



CY CERGY PARIS UNIVERSITÉ
DU DATA ANALYST
UE RÉDACTION D'UN RAPPORT D'ANALYSE

Variations inter-disciplinaires des pratiques de citation dans les thèses

QUENTIN FOUCHÉ

9 septembre 2022

Résumé

Ce rapport présente une analyse des citations de références bibliographiques dans un corpus de 22 889 thèses électroniques, disponibles sur le site d'archive ouverte HAL. L'objectif était d'étudier les variations du nombre et du type de références citées par thèse en fonction de la discipline de la thèse. Les résultats ont montré que le nombre de références citées par thèse diffère significativement selon la discipline. En particulier, les doctorants de sciences humaines et sociales (SHS) ont cité plus de références que ceux des sciences dures. Une différence significative a également été retrouvée dans le type de référence citée : les citations d'articles scientifiques sont proportionnellement plus nombreuses en sciences dures qu'en SHS, contrairement aux citations de livres et chapitres de livres. Ces différences de pratique de citation entre les disciplines peuvent être reliées aux différences de normes dans la rédaction des publications scientifiques, ainsi qu'aux spécificités rhétoriques de chaque discipline.

Table des matières

Liste des figures	2
Liste des tableaux	3
Introduction	4
Méthodes	5
Présentation du jeu de données	5
Étude des données manquantes	5
Analyses statistiques	6
Résultats	8
Nombre de références citées par thèse et discipline	8
Type de référence citée et discipline	8
Nombre de références citées par thèse et type de référence	9
Discussion	10
Références	12
Annexe 1	14
Annexe 2	17

Liste des figures

Figure 1	Répartition des valeurs présentes et manquantes	7
Figure 2	Nombre de valeurs manquantes par combinaison de variables	7
Figure 3	Nombre de références citées par thèse en fonction de la discipline . .	8
Figure 4	Pourcentage de références citées par thèse en fonction de la discipline et du type de référence	9

Liste des tableaux

Tableau 1	Description des variables analysées	5
Tableau 2	Nombre de thèses par discipline	6
Tableau 3	Liste des disciplines du conseil national des universités	14
Tableau 4	Liste des disciplines du conseil national des universités (suite) . . .	15
Tableau 5	Liste des disciplines du conseil national des universités (suite) . . .	16
Tableau 6	Tests de Shapiro sur le nombre de références citées par thèse . . .	17
Tableau 7	Tests de Shapiro sur le pourcentage d'articles cités par thèse . . .	17

Introduction

En sciences, citer les sources utilisées pour énoncer une idée, une méthodologie ou encore des résultats d'expériences, dans le cadre d'une communication écrite ou orale, est indispensable pour permettre aux autres chercheurs d'évaluer la véracité de l'information. Cette pratique permet également aux chercheurs de placer leurs recherches dans le contexte de celles de leur communauté, tout en s'en démarquant en soulignant les aspects encore inconnus qui sont abordés dans leurs propres contributions (Hyland, 1999). La citation des sources se retrouve dans toutes les productions scientifiques écrites : articles de journal scientifique, livres, thèses, rapports de stage, compte-rendus de conférence... Parmi toutes ces productions, les thèses font partie de celles où sont citées le plus de références, du fait du volume élevé d'informations qu'elles contiennent. Cela en fait un objet d'étude particulièrement pertinent pour analyser la manière dont les chercheurs (dans ce cas-ci, des étudiants en doctorat) citent les travaux de leurs pairs.

Plusieurs études se sont déjà intéressées à la mobilisation des références bibliographiques dans les thèses. Ces travaux ont porté, entre autres, sur le nombre de références par thèse, le type de référence utilisée (*e.g.*, articles, livres, etc.), l'année de publication, le journal ou l'institution dans lequel a été publié la référence, ou encore la langue de la référence (Feyereisen et Spoiden, 2009 ; Gao et al., 2009 ; Hoffmann et Doucette, 2012 ; Kumar et Doa, 2009 ; Vallmitjana et Sabaté, 2008). D'autres études se sont plutôt focalisées sur l'analyse des mots et expressions utilisées pour citer les références dans le texte de la thèse (Soler-Monreal et Gil-Salom, 2011 ; Thompson, 2005).

Le plus souvent, les analyses de citations se font sur les thèses réalisées dans une même université ou dans une même discipline (Hoffmann & Doucette, 2012). Les comparaisons inter-disciplinaires, en particulier, sont plus rares, ou portent sur un nombre restreint de disciplines (*e.g.*, quatre disciplines comparées dans Gao et al., 2009). Du fait des différences de méthodologie entre les études réalisées dans des disciplines différentes, la comparaison directe de leurs résultats est souvent difficile, voire impossible (Hoffmann et Doucette, 2012). Pourtant, analyser les variations inter-disciplinaires dans la mobilisation des références permettrait de mieux comprendre le rôle du processus de citation dans la transmission des connaissances. À l'échelle des articles scientifiques, des différences de pratique entre disciplines ont déjà été identifiées : les auteurs qui publient en sciences humaines et sociales (SHS) citent davantage de références que ceux publiant en sciences dites dures (*i.e.*, mathématiques, chimie, physique, biologie, informatique et sciences de l'ingénieur) (Hyland, 1999). Cette différence se retrouve-t-elle à l'échelle des thèses ? Plus généralement, comment la mobilisation des références bibliographiques dans une thèse varie-t-elle selon la discipline de la thèse ?

Afin de répondre à cette problématique, cette présente étude s'est intéressée aux références bibliographiques citées dans un corpus de plusieurs milliers de thèses, réalisées dans onze disciplines différentes. L'approche, à la fois descriptive et hypothético-déductive, a consisté à analyser les variations du nombre et du type de références citées par thèse en fonction de la discipline de la thèse, en se focalisant en particulier sur les différences entre les SHS et les sciences dures. Au total, trois hypothèses ont été testées : (1) le nombre de références citées est plus élevé dans les thèses de SHS que dans celles de sciences dures ;

(2) parmi les références citées dans une thèse, la proportion d'articles scientifiques est plus élevée en sciences dures qu'en SHS, et inversement pour les livres et chapitres de livres ;
 (3) plus le nombre de références citées est faible, plus la proportion de livres et chapitres de livres cités parmi ces références est élevée.

Méthodes

Présentation du jeu de données

Le jeu de données analysé décrit les métadonnées des références bibliographiques citées dans 22 889 thèses de science, disponibles en ligne sur le site d'archive ouverte HAL (<https://hal.archives-ouvertes.fr>). Il est composé de 1 821 132 références et de quatre variables : l'identifiant de la thèse dans HAL (*i.e.*, son numéro TEL), la discipline de la thèse, le type de la référence, et son titre (Tableau 1). Toutes ces variables contiennent des données textuelles. Les disciplines sont catégorisées selon les groupes définis par le conseil national des universités (Annexe 1). Onze d'entre eux sont représentés dans ce jeu de données : les groupes I à X et le groupe pharmacie (dans la suite du texte, ces différents groupes sont désignés par le terme discipline). Les références citées dans les thèses sont de 11 types différents : article scientifique, livre, chapitre de livre, papier de conférence, thèse, rapport de stage, page web, manuscrit non publié, brevet, interview, film et communication personnelle.

Tableau 1. Description des variables analysées

Variable	Nombre de valeurs uniques	Données manquantes (%)
Identifiant de la thèse	22899	0
Discipline de la thèse	11	0
Type de la référence	12	14.8
Titre de la référence	1437703	10.5

Le nombre de thèses présentes dans ce jeu de données est de 29 en pharmacie et varie entre 229 et 9430 dans les disciplines I à X (Tableau 2). En raison du faible nombre de thèses de pharmacie par rapport aux autres disciplines, les références citées dans ces thèses n'ont pas été prises en compte dans les analyses.

Tableau 2. Nombre de thèses présentes dans le jeu de données par discipline

Discipline	Nombre de thèses
I	229
II	1359
III	654
IV	1003
V	2851
VI	2768
VII	1235
VIII	1551
IX	9430
X	1780
Pharmacie	29

Étude des données manquantes

Deux des quatre variables du jeu de données comportent un pourcentage non nul de valeurs manquantes : 14.8% pour le type de référence et 10.5% pour le titre des références (Tableau 1). Les références dont le type est inconnu ont été gardées telles quelles dans l'analyse du nombre de références par thèse, puis supprimées pour étudier les liens entre type de référence, nombre total de références et discipline. Les données liées au titre des références, quant à elles, n'ont pas été utilisées.

La répartition des valeurs manquantes dans le type et le titre des références est représentée dans la Figure 1. Les références y ont été triées de sorte à afficher en premier celles dont le type est connu. Cette répartition ne montre pas de régularité particulière dans la localisation des données manquantes entre ces deux variables (Figure 1). Ce constat est confirmé avec la Figure 2, qui indique le nombre de valeurs manquantes présentes uniquement dans le type de référence, uniquement dans le titre, ou simultanément dans les deux. Seules 14% des références n'ont à la fois pas de type et de titre, ce qui suggère une absence de lien dans le caractère manquant des données entre ces deux variables.

Analyses statistiques

Le nombre total de références citées par thèse et le pourcentage de chaque type de référence cité dans une thèse sont les deux paramètres sur lesquels ont porté les analyses. Pour étudier les variations de la mobilisation des références en fonction de la discipline, le nombre de références citées par thèse a d'abord été comparé entre les disciplines en utilisant le test de Kruskal-Wallis. Ce test non-paramétrique a été préféré à l'ANOVA car quelle que soit la discipline, la distribution du nombre de références citées par thèse ne suivait pas une loi normale (Annexe 2). Pour les mêmes raisons, ce même test a été appliqué pour

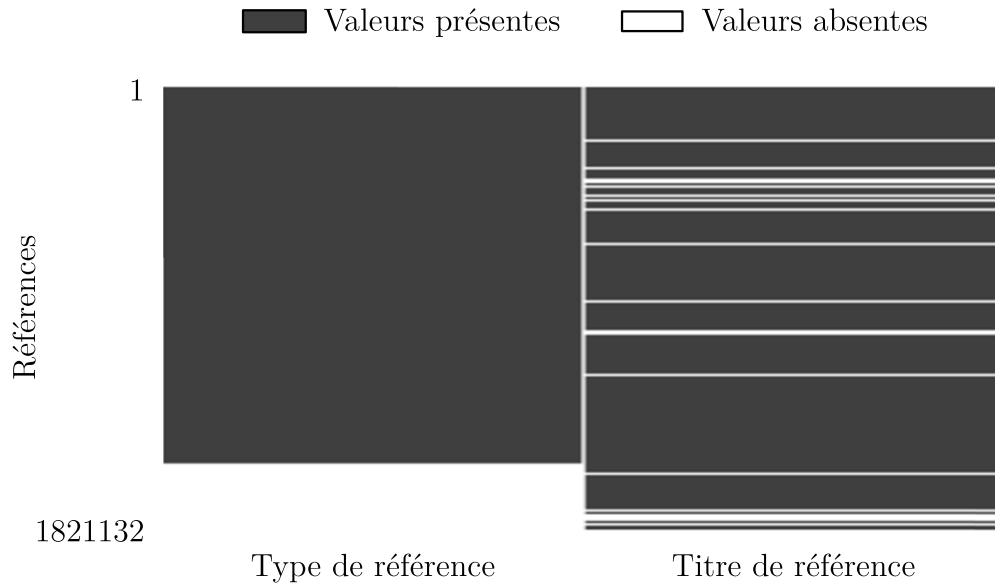


Figure 1. Répartition des valeurs présentes (en noir) et manquantes (en blanc) dans le type et le titre des références, pour toutes les références présentes dans le jeu de données.

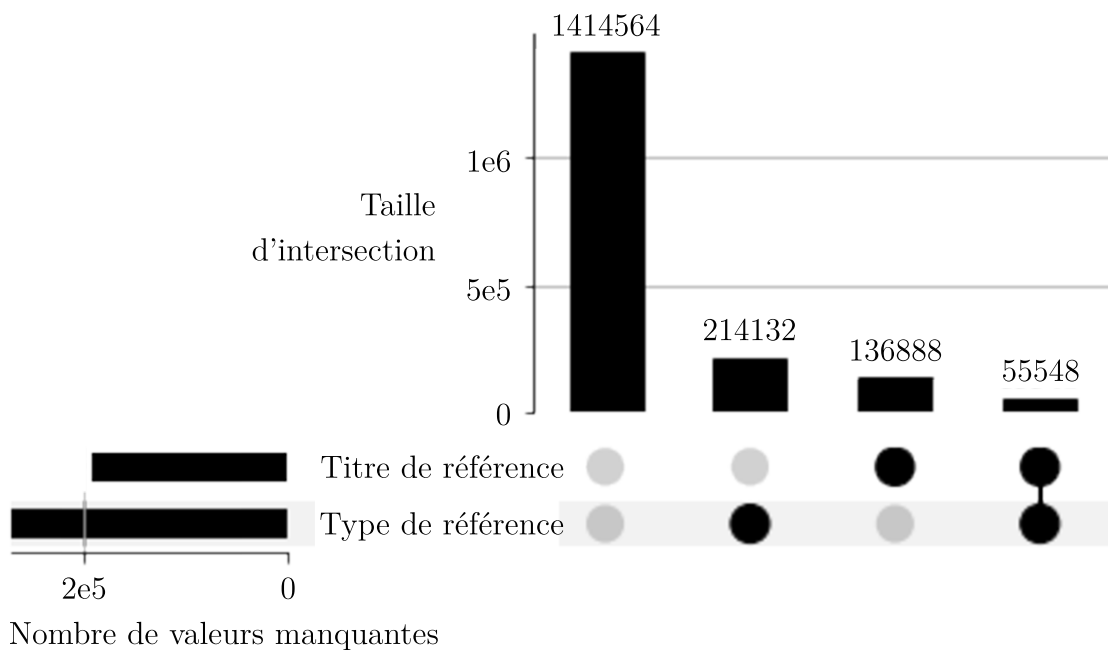


Figure 2. Nombre de valeurs manquantes dans le type et le titre des références. Les points colorés en noir indiquent dans quelle combinaison de variables sont situées les valeurs manquantes.

analyser les différences inter-disciplinaires dans le pourcentage d'articles scientifiques cités par thèse. Enfin, pour étudier le lien entre le pourcentage de livres et chapitres cités et le nombre total de références citées par thèse, une régression linéaire a été appliquée entre ces deux variables, associée à un test de corrélation de Pearson, en séparant les thèses de

SHS de celles de sciences dures. Toutes les analyses ont été réalisées avec le langage de programmation Python (v3.10.4). Le seuil de signicativité des tests a été fixé à $\alpha = 0.05$.

Résultats

Nombre de références citées par thèse et discipline

Le nombre moyen de références citées par thèse est représenté dans la Figure 3. Ce nombre diffère de manière significative entre les disciplines (test de Kruskal-Wallis : $H = 850$, $P < 0.001$). Plus précisément, les disciplines dans lesquelles le nombre de références est le plus élevé (entre 100 et 110 références citées par thèse en moyenne) sont les disciplines I, II, III et IV, c'est-à-dire toutes celles correspondant aux sciences humaines et sociales. Les doctorants des disciplines associées aux sciences dures citent en moyenne 10 à 35 références de moins que ceux des SHS (*i.e.*, entre 75 et 90 références par thèse).

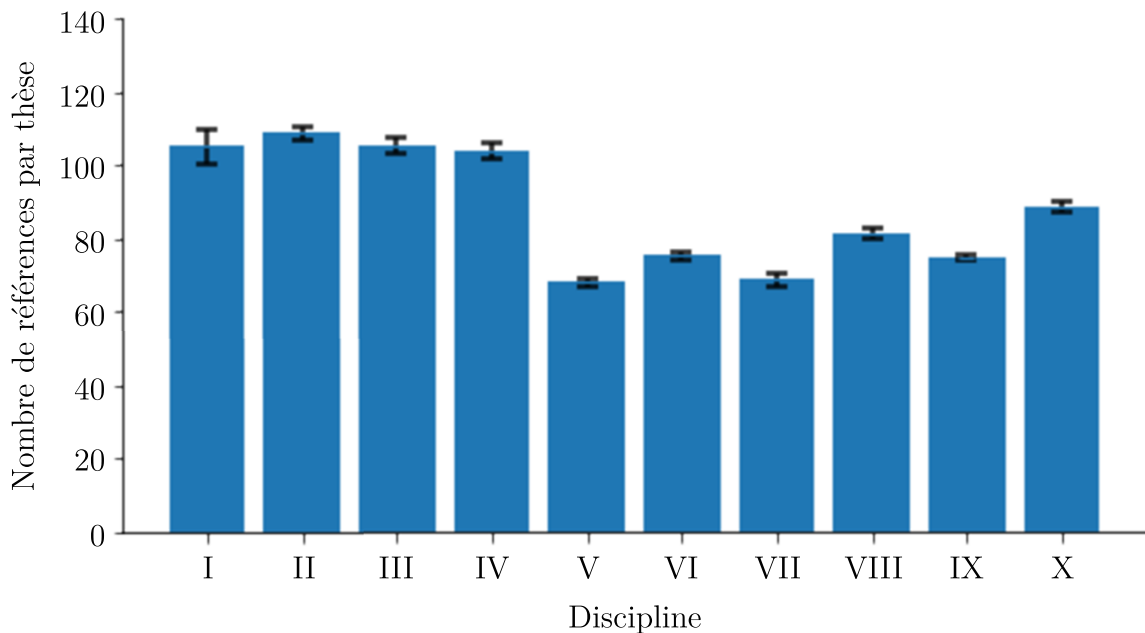


Figure 3. Nombre moyen de références citées par thèse en fonction de la discipline de la thèse. Les barres d'erreurs représentent l'erreur standard à la moyenne. Le nombre de thèses dans chaque discipline est donné dans le Tableau 2.

Type de référence citée et discipline

La Figure 4 représente le pourcentage moyen de références citées par thèse en fonction de la discipline et du type de référence. Le type le plus représenté quelle que soit la discipline (excepté la discipline III) est l'article de journal scientifique (entre 38 et 93% des

références). Viennent ensuite les livres et chapitres de livres, qui représentent respectivement entre 3 et 35% et entre 2 et 16% des références. Les autres types de référence sont comparativement très peu cités (moins de 2% des références), à l'exception des papiers de conférence, qui représentent de 0.3 à 7% des citations et jusqu'à 14% pour la discipline V.

L'analyse du pourcentage d'articles scientifiques cités par thèse montre que celui-ci diffère significativement entre les disciplines (test de Kruskal-Wallis : $H = 6606$, $P < 0.001$). Il est le moins élevé dans les disciplines de SHS (entre 38 et 65%) ainsi que dans la discipline V (58%), qui correspond aux mathématiques et à l'informatique. Les thèses des autres disciplines de sciences dures comportent entre 76 et 93% de citations d'articles. À l'inverse, le pourcentage de livres et chapitres de livres est plus élevé dans les disciplines de SHS (entre 33 et 60%) que dans celles de sciences dures (entre 6 et 24%).

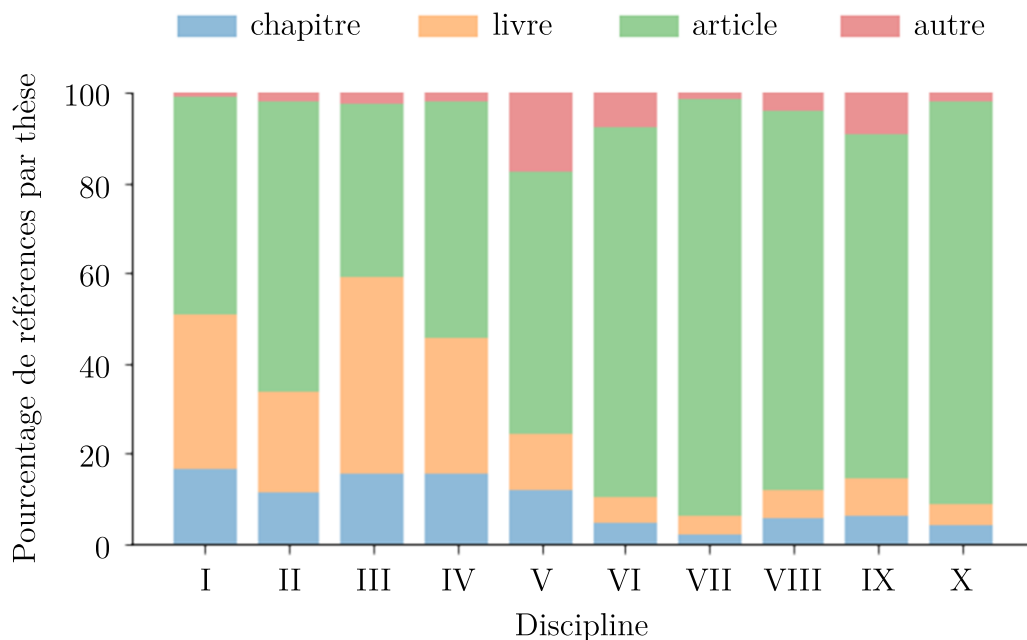


Figure 4. Pourcentage moyen de références citées par thèse en fonction de la discipline et du type de référence. Le nombre de thèses dans chaque discipline est donné dans le Tableau 2. La catégorie autre regroupe les types suivants : papiers de conférence, thèses, rapports de stage, pages web, manuscrits non publiés, brevets, interviews, films et communications personnelles.

Nombre de références citées par thèse et type de référence

La relation entre le pourcentage de livres et chapitres de livres cités dans une thèse et le nombre total de références citées dans la thèse a été analysée à l'aide d'une régression linéaire. Puisqu'il existe une différence nette dans le pourcentage de livres et chapitres cités par thèse entre les SHS et les sciences dures (cf. section précédente), la régression linéaire a été appliquée séparément sur chacun de ces deux champs disciplinaires.

En ce qui concerne les thèses de SHS, une corrélation positive significative est observée entre le pourcentage de livres et chapitres et le nombre de références (Test de corrélation de Pearson : $r = 0.041$, $P = 0.021$). La droite de régression a pour équation : $y = 42.06 + 0.017x$. La pente de cette droite indique que pour une augmentation de 100 références citées dans une thèse, le pourcentage de livres et chapitres cités augmentera en moyenne de 1.7%. L'erreur standard résiduelle du modèle, qui donne la différence typique entre une valeur prédite par le modèle et une valeur réelle, est de 23.26%¹.

Contrairement aux thèses de SHS, une corrélation négative significative est observée entre le pourcentage de livres et chapitres cités et le nombre total de références citées par thèse (Test de corrélation de Pearson : $r = -0.043$, $P < 0.001$; équation de la droite de régression : $y = 14.80 - 0.012x$). Selon cette régression, le pourcentage de livres et chapitres cités diminuera en moyenne de 1.2% toutes les 100 références ajoutées dans la thèse. L'erreur standard résiduelle du modèle est de 13.32%.

Discussion

Les analyses présentées dans ce rapport ont démontré tout d'abord qu'il existe un lien significatif entre le nombre de références citées dans une thèse et la discipline de la thèse, les thèses de SHS comportant davantage de références que celles de sciences dures. Un second lien entre mobilisation des références et discipline a été observé au niveau du type de référence citée : la proportion d'articles cités est plus élevée et celle de livres et chapitres plus faible en sciences dures qu'en SHS. Enfin, il existe une faible corrélation, positive en SHS et négative en sciences dures, entre le pourcentage de livres et chapitres cités et le nombre total de références citées dans une thèse. Ces trois résultats sont discutés un à un dans la présente section, en les confrontant aux hypothèses de départ ainsi qu'aux données disponibles dans la littérature. Certaines limites de l'étude sont ensuite exposées, avant de terminer sur des perspectives pour de futures recherches.

Le nombre plus élevé de références citées dans les thèses de SHS par rapport aux sciences dures confirme la première hypothèse de l'étude. Ce résultat est cohérent avec le plus grand nombre de citations en SHS dans les articles scientifiques, mis en évidence dans une précédente étude (Hyland, 1999). Selon l'auteur de ces travaux, cette différence de pratique peut en partie s'expliquer par des différences de normes dans la rédaction des publications. En sciences dures par exemple, les chercheurs réalisent leurs expériences en suivant des protocoles standardisées, généralement connus de leurs pairs, qu'il n'est pas nécessaire de référencer dans les publications (Hyland, 1999). Cette norme n'est pas retrouvée en SHS, où d'ailleurs les études expérimentales sont moins fréquentes. Par ailleurs, alors que l'argumentation en sciences dures s'appuie souvent sur des études proches, en SHS elle fait appel à un plus grand nombre de sujets, et donc de sources (Hyland, 1999). Les interprétations y sont aussi davantage soumises à débat, ce qui augmente le nombre de références à citer. Ces différences de pratique entre SHS et sciences dures pourraient cependant varier selon les régions, comme le suggère l'étude de Gao et al. (2009) sur des thèses réalisées en Chine, dans laquelle aucune différence notable n'a

¹L'erreur standard résiduelle a été calculée à l'aide d'une fonction créée par l'auteur sous Python.

été observée dans le nombre de références citées entre sciences dures et SHS.

En plus de citer plus de sources, les doctorants en SHS se réfèrent également à une plus grande proportion de livres et chapitres de livres, et moins aux articles de revues scientifiques, que les doctorants en sciences dures. Ce résultat, qui confirme la deuxième hypothèse de l'étude, est également retrouvé dans les travaux de Gao et al. (2009) : ces auteurs ont observé un écart de 30% d'articles cités en moins dans les thèses de SHS par rapport aux autres disciplines. Le fait que davantage de livres et chapitres soient cités en SHS peut être simplement dû à une plus grande abondance de ce type de ressources qu'en sciences dures. En philosophie par exemple, la plupart des communications se font sous forme de livres (Cullars, 1999). Par comparaison, en sciences dures, les sources ayant la plus grande valeur de preuve sont les articles de journaux à comité de lecture, qui sont systématiquement évalués par d'autres chercheurs avant d'être publiés.

La troisième hypothèse de l'étude portait moins sur les variations liées à la discipline que sur les différences inter-individuelles de pratique de citation. Elle supposait une corrélation négative entre le pourcentage de livres et chapitres cités et le nombre total de références citées par thèse. Les résultats tendent à confirmer cette hypothèse en sciences dures, où une corrélation négative a été observée. Cette corrélation pourrait être due à des différences d'utilisation des sources entre les livres et les articles. Par exemple, les livres pourraient être davantage utilisés pour supporter des propositions générales, notamment dans l'introduction des thèses, qui autrement nécessiteraient la citation d'un grand nombre d'articles. Citer un livre permettrait ainsi de faire l'économie de plusieurs citations d'articles, et donc de citer moins de références dans la thèse. En SHS, une corrélation positive a été observée entre le nombre de références par thèse et le pourcentage de livres et chapitres cités, réfutant l'hypothèse de départ. Ces deux corrélations, négative en sciences dures et positive en SHS, restent toutefois difficile à interpréter, étant donné la faible amplitude de la corrélation (la proportion de livres et chapitres cités diminue de seulement 1.2% en sciences dures et augmente de 1.7% en SHS toutes les 100 références) et la très faible adéquation du modèle aux données (l'écart moyen entre une valeur prédite par le modèle et les valeurs réelles est de 13.32% en sciences dures et 23.26% en SHS).

Dans l'ensemble, cette étude a mis en évidence des différences nettes de pratique de citation entre les doctorants de SHS et ceux des sciences dures. Si les analyses ont porté sur un nombre conséquent de thèses, elles n'ont cependant pas pris en compte un certain nombre de facteurs qui auraient permis d'affiner les résultats. Par exemple, la prise en compte de l'année de soutenance des thèses permettrait d'analyser l'évolution temporelle des différences de citation entre disciplines. Cela permettrait notamment de savoir si ces différences se sont accrues ou ont au contraire régressé avec le développement du numérique. Des informations sur la région géographique, ou sur l'institution dans laquelle a été réalisée la thèse, permettraient d'identifier d'éventuelles disparités spatiales dans les pratiques, comme le suggèrent les différences de résultats obtenus par Gao et al. (2009). Une autre limite de cette étude est la prise en compte exclusive des thèses électroniques. Bien que simplifiant grandement l'acquisition des données, l'analyse se trouve privée d'un grand nombre de thèses, la plupart plus anciennes, qui pourraient ne pas suivre les mêmes tendances que celles observées ici (Hoffmann et Doucette, 2012).

Afin d'approfondir les recherches sur les différences de pratique de citation entre SHS

et sciences dures, des études supplémentaires pourraient s'intéresser aux variations des pratiques vis-à-vis de l'âge des citations (*i.e.*, la différence entre la date de soutenance de la thèse et la date de publication de la référence) (Hoffmann et Doucette, 2012). Il est en effet probable que les références citées soient bien plus anciennes en philosophie ou en histoire, où les sources peuvent remonter jusqu'à l'Antiquité, qu'en mathématiques, physique ou biologie, où les connaissances acquises par le passé deviennent systématiquement obsolètes une fois remplacées par des connaissances plus récentes. Le type de référence citée pourrait également varier selon l'ordre de citation dans la thèse. Par exemple, les livres sont-ils davantage cités dans l'introduction des thèses en sciences dures par rapport aux SHS ? Enfin, comme cela a été observé dans les articles scientifiques, des différences entre disciplines pourraient exister dans la manière dont les références sont reportées dans le texte de la thèse (Hyland, 1999).

Références

Cullars, J. M. (1998). Citation characteristics of English-language monographs in philosophy. *Library & information science research* 20(1), 41-68. [https://doi.org/10.1016/S0740-8188\(98\)90005-6](https://doi.org/10.1016/S0740-8188(98)90005-6)

Feyereisen, P. et Spoiden, A. (2009). Can local citation analysis of master's and doctoral theses help decision-making about the management of the collection of periodicals? A case study in psychology and education sciences. *The Journal of Academic Librarianship* 35(6), 514-522. <https://doi.org/10.1016/j.acalib.2009.08.018>

Gao, S. J., Yu, W. Z. et Luo, F. P. (2009). Citation analysis of PhD thesis at Wuhan University, China. *Library Collections, Acquisitions, and Technical Services* 33(1), 8-16. <https://doi.org/10.1016/j.lcats.2009.03.001>

Hoffmann, K. et Doucette, L. (2012). A review of citation analysis methodologies for collection management. *College & Research Libraries* 73(4), 321-335. <https://doi.org/10.5860/crl-254>

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics* 20(3), 341-367. <https://doi.org/10.1093/applin/20.3.341>

Kumar, H. A. et Dora, M. (2011). Citation analysis of doctoral dissertations at IIMA: A review of the local use of journals. *Library Collections, Acquisitions, and Technical Services* 35(1), 32-39. <https://doi.org/10.1016/j.lcats.2011.03.002>

Monreal, C. S. et Salom, L. G. (2011). A cross-language study on citation practice in PhD theses. *International Journal of English Studies* 11(2), 53-75. <https://doi.org/10.6018/ijes/2011/2/149641>

Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes* 4(4), 307-323. <https://doi.org/10.1016/j.jeap.>

2005.07.006

Vallmitjana, N. et Sabaté, L. G. (2008). Citation analysis of Ph. D. dissertation references as a tool for collection management in an academic chemistry library. *College & Research Libraries* 69(1), 72-82. <https://doi.org/10.5860/crl.69.1.72>

Annexe 1

Tableau 3. Liste des disciplines du conseil national des universités

Groupe	Section	Discipline
I	01	Droit privé et sciences criminelles
I	02	Droit public
I	03	Histoire du droit et des institutions
I	04	Science politique
II	05	Sciences économiques
II	06	Sciences de gestion
III	07	Sciences du langage : linguistique et phonétique générales
III	08	Langues et littératures anciennes
III	09	Langue et littérature françaises
III	10	Littératures comparées
III	11	Langues et littératures anglaises et anglo-saxonnes
III	12	Langues et littératures germaniques et scandinaves
III	13	Langues et littératures slaves
III	14	Langues et littératures romanes : espagnol, italien, portugais, autres langues romanes
III	15	Langues et littératures arabes, chinoises, japonaises, hébraïques, d'autres domaines linguistiques
IV	16	Psychologie, psychologie clinique, psychologie sociale
IV	17	Philosophie
IV	18	Architecture (ses théories et ses pratiques), arts appliqués, arts plastiques, arts du spectacle, épistémologie des enseignements artistiques, esthétique, musicologie, musique, sciences de l'art
IV	19	Sociologie, démographie
IV	20	Anthropologie biologique, ethnologie, préhistoire
IV	21	Histoire, civilisation, archéologie et art des mondes anciens et médiévaux
IV	22	Histoire et civilisations : histoire des mondes modernes histoire du monde contemporain, de l'art, de la musique
IV	23	Géographie physique, humaine, économique et régionale
IV	24	Aménagement de l'espace, urbanisme

Tableau 4. Liste des disciplines du conseil national des universités (suite)

Groupe	Section	Discipline
V	25	Mathématiques
V	26	Mathématiques appliquées et applications des mathématiques
V	27	Informatique
VI	28	Milieus denses et matériaux
VI	29	Constituants élémentaires
VI	30	Milieus dilués et optique
VII	31	Chimie théorique, physique, analytique
VII	32	Chimie organique, minérale, industrielle
VII	33	Chimie des matériaux
VIII	34	Astronomie, astrophysique
VIII	35	Structure et évolution de la Terre et des autres planètes
VIII	36	Terre solide : géodynamique des enveloppes supérieures, paléo-biosphère
VIII	37	Météorologie, océanographie physique et physique de l'environnement
IX	60	Mécanique, génie mécanique, génie civil
IX	61	Génie informatique, automatique et traitement du signal
IX	62	Energétique, génie des procédés
IX	63	Génie électrique, électronique, photonique et systèmes
X	64	Biochimie et biologie moléculaire
X	65	Biologie cellulaire
X	66	Physiologie
X	67	Biologie des populations et écologie
X	68	Biologie des organismes
X	69	Neurosciences
XII	70	Sciences de l'éducation
XII	71	Sciences de l'information et de la communication
XII	72	Epistémologie, histoire des sciences et des techniques
XII	73	Cultures et langues régionales
XII	74	Sciences et techniques des activités physiques et sportives

Tableau 5. Liste des disciplines du conseil national des universités (suite)

Groupe	Section	Discipline
Théologie	76	Théologie catholique
Théologie	77	Théologie protestante
Pharmacie	85	Sciences physico-chimiques et technologies pharmaceutiques
Pharmacie	86	Sciences du médicament
Pharmacie	87	Sciences biologiques pharmaceutiques

Annexe 2

Tableau 6. Résultats des tests de Shapiro appliqués pour analyser la normalité de la distribution du nombre de références citées par thèse, pour chaque discipline.

Discipline	W	P
I	0.95	< 0.001
II	0.97	< 0.001
III	0.98	< 0.001
IV	0.96	< 0.001
V	0.94	< 0.001
VI	0.95	< 0.001
VII	0.88	< 0.001
VIII	0.94	< 0.001
IX	0.95	< 0.001
X	0.93	< 0.001

Tableau 7. Résultats des tests de Shapiro appliqués pour analyser la normalité de la distribution du pourcentage d'articles cités par thèse, pour chaque discipline.

Discipline	W	P
I	0.97	< 0.001
II	0.98	< 0.001
III	0.96	< 0.001
IV	0.99	< 0.001
V	0.98	< 0.001
VI	0.84	< 0.001
VII	0.70	< 0.001
VIII	0.86	< 0.001
IX	0.93	< 0.001
X	0.72	< 0.001