

Comparaison des approches pour l'extraction de phrases-clés

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

Séminaire doctoral ObTIC
SCAI, salle du Conseil
Paris, le 11 avril 2025



Table des matières

- 1. Contributions de Charcot
- 2. Évaluation quantitative
- 3. Évaluation qualitative
- 4. Conclusion

- 1. Contributions de Charcot
- 2. Évaluation quantitative
- 3. Évaluation qualitative
- 4. Conclusion

Termes inventés par Charcot : référence

paralysie agitante	maladie de Parkinson
ataxie locomotrice progressive	tabes dorsalis
arthropathies tabétiques	arthropathie de Charcot
trépidation épileptoïde du pied	clonus
sclérose en plaques disséminées	sclérose multiple
sclérose latérale amyotrophique	maladie de Charcot / Lou Gehrig
idée(s) fixe(s), maladie des tics convulsifs	hystérie et syndrome de Tourette
mouvements involontaires	chorées, athétose
incapacité d'être debout / de marcher	astasie-abasie
atrophie musculaire progressive	maladie Charcot-Marie-Tooth

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

≠ termes transmis : hystérie
épilepsie
hypnose

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	●○○○	○○○○○○○	○○	

- Contributions de Charcot
- Évaluation quantitative
- Évaluation qualitative
- Conclusion

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	○●○○	○○○○○○○	○○	
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

Ljūdmila PETKOVIĆ	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral ObTIC, 13/03/2025	5 / 20
-------------------	---	--------------------------------------	--------

Ljūdmila PETKOVIĆ	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral ObTIC, 13/03/2025	6 / 20
-------------------	---	--------------------------------------	--------

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	○○●○	○○○○○○○	○○	
Le domaine potentiellement impactant : hystérie					

Terme	TF-IDF (TermSuite)	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 épileptique 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>astésie-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. 1 – Les scores de pertinence pour les termes de référence à partir du corpus «Charcot».

Moyenne pour tous les termes combinés : 0,3280

Ljūdmila PETKOVIĆ	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral ObTIC, 13/03/2025	7 / 20
-------------------	---	--------------------------------------	--------

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	○○○●	○○○○○○○	○○	
Autre domaine potentiellement impactant : 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 épileptique 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
<i>tics convulsifs</i>	NA	0,1293	0,8385	0,8331	0,4502¹
<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>astésie-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. 2 – Les scores de pertinence pour les termes de référence à partir du corpus «Autres».

Moyenne pour tous les termes combinés : 0,3214

¹ Le score le plus élevé parmi les termes inventés par Charcot et Inoué.

Ljūdmila PETKOVIĆ	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral ObTIC, 13/03/2025	8 / 20
-------------------	---	--------------------------------------	--------

1. Contributions de Charcot

2. Évaluation quantitative

3. Évaluation qualitative

4. Conclusion

- on utilise un terme dans le contexte où on cite Charcot
 - terme exact
 - synonyme ou terme relevant du champ conceptuel en question
 - ensembles de mots liés à un certain concept (COSTACHESCU, 2024)
pied tabétique → forme particulière d'arthropathies tabétiques

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

Tab. 3 – Concordance des termes médicaux faisant référence à Charcot – corpus Autres.

Terme	Contexte
tremblement	[...] Charcot présentait, dans son amphithéâtre, pour démontrer les caractères oscillatoires des diverses variétés de tremblements.
nystagmus	SCLÉROSE EN PLAQUES [...] [...], nystagmus, [...]. 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 cylindrxes [...]
embarras parole	Lorsqu'on se trouve en présence d'un malade ayant de l'embarras de la parole [...] A. La réponse à la première proposition n'est nullement embarrassante, si l'on veut se rappeler ces paroles de M. le professeur Charcot : [...]
sclérose latérale amyotrophique	[...] sclérose latérale amyotrophique, maladie découverte par mon illustre maître Charcot.
tics convulsifs	désignée par M. Charcot sous le nom de maladie des tics convulsifs
atrophie musculaire progressive	Charcot et Marie ont décrit la « forme particulière d'atrophie musculaire progressive »
aphasie	Lorsqu'il y a, dit M. le professeur Charcot (1), suppression de la mémoire pour l'articulation des mots, c'est l'aphasie motrice d'articulation ou aphasie de Broca qui se présente.

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

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' <u>astasie</u> et d' <u>abasie</u> . 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'hystérie traumatique [...]
<i>épilepsie</i>	M. Charcot a décrit avec le plus grand soin l'épilepsie partielle d'origine syphilitique [...]
<i>hypnose</i>	Les trois états de l'hypnose décrits par M. Charcot sont devenus classiques, [...]
<i>systématisation de l'organisation de la moelle 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. 5 – Concordance des termes médicaux faisant référence à Charcot – corpus Autres (fin).

- 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

Terme	Cooccurrent	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	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
systématisation de l'organisation de la moelle épinière	NA	NA	NA	NA	NA
localisations cérébrales	Charcot	2 968	9	7	5,3

Tab. 6 – Analyse des cooccurrences des termes médicaux à partir du corpus Autres.

Terme	Contexte
<i>trépidation épileptoïde du pied (clonus)</i>	On [Babinski] le désigne alors sous la dénomination de « clonus du pied », « trépidation épileptoïde du pied »
<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 nystagmus 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 troubles fonctionnels de la parole et notamment sur le bégaiement.
<i>athétose</i>	[...] nous avons affaire au symptôme désigné par M. W. Hammond sous le nom d'athétose.
<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'épilepsie à laquelle il a attaché son nom, [...]
<i>hypnose</i>	Selon Braid , l'hypnose est caractérisée par des phénomènes mentaux et physiques, particuliers à cette condition.

Tab. 7 – Concordance des termes médicaux faisant référence à d'autres médecins – corpus Autres.

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	○○○○	○○○○○○○○	●○	

- 1. Contributions de Charcot
- 2. Évaluation quantitative
- 3. Évaluation qualitative
- 4. Conclusion

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	○○○○	○○○○○○○○	○●	

Conclusion

- màj nécessaire...
- ① les termes les plus impactants
 - pour Charcot : *hystérie*
 - pour les Autres : *syndrome de Tourette*
 - ② absence de cooccurrent Charcot explicable :
 - épilepsie : terme créé par J. H. Jackson
 - hypnose : terme créé par J. Braid
 - [...]
 - les résultats sont alignés avec les faits historiques

Ljūdmila PETKOVIC	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral OeTIC, 13/03/2025	17 / 20
-------------------	---	--------------------------------------	---------

Ljūdmila PETKOVIC	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral OeTIC, 13/03/2025	18 / 20
-------------------	---	--------------------------------------	---------

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	○○○○	○○○○○○○○	○○	

Références I

 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 p. 4*).

 COSTACHESCU, A. (2024). *Les champs linguistiques*. <https://www.lingvistica.ro/imagini/Champs-semantiques.pdf>. Consulté le 11 avril 2025 (*voir p. 10*).


 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.


 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:TCuXWzGHjo31RdxGR9jECRG2rZzqvOK3G0zHF7yAa2NfxtdDFqxe-MmSHMC6e80FiFxI4sLj2aW60yDk, p. 129-146.

Ljūdmila PETKOVIC	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral OeTIC, 13/03/2025	19 / 20
-------------------	---	--------------------------------------	---------

Table des matières	Contributions de Charcot	Évaluation quantitative	Évaluation qualitative	Conclusion	Références
○	○○	○○○○	○○○○○○○○	○○	

Références II

 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>.

 WALUSINSKI, O. (2025). *Jean-Martin Charcot's Birth Bicentennia*. <http://ishn.org/>. Consulté le 11 avril 2025 (*voir p. 4*).

Ljūdmila PETKOVIC	Comparaison des approches pour l'extraction de phrases-clés	Séminaire doctoral OeTIC, 13/03/2025	20 / 20
-------------------	---	--------------------------------------	---------