UNIVERSITÉ DE GENÈVE
Faculté des Lettres
Département de linguistique

Analyse des champs lexicaux dans le fonds patrimonial de Jean-Martin Charcot

Mémoire présenté en vue de l'obtention du

Certificat de spécialisation en Linguistique

Étudiante :
Ljudmila PETKOVIĆ
N° de matricule : 19-337-757
ljudmila.petkovic@etu.unige.ch

Directeur :
Prof. Dr Christopher LAENZLINGER

Co-directeur :
Luka NERIMA

29 octobre 2024

Table des matières

Introduction

Ce mémoire, à la jonction de l'histoire des sciences et de la linguistique computationnelle, propose une étude interdisciplinaire dont l'objectif est la valorisation numérique du fonds patrimonial de Jean-Martin Charcot, fondateur de la neurologie moderne au XIX^{e} siècle en France [1]. À ce titre, nous nous intéressons tout particulièrement à l'analyse de la genèse et de la migration du discours médical de pathologie anatomique, de neurologie et psychologie de Charcot dans les écrits réalisés en collaboration et dans les écrits de ses disciples et continuateurs. Si l'importance de ses travaux scientifiques est un sujet largement étudié du point de vue théorique (Bogousslavsky, 2011 ; Broussolle *et al.*, 2012 ; Camargo *et al.*, 2024), cet aspect reste inexploré dans une perspective quantitative.

Ce travail se veut donc un premier pas vers l'établissement de l'édition numérique du corpus Charcot issu de son fonds volumineux et d'une grande importance scientifique. D'abord, les traitements sous-jacents (notamment, la lemmatisation et l'indexation des textes par les lemmes) nous ont permis de produire une transcription interrogeable dans notre cadre de recherche grâce aux outils développés au sein de l'équipe-projet OBTIC [2]. L'objectif de cette démarche est d'y analyser le discours médical de Charcot à travers l'extraction des expressions à mots multiples (Nerima *et al.*, 2006, p. 96) [3], qui constituent potentiellement des champs lexicaux et des savoirs en circulation. Ensuite, nous comparons les textes écrits par Charcot avec ceux de ses collaborateurs et successeurs, *via* les concepts-clés liés à son discours scientifique. Nous considérons que ces concepts correspondent aux termes reflétant des contributions que Charcot a apportées à la compréhension et à la caractérisation des pathologies neurologiques (*hystérie, sclérose en plaques disséminées*, etc.).

Les expériences menées sur le transfert des concepts d'un corpus à l'autre se basent

1. Le présent travail fait également partie du projet doctoral de l'autrice en cours, dans le cadre du programme *Instituts et Initiative* de l'Observatoire des Patrimoines de l'Alliance Sorbonne Université : https://theses.fr/s382733.

2. https://obtic.sorbonne-universite.fr/

3. angl. *multi-word expressions*, se déclinant sous la forme suivante, entre autres : SUBSTANTIF + ADJECTIF + ADJECTIF. Exemple : la pathologie *sclérose latérale amyotrophique*.

sur deux mesures statistiques et l'une de l'apprentissage profond [4] . Enfin, une proposition de l'extraction des phrases-clés à l'aide de l'apprentissage profond et de l'analyse sémantique des passages contenant les concepts médicaux est formulée. Au-delà du cas de Charcot, ce travail vise à établir un protocole permettant d'appréhender la circulation de concepts de manière automatisée.

Le présent mémoire est structuré en quatre parties principales : après l'introduction, nous traçons l'évolution du progrès médical dans le cadre épistémologique, où Charcot a joué un rôle important, avant de présenter ses contributions principales (chapitre 1). Dans le chapitre 2, nous soulignons les aspects des circulations des savoirs et proposons une revue de la littérature portant sur ce sujet du point de vue numérique. Le fonds Charcot et la constitution du corpus de recherche correspondant sont abordés dans le chapitre 3. Ensuite, le chapitre 4 présente les premières tentatives d'analyse computationnelle de l'impact de Charcot sur ses élèves et collègues, ainsi que les limites de ces approches, en proposant de nouvelles méthodes pour la quantification de la pertinence des expressions polylexicales. Enfin, le dernier chapitre est consacré à la conclusion du travail et aux pistes pour des recherches futures.

4. angl. *deep learning*.

Chapitre 1 La rupture épistémologique en médecine : la notion d'*hystérie*

1.1 La rupture comme source du progrès scientifique

« Les vraies révolutions sont lentes et elles ne sont jamais sanglantes. »

— Anouilh (1956)

La science progresse en corrigeant constamment les erreurs, c'est-à-dire que les erreurs précèdent nécessairement l'établissement de la connaissance scientifique. Bien que ce processus de correction des erreurs puisse être observé de manière diachronique, il est de nature circulaire. En outre, si une doctrine devient obsolète avec le temps et l'avènement des technologies avancées permettant de recueillir de nouvelles preuves, une doctrine actuellement en vigueur deviendra tout de même à son tour obsolète à un moment [1].

Un tel cycle des observations empiriques peut être bouleversé, selon Bachelard (1934, p. 26), par la « rupture et non pas continuité entre l'observation et l'expérimentation ». Autrement dit, la rupture épistémologique survient lors d'un renversement fondamental dans la façon d'établir une connaissance dans un domaine particulier. De fait, ce phénomène caractérise une « révolution scientifique » (Koyré, 1957, p. 2), terme apparenté avec celui du « changement de paradigme », introduit par Kuhn (1962, p. 66). D'après ce dernier, les *paradigmes* désignent les « découvertes scientifiques universellement reconnues qui, pour un temps, fournissent à une communauté de chercheur·euse·x·s des problèmes types et des solutions ».

1. L'un des exemples le plus connu de l'obsolescence scientifique est sans doute le passage du modèle géocentrique de l'univers, défendu par Aristote et Ptolémée (selon lesquels la Terre est immobile au centre de l'Univers), à la conception héliocentrique de Nicolas Copernic, qui affirmait que la Terre tournait autour du Soleil.

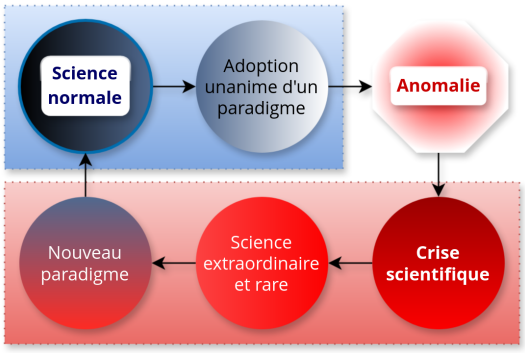

FIGURE 1.1 – Conception kuhnienne du progrès scientifique, adapt. de Amiri (2012).

Dans cette optique, la structure des révolutions scientifiques désigne un modèle épistémique constitué des épisodes non cumulatifs du développement scientifique (Figure 1.1), marqués par des passages radicaux d'un paradigme à un autre. Le nouveau paradigme ne désigne donc pas une extension de l'ancien paradigme ; au contraire, ce dernier est entièrement ou partiellement remplacé par un nouveau paradigme incompatible avec le précédent.

Cela est bel et bien un signe de l'*émergence* d'une nouvelle théorie ou découverte, tout en prouvant que le développement historique des théories est fondamentalement discontinu. Dans un esprit similaire, Bachelard (1970, p. 72) souligne :

> « Il ne saurait y avoir de vérité *première*. Il n'y a que des erreurs *premières*. On ne doit donc pas hésiter à inscrire à l'actif du sujet son expérience essentiellement malheureuse. La première et la plus essentielle fonction de l'activité du sujet est de se tromper. Plus complexe sera son erreur, plus riche sera son expérience. L'expérience est très précisément le souvenir des erreurs rectifiées. L'être pur est l'être détrompé. »

Un exemple du changement de paradigme est l'évolution du terme *hystérie*, introduit par Hippocrate dans l'Antiquité au V^e^ s. av. J.-C., qui expliquait cette maladie par un déplacement de l'utérus dans le corps féminin [2]. Au Moyen Âge, surtout à partir du XIII^e^ s., les *hystériques* étaient considérées par l'Église comme possédées par le diable et, par conséquent, chassées, torturées ou soumises aux exorcismes dans une perspective religieuse (Tasca *et al.*, 2012, p. 113). Néanmoins, certains scientifiques de la Renaissance commencent progressivement à s'éloigner de l'étiologie démonologique de cette maladie ; un cas notable est celui du médecin Charles Le Pois (1563-1633), qui fut le premier à désigner le cerveau, et plus précisément, le *sensorium commune* [3], comme siège de la maladie hystérique en 1618, en associant l'hystérie autant aux hommes qu'aux femmes (Wright, 1980, p. 235) [4].

Pour mieux comprendre l'importance de ce changement de pensée radical, il convient également de souligner que notre compréhension actuelle du système nerveux central

2. Ce terme est issu du mot grec ὑστέρα, par le latin *hystera*, « matrice ». Par dérivation, le terme « hystérique » se référait à une personne « (femme) malade de l'utérus », selon Rey (1998, p. 1767).

3. ce que Kant (1863, p. 452). appelle plus tard « siège commun de la sensibilité » pour désigner l'ensemble des perceptions.

4. Le Pois (1618, p. 101) a noté que les symptômes communément appelés hystériques se référaient à l'épilepsie, mais qu'il était prouvé que l'épilepsie elle-même était une maladie *idiopathique* (existant par elle-même, sans lien avec une autre maladie) de la tête, et non pas provoquée par les troubles de l'utérus ou des intestins.

est basée sur les premières descriptions faites de manière rigoureuse par Constanzo Varolio (1543-1575) au XVI^e^ s. (Tubbs *et al.*, 2008, p. 734) [5]. À l'époque des Lumières en Angleterre (fin XVII^e^ – début XVIII^e^ s.), Thomas Willis (1621-1675), créateur du terme *neurologia* [6] en 1664 (Monteiro *et al.*, 2021, p. 2), maintint et développa cette conception en caractérisant cette maladie comme principalement convulsive en raison des explosions des « esprits animaux » dans le cerveau (Willis, 1681, p. 1). Enfin, l'histoire de la neurologie trouve son ancrage à la fin du XIX^e^ siècle dans les travaux de Jean-Martin Charcot (1825-1893). Ce n'est qu'à cette période que la maladie en question a été systématiquement traitée comme un trouble neurologique (Tasca *et al.*, 2012, p. 114). La sous-section 1.2 évoque certains de ses apports principaux dans le domaine scientifique.

1.2 Jean-Martin Charcot : un médecin à l'aube de la neurologie moderne

Figure emblématique et directeur de l'illustre École de la Salpêtrière (basée à l'actuelle hôpital de la Pitié-Salpêtrière à Paris), Charcot a laissé une trace indélébile dans le domaine de la neurologie. Il est essentiellement connu pour ses études portant sur les troubles névrotiques, notamment l'hystérie. Selon lui, l'hystérie découle d'une dégénérescence héréditaire du système nerveux, en montrant qu'elle est en fait plus fréquente chez les hommes que chez les femmes (Tasca *et al.*, 2012, p. 114). Charcot a été reconnu pour ses travaux de recherche sur l'hypnose qu'il a utilisée afin d'induire l'état modifié de conscience d'un sujet, permettant l'analyse des symptômes hystériques et leur traitement. Son nom est également associé aux descriptions de nombreuses pathologies connues aujourd'hui, comme la *maladie de Parkinson*, la *sclérose en plaques disséminées*, abr. *SEP* (ou *sclérose multiple*), la *sclérose latérale amyotrophique*, abr. *SLA* (soit la *maladie de Charcot*, ou *maladie Lou-Gehrig*) etc [7].

Ces explorations des abîmes de l'esprit humain lui ont valu de nombreuses appellations : à part avoir été globalement considéré comme le père de la neurologie française et moderne (Teive *et al.* 2022, p. 761 ; Broussolle *et al.* 2012, p. 301), d'autres noms symboliques lui ont été associés, notamment « Napoléon des névroses », « Paganini de l'hystérie » (Mirbeau & Michel 1995, p. 124), ou même « César de la Faculté » (Camargo *et al.*, 2024, p. 1109). Dans la même lignée de pensée, l'École de la Salpêtrière était caractérisée comme la « Mecque de la neurologie » grâce aux activités de Charcot (Teive *et al.* 2014, p. 637 ; Goetz 2017, p. 628 ; Camargo *et al.* 2024, p. 1100). En outre, de nombreuses références à Charcot et des descriptions d'attaques hystériques figurent non seulement

5. Il s'agit de l'identification et de la description de la structure cérébrale agissant comme un relai entre le cerveau et le cervelet, appelée *pont* (lat. *pons*) par Varolio (1573), soit *pont de Varole* (lat. *pons Varolii*), en l'honneur du célèbre anatomiste, qui fut le premier à examiner le cerveau de sa base vers le haut.

6. Terme présent dans Willis (1664).

7. Pour un aperçu détaillé des contributions majeures de Charcot dans le domaine de la médecine, voir Camargo *et al.* (2024, p. 1102).

dans la littérature médicale, mais aussi dans des romans naturalistes français et européens, notamment en Pays-Bas, Russie, pays scandinaves, Espagne, Italie et Allemagne (Koehler, 2013).

Charcot a créé un véritable réseau scientifique autour de soi grâce à ses idées novatrices qui ont eu un grand retentissement parmi ses collaborateurs, élèves et savants polymathes. Parmi eux, nous ne nommons que quelques figures majeures souvent citées dans la littérature (Gomes & Engelhardt 2013, p. 816 ; Bogousslavsky 2014, p. 55 ; Camargo *et al.* 2024, p. 1100), notamment :

- Paul Richer (1849-1933), anatomiste, neurologue et sculpteur, qui a résumé les premières études de Charcot sur l'hystérie dans ses *Études cliniques sur l'hystéro-épilepsie ou grande hystérie* ;
- Georges Gilles de la Tourette (1857-1904), psychiatre et neurologue, qui a décrit les symptômes de la *maladie des tics*, renommée *syndrôme de Tourette* en son hommage par Charcot lui-même ;
- Pierre Janet (1839-1916), philosophe, neurologue et psychiatre, concepteur des termes *dissociation* et *sous-conscient* ;
- Désiré Magloire Bourneville (1840-1909), homme politique et neurologue, qui a publié le premier tome de l'ouvrage monumental l'*Iconographie photographique de la Salpêtrière*, dédiée à l'hystérie, sous l'égide de Charcot ;
- Joseph Babinski (1857-1932), neurologue et neurobiologiste, concepteur du terme *pithiatisme*, qui a découvert le réflexe cutané plantaire, appelé également *signe de Babinski*.

L'impact colossal de Charcot sur sa propre discipline se reflète aussi dans le changement d'intérêt radical du célèbre psychanalyste Sigmund Freud (1856-1939), caractérisé par le passage de la neurologie générale à l'hystérie, l'hypnose et d'autres troubles psychologiques. En effet, son séjour dans le service de Charcot à Paris en 1885-1886 a donné lieu au développement de la théorie psychanalytique (Camargo *et al.*, 2018, p. 41). Néanmoins, certains scientifiques ont fortement contesté le raisonnement scientifique de Charcot, comme le neurologue Hippolyte Bernheim (1840-1919) avec l'École de Nancy pendant les années 1880-1890. Cette polémique porte sur la nature de l'hypnose qui, pour Charcot, représentait un état pathologique propre aux hystériques, et non pas un état de sommeil obtenu par suggestion qui est susceptible d'applications thérapeutiques (et donc, applicable à pratiquement n'importe qui), comme le soutenait Bernheim (1891, pp. 130–131).

Étant donné l'importance des travaux de Charcot et ses contributions dans le domaine de la neurologie et au-delà, nous souhaitons explorer la notion de la circulation des savoirs au prisme du numérique à travers son impact. Avant d'aborder la question d'opérationnalisation de son impact, nous tenons d'abord à décortiquer les mécanismes à l'origine des circulations des savoirs à grande échelle, ainsi que de définir la notion d'un « concept » pouvant véhiculer les informations importantes concernant les circulations en question.

Chapitre 2 Pister la circulation du discours médical au prisme du numérique

2.1 Modalités des circulations des savoirs

De nombreux·ses chercheur·se·s·x partagent le point de vue selon lequel la notion de « circulation des savoirs » constitue un champ de recherche vaste, ainsi qu'un nouveau paradigme de la connaissance depuis le début du XXI[e] siècle et l'avènement du Web 2.0 [1] (Landais, 2014 ; Quet, 2014). Le terme en question reste toutefois assez complexe en raison de visions différentes sur la façon de le définir. Afin d'éclairer cette problématique, Quet (2014, pp. 221–222) souligne trois aspects suivants :

1. **Éléments de la circulation**. Qu'est-ce qui circule ?
   - individus (savants, techniciens, traducteurs, etc.) ;
   - objets matériels (instruments scientifiques, ouvrages etc.) :
   - constructions symboliques (théories, concepts etc.).
2. **Conceptions de la circulation et méthodes de son analyse** ;
   - définition de la circulation comme « traduction », « diffusion », « accès » ou « succès » ;
   - critères méthodologiques possibles pour étudier la circulation p. ex. d'une théorie :
     – circulations géographiques des principaux concepteurs qu'on lui reconnaît ;
     – circulations et lectures des textes produits par leurs concepteurs ;
     – usages et applications analogiques qui en sont faits dans d'autres domaines.
   - enjeux d'articulation de ces différents niveaux d'observation du point de vue méthodologique et de celui de la production du texte de recherche, dans le cas des croisements de ces niveaux.
3. **Conceptions analytiques et normatives des savoirs**
   - affaiblissement des catégories des « savoirs profanes » et « savoirs scientifiques », ainsi que de l'opposition entre eux ;

1. Cette phase de l'évolution du Web se caractérise notamment par la transformation majeure de l'Internet en vue du développement des réseaux sociaux, des blogs et des sites participatifs, tout en permettant aux utilisateur·trice·s·x de créer, partager et interagir avec du contenu Web. Nous traversons actuellement l'ère du Web 3.0 qui repose sur des technologies telles que la chaîne de blocs (angl. *blockchain*), le *NFT* (angl. *non-fungible token*), l'intelligence artificielle, métavers et le Web sémantique (Varet, 2023).

- revalorisation des savoirs implicites et de la dimension pratique des connaissances ;
- glorification de la circulation comme porteuse de valeurs *a priori* positives : confrontation à l'autre, hybridation, production de nouveauté, etc.

Dans le cadre de l'analyse de l'impact scientifique de Charcot, nous étudions *in fine* la circulation de ses théories et des concepts médicaux dont il était inventeur (p. ex. *SLA*) et transmetteur (p. ex. *hystérie*) [2].

2.2 À partir de quel moment un concept devient-il pertinent ?

Le mot « concept » est un terme générique qui renvoie à un grand nombre de théories provenant de divers domaines de pensée, sans qu'il en existe une qui soit exhaustive et universellement acceptée. D'après Lecourt (1999, p. 224), l'invention de l'entité du concept remonte à l'ère d'Aristote, qui l'a caractérisé comme une abstraction, un mode de connaissance à la fois médiat et général, et comme mode de classification entre le genre et l'espèce (*intension* et *extension*). En revanche, selon les linguistes, un concept a une structure double, constituée du sens linguistique et culturel. Sa couche intérieure est constituée du noyau étymologique sur lequel repose ensuite la couche périphérique qui hérite les éléments formés par la culture, les traditions et les expériences humaines [3]. Il peut être exprimé par de différentes éléments du langage, soit : lexèmes, idiomes, collocations, phrases ou textes entiers (Nemickienė, 2011, p. 5). Dans le domaine du traitement automatique des langues (TAL), le terme « concept » peut s'apparenter à celui des « entités nommées », comme en témoignent les recherches sur l'extraction automatique de la terminologie biomédicale (Jolly *et al.*, 2024 ; Navarro *et al.*, 2023). Un concept d'un domaine de connaissance peut faire partie d'un thésaurus, liste organisée de termes contrôlés et normalisés, auquel cas le concept est appelé « descripteur ». (Rennesson *et al.*, 2020, p. 16).

Afin de pouvoir analyser les concepts médicaux liés à Charcot, il paraît important de déterminer à partir de quel moment un mot ou un groupe de mots devient un concept en sciences humaines et sociales (ci-après SHS). Du point de vue de l'histoire des concepts (allem. *Begriffsgeschichte*), cette transformation survient lorsqu'un seul mot comprend toute la gamme des significations dérivées d'un contexte sociopolitique (Koselleck & Richter, 2011, p. 258). À titre d'exemple, le concept d'un *état* ne peut être interprété qu'à travers ses différents constituants, dont *souveraineté territoriale, législation, fiscalité*, parmi maints d'autres. Les concepts sont donc les concentrations par défaut ambiguës d'une multitude de contenus sémantiques, uniquement interprétables et indéfinissables, par

2. Comme déjà expliqué dans la partie 1.1, Charcot n'a pas inventé ce terme, mais en réinterprété le sens.

3. En linguoculturologie, on retrouve le terme « concept linguo-culturel » qui reflète cette nature double du concept.

contraste avec des significations des mots qui peuvent être définies de manière exacte (Koselleck & Richter, 2011, p. 20).

De plus, les concepts comme *histoire* ou *progrès* sont caractérisés comme « collectifs singuliers » qui marquent un passage du domain concret d'un individu (plusieurs *histoires* et *progrès* individuels) au domain abstrait et général du collectif social (une *histoire* ou un *progrès* général ou collectif). Ce phénomène linguistique, ainsi que la création des concepts comme *industrie, usine, classe moyenne* etc., reflète un changement de paradigme dans l'organisation sociale survenu lors des révolutions politiques et industrielles (Hobsbawm, 2010, p. 1). Cela traduit donc le lien fort entre l'histoire du langage et l'histoire des idées. Cette période charnière est nommée *Sattelzeit* [4] (Koselleck & Richter, 2011, p. 8), entre 1750 et 1830, durant laquelle les concepts historiques deviennent abstraits, singularisés, respatialisés et retemporalisés.

Ces considérations peuvent s'appliquer à d'autres constructs en SHS, comme *travail*, *intelligencija*, *Ancien Régime*, *avant-garde*, *Occident* etc. Elles ont acquis le statut des concepts « nomades » en raison de leur circulation spatio-temporelle et linguistique (Ghermani, 2011, p. 117). Plusieurs questionnements ont été soulevés par la même autrice à l'égard de leur émergence, notamment pour déterminer à quel moment un concept devient une entrée dans un dictionnaire des SHS : « *Pourquoi un concept fait-il son entrée dans un dictionnaire ? Au terme de quel processus ? À l'inverse, comment cette percée lexicale est-elle parfois impossible ou refusée ?* ». Les processus permettant à un concept d'obtenir le statut de scientificité sont la propagation, la bifurcation, la capture [5], mais aussi les pratiques scientifiques conduisant aux masquages de sens (p. ex. dans le cas du terme « confession [religieuse] », dont le sens varie en fonction du pays dans lequel il est utilisé).

Comme nous avons pu voir, l'histoire des concepts concerne principalement les manifestations de conflits sociopolitiques particuliers qui doivent être compris dans leur contexte approprié (p. ex. les mots comme *liberté* ou *démocratie* portent la connotation polémique dont le sens ne peut être précisé qu'à travers leurs antithèses). Toutefois, nous considérons que cette théorie pourrait nous permettre de formaliser une approche pour tracer l'évolution des concepts médicaux. Dans cette optique, ces concepts auront le rôle des vecteurs de la crise conceptuelle, ce qui représenterait une forme de *Sattelzeit* dans le domaine de la médecine : autrement dit, ces concepts sont détournés de leurs sens initiaux neutres (descriptions des pathologies) vers ceux exerçant un certain impact sur la communauté scientifique.

4. Trad. allem. « époque de selle ».

5. Termes employés par Stengers (1987, pp. 8–22), représentatrice de la conception constructiviste du savoir scientifique.

2.3 Études numériques des circulations culturelles

Incontestablement, l'époque actuelle est profondément marquée par le « déluge des données », phénomène représentatif de la quatrième paradigme de la science, selon Jim Gray (Hey *et al.*, 2009, p. 30). Par conséquent, les projets numériques sont aujourd'hui « pilotés par les données » [6] et ceux qui sont centrés sur les explorations des circulations culturelles au prisme du numérique se concrétisent à grande échelle. Sont fortement axés sur cette thématique : (i.) certains chaires universitaires, notamment celle des Humanités numériques à l'université de Genève (Joyeux-Prunel & Gabay, 2022) ; (ii.) de divers évènements scientifiques, comme la journée d'étude « Circulation des écrits littéraires de la première modernité et humanités numériques » [7], les colloques Humanistica 2023 [8], ACFAS 2023 [9] etc., ou bien (iii.) des revues entières, par exemple *Mots. Les langages du politique*, dont les articles portent sur les thématiques aussi diverses que les circulations textuelles internationales du discours complotiste des « Illuminati » (Chaudet, 2022), du discours « conspirationniste » sur Twitter (Giry & Nouvel, 2022) etc.

Ce mémoire est basé sur la contribution de Petkovic *et al.* (2023) s'inscrivant dans l'optique de l'exploration des circulations des concepts médicaux. Nous souhaitons mesurer informatiquement l'impact scientifique des travaux de Charcot sur ses collaborateurs et successeurs, membres de son réseau scientifique. Cette mesure se fonde sur l'analyse des concepts-clés en matière de son discours scientifique, et plus particulièrement sur l'opérationnalisation du terme « influence », définie ici comme une intertextualité [10] unidirectionnelle, allant des écrits de Charcot (ci-après corpus « Charcot ») vers ceux de ses collaborateurs et successeurs (ci-après corpus « Autres »). Il s'agit donc *in fine* d'aborder computationnellement la question des circulations, non pas des artefacts matériels comme les manuscrits (Gabay *et al.*, 2021) et les images (Joyeux-Prunel, 2019), mais des phénomènes textuels complexes (Manjavacas *et al.*, 2019) ayant une dimension théorique forte.

La question de recherche sous-tendant ce mémoire s'approche tangentiellement des travaux de Riguet (2018) et de Roe *et al.* (2023). Le premier travail porte sur la réception de la pensée scientifique du physiologiste français Claude Bernard dans la critique littéraire, illustrée par l'alignement des textes de Bernard avec des ouvrages de critique littéraire. Le second article porte sur la détection de réemplois textuels à grande échelle et l'analyse de réseaux pour identifier les « influenceurs » dans les ouvrages français du siècle des Lumières.

6. Traduction du terme *data-driven* introduit par Johns (1991), issu de l'expression *data-driven learning*.
7. https://www.fabula.org/actualites/86846/circulation-des-ecrits-litteraires-de-la-premiere-modernite-et-humanites-numeriques.html
8. https://humanistica2023.sciencesconf.org/
9. https://www.crihn.org/nouvelles/2022/12/11/colloque-de-la-transformation-des-sciences-humaines-par-les-humanites-numeriques-acfas-2023/
10. Nous nous appuyons sur la définition de l'intertextualité dans la littérature, où ce terme désigne « la perception, par le lecteur, de rapports entre une œuvre et d'autres qui l'ont précédée ou suivie » (Riffaterre, 1980, p. 4).

Chapitre 3 Valorisation du fonds Charcot

3.1 Description du fonds Charcot

Le fonds patrimonial de Jean-Martin Charcot est conservé à la Bibliothèque de Neurosciences Jean-Martin Charcot par la Bibliothèque numérique patrimoniale de Sorbonne Université (BSU) [1]. Ce fonds regroupe des ouvrages suivants :

- fonds historique Charcot (bibliothèque personnelle de Charcot) : ouvrages, périodiques, collection de thèses et de tirés à part, manuscrits, observations, collection neurologique couvrant la seconde partie du XIX^{e} siècle, fonds bibliophilique ancien ;
- collections de la bibliothèque des Internes de la Salpêtrière : ouvrages, périodiques, thèses en neurologie et psychiatrie pour la période 1800-1950 ;
- donations en ouvrages du docteur Achille Souques.

Dans un souci de préservation d'ouvrages originaux et de valorisation de collections ayant un caractère iconographique notable, une partie de ce fonds a été numérisée. Ces archives numérisées sont disponibles sur le portail numérique SorbonNum [2], porte d'entrée unique vers les collections scientifiques patrimoniales et numériques de Sorbonne Université, ainsi que sur Gallica, bibliothèque numérique de la Bibliothèque nationale de France (BnF) [3].

Le fonds numérisé a été décrit et divisé par la BSU en quatre grandes typologies de documents :

1. Fonds iconographique
   - **Album des internes** : Album des promotions annuelles d'internes, photographiées et classées par établissements de l'Assistance Publique, entre 1860 et 1963 ;
   - **Photographies sur les aliénés de Bicêtre par Désiré Magloire Bourneville** :

1. https://www.sorbonne-universite.fr/bu/decouvrir-nos-bibliotheques/la-bibliotheque-charcot.
2. anc. Jubilothèque, https://patrimoine.sorbonne-universite.fr/collection/Fonds-Charcot
3. https://gallica.bnf.fr/services/engine/search/sru?operation=searchRetrieve&version=1.2&query=%28gallica%20all%20%22Charcot%2C%20Jean-Martin%22%29&lang=fr&suggest=0.

deux albums présentant les photographies des « petits enfants anormaux » hospitalisés à Bicêtre dans le service du docteur Bourneville, collaborateur de Charcot.

2. Leçons et manuscrits des leçons de Charcot
   - **Manuscrits des leçons et observations de Charcot (1825-1893)** : leçons orales de Charcot, rédigées intégralement de sa main et annotées ;
   - **Leçons de Charcot** : numérisation des volumes de l'*Œuvre Complète* de Charcot consacrés au système nerveux et à l'enseignement clinique, comme par exemple les célèbres leçons du Mardi, sur l'hystérie notamment.
3. Périodiques
   - ***Les Recherches cliniques et thérapeutiques sur l'épilepsie, l'hystérie et l'idiotie (1872-1903)*** de Bourneville. Y est retracée toute l'activité du Service des Enfants Idiots, à la Salpêtrière puis à Bicêtre, par le biais des compte-rendu illustrés de photographies et rédigés par Bourneville ;
   - ***Revue de l'Hypnotisme (1887-1910)*** : périodique consacré à l'hypnotisme que Charcot a réhabilité, publiant les principaux articles théoriques sur cette discipline ;
   - ***Journal du magnétisme (1845-1861)*** : la collection reflète les recherches sur le magnétisme, renouvelées au milieu du XIX^{e} siècle ;
   - ***Revue photographique des hôpitaux de Paris (1869-1872)***. Première revue exposant les applications de la photographie à la médecine, notamment la médecine hospitalière, à travers les études menées à l'Hôpital Saint-Louis, et à la Salpêtrière ;
   - ***Iconographie Photographique de la Salpêtrière (1875-1879)***. La collection présente les observations de patientes examinées à la Salpêtrière, accompagnées de photographies d'Albert Londe, directeur du service photographique, présentant les divers stades de la crise d'hystérie ;
   - ***Nouvelle Iconographie de la Salpêtrière (1888-1918)***. La revue est fondée sous la direction de Charcot par Paul Richer, Gilles de la Tourette et Albert Londe. Elle réunit la collection de clichés constituée à la Salpêtrière a pour but la représentation objective des pathologies observées. Elle prend la relève de l'*Iconographie Photographique de la Salpêtrière*. Les articles sont illustrés de photographies, de dessins et de lithographies ;
   - ***Archives de neurologie (1880-1907)***. Sous-titrée « Revue trimestrielle des maladies nerveuses et mentales », les Archives de neurologie sont publiées sous la direction de Charcot par Bourneville. La revue édite, groupe, catégorise et compare la masse des travaux de pathologie nerveuse. Les *Archives de neurologie* sont devenues bisannuelles en 1881.
4. Ouvrages de la bibliothèque de Charcot
   - **Collection d'atlas d'anatomie et de pathologie du système nerveux**, publiés durant le XIX^{e} siècle. L'iconographie de ces ouvrages est remarquable, à commencer par l'*Atlas de Vicq d'Azyr*, médecin du roi Louis XVI ;
   - **Traités**. Cette collection regroupe à la fois des traités sélectionnés dans la biblio-

thèque de Charcot (comme l'*Opera omnia*... de Thomas Willis, 1682, comportant des gravures), des atlas et des textes significatifs des successeurs de Charcot, issus de la bibliothèque des Internes de la Salpêtrière (par exemple l'*Anatomie des centres nerveux* des Déjerine).

3.2 Constitution du corpus Charcot

Le corpus de travail est constitué de 201 documents OCRisés (sans post-correction), fournis au format XML par la BSU. Nous avons procédé, dans un premier temps, à une restructuration des textes en XML-TEI [4] à l'aide de l'outil TEINTE [5], afin de permettre la fouille avancée du corpus Charcot à travers des outils développés au sein de l'équipe-projet OBTIC. D'une part, le moteur de recherche OBVIE [6] permet de repérer des textes similaires par ordre de pertinence à partir des termes en commun. D'autre part, l'algorithme TEXTPAIR génère une liste de passages similaires, c'est-à-dire les séquences de mots qui se chevauchent (n-grammes de mots) pour chaque texte, en comparant ensuite ces résultats avec ceux de séquences dans d'autres textes [7].

Afin de mesurer l'impact de Charcot sur son entourage et d'analyser la circulation de concepts véhiculés dans le corpus, nous avons commencé par séparer les documents rédigés par Charcot de ceux rédigés par ses co-auteurs (p. ex. Bourneville) ou les auteurs thématiquement proches de lui (p. ex. son élève Gilles de la Tourette). Nous avons obtenu respectivement 68 (corpus « Charcot ») et 133 (corpus « Autres ») documents, comme présenté dans le tableau 3.1. Les deux corpus issus du fonds Charcot sont librement disponibles et interrogeables sur les deux plateformes OBVIE [8] et TEXTPAIR [9].

Corpus	Nb de documents	Nb de tokens
Charcot textes rédigés par Charcot	68	12 190 649 (38,12 %)
Autres textes rédigés par les membres de son réseau scientifique	133	19 788 830 (61,88 %)
TOTAL	**201**	**31 979 479** (100 %)

TABLEAU 3.1 – Répartition du fonds Charcot selon les auteurs.

4. Originalement, ces fichiers ne contenaient que les balises `<doc>`, `<id_doc>` et `<pages>`.
5. https://github.com/OBVIL/teinte_obtic
6. https://obtic.huma-num.fr/obvie/. Pour d'amples informations sur le fonctionnement de cet outil, *cf.* Alrahabi (2022).
7. https://artfl-project.uchicago.edu/text-pair.
8. https://obtic.huma-num.fr/obvie/charcot/?view=corpus
9. https://anomander.uchicago.edu/

Chapitre 4 Résultats

4.1 Exploration du corpus Charcot : OBVIE et TextPair

Une première exploration du corpus Charcot à travers l'application OBVIE nous a permis d'identifier les substantifs les plus importants de chaque corpus en utilisant les fréquences brutes ou des méthodes plus fines comme TF-IDF, BM25 (détaillées dans la partie 4.2), $\chi 2$ ou le Test Gamma. Cependant, l'application ne permet pas de quantifier la pertinence des expressions polylexicales, soit les n-grammes de mots, très fréquentes dans les deux corpus et dont la décomposition entraînerait une perte d'information (p. ex. le terme polysémique « bulbe » qui a une valeur spécifique dans l'expression figée *bulbe rachidien*). En observant la figure 4.1, nous constatons que l'abscisse donne l'information sur les dates de publication des ouvrages compris dans les corpus, alors que l'ordonnée indique le nombre d'occurrences par million de mots, soit *parties par million* (*ppm*) [1].

Concernant l'alignement des séquences similaires aux deux corpus, TextPair nous a permis, par une lecture attentive, de faire des comparaisons entre les textes et de rechercher des termes au sein des passages similaires, malgré le nombre de résultats assez conséquent (*cf.* la figure 4.2). En raison de sa capacité de détecter les passages similaires, notamment les citations directes, les plagiats ou les réemplois, ce logiciel, ainsi qu'un autre logiciel de détection de plagiat, peuvent nous servir de *baseline* pour comparer leurs résultats avec ceux proposés dans la partie 4.2.

4.2 Extraction des phrases-clés : méthodes statistiques

Afin de surmonter les limites rencontrées avec ces deux outils, nous avons proposé une nouvelle méthode pour identifier des concepts dans les deux corpus en nous basant sur le poids de leur apparition, calculé selon trois différentes mesures de pondération [2] :

1. *Cf.* le guide d'utilisation d'OBVIE détaillé : https://obtic.huma-num.fr/obvie//static/aide.html.
2. Le code est disponible en ligne : https://github.com/ljpetkovic/Charcot_circulations.

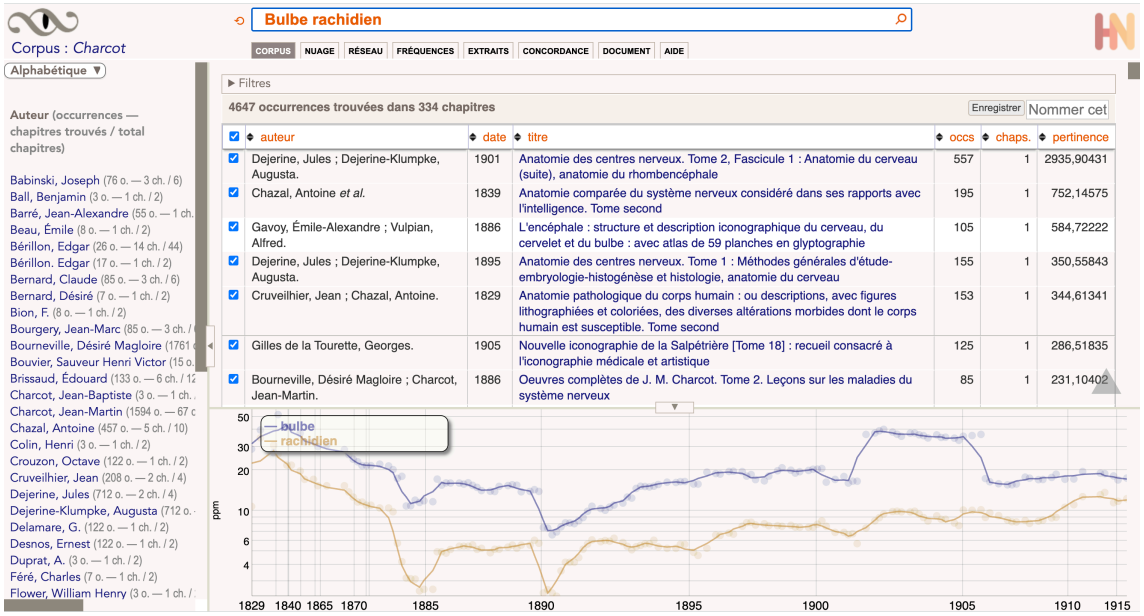

FIGURE 4.1 – Distribution des fréquences des tokens avec la frise chronologique pour ceux constituant l'expression « bulbe rachidien » (issus du corpus « Charcot » et du corpus « Autres ») dans le logiciel OBVIE.

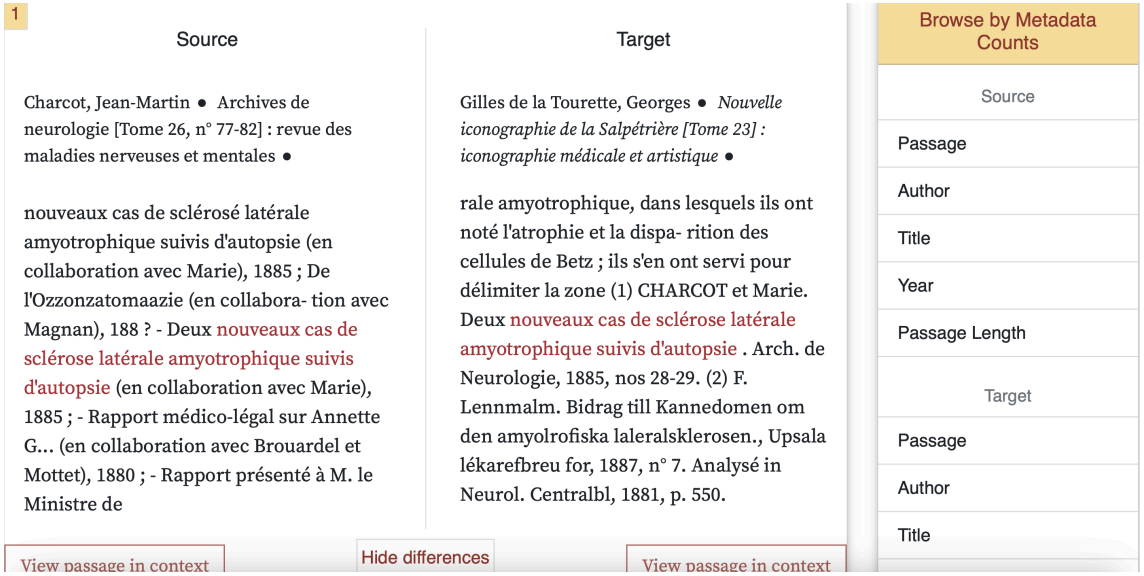

FIGURE 4.2 – Alignement et comparaison d'un texte de Charcot à celui de Georges Gilles de la Tourette (le seul résultat) en lançant la requête *sclérose latérale amyotrophique*.

- TF-IDF (Robertson & Jones, 1976) est une méthode qui permet d'évaluer l'importance d'un terme contenu dans un document relativement à un corpus plus large en récompensant la fréquence des termes, sans tenir compte des variations de longueur du document ;
- BM25 est une fonction de classement qui classe un ensemble de documents en fonction des termes de requête apparaissant dans chaque document, quelle que soit l'interrelation entre les termes de requête au sein d'un document (par exemple, leur proximité relative). Il s'agit d'une tentative d'amélioration de TF-IDF, notamment pour prendre en compte divers facteurs tels que la longueur du document et les

problèmes engendrés par la possible saturation des termes (Robertson *et al.*, 2009, p. 355) ;

- BERT (Devlin *et al.*, 2019) est un modèle pré-entraîné qui utilise l'apprentissage profond non-supervisé sur de grandes quantités de données textuelles pour apprendre des représentations de mots et de phrases, et comprendre le contexte et la sémantique. Il est basé sur l'architecture des *transformeurs*, qui est un type de grands modèles de langue utilisé pour le TAL.

La liste des concepts retenus pour l'étude est composée de termes ou expressions popularisés par Charcot, comme *hystérie*, *sclérose latérale* etc. (Camargo *et al.*, 2024, p. 1102) [3]. Pour chaque entrée, nous avons pris en compte les formes du singulier et du pluriel obtenues grâce à des expressions régulières. La liste est produite de façon supervisée et provient du croisement entre la liste des termes obtenus avec OBVIE et l'index d'une édition des œuvres complètes de (Charcot, 1892, pp. 493–507), dont nous avons retiré les termes génériques (*os*, *cerveau*, etc.).

Comme nous pouvons l'observer sur la figure 4.3, la mesure BM25 révèle une intensification du lexique de Charcot dans le corpus « Autres ». Plus précisément, tous les termes évalués sont identifiés comme plus signifiants dans le discours des « Autres » que dans celui de Charcot, les scores étant plus élevés pour 14 termes (sur 14 évalués) utilisés par le réseau de Charcot. D'ailleurs, d'après le tableau 5 (en annexe), c'est la seule mesure dont les valeurs témoignent clairement d'un lexique partagé entre Charcot et ses successeurs et collaborateurs, *a contrario* des deux autres mesures, où le rapport en question est inversé (la grande majorité des termes étant plus pertinente dans le discours de Charcot, et son impact étant donc moins accentué). Concrètement, les termes les plus pertinents semblent être *sclérose en plaque disséminées* (score 0,83), *paralysie rhumatismale* (0,68), *atrophie progressive* (0,53) et *arthrite déformante* (0,50).

D'autre part, nous avons utilisé BERT pour mesurer le poids des termes dans les deux corpus. Bien que ce type de modèle ne fournisse pas directement de poids pour les mots, nous pourrions en extraire des informations utiles pour estimer l'importance ou le poids des mots dans les textes. Différentes approches sont utilisées pour obtenir une représentation de l'importance des mots, en exploitant des plongements lexicaux et des mécanismes d'attention (Vaswani *et al.*, 2023). Pour ce travail en cours, nous avons utilisé le modèle `bert-base-multilingual-cased`. Les premiers résultats obtenus se trouvent dans le tableau 5 et restent à améliorer. Cependant, nous avons observé que les termes les plus pertinents pour le discours de Charcot étaient ceux qui désignent les noms des différentes pathologies (*diplopie*, *myélite partielle*, *état de mal épileptique*, *paralysie labio-glosso-laryngée* etc.), contrairement à d'autres notions plus abstraites (*vicieuses*, *délire*, *miracle*) qui sont prédominantes dans le corpus « Autres » (termes non renseignés dans le tableau en question). La présence de ce dernier type de notion n'est pas étonnant, étant donné que Charcot aborde la question des guérisons miraculeuses dans ses recherches [4].

3. *Cf.* la liste exhaustive des termes et des expressions popularisés par Charcot en annexe.
4. Voir notamment son œuvre *La foi qui guérit* (Charcot, 1897).

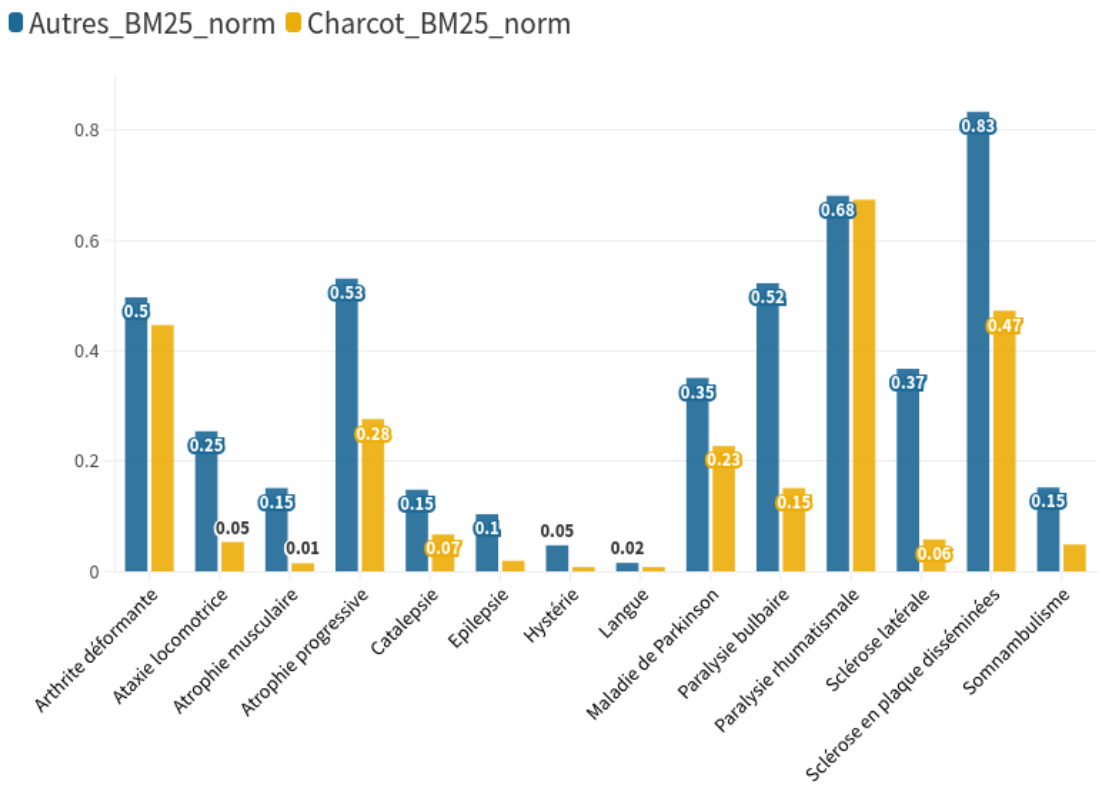

FIGURE 4.3 – Visualisation de pertinence des concepts dans les deux corpus suivant la métrique BM25. Les valeurs des concepts associées au corpus « Autres » sont représentées en bleu, alors que celles du corpus « Charcot » en jaune.

4.3 Extraction des phrases-clés : méthode à base d'apprentissage profond

En complément de la méthode du calcul de pertinence des termes médicaux fournis de manière supervisée (partie 4.2), nous exposons ici des résultats de l'approche non-supervisée pour extraire des mots/phrases-clés pertinents à partir de nos deux corpus [5]. L'objectif de cette approche est de détecter les termes communs entre les deux corpus et de montrer la répartition des termes les plus pertinents dans le réseau de Charcot. Deux algorithmes librement disponibles sont présentés ici pour illustrer cette dernière approche : `keybert` (Grootendorst *et al.*, 2023) [6] et `keyphrase-vectorizers` [7].

4.3.1 Librairie `keybert`

Cette librairie Python permet d'exploiter les plongements de mots (angl. *word embeddings*) du type BERT pour générer des mots/phrases-clés les plus similaires à un document. La figure 4.4 illustre la chaîne de traitement appliquée à nos deux corpus :

5. *Cf.* le dépôt GitHub https://github.com/ljpetkovic/Charcot_KeyBERT_Keyphrase-Vectorizers/.
6. https://maartengr.github.io/KeyBERT/
7. https://pypi.org/project/keyphrase-vectorizers/

1. les corpus « Charcot » et « Autres » sont utilisés comme les données d'entrée au format `.txt` ;
2. les documents d'entrée ont été tokenisés en phrases-clés candidates avec la fonction `CountVectorizer` ;
3. les plongements des documents et de leurs phrases-clés candidates ont été générés par le modèle de langue `sentence-transformers` ;
4. la similarité cosinus a été calculée entre les documents d'entrée et les phrases-clés candidates, où celles avec les scores les plus élevés sont extraites.

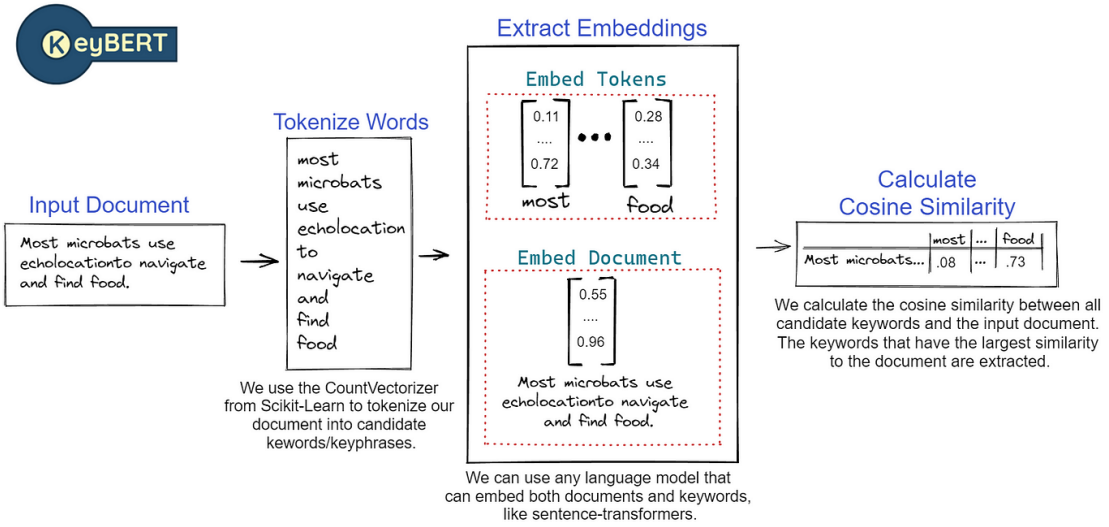

FIGURE 4.4 – *Pipeline* de la librairie `keybert` [8].

Une première tentative de génération des phrases-clés les plus pertinentes dans les deux corpus n'a produit que deux termes : ARTICULATION DE [*sic*] ÉPAULE et PARALYSIE FACIALE PÉRIPHÉRIQUE. Par ailleurs, en observant les 15 phrases-clés les plus pertinentes dans le corpus « Autres » (figure 4.5), nous constatons un manque de diversification des résultats et des phrases-clés qui se ressemblent (*la sensibilité tactile*, *sensibilité tactile au*, *la sensibilité tend* etc.) [9]. Un autre problème observé était la non-grammaticalité des phrases-clés extraites (*sémi lunaire segment*, *prière le malade* etc.), ce qui nous a incités à tester une approche plus fine, décrite dans la partie 4.3.2.

4.3.2 Approche *PatternRank*

Cette approche exploite la librairie `keyphrase-vectorizers` qui offre la possibilité d'extraire les phrases-clés pertinentes et spécifiques à l'aide des balises de parties de discours. Cela nous a paru comme une piste intéressante, étant donné que les termes médicaux (surtout ceux plus pointus) que l'on souhaitait extraire étaient généralement des

8. Illustration reprise de https://maartengr.github.io/KeyBERT/guides/quickstart.html#installation.

9. Pour assurer que les phrases-clés ne se ressemblent pas, il faut utiliser le paramètre `use_mmr` et spécifier sa valeur entre 0 et 1.

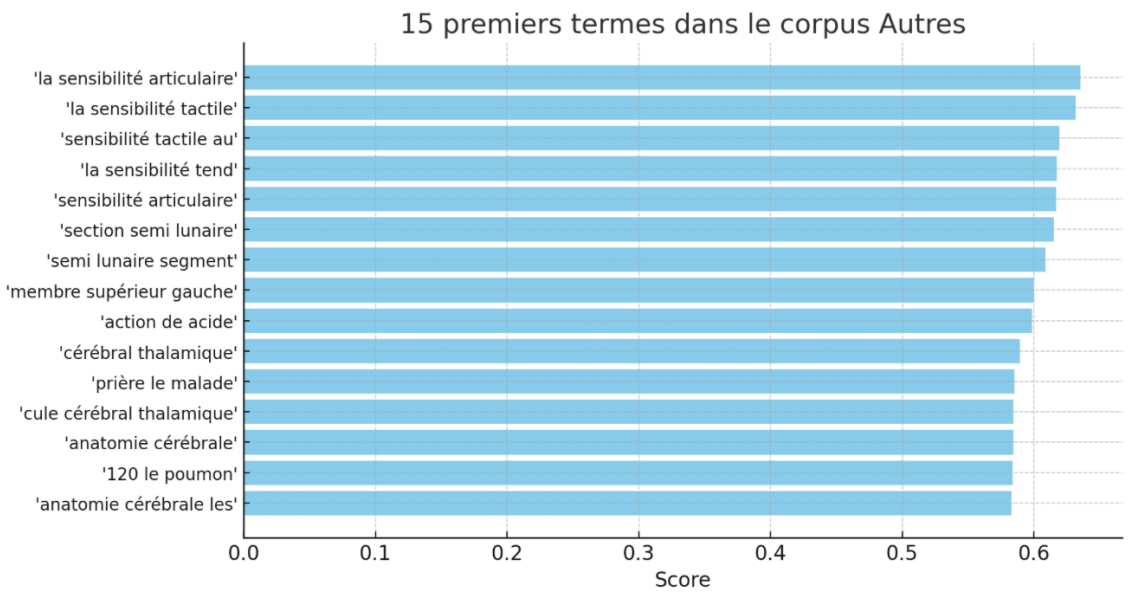

FIGURE 4.5 – 15 termes les plus pertinents dans le corpus « Autres » selon `keybert`.

n-grammes constitués des substantifs, suivis d'un ou plusieurs adjectifs (p. ex. *sclérose latérale amyotrophique*). Voici les étapes de la chaîne de traitement de l'approche *PatternRank* (figure 4.6) :

1. les corpus « Charcot » et « Autres » sont utilisés comme les données d'entrée au format `.txt` ;
2. les tokens ont été extraits et étiquetés avec les balises de partie du discours et les expressions régulières `<N.*>+<ADJ.*>*` (sans utiliser le paramètre `use_mmr`) ;
3. les tokens ont été sélectionnés selon les balises de partie de discours souhaitées et gardés comme les phrases-clés candidates ;
4. les plongements des documents et de leurs phrases-clés candidates ont été générés par le modèle de langue (en l'occurrence `flair` [10]) ;
5. les similarités cosinus ont été calculées entre ces deux types de plongements, et les phrases-clés candidates ont été triées par ordre décroissant ;
6. les phrases-clés les plus représentatives ont été extraites.

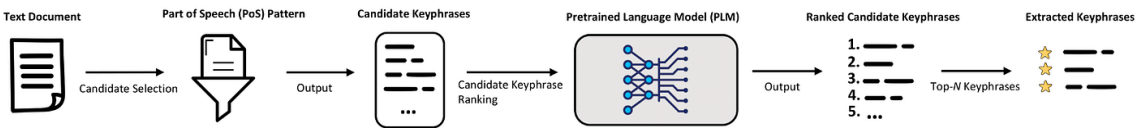

FIGURE 4.6 – *Workflow* de la méthode *PatternRank* (Schopf *et al.*, 2022, p. 2).

La figure 4.7 nous informe sur les 15 termes les plus pertinents et fréquents, extraits avec la librairie `keyphrase-vectorizers`, que l'on retrouve dans les deux corpus. Malgré certains tokens tronqués, très probablement en raison d'un OCR imparfait (*ments* → *mouvements*, *decins* → *médecins* etc.), nous observons une diversification des résultats.

10. https://github.com/flairNLP/flair

Après cela, il reste la question de mieux comprendre le rôle des phrases-clés extraites dans les écrits de l'entourage de Charcot et/ou si elles sont vraiment significatives ou pas.

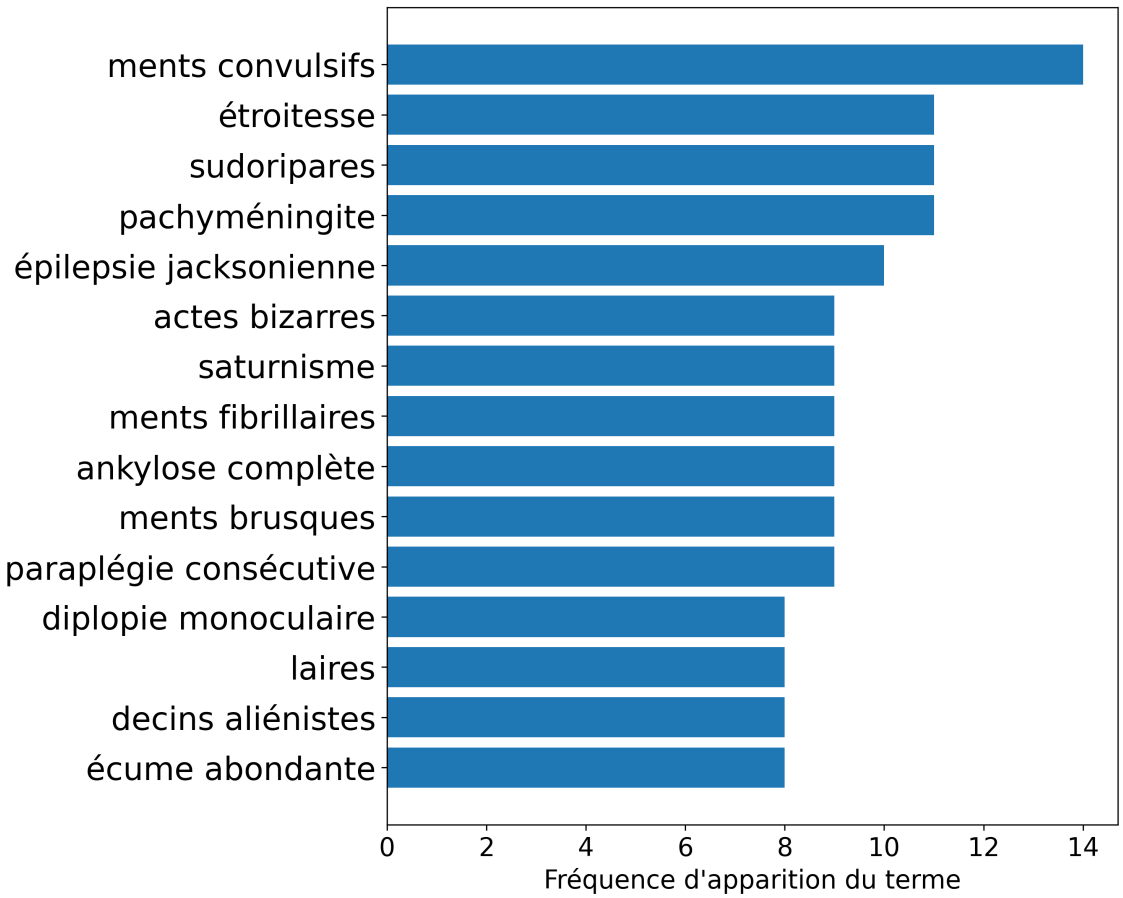

FIGURE 4.7 – Les 15 termes les plus fréquents extraits avec `keyphrase-vectorizers` et partagés par les deux corpus.

Conclusion

Ce travail constitue la première phase d'exploration du corpus de Charcot qui vise à mesurer le degré d'intertextualité entre le discours médical de Charcot et celui de son réseau scientifique, en repérant les termes les plus importants de son œuvre dans les textes de ses collègues et successeurs. Les deux outils, OBVIE et TextPair, nous offrent des fonctionnalités avancées de recherche et de comparaison de textes dans le cadre d'une analyse de textes assistée par ordinateur ; or, ils ne proposent pas de fonctionnalité de lecture distante permettant de rendre compte de l'impact de Charcot sur son réseau scientifique à travers la pertinence des concepts principaux de ses travaux. Pour pallier ce problème, nous avons d'abord opté pour une approche statistique et supervisée, dans le cadre de laquelle nous avons quantifié les concepts polylexicaux dans les deux corpus selon trois différentes métriques de pondération : TF-IDF, BM25 et BERT. Nous avons ensuite comparé cette approche avec celle basée sur l'apprentissage profond et non-supervisé, en extrayant les phrases-clés les plus pertinentes avec les outils `keybert` et `keyphrase-vectorizers`. Cela nous a permis de détecter les termes communs entre les deux corpus et enfin de montrer la répartition des termes les plus pertinents dans le réseau de Charcot. La visualisation des résultats nous a permis d'observer des phénomènes qu'il serait nécessaire de valider auprès de spécialistes de Charcot. Ces expériences répondent donc partiellement à notre question de recherche, puisqu'elles ne comprennent pas de dimension chronologique de l'impact des concepts médicaux sur le long terme.

Pour la suite, deux pistes de recherche devraient être suivies : dans un premier temps, opérer une analyse sémantique des passages avec l'outil Ariane qui contiendraient les concepts médicaux, afin d'étudier les différentes modalités de prise en charge énonciative : opinions, accords, désaccords, définitions, etc. (Alrahabi, 2021). En effet, reprendre un terme ne veut pas dire y adhérer : on peut citer pour dire, par exemple, que l'on n'est pas d'accord. Il serait donc pertinent d'annoter le corpus Charcot avec Ariane, retenir les passages qui contiennent le plus de catégories « opposées » (puisque ses termes ont suscité du débat et de la polémique), y identifier les phrases-clés avec `keybert` ou `keyphrase-vectorizers`, et les comparer enfin avec les termes de l'index utilisé dans la partie 4.2. Comme deuxième piste, nous proposons d'améliorer le texte issu de l'OCR

à l'aide d'une approche basée sur l'apprentissage profond [1] et évaluer l'impact de la correction orthographique de notre corpus sur ces résultats.

1. Cette approche constituerait une suite du travail effectué dans le cadre de la correction automatique de l'OCR à l'aide du modèle de langue statistique issu du logiciel `JamSpell` (Petkovic *et al.*, 2022).

Annexe

Liste des termes et expressions popularisées par Charcot

amblyopie hystérique
achromatopsie hystérique
amyotrophie protopathique
amyotrophie symptomatique
analgésie
anesthésie
angioneuroses
apoplexie spinale
arthrite déformante
arthropathie des ataxiques
articulations
ataxie locomotrice
atrophie musculaire
atrophie progressive
attaque-accès
attaque apoplectiforme
attaque hystérique
attitude passionnelles
attraction
aura hystérique
avant-mur
bromure de camphre
bulbe rachidien
capsule interne
capsule surrénale
catalepsie
cellule nerveuse
chloroforme
chorée
chorée rythmique hystérique
cicatrice vicieuse
cirrhose de muscles
cœlialgie
compression de l'ovaire
congestion
contractilité électrique
contracture hystérique permanente
contracture permanente
contracture tardive
contracture des uretères
convulsionnaire
convulsion
corde du tympan
corps granuleux
corps opto-strié
courant électrique
crise gastrique
danse
décubitus aigu
dégénération cireuse
délire
diplopie
dynamométrie
ecchymoses
ecthyma
electro-diagnostic
embolie
encéphalite
endocardite
épilepsie
épilepsie spinale
éruption
érythème pernios
escarre des fesses
escarre sacrée
état de mal épileptique
état de mal hystéro-épileptique
excitabilité
faisceau radiculaire interne
faradisation
fève de calabar
glossy skin
globe hystérique
griffe
hématomyélie
hémianesthésie hystérique
hémianesthésie de cause encéphalique
hémichorée
hémiopie
hémiparaplégie
hémiplégie
histologie
hypérémié
hyperesthésie ovarienne

hystérie
hystérie épileptiforme
hystérie ovarienne
hystérie grave
hystérie locale
hystérie infantile
hystérie locale traumatique
hystéro-épilepsie
immobilisation
incoordination motrice
ischémie
ischurie
latéropulsion
lèpre
lésion
lésion oculaire
maladie de Parkinson
méningite ascendante
méningite cervicale
métalloscopie
miracle
moëlle épinière
myélite aiguë centrale
myélite partielle
myélite traumatique
myodynie
myopathie
néphro-cystite
néphrotomie
nerf dilatateur
nerf facial
nerf glandulaire
nerf sciatique
nerf sécréteur
nerf trijumeau
nerf trophique
nerf vaso-moteur
névrite
névroglie
nitrite d'amyle
nystagmus
oblitération
oligurie hystérique
ovarie hystérique
paralysie agitante
paralysie bulbaire
paralysie consécutive
paralysie générale progressive
paralysie générale spinale
paralysie hystérique
paralysie infantile
paralysie labio-glosso-laryngée
paralysie pseudo-hypertrophique
paralysie rhumatismale
paraplégie traumatique
parésie
petit mal
phlegmon
pied bot
putamen
rémission
rétention
rétropulsion
rigidité
salivation
sclérodermie
sclérose fasciculée
sclérose descendante
sclérose latérale
sclérose postérieure
sclérose en plaque disséminées
somnambulisme
tarentisme
torticolis
thermoanesthésie
tremblement
trépidation
trismus
trouble trophiques
tubercule de la moëlle
tympanisme
urticaire
vertige
vision
vomissement hystérique
vomissement urémique
vomissement de sang
zona

Terme	*Charcot*				*Autres*			
	Fréquence	TF-IDF	BM25	BERT	Fréquence	TF-IDF	BM25	BERT
Arthrite déformante	30	0,16	0,45	0,80	24	0,02	**0,50**	0,40
Ataxie locomotrice	559	0,35	0,05	0,83	169	0,08	0,25	0,39
Atrophie musculaire	1105	0,20	0,02	0,84	1465	0,43	0,15	0,42
Atrophie progressive	40	0,14	0,27	0,72	22	0,02	**0,53**	0,39
Catalepsie	681	0,54	0,07	0,88	975	0,28	0,15	0,39
Épilepsie	414	0,09	0,02	0,78	577	0,12	0,10	0,41
Hystérie	5775	0,51	0,01	0,74	4934	0,45	0,05	0,41
Langue	2695	0,24	0,01	0,72	3591	0,11	0,02	0,41
Maladie de Parkinson	75	0,21	0,23	0,81	130	0,09	0,35	0,37
Paralysie bulbaire	149	0,27	0,15	0,89	93	0,09	0,52	0,40
Paralysie rhumatismale	8	0,07	0,67	0,86	14	0,02	**0,68**	0,44
Sclérose latérale	445	0,30	0,06	0,88	127	0,09	0,37	0,41
Sclérose en plaque disséminées	45	0,25	0,47	0,87	12	0,02	**0,83**	0,40
Somnambulisme	847	0,49	0,05	0,89	3410	**1**	0,15	0,43

Tableau 1 – Calcul de pertinence des concepts selon les métriques TF-IDF, BM25 et BERT dans les corpus « Charcot » et « Autres ».

Bibliographie

Alrahabi, M. (2021). Ariane : dispositif de fouille et de lecture synthétique de textes. In *DigitAl Humanities and cuLtural herItAge : data and knowledge management and analysis (Atelier Dahlia)*. https://hal.science/hal-03167271. (page 21)

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/. (page 13)

Amiri, V. V. (24 novembre 2012). T. S. Kuhn. *Histo Philo Sciences*. https://histoirephilosciences.wordpress.com/depuis-le-20eme-siecles/une-nouvelle-epistemologie/t-s-kuhn/. (page 4)

Anouilh, J. (1956). *Pauvre Bitos ou le dîner de têtes*. Gallimard, coll. « Folio », n° 301. https://archive.org/details/anouilh-pauvre-bitos-ou-le-diner-de-tetes-1979. (page 3)

Bachelard, G. (1934). *La formation de l'esprit scientifique : contribution à une psychanalyse de la connaissance*. Vrin. https://gastonbachelard.org/wp-content/uploads/2015/07/formation_esprit.pdf. (page 3)

Bachelard, G. (1970). *Idéalisme discursif*. Vrin, présentation de Georges Canguilhem : Paris. https://www.academia.edu/27217437/BACHELARD_Gaston_%C3%89tudes_Vrin_1970_. (page 4)

Bernheim, H. (1891). *De la suggestion et de ses applications à la thérapeutique*. Paris : Octave Doin. https://gallica.bnf.fr/ark:/12148/bpt6k97805169. (page 6)

Bogousslavsky, J. (2011). *Following Charcot : A Forgotten History of Neurology and Psychiatry*, volume 29. Karger Medical and Scientific Publishers. https://nah.sen.es/en/issues/lastest-issues/135-journals/volume-2/issue-2/270-the-mysteries-of-hysteria. (page 1)

Bogousslavsky, J. (2014). The Mysteries of Hysteria. *Neurosciences and History*, 2(2), 54–73. https://nah.sen.es/vmfiles/abstract/NAHV2N2201454_73EN.pdf. (page 6)

Broussolle, E., Poirier, J., Clarac, F., & Barbara, J.-G. (2012). Figures and institutions of the neurological sciences in Paris from 1800 to 1950. Part III : Neurology. *Revue Neurologique*, 168(4), 301–320. https://doi.org/10.1016/j.neurol.2011.10.006. (pages 1, 5)

Camargo, C. H. F., Coutinho, L., Correa Neto, Y., Engelhardt, E., Maranhão Filho, P., Walusinski, O., & Teive, H. A. G. (2024). Jean-Martin Charcot : the polymath. *Arquivos de Neuro-psiquiatria*, 81, 1098–1111. https://www.thieme-connect.de/products/ejournals/pdf/10.1055/s-0043-1775984.pdf. (pages 1, 5, 6, and 16)

Camargo, C. H. F., Marques, P. T., de Oliveira, L. P., Germinian, F. M., de Paola, L., & Teive, H. A. G. (2018). Jean-Martin Charcot's Influence on Career of Sigmund Freud, and the Influence of this Meeting for the Brazilian Medicine. *Revista Brasileira de Neurologia*, 54(2). https://docs.bvsalud.org/biblioref/2018/07/907032/revista542v4-artigo6.pdf. (page 6)

Charcot, J. M. (1892). *Œuvres complètes de J. M. Charcot. Leçons sur les maladies du système nerveux.*, volume 1. Bureaux du progrès medical. https://patrimoine.sorbonne-universite.fr/viewer/3468/?offset=1#page=2&viewer=picture&o=&n=0&q=. (page 16)

Charcot, J.-M. (1897). *La foi qui guérit*. F. Alcan (Paris). https://gallica.bnf.fr/ark:/12148/bpt6k68008w. (page 16)

Chaudet, C. (2022). Les « Illuminati » du pamphlet au roman : circulations dun discours complotiste à grande échelle depuis le tournant du XIXe siècle. *Mots. Les langages du politique*, (pp. 19–36). https://www.cairn.info/revue-mots-2022-3-page-19.htm. (page 10)

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert : Pre-training of Deep Bidirectional Transformers for Language Understanding. (page 16)

Gabay, S., Petkovic, L., Bartz, A., Levenson, M. G., & Du Noyer, L. R. (2021). Katabase : À la recherche des manuscrits vendus. In *Humanistica 2021* (pp. 1–7). https://hal.science/hal-03066108. (page 10)

Ghermani, N. (2011). Confessions. In O. Christin (Ed.), *Dictionnaire des concepts nomades en sciences humaines* (pp. 117–133). Métailié. https://www.academia.edu/5335160/_Confession_. (page 9)

Giry, J. & Nouvel, D. (2022). Étudier les discours « conspirationnistes » et leur circulation sur Twitter : Les théories du complot comme objets du traitement automatique du

langage et de l'analyse des données textuelles. *Mots. Les langages du politique*, (pp. 37–55). https://www.cairn.info/revue-mots-2022-3-page-37.htm. (page 10)

Goetz, C. (2017). Charcot : Past and present. *Revue Neurologique*, 173(10), 628–636. https://doi.org/10.1016/j.neurol.2017.04.004. (page 5)

Gomes, M. d. M. & Engelhardt, E. (2013). Jean-Martin Charcot, father of modern neurology : an homage 120 years after his death. *Arquivos de Neuro-Psiquiatria*, 71, 815–817. https://doi.org/10.1590/0004-282X20130128. (page 6)

Grootendorst, M., Mishra, A., Matsak, A., OysterMax, Govil, P., Ogura, Y., Warmerdam, V. D., & yusuke1997 (2023). Maartengr/keybert : v0.8. https://doi.org/10.5281/zenodo.8388690. (page 17)

Hey, T., Tansley, S., & Tolle, K. M. (2009). Jim Gray on eScience : A Transformed Scientific Method. In T. Hey, S. Tansley, & K. M. Tolle (Eds.), *The Fourth Paradigm*. Microsoft Research. http://languagelog.ldc.upenn.edu/myl/JimGrayOnE-Science.pdf. (page 10)

Hobsbawm, E. (2010). *Age of revolution : 1789-1848*. Hachette UK. https://files.libcom.org/files/Eric%20Hobsbawm%20-%20Age%20Of%20Revolution%201789%20-1848.pdf. (page 9)

Johns, T. F. (1991). Should You be Persuaded. Two Samples of Data-Driven Learning Materials. https://api.semanticscholar.org/CorpusID:53988458. (page 10)

Jolly, A., Pandey, V., Singh, I., & Sharma, N. (2024). Exploring Biomedical Named Entity Recognition via SciSpacy and BioBERT models. *The Open Biomedical Engineering Journal*, 18(1). https://doi.org/10.2174/0118741207289680240510045617. (page 8)

Joyeux-Prunel, B. (2019). Visual Contagions, the Art Historian, and the Digital Strategies to Work on Them. *Artl@s Bulletin*, 8(3), 128–144. https://docs.lib.purdue.edu/artlas/vol8/iss3/8/. (page 10)

Joyeux-Prunel, B. & Gabay, S. (2022). Circulations des savoirs, de la recherche à l'enseignement. *Arabesques*. https://doi.org/10.35562/arabesques.2847. (page 10)

Kant, É. (1863). *Anthropologie d'un point de vue pragmatique* (trad. J. Tissot). Librairie Ladrange (originalement publié en 1798). https://fr.wikisource.org/wiki/Page:Kant_-_Anthropologie.djvu/452. (page 4)

Koehler, P. J. (2013). Chapter 6 – Charcot, La Salpêtrière, and Hysteria as Represented in European Literature. In S. Finger, F. Boller, & A. Stiles (Eds.), *Literature, Neurology, and Neuroscience : Neurological and Psychiatric Disorders*, volume 206 of *Progress in Brain Research* (pp. 93–122). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780444633644000235. (page 6)

Koselleck, R. & Richter, M. (2011). Introduction and Prefaces to the Geschichtliche Grundbegriffe : (Basic Concepts in History : A Historical Dictionary of Political and Social Language in Germany). *Contributions to the History of Concepts*, 6(1), 1–37. https://www.berghahnjournals.com/view/journals/contributions/6/1/choc060102.xml. (pages 8, 9)

Koyré, A. (1957). *From the Closed World to the Infinite Universe*, volume 1. Baltimore, Johns Hopkins Press. https://archive.org/details/fromclosedworldt0000koyr/page/2/mode/2up?q=%22revolution%22. (page 3)

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press. https://www.lri.fr/~mbl/Stanford/CS477/papers/Kuhn-SSR-2ndEd.pdf. (page 3)

Landais, É. (2014). « Frédéric Darbellay, éd., *La circulation des savoirs. Interdisciplinarité, concepts nomades, analogies, métaphores* » : Berne, P. Lang, 2012, 245 pages. *Questions de communication*, 26, 331–333. https://doi.org/10.4000/questionsdecommunication.9367. (page 7)

Le Pois, C. (1618). *Selectiorum observationum et consiliorum de praetervisis hactenus morbis affectibusque praeter naturum, ab aqua seu serosa colluvie et diluvie ortis, liber singularis.* Authore Carolo Pisone, Ponte ad Monticulum, apud Carolum Mercatorem. https://archive.org/details/BIUSante_05814/page/n3/mode/2up. (page 4)

Lecourt, D., Ed. (1999). *Dictionnaire d'histoire et philosophie des sciences*. Puf. https://www.librairiedalloz.fr/livre/9782130544999-dictionnaire-d-histoire-et-philosophie-des-sciences-4e-edition-dominique-lecourt/. (page 8)

Manjavacas, E., Long, B., & Kestemont, M. (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* (pp. 104–114). Minneapolis, USA : Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-2514. (page 10)

Mirbeau, O. & Michel, P. (1995). *Chroniques du diable*, volume 555. Presses Univ. Franche-Comté. https://mirbeau.asso.fr/darticlesfrancais/Marquer-Mirbeau%20et%20Charcot.pdfen. (page 5)

Monteiro, F., Nardi, A., & Gomes, M. (2021). The 400th anniversary of the birth of Thomas Willis (1621-1675) : an invaluable contributor to neuroscience. *Revista Brasileira de Psiquiatria*, 44. https://doi.org/10.1590/1516-4446-2021-2159. (page 5)

Navarro, D. F., Ijaz, K., Rezazadegan, D., Rahimi-Ardabili, H., Dras, M., Coiera, E., & Berkovsky, S. (2023). Clinical named entity recognition and relation extraction using natural language processing of medical free text : A systematic review. *International Journal of Medical Informatics*, 177, 105122. https://doi.org/10.1016/j.ijmedinf.2023.105122. (page 8)

Nemickienė, Ž. (2011). "Concept" in Modern Linguistics : the Component of the Concept "Good". *Filologija*, 16, 26–36. https://core.ac.uk/outputs/62656539?source=oai. (page 8)

Nerima, L., Seretan, V., & Wehrli, E. (2006). Le problème des collocations en TAL. *Nouveaux cahiers de linguistique française*, 27, 95–115. https://access.archive-ouverte.unige.ch/access/metadata/fc3fad28-5b90-42ec-bea5-0c6d54cb5452/download. (page 1)

Petkovic, L., Alrahabi, M., & Glenn, R. (2022). Impact de la correction automatique de locr/htr sur la reconnaissance dentités nommées dans un corpus bruité. *Journal of Information Sciences*, 21(2), 42–57. https://doi.org/10.34874/IMIST.PRSM/jis-v21i2.36599. (page 22)

Petkovic, L., Alrahabi, M., & Roe, G. (2023). Circulation du discours médical de jean-martin charcot. In *Humanistica 2023*. https://hal.science/HUMANISTICA-2023/hal-04107099v1. (page 10)

Quet, M. (2014). « Frédéric Darbellay, *La circulation des savoirs. Interdisciplinarité, concepts nomades, analogies, métaphores* ». *Revue d'anthropologie des connaissances*, 8(8-1). https://doi.org/10.3917/rac.022.0221. (page 7)

Rennesson, M., Georget, M., Paillard, C., Perrin, O., Pigeotte, H., & Tête, C. (2020). Le thésaurus, un vocabulaire contrôlé pour parler le même langage. *Médecine Palliative*, 19(1), 15–23. Documentation et pratiques documentaires en soins palliatifs. Coordonné par Caroline Tête. (page 8)

Rey, A. (1998). *Dictionnaire historique de la langue française. Tome 2.* Le Robert. https://www.plouffe.fr/simon/Dictionnaires/Le%20Robert%20Dictionnaire%20Historique%202a.pdf. (page 4)

Riffaterre, M. (1980). La trace de l'intertexte. *Pensée (La) Paris*, (215), 4–18. https://api.semanticscholar.org/CorpusID:170902390. (page 10)

Riguet, M. (2018). L'impact de la physiologie dans la critique littéraire de la fin du XIX^{ème} siècle : l'exemple de Claude Bernard. *Epistémocritique : Littérature et savoirs*. https://hal.science/hal-01903871. (page 10)

Robertson, S., Zaragoza, H., *et al.* (2009). The probabilistic relevance framework : Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4), 333–389. https://dx.doi.org/10.1561/1500000019. (page 16)

Robertson, S. E. & Jones, K. S. (1976). Relevance Weighting of Search Terms. *Journal of the American Society for Information science*, 27(3), 129–146. https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.4630270302?casa_token=TfyVkMGkDQsAAAAA:TCuXWzGHjo31RdxGR9jECRG2rZzqv0K3GOzHF7yAa2NfxtDFqxe-MmSHMC6e8OFiFxI4sLj2aW6OyDk. (page 15)

Roe, G., Fedchenko, V., & Nicolosi, D. M. (2023). Enlightenment Inflluencers : Networks of Text Reuse in 18th-century France. In *Digital Humanities 2023* (pp. 296–299). https://doi.org/10.5281/zenodo.8107964. (page 10)

Schopf, T., Klimek, S., & Matthes, F. (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* : SCITEPRESS – Science and Technology Publications. http://dx.doi.org/10.5220/0011546600003335. (page 19)

Stengers, I. (1987). *D'une science à l'autre. : Des concepts nomades*. Seuil. https://archive.org/details/dunesciencealaut0000unse. (page 9)

Tasca, C., Rapetti, M., Carta, M. G., & Fadda, B. (2012). Women And Hysteria In The History Of Mental Health. *Clinical Practice & Epidemiology in Mental Health : CP & EMH*, 8, 110–119. https://doi.org/10.2174/1745017901208010110. (pages 4, 5)

Teive, H. A. G., Coutinho, L., Camargo, C. H. F., Munhoz, R. P., & Walusinski, O. (2022). Thomas Willis' legacy on the 400th anniversary of his birth. *Arquivos de Neuro-Psiquiatria*, 80, 759–762. https://doi.org/10.1055/s-0042-1755278. (page 5)

Teive, H. A. G., Germiniani, F., Munhoz, R. P., & Paola, L. d. (2014). 126 hysterical years - the contribution of Charcot. *Arquivos de Neuro-Psiquiatria*, 72, 636–639. https://doi.org/10.1590/0004-282x20140068.PMID:25098481. (page 5)

Tubbs, R. S., Loukas, M., Shoja, M. M., Apaydin, N., Ardalan, M. R., Shokouhi, G., & Oakes, W. J. (2008). Costanzo Varolio (Constantius Varolius 1543–1575) and the Pons Varolli. *Neurosurgery*, 62(3), 734–737. https://doi.org/10.1227/01.neu.0000317323.63859.2a. (page 5)

Varet, V. (2023). Les nouvelles modalités numériques : *blockchain*, Web 3.0, NFT, métavers... *Legipresse*, 68(HS1), 59–70. https://doi.org/10.3917/legip.hs68.0059. (page 7)

Varolio, C. (1573). *De nervis opticis nonnullisq : aliis praeter communem opinionem in humano capite obseruatis.* Patavii : apud P. et A. Meiettos fratres. https://gallica.bnf.fr/ark:/12148/bpt6k325486q. (page 5)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention Is All You Need. https://arxiv.org/abs/1706.03762. (page 16)

Willis, T. (1664). *Cerebri anatome : cui accessit nervorum descriptio et usus.* Londini : Typis Ja. Flesher, impensis Jo. Martyn & Ja. Allestry, apud insigne Campanæ in Cœmeterio, D. Pauli. https://books.google.fr/books/?id=L2xEAAAAcAAJ&pg=PP9#v=onepage&q&f=false. (page 5)

Willis, T. (1681). *An Essay of the Pathology of the Brain and Nervous Stock in which Convulsive Diseases are Treated of.* London : Printed by J. B. for T. Dring. https://quod.lib.umich.edu/e/eebo/A66496.0001.001?rgn=main;view=fulltext. (page 5)

Wright, J. P. (1980). Hysteria and Mechanical Man. *Journal of the History of Ideas*, 41(2), 233–247. https://doi.org/10.2307/2709458. (page 4)