debout / de marcher  
 atrophie musculaire progressive

**maladie de Parkinson**  
***tabes dorsalis***  
**arthropathie de Charcot**  
**clonus**  
**sclérose multiple**  
**maladie de Charcot / Lou Gehrig**  
**syndrome de Tourette**  
**chorées, athétose**  
**astasie-abasie**  
**maladie Charcot-Marie-Tooth**

(WALUSINSKI, 2025; CAMARGO et al., 2024)

≠ termes transmis : **hystérie**  
**épilepsie**  
**hypnose**

- 1 Contributions principales de Charcot
- 2 Approches comparées
- 3 Conclusion

# Approches comparées

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

Traitements effectués en local (1,2) et *via* la plateforme MeSU<sup>a</sup> (3).

---

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

# Critères de comparaison des approches

- prise en compte des synonymes des termes
  - ex. *paralysie agitante* → *maladie de Parkinson*
- recenser le score le plus élevé sur le terme ou sur son synonyme
- les méthodes classiques vs. celles de l'état de l'art

# Le domaine potentiellement impactant : **syndrome de Tourette**

Terme	TF-IDF (TermSuite)	TF-IDF	BM25	PatternRank
<i>maladie de Parkinson</i>	0,05	0,0775	0,333	0,7936
<i>ataxie locomotrice progressive</i>	0,32	0,0386	0,4877	0,7431
<i>arthropathies tabétiques</i>	0,33	0,0934	0,4928	0,7506
<i>trépidation épileptoïde du pied</i>	0,0198	0,1227	0,2919	0,7597
<i>sclérose en plaques disséminées</i>	NA	0,178	0,8089	NA
<i>tremblement</i>	NA	0,1686	0,0362	0,7683
<i>nystagmus</i>	0,0243	0,1326	0,146	0,7474
<i>embarras parole</i>	NA	NA	0,0018	0,9347
<i>sclérose latérale amyotrophique</i>	NA	0,044	0,6586	NA
<b><i>tics convulsifs</i></b>	<b>NA</b>	<b>0,1293</b>	<b>0,8385</b>	<b>0,8331</b>
<i>atrophie musculaire progressive</i>	0,40	0,1118	0,3489	0,8053
<i>aphasie</i>	0,0587	0,2245	0,1334	0,7960
<i>astisie-abasie</i>	NA	0,0478	0,3565	0,7375
<i>athétose</i>	NA	0,2029	0,274	0,8068
<i>chorées</i>	NA	0,1336	0,0701	0,8047
<i>hystérie</i>	0,2724	0,3711	0,0442	0,8018
<i>épilepsie</i>	NA	0,164	0,0247	0,8199
<i>hypnose</i>	0,3543	1	0,2922	0,7738

Tab. 1 – Les scores de pertinence pour les termes de référence à partir du corpus « Autres ».

# Analyse comparative des approches employées

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

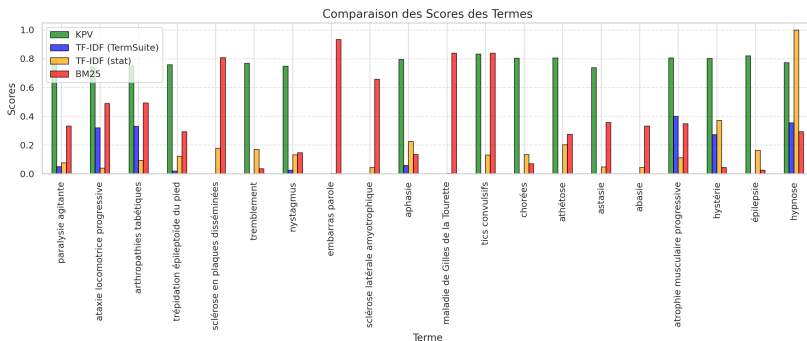


Fig. 1 – Visualisation des scores de pertinences pour chaque terme de référence



# Chronologie d'une locution : indice de croissance de l'impact?

- évolution de la fréquence des termes au sein des deux corpus
- convergence entre des termes : fin XIX<sup>e</sup>, début XX<sup>e</sup> s.
  - ppm : nombre d'occurrences par million de mots

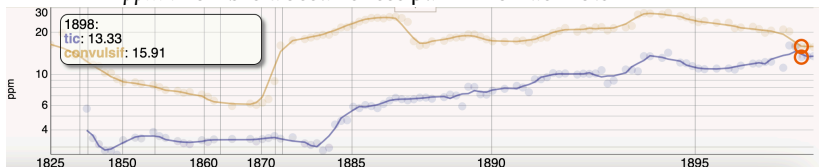


Fig. 2 – Chronologie de la fréquence du terme *tic convulsif*.

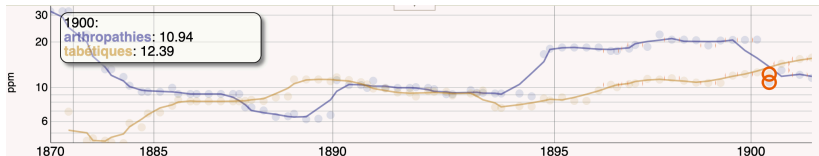


Fig. 3 – Chronologie de la fréquence du terme *arthropathies tabétiques*.

- 1 Contributions principales de Charcot
- 2 Approches comparées
- 3 Conclusion**

# Conclusion

- 1 *PatternRank* : la méthode la plus fiable par rapport aux autres méthodes?
  - 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 → syndrome de Tourette.
- 3 Absence des scores pour les termes comme *SEP* et *SLA* :
  - solution : chercher leurs symptômes ou leurs descriptions :
    - *amyotrophie spinale progressive, secousses nystagmiques...*

# Références I



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. <https://www.thieme-connect.de/products/ejournals/pdf/10.1055/s-0043-1775984.pdf>, p. 1098-1111 (voir p. 3).



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. 5).



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:TCuXWzGHjo31RdxGR9jECRG2rZzqvOK3G0zHF7yAa2NfxtdFqxe-MmSHMC6e80FiFxI4sLj2aW60yDk](https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.4630270302?casa_token=TfyVkMGkDQsAAAAA:TCuXWzGHjo31RdxGR9jECRG2rZzqvOK3G0zHF7yAa2NfxtdFqxe-MmSHMC6e80FiFxI4sLj2aW60yDk), p. 129-146 (voir p. 5).



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*. <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 p. 5).

## Références II



WALUSINSKI, O. (2025). *Jean-Martin Charcot's Birth Bicentennia*. <http://ishn.org/>. Consulté le 13 mars 2025 (voir p. 3).