

Université de Montréal

**Les remerciements et leurs fonctions dans le
système de la reconnaissance scientifique**

par Adèle Paul-Hus

École de bibliothéconomie et des sciences de l'information
Faculté des arts et des sciences

Thèse présentée
en vue de l'obtention du grade de Philosophiae doctor (Ph. D.)
en sciences de l'information

Décembre, 2019

© Adèle Paul-Hus, 2019

Université de Montréal
École de bibliothéconomie et des sciences de l'information

Cette thèse intitulée
Les remerciements et leurs fonctions dans le système de la reconnaissance scientifique

présentée par
Adèle Paul-Hus

A été évaluée par un jury composé des personnes suivantes

Lyne Da Sylva, Présidente-rapporteuse
Vincent Larivière, Directeur de recherche
Nadine Desrochers, Codirectrice
Christine Dufour, Membre du jury
David Pontille, Examineur externe

Résumé

Dans la communauté scientifique, le statut d'auteur est associé à la reconnaissance et à la crédibilité auprès des pairs puisque l'autorat constitue un facteur essentiel dans la prise de décisions pour l'embauche et la promotion d'un chercheur, ainsi que dans l'attribution de prix et de financement. Par ailleurs, dans les publications savantes, il est pratique courante de remercier les individus, institutions et organismes subventionnaires qui ont contribué de diverses façons aux travaux de recherche menant à la publication. Bien qu'ils puissent être considérés comme un simple acte de courtoisie, les remerciements peuvent également être perçus comme un marqueur de capital symbolique. Les remerciements permettent de souligner des contributions extrêmement variées, de nature cognitive, technique et sociale, mais qui ne satisfont généralement pas les critères d'autorat. En ce sens, les remerciements permettent de mieux comprendre le rôle des individus et des organisations qui ont contribué à la recherche. Dans ce contexte, notre recherche vise à décrire les fonctions des remerciements dans le système de la reconnaissance scientifique pour mieux comprendre leur valeur, du point de vue du crédit scientifique, ainsi qu'à explorer les facteurs qui peuvent influencer leurs caractéristiques dans le contexte des pratiques d'attribution du crédit scientifique.

Notre étude s'intéresse d'abord à la distribution du crédit scientifique parmi l'ensemble des contributeurs d'un projet de recherche, soit les auteurs et les remerciés. L'analyse du nombre de contributeurs par article montre que les importantes différences disciplinaires traditionnellement observées en matière de collaboration sont grandement réduites lorsque les remerciés sont pris en compte. Nos analyses des caractéristiques des individus remerciés montrent que les disparités de genre, maintes fois démontrées en matière d'autorat, sont également présentes dans les remerciements où les femmes sont sous-représentées. De plus, les femmes ont tendance à remercier une proportion plus élevée de femmes que les hommes. En ce qui concerne le statut académique, nos résultats montrent que les individus remerciés qui ont déjà publié ont tendance à occuper une position plus élevée dans la hiérarchie académique que le reste des auteurs.

Notre analyse comparative du contenu des remerciements souligne d'importantes tendances disciplinaires en matière de types de contributions remerciées. Nos résultats fournissent la première démonstration empirique à grande échelle des variations qui existent entre les disciplines dans le contenu des remerciements. Finalement, notre analyse qualitative des termes contenus dans les remerciements met en lumière trois fonctions principales des remerciements : la description des contributions, la déclaration des responsabilités, ainsi que l'expression de la voix de l'auteur.

Mots-clés : autorat, capital symbolique, collaboration, contribution, crédit scientifique, reconnaissance scientifique, remerciements

Abstract

In the scientific community, authorship is associated with recognition and credibility and thus plays a central role in decisions related to the hiring and promotion of a researcher, and in the attribution of prizes and funding. In scholarly publications, it is also common practice to acknowledge individuals, institutions, and funding organizations that contributed in various ways to a research project. Although they can be considered as a simple scholar's courtesy, acknowledgements can also be perceived as a marker of symbolic capital. Acknowledgements shed light on extremely diverse contributions, of cognitive, technical and social nature, that generally do not meet authorship criteria. As such, acknowledgements allow for a better understanding of the role played by individuals and organizations that contribute to research. In this context, our research aims to describe acknowledgements' functions within the reward system of science to better understand their value, in terms of scientific credit, and to explore factors that can influence their characteristics in the context of credit attribution practices.

Our study addresses scientific credit distribution among all contributors of a research project, authors and acknowledgees. Our analyses of the number of contributors per paper show that the important disciplinary differences in team size, which traditionally characterize collaboration, are greatly reduced when considering acknowledgees as contributors. Our analyses of acknowledgees' sociodemographic characteristics show that gender disparities, repeatedly demonstrated in terms of authorship, are also present in the acknowledgements, where women are also under-represented. Moreover, women tend to acknowledge a higher proportion of women than men do. Regarding academic status, our results show that acknowledgees who have already published tend to have a higher position in the academic hierarchy than the rest of authors.

Our comparative analysis of acknowledgements' content highlights important disciplinary trends in terms of types of contributions mentioned. Our results constitute the first large-scale empirical demonstration of disciplinary variations in the content of acknowledgements. Finally, our qualitative analysis of acknowledgements' content emphasises three main functions of acknowledgements: the description of contributions, the responsibility disclaimers, and the expression of the authorial voice.

Keywords: acknowledgements, authorship, collaboration, contribution, reward system of science, scientific credit, symbolic capital

Table des matières

Résumé.....	i
Abstract.....	iii
Table des matières.....	v
Liste des tableaux.....	x
Liste des figures.....	xii
Remerciements.....	xiii
Introduction.....	1
Contexte et problématique.....	1
Objectifs.....	4
Questions de recherche.....	5
Chapitre 1. Revue de la littérature.....	7
1.1 La reconnaissance dans le champ scientifique.....	7
1.1.1 La structure normative de la science.....	7
1.1.2 Le système de la reconnaissance scientifique.....	11
1.1.3 Le champ scientifique.....	12
1.1.4 Le triangle de la reconnaissance.....	15
1.2 L'autorat scientifique.....	16
1.2.1 L'évolution des pratiques d'autorat.....	16
1.2.2 Les fonctions de l'autorat.....	19
1.3 Les facteurs qui influencent les pratiques d'attribution du crédit.....	22
1.3.1 Discipline.....	23
1.3.2 L'âge et le statut académiques.....	26

1.3.3 Le genre	27
1.4 Les remerciements dans le contexte de la communication savante	28
1.4.1 Cinquante ans de recherche sur les remerciements.....	29
1.4.2 Reconnaissance, crédit et remerciements	32
Chapitre 2. Méthodologie	38
2.1 Devis mixte	38
2.2 Collecte de données	40
2.2.1 Données de remerciements	40
2.3 Analyse des données de remerciements.....	45
2.3.1 Traitement automatique du langage naturel.....	45
2.3.2 Analyses bibliométriques.....	49
2.3.3 Analyse de correspondance.....	53
2.3.4 Analyse qualitative de contenu	55
2.4 Adéquation entre le devis méthodologique et les objectifs de la recherche	58
2.5 Qualité de la recherche.....	60
2.5.1 Volet quantitatif	60
2.5.2 Volet qualitatif	61
Chapitre 3. Résultats	63
3.1 La distribution du crédit scientifique	63
Abstract.....	65
Introduction.....	66
Material and methods.....	70

Results.....	72
Discussion and conclusion.....	78
Acknowledgement	81
References.....	81
3.2 Les individus remerciés	86
Abstract.....	89
Introduction.....	90
Material and methods.....	93
Results.....	98
Discussion.....	101
Conclusion	104
Acknowledgements.....	105
References.....	105
Appendix.....	111
3.3 Les types de contributions qui mènent aux remerciements	113
Abstract.....	116
Introduction.....	117
Data and Methods	120
Results.....	125
Discussion and Conclusion.....	131
Acknowledgements.....	134
References.....	134

Supporting Information.....	142
3.4 Le contexte d'utilisation des termes les plus fréquents dans les remerciements	161
Abstract.....	163
Introduction.....	163
Objective and Research Questions.....	167
Data and Methods	167
Results.....	171
Discussion and Conclusion.....	177
Acknowledgement	179
References.....	179
Appendix.....	186
Chapitre 4. Discussion	191
4.1 Les fonctions des remerciements dans le système de la reconnaissance scientifique ..	191
4.2 Les limites du co-autorat comme mesure de la collaboration.....	194
4.3 L'effet Saint-Matthieu et l'effet Matilda	197
4.4 Les différences disciplinaires.....	199
4.5 Les impacts de l'indexation à grande échelle des remerciements	201
4.6 Limites de la recherche	203
Conclusion	206
Apports de la recherche	208
Apports empiriques.....	208
Apports méthodologiques	208

Apports pratiques	209
Recherches futures	210
Bibliographie.....	212
Annexe 1. Données de l'article 1	241
Annexe 2. Données de l'article 2	245
Annexe 3. Données de l'article 3	246
Annexe 4. Données de l'article 4.....	247

Liste des tableaux

Tableau 1. Nombre d'articles qui inclut des remerciements et pourcentage du total des articles publiés, par année de publication et par discipline	44
Tableau 2. Performance de l'algorithme Wiki-Gendersort comparée à la performance de quatre bases de données associant un genre à des prénoms (adapté de Bérubé et al., 2020).	52
Tableau 3. Nombre et proportion des occurrences de noms personnels (auteurs et remerciés), par genre.....	53
Tableau 4. Arbre de codage : catégories de contenu des remerciements et leur définition.....	57

Article 1

Table 1. Number of 2015 papers, number (and percentage) of papers with acknowledgements, and number (and percentage) of papers with acknowledgees	73
---	----

Article 2

Table 1. Number of papers and acknowledgements mentioning individual names, by discipline	94
Table 2. Number of distinct name strings and distinct name strings who are also authors in WoS, by discipline	96
Table 3. Proportion of acknowledgees who are female as a function of the gender of the leading authors, by discipline.....	99
Table 4. Descriptive statistics for the number of publications, by discipline	111
Table 5. Descriptive statistics for the total field-normalized citations, by discipline.	111
Table 6. Descriptive statistics for the academic age, by discipline.....	112
Table 7. Descriptive statistics for the percentages of leading authorships, by discipline	112

Article 3

Table 1. Explained and cumulative variance for each axis.....	124
Table 2. Relative contributions of the factor to the element for disciplines (expressed as a percentage)	125
Table 3. Number of papers indexed in WoS (all and with funding acknowledgements) and percentage of papers with funding acknowledgements, by discipline (2015)	126
Table 4. Frequency of the 214 most frequent noun phrases, by discipline.....	143
Table 5. Quality of representation of the rows (cumulative contribution for each NP)	152
Table 6. Quality of representation of the columns (cumulative contribution for each discipline).....	160

Article 4

Table 1. Categories and definitions.....	170
Table 2. Acknowledgement words coding results	172
Table 3. References of the acknowledgement excerpts cited	186

Liste des figures

Figure 1. Devis méthodologique.....	38
Figure 2. Extrait de la notice Web of Science pour la publication http://dx.doi.org/10.1016/j.carbpol.2016.08.078	41
<i>Article 1</i>	
Figure 1. Cumulative distribution of papers (with acknowledgements) (%), as a function of numbers of authors.....	74
Figure 2. Distribution of papers by number of authors (a) and acknowledgees (b).....	75
Figure 3. Mean number of authors and acknowledgees, by discipline	77
Figure 4. Mean number of acknowledgees by number of authors	78
<i>Article 2</i>	
Figure 1. Proportion of female authors, acknowledgees, and acknowledgees who are also authors by discipline.	98
Figure 2. Distribution of researchers (all WoS authors, and acknowledgees who are also authors) by number of publications, total field-normalized citations, academic age, and percentage of leading authorships.....	100
<i>Article 3</i>	
Figure 1. Bidimensional Correspondence Analysis for acknowledgements patterns by discipline (plane 1-2).....	129
Figure 2. Bidimensional Correspondence Analysis for acknowledgements patterns by discipline (plane 3-4).....	130
Figure 3. Frequency distribution of noun phrases found in acknowledgements	142

Remerciements

Merci d'abord à mes directeurs de recherche, Vincent Larivière et Nadine Desrochers, pour leur accompagnement et leur support indéfectible au cours des cinq dernières années. Leur présence et leur soutien ont contribué à faire de mon parcours doctoral une expérience positive et enrichissante.

Je remercie également mes collaborateurs, Rodrigo Costas, Adrián A. Díaz-Faes, Philippe Mongeon et Maxime Sainte-Marie pour leur aide et leur contribution à mes travaux de recherche.

Je tiens également à remercier les membres du jury, Lyne Da Sylva, Christine Dufour et David Pontille, ainsi que les membres de mon comité de recherche, Clément Arsenault et Audrey Laplante pour leur relecture attentive de ma thèse et leurs suggestions.

Pour leur soutien financier, je remercie le Conseil de recherches en sciences humaines du Canada, les Fonds de recherche du Québec – Société et culture, la Chaire de recherche du Canada sur les transformations de la communication savante, l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal, le Centre interuniversitaire de recherche sur la science et la technologie, la Beta Phi Mu Honor Society, ainsi que l' Association for Information Science and Technology.

Merci à toute l'équipe de la Chaire de recherche du Canada sur les transformations de la communication savante, et mes collègues du doctorat, pour leur présence, leur soutien moral et les nombreux échanges (sérieux ou moins). Une mention spéciale à Antoine Archambault, Nicolas Bérubé, Timothy Bowman, Sarah Cameron-Pesant, Gita Ghiasi, Stefanie Haustein, Philippe Mongeon, Diane Marie Plante, Maxime Sainte-Marie et Fei Shu.

Finalement, merci à Laurent et Mathilde d'être là et de donner un sens à tout ça.

Introduction

Dans la première publication qu'il consacre à l'étude des remerciements en 1991, Blaise Cronin positionne les remerciements aux côtés des citations et de l'autorat au sein du système de la reconnaissance scientifique. Plus de deux décennies plus tard, les remerciements occupent toujours un espace périphérique dans ce système. Notre recherche vise à décrire les fonctions des remerciements dans le système de la reconnaissance scientifique et à explorer les facteurs qui peuvent influencer leurs caractéristiques dans le contexte des pratiques d'attribution du crédit scientifique.

Dans les sections suivantes, nous présentons la problématique à l'origine de notre recherche et les objectifs de notre étude. Nous présentons ensuite la revue de la littérature, qui vise à ancrer notre étude dans un cadre conceptuel pertinent et à la situer au sein de la littérature existante. Une description détaillée est faite de la méthodologie utilisée. Les résultats sont ensuite présentés sous la forme de quatre articles de recherche publiés dans des revues savantes et d'une discussion qui permet de faire la synthèse de ces résultats, ainsi qu'une conclusion.

Contexte et problématique

La communauté scientifique est soumise à un ensemble de normes desquelles émerge le système de reconnaissance scientifique (*reward system of science*) tel que proposé par Merton (1973). Au cœur de ce système, la notion d'auteur scientifique est associée au capital scientifique — cette forme de capital symbolique ayant une valeur spécifique dans le champ scientifique (Bourdieu, 2001). Dans le champ scientifique, le capital symbolique associé à un travail savant est typiquement octroyé par la signature d'une publication, c'est-à-dire l'attribution du statut d'auteur. Pour un chercheur, le statut d'auteur est associé à la reconnaissance et à la crédibilité auprès de ses pairs puisque l'autorat constitue un facteur essentiel dans la prise de décisions pour l'embauche et la promotion d'un chercheur, ainsi que dans l'attribution de prix et de financement.

Toutefois, dans les publications savantes, il est également pratique courante de remercier les individus, institutions et organismes subventionnaires qui ont contribué de diverses façons

aux travaux de recherche menant à la publication (Cronin, 1995). En effet, on retrouve aujourd'hui des remerciements dans plus de la moitié des articles scientifiques (Paul-Hus, Desrochers et Costas, 2016). Bien qu'ils puissent être considérés comme un simple acte de courtoisie (*scholar's courtesy* [Cronin, 1995]), les remerciements peuvent également être perçus comme un marqueur de capital symbolique. Les remerciements permettent de souligner des contributions extrêmement variées, de nature cognitive, technique et sociale, mais qui ne satisfont généralement pas les critères d'autorat. En ce sens, les remerciements permettent souvent de comprendre le rôle des individus et des organisations qui ont contribué à la recherche. Cependant, les critères permettant de déterminer le statut d'un contributeur à un projet de recherche — auteur ou « remercié » — varient grandement selon la discipline étudiée et la position hiérarchique des individus dans le champ scientifique (Birnholtz, 2006). Cronin (1995) a notamment montré que les conflits émergeant d'interprétations divergentes des critères d'autorat ne sont pas un fait rare dans le domaine scientifique, certains contributeurs se sentant lésés par leur statut de remercié plutôt que d'auteur.

Les remerciements offrent également de l'information quant aux réseaux de collaboration entre chercheurs, révélant en quelque sorte une portion du « collège invisible » qui lie les chercheurs les uns aux autres (*invisible college* [Crane, 1972; Price, 1963; Price et Beaver, 1966]). Les remerciements ouvrent ainsi une fenêtre sur les collaborations informelles permettant d'aller au-delà des analyses bibliométriques traditionnelles de citations et d'autorat (Costas et van Leeuwen, 2012; Cronin, 1995) et de mieux cerner les pratiques d'attribution du crédit scientifique.

Dans le contexte de la communication savante, les remerciements ont une réputation ambiguë, d'abord en raison de leur nature parfois personnelle (Hyland, 2003), de leur manque de standardisation (Cronin, 2014) ainsi que du fait qu'ils soient perçus, dans certains cas du moins, comme des gestes intéressés, profitant davantage à leur auteur qu'aux remerciés eux-mêmes (Coates, 1999). Certains auteurs pourraient en effet tirer avantage à s'associer, par le truchement des remerciements, à des chercheurs de renom. Malgré cette réputation, les remerciements ont été au cœur de plus de deux cents publications savantes (articles scientifiques, notes éditoriales, chapitres de livre et thèses de doctorat) depuis les années 1960 (Desrochers, Paul-Hus et Pecoskie, 2017). L'étude des remerciements ne constitue donc plus un

champ de recherche émergent; mais se déploie plutôt sous de multiples formes et emprunte à diverses disciplines et approches théoriques. Toutefois, aucun consensus clair quant à la valeur et aux fonctions des remerciements ne semble se dégager de cette littérature. La nature variée et multidisciplinaire des études s'étant penchées sur les remerciements contribue à la fois à la richesse de ce corpus et à sa complexité. Alors qu'une majorité de ces études présentent des données empiriques, certaines proposent plutôt des contributions théoriques et critiques. Cette littérature n'en demeure pas moins cohérente en ce qui a trait aux aspects fondamentaux des remerciements qui méritent d'être analysés : le contenu des remerciements (*qui* et *quoi* remercier) ainsi que la forme des remerciements (*où* et *comment* remercier) (Desrochers, Paul-Hus et Pecoskie, 2017). Une analyse s'appuyant sur des données empiriques agrégées à grande échelle et couvrant des disciplines variées est aujourd'hui nécessaire afin de comprendre la valeur et les fonctions des remerciements dans le contexte scientifique.

Depuis 2008, la base de données bibliographiques Web of Science (WoS) collecte les informations qui se trouvent dans la section des remerciements (*acknowledgements*) des articles scientifiques indexés dans le *Science Citation Index*. Les informations ainsi recueillies permettent, entre autres, d'identifier les sources de financement d'une publication (Web of Science, 2009). Dès les années qui suivent l'ajout des remerciements dans les notices bibliographiques de WoS, plusieurs publications mettent à profit ces données nouvellement disponibles à grande échelle pour étudier l'impact du financement en termes de publications et de citations (p. ex. Boyack et Jordan, 2011; Lewison, 2009; Lyubarova, Itagaki et Itagaki, 2009; Rigby, 2011; Shapira et Wang, 2010). On peut désormais distinguer une branche à part entière dans la littérature sur les remerciements dans laquelle les remerciements constituent uniquement une source de données pour l'étude du financement de la recherche scientifique et où la valeur symbolique associée aux remerciements n'est que rarement abordée (Desrochers, Paul-Hus et Pecoskie, 2017). Leur présence semble toutefois hautement variable selon les disciplines étudiées (Paul-Hus, Desrochers et Costas, 2016).

Afin de mieux comprendre le capital symbolique associé aux remerciements, notre recherche vise à décrire les fonctions des remerciements, ainsi qu'à explorer les facteurs qui peuvent influencer le contenu des remerciements et plus globalement les pratiques d'attribution du crédit dans le monde scientifique. Mieux comprendre le rôle des remerciés, notamment par

rapport à celui des auteurs, dans le processus de la recherche et de la communication scientifique, revêt une grande pertinence pour l'ensemble de la communauté scientifique, pour laquelle le concept d'auteur est d'une importance capitale. L'évaluation des chercheurs, que ce soit dans le cadre de promotions ou de demandes de subvention, repose sur la notion centrale de l'auteur scientifique. Ainsi, que l'évaluation soit faite par les pairs ou par la compilation d'indicateurs bibliométriques, elle s'appuie sur la production d'un chercheur, c'est-à-dire les publications auxquelles il a contribué en tant qu'auteur. Bien que les remerciements aient suscité nombre d'études et qu'il ait été suggéré par plusieurs de les considérer aux côtés de l'autorat et/ou des citations (notamment Costas et van Leeuwen, 2012; Cronin et Weaver, 1995; Mackintosh, 1972; McCain, 2018), il demeure que les remerciements ont peine à se positionner clairement dans le système de la reconnaissance scientifique.

Objectifs

Notre recherche doctorale vise à décrire les fonctions des remerciements dans le système de la reconnaissance scientifique pour mieux comprendre leur valeur du point de vue du crédit scientifique, ainsi qu'à explorer les facteurs qui peuvent influencer leurs caractéristiques dans le contexte des pratiques d'attribution du crédit scientifique. Les fonctions des remerciements sont ici opérationnalisées par la description et la caractérisation du contenu des remerciements en termes d'individus et de contributions, dans la perspective du système de la reconnaissance scientifique où le principal mode d'attribution du crédit est l'autorat. Cette étude s'appuie donc sur le cadre conceptuel du système de la reconnaissance scientifique de Merton (1973), du capital symbolique de Bourdieu (1975) et du triangle de la reconnaissance dans lequel Cronin (1995) positionne les remerciements aux côtés de l'autorat et des citations.

Notre recherche se déploie autour de trois axes principaux : 1) l'attribution du crédit scientifique, 2) les individus remerciés et 3) les contributions qui mènent aux remerciements. Les remerciements indexés dans la base de données Web of Science seront donc analysés selon deux approches : l'analyse quantitative à grande échelle qui vise d'abord à décrire le contenu des remerciements à l'aide d'un traitement automatique du langage naturel qui permet d'extraire les syntagmes nominaux (contributions remerciées) et les entités nommées (individus remerciés), puis à mettre en relation les caractéristiques des remerciements avec différentes

métadonnées des articles qui les contiennent (auteurs et disciplines). Dans un deuxième temps, une analyse qualitative du contenu des remerciements permettra de contextualiser les résultats de l'approche quantitative.

Questions de recherche

Nos objectifs se traduisent par quatre questions de recherche (QR) qui sont alignées sur les trois axes principaux de notre projet.

QR1. De quelle façon le crédit scientifique est-il distribué au sein des équipes de recherche par les individus ayant contribué à une publication scientifique?

A. Quelle proportion des contributeurs est créditée comme auteurs et quelle proportion est créditée comme individus remerciés?

i. Comment cela varie-t-il selon les disciplines?

B. Comment le nombre d'individus remerciés varie-t-il en fonction du nombre d'auteurs signant une publication?

i. Comment cela varie-t-il selon les disciplines?

QR2. Quelles sont les caractéristiques des individus remerciés dans les publications savantes?

A. Quel est le genre¹ des individus remerciés?

i. Quelle proportion des individus remerciés est de genre féminin?

a. Comment cela varie-t-il en fonction du genre des auteurs?

b. Comment cela varie-t-il selon les disciplines?

B. Quel est le statut académique des individus remerciés?

i. Quelle proportion des individus remerciés est également auteur de publications scientifiques?

¹ Le genre des individus remerciés réfère aux caractéristiques linguistiques de leur nom et non au genre des individus eux-mêmes.

- a. Combien de publications ces individus remerciés ont-ils signées en tant qu'auteurs?
 - b. Sur quelle proportion de ces publications les individus remerciés occupent-ils une position d'auteur principal (premier auteur ou auteur de correspondance)?
 - c. Combien de citations ces individus remerciés ont-ils reçues en tant qu'auteurs?
 - d. Quel est l'âge académique de ces individus remerciés?
- ii. Comment ces indicateurs varient-ils selon les disciplines?

QR3. Quels types de contributions mènent aux remerciements?

- A. Quels syntagmes nominaux sont les plus fréquents dans les remerciements?
- B. Comment cela varie-t-il selon les disciplines?

QR4. Dans quels contextes sont utilisés les termes les plus fréquents dans les remerciements?

- A. Comment cela varie-t-il selon les disciplines?
- B. L'analyse qualitative de ces termes peut-elle révéler de nouvelles avenues pour une utilisation contextualisée des remerciements dans les études à grande échelle?

Quatre articles permettent d'aborder les quatre questions de recherche et se détaillent de la façon suivante :

- L'article 1 *The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements* vise à répondre à la QR1.
- L'article 2 *Who are the acknowledgees? An analysis of gender and academic status* vise à répondre à la QR2.
- L'article 3 *Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences* vise à répondre à la QR3.
- L'article 4 *Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015* vise à répondre à la QR4.

Chapitre 1. Revue de la littérature

Ce chapitre vise à présenter le contexte théorique sur lequel notre projet s'appuie et à positionner notre thèse au sein de la littérature existante. Nous présentons d'abord les travaux sur le champ scientifique et son système de reconnaissance. À la lumière de ce cadre théorique formé de la rencontre de trois positionnements théoriques, nous examinons les fonctions de l'autorat dans le système de la reconnaissance scientifique ainsi que l'évolution des pratiques d'autorat dans le temps, puisque les remerciements sont étroitement liés à ces pratiques. Nous aborderons par la suite les principaux facteurs qui influencent plus largement les pratiques d'attribution du crédit scientifique. Finalement, nous ferons un survol de la littérature s'étant intéressée aux remerciements dans le contexte de la communication savante.

1.1 La reconnaissance dans le champ scientifique

Les travaux des sociologues Robert K. Merton et Pierre Bourdieu se situent aux fondements de la sociologie des sciences. Les visions complémentaires du système de la reconnaissance scientifique de Merton et du capital scientifique de Bourdieu serviront d'assises théoriques à l'étude des fonctions des remerciements dans le contexte de la communication savante. Les travaux de Blaise Cronin permettent quant à eux de faire le pont entre le cadre théorique offert par Merton et Bourdieu et les remerciements comme objet d'étude. En effet, ce sont les travaux de Cronin qui permettent d'associer les remerciements à l'attribution du crédit scientifique et de résolument les positionner au sein du système de la reconnaissance scientifique. Les pages qui suivent présentent la structure normative de la science et le système de la reconnaissance scientifique proposés par Merton, la perspective de Bourdieu élaborée autour des notions de champ, d'*illusio*, et de capital, ainsi que le concept de triangle de la reconnaissance proposé par Cronin.

1.1.1 La structure normative de la science

Selon Merton (1973), l'institution scientifique est une structure sociale qui s'est dotée de normes et d'un système de récompense qui lui est propre. La principale forme de récompense, dans ce système, est la reconnaissance (*recognition*) accordée par les pairs pour l'originalité des travaux

et découvertes d'un chercheur. Cette reconnaissance est un témoignage socialement validé du fait qu'un chercheur a rempli les exigences associées à son rôle dans l'institution scientifique.

Pour comprendre le système de la reconnaissance scientifique envisagé par Merton, il faut d'abord examiner la structure normative de laquelle il découle et avec laquelle il interagit pour former l'institution sociale scientifique. Au début des années 1940, Merton propose une structure normative de la science (Merton, 1942; 1973). Ayant évolué depuis le XVI^e siècle d'une institution visant en partie à justifier et à valider l'existence de Dieu, l'institution scientifique moderne est devenue, au fil des siècles, autoportante, ne nécessitant pas de justification autre que l'avancement des connaissances scientifiques elles-mêmes. L'institution scientifique moderne s'est alors dotée d'une structure sociale autonome, un éthos propre au monde scientifique. L'éthos de la science est un ensemble de règles, de prescriptions, de mœurs, de croyances, de valeurs et de présuppositions liant les scientifiques les uns aux autres ainsi qu'à leur communauté. Quatre normes scientifiques, ou impératifs institutionnels constituent l'éthos de la science : l'universalisme, le communalisme, le désintéressement et le scepticisme organisé² (Merton 1942; 1973). Selon Merton, ces normes seraient, dans une certaine mesure, intériorisées par les chercheurs, et orienteraient leurs comportements. Plusieurs critiques reprocheront à Merton de présenter une vision idéalisée de l'institution scientifique comme un ensemble social relativement harmonieux où chaque élément occupe une place nécessaire (Bourdieu, 2001; Knorr-Cetina, 1991; Turner, 2007). Ainsi, les normes de Merton peignent un portrait de l'institution scientifique idéale, et devraient en ce sens être envisagées comme des normes prescriptives plutôt que descriptives.

Universalisme

Selon Merton, la norme d'universalisme stipule que la confirmation ou le rejet d'une découverte en tant que connaissance scientifique certifiée par la communauté ne doit en aucun cas dépendre des attributs sociodémographiques de son découvreur. L'ethnicité, la nationalité, le genre, l'appartenance religieuse, la classe sociale et les qualités personnelles d'un chercheur sont tous des attributs non pertinents face à la véracité d'un énoncé scientifique. Merton mentionne à cet

² Notre traduction de *universalism, communism, disinterestedness* et *organized skepticism*.

effet que l'objectivité scientifique exclut le particularisme, c'est-à-dire que des critères particuliers (non universels) de validité ne peuvent être invoqués pour valider ou invalider une découverte. Les découvertes et les énoncés scientifiques doivent donc être jugés selon un cadre de référence universel; en somme, toute découverte scientifique doit être étudiée indépendamment de son découvreur et de ses attributs particuliers.

Dans les faits, la norme est toutefois fréquemment contredite. En effet, dans la pratique, il s'avère difficile, voire impossible, de faire abstraction du statut d'un chercheur au sein de la hiérarchie académique lors de l'évaluation de la véracité d'une découverte, que ce biais soit conscient ou non (Blank, 1991; Bornmann, 2011; Ziman, 2000).

Communalisme

Selon la norme de communalisme, la science est un bien commun. Les produits de la recherche scientifique sont le résultat de la collaboration sociale et constituent donc un bien public. Par conséquent, les droits individuels du producteur devraient être limités. Ainsi, selon la vision de Merton, les droits de propriété intellectuelle d'un chercheur devraient se limiter à la reconnaissance (*recognition*) et à l'estime des pairs. Selon Merton, le secret scientifique et la notion de propriété privée illustrent donc l'antithèse de la norme du communalisme.

Toutefois, dans le contexte scientifique contemporain, la multiplication des brevets (Henderson, Jaffe et Trajtenberg, 1998a, 1998b; Mowery et al., 2001) et les coûts croissants exigés pour l'accès à de nombreuses revues savantes — qui permettent en retour aux grands éditeurs savants d'engranger des profits faramineux (Larivière, Haustein et Mongeon, 2015) — constituent des exemples flagrants de situations où la propriété privée est privilégiée au détriment de la norme de communalisme. En contrepartie, le mouvement du libre accès aux publications scientifiques (voir le *Budapest Open Access Initiative* (2002) qui présente les principes fondateurs du mouvement) et les initiatives de science citoyenne (Franzoni et Sauermann, 2014) sont des exemples patents de la mise en application de cette norme.

Désintéressement

Tel que défini par Merton, l'impératif institutionnel du désintéressement stipule que l'intérêt personnel des chercheurs ne devrait pas guider leurs comportements; la quête de la vérité devrait

être l'unique motivation du scientifique. Selon cette norme, seule la vérité doit primer dans l'investigation scientifique. Merton précise néanmoins que le désintéressement au niveau institutionnel n'est pas nécessairement applicable au niveau motivationnel. Le désintéressement ne découlerait donc pas des qualités morales des chercheurs telles que l'altruisme et l'honnêteté, mais plutôt du système normatif de l'institution qui récompense les travaux et les découvertes scientifiquement valides.

Cependant, dans le contexte contemporain, la compétition entre les chercheurs, intensifiée par l'importance accordée à l'originalité et à la priorité des découvertes en tant que critères de réussite, peut créer un incitatif à la fraude dans le but d'éclipser la compétition. Dans les faits, plusieurs stratégies peuvent être adoptées dans l'optique de favoriser son autopromotion (p. ex. formation de cartel de citations, saucissonnage des publications [Martin, 2016]).

Scepticisme organisé

Selon Merton, la norme de scepticisme organisé stipule qu'*a priori*, un énoncé scientifique est faux jusqu'à preuve du contraire. Les découvertes scientifiques sont donc soumises à un examen critique par les pairs avant d'être acceptées et certifiées comme connaissance scientifique. Une caractéristique fondamentale de la science, qui découle directement de cette norme, est le fait que les connaissances scientifiques peuvent être réévaluées et invalidées en tout temps. Selon Merton, la vérité, au sens scientifique, ne se trouve pas auprès d'une autorité, mais nécessite plutôt preuves et expérimentations.

La compétition au sein de la communauté scientifique favorise le scepticisme par le processus de révision par les pairs qui fait intervenir des chercheurs portant le double chapeau de juge et partie. Ces derniers ont donc intérêt à examiner minutieusement les découvertes revendiquées par les autres membres de la communauté puisque ces découvertes seront des découvertes en moins pour eux. Cet intérêt est double donc en raison de leur statut de collègue, participant à atteindre un objectif commun (l'avancement des connaissances), mais également de concurrent dans le système compétitif de la science (Bourdieu, 2001). Par contre, la nature sociale de l'institution scientifique et son haut degré de hiérarchisation peut nuire à la mise en application de la norme de scepticisme organisé. En effet, la position hiérarchique occupée par

un chercheur peut avoir un effet sur son autorité apparente. Bien que la vérité scientifique soit au-dessus de toute hiérarchie sociale, il est aisé d'imaginer qu'un jeune chercheur puisse hésiter à remettre en cause l'autorité d'un chercheur renommé en raison même de sa position dans la hiérarchie académique, d'autant plus si son travail remet en cause la science dite « normale » à l'intérieur d'un paradigme donné (Kuhn, 1962; 2012).

1.1.2 Le système de la reconnaissance scientifique

C'est à la fin des années 1950 que Merton formalise l'idée du système de la reconnaissance scientifique (Merton, 1957; 1973). Au cœur de ce système, la notion de reconnaissance constitue le moteur des actions des chercheurs. Selon Merton, les normes scientifiques tendent toutes vers un même objectif : l'avancement des connaissances. Dans ce contexte, le rôle du chercheur est donc de participer à cet objectif par l'originalité de ses travaux, puisque c'est bien par l'originalité, par de petits ou de plus grands incréments, que les connaissances s'accumulent et avancent. Ainsi, la reconnaissance et l'estime des pairs vont aux scientifiques qui participent à atteindre l'objectif de l'institution scientifique par leurs contributions au stock des connaissances. Le système de la reconnaissance scientifique est donc un système dont l'institution scientifique s'est dotée en vue de récompenser ceux qui se conforment à ses normes et y remplissent leur rôle.

Le système décrit par Merton a évolué au cours de l'histoire et s'adapte aujourd'hui encore aux réalités changeantes de la communauté scientifique. Les nombreuses transformations de l'entreprise scientifique de l'après-guerre entraînent notamment une augmentation de la collaboration et donc du nombre d'auteurs signant les publications scientifiques (Biagioli, 2003; Cronin, 2001). Dès la seconde moitié du XX^e siècle, en réaction à l'augmentation croissante du taux de co-autorat, de nouvelles pratiques d'ordonnement des noms dans les listes d'auteurs sont mises en place afin d'augmenter la visibilité des principaux responsables d'une publication, qui récoltent ainsi une plus grande part de reconnaissance (Zuckerman, 1968).

Selon Merton (1973), dans le monde scientifique, les droits de propriété pour une idée ou une découverte se limitent à la reconnaissance accordée par les pairs à la contribution spécifique et distinctive d'un chercheur. De nombreuses formes de reconnaissance publique

permettent toutefois de récompenser les chercheurs tels que les prix et les médailles, les postes dans des institutions prestigieuses, les citations comme indicateur de l'impact de la recherche.

Pour la majorité des chercheurs, publier les résultats de travaux de recherche devient l'équivalent symbolique de faire une découverte significative. Ainsi, la signature de publications savantes est intimement associée à la reconnaissance puisque c'est d'abord l'attribution du statut d'auteur qui permet de réclamer le crédit pour une idée ou une découverte originale. L'importance accordée à la priorité dans le système de la reconnaissance savante, et donc à la publication de résultats originaux, est la conséquence des normes institutionnelles de la science. En effet, selon Merton (1973), les normes d'universalisme, de communalisme, de désintéressement et de scepticisme organisé exercent sur les chercheurs une pression qui les pousse à affirmer leur propriété intellectuelle puisque c'est par la démonstration de la priorité que les chercheurs peuvent acquérir la reconnaissance symbolique pour leur contribution à l'avancement des connaissances.

Bien que les travaux de Merton soient aujourd'hui considérés comme incontournables en sociologie des sciences, sa vision de l'institution scientifique et des normes qui la gouvernent a fait l'objet de nombreuses critiques (Saint-Martin, 2013), notamment de la part de Bourdieu.1.1.3 Le champ scientifique

En 1975, Bourdieu propose une interprétation distincte du système de la reconnaissance scientifique, qu'il définit à l'aide des concepts de champ, d'*illusio* et de capital symbolique, concepts qui sont utiles de manière plus générale pour comprendre l'ensemble de la société. Alors que Merton propose une vision quelque peu idéalisée de la communauté scientifique où les chercheurs seraient tous unis par un objectif commun, l'avancement des connaissances, Bourdieu considère plutôt le champ scientifique comme un espace de luttes pour l'autorité scientifique.

Le champ

Pour Bourdieu, un champ est un espace social ayant ses propres normes, ses propres façons de faire — son *habitus* — et ses propres barrières à l'entrée. L'*habitus* est un système de schèmes de conduites et de dispositions propres à un champ. L'*habitus* scientifique constitue le

fondement des pratiques et des enjeux du champ scientifique. Les agents, les chercheurs dans le champ scientifique, intériorisent et incorporent ces dispositions, si bien qu'elles deviennent en grande partie inconscientes. L'*habitus* peut alors être décrit comme l'intériorisation des structures sociales par les agents (Bourdieu, 1984).

Chaque champ possède des enjeux, des objets et des intérêts qui lui sont spécifiques. Le champ scientifique est un champ social comme un autre en ce sens qu'il est un espace de luttes de pouvoir, avec ses rapports de forces, ses intérêts et ses profits (son capital symbolique) qui revêtent tous une forme spécifique. L'autorité scientifique, cette « espèce de capital symbolique » (Bourdieu, 2001) qui peut être accumulé, transmis et reconverti en d'autres espèces de capital (économique ou social, par exemple), se situe au cœur des luttes dans le champ scientifique. Selon Bourdieu, l'autorité scientifique est un enjeu fondamentalement double en ce qu'il se définit à la fois comme une capacité technique (la possession de connaissances scientifiques) et comme un pouvoir social. L'autorité scientifique est donc la compétence de parler et d'agir de façon conforme et adaptée aux pratiques du champ (Bourdieu, 1975).

L'*illusio*

L'*illusio* est ce qui constitue le champ comme espace du jeu. Pour entrer dans cet espace, il faut d'abord croire dans le jeu et dans ses règles. L'*illusio*, comme droit d'entrée dans un champ, implique alors la « croyance non seulement dans les enjeux, mais aussi dans le jeu lui-même, c'est-à-dire dans le fait que le jeu en vaut la chandelle, vaut la peine d'être joué » (Bourdieu, 2001, p. 102). L'intérêt premier dans un champ n'est donc pas la quête de capital économique, mais plutôt le fait de s'investir dans le jeu d'un champ. L'*illusio* est donc ce qui donne « sens et direction » (Chevallier et Chauviré, 2010, p. 83) à l'existence d'un agent dans un champ. Dans le champ scientifique, c'est en adoptant des pratiques conformes aux règles du jeu, c'est-à-dire aux normes définies par la communauté, qu'un chercheur est en mesure d'acquérir et d'accumuler du capital ayant une valeur spécifique au champ, du capital symbolique.

Le capital

Le capital est l'enjeu central de toutes les luttes dans un champ. Les différentes espèces de capital constituent alors les enjeux spécifiques des différents champs sociaux. De manière

générale, Bourdieu définit le capital comme toute espèce de travail accumulé qui a le potentiel de produire des profits qui pourront à leur tour se reproduire (Bourdieu, 1986). La structure de distribution des différents types de capitaux, à un moment donné, représente la structure du champ, c'est-à-dire l'ensemble des contraintes qui gouvernent le fonctionnement social, les pratiques et les stratégies d'action des agents dans ce champ. Le capital se présente sous trois espèces différentes : le capital économique, qui peut être converti de façon directe et immédiate en argent; le capital culturel, qui peut être institutionnalisé sous forme de diplômes ou qualifications scolaires; et le capital social qui est quant à lui constitué des relations sociales d'un agent (Bourdieu, 1986). Le capital symbolique ne serait pas une espèce particulière de capital, mais plutôt l'espèce sous laquelle le capital se déploie dans un champ spécifique. Bourdieu (2003) dira que :

[t]oute espèce de capital (économique, culturel, social) tend (à des degrés différents) à fonctionner comme capital symbolique (en sorte qu'il vaudrait peut-être mieux parler, en toute rigueur, *d'effets symboliques du capital*) lorsqu'il obtient une reconnaissance explicite ou pratique, celle d'un *habitus* structuré selon les mêmes structures que l'espace où il s'est engendré. (p. 347)

Le capital scientifique est donc, selon Bourdieu, une espèce de capital symbolique propre au champ scientifique fondé sur la connaissance et la reconnaissance des pairs d'un chercheur, pairs qui sont également ses concurrents dans le contexte de compétition pour l'autorité scientifique dans le champ scientifique. En effet, étant donné le degré de codification élevé du champ scientifique, seuls les agents d'un même champ — c'est-à-dire les autres producteurs de connaissances agissant selon les mêmes règles du jeu — ont la capacité d'évaluer les contributions scientifiques d'autres chercheurs. Ce degré de codification contribue ainsi à la grande autonomie du champ scientifique. Les pairs, qui sont à la fois juges et parties, possèdent donc les dispositions et les catégories de perceptions adéquates et spécifiques au champ scientifique leur permettant de faire les distinctions pertinentes pour évaluer (et valider ou invalider) les travaux et les idées scientifiques d'un concurrent (Bourdieu, 2001).

Dans *Homo academicus* (1984), Bourdieu mentionne déjà les citations comme outils permettant de mesurer le capital scientifique, mais c'est véritablement Cronin qui ancrera le

concept de reconnaissance scientifique dans l'empirisme. Dans son livre *The Hand of Science* (2005), Cronin réunit ses principaux travaux bibliométriques abordant les pratiques de la communication savante et les formes de reconnaissance qui y sont rattachées.

1.1.4 Le triangle de la reconnaissance

En 1993, Cronin propose le concept du triangle de la reconnaissance (*reward triangle* [Cronin et Weaver-Wozniak, 1993; Cronin, 1995]) composé de l'autorat, des citations et des remerciements. Le triangle émerge des trois statuts qu'un individu peut avoir dans une publication scientifique : l'auteur, l'individu cité et l'individu remercié. Cronin postule que l'autorat et les citations ne dévoilent qu'une portion incomplète de la contribution d'un individu au sein du système scientifique (Cronin, 1995). En effet, les remerciements constitueraient selon lui un vecteur supplémentaire à considérer dans l'évaluation des activités d'un chercheur et une forme non négligeable d'attribution de crédit scientifique. Lorsqu'évaluée à l'aide de méthodes bibliométriques, la performance des chercheurs est typiquement mesurée à l'aide de deux indicateurs : la production et l'impact, respectivement calculés à l'aide du nombre de publications signées par un chercheur et du nombre de citations récoltées par ces publications. La pierre angulaire du système de reconnaissance scientifique est formée de l'intersection entre l'autorat et l'influence intellectuelle, c'est-à-dire l'intersection de la production savante et de l'impact d'un chercheur (Cronin, 1995). Cronin et Weaver-Wozniak (1993) soulignent toutefois l'ironie du fait qu'un auteur peut récolter la reconnaissance associée à une citation, peu importe la nature de la citation — que celle-ci soit infirmative ou superficielle — alors que le même individu qui reçoit un remerciement pour une contribution conceptuelle ou méthodologique à une publication ne récoltera en retour aucun capital symbolique, les remerciements n'étant pas considérés comme une forme de reconnaissance formelle au sein de la communauté scientifique.

Ainsi, selon Cronin, à la base du triangle de la reconnaissance, aux côtés de l'autorat et des citations, les remerciements offrent une fenêtre unique sur certaines contributions qui n'atteignent pas le niveau de l'autorat (*sub-authorship collaboration*, [Patel, 1973, p. 81]). Les remerciements peuvent donc constituer un important indicateur de la collaboration scientifique informelle permettant d'identifier les réseaux d'influences au sein des communautés scientifiques (Cronin, 2001) et d'obtenir une image plus complète du nombre d'individus —

auteurs et remerciés — impliqués dans la production de connaissances scientifiques (Heffner, 1981). Avec plus de vingt-cinq publications consacrées aux remerciements entre 1991 et 2014, les travaux de Blaise Cronin ont sans contredit permis de justifier la place des remerciements dans la discussion sur les formes de reconnaissances associées à la communication savante.

1.2 L'autorat scientifique

Tel qu'exposé par Merton et Bourdieu, dans le monde scientifique, le crédit associé à une idée ou une découverte n'équivaut pas à une compensation monétaire ni à la propriété intellectuelle au sens juridique (Biagioli, 1998), mais s'inscrit plutôt dans un système de la reconnaissance où le crédit pour un travail est symbolique et s'appuie principalement sur les publications signées par un chercheur, c'est-à-dire l'octroi du statut d'auteur. En ce sens, l'autorat constitue le principal mode d'attribution du crédit scientifique, le « *undisputed coin* » du système de la reconnaissance scientifique (Cronin, 2001, p. 559). La prochaine section présente plus spécifiquement les fonctions de l'autorat dans le système de la reconnaissance scientifique ainsi que l'évolution des pratiques d'autorat au cours des derniers siècles.

1.2.1 L'évolution des pratiques d'autorat

Bien qu'elle soit aujourd'hui indissociable de la publication d'un travail savant, la signature de l'auteur n'a pas toujours été considérée comme un élément essentiel de la communication scientifique. Au Moyen Âge, c'était au noble, plutôt qu'à celui qui avait écrit le texte, d'apposer sa signature sur une publication qu'il avait commanditée afin d'attester de la véracité des faits rapportés et de légitimer la publication. De nos jours, l'affiliation institutionnelle remplace le rang social dans le rôle d'attestation de l'honnêteté intellectuelle d'un chercheur (Shapin, 1989). Les fonctions et les pratiques de l'autorat ont ainsi évolué au fil de l'histoire scientifique, de l'alchimie des premières heures pratiquée par des amateurs à l'institutionnalisation et la professionnalisation de la science au XVII^e siècle et la croissance de la collaboration (Merton, 1973) jusqu'à l'émergence de l'hyperautorat après la Seconde Guerre mondiale (Cronin, 2001).

Jusqu'au XVII^e siècle, le mode dominant de communication savante est le courrier. À cette époque, de vastes réseaux épistolaires permettent la diffusion des connaissances

scientifiques à une communauté restreinte. Les lettres sont souvent rédigées sous forme codée, en plus d'être datées et signées afin de garantir la priorité des découvertes. À la fin du XVII^e siècle, les réunions de la *Royal Society of London* constituent le premier lieu de diffusion des résultats d'expérimentations scientifiques. C'est précisément de ces réunions que naîtra l'une des premières revues scientifiques³ en mars 1665, les *Philosophical Transactions of the Royal Society of London* (Pontille, 2004). La notion d'auteur scientifique apparaît donc dans ce contexte de diffusion publique des connaissances scientifiques au sein des premières sociétés et revues savantes. Ainsi, l'auteur scientifique signe ses publications afin de revendiquer l'originalité et la priorité de ses travaux. L'autorat permet en retour au chercheur de recevoir une forme de rétribution symbolique, la reconnaissance de ses pairs. En revanche, l'autorat scientifique n'est traditionnellement pas lié à une rétribution monétaire. L'autorat scientifique précède en effet l'émergence de la notion de droits d'auteurs qui, élaborée plus tard au XVIII^e siècle, constituera le fondement de la définition de l'auteur littéraire (Chartier, 2003; Pontille, 2004).

Le modèle traditionnel de l'autorat – où un seul auteur reçoit l'ensemble du crédit pour le texte qu'il signe – sera dominant en science pendant plusieurs siècles (Greene, 2007; Price, 1963; Wuchty, Jones et Uzzi, 2007). Le processus de professionnalisation de la science au XIX^e siècle s'accompagne de nouvelles pratiques en matière de signature scientifique : l'autorat multiple (co-autorat). Ce n'est toutefois qu'au XX^e siècle que le co-autorat devient une pratique fréquente (Beaver et Rosen, 1978, 1979a). En effet, après la Seconde Guerre mondiale, la science moderne connaît un développement exponentiel; on assiste à une transition de la *little science* à la *big science*⁴ (Price, 1963). Les gouvernements se mettent alors à financer massivement les projets scientifiques à grande échelle (Cronin, 2005). Dans plusieurs domaines,

³ Les *Philosophical Transactions* sont souvent considérées comme la première revue savante bien qu'en janvier 1665, trois mois avant la parution du premier numéro de la publication londonienne, le *Journal des Scavans* de l'Académie royale des sciences est lancé. Cette revue se consacre toutefois à la publication de comptes rendus de livres et moins à la publication d'expérimentations scientifiques (Pontille, 2004).

⁴ Terme utilisé en histoire des sciences pour désigner les changements scientifiques de l'après-guerre, caractérisé par des projets scientifiques à très grande échelle.

le travail scientifique devient synonyme de collaboration (Galison, 2003). La spécialisation et la complexification technologique de nombreux objets de recherche nécessitent désormais le recours à des équipes formées d'un grand nombre de chercheurs d'origines disciplinaires diverses. Cette interdépendance structurelle a pour résultat une réorganisation du travail scientifique qui aura un impact important sur la notion et les pratiques d'autorat. L'autorat scientifique passe alors d'un modèle traditionnel, où l'auteur unique est la norme, à un fractionnement de l'autorat dans le modèle actuel où le co-autorat domine dans la majorité des disciplines (Beaver et Rosen, 1979b; Wuchty, Jones et Uzzi, 2007).

L'importante augmentation du nombre moyen d'auteurs par article depuis les années 1950 témoigne, en partie, de cette transformation de l'organisation scientifique en entreprise majoritairement collaborative (Cronin, 2001). On parle aujourd'hui de « l'impératif de collaboration » (Bozeman et Boardman, 2014) qui fait désormais office de norme dans un nombre croissant de disciplines scientifiques. Un facteur supplémentaire ne peut cependant être écarté lorsque vient le temps d'expliquer l'augmentation du nombre d'auteurs par article : un assouplissement des critères d'autorat. En effet, bien que la complexification et la spécialisation du travail scientifique ne soient plus à prouver, il demeure aujourd'hui difficile de départager la part de l'augmentation du nombre d'auteurs par publication qui est due à une augmentation réelle du nombre de collaborations de celle qui serait due aux transformations des pratiques d'autorat et d'un assouplissement des critères en matière de signature scientifique. Les phénomènes d'auteur honorifique ou invité (nommer un individu comme auteur alors qu'il n'a pas contribué à la recherche) et d'auteur fantôme (ne pas nommer un individu comme auteur alors qu'il a contribué à la recherche) constituent des exemples concrets de cet assouplissement des critères d'autorat qui peuvent être considérés comme des pratiques d'autorat problématiques ou peu éthiques (Flanagin et al., 1998; Sismondo, 2009; Teixeira da Silva et Dobránszki, 2015; Wislar et al., 2011).

À la fin des années 1990, la notion d'autorat fait l'objet d'un important débat dans le domaine biomédical. La forme traditionnelle de l'auteur unique apparaît comme étant de plus en plus désuète. En effet, le modèle traditionnel s'avère inadapté à mesure que le nombre moyen d'auteurs par article augmente, puisque le crédit et la responsabilité associés à l'autorat s'en

trouvent à la fois obscurcis et dilués. Un modèle plus conforme à la nouvelle réalité devrait plutôt prendre en compte les divers types de contributions qui forment l'entreprise scientifique moderne, où la collaboration est plus souvent la norme que l'exception. Dans ce contexte, Rennie, Yank et Emanuel (1997) proposent un changement conceptuel radical : remplacer la notion d'auteur par celle de contributeur. Le modèle des contributeurs suppose de décrire de façon explicite la contribution de tous les individus ayant pris part à un projet de recherche permettant ainsi d'éliminer la distinction artificielle qui sépare les auteurs des contributeurs non-auteurs qui sont mentionnés dans les remerciements. Selon Rennie, Yank et Emmanuel (1997), l'abandon du concept d'auteur pour celui de contributeur reflète une conception beaucoup plus actuelle et représentative de la science moderne. Sans parler de contributeur, Resnik (1997) propose au même moment un nouveau système d'attribution du crédit scientifique afin de mieux définir la contribution et les responsabilités de chacun des membres d'une équipe de recherche. À la manière des crédits qui défilent dans un générique de film, la proposition de Resnik suggère l'ajout de la description détaillée du rôle de chaque individu impliqué dans la production d'une publication scientifique.

Dans les années suivantes, plusieurs revues importantes du domaine biomédical (p. ex. JAMA, Lancet, Annals of Internal Medicine, British Medical Journal, American Journal of Public Health, et plus récemment PLOS One) ont adopté le concept du contributeur. Ce modèle n'a toutefois pas réussi à supplanter le modèle traditionnel. On le comprend à la lumière des fonctions de l'atorat et de la reconnaissance qui lui est accordée dans le champ scientifique. Les notions d'auteur, de contributeur ainsi que de remercié se côtoient donc aujourd'hui dans le système de la communication savante.

1.2.2 Les fonctions de l'atorat

En science, l'authentification des connaissances passe par la validation collective des faits publiés. Pour ce faire, les connaissances doivent être revendiquées par leurs créateurs; le lien d'atorat permet donc d'unir le créateur à son texte afin d'en faire une source de connaissances certifiée par les pairs (Pontille, 2004). Au-delà de l'authentification, l'atorat remplirait trois principales fonctions dans le champ scientifique : l'attribution du crédit, de la responsabilité et du capital.

Premièrement, l'autorat permet l'attribution du crédit pour une découverte ou une idée. L'autorat est généralement accepté comme le principal mode de reconnaissance des contributions d'un chercheur au stock des connaissances (Biagioli, 2003; Birnholtz, 2006; Bourdieu, 2001; Merton, 1973). Avoir son nom dans la liste des auteurs d'un article scientifique permet de signaler une contribution significative à une découverte originale ou une idée. Par la signature d'une publication, un chercheur revendique l'originalité et la priorité d'une découverte et peut en retour obtenir la reconnaissance de ses pairs pour cette découverte (Merton, 1973).

Deuxièmement, l'autorat permet l'attribution de la propriété d'une découverte ou d'une idée. La propriété n'est pas entendue ici dans le sens d'un droit d'auteur associé à des redevances financières, mais fait plutôt référence à la notion de responsabilité liée à une découverte ou une idée, incluant la responsabilité des erreurs et des controverses (Birnholtz, 2006). Une spécificité de la question de l'attribution du crédit en science est que l'octroi du crédit est inséparable de l'attribution de la responsabilité. Dès lors, lorsqu'un chercheur reçoit le crédit pour une découverte ou une idée, il doit en retour en garantir la responsabilité (Biagioli, 1998).

Troisièmement, l'autorat permet l'existence d'une économie de la réputation. Dans les descriptions du système de la reconnaissance savante de Merton (1973) et de Bourdieu (2001), les récompenses prennent la forme d'actes de reconnaissance symbolique publics tels que les prix et les distinctions honorifiques, les citations, ainsi que les remerciements, dans une moindre mesure (Cronin, 1995). Ces actes de reconnaissance s'appuient toutefois sur les contributions qu'un chercheur a faites à l'avancement des connaissances et passe donc généralement par la reconnaissance de l'originalité et de la priorité des travaux qu'un chercheur signe et réclame comme étant siens par l'autorat. L'acquisition et l'accumulation de capital dans le champ scientifique se font donc par la signature de publications savantes. Ainsi, l'autorité scientifique d'un chercheur n'est pas établie par la possession de capital économique, mais plutôt de capital symbolique. Le capital symbolique est le fruit de la reconnaissance par des tiers de la légitimité de la position de celui qui en est le possesseur (Bourdieu, 2001). L'autorat est donc le principal mode d'attribution du capital symbolique dans le système de la reconnaissance savante. Pour Bourdieu, accumuler du capital scientifique — cette espèce de capital symbolique propre au champ scientifique — c'est se faire un nom, être visible et être reconnu pour son autorité

scientifique. Plusieurs stratégies visent à maximiser cette visibilité, notamment les pratiques d'ordonnement des noms d'auteurs signant les articles scientifiques qui permettent de « minimiser la perte de valeur distinctive » (Bourdieu, 1975, p. 99), conséquence de la division et de la spécialisation croissantes du travail scientifique depuis les années 1950 (Cronin, 2001; Galison, 2003; Zuckerman, 1968). Ces pratiques d'ordonnement varient d'une discipline à l'autre, reflétant les pratiques disciplinaires en matière de division du travail et de distribution du crédit scientifique. De la même façon, la mention dans les remerciements permet de créditer la contribution d'un individu sans pour autant partager avec ce dernier le capital symbolique associé à l'autorat d'une publication savante.

L'évaluation d'un chercheur, que ce soit dans le cadre de l'embauche, d'une promotion ou d'un concours de financement, s'appuie généralement sur les publications signées en tant qu'auteur, sur le financement obtenu, ainsi que sur les distinctions remportées qui permettent de mesurer la valeur de ses travaux, c'est-à-dire ses apports en connaissances et la reconnaissance obtenue (Bourdieu, 2001; Pontille, 2004). L'acquisition de capital symbolique permet ainsi à un chercheur d'obtenir d'autres espèces de capitaux (économique, social et culturel) par la conversion de cette rétribution symbolique, qui elle s'appuie sur les travaux où le chercheur est reconnu comme auteur.

Il est toutefois important de souligner la nature circulaire des processus d'attribution et de conversion du capital dans le champ scientifique où le capital symbolique permet de générer du capital économique (financement) nécessaire à la production de nouveaux biens scientifiques menant en retour à plus de capital symbolique. Bien que l'autorat occupe une place centrale dans le processus d'attribution du capital symbolique, la capacité de produire et de signer des publications originales s'avère tout de même étroitement liée au capital économique dont dispose un chercheur. En effet, un chercheur doit être en mesure de mobiliser des ressources financières et matérielles pour produire des biens symboliques (publications savantes signées en tant qu'auteur) lui assurant en retour la reconnaissance de ses pairs et lui permettant d'amasser plus de capital symbolique, qui pourra à son tour être converti en d'autres espèces de capital (et en ressources). Au final, l'autorat constitue la principale monnaie d'échange

symbolique dans le champ scientifique, permettant en retour d'accéder aux ressources nécessaires à la production scientifique.

1.3 Les facteurs qui influencent les pratiques d'attribution du crédit

Les fonctions de l'autorat au sein du champ scientifique et leur importance dans le système de la reconnaissance scientifique expliquent que les critères d'attribution du crédit scientifique ne reposent pas uniquement sur la nature de la contribution d'un individu : d'autres facteurs entrent en jeu, notamment le contexte disciplinaire ainsi que des facteurs sociodémographiques, tels que le genre, l'âge et le statut académique, qui sont liés à la position hiérarchique d'un individu au sein de son équipe de recherche et plus largement dans le champ scientifique.

En effet, la structure du champ scientifique entraîne ce que Merton (1968; 1973) désigne sous le nom d'effet Saint-Matthieu (*Matthew effect*). L'effet Saint-Matthieu consiste à attribuer une part de reconnaissance disproportionnée à la contribution de chercheurs renommés et une part démesurément faible aux chercheurs moins renommés. Ainsi, dans un effet circulaire, les chercheurs éminents reçoivent plus de crédit pour leur contribution alors que les chercheurs relativement inconnus tendent à recevoir trop peu de crédit pour leur contribution. L'effet Saint-Matthieu démontre que la communauté scientifique est un espace de compétition imparfait, où les chances de chacun ne sont pas égales et qui a pour conséquence le renforcement des positions dominantes au détriment des positions dominées (Bourdieu, 2001). Dans cette structure hiérarchisée du champ scientifique, l'autorité est distribuée en fonction de la production scientifique des chercheurs. Le dominant est celui qui occupe une place avantageuse dans la structure, en raison de son capital, position qui lui permet en retour d'accumuler plus de capital et d'ainsi maintenir la structure de distribution des forces (Bourdieu, 2001).

Les travaux de Merton (1973), Cole et Cole (1973) et Zuckerman (1977) sur les effets de la stratification sociale en science ont montré que de nombreuses variables sociodémographiques déterminent et influencent la position hiérarchique d'un chercheur. Le contexte disciplinaire, ainsi que l'âge, le statut académique (étudiant, chercheur établi, technicien), le genre d'un chercheur et le contexte national et culturel sont tous des facteurs à prendre en compte pour comprendre les pratiques d'attribution du crédit dans le champ

scientifique. En effet, ce sont les dominants qui déterminent l'attribution du crédit dans une équipe de recherche : qui signera une publication en tant qu'auteur et qui sera simplement remercié pour sa contribution. Dans le monde scientifique, « la capacité de signer semble dépendre en grande partie du statut social des personnes, elle est intimement liée à l'autorité institutionnelle » (Shapin 1989, traduction de Pontille, 2004, p. 130). Bozeman et Youtie (2016) ont étudié, à l'aide de données qualitatives, les facteurs pouvant expliquer l'inclusion et l'exclusion qualifiée comme « injustifiée » d'individus comme co-auteurs. Les facteurs invoqués incluent notamment les caractéristiques de la discipline, les dynamiques de pouvoir au sein des équipes de recherche (qui incluent les variables de la séniorité et du genre des individus) et la dispersion géographique des collaborateurs.

Dans une récente enquête effectuée auprès de 6 673 chercheurs internationaux, Smith et al. (2019) ont montré que la discipline, le rang académique, ainsi que le genre des chercheurs avaient un effet significatif sur les désaccords entourant l'attribution du statut d'auteur dans une équipe de recherche. De plus, de façon plutôt paradoxale, les chercheurs collaborant au sein d'équipes multidisciplinaires ont rapporté avoir vécu moins de désaccord dans les décisions entourant l'autorat. Les auteurs avancent plusieurs hypothèses pouvant expliquer cette situation, notamment une communication accrue dans les équipes multidisciplinaires et une plus grande valorisation de la diversité.

1.3.1 Discipline

Dans le contexte de la recherche scientifique, écrire un texte ne signifie pas *de facto* le signer et signer ne nécessite pas toujours qu'il y ait eu un acte d'écriture. Les critères d'attribution du crédit scientifique et les problèmes provoqués par les très longues listes de co-auteurs sont marqués par d'importantes différences disciplinaires. Les pratiques d'autorat se déploient selon un spectre ayant à l'un de ses extrêmes les sciences naturelles empiriques et à l'autre les sciences humaines théoriques. Dans la plupart des disciplines empiriques, les articles signés par un seul auteur sont en voie de disparition. La production de nouvelles connaissances semble y nécessiter la coopération de plusieurs personnes. Dans ces disciplines, la spécialisation et la division technique et cognitive du travail sont telles que chaque membre d'une équipe est totalement dépendant des autres pour en arriver à une production complète.

Par ailleurs, il semble exister un fossé dans plusieurs disciplines empiriques entre « ceux qui pensent » et « ceux qui manipulent » (Shapin, 1989). Ainsi, malgré la spécialisation et la professionnalisation des activités scientifiques, certaines tâches techniques (manipulations et collectes de données, notamment) ne sont pas jugées suffisantes pour permettre à un individu d'accéder au statut d'auteur et seront, dans certaines disciplines, automatiquement rangées du côté des remerciements (Pontille, 2004).

Deux domaines de recherche se distinguent en matière de co-autorat : la physique des particules et les sciences biomédicales. En 2001, Cronin introduit le néologisme « hyperautorat » (*hyperauthorship*) pour désigner ces collaborations qui réunissent plusieurs centaines d'auteurs sur une même publication. Dans le domaine biomédical, les essais cliniques qui font intervenir de nombreux centres et institutions de recherche représentent l'exemple par excellence d'hyperautorat. Lorsque des centaines, voire des milliers de collaborateurs signent une publication, il devient quasi impossible d'identifier qui a fait quoi et surtout qui est responsable de quels aspects. Bien que les institutions de recherche, les organismes de financement et de nombreux comités éditoriaux aient mis en place des définitions et des critères d'autorat visant à baliser l'attribution du statut d'auteur, ces définitions ne sont pas uniformes, sont difficiles à appliquer et demeurent souvent méconnues (Bosch et al., 2012; Smith et Williams-Jones, 2012; Teixeira da Silva et Dobránszki, 2015). Dans une récente étude portant sur les définitions de l'autorat en sciences sociales, Chang (2018) a montré que plus de la moitié des 1 065 revues analysées n'avaient pas de définition établie de l'autorat et que seulement 3,8 % de ces revues avaient intégré une définition de l'autorat dans leur section d'instruction pour les auteurs.

Les lignes directrices les plus citées en matière d'autorat sont celles proposées par l'International Committee of Medical Journal Editors (ICMJE). Les critères de l'ICMJE distinguent de façon explicite les contributions intellectuelles des contributions techniques ou manuelles :

- 1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
- 2) Drafting the work or revising it critically for important intellectual content; AND

3) Final approval of the version to be published; AND

4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. (ICMJE, 2018)

Selon Biagioli (2003), plutôt que de redéfinir la notion même de l'auteur scientifique à la lumière des conditions actuelles de l'entreprise scientifique, c'est-à-dire les collaborations fréquentes et à grande échelle, l'ICMJE, par sa définition, renforce la notion de l'auteur individuel. Selon l'ICMJE, ce sont les contributions intellectuelles d'un individu qui peuvent le qualifier pour l'atorat, les autres formes de contributions techniques et non-intellectuelles devant être reléguées aux remerciements.

Bien que considéré problématique dans le domaine biomédical, la pratique de l'hyperatorat est tout à fait acceptable en physique des particules, domaine dans lequel la signature scientifique est très clairement encadrée. La collaboration du Collider Detector au Fermilab (CDF) en constitue un bon exemple. Le CDF a élaboré ses propres règles en matière d'attribution du crédit et de signature scientifique : toutes les publications produites par la collaboration doivent inclure tous les noms apparaissant sur la Liste standard d'auteurs. La Liste standard inclut le nom de tous les membres actifs de la collaboration, en ordre alphabétique. Pour être considéré comme membre actif, un chercheur doit consacrer au moins 50 % de son temps de recherche aux expérimentations du CDF. Tous les membres actifs sont ajoutés à la Liste standard d'auteurs après un an de travail au sein de la collaboration et le nom d'un chercheur est retiré seulement un an après son départ. La logique qui sous-tend de telles pratiques repose sur la présupposition qu'aucune recherche n'aurait pu être effectuée sans la contribution de chacun des individus. Cette reconnaissance du rôle de chacun vise également à encourager l'engagement et la contribution de l'ensemble des membres de la collaboration.

Les exigences bureaucratiques du système de signature scientifique en physique des particules mettent en lumière le fossé qui sépare la conception de l'atorat scientifique en sciences biomédicales, tel qu'illustré par les critères de l'ICMJE, et la conception de l'atorat spécifique à la physique des particules. En ce qui concerne l'attribution de la responsabilité, la collaboration du CDF possède un protocole détaillé de révision interne avant même la soumission d'un article à la communauté externe. Après trois rondes de révision interne, les

chercheurs qui ne sont pas satisfaits du produit fini peuvent retirer leur nom de la liste d'auteurs apparaissant sur l'article. Ainsi, le crédit et la responsabilité sont gérés de façon indépendante en physique des particules alors que ces deux aspects sont intrinsèquement indissociables dans le domaine biomédical (Biagioli, 2003).

À l'autre extrémité du spectre disciplinaire, les pratiques d'autorat dans les humanités semblent suivre des règles complètement différentes puisque l'auteur unique y est encore aujourd'hui chose commune (Cronin, 2001; Larivière, Gingras et Archambault, 2006; Ossenblok, Verleysen et Engels, 2014; Pontille, 2004). Cet état de fait peut aisément s'expliquer par les objets d'étude et les méthodes employées dans les disciplines du domaine des humanités qui n'impliquent habituellement que peu de division du travail et d'investissement de ressources (financières, humaines et technologiques) comparables à celles qui doivent aujourd'hui être déployées dans les sciences naturelles et biomédicales, par exemple.

Entre ces deux extrêmes, les pratiques d'autorat se déclinent de multiples façons, selon les traditions d'une discipline ou une spécialité donnée ou l'organisation du travail nécessaire à la production de nouvelles connaissances. Ainsi, les pratiques d'autorat dans les disciplines des sciences sociales se révèlent très variées. Par exemple, le nombre moyen d'auteurs par article dans les disciplines de l'économie, de la gestion et de la finance est aujourd'hui d'environ 2,5, alors qu'en droit et en science politique, un seul auteur signe encore aujourd'hui la majorité des articles scientifiques publiés (Henriksen, 2016). D'autres disciplines présentent des portraits d'autant plus variés en fonction du niveau de spécialité analysé. Par exemple, la psychologie, inclut à la fois des spécialités comme la psychologie clinique où le travail scientifique nécessite une plus grande collaboration et où le nombre moyen d'auteurs par article est élevé, mais également la psychoanalyse, où le chercheur travaille et signe majoritairement seul ses publications (Henriksen, 2016). Une même tendance caractérise toutefois l'ensemble des sciences sociales avec un nombre moyen d'auteurs par article à la hausse au cours des dernières décennies (Henriksen, 2016; Ossenblok, Verleysen et Engels, 2014).

1.3.2 L'âge et le statut académiques

L'âge académique a un effet sur la capacité d'un individu à revendiquer le statut d'auteur ainsi que sa position dans la liste d'auteurs. Ainsi, à contribution égale, le chercheur en début de

carrière n'a généralement pas la même autorité au sein d'une collaboration qu'un chercheur établi. Cette position hiérarchique ne lui offre donc souvent pas l'autorité pour décider de sa place dans la liste d'auteurs. Paradoxalement, en raison même de son âge académique et du fait que sa renommée reste à bâtir, le jeune chercheur accordera généralement plus d'importance à l'autorat et à sa position dans la liste d'auteur qu'un chercheur déjà renommé (Merton, 1973). L'âge académique peut également être lié au statut de l'individu : le poids hiérarchique d'un étudiant n'est pas le même que celui d'un professeur émérite, ni même que d'un technicien. Une stratification très fine, reflet du haut degré de codification du champ scientifique (Bourdieu 2001), se découpe ainsi entre les assistants de recherche, les techniciens, les étudiants de premier cycle, les étudiants des cycles supérieurs, les postdoctorants, les professeurs adjoints, agrégés, titulaires et émérites.

Shapin (1989) a montré que le rôle essentiel des techniciens dans le développement de la science moderne au XVII^e siècle est passé totalement sous silence dans l'histoire des sciences, les contributions de ces derniers ayant été à l'époque complètement oblitérées des publications scientifiques. Encore aujourd'hui, dans de nombreux contextes disciplinaires (tout dépendant également des cultures organisationnelles spécifiques), une distinction rigide demeure entre les agents performant des contributions intellectuelles et ceux assignés aux contributions techniques. Dans un tel contexte, les techniciens, en raison de leur statut, ne signeront pas la publication à laquelle ils ont contribué.

1.3.3 Le genre

Le genre des chercheurs est également une variable qui intervient dans la stratification sociale de l'institution scientifique. Malgré de nombreuses avancées et une conscientisation croissante, les femmes sont encore victimes de discrimination liée au genre. Les études sont nombreuses à montrer les disparités de genre : les femmes signent en proportion moins d'articles scientifiques et récoltent moins de citations (Larivière et al., 2013), elles obtiennent moins de financement (Kaatz et al., 2016; Ley et Hamilton, 2008), elles jouent plus fréquemment un rôle de soutien dans les équipes de recherche où elles effectuent plus souvent des contributions de nature technique (Macaluso et al., 2016), elles tendent à occuper une position plus périphérique dans les réseaux de collaboration (Ghiasi, Larivière et Sugimoto, 2015), elles obtiennent en

proportion moins de prix et distinctions honorifiques (Lincoln et al., 2012) et ont en moyenne un salaire moins élevé que les hommes (Shen, 2013).

L'effet Matilda peut en partie expliquer ces nombreuses inégalités. L'effet Matilda désigne la tendance à minimiser ou même à totalement ignorer les contributions scientifiques faites par des femmes, contributions qui dans de nombreux cas seront attribuées à leurs collègues masculins (Rossiter, 1993). Que ce soit par le déni ou l'ignorance (et donc la non-citation) d'une publication signée par une femme, par la minimisation de la contribution spécifique d'une femme dans une publication cosignée avec des hommes ou par le fait qu'une femme n'accède tout simplement pas au statut d'auteur et est reléguée aux remerciements, l'effet Matilda tend ainsi à appuyer l'influence du genre dans l'attribution du crédit pour une idée ou une découverte.

1.4 Les remerciements dans le contexte de la communication savante

L'objectif de notre recherche doctorale est de décrire les fonctions des remerciements dans un contexte de communication savante, afin de mieux comprendre leur importance comme marqueur de crédit scientifique. Cette section vise donc, dans un premier temps, à faire un survol de la littérature scientifique s'étant intéressée aux remerciements, afin de mieux situer notre projet, et dans un deuxième temps à présenter plus en détails la littérature ayant abordée les remerciements du point de vue du crédit scientifique. Nous présentons d'abord les principales caractéristiques du corpus des études portant sur les remerciements dans le contexte de la communication savante : disciplines, méthodes, types de publications et principales perspectives du champ d'études. Puis, nous détaillons de façon plus approfondie la perspective des remerciements comme forme d'attribution du crédit scientifique, cette dernière étant la plus pertinente dans le contexte de notre recherche doctorale puisque notre objectif est centré sur l'angle du crédit scientifique.

1.4.1 Cinquante ans de recherche sur les remerciements⁵

Dès 1967, on voit apparaître les remerciements aux côtés de l'autorat et des citations dans une étude bibliométrique de la communication scientifique (Parker, Paisley et Garrett, 1967). Depuis, plus de 200 publications scientifiques se sont intéressées aux remerciements dans le contexte de la communication savante. Les disciplines ayant contribué à ce champ de recherche sont nombreuses et variées, allant de la psychologie aux sciences naturelles et biomédicales en passant par l'anthropologie, la linguistique et l'éthique. Toutefois, les études dédiées à l'analyse des remerciements proviennent majoritairement des sciences humaines et sociales, avec une contribution dominante des sciences de l'information (46% des études dédiées aux remerciements et indexées dans la base de données Web of Science proviennent de la catégorie *Information Science & Library Science*). L'informatique et l'étude des sciences et technologies, deux champs de recherche souvent proches des sciences de l'information, sont également très présents dans le corpus de la littérature sur les remerciements (32% des études du corpus indexées dans la base de données Web of Science proviennent de la catégorie *Computer Science* et 9% des catégories *Science & Technologie* et *History & Philosophy of Science*).

En ce qui a trait aux méthodes utilisées pour étudier les remerciements, les études empiriques basées sur les données textuelles sont largement les plus fréquentes et représentent 64% des publications du corpus. Une très forte majorité d'études ont utilisé des méthodes d'analyses quantitatives, seules ou en combinaison avec d'autres méthodes, elles sont présentes

⁵ Cette section s'appuie sur une revue systématique de la littérature réalisée en collaboration avec Nadine Desrochers et Jen Pecoskie, publiée dans *Journal of the Association for Information Science and Technology* (Desrochers, Paul-Hus et Pecoskie, 2017). Afin d'obtenir le corpus final de 229 publications scientifiques sur les remerciements, nous avons effectué des requêtes dans 10 bases de données bibliographiques. Des 1 521 notices bibliographiques repérées grâce à ces requêtes, seules 160 notices ont été retenues dans le corpus. Les raisons d'exclusion incluent notamment les doublons, l'utilisation du terme « remerciement » dans un sens homonyme non pertinent à notre étude, et les notices incomplètes ou erronées. De plus, seules les publications en français et en anglais ont été retenues dans le corpus. Aucune limite chronologique ou géographique n'a été appliquée au corpus. Finalement, 95 notices additionnelles ont été incluses au corpus final et proviennent d'autres sources et méthodes de recherche : analyse manuelle des bibliographies de publications déjà incluses au corpus, connaissance personnelle et sérendipité.

dans 71% des publications du corpus. Les analyses quantitatives se limitent dans une large proportion aux statistiques descriptives (66% des publications) — c'est-à-dire le nombre d'articles d'un corpus qui inclut des remerciements — mais incluent parfois l'utilisation d'indicateurs bibliométriques et dans une moindre mesure de statistiques inférentielles (24% des publications). Près de la moitié des études empiriques (47%) incluent aussi des analyses qualitatives où le contenu des remerciements est codé manuellement à l'aide d'un arbre de codage inductif ou prédéterminé (p. ex. typologie des remerciements empruntée à la littérature existante). Une proportion plus faible, mais toutefois non négligeable, des études empiriques ont utilisé des méthodes d'analyses linguistiques (13% des publications du corpus) où les courts textes de remerciements sont analysés sous un angle pragmatique, sémantique ou structurel.

D'autres études ont utilisé des méthodes s'appuyant non pas sur les données textuelles des remerciements, mais plutôt sur des données recueillies dans le cadre d'enquêtes et d'entrevues visant à explorer les pratiques de remerciements et l'opinion des auteurs ou remerciés à l'égard des remerciements, ces études représentent 15% du corpus. La littérature sur les remerciements compte également des guides de bonnes pratiques et lignes directrices qui visent à établir des normes pour encadrer les pratiques d'attribution du crédit et la rédaction de remerciements en contexte scientifique (p. ex. Alwi, 2007; Baggs, 2008; Glass, 1992; Graf et al., 2007; Hess et al., 2015; Swales et Feak, 2000), ces études représentent 6% du corpus. Les remerciements demeurent toutefois à ce jour un espace très peu balisé. À cet effet, de nombreuses études à travers le temps mentionnent le manque de standardisation des remerciements comme une des principales limites à l'analyse de ces courts textes :

The format of acknowledgement varies from field to field and from journal to journal. As noted, persons and institutional sources may be listed in the methods and materials section of an article or explicitly thanked in an acknowledgements section. (McCain, 1991, p. 506)

Since there are no established formats for acknowledgements in papers, as there are for citations, expressions of gratitude vary greatly and sometimes it was difficult to identify the correct type of support, and even more difficult, the correct funding organisation. (Jeschin, Lewison et Anderson, 1995, p. 238)

The first source of simple error may arise through the misspelling of the names of funding bodies and potentially the names of grants and grant codes [...] A second difficulty will be that researchers will not correctly remember the funding bodies and grants that they used to support the research. (Rigby, 2011, p. 368-369)

In using and interpreting funding acknowledgement data for large-scale bibliometric analyses, some caveats and limitations need to be recognized. Funding arrangements can be complex: more than one grant number from a single funding agency as well as multiple funding agencies can be reported as supporting the research presented in a single paper. Additionally, not all publications indexed by WoS contain funding information. [...] The non-reporting of research awards may also be due to intentional or unintentional neglect by authors. (Wang et Shapira, 2011, p. 567)

The isolation of the funding brought by collaborators seems not to be possible at the moment according to the way in which authors acknowledge funding (not always each author indicate his/her funding acknowledgement individually) and how it is reflected in bibliographic databases (even if the information is available, there is no direct link between the author and his/her individual funding acknowledgement. (Costas et Yegros-Yegros, 2013, p. 94)

La signification ambiguë des remerciements dans le contexte scientifique vient également du fait qu'ils soient perçus, dans certains cas du moins, comme des gestes intéressés, profitant davantage à leur auteur qu'aux remerciés eux-mêmes, comme en témoignent les extraits suivants :

This is why the mention of other, especially “older,” anthropologists’ names in acknowledgements may be viewed as a strategy for gaining professional attention. (Ben-Ari, 1987, p. 67)

Acknowledgements are thus pointedly self-serving gestures, and they contain genuine and appropriate expressions of real gratitude, but as all authors sense, they simultaneously prepare the reader to judge sympathetically the author’s text. (Coates, 1999, p. 255)

Acknowledgments are permeated by hyperbole, effusiveness, overstatement, and exaggeration. (Hollander, 2001, p. 64)

But although academics appear to subscribe to a governing etiquette of acknowledging practices [...], expressing thanks to others is not an entirely altruistic practice, and it is this potential for flattery and self-promotion which has attracted criticism to acknowledgements. (Hyland, 2003, p. 245)

Acknowledgements have been discussed as a form of patronage in scholarly communication, where the reality of the past may be purposefully glossed over and where the author could be looking toward the possibility of receiving future favors. (Forzetting, 2010, p. 4)

Néanmoins, trois perspectives principales se dégagent de la littérature sur les remerciements dans le contexte de la communication savante : les remerciements comme genre académique, les remerciements comme outil bibliométrique et les remerciements comme forme d’attribution de crédit (Desrochers, Paul-Hus et Pecoskie, 2017). Les études s’étant intéressées à la forme, à

la structure et au style des remerciements en tant que genre proviennent surtout du domaine linguistique (p. ex. Al-Ali, 2010; Gesuato, 2004; Giannoni, 2002; Lasaky, 2011; Loan, 2017; Yang et al., 2012; Zhao et Jiang, 2010). Cette perspective caractérise 22% des publications du corpus. La deuxième perspective, les remerciements comme outil bibliométrique, regroupe toutes les études où les remerciements sont analysés aux côtés d'autres indicateurs bibliométriques dans l'évaluation d'un domaine ou d'un objet de recherche (p. ex. Harter et Hooten, 1992; Innes, 2006; Lewison et Markusova, 2010; Murray, 1982; Mwendera et al., 2017; Tiew, 1998). Près de 50% des publications du corpus se situent dans cette perspective qui inclut notamment les études s'intéressant au financement de la production scientifique. Ces études utilisent généralement les remerciements comme source de données pour mesurer les tendances dans le financement et l'effet du financement sur la productivité et sur l'impact de chercheurs, de groupes de recherche ou d'institutions (p. ex. Crawford et Biderman, 1970; Gök, Rigby et Shapira, 2016; Markusova, Libkind et Aversa, 2012; Mejia et Kajikawa, 2018; Morillo, 2016; Morley, Cameron et Bazermore, 2016; Rong, Grant et Ward, 1989; Shah et al., 2014; Stehr et Forkel, 2013; Zhou et Tian, 2014). Finalement, les remerciements comme forme d'attribution du crédit constituent l'angle dominant du champ de recherche des remerciements d'un point de vue quantitatif puisque 66% des publications ont adopté cette perspective. Cette perspective inclut les publications qui positionnent les remerciements aux côtés de l'autorat, des citations et des mentions de contribution dans l'écosystème de l'attribution du crédit scientifique.

1.4.2 Reconnaissance, crédit et remerciements

En 1972, Mackintosh publie une thèse doctorale en sociologie consacrée aux remerciements dans la communication savante. Avec cette contribution, Mackintosh campe l'étude des remerciements dans la sociologie des sciences, s'ancrant dans le cadre conceptuel du système de la reconnaissance scientifique de Merton (1973). Mackintosh (1972) propose dans sa thèse une première typologie qui classe les remerciements selon trois catégories : les remerciements relatifs aux 1) installations et équipements (*facilities*), 2) à l'accès aux données (*access to data*) ainsi que 3) l'aide de la part d'individus (*help rendered by individuals*).

Reprenant à son tour la perspective du système de la reconnaissance scientifique de Merton, Cronin est sans contredit l'auteur ayant le plus contribué à l'étude des remerciements dans le contexte de la communication savante, avec plus d'une vingtaine de publications

scientifiques consacrées à l'analyse des remerciements au cours des vingt-cinq dernières années (Desrochers, Paul-Hus et Pecoskie, 2017). Cronin suggère que les remerciements soient considérés aux côtés de l'autorat et des citations au sein du triangle de la reconnaissance (Cronin et Weaver-Wozniak 1993; Cronin, 1995) puisque ces trois éléments permettent d'attribuer une forme de crédit scientifique. En effet, Cronin postule que l'autorat et les citations n'offrent qu'un portrait partiel des contributions et influences ayant mené à une publication savante et que les remerciements devraient être considérés comme une forme de reconnaissance scientifique.

Dans leur étude comparant la collaboration en sciences économiques et en biologie, Laband et Tollison (2000) mesurent la collaboration formelle par le nombre de co-auteurs signant une publication et la collaboration informelle par le nombre d'individus mentionnés dans la rubrique des remerciements. Les auteurs montrent que bien que la collaboration formelle soit plus fréquente en biologie, la collaboration intellectuelle informelle est plus importante en sciences économiques.

Patel (1973) constate pour sa part que certains détails révélateurs des dynamiques du travail de groupe ne sont conservés nulle part ailleurs que dans les remerciements d'une publication scientifique. Selon la perspective du triangle de la reconnaissance, les remerciements, comme l'autorat et les citations, témoignent du fait que l'écriture scientifique ne survient pas dans un vacuum sociocognitif. Cronin nous rappelle ainsi que l'écriture, dans de nombreux domaines du champ scientifique, est un acte social (Cronin, 2005). Les remerciements permettraient dans certains cas de révéler les alliances académiques et généalogies d'influences d'un chercheur. En ce sens, les remerciements participent à la construction de la réputation d'un chercheur par ses associations et relations alléguées, par la mise en scène de son capital symbolique (Waquet, 2005, p. 361). C'est dans ce contexte que Cronin et Weaver-Wozniak (1993) envisageaient la création d'un index des remerciements, similaire aux index de citations, dans le but d'inclure les remerciements dans les exercices d'évaluation bibliométriques des chercheurs. L'index des remerciements, tel qu'imaginé par Cronin et ses collègues dans les années 1990, ne verra jamais le jour. Toutefois, McCain (2018) a récemment repris l'idée d'un index des remerciements. McCain (2018) étudie la faisabilité de créer un *Personal Acknowledgements Index*, mais identifie les nombreuses difficultés liées à la création d'un tel index.

Cependant, depuis 2008, les remerciements qui contiennent une mention de financement sont indexés et disponibles à l'échelle agrégée dans la base de données bibliographiques Web of Science (WoS) sous le champ *Funding Text* pour tous les articles indexés dans le *Science Citation Index (SCI)*. Les remerciements contenus dans les articles indexés dans le *Social Sciences Citation Index (SSCI)* ont commencé à être collectés plus récemment, en 2015 (Paul-Hus, Desrochers et Costas, 2016).

Se positionnant clairement dans le cadre conceptuel du triangle de la reconnaissance de Cronin, Costas et van Leeuwen (2012) présentent la première analyse bibliométrique à grande échelle des données de remerciements indexées par WoS. L'analyse de la relation entre les citations et les remerciements montre que la présence de remerciements est associée à un nombre moyen plus élevé de citations et à des revues ayant un facteur d'impact plus élevé. Les résultats de Costas et van Leeuwen démontrent la pertinence de l'analyse des remerciements en ce qu'elle offre de nouvelles opportunités pour l'étude de l'influence scientifique et des relations entre chercheurs, institutions et organismes subventionnaires.

Comme Mackintosh (1972) et McCain (1991) avant lui, Cronin propose en 1991 une typologie des remerciements, qui sera modifiée au fil des ans et prendra finalement la forme d'une typologie en six catégories de soutien : 1) conceptuel, 2) éditorial, 3) financier, 4) technique, 5) moral et 6) indéterminé⁶ (Cronin, Shaw et LaBarre, 2004). Plusieurs autres chercheurs proposeront également leur propre classification des remerciements (p. ex. Ben-Ari, 1987; Giannoni, 2006a; Giles et Councill, 2004; Heffner, 1981). Toutefois, seul le modèle élaboré par Cronin (1991; Cronin, Shaw et La Barre, 2004) se démarquera et sera repris, dans une version ou une autre, par plusieurs auteurs (mentionnons entre autres Giannoni, 1998; Hyland, 2003; 2004; Koley et Sen, 2013; Rattan, 2013; Salager-Meyer et al., 2006; Salager-Meyer, Alcaraz-Ariza et Berbesí, 2009; Tiew et Sen, 2002). Cependant, à ce jour, aucune étude n'a permis de démontrer l'applicabilité d'une telle typologie sur des données agrégées à grande échelle.

⁶ Notre traduction des catégories suivantes : « conceptual; editorial; financial; instrumental/technical; moral; and unknown ».

Les enquêtes abordant la question des remerciements auprès de la communauté scientifique, bien que fort révélatrices des pratiques des chercheurs en matière d'attribution du crédit et fréquemment reprises et citées, constituent une méthode moins souvent privilégiée par les chercheurs s'étant penchés sur les remerciements. L'étude pionnière de Spiegel et Keith-Spiegel (1970) s'intéresse aux pratiques d'autorat et de remerciements ainsi qu'à l'éthique de l'attribution du crédit en psychologie. Déjà, en 1970, les répondants sondés disaient être favorables à l'idée d'une section détaillant la contribution de chaque membre de l'équipe, auteurs et remerciés, assurant ainsi à chacun une juste reconnaissance de la contribution faite qui, lorsque jugée mineure, est trop fréquemment non reconnue. Plusieurs études reprendront l'instrument de collecte proposé par Spiegel et Keith-Spiegel (1970) : une décennie plus tard, Bridgewater, Bornstein et Walkenbach (1981) mettent à jour l'enquête de Spiegel et Keith-Spiegel et corroborent les résultats obtenus par l'étude originale alors que Werley et al. (1981), Waltz, Nelson et Chambers (1985), Templeton Gay, Griffith Lavender et McCard (1987), Butler et Ginn (1998) et Ginn et Butler (1998) reprendront tout le questionnaire développé par Spiegel et Keith-Spiegel et l'appliqueront dans le contexte des sciences infirmières et autres sciences de la santé. De manière générale, deux constats se dégagent de ces enquêtes : 1) le crédit devrait uniquement être attribué en fonction de la contribution d'un individu et non en fonction de son rang académique (étudiant, technicien, professionnel ou professeur) et 2) l'autorat et le contenu des remerciements devraient être discutés avec les individus impliqués dans un projet de recherche dès le début du projet.

Toujours dans les années 1970, Heffner (1979) emprunte une voie différente et sonde plus de 200 individus remerciés dans des publications scientifiques. Les résultats de son enquête montrent qu'en dépit des pratiques prescrites, le crédit n'est pas toujours attribué en fonction de critères universels en contexte collaboratif. En effet, selon les répondants, le genre et le statut académique (le fait de détenir ou non un doctorat) des individus semblent avoir une influence sur leur inclusion ou non dans la liste des auteurs.

Des années plus tard, en 1994, Cronin et Overfelt publient les résultats d'un sondage effectué auprès de 278 chercheurs américains concernant leurs pratiques ainsi que leurs attentes et opinions en matière de remerciements. Bien qu'il semble y avoir consensus en ce qui a trait à la fonction des remerciements — ils permettent de créditer de façon appropriée les

contributions non triviales qui sont toutefois insuffisantes pour mériter le statut d'auteur — très peu de répondants souhaitent voir les remerciements inclus dans le processus d'évaluation des chercheurs.

En 1995, Cronin publie *The scholar's courtesy: The role of acknowledgement in the primary communication process*, un ouvrage phare pour l'étude des remerciements dans le contexte de la communication savante qui rassemble et synthétise plusieurs de ses études précédentes sur les remerciements et permet d'asseoir le concept de triangle de la reconnaissance.

En 2002, à l'aide d'entrevues et de données bibliométriques, Laudel (2002) montre que certains types de contributions sont seulement crédités dans les remerciements alors qu'une part non négligeable des contributions demeure complètement invisible, ne menant ni au statut d'auteur ni aux remerciements. Plus récemment, dans une enquête s'appuyant sur 2 327 répondants, Jabbedhari et Walsh (2017) montrent que les collaborateurs non-auteurs sont fréquents et que plus de la moitié des publications étudiées inclut au moins un collaborateur non-auteur.

Dans leur enquête s'intéressant aux conflits en matière d'attribution de crédit scientifique et effectuée auprès de 600 chimistes américains, Seeman et House (2010a, 2010b, 2015) montrent que pour une majorité des répondants, une mention dans les remerciements constitue la forme de crédit la plus appropriée pour une contribution essentielle, mais somme toute limitée. Les auteurs soulignent toutefois que bien qu'un remerciement soit présent, tangible et réel, il demeure simultanément invisible, les remerciements pour les contributions individuelles n'étant répertoriés nulle part de façon systématique et n'étant pris en compte dans aucune évaluation (Seeman et House, 2010a).

Après plus de cinquante ans de recherche, la littérature n'offre toujours pas de consensus quant à la valeur des remerciements dans le contexte du système de la reconnaissance scientifique. De nombreuses études se sont intéressées à la description du contenu des remerciements, mais aucune n'a été en mesure de comparer, à l'aide de données à grande échelle, les différentes pratiques disciplinaires en ce qui a trait aux contributions et aux individus remerciés. De plus, plusieurs études quantitatives récentes soulignent l'importance des analyses

qualitatives qui offrent une compréhension contextualisée du contenu des remerciements, complémentaires aux analyses à grande échelle.

Dans ce contexte, notre projet de recherche propose d'allier les méthodes quantitatives et qualitatives pour analyser le contenu des remerciements et tenter de mieux comprendre la valeur des remerciements en termes de crédit scientifique. Le prochain chapitre permet de détailler la méthodologie utilisée dans le cadre de notre recherche doctorale.

Chapitre 2. Méthodologie

Ce chapitre permet de détailler la méthodologie utilisée pour répondre à nos objectifs et questions de recherche. Nous présentons d’abord le devis méthodologique employé, suivi de la description de la collecte et de l’analyse des données. Finalement, nous présentons les mesures mises en place afin d’assurer la qualité de la recherche.

2.1 Devis mixte

Un devis mixte nous permet d’aborder les trois axes de notre recherche et de répondre aux objectifs descriptifs et exploratoires de notre projet. Tel qu’illustré dans la Figure 1, dans la première phase du projet nous avons procédé à l’analyse linguistique du contenu des textes de remerciements par traitement automatique du langage naturel afin d’en extraire des unités significatives de sens en regard de nos questions de recherche. Cette analyse, réalisée en deux temps, visait à extraire les entités nommées (individus remerciés) et les syntagmes nominaux (contributions ayant fait l’objet d’une reconnaissance).

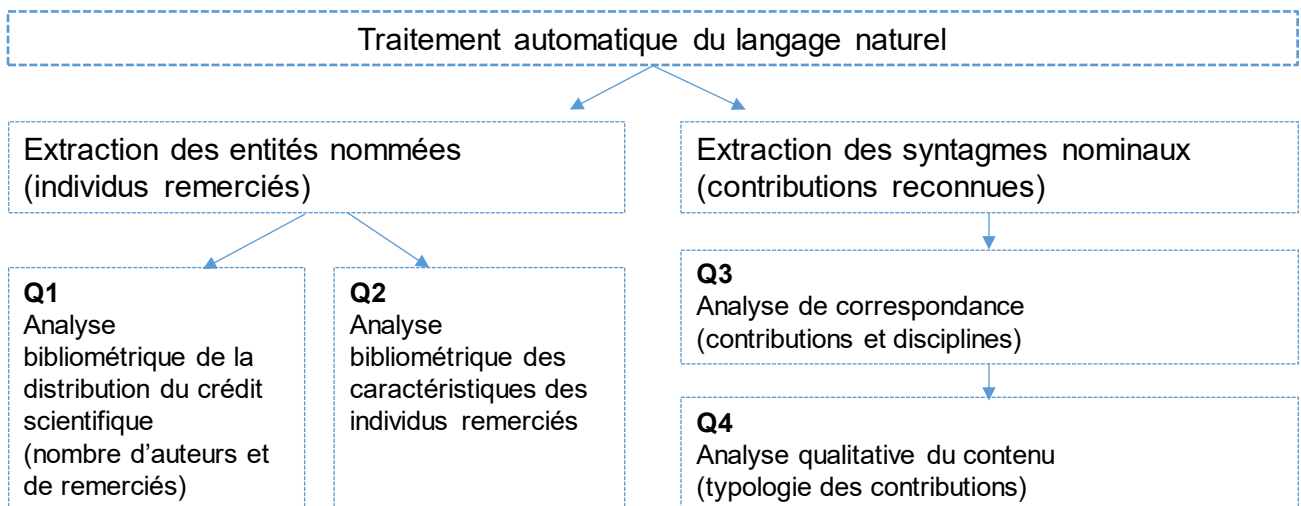


Figure 1. – Devis méthodologique

Dans une deuxième phase, les analyses bibliométriques et l’analyse de correspondance nous ont permis de mettre en relation le contenu des remerciements avec les différentes métadonnées des

articles qui les contiennent (auteurs, disciplines). Alors que l'analyse bibliométrique du nombre d'auteurs et de remerciés permet d'obtenir un portrait global des pratiques d'attribution du crédit scientifique, l'analyse bibliométrique des caractéristiques des individus remerciés permet d'aller plus loin et de décrire les individus remerciés en termes de genre et de statut académique.

La troisième phase a pour objectif de présenter les résultats quantitatifs de l'analyse de correspondance effectuée sur les syntagmes nominaux (contributions reconnues) en les contextualisant par une analyse qualitative de contenu d'un sous-ensemble du corpus des remerciements. L'analyse qualitative de contenu permet donc d'établir une typologie des contributions reconnues à la lumière à la fois des résultats de l'analyse à grande échelle et de l'analyse contextualisée des textes.

Notre thèse par articles intègre dès lors les trois axes de notre projet de recherche : 1) l'attribution du crédit scientifique, 2) les individus remerciés et 3) les contributions qui mènent aux remerciements. Quatre articles scientifiques visent à répondre à nos quatre questions de recherche (QR). L'article *The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements* (Paul-Hus et al., 2017a) aborde la QR1 qui concerne l'attribution du crédit scientifique en contexte collaboratif. L'article *Who are the acknowledgees? An analysis of gender and academic status* (Paul-Hus et al., 2020) permet de répondre à la QR2 et s'intéresse aux individus remerciés et investigate leurs caractéristiques (genre et statut académique). Deux articles s'intéressent aux contributions reconnues dans les remerciements : l'article *Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences* (Paul-Hus et al., 2017b) aborde la QR3 et présente une analyse à grande échelle des contributions remerciées; alors que l'article *Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015* (Paul-Hus et Desrochers, 2019) permet de répondre à la QR4 et propose une typologie des contributions remerciées.

2.2 Collecte de données

2.2.1 Données de remerciements

Notre étude porte sur les remerciements inclus dans les articles scientifiques puisque ces documents sont issus du processus de révision par les pairs et sont, en bibliométrie, généralement considérés comme des contributions originales aux connaissances scientifiques (Larivière et Sugimoto, 2018; Moed, 2005). Pour décrire les fonctions des remerciements dans le système de la reconnaissance scientifique et explorer les facteurs qui peuvent influencer leurs caractéristiques, nous avons analysé le contenu des remerciements disponibles à l'échelle agrégée dans la base de données bibliographiques Web of Science (WoS) de Clarivate Analytics.

Deux bases de données bibliographiques sont principalement utilisées en bibliométrie : Web of Science (WoS) de Clarivate Analytics et Scopus d'Elsevier (Mongeon et Paul-Hus, 2016). En plus des données bibliographiques traditionnelles, ces deux bases de données ont la particularité d'indexer les références des articles et les adresses institutionnelles de tous les auteurs. WoS collecte et indexe, depuis août 2008, les remerciements présents dans les articles scientifiques répertoriés dans la base de données *Science Citation Index Expanded* (SCIE). Depuis 2015, les remerciements des articles répertoriés dans la base de données *Social Sciences Citation Index* (SSCI) sont également collectés (Paul-Hus, Desrochers et Costas, 2016). C'est en réponse aux exigences de nombreux organismes subventionnaires visant l'identification formelle des sources de financement d'une publication que WoS a commencé à indexer le contenu des remerciements en août 2008. L'objectif de WoS est de fournir des données qui permettent aux chercheurs et aux organismes subventionnaires d'évaluer l'impact du financement en termes de publications et de citations. Toutefois, c'est l'indexation du plein texte des remerciements qui permet d'aller au-delà des analyses de financement et d'étudier le contenu des remerciements en termes d'unités et de contributions remerciées :

Web of Science indexes the Funding Agency and, if available, any grant numbers. We also index the source text from the original article for understanding of the context of the acknowledgement. (Web of Science, 2009)

En effet, les données de remerciements collectées et indexées par WoS sont structurées en trois champs : *Funding Text*, *Funding Agency* et *Grant Number*. Le champ *Funding Text* contient le

texte des remerciements tel qu'il apparaît dans l'article dont il est issu; par conséquent il contient les informations relatives au financement, mais également tous les autres types de contributions reconnues par les auteurs. Le champ *Funding Agency* contient le nom des organismes ou des institutions remerciées pour leur contribution financière et le champ *Grant Number* contient le numéro de la subvention ou du programme de financement, si ce dernier est identifié dans les remerciements. La Figure 2 présente un exemple des données de remerciements, telles qu'indexées par WoS.

Funding

Funding Agency	Grant Number
China Scholarship Council	
University of Queensland	
National Science Foundation of China	C1304013151101138

Close funding text

The authors would like to thank Mr Douglas Lush, Queensland Department of Agriculture and Fisheries at the Leslie Research Facility, Toowoomba, Australia, for supplying the barley samples. We also appreciate the help of Dr. Alex Wu for model fitting and Prudence Powell for SEC analysis. Wenwen Yu gratefully acknowledges the support of the China Scholarship Council and the University of Queensland. We acknowledge the support of a National Science Foundation of China grant C1304013151101138.

Figure 2. – Extrait de la notice Web of Science pour la publication

<http://dx.doi.org/10.1016/j.carbpol.2016.08.078>

Les données de remerciements ne sont toutefois pas collectées de façon systématique par WoS. En effet, les remerciements d'une publication sont collectés et indexés seulement s'ils incluent une mention de financement. De plus, l'indexation des remerciements varie en fonction de l'année de publication, du type de document, de la langue de publication et de l'index considéré (Paul-Hus, Desrochers et Costas, 2016). À partir de 2008, les remerciements de tous les articles originaux (*article*) et articles de synthèse (*review*) inclus dans le *Science Citation Index Expanded* (SCIE) sont collectés et indexés par WoS, mais seulement à condition qu'ils

contiennent une mention de financement et qu'ils soient rédigés en anglais. En 2015, WoS a élargi sa politique d'indexation des remerciements aux articles originaux et de synthèse du *Social Sciences Citation Index* (SSCI) (Paul-Hus, Desrochers et Costas, 2016; Thomson Reuters, 2015). L'*Arts & Humanities Citation Index* (A&HCI) est désormais le seul index de WoS pour lequel les remerciements ne sont pas disponibles. Malgré toutes ces limites, WoS demeure à ce jour la source bibliométrique la plus exhaustive en ce qui concerne les remerciements. Depuis 2013, la base de données bibliographique Scopus d'Elsevier collecte également les données relatives au financement à partir du texte des remerciements. Toutefois, Scopus n'indexe le plein texte des remerciements qui accompagne les informations concernant le financement que depuis août 2017 (Elsevier, 2017).

Compte tenu de ces limitations, la population accessible dans le cadre de ce projet de recherche est limitée aux remerciements rédigés en anglais, contenant des informations de financement et provenant des articles originaux (*article*) et de synthèse (*review*) indexés dans WoS. Afin de pouvoir comparer l'ensemble des disciplines des sciences naturelles, biomédicales et du génie (SCIE) aux sciences sociales (SSCI), les données ont été restreintes aux publications parues depuis 2015. Deux jeux de données ont été analysés dans le cadre de notre thèse : les remerciements provenant des articles publiés en 2015 pour l'analyse bibliométrique de la distribution du crédit scientifique et l'analyse des contributions reconnues (analyse de correspondance et analyse qualitative du contenu) et les remerciements provenant des articles publiés entre 2015 et 2017 pour l'analyse bibliométrique des individus⁷. Le tableau 1 présente la répartition de ces articles par discipline ainsi que la proportion que représente ce sous-ensemble par rapport à l'ensemble des articles publiés pour la période étudiée. Le Tableau 1 montre par ailleurs que les pourcentages d'articles contenant des remerciements indexés dans WoS sont relativement stables par disciplines aux cours des trois années d'observation.

Une version locale de la base de données WoS, hébergée sur un serveur SQL à l'Observatoire des sciences et des technologies (OST) de l'Université du Québec à Montréal, a

⁷ Le deuxième jeu de données a été élargi et couvre les années 2015 à 2017 puisque cette portion de l'analyse a été effectuée plus tard durant le projet de recherche et des données plus récentes étaient alors disponibles.

été optimisée afin de répondre aux requêtes bibliométriques. Les données de remerciements, ainsi que les métadonnées des articles qui les contiennent, ont été recueillies à l'aide d'une requête visant à sélectionner l'ensemble des articles publiés en 2015 (puis entre 2015 et 2017 pour le second jeu de données) pour lequel le champ des remerciements (*Funding Text*) est non nul. Outre les remerciements, les métadonnées recueillies comprennent le nom des auteurs de l'article, leur adresse d'affiliation institutionnelle, le titre de la revue dans lequel l'article est publié (ce qui permet d'attribuer une discipline à chaque article), les références bibliographiques ainsi que les citations reçues par l'article.

Tableau 1. – Nombre d’articles qui inclut des remerciements et pourcentage du total des articles publiés, par année de publication et par discipline⁸

	2015			2016			2017			Total		
	N articles	N articles avec remerciements	%	N articles	N articles avec remerciements	%	N articles	N articles avec remerciements	%	N articles	N articles avec remerciements	%
Biologie	105 279	76 281	72%	111 389	82 865	74%	114 923	86 912	76%	331 591	246 058	74%
Champs professionnels	41 015	12 552	31%	43 323	15 399	36%	44 095	15 871	36%	128 433	43 822	34%
Chimie	151 947	123 806	81%	152 985	124 710	82%	148 599	123 245	83%	453 531	371 761	82%
Génie et technologie	241 124	165 590	69%	263 364	179 402	68%	280 871	203 280	72%	785 359	548 272	70%
Mathématique	49 997	35 390	71%	49 905	33 708	68%	51 145	35 150	69%	151 047	104 248	69%
Médecine Clinique	389 311	218 367	56%	404 251	223 123	55%	400 130	229 706	57%	1 193 692	671 196	56%
Physique	124 556	95 676	77%	126 022	93 847	74%	124 058	97 411	79%	374 636	286 934	77%
Psychologie	31 286	15 085	48%	32 267	17 192	53%	32 092	17 822	56%	95 645	50 099	52%
Recherche biomédicale	189 066	158 067	84%	200 399	164 047	82%	199 219	164 805	83%	588 684	486 919	83%
Sciences de la santé	37 309	18 703	50%	38 287	20 523	54%	39 104	21 919	56%	114 700	61 145	53%
Sciences de la terre	92 238	72 922	79%	100 365	78 306	78%	102 810	82 368	80%	295 413	233 596	79%
Sciences sociales	50 420	16 972	34%	52 466	20 765	40%	52 433	22 098	42%	155 319	59 835	39%
Total	1 503 548	1 009 411	67%	1 575 023	1 053 887	67%	1 589 479	1 100 587	69%	4 668 050	3 163 885	68%

⁸ Classification disciplinaire présentée à la p.59.

2.3 Analyse des données de remerciements

Dans cette section du chapitre, nous présentons les différentes analyses effectuées sur les données de remerciements afin de répondre à nos questions de recherche. Le traitement automatique du langage naturel permet d'extraire les entités nommées et les syntagmes nominaux des textes de remerciements. Les analyses bibliométriques visent à croiser les données extraites des remerciements avec des variables bibliométriques. L'analyse de correspondance permet d'analyser les syntagmes nominaux les plus fréquemment utilisés dans les remerciements en fonction des disciplines. Finalement, l'analyse qualitative de contenu vise à contextualiser les résultats obtenus par l'analyse de correspondance.

2.3.1 Traitement automatique du langage naturel

Les méthodes utilisées en traitement automatique du langage naturel proviennent de la linguistique, de l'informatique et de l'intelligence artificielle (Jurafsky et Martin, 2009). Notre objectif n'est donc pas ici de dresser un portrait exhaustif de ce champ de recherche, mais plutôt de brièvement présenter la méthode utilisée pour analyser le contenu des remerciements.

Afin d'identifier et de distinguer les individus remerciés et les types de contributions mentionnées dans les remerciements, l'analyse linguistique effectuée dans le cadre de ce projet se concentre sur les entités nommées pour les individus remerciés et sur les syntagmes nominaux pour les contributions reconnues.

Extraction des entités nommées⁹

De façon générale, les entités nommées désignent les noms propres qui réfèrent aux personnes, aux organisations (institutions, compagnies, organismes de financement) et aux lieux (villes, provinces, régions et pays). Dans le cadre de notre projet de recherche, seuls les noms de personnes référant à des individus ont été retenus. Les noms de personnes figurant dans le nom d'entités collectives ont donc été enlevés (p. ex. Fondation Marie Curie). Afin d'extraire les individus remerciés des textes des remerciements, nous avons

⁹ Certains passages de cette sous-section sont traduits et adaptés de Paul-Hus et al. (2017a).

utilisé le module Stanford *Named Entity Recognizer* (NER) (Finkel et al., 2005) du *Natural Language ToolKit* (Bird, Klein et Loper, 2009). Le Stanford NER permet d'identifier les séquences de mots qui réfèrent à un nom propre et d'y apposer une étiquette selon les trois classes suivantes : personne, organisation et lieu. Il est à noter que cet outil est développé et testé pour une performance maximale sur les textes rédigés en anglais. La performance de l'algorithme pour les entités nommées dans d'autres langues peut donc s'en voir réduite. Toutefois, le Stanford NER demeure un des systèmes d'identification d'entités nommées, disponibles en libre accès, les plus robustes et efficaces (Ratinov et Roth, 2009).

La liste des entités nommées extraites des remerciements a ensuite été nettoyée afin d'en éliminer les entités non-humaines. Ce nettoyage s'est fait en plusieurs étapes. D'abord, les noms incomplets ont été enlevés (i.e. les entités contenant uniquement un prénom, un nom de famille, ou une initiale) afin de conserver seulement les entités composées d'au moins une initiale et un nom de famille. Afin d'éliminer les noms ne référant pas à des personnes, les noms de famille contenu dans notre liste d'entités nommées ont été comparés à une liste de noms de famille validés comme appartenant à des personnes, c'est-à-dire la liste de tous les noms de famille d'auteurs apparaissant sur une publication indexées dans WoS pour la période 1900 à 2016. Cette liste de 2 649 212 noms de famille distincts a donc été utilisée comme liste de référence pour valider les noms de famille de notre liste d'entités nommées. Une entité nommée pour laquelle il n'y aurait aucune correspondance avec cette liste de référence serait donc considérée comme ne référant pas à une personne et éliminée. Par la suite, un nettoyage manuel a été effectué afin d'éliminer toutes entités nommées restantes qui ne réfèraient pas à un individu (comme dans le cas de noms de bourses, de fondations, d'organisations ou d'institutions). Finalement, les remerciements incluent souvent le nom d'un ou de plusieurs auteurs signant l'article duquel le remerciement a été extrait. Afin d'éviter les faux positifs et de compter ces auteurs comme des remerciés, nous avons enlevé les noms d'auteurs se désignant eux-même dans les remerciements d'un article spécifique de notre liste d'entités nommées. Lorsqu'un ou plusieurs noms extraits des remerciements d'un article X correspondait au nom d'un des auteurs (première initiale et nom de famille) de ce même article, ce nom a été enlevé de la liste des remerciés provenant de cet article spécifique, tel qu'illustré par l'exemple suivant :

Article X

Auteurs: J. Zhang, X. Feng et Y. Xu

Remerciement: « Jinsong Zhang, Xiao Feng, and Yong Xu contributed equally to this work. The authors would like to thank Xiang Zhou for data collection and Li Yu for data processing. »

Liste d'entités nommées extraites pour l'article X: Jinsong Zhang, Xiao Feng, Yong Xu, Xiang Zhou et Li Yu.

Liste de remerciés nettoyée pour l'article X: Xiang Zhou et Li Yu

Extraction des syntagmes nominaux¹⁰

Les syntagmes nominaux sont des groupes de mots centrés sur un nom donné et qui, avec zéro ou plusieurs constituants de diverses catégories syntaxiques, exercent la même fonction grammaticale que les noms simples; un exemple serait le syntagme nominal « *financial support* » par rapport au nom simple « *support* ».

Plusieurs étapes de prétraitement linguistique ont été nécessaires pour extraire efficacement les syntagmes nominaux des courts textes de remerciements. Tout d'abord, le texte des remerciements extrait du champ *Funding Text* a été segmenté en mots à l'aide du *Penn TreeBank Tokenizer* (Marcus, Marcinkiewicz, Santorini, 1993; Marcus et al., 1994) du *Natural Language ToolKit* (Bird, Klein et Loper, 2009). Afin d'identifier les noms communs présents dans le corpus, chaque texte de remerciements a ensuite été analysé morphologiquement et syntaxiquement en utilisant le *Stanford Log-Linear Part-of-Speech (POS) Tagger* (Toutanova et Manning, 2000; Toutanova et al., 2003).

En ce qui concerne l'identification et l'extraction des syntagmes nominaux, de nombreux outils sont actuellement disponibles puisque la caractérisation des syntagmes nominaux demeure un domaine actif de recherche en linguistique. La segmentation des syntagmes nominaux a été ici implémentée à l'aide d'une version modifiée de l'ensemble de règles grammaticales pour l'extraction de mots-clés conçu par Kim et al. (2010). Cet ensemble se compose de deux règles. La première vise à détecter les composantes nominales d'un syntagme nominal, allant d'un seul nom, tel que « *funding* », à une

¹⁰ Certains passages de cette sous-section sont traduits et adaptés de Paul-Hus et al. (2017b).

séquence de noms ou encore d'adjectifs se terminant par un nom, comme dans le « *seismic measurement open access software package* ». La deuxième règle vise à fusionner toutes paires de composantes nominales consécutives identifiées par la première règle et séparées par toute préposition ou conjonction de coordination (« *of* », « *for* », « *with* », « *in* », « *if* », etc.), comme dans l'extrait « *technical assistance in bacterial challenge experiment degeneration* ». Cependant, telle que formulée par Kim et al. (2010), cette deuxième règle ne permet qu'une seule fusion, ce qui empêche l'identification complète de syntagmes formés de plus de deux composantes nominales tels que « *strategy for gene discovery in schizophrenia* » ou « *coating of functionalized polysaccharide with embedded nanoparticles* ». Afin d'extraire correctement de tels syntagmes nominaux, la deuxième règle de Kim et al. a été modifiée afin de permettre la fusion multiple de ce type de composantes.

Afin de nettoyer la liste des syntagmes nominaux extraits, nous avons d'abord éliminé tous ceux ne contenant qu'une référence à un numéro de financement. Les syntagmes nominaux composés d'une seule lettre ont également été éliminés. Puis, pour chaque syntagme nominal extrait, un score de fréquence a été généré en fonction du nombre d'occurrences du syntagme nominal dans le corpus. Afin d'augmenter l'efficacité des traitements et des analyses ultérieures, nous avons appliqué un seuil minimal de deux occurrences aux syntagmes nominaux de notre liste pour réduire de façon significative la taille de la liste de syntagme nominaux. Tous les syntgemes nominaux qui apparaissaient seulement une fois dans notre corpus ont donc été éliminés de notre liste. En effet, les syntagmes nominaux n'ayant qu'une occurrence dans un corpus donné (*hapax legomena*) représentent habituellement une très large proportion de l'ensemble des syntagmes nominaux du corpus et sont considérés, de par leur faible fréquence individuelle, comme ayant peu de valeur pour les techniques computationnelles dans le contexte de traitements automatique du langage naturel (Jurafsky et Martin, 2009).

2.3.2 Analyses bibliométriques¹¹

Afin d'arriver à une analyse significative des remerciements et d'être en mesure de répondre à nos questions de recherche, le contenu des remerciements a été croisé avec certaines variables bibliométriques. En effet, la littérature montre que les pratiques d'attribution du crédit scientifique varient notamment en fonction des disciplines étudiées, du statut académique et du genre des individus concernés. Le contenu des remerciements, c'est-à-dire les individus remerciés ainsi que les contributions mentionnées, a donc été analysé en fonction de la discipline d'une publication, du nombre d'auteurs signant la publication, du statut académique des auteurs et des remerciés et du genre des auteurs et des remerciés.

Classification disciplinaire

La classification disciplinaire de la National Science Foundation (NSF, 2006) a été utilisée afin d'attribuer une discipline à chacun des articles inclus dans nos analyses. La classification de la NSF est composée de 144 spécialités regroupées en quatorze disciplines. Contrairement à la classification disciplinaire du WoS qui attribue à chaque article entre une et quatre spécialités disciplinaires, la classification de la NSF attribue une seule discipline et une seule spécialité (sous-discipline) à chaque revue, évitant ainsi de compter plusieurs fois un même article assigné à plusieurs disciplines dans WoS. Dans le cadre de nos analyses, nous utilisons les douze disciplines de la classification de la NSF qui caractérisent les publications incluses dans le SCI-E et le SSCI de WoS : biologie, champs professionnels, chimie, génie et technologie, mathématiques, médecine clinique, physique, psychologie, recherche biomédicale, sciences de la santé, sciences de la terre et sciences sociales¹².

¹¹ Certains passages de cette sous-section sont traduits et adaptés de Paul-Hus et al. (2017a; 2020).

¹² Notre traduction des disciplines de la NSF : « Biology, Professional Fields, Chemistry, Engineering and Technology, Mathematics, Clinical Medicine, Physics, Psychology, Biomedical Research, Health, Earth and Space, and Social Sciences ».

Déterminer le statut académique des individus remerciés

Le statut académique des individus remerciés est opérationnalisé dans notre étude à l'aide de quatre indicateurs bibliométriques : 1) le nombre d'articles publiés (en tant qu'auteur) par un individu remercié, 2) le nombre total de citations obtenues par ces articles et normalisé selon la discipline, 3) l'âge académique des individus remerciés et 4) la proportion des articles pour laquelle un individu remercié occupe une position d'auteur principal. L'ensemble de ces indicateurs ne s'appliquent évidemment qu'aux individus remerciés ayant signé à titre d'auteur au moins un article indexé dans la base de données WoS.

Déterminer si un individu remercié est un auteur s'est avéré être une tâche complexe étant donné que nous ne disposons que du nom des individus remerciés. Pour déterminer le statut d'auteur de ces individus, nous avons d'abord désambiguïsé automatiquement tous les noms d'auteurs inclus dans WoS à l'aide de l'algorithme de Caron et van Eck (2014). Pour chaque nom d'individu remercié, nous avons identifié tous les auteurs désambiguïsés qui portaient le même nom. Nous avons considéré qu'une paire individu remercié-auteur désambiguïsé était valide lorsqu'une seule paire remercié-auteur existait pour une même discipline *ou* avec la même affiliation institutionnelle pour l'auteur et une des affiliations institutionnelles apparaissant sur l'article dans lequel l'individu est remercié. Avec ce critère de validation des paires, nous favorisons la précision plutôt que le rappel et visons ainsi à réduire le risque de faux positifs au sein de nos paires validées puisque les individus avec des noms très communs (pour qui plusieurs paires remercié-auteur auraient existés pour une même discipline ou avec une même affiliation) sont presque systématiquement exclus de notre jeu de données final. Une fois qu'une paire individu remercié-auteur désambiguïsé est identifiée et validée, nous disposons du profil de publication d'un individu et sommes en mesure de calculer nos indicateurs bibliométriques.

Le nombre d'articles signés en tant qu'auteur par les individus remerciés permet de quantifier la production scientifique des individus. Les publications sont ici limitées aux articles originaux et articles de synthèse et sont comptées de façon unitaire (un article = une unité pour chaque auteur indiqué dans l'article) (Larivière et Sugimoto, 2018).

Le nombre de citations reçues par une publication est généralement considéré comme un indicateur de l'impact de cette publication dans la communauté scientifique. En effet, en bibliométrie on considère que les publications les plus souvent citées dans les bibliographies de documents savants ont généralement un plus grand impact sur la production de nouvelles connaissances que les publications moins souvent citées (Larivière et Sugimoto, 2018). Toutefois, les pratiques de citation varient grandement d'une discipline à l'autre. Nous utilisons donc le total de citations normalisé par discipline afin de pouvoir comparer les disciplines entre elles et de quantifier l'impact obtenu par les publications signées par un individu remercié.

L'âge académique, qui permet de mesurer la séniorité d'un individu, est quant à lui calculé à l'aide de l'année de publication de la première publication signée en tant qu'auteur par un individu remercié (Radicchi et Castellano, 2013). Les publications d'un individu sont limitées aux articles originaux et articles de synthèse, publiés depuis 1980 (puisque avant cette date les données d'auteurs désambiguïsées ne sont pas disponibles). Dans le cadre de notre analyse, l'âge académique maximal des auteurs est donc de 36 ans.

La position dans la liste d'auteurs d'un article permet généralement de nous informer sur la nature et l'importance de la contribution d'un auteur. Dans le cadre de notre étude, nous considérons le premier auteur et l'auteur de correspondance (qui peuvent être des personnes distinctes ou la même personne) comme les auteurs principaux d'un article. En effet, les premiers auteurs sont souvent associés à la plus grande part de tâches effectuées dans une étude (Larivière et al., 2016) alors que les auteurs de correspondance sont souvent associés aux tâches de conception et de supervision d'une étude (Mattson, Sundberg et Laget, 2011). La proportion des articles pour laquelle un individu remercié occupe une position d'auteur principal nous aide donc à mieux caractériser le statut académique d'un individu remercié.

Ces quatre indicateurs permettent d'obtenir une caractérisation assez précise du statut académique des individus remerciés en matière de productivité, d'impact, de séniorité et de leadership et nous fournissent une évaluation indirecte de la position des individus remerciés au sein de la hiérarchie académique. Les résultats pour ces indicateurs sont présentés sous forme de distributions de valeurs. Afin de pouvoir analyser et comparer

les résultats de ces indicateurs chez les remerciés qui sont également auteurs, nous utilisons les distributions de valeurs de ces mêmes indicateurs chez tous les auteurs ayant publiés au moins un article (original ou de synthèse) indexé dans WoS au cours de la période 2015-2017. Pour ce qui est de la catégorisation disciplinaire, chaque auteur est assigné à la discipline pour laquelle il ou elle a le plus grand nombre de publications au cours de sa carrière. Lorsque deux disciplines sont ex aequo, une des deux disciplines a été choisie de manière aléatoire afin d'assigner une seule discipline par auteur.

Déterminer le genre des auteurs et des individus remerciés

Le genre des auteurs et des individus remerciés est déterminé à l'aide de l'algorithme Wiki-Gendersort (Bérubé et al., 2020) qui s'appuie uniquement sur les prénoms pour identifier le genre des personnes. L'algorithme utilise les données de Wikipédia pour attribuer un genre à un prénom. Ainsi, si une page Wikipédia a pour titre un prénom (seul ou suivi d'un nom de famille), l'ensemble des pronoms personnels et des pronoms possessifs contenus dans l'introduction de la page est analysé. Un genre est assigné à un prénom donné lorsque le même genre est attribué à 75% ou plus des occurrences rencontrées dans les 20 premières pages Wikipédia associées à ce nom. Si moins de 75% des pages analysées sont attribués à l'un ou l'autre des genres, le prénom est identifié comme unisexe. Bérubé et al. (2020) montre que l'algorithme Wiki-Gendersort obtient une performance comparable aux autres méthodes disponibles dans la littérature. Tel qu'illustré dans le Tableau 2, l'algorithme Wiki-Gendersort attribue correctement le genre dans plus de 97 % des 130 645 prénoms classés par l'algorithme, lorsque la performance de l'algorithme est comparée aux résultats de quatre bases de données qui associent un genre à des prénoms.

Tableau 2. – Résultats de l'algorithme Wiki-Gendersort comparées à quatre bases de données associant un genre à des prénoms (adapté de Bérubé et al., 2020)

Base de données	Noms identifiés (%)	Genre correctement attribué (%)
GenderChecker	90,65	98,39
Gender.c	95,85	98,95
NamSor	89,39	97,07
U.S. Census	97,28	99,91

La performance de l'algorithme Wiki-Gendersort varie cependant en fonction de la langue d'origine du prénom analysé. En effet, les algorithmes et bases de données d'association

de genre à un prénom ont généralement plus de difficulté à assigner un genre aux prénoms asiatiques, plus particulièrement aux prénoms chinois (Santamaría et Mihaljevic, 2018). Il en résulte que la proportion des prénoms chinois classés dans la catégorie « genre inconnu » par Wiki-Gendersort est plus élevée comparativement aux prénoms d'autres origines. Toutefois, cette limite est partagée par l'ensemble des bases de données d'assignation de genre analysées.

Le tableau 3 présente les résultats du processus d'assignation de genre aux noms d'auteurs et d'individus remerciés inclus dans notre jeu de données, à l'aide de l'algorithme Wiki-Gendersort. Au total, un genre a été assigné à 69 % des occurrences de noms personnels (auteurs et remerciés) dans notre jeu de données. Nos analyses utilisant la variable genre portent donc sur les occurrences de noms personnels pour lesquels un genre a pu être assigné.

Tableau 3. – Nombre et proportion des occurrences de noms personnels, par genre

	Hommes		Femmes		Inconnu		Total N
	N	%	N	%	N	%	
Auteurs	2 965 510	48.78%	1 299 618	21.38%	1 813 637	29.84%	6 078 765
Remerciés	1 549 304	46.08%	701 427	20.86%	1 111 160	33.05%	3 361 891
Total	4 514 814	47.43%	2 001 045	21.20%	2 924 797	31.44%	9 440 656

2.3.3 Analyse de correspondance¹³

L'analyse de correspondance est une méthode d'analyse multivariée qui permet d'examiner la relation entre les catégories de deux variables discrètes. L'objectif de l'analyse de correspondance est de révéler la structure sous-jacente à une matrice de données complexe. Au cœur de cette méthode, on retrouve le besoin de réduire un espace multidimensionnel en un espace bidimensionnel tout en conservant le plus d'information quant à la structure d'association des variables à l'étude (Beh et Lombardo, 2014). L'analyse de correspondance permet de représenter visuellement les associations entre des variables catégorielles sous forme de points situés dans un espace bidimensionnel, facilitant ainsi l'interprétation des résultats. Ainsi, les catégories ayant des distributions

¹³ Certains passages de cette sous-section sont traduits et adaptés de Paul-Hus et al. (2017b).

similaires seront représentées par des points rapprochés dans l'espace alors que les catégories présentant des distributions dissimilaires seront éloignées (Clausen, 1998).

Cette méthode a d'abord été proposée par Hirschfeld (1935), mais plusieurs chercheurs ont contribué aux développements de la méthode dans les années qui ont suivi, notamment Fisher (1940) et Guttman (1941). L'analyse de correspondance a été popularisée en France grâce aux travaux de Benzécri (1980), repris entre autres par Escofier et Pagès (2008) ainsi que par Lebart et Salem (1994). Toutefois, c'est l'application qu'en a faite Pierre Bourdieu dans ses analyses sociologiques (1979; 1984) qui explique en grande partie l'intérêt que portent aujourd'hui les sciences sociales à cette méthode.

Dans le cadre de notre projet, l'analyse de correspondance a été appliquée à l'aide du programme MultBiplot (Vicente-Villardón, 2015). L'analyse débute par la transformation des fréquences d'un tableau croisé en proportion afin d'obtenir, pour chaque ligne, un profil de fréquences relatives. Appliqué au corpus des remerciements pour notre projet, le tableau croisé est formé des syntagmes nominaux les plus fréquents du corpus en fonction des disciplines. De la même façon, le profil d'une colonne est obtenu en calculant la fréquence relative de chacun des syntagmes nominaux inclus dans l'analyse. Les lignes (et les colonnes) ayant des profils similaires se traduisent par des points rapprochés dans la représentation visuelle alors que les lignes (et les colonnes) ayant des profils différents produisent des points éloignés les uns des autres. En d'autres mots, si deux disciplines se trouvent à proximité l'une de l'autre dans un plan bidimensionnel, c'est qu'elles sont définies par des modèles (*patterns*) de remerciements similaires.

Il est toutefois important de noter que la proximité entre les points est mesurée en distances du χ^2 , ce qui signifie que le poids assigné à chaque élément est inversement proportionnel aux effectifs totaux d'une ligne ou d'une colonne. Ainsi, les catégories (lignes ou colonnes) qui ont moins d'occurrences totales contribuent relativement plus aux distances entre les points que les catégories qui ont plus d'occurrences (Beh et Lombardo, 2014). Dans le contexte de l'analyse de correspondance, les distances du χ^2 sont étroitement liées au concept de variance, qui est désigné par le terme inertie. L'inertie totale est une mesure de l'étendue de la dispersion des points (lignes et colonnes) par rapport au centre du plan bidimensionnel. L'inertie totale est décomposée en un ensemble de valeurs propres

(*eigenvalues*). Le nombre de valeurs propres nécessaires pour expliquer l'inertie totale est égal au nombre de dimensions nécessaire pour représenter les relations entre les points. Pour arriver à une analyse de correspondance parcimonieuse, on doit tenter d'expliquer la plus grande part de l'inertie à l'aide du plus petit nombre de dimensions, on vise alors à maximiser le ratio « données expliquées/paramètres requis ».

Dans le cas du présent projet, les syntagmes nominaux les plus fréquents se trouvent près de l'origine du plan alors que les syntagmes nominaux moins fréquents se trouvent plus loin des axes. Étant donné le biais de sélection du WoS pour les remerciements qui contiennent une mention de financement, il est attendu que les syntagmes nominaux relatifs au financement soient plus fréquents que ceux qui désignent d'autres types de contributions (techniques ou intellectuelles par exemple). L'analyse de correspondance permet donc de mettre l'accent sur les modèles de remerciements moins fréquents et potentiellement plus distinctifs.

2.3.4 Analyse qualitative de contenu¹⁴

Les types de contributions reconnues dans les remerciements, identifiés par le processus de traitement automatique du langage naturel et examinés au niveau macro par l'analyse de correspondance, ont ensuite fait l'objet d'une analyse qualitative de contenu. Plus précisément, l'analyse de correspondance a mis en lumière les tendances disciplinaires en matière de contributions remerciées alors que l'objectif de l'analyse qualitative est de contextualiser les résultats obtenus par les méthodes quantitatives en retournant au texte original des remerciements. En effet, dans certains cas, le sens d'un syntagme nominal est étroitement lié à son contexte, notamment disciplinaire. Par exemple, le syntagme nominal « speaker » réfère à plus d'une signification dans le contexte des remerciements. En médecine clinique, le terme « speaker » est principalement utilisé dans les déclarations de conflits d'intérêts, comme dans l'extrait suivant : « Dr. Aversa has received payments for the provision of [...] Participation as a speaker at scientific congresses: Aegerion, AstraZeneca, MSD, Mediolanum, Sanofi, Amgen. » (Aversa, 2015). Dans d'autres

¹⁴ Certains passages de cette sous-section sont traduits et adaptés de Paul-Hus et Desrochers (2019).

disciplines, par contre, le même terme est plutôt utilisé pour remercier un individu pour une contribution précise, comme dans l'exemple suivant où les auteurs soulignent la contribution d'un anglophone à la rédaction ou la révision du manuscrit : « we also would like to thank Dr. Paul Hobson, native speaker, for revision of the manuscript » (Gastón et al., 2015).

L'analyse qualitative s'est concentrée sur les syntagmes nominaux les plus fréquents du corpus des remerciements, tel qu'identifiés par l'analyse de correspondance. Les 214 syntagmes nominaux les plus fréquents¹⁵ du corpus des remerciements ont d'abord été décomposés en mots afin de pouvoir éliminer les mots occurants dans plus d'un syntagme nominal; ces mots redondants ont ensuite été exclus des mots unitaires (p. ex. « *technical assisance* » est décomposé en « technical » et « assistance » et comme le mot « assistance » est présent dans plus d'un syntagme nominal de notre corpus, une seule occurrence est conservée). Nous obtenons alors une liste de 154 mots distincts. Vingt mots ont ensuite été sélectionnés dans cette liste suivant les principes de l'échantillonnage par choix raisonné. Les cas retenus pour le codage ont été sélectionnés puisqu'ils offraient des manifestations utiles et pertinentes du phénomène à l'étude (Patton, 2002). Ainsi, la sélection des mots inclus dans notre échantillon s'appuie sur les résultats de l'analyse de correspondance qui met notamment en lumière les énoncés et déclarations préformulés comme dans l'exemple suivant : « *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript* ». Une attention particulière a donc été accordée aux mots fréquemment utilisés dans ce genre d'énoncés (p. ex. *analysis, collection, design, preparation*). L'échantillonnage a également été orienté vers les mots à caractère polysémique, ceux-ci pouvant potentiellement mener à des significations contextuelles différentes.

L'arbre de codage développé pour l'analyse qualitative a été construit de façon inductive par deux codeurs, en suivant le principe de réconciliation dans tous les cas de désaccord (« *negotiated agreement* », Campbell et al., 2013, p. 305; voir aussi Hruschka et

¹⁵ Les 214 syntagmes nominaux les plus fréquents ont été définis par un seuil minimal de 2 000 occurrences dans le corpus.

al., 2004; Schreier, 2012). L'arbre de codage, présentée dans le Tableau 4, est composé de 13 catégories. Le codage a ensuite été effectué selon le principe de saturation théorique en s'appuyant sur un échantillon qualitatif de textes de remerciements contenant un mot donné. De plus, l'échantillon est stratifié par discipline afin de pouvoir rendre compte de potentielles différences disciplinaires dans l'utilisation des mots. Les vingt mots de l'échantillon ont été codés dans un minimum de quinze phrases extraites du texte de remerciement original pour chacune des douze disciplines incluses dans notre analyse, pour un total de 3 754 cas codés. Les vingt mots ont été codés dans leur contexte de remerciements original, en utilisant la phrase comme unité d'analyse. Le codage est guidé par la question : « Dans quel contexte est utilisé ce mot? ». Un seul code a été appliqué à chacune des occurrences codées avec pour objectif de qualifier le contexte dans lequel le mot est utilisé.

Tableau 4. – Arbre de codage : catégories de contenu des remerciements et leur définition

Category	Definition	Example
Financial disclosure	Includes all types of funding and financial support or assistance.	“The financial assistance of the National Research Foundation (NRF grant: Unlocking the future- FA2007043000003) towards this research is hereby acknowledged.”
Conflict of interest	Refers to potential or actual conflict of interest or the absence of conflict of interest, which can be financial or otherwise.	“P.A.P. has an equity interest in Digital Proteomics, LLC, a company that may potentially benefit from the research results.”
Disclaimer	Responsibility disclaimer that content/opinions/conclusions are those of the author(s) solely and not of the funder or of another organization.	“The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.”
Ethics	Refers to ethical review, ethical approval of the research; can include some form of "seal of approval" by agencies.	“Institutional review board approval was granted under University IRB PRO12110345.”
Peer communication	Refers to intellectual contribution and communication with colleagues and peers, trusted assessors. Includes the process of comments, feedback, suggestions and peer review.	“Tim Birt, David Anderson, Anna Tigano, Rebecca Taylor, Nathaniel Clark, Catherine Dale and Raphael Lavoie provided insightful discussions.”
Investigation and Analysis	Refers to specific tasks such as the collection, treatment and analysis of data; the cycle of pre-writing work.	“Thanks Dr. Dongliang Li and Dr. Jianjun Cao from Nanjing Xiaozhuang University,

		for their help on field work and data analysis.”
Supervision and Management	Tasks and roles related to supervision, leadership and management responsibilities.	“Research included in this review was partly completed at the University of Newcastle, Australia, under the supervision of Dr John Clulow and Dr Micheal Mahony.”
Materials and Resources	Refers to all kinds of study materials, samples, computing resources, infrastructure, physical installations and instrumentation and reagents. People as objects of study (such as patients or population/sample) are also included.	“We thank Calcul Quebec and Compute Canada for access to the Mammoth supercomputer.”
Writing	Includes creation and/or presentation of the published work: original draft preparation, contribution to the writing itself; can include the creation of visualizations, maps, figures, tables, and illustrations.	“We thank Donald Cochrane, University of Saskatchewan, for his writing assistance”
Dissemination	Includes project, documents, and other forms of dissemination, such as conference presentations. Includes issues linked to cost of publication and open-access models.	“Data and supporting materials necessary to reproduce the numerical results will be available at www.hobolt.com upon publication.”
Organization	Refers to institutions or organizations, research centres, research groups, research chairs (can include funding organizations).	“The second author would like to thank Guangxi Experiment Center of Information Science.”
Combination	Two or more clear categories combined.	“The authors are grateful to the two referees and the editor for comments and suggestions and to Alfio Viola (University of Catania) for SEM assistance.”
Vague or other	Meaning cannot be inferred or is not covered by any other categories.	“We thank Jonas Klevas and Dainius Prakapavicius for their contribution during various stages of the paper preparation.”

2.4 Adéquation entre le devis méthodologique et les objectifs de la recherche

Notre recherche doctorale vise à décrire les fonctions des remerciements dans le système de la reconnaissance scientifique pour mieux comprendre leur valeur du point de vue du crédit scientifique, ainsi qu’à explorer les facteurs qui peuvent influencer leurs caractéristiques dans le contexte des pratiques d’attribution du crédit scientifique. De manière plus spécifique, notre recherche vise à décrire la façon dont le crédit scientifique

est distribué entre les auteurs et les remerciés qui ont contribué à une publication scientifique (QR1), les caractéristiques des individus remerciés dans les publications savantes, en termes de genre et de statut académique (QR2), les types de contributions mentionnées dans les remerciements (QR3) et les contextes d'utilisation des termes les plus fréquemment employés dans les remerciements (QR4).

La première phase de traitement sur les données de remerciements vise à extraire des unités significantes (entités nommées et syntagmes nominaux) des chaînes de textes par traitement automatique du langage naturel. L'extraction des entités nommées et les phases de traitement et de nettoyage qui s'en suivent permettent ainsi d'obtenir le nombre d'individus remerciés par article. Ces données, couplées aux données bibliométriques (classification disciplinaire et nombre d'auteurs par article) nous permettent de répondre à la QR1. Pour répondre à la QR2, nous reprenons les phases de traitements permettant d'extraire les entités nommées et d'obtenir la liste des individus remerciés dans chaque article, et ajoutons deux phases de traitements supplémentaires aux données, soit l'identification du genre des auteurs et des remerciés à l'aide d'un algorithme associant un genre aux prénoms, et la compilation de quatre indicateurs bibliométriques visant à caractériser le statut académique des remerciés (le nombre d'articles signés, le nombre total de citations obtenues par ces articles et normalisé selon la discipline, l'âge académique des remerciés et la proportion des articles pour laquelle un remercié occupe une position d'auteur principal).

En parallèle avec ces phases de traitements sur les entités nommées, nous procédons à l'extraction des syntagmes nominaux. Encore une fois, des phases de traitement et de nettoyage supplémentaires doivent être appliquées aux données extraites afin d'en obtenir des unités significantes en regard de nos questions de recherche. Pour répondre à la QR3, notre analyse des syntagmes nominaux doit permettre d'aller au-delà de la simple liste des syntagmes nominaux les plus fréquents. En effet, les syntagmes nominaux ayant le plus d'occurrences dans l'ensemble du corpus ne sont pas nécessairement les plus parlants pour décrire le contenu des remerciements et les variations disciplinaires. De plus, la fréquence des syntagmes nominaux varie énormément d'une discipline à l'autre. L'analyse de correspondance permet ainsi de faire ressortir les termes les plus significatifs pour chaque

discipline et de répondre à la QR3. Finalement, l'analyse qualitative du contexte d'utilisation des syntagmes nominaux identifiés par l'analyse de correspondance permet de répondre à la QR4. Ainsi, l'analyse qualitative de contenu permet de contextualiser les résultats de l'analyse quantitative et d'approfondir les questions soulevées par cette analyse, notamment en ce qui concerne l'utilisation de phrases et déclarations préformulées. Pour ce faire, l'échantillon des syntagmes nominaux les plus fréquents est réduit en mots distincts et un échantillonnage par choix raisonné (basé sur les résultats quantitatifs) est effectué pour sélectionner les mots qui feront l'objet du codage qualitatif. L'analyse qualitative de contenu permet de décrire de façon détaillée et contextualisée les différentes utilisations des termes les plus fréquemment employés dans les remerciements.

2.5 Qualité de la recherche

Notre devis de recherche inclut plusieurs mesures mises en place afin d'assurer la qualité de la recherche. D'abord, le devis mixte, de par la complémentarité des méthodes quantitatives et qualitatives, permet de tirer profit des forces spécifiques à chacune des approches. En effet, les objectifs de recherche de niveaux descriptif et explicatif appellent des paliers multiples d'analyse et nécessitent une méthodologie à la fois quantitative pour obtenir une perspective macro (et donc généralisable) et qualitative pour obtenir une perspective micro (et contextualisée) (Pluye et al., 2012). Plusieurs critères permettent d'évaluer la qualité d'une recherche : la validité interne (qui désigne la valeur de vérité d'une recherche), la validité externe (qui vise l'applicabilité des résultats de la recherche), la fidélité (qui vise la neutralité de la recherche) ainsi que la constance des mesures utilisées (qui vise l'objectivité de la recherche).

2.5.1 Volet quantitatif

La validité interne, qui vise à assurer la véracité des résultats de la recherche (Fortin et Gagnon, 2016), est assurée dans notre projet par une bonne compréhension du phénomène étudié. En effet, l'objectif du critère de validité interne est d'assurer la valeur de vérité de la recherche, de s'assurer que les variables étudiées sont bien celles en causes dans le phénomène observé (Creswell, 2009). La validité est donc assurée en amont par notre revue systématique de la littérature scientifique sur les remerciements dans le contexte de la

communication savante (Desrochers, Paul-Hus et Pecoskie, 2017) qui nous procure une compréhension approfondie du phénomène étudié et des variables à prendre en compte ainsi qu'une connaissance approfondie du corpus utilisé grâce à une analyse détaillée des données de remerciements indexées par WoS (Paul-Hus, Desrochers et Costas, 2016). En aval, les analyses bibliométriques des caractéristiques des individus remerciés permettent d'approfondir l'analyse de l'attribution du crédit scientifique effectuée dans un premier temps, en ajoutant des variables pertinentes à la compréhension des pratiques d'attribution du crédit, soit le genre, le statut académique et la discipline des individus remerciés. De plus, la phase d'analyse qualitative des syntagmes nominaux contenus dans les textes de remerciements vise justement à assurer la validité des résultats obtenus par l'analyse de correspondance grâce à la contextualisation et au retour aux données brutes.

Le critère de validité externe renvoie quant à lui à la notion de généralisabilité ou la transférabilité des résultats de la recherche. La population accessible de remerciements se limite aux articles scientifiques indexés par WoS depuis 2008. De 2008 à 2014, seuls les remerciements des articles du *Science Citation Index Expanded* sont collectés par WoS et c'est seulement depuis 2015 que la base de données donne accès aux remerciements provenant du *Social Sciences Citation Index*. Étant donné l'importance de la variable disciplinaire, l'échantillon analysé est restreint aux remerciements tirés d'articles publiés depuis 2015 afin d'assurer la généralisabilité des résultats à l'ensemble des disciplines des sciences naturelles, biomédicales et sociales.

Les critères de fidélité et d'objectivité peuvent être définis comme la stabilité et la constance des résultats à travers le temps, les lieux et les chercheurs (Creswell, 2009; Pickard, 2007). La fidélité et l'objectivité de notre recherche sont quant à elles assurées par une définition précise des concepts et variables étudiés, des mesures employées ainsi qu'une documentation détaillée de la recherche (Creswell, 2009).

2.5.2 Volet qualitatif

La crédibilité de la phase qualitative de l'analyse des syntagmes nominaux, qui tout comme la validité interne du volet quantitatif réfère à la valeur de vérité de la recherche, est assurée par une description dense et détaillée (*thick or rich description*) (Patton, 2015) des résultats.

Les critères de fiabilité et de confirmabilité, qui sont le pendant qualitatif des critères de fidélité et d'objectivité du volet quantitatif, sont quant à eux assurés par la réconciliation entre les codeurs (Campbell et al., 2013), ainsi que par la description explicite et détaillée des méthodes et procédures de codage (Creswell, 2009). La transférabilité, mesure de la validité externe de la recherche en contexte qualitatif, est assurée par la taille et la diversité disciplinaire de l'échantillon analysé qui permettent d'obtenir un échantillon conceptuellement représentatif de l'ensemble des cas possibles (Krippendorff, 2004, p. 84), ainsi que par la documentation détaillée du contexte de l'étude, de la population étudiée et des processus et méthode de codage mis en place (Creswell, 2009).

Chapitre 3. Résultats

3.1 La distribution du crédit scientifique

Article 1

The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements

Publié dans *Journal of Informetrics* (2017); 11(1) : 80-87.

<https://doi.org/10.1016/j.joi.2016.11.005>

Cet article est inclus dans la présente thèse avec l'autorisation des co-auteurs et de l'éditeur.

© Elsevier, 2017. Cette version du manuscrit est rendue disponible sous la licence CC-BY-NC-ND 4.0 <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Contributions des auteurs (selon la taxonomie CRediT¹⁶)

Adèle Paul-Hus: Conception de l'étude, méthodologie, investigation, analyse des données, visualisation, rédaction de l'ébauche initiale, révision critique du manuscrit et acquisition de soutien financier

Philippe Mongeon: Méthodologie, gestion des données, investigation, calcul/computation et révision critique du manuscrit

Maxime Sainte-Marie: Méthodologie, investigation, calcul/computation et révision critique du manuscrit

Vincent Larivière: Méthodologie, supervision, ressources, révision critique du manuscrit et acquisition de soutien financier

Contributions de l'auteure principale :

Dans le cadre de ce projet, j'ai d'abord défini les objectifs et questions de recherche spécifiques à cet article. J'ai ensuite procédé à la revue de la littérature. Sous la supervision de mon directeur de recherche Vincent Larivière, j'ai élaboré le devis méthodologique de l'étude. J'ai ensuite défini les critères d'inclusion et d'exclusion des données bibliométriques (incluant les remerciements) à récolter dans la base de données Web of Science pour pouvoir répondre à mes questions de recherche. Suivant ces critères, Philippe Mongeon a élaboré des requêtes SQL pour récolter les données bibliométriques nécessaires (incluant les textes de remerciements). Maxime Sainte-Marie a ensuite implémenté un algorithme permettant d'extraire les noms propres des chaînes de textes de remerciements. Sous ma supervision, Maxime Sainte-Marie a procédé à plusieurs itérations et corrections de l'algorithme afin qu'il réponde véritablement aux critères que j'avais établis. Puis, j'ai fait le nettoyage manuel final des noms propres extraits des remerciements. J'ai ensuite fait l'analyse des données, les visualisations et l'interprétation des résultats, qui ont été révisés par Philippe Mongeon et Vincent Larivière. J'ai finalement rédigé l'ébauche originale de l'article, qui a été révisée et commentée par mes trois co-auteurs.

¹⁶ <https://casrai.org/credit/>

The sum of it all: revealing collaboration patterns by combining authorship and acknowledgements

Adèle Paul-Hus¹, Philippe Mongeon¹, Maxime Sainte-Marie¹, Vincent Larivière²

¹adele.paul-hus@umontreal.ca, philippe.mongeon@umontreal.ca,
maxime.sainte-marie@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal
PO Box 6128, Downtown Station, Montreal, Quebec, H3C 3J7 (Canada)

² vincent.lariviere@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal
PO Box 6128, Downtown Station, Montreal, Quebec, H3C 3J7 (Canada) and Observatoire
des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la
Science et la Technologie (CIRST), Université du Québec à Montréal, PO Box 8888,
Downtown Station, H3C 3P8 Montréal, Qc. (Canada)

Abstract

Acknowledgments are one of many conventions by which researchers publicly bestow recognition towards individuals, organizations and institutions that contributed in some way to the work that led to publication. Combining data on both co-authors and acknowledged individuals, the present study analyses disciplinary differences in researchers' credit attribution practices in collaborative context. Our results show that the important differences traditionally observed between disciplines in terms of team size are greatly reduced when acknowledgees are taken into account. Broadening the measurement of collaboration beyond co-authorship by including individuals credited in the acknowledgements allows for an assessment of collaboration practices and team work that might be closer to the reality of contemporary research, especially in the social sciences and humanities.

Keywords: collaboration, co-authorship, acknowledgements, credit attribution

Introduction

Acknowledgments are one of many conventions by which researchers give credit and publicly share gratitude and recognition towards individuals, organizations and institutions that contributed to the work that led to publication. Although they could be perceived as the “scholar’s courtesy” (Cronin, 1995), acknowledgements convey rich information that can shed light on researchers’ collaborative activities that cannot be revealed by analysing co-authorship. In that sense, acknowledgements can be conceived as markers of symbolic capital (Bourdieu, 1975) that complements authorship, and have been included as a component of the “reward triangle” alongside authorships and citations (Cronin & Weaver-Wozniak, 1993). In most natural and biomedical sciences disciplines, teamwork constitutes the norm rather than the exception (Cronin, 2004; Wuchty, Jones & Uzzi, 2007). Henriksen (2016) and Larivière, Gingras and Archambault (2006) have further shown that the rise in research collaborations also extends to most social sciences disciplines, in terms of average number of authors, share of co-authored articles, as well as international collaboration. However, these results, as most bibliometric investigations of collaboration, are limited to formal collaborations as measured by co-authorship. Indeed, as highlighted by Katz and Martin (1997), many instances of collaboration do not lead to co-authorship, while indirect interactions between researchers might actually do. This has led them to conclude that co-authorship is a “rather imperfect or partial indicator of research collaboration between individuals ” (Katz and Martin, 1997, p. 11).

Laudel (2002) also challenged that traditional bibliometric practice of using co-authorships as a proxy for research collaboration and identified six types of research collaborations associated to distinct patterns of rewards. Based on interviews with researchers and an analysis of 133 publications, Laudel (2002) showed that, while some contributions were associated with authorship, one third of all contributions analysed were only rewarded by acknowledgements and about half of contributions were not associated to any public recognition and were thus invisible in formal communication channels. More recently, Ponomariov and Boardman (2016) surveyed academic researchers on their relationship with their collaborators and showed that in many instances, collaboration does

not entail co-authorship, a finding which leads the authors to suggest using data that go beyond co-authorship when studying collaboration.

Types of contributions that get rewarded by authorship vary in their nature but also by field, discipline and specific teamwork culture (Larivière et al., 2016). High Energy Physics (HEP) represents a telling example of discipline-specific authorship attribution practices, with projects typically involving thousands of individuals and almost as many institutions. In that context, specific guidelines govern authorship. For instance, all members the project are included in a standard author list and each paper emerging from the project will be alphabetically co-authored by all those on the list (Biagioli, 2003, Birnholtz, 2006). In 2015, a new record for the largest number of authors on a single research article has been set by a HEP publication, co-signed by more 5,000 individuals (Castelvecchi, 2015). A contrasting example is found in medical research, where the notion of authorship is closely linked to responsibility and accountability. Given the dangerous consequences associated to fraud in those disciplines and its rising co-authorship rates, the International Committee of Medical Journal Editors (ICMJE) published, for the first time in 1988, the *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. Updated in 2015, the ICMJE criteria recommends that authorship be based on:

- substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work, AND
- drafting the work or revising it critically for important intellectual content, AND
- final approval of the version to be published, AND
- agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. (ICMJE, 2015, p.2).

Moreover, “contributors who meet fewer than all 4 of the above criteria for authorship should not be listed as authors, but they should be acknowledged” (ICMJE, 2015, p.3). This suggests that when the ICMJE guidelines are strictly followed, many contributions

may be insufficient to warrant authorship and should rather be rewarded by an acknowledgement only.

Contrasting with the ICMJE authorship guidelines, Rennie, Yank and Emmanuel (1997) proposed that the notion of author is “outmoded”, and that it cannot appropriately account for credit and responsibility in multi-authors publications¹⁷. They proposed a system where the notion of contributorship would replace the notion of authorship. The main objective of their proposition was to ensure more equitable and reliable credit and responsibility attribution practices, where all collaborators would systematically disclose their specific contributions. This radical alternative would eliminate “the artificial distinction, mostly of a social nature, between authors and non-author contributors—that is, between ‘authors’ and ‘acknowledgees’” (Rennie, Yank & Emmanuel, 1997, p. 584). Almost two decades later, the contributorship model, as envisioned originally, has not been implemented anywhere. However, many journals, mostly in the medical field, now include contribution statements (e.g. Nature, PNAS, the British Medical Journal and the PLOS series of journals).

Notwithstanding their potential to reveal often invisible contributions to research, the current format of acknowledgements limits their use. As highlighted by McCain (1991), “[t]he format of acknowledgment varies from field to field and from journal to journal. As noted, persons and institutional sources may be listed in the methods and materials section of an article or explicitly thanked in an acknowledgements section” (p.506). This lack of standardization—highlighted by many researchers (e.g. Cronin, 1995; Giles & Council, 2004; Mackintosh, 1972; Parker, Paisley & Garrett, 1967)—has contributed to the ambiguous reputation of acknowledgements in the scientific community. This unstandardized space of thanking leads to very heterogeneous testimonies of gratitude, and

¹⁷ It should be noted that the ICMJE authorship guidelines were slightly different at the time of Rennie, Yank and Emmanuel proposal and consisted of the following: “Authorship credit should be based only on substantial contributions to (a) conception and design, or analysis and interpretation of data; and to (b) drafting the article or revising it critically for important intellectual content; and on (c) final approval of the version to be published. Conditions (a), (b), and (c) must all be met.” (ICMJE, 1997:311).

contributions getting rewarded by an acknowledgement can be even more heterogeneous than those leading to authorship. On the one hand, Cronin, et al.'s (1993) classification of acknowledgements ranges from conceptual and intellectual contributions to provision of financial support, access to data and materials, technical assistance and manuscript preparation; these same types of contributions can be sufficient to warrant authorship in certain contexts. On the other hand, contributions that could be perceived as trivial or hardly relevant in light of most authorship criteria can lead to authorship in some instances. For example, in a recent article, one of the authors' contribution consisted in driving the car during the data collection process¹⁸. Similarly, several studies have reported a high prevalence of honorific (or gift) authorship, where researchers who did not make a substantial contribution to a work (or did not contribute at all in some cases), but are included in the author list (e.g. Flanagin et al., 1998; Marušić, Bošnjak & Jerončić, 2011; Wislar et al., 2011). This diversity of disciplinary, but also individual, authorship attribution practices—some of which being more inclusive than others—can induce artificial distinctions in team size and collaboration, as measured by co-authorship. This highlights the need for new methods that transcend such limitations and provide a more accurate assessment of collaboration in research. This paper attempts to do so by combining acknowledgements and authorship data in order to explore the potential of acknowledgements to reveal collaboration practices going beyond authorship analysis.

Since Giles and Council (2004) pioneering analysis of more than 180,000 acknowledgements found in computer science papers, no large-scale investigation of individuals acknowledged were performed. The present study aims at filling this gap, by analysing more than 1,000,000 scholarly documents containing acknowledgements. In order to extend the notion of collaboration beyond authorship, we analyse formal and informal collaborations. Our unit of analysis thus includes broader types of contributions that are credited by authorship in certain contexts and not in others, when they are made

¹⁸ <https://twitter.com/igoodfel/status/732927411650744320>

visible in acknowledgements. The individuals involved in a research paper, and credited for it (either formally by authorship or informally by a mention in the acknowledgements) will thus for the purpose of this study be designated as contributors. Collaboration is hence defined inclusively and operationalized as papers having at least two contributors credited on a paper, mentioned either in the byline or the acknowledgements text. The objective of this study is to compare the credit attribution practices of researchers in natural, medical and social sciences. More specifically, we aim at answering the following research questions:

- How many contributors are credited on scholarly publications and how does this vary by discipline?
- What share of contributors is credited as authors and what share is credited as acknowledgees and how does it vary by discipline?
- How does the number of acknowledgees vary as a function of the number of authors signing a scholarly publication and how does this relationship vary by discipline?

Material and methods

Data

Data for this study were drawn from Web of Science (WoS) Science Citation Index Expanded (SCI-E) and Social Sciences Citation Index (SSCI), which include acknowledgement data. These acknowledgement data are structured in three fields: the 'Funding Text' (FT), 'Funding Agency' (FO) and 'Grant Number' (FG). FT is the full text of acknowledgements, as it appears in the paper from which it is retrieved. However, as shown by Paul-Hus, Desrochers and Costas (2016), acknowledgements texts are collected and indexed by WoS only if they include funding information. The sum of contributors (authors and acknowledgees) here analysed is consequently limited to publications where a source of funding is acknowledged.

Although WoS started the collection of acknowledgements data in August 2008 for SCI-E articles and reviews, the collection of these data only started in 2015 for SSCI publications (Paul-Hus, Desrochers & Costas, 2016). A dataset of acknowledgement texts

was derived from all 2015 articles and reviews from all disciplines covered by SCI-E and SSCI: Biology, Biomedical Research, Chemistry, Clinical Medicine, Earth and Space, Engineering and Technology, Health, Mathematics, Physics, Professional Fields, Psychology and Social Sciences. The dataset includes a total of 1,009,411 papers with acknowledgements texts, which corresponds to 67.1% of all articles and reviews published in 2015 (Table 1). Discipline assignment was done using the NSF field classification of journals (National Science Foundation, 2006); since the NSF classification assigns only one discipline specialty to each journal, this prevents the double counting of papers.

Analysis

In order to obtain the number of individuals acknowledged per paper, the Stanford Named Entity Recognizer (NER) (Finkel, Grenager & Manning; 2005) module of the Natural Language ToolKit (NLTK) (Bird, 2009) was used on each string of acknowledgment text retrieved from the FT field. Application of the Stanford NER algorithm and selection of all named entities tagged as 'person', led to the extraction of 817,125 distinct person names.

The list of named entities extracted from the acknowledgements was then cleaned in order to eliminate non-human entities. This was done in several steps: incomplete names were first removed from the list (entities containing only a first or last name, or only initials), retaining only entities composed of at least one initial and one last name. In order to remove names not designating actual persons, the list was compared to the list of last names of all authors appearing on publications from 1900 to 2016 in WoS indexes, which includes 2,649,212 distinct last names. This WoS authors list was thus used here as a person-name benchmark list. Entities with no match in this list were considered as not referring to actual individuals and were removed. A further manual cleaning step was done to remove all remaining names that did not refer to individual persons such as grant, foundation, organization and institution names. Examples of such names removed by manual cleaning include: Frederick Banting (grant), Marie Curie (grant and foundation), Boehringer Ingelheim (organization) and Instituto de Salud Carlos III (institution). Finally, acknowledgements often contain the name(s) of the author(s) signing the paper from which the acknowledgements were retrieved. When the name(s) extracted from the acknowledgements of a paper X matched the name of one of the author appearing in the

byline of that paper X (using the first initial and the last name), this name was removed from the acknowledgees list for that specific paper, such as in the example below:

Paper X

Authors: J. Zhang, X. Feng and Y. Xu

Acknowledgements text: “Jinsong Zhang, Xiao Feng, and Yong Xu contributed equally to this work [...]”

The final list of acknowledgments extracted names includes 810,525 distinct names appearing in 362,767 papers.

Results

Table 1 presents, by discipline, the number of 2015 articles and reviews, the number (and percentage) of those with acknowledgements, and the number (and percentage) of those that contain at least one acknowledged individual. The proportion of papers in which the acknowledgements include the mention of individuals ranges from 12% (Professional Fields) to 45% (Earth and Space), with an average of 24% all disciplines considered. However, the size of disciplines, in terms of absolute number of papers, varies greatly. Indeed, only 17% of Clinical Medicine papers and 31% of Biomedical Research papers include acknowledgements of specific persons, yet they have the highest number of papers, with respectively 67,019 and 59,142 papers that include the mention of specific individuals in their acknowledgements. The following analysis will focus on the subset of papers that includes funding acknowledgements indexed in WoS, for a total of 1,009,411 papers.

Table 1. – Number of 2015 papers, number (and percentage) of papers with acknowledgements, and number (and percentage) of papers with acknowledgees

Discipline	All papers N	Papers with acknowledgements		Papers with acknowledgees		
		N	%	N	% _{ack.}	% _{total}
Earth & Space	92,238	72,922	79.1	41,633	57.1	45.1
Biology	105,279	76,281	72.5	43,365	56.8	41.2
Biomedical Research	189,066	158,067	83.6	59,142	37.4	31.3
Physics	124,556	95,676	76.8	35,063	36.6	28.2
Psychology	31,286	15,085	48.2	7,736	51.3	24.7
Chemistry	151,947	123,806	81.5	36,583	29.5	24.1
Social Sciences	50,420	16,972	33.7	9,291	54.7	18.4
Engineering and Technology	241,124	165,590	68.7	43,899	26.5	18.2
Clinical Medicine	389,311	218,367	56.1	67,019	30.7	17.2
Mathematics	49,997	35,390	70.8	8,314	23.5	16.6
Health	37,309	18,703	50.1	5,651	30.2	15.1
Professional Fields	41,015	12,552	30.6	5,071	40.4	12.4
Total	1,503,548	1,009,411	67.1	362,767	35.9	24.1

Figure 1 presents the cumulative distribution of papers (with acknowledgements) as a function of the number of author(s) they *at least* contain, by discipline. Three groups of disciplines can be distinguished. The first group, which includes Social Sciences, Mathematics and Professional Fields, has the lowest number of authors per paper, with more than 85% of papers having four authors or less, and with proportions of single authored papers ranging from almost 15% (Professional Fields) to more than 25% (Social Sciences). The group composed of Biomedical Research and Clinical Medicine has the highest levels of co-authorship, with 90% of papers having 12 authors or less, and less than 2% having only one author. The remaining disciplines, mostly from the natural sciences, can be found between those two groups.

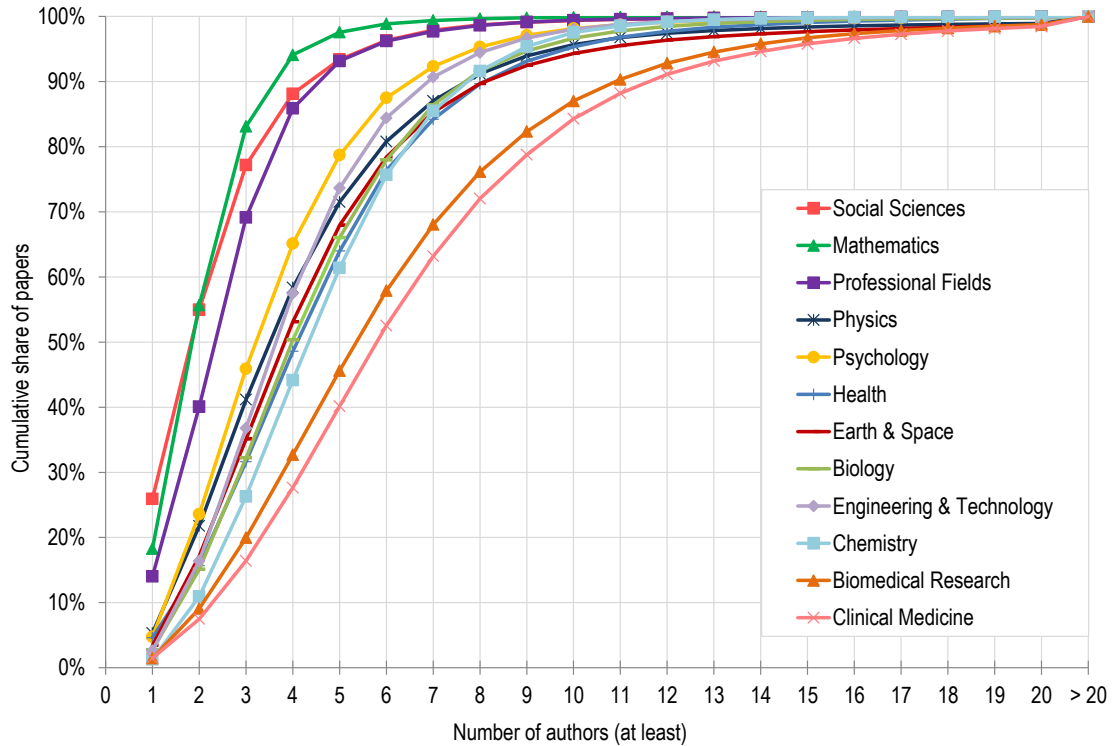


Figure 1. – Cumulative distribution of papers (with acknowledgements) (%), as a function of numbers of authors

Figure 2 presents, for all disciplines combined, the distribution of papers (on a log scale) by number of authors (a) and acknowledgees (b), highlighting the skewness of the data. In both cases, the figure clearly shows that the highest proportion of papers is signed by less than 15 authors and acknowledges less than 10 persons. Moreover, both graphs present long-tailed distributions where extreme values of authors and acknowledgees per paper are highly dispersed. Given these data characteristics, the median would generally appear as a more robust measure to describe such distributions. However, because of the high proportion of papers that bears no acknowledgement to specific individuals, the median value of acknowledgees per paper is zero in most disciplines. In this context, the mean value of authors and acknowledgees per paper is deemed the most appropriate measure to describe in a meaningful way the dataset at hand.

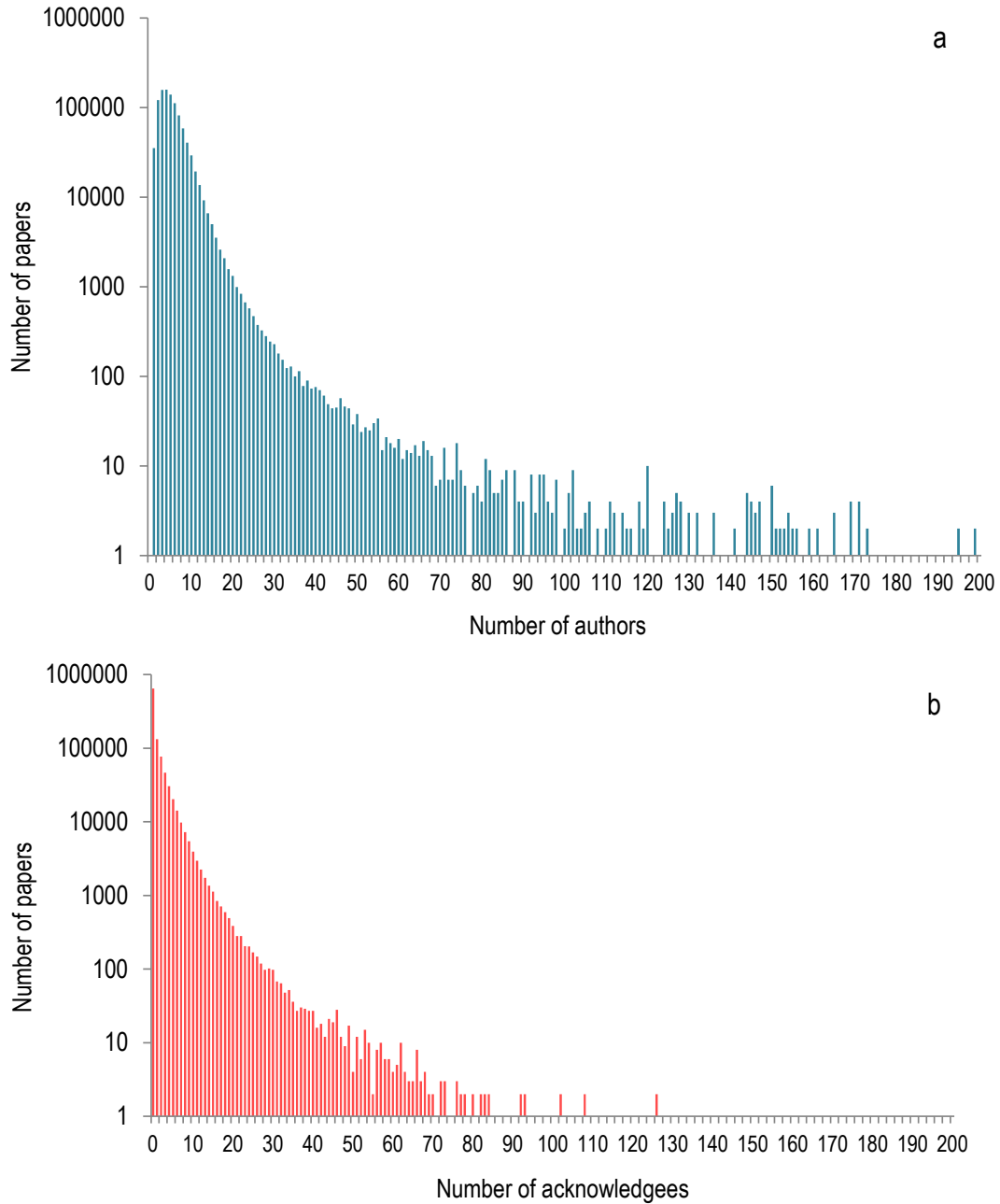
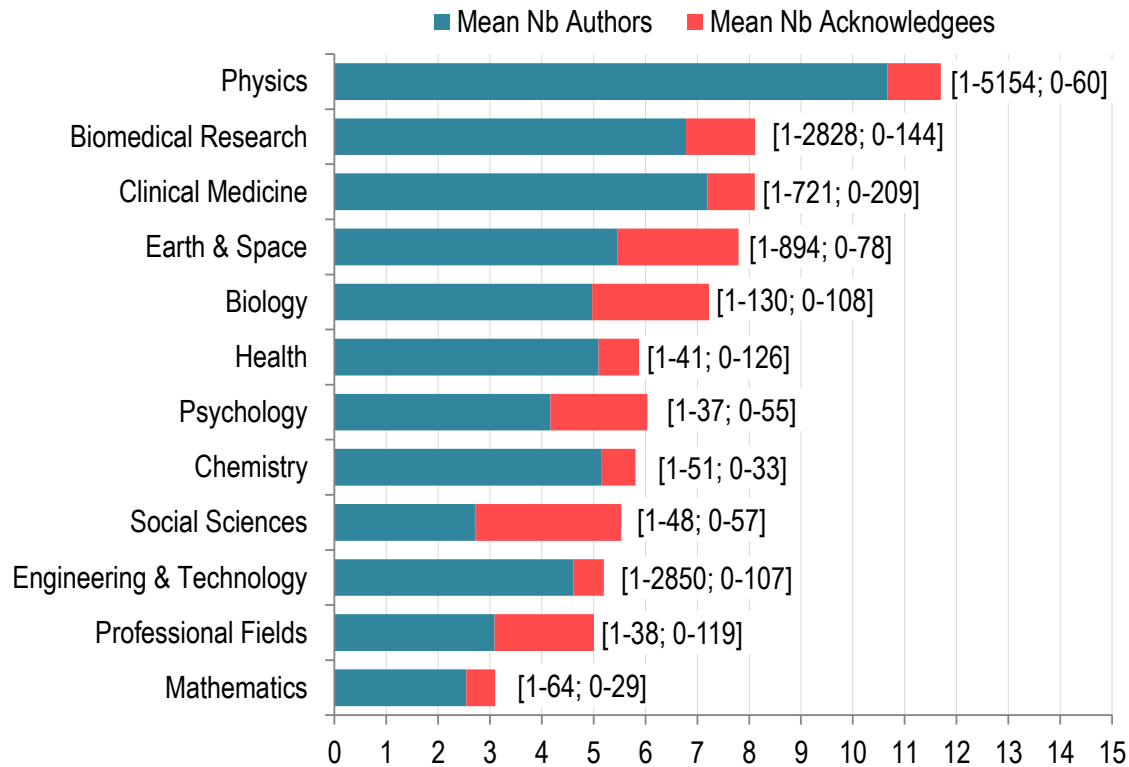


Figure 2. – Distribution of papers by number of authors (a) and acknowledgees (b)

Figure 3 presents the mean number of acknowledgees and authors per paper, for the 12 disciplines. The mean number of contributors (the sum of authors and acknowledgees per paper) ranges between 3.1 (Mathematics) and 11.7 (Physics). Figure 3 also displays the variability of the number of contributors per paper with the minimum and maximum number of authors and number of acknowledgees per paper shown in square brackets for

each discipline. Physics is by far the discipline where the mean number of contributors is the highest. However, this high number of contributors is mostly attributable to authors, since Physics papers are on average signed by more than 10 authors (10.7) but only acknowledge one person on average. Biomedical Research and Clinical Medicine both have on average more than eight contributors per paper and are similar to Physics in terms of proportion of authors and acknowledgees. In the natural sciences, Earth and Space and Biology differentiate from the other disciplines with about one third of the contributors being acknowledgees. As expected Social Sciences and Professional Fields have low average numbers of authors but, in turn, their average number of acknowledgees is similar to that of Earth and Space and Biology. Consequently, these two disciplines stand out with their high acknowledgee/author ratio. In Social Sciences, the mean number of acknowledgees per paper (2.8) even exceeds the mean number of authors per paper (2.7).

As a result, disciplines from the social sciences (Health, Psychology, Social Sciences and Professional Fields), which traditionally exhibit much lower level of collaboration when solely considering co-authorship (Larivière, Gingras & Archambault, 2006; Wuchty, Jones & Uzzi, 2007) are displaying mean numbers of contributors that are comparable to what is observed in Chemistry and Engineering and Technology, both natural sciences. Overall, Figure 3 shows that the important differences traditionally observed between all disciplines in terms of team size as measured by co-authorship ($M = 4.98$, $SD = 2.15$, $RSD = 43\%$) are greatly reduced when acknowledgees are taken into account ($M = 6.68$, $SD = 2.08$, $RSD = 31\%$). This result might indicate that, when considering team size, disciplinary patterns might reflect differences in authorship attribution practices more than actual collaboration practices. In that sense, it suggests that disciplinary differences usually observed in the collaboration level might be amplified by the way we measure collaboration.



Note: the numbers in the brackets represent the range of the number of author (left) and of the number of acknowledges (right)

Figure 3. – Mean number of authors and acknowledges, by discipline

For all disciplines, the mean number of acknowledges decreases or remains stable as the number of authors per paper increases (Figure 4). Moreover, in all disciplines, with the exception of Clinical Medicine and Mathematics, the mean number of acknowledges is the highest for papers signed by a single author. In Biology, Social Sciences and Professional Fields, a clear decreasing trend is observed in terms of average number of acknowledges, which implies that as more contributors get credited as authors on collaborative papers, less get acknowledged. For all other disciplines, the mean number of acknowledges remains stable as the number of authors increases. These trends further support the idea that the lower mean number of authors per paper observed in some disciplines is partly due to less inclusive authorship attribution practices. Mathematics is the exception, standing out as having both the lowest mean number of authors and the lowest mean number of acknowledges, even for single authored papers.

It should be noted that the relation between the mean number of acknowledgees and the number of authors is presented for values between one and nine on the authors axis since most disciplines have their biggest share papers in that subset, ranging between 100% of papers in Mathematics that have nine authors or less, and 80% of papers in Clinical Medicine that have nine authors or less (see Figure 1 for the complete cumulative distribution). Beyond nine authors, in many disciplines, the number of papers is too small to allow for robust measures.

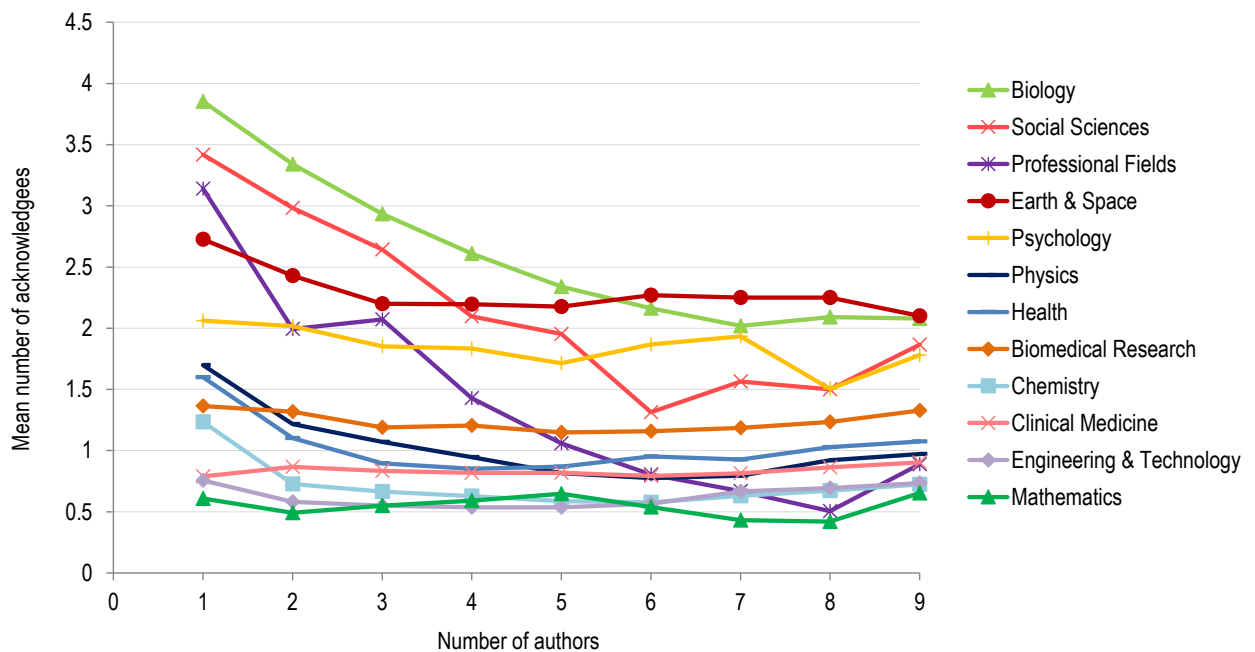


Figure 4. – Mean number of acknowledgees by number of authors

Discussion and conclusion

Over the last decades, scientific collaboration has been the focus of hundreds of bibliometric analyses. However, these analyses have almost all relied on co-authorship as an indicator of collaboration, an operationalization which has been shown to have limitations (Katz & Martin, 1997; Laudel, 2002; Lundberg et al., 2006). Broadening the measurement of collaboration to include individuals mentioned in the acknowledgements of scholarly publications allows for an assessment of collaboration practices that might be closer to the reality of contemporary research. Our results show that disciplinary

differences in collaborative activities are actually much less important, as scholars in the social sciences are collaborating much more than what co-authorship alone suggests. This also confirms that the lone scholar has become an endangered species in most disciplines, including the social sciences. In fact, our data shows that in 40% of publications signed by only one author (restricted to the subset of papers where acknowledgements have been indexed), single authors are not alone since they acknowledge specific individuals who contributed to their research.

One limitation of this study is related to the data source. As mentioned in the methods section, acknowledgements are collected and indexed in WoS only when they contain funding information, thus creating a bias toward funded research projects. Moreover, our analyses are restricted to the Science Citations Index Expanded (SCI-E) and the Social Sciences Citations Index (SSCI) and do not cover publications from Arts and Humanities, which are considered as disciplines with lower collaborative practices (Larivière, Gingras & Archambault, 2006). Yet, WoS still constitutes the most comprehensive source for acknowledgements data. As shown in Table 1, acknowledgements are not evenly distributed among disciplines. Since acknowledgements are not collected systematically, we cannot conclude that acknowledgements are less frequent in those disciplines that exhibit lower shares of funding acknowledgements. We cannot exclude the possibility that a certain number of papers, not analysed in the present study, includes acknowledgements without specifically mentioning funding source. However, our dataset still represents more than two thirds of all articles and reviews published in 2015, a sample size large enough to ensure the robustness of our findings.

Another potential limitation is related to the nature of contributions acknowledged. While most acknowledgements are made to individuals who have actively contributed to the work that led to a publication, there may also be cases where authors acknowledge their partners and family for their support, or other types of “non-scientific” contributions. Nevertheless, our results provide strong evidence of the existence of disciplinary differences not only in terms of authorship practices, but also in terms of acknowledgement practices. Furthermore, these practices appear to influence each other, highlighting the necessity of taking both into account when measuring collaboration in research. However,

by assessing the number of individuals involved in the production of scientific publications, we do not weight the value of contributions leading to acknowledgement as equivalent to the ones leading to authorship but rather aim at taking into account the high variability of authorship and acknowledgement practices. In the end, our results suggest that disciplinary differences traditionally observed in terms of team sizes and collaborative activity might be, at least in part, an artifact of the indicator we use to measure collaboration and not a truthful reflection of team size variation between disciplines.

PLOS journals recently introduced a new taxonomy of contributions providing standardized and fine-grained information that “makes transparent who participated and the roles they played” (Atkins, 2016). This new taxonomy has the objective “to know and unambiguously credit [those] who participated in the work being published and forms the base for plans to eventually provide credit to all participants in the research outputs ecosystem” (Atkins, 2016). In order to achieve such unambiguous credit attribution, standardized authorship and acknowledgements criteria would need to be applied and followed rigorously and uniformly across disciplines which, as we have demonstrated in this study, is far from being the case. Moreover, many studies have shown the risks of introducing bias in reporting contributions when authorship and contributorship statements are restricted to a pre-determined taxonomy (Bates et al., 2004; Ivaniš et al., 2008; Marušić et al., 2006). The reliability of such disclosures of contributions tends to be affected by one’s own autobiographical memory and perceived value of contribution (Ilakovac et al., 2006; Ivaniš et al., 2011). Nevertheless, the new credit taxonomy from PLOS constitutes a further step towards transparency and accountability for all team members involved in research projects. In turn, a standardized system providing the description of all participants’ contributions could lead to a more equitable distribution of credit and reward—which could also contribute to provide a more accurate portrait of what contemporary research involves in terms of humans and materials resources, especially in the social sciences and humanities.

Acknowledgement

This research was supported by the Social Sciences and Humanities Research Council of Canada: Joseph-Armand Bombardier CGS Doctoral Scholarships (Paul-Hus, Mongeon); Insight Development Grant (Larivière).

References

- Atkins, H. (2016). *Author Credit: PLOS and CRediT Update*. The Official PLOS Blog. Retrieved from <http://blogs.plos.org/plos/2016/07/author-credit-plos-and-credit-update/>
- Bates, T., Anić, A., Marusić, M. & Marusić, A. (2004). Authorship criteria and disclosure of contributions: comparison of 3 general medical journals with different author contribution forms. *JAMA*, 292(1), 86–88. <https://doi.org/10.1001/jama.292.1.86>
- Biagioli, M. (2003). Rights or rewards? Changing frameworks of scientific authorship. In P. Galison & M. Biagioli (Eds.), *Scientific authorship : credit and intellectual property in science* (pp. 255–279). New York, NY: Routledge.
- Bird, S., Klein, E. & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*. Sebastopol, CA: O'Reilly Media.
- Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, 57(13), 1758–1770. <http://doi.org/10.1002/asi.20380>
- Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, 14(6), 19-47.
- Castelvecchi, D. (2015). Physics paper sets record with more than 5,000 authors. *Nature News*. <http://doi.org/10.1038/nature.2015.17567>
- Cronin, B. (2004). Bowling alone together: Academic writing as distributed cognition. *Journal of the American Society for Information Science and Technology*, 55(6), 557-560. <http://doi.org/10.1002/asi.10406>

- Cronin, B. (1995). *The scholar's courtesy : the role of acknowledgement in the primary communication process*. London: Taylor Graham.
- Cronin, B. & Weaver-Wozniak, S. (1993). Online access to acknowledgements. In M. E. Williams (Ed), *Proceedings of the Fourteenth National Online Meeting 1993*. (p.93-98). Medford: NY, Learned Information, Inc.
- Cronin, B., McKenzie, G., Rubio, L. & Weaver-Wozniak, S. (1993). Accounting for influence: Acknowledgments in contemporary sociology. *Journal of the American Society for Information Science*, 44(7), 406–412.
- Finkel, J.R., Grenager T. & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In K. Knight, H.T. Ng and K. Oflazer (Eds), *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (ACL 2005, pp. 363-370). Ann Arbor, MI, Association for Computational Linguistics.
- Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D. & Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *JAMA*, 280(3), 222–224. <http://doi.org/10.1001/jama.280.3.222>
- Giles, C. L. & Councill, I. G. (2004). Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17599–604. <http://doi.org/10.1073/pnas.0407743101>
- Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455-476. <http://doi.org/10.1007/s11192-016-1849-x>
- Illakovac, V., Fister, K., Marušić, M. & Marušić, A. (2007). Reliability of disclosure forms of authors' contributions. *Canadian Medical Association Journal*, 176(1), 41–46. <https://doi.org/10.1503/cmaj.060687>

- International Committee of Medical Journal Editors. (1997). Uniform requirements for manuscripts submitted to biomedical journals. *New England Journal of Medicine*, 336(4), 309–315.
- International Committee of Medical Journal Editors. (2015, December). Recommendations. Retrieved from <http://www.icmje.org/icmje-recommendations.pdf>
- Ivaniš, A., Hren, D., Marušić, M. & Marušić, A. (2011). Less work, less respect: Authors' perceived importance of research contributions and their declared contributions to research articles. *PLOS ONE*, 6(6), e20206.
<https://doi.org/10.1371/journal.pone.0020206>
- Ivaniš, A., Hren, D., Sambunjak, D., Marušić, M. & Marušić, A. (2008). Quantification of authors' contributions and eligibility for authorship: Randomized study in a general medical journal. *Journal of General Internal Medicine*, 23(9), 1303–1310.
<https://doi.org/10.1007/s11606-008-0599-8>
- Katz, J. S. & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1-18.
- Larivière, V., Gingras, Y. & Archambault, É. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*, 68(3): 519-533.
- Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A. & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417–435. <https://doi.org/10.1177/0306312716650046>
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1), 3-15. <http://doi.org/10.3152/147154402781776961>
- Lundberg, J., Tomson, G., Lundkvist, I., Skar, J. & Brommels, M. (2006). Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics*, 69(3), 575–589.
<https://doi.org/10.1007/s11192-006-0170-5>

- Mackintosh, S. H. (1972). *Acknowledgment patterns in sociology* (Ph.D. Thesis). University of Oregon, Michigan.
- Marušić, A., Bošnjak, L. & Jerončić, A. (2011). A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS ONE*, 6(9), e23477. <http://doi.org/10.1371/journal.pone.0023477>
- Marušić, A., Bates, T., Anić, A. & Marušić, M. (2006). How the structure of contribution disclosure statements affects validity of authorship: a randomized study in a general medical journal. *Current Medical Research and Opinion*, 22(6), 1035–1044. <https://doi.org/10.1185/030079906X104885>
- McCain, K. W. (1991). Communication, competition, and secrecy: The production and dissemination of research-related information in genetics. *Science, Technology & Human Values*, 16(4), 491–516. <http://doi.org/10.1177/016224399101600404>
- National Science Foundation. (2006). Science and engineering indicators. Chapter 5: Academic research and development. Data and terminology. Retrieved from <http://www.nsf.gov/statistics/seind06/c5/c5s3.htm#sb1>
- Parker, E. B., Paisley, W. J. & Garrett, R. L. (1967). Bibliographic citations as unobtrusive measures of scientific communication. In *National Science Foundation Grant Gn-434*. Stanford University. Institute for Communication Research.
- Paul-Hus, A., Desrochers, N. & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, 108(1), 167-182. <http://doi.org/10.1007/s11192-016-1953-y>
- Ponomariov, B. & Boardman, C. (2016). What is co-authorship? *Scientometrics*, 109(3), 1939-1963. <https://doi.org/10.1007/s11192-016-2127-7>
- Rennie, D., Yank, V. & Emanuel, L. (1997). When authorship fails: A Proposal to make contributors accountable. *JAMA*, 278(7), 579–585. <http://doi.org/10.1001/jama.1997.03550070071041>

Wislar, J. S., Flanagan, A., Fontanarosa, P. B. & DeAngelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ*, *343*, d6128. <http://doi.org/10.1136/bmj.d6128>

Wuchty, S., Jones, B. F. & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316* (5827), 1036–1039. <http://doi.org/10.1126/science.1136099>

3.2 Les individus remerciés

Article 2

Who are the acknowledgees? An analysis of gender and academic status

Accepté pour publication dans *Quantitative Science Studies* (2020).

https://doi.org/10.1162/qss_a_00036

Cet article est inclus dans la présente thèse avec l'autorisation des co-auteurs.

© 2020 Paul-Hus, Mongeon, Sainte-Marie et Larivière

Contributions des auteurs (selon la taxonomie CRediT¹⁹)

Adèle Paul-Hus: Conception de l'étude, méthodologie, investigation, analyse des données, visualisation, rédaction de l'ébauche initiale, révision critique du manuscrit et acquisition de soutien financier

Philippe Mongeon: Méthodologie, gestion des données, investigation, calcul/computation et révision critique du manuscrit

Maxime Sainte-Marie: Méthodologie, investigation, calcul/computation, visualisation et révision critique du manuscrit

Vincent Larivière: Supervision, ressources, révision critique du manuscrit et acquisition de soutien financier

Contributions de l'auteure principale: Comme pour l'article 1, j'ai d'abord défini les objectifs et questions de recherche spécifiques à cet article. J'ai ensuite procédé à la revue de la littérature ciblée pour ces questions de recherche. J'ai élaboré le devis méthodologique de l'étude, qui a été approuvé par mon directeur de recherche, Vincent Larivière. J'ai ensuite défini les critères d'inclusion et d'exclusion des données bibliométriques (incluant les remerciements) à récolter dans la base de données Web of Science pour pouvoir répondre à mes questions de recherche. Suivant ces critères, Philippe Mongeon a élaboré des requêtes SQL pour récolter les données bibliométriques nécessaires. Maxime Sainte-Marie a ensuite mis à jour et implémenté un algorithme permettant d'extraire les noms propres des chaînes de textes de remerciements; il s'agissait ici d'une mise à jour de la méthode utilisée pour l'article 1 et c'est sous ma direction qu'il a effectué les changements. Puis, j'ai fait le nettoyage manuel final des noms propres extraits des remerciements. Philippe Mongeon a ensuite implémenté un algorithme d'attribution du genre en fonction du prénom afin d'identifier le genre des auteurs et des individus remerciés. Puis, suivant mes indications, Philippe Mongeon a compilé les indicateurs bibliométriques de statut académique des auteurs et des remerciés. J'ai ensuite fait l'analyse des données, les visualisations (à l'exception de la Figure 2 qui a été réalisée

¹⁹ <https://casrai.org/credit/>

par Maxime Sainte-Marie, sous ma direction) et l'interprétation des résultats, qui ont été révisés par Philippe Mongeon et Vincent Larivière. J'ai finalement rédigé l'ébauche originale de l'article, qui a été révisée et commentée par mes trois co-auteurs.

Who are the acknowledgees? An analysis of gender and academic status

Adèle Paul-Hus¹, Philippe Mongeon², Maxime Sainte-Marie² and Vincent Larivière³

¹ adele.paul-hus@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal, PO Box 6128, Downtown Station, Montreal, Quebec, H3C 3J7 (Canada) and Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), CP 8888, Succ. Centre-Ville, H3C 3P8 Montréal, Qc. (Canada)

² philippe.mongeon@ps.au.dk ; msaintemarie@ps.au.dk

Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Aarhus, Denmark

³ vincent.lariviere@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal, PO Box 6128, Downtown Station, Montreal, Quebec, H3C 3J7 (Canada) and Université du Québec à Montréal, Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Observatoire des Sciences et des Technologies (OST), CP 8888, Succ. Centre-Ville, H3C 3P8 Montréal, Qc. (Canada)

Abstract

Acknowledgements found in scholarly papers allow for credit attribution of non-author contributors. As such, they are associated with a different kind of recognition than authorship. While several studies have shown that social factors affect authorship and citation practices, few analyses have been performed on acknowledgements. Based on 878,250 acknowledgees mentioned in 291,167 papers published between 2015 and 2017, this study analyses the gender and academic status of individuals named in the acknowledgements of scientific papers. Our results show that gender disparities generally found in authorship can be extended to acknowledgements, and that women are even more underrepresented in acknowledgements section than in authors' lists. Our findings also show that women acknowledge proportionally more women than men do. Regarding academic status, our results show that acknowledgees who have already published tend to have a higher position in the academic hierarchy compared to all of WoS authors. Taken together, these findings suggest that acknowledgement practices might be associated with academic status and gender.

Keywords: acknowledgement, authorship, scientific credit, gender, academic status

Introduction

Acknowledgements found in scientific papers are a public testimony of authors' gratitude and recognition that can reveal contributions of varied nature made by individuals, institutions and organizations. As such, acknowledgements have been positioned alongside authorship and citations as a form of scientific recognition in the "reward triangle" (Cronin, 1995). They also allow for the division of credit among authors and other contributors named in acknowledgements section. In this sense, acknowledgements can illuminate "sub-authorship collaboration" (Patel, 1973, p.81). In the reward system of science (Merton, 1973)—where authorship constitutes the main means to accumulate "symbolic capital" (Bourdieu, 1975)—a mention in the acknowledgement is not associated with the same kind of recognition as authorship. Moreover, given the hierarchical structure of the scientific community, it can be difficult to discern the reason justifying one's presence (or absence) in the authors' list since credit attribution can be difficult to disentangle from one's status within the hierarchy (Heffner, 1979). Credit attribution does not exclusively rely on the nature of contributions made and numerous other factors come into play, such as disciplinary context, structure of the project (Jabbehdari & Walsh 2017) and one's position in the academic hierarchy (Cole & Cole, 1973; Merton, 1973; Zuckerman, 1977). Among those factors, gender, seniority, and academic status have been shown to have an effect on inclusion in authorship lists (Haeussler & Sauerma, 2013; Larivière et al. 2016; Lissoni, Montobbio & Zirulia, 2013). A recent survey of 6,673 researchers provided evidence that discipline, academic rank, and gender were all affecting, to various degrees, authorship disagreements in research teams (Smith et al., 2019).

Non-author collaborators

Authorship criteria have been the subject of numerous discussions in the last decades (e.g., Marušić, Bošnjak & Jerončić, 2011; Sismondo, 2009; Wager, 2009; Wislar et al., 2011). However, collaborators who are not authors have received less attention. Shapin's seminal work (1989) has shown that the essential role played by technicians in the scientific development of the seventeenth century has been completely obliterated from the history of science, as their contributions were not recorded anywhere—reflecting their invisible

status at the time. The professionalization of science has transformed the organisation of scientific work; yet technicians' contributions remain invisible in many ways.

Heffner (1979) was one of the first to investigate the credit allocation in science using acknowledgements as recognition for contributions. Based on a questionnaire completed by 207 individuals named in acknowledgements of scientific papers (acknowledgees) from social and natural sciences, Heffner found that publication credit is not always attributed on the basis of universalistic principles, and that 12% of respondents reported having been excluded from the authors' list when they felt their contribution warranted authorship. Female PhDs were twice more likely than any other group of his sample (male and non-PhDs) to express exclusion from the authorship list when they believed they should have been named as an author.

Laband and Tollison (2000), Laudel (2002), Ponomariov and Boardman (2016), and Bozeman and Youtie (2016) analyzed collaboration beyond lists of authors. Laband and Tollison (2000) compared the number of co-authors (formal collaboration) and the number of individuals mentioned in the acknowledgements (informal collaboration) in economics and biology. Although formal co-authorship was more frequent in biology, informal collaboration appeared as more prominent in economics, demonstrating that disciplinary practices can affect collaboration in its forms and rewards. Based on interviews and publication analysis of 101 researchers, Laudel (2002) found that one third of all collaborators were non-authors, and were only mentioned in acknowledgements. Moreover, about half of the contributions were not publicly credited, and therefore remained invisible in formal communication channels. Ponomariov and Boardman (2016) surveyed 1,581 academic researchers and asked about their relationship with their closest collaborators. They show that there are numerous dimensions to co-authorship and that collaboration often does not lead to co-authored papers. The authors conclude that the "fluid content and boundaries of collaborations" (p. 1959) call for data that go beyond traditional co-authorship lists. Bozeman and Youtie (2016) interviewed and analyzed online posts of US researchers on factors relating to the perceived unwarranted exclusion and inclusion from authors' list. Their analysis shows that a few interacting variables can explain the perceived exclusion from authorship: the geographic separation of

collaborators (especially, the relocation of less experienced individuals), differential in power and experience, disagreements about the value of technical contributions, and gender dynamics.

More recently, Jabbehdari and Walsh (2017), and Paul-Hus et al. (2017a) investigated the presence non-author collaborators (i.e. those who contributed to a project but do not appear as authors) across research fields. Based on a survey of 1,643 authors, Jabbehdari and Walsh (2017) found that non-author collaborators are not rare and that their presence varies by discipline. The highest rates of non-author collaborators occurred in engineering and in agricultural sciences, while the lowest occurred in computer science and mathematics, and in physics and space science. Analyzing 362,767 scientific papers and their acknowledgements, Paul-Hus et al. (2017a) found that the mean numbers of acknowledgees (non-author collaborators) per paper were the highest in social sciences, biology, and earth and space, and the lowest in mathematics, and in chemistry. These findings show that traditional differences observed between disciplines in terms of team size are greatly reduced when acknowledgees are taken into account.

The gender gap in acknowledgements

Few studies have looked at the gender of individuals named in the acknowledgements of scientific publications. Coates (1999), Hoder-Salmon (1978), Lewis-Beck (1980), and have discussed the gender issue of scientific credit distribution looking at the contributions of spouses through the analysis scholarly books' acknowledgements.

Moore (1984) investigated the effect of authors' gender on the content of their acknowledgements, more specifically on the gender of those acknowledged. In a study based on 300 male-authored and 70 female-authored psychology books, Moore (1984) found that while men acknowledged mainly other men, women acknowledged the contributions of both genders. In another analysis, based on 684 psychology articles, Moore (1984) found a lower proportion of female acknowledgees, especially among male authors. Sugimoto and Cronin (2012) obtained the same results while analyzing the scholarly production of six important information scientists. The authors included in their sample were more likely to acknowledge individuals of the same sex, which led Sugimoto and

Cronin to conclude that “scholars are more likely to seek (and acknowledge) collaboration, consultation, and guidance from same-sex colleagues” (p.463).

Looking at the gender of authors and acknowledgees in women's studies, Cronin, Martinson and Davenport (1997) found, as expected given the field, that a vast majority of authors are women (93% of a total of 1,504 authors). When looking at the gender of acknowledgees, they found that 66% of the individuals mentioned in the acknowledgments are women, 20% men, the remainder being either unidentified or unknown. The results also show a higher mean number of acknowledgees per paper in women's studies than in philosophy, history, psychology and sociology. More recently, Dung et al. (2019) highlighted women’s hidden contributions to the field of theoretical population genetics by analyzing programmers acknowledged within articles published between 1970 and 1990 in the journal *Theoretical Population Biology*. Results (Dung et al., 2019) showed that women are significantly more present within the acknowledged programmers (43.2% of women) as compared to authors (7.4% of women).

Objectives

Focusing on acknowledgees as non-author collaborators, the objective of this paper is to better understand how gender and academic status may associate with credit attribution practices in the context of acknowledgements. More specifically, the first goal of this paper is to measure the percentage of acknowledgees who are women, and to assess whether this percentage varies across disciplines, and as a function of the gender of the leading authors. The second goal of this paper is to characterize the academic status of acknowledgees who are also authors of other scientific publications (academic age, number of publications, citations, and leading role).

Material and methods

Data

This study is based on all acknowledgements extracted from articles and reviews published between 2015 and 2017, and indexed in the Science Citation Index Expanded (SCI-E) and

Social Sciences Citation Index (SSCI) from Clarivate Analytics' Web of Science (WoS)²⁰. Access to the WoS data in a relational database format was provided by the *Observatoire des sciences et des technologies* (<http://www.ost.uqam.ca>). The dataset used in the present analysis was extracted from the full text of the acknowledgements sections of papers, and includes a total of 1,045,131 acknowledgements from as many papers. The dataset covers all disciplines included in the SCI-E and SSCI. Discipline of papers were assigned using the NSF field classification of journals (National Science Foundation, 2006); the NSF classification assigns only one discipline specialty to each journal, thus preventing multiple counting of multidisciplinary papers.

Analysis

The extraction of individual names from acknowledgement texts was done using the Stanford Named Entity Recognizer (NER) (Finkel, Grenager & Manning, 2005) module of the Natural Language ToolKit (NLTK) (Bird, 2009). To obtain the number of acknowledgees per paper, the algorithm was applied on each string of acknowledgement text retrieved from WoS and all named entities tagged as 'person' were selected²¹.

Several data cleaning procedures were undertaken in order to eliminate non-human entities from the list of extracted names. First, incomplete names were removed from the list (occurrences containing only a first or last name, or only initials), retaining only occurrences composed of a complete name (i.e. at least one initial and one last name). We then manually removed all remaining names that did not refer to individual persons such as grant, foundation, organization, and institution names. Examples of such names removed by manual cleaning include Frederick Banting (grant), Marie Curie (grant and foundation), Boehringer Ingelheim (organization) and Instituto de Salud Carlos III (institution).

²⁰ Acknowledgements data are collected and indexed in the WoS only if they include funding information (Paul-Hus, Desrochers & Costas, 2016).

²¹ This extraction procedure was used in a previous round of analysis on a smaller dataset (Paul-Hus et al., 2017a).

Since acknowledgements often contain the name(s) of the author(s) signing the paper from which the acknowledgements are retrieved, a final step of cleaning was necessary. When the name(s) extracted from the acknowledgements of a paper X matched the name of one of the authors appearing in the byline of that paper X (using the first initial and the last name), this name was removed from the acknowledgees list for that specific paper, such as in the example below:

Paper X

Authors: J. Zhang, X. Feng and Y. Xu

Acknowledgements text: “Jinsong Zhang, Xiao Feng, and Yong Xu contributed equally to this work [...]”

Determining the gender of authors and acknowledgees

For the purpose of our analysis, we consider first and corresponding authors as lead authors of a paper, since first authors are generally associated with the highest proportion of tasks performed in a paper (Larivière et al., 2016), and the corresponding author is assigned to the author responsible for correspondance and is often associated to the initial conception and supervision of the research project (Mattsson, Sundberg and Laget; 2011). If both the first and corresponding authors are women, the paper is considered female-led, if both are men, the paper is considered male-led and if first and corresponding author are of different genders, the paper is considered mixed.

The gender assignation of personal names (authors and acknowledgees) was done using the Wiki-Gendersort algorithm (Bérubé et al., 2020). Using Wikipedia pages to get gender information, this algorithm increases reliability of gender assignation by examining the first names of the names covered by Wikipedia and counting the number of masculine and feminine pronouns in the introduction section of the first twenty pages. Gender is assigned to a first name when the same gender was attributed to 75% of Wikipedia pages. No gender is assigned when this threshold is not met. As shown in Table 1, using the Wiki-Gendersort algorithm, we were able to identify the gender of 67% of all personal names mentioned in the acknowledgements of our dataset, and of 70% of the authors. The remaining are classified as unknown gender, which includes unisex names. Our analysis of

the gender variable uses occurrences of individual names for which a gender could be assigned (Female or Male).

Table 1. – Number of papers and acknowledgements mentioning individual names, by discipline

Discipline	N papers	N mentions of person's names (acknowledgements)						Authorships	
		Female		Male		Unknown		N	% with identified gender
		N	%	N	%	N	%		
Biology	124 234	116 336	20.6%	245 110	43.5%	202 656	35.9%	606 885	73.2%
Biomedical Research	177 916	139 862	23.0%	248 762	41.0%	218 193	36.0%	1 304 857	78.2%
Chemistry	101 983	38 730	17.2%	108 015	48.0%	78 319	34.8%	539 223	74.7%
Clinical Medicine	179 725	153 204	30.9%	212 903	42.9%	130 490	26.3%	1 338 073	77.1%
Earth and Space	124 087	78 717	16.2%	246 523	50.6%	161 696	33.2%	699 914	59.9%
Eng. and Tech.	131 867	33 706	11.7%	141 103	49.1%	112 576	39.2%	627 025	65.2%
Health	16 191	25 281	47.1%	20 556	38.3%	7 854	14.6%	80 641	85.1%
Mathematics	25 272	4 883	8.7%	31 010	55.0%	20 521	36.4%	64 948	72.2%
Physics	96 156	17 300	6.3%	113 884	41.7%	142 056	52.0%	601 672	39.5%
Professional Fields	15 828	20 151	27.0%	46 378	62.1%	8 101	10.9%	46 415	83.4%
Psychology	22 364	34 535	42.6%	33 583	41.4%	12 905	15.9%	91 815	88.6%
Social Sciences	29 508	38 722	24.8%	101 477	65.1%	15 793	10.1%	77 497	84.4%
Total	1 045 131	701 427	20.9%	1 549 304	46.1%	1 111 160	33.1%	6 078 965	70.2%

Determining the academic status of the acknowledgees

Determining whether an acknowledgee is the author of at least one WoS-index article or review is not an easy task, given that we only have acknowledgee's names, and no institutional or disciplinary affiliation (except that of the acknowledging paper). First, all authors names from WoS database were disambiguated using the Caron and van Eck algorithm (2014). Then, for each acknowledgee name, we found all unique disambiguated authors with the same name. We considered an author-acknowledgee match to be valid when there was only one author-acknowledgee pair found in the same discipline *or* with the same institutional affiliation as the acknowledging paper. We thus focus on precision over recall, as individuals with very common names are almost systematically excluded. Following this procedure, a total of 520,932 distinct acknowledgees with at least one WoS publication (article or review) were found.

Table 2. – Number of distinct name strings and distinct name strings who are also authors in WoS, by discipline

Discipline	N distinct acknowledgees	N distinct acknowledgees also authors in WoS	
		N	%
Biology	349 623	77 671	22.2%
Biomedical Research	380 767	105 490	27.7%
Chemistry	135 240	38 641	28.6%
Clinical Medicine	279 569	61 073	21.8%
Earth and Space	263 375	79 207	30.1%
Engineering and Technology	166 995	44 395	26.6%
Health	36 689	8 996	24.5%
Mathematics	25 955	10 471	40.3%
Physics	138 884	46 304	33.3%
Professional Fields	47 127	12 261	26.0%
Psychology	53 684	11 954	22.3%
Social Sciences	93 554	24 469	26.2%
All disciplines	1 585 389	520 932	32.9%

For each acknowledgee identified as an author, we use the following indicators as proxies for their academic status:

- Academic age (2017 minus the publication year of the first paper)²²
- Number of publications (all publications published until 2017)
- Total field-normalized citations (based on the aforementioned NSF classification, the total was calculated as the sum of field-normalized citation scores)
- Share of the acknowledgees' publications for which he or she has a leading role (first or corresponding author).

²² Data for the academic age is limited to papers published after 1980, since disambiguated authors' data are not valid before this year. The maximum academic age of an author is thus 36 years old.

These indicators are used to measure acknowledgees' position within the academic hierarchy. Results for these indicators are presented as distribution of values. In order to compare results for acknowledgees who are also authors, we use the distributions of all authors who had published at least one article in WoS between 2015 and 2017, each author being assigned to the discipline in which he/she has the highest number of publications throughout his/her career. In case of a tie, one of the tied disciplines was chosen randomly.

Results

Gender

Figure 1 compares the percentage of women among all authors, acknowledgees, and the subset of acknowledgees who are also authors, by discipline. It shows that the well-known gender gap found in authorship (Larivière et al. 2013; West et al., 2013) is also present in acknowledgements.

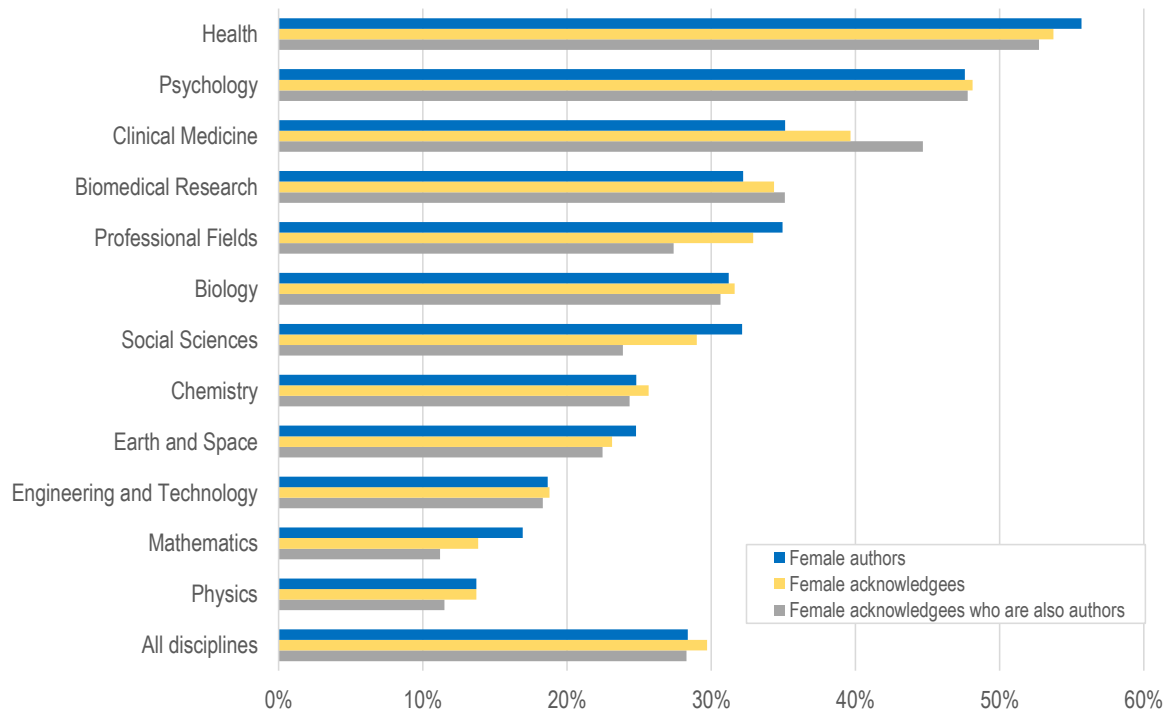


Figure 1. – Proportion of female authors, acknowledgees, and acknowledgees who are also authors by discipline.

Women represent less than half of both authors and acknowledgees in all disciplines, with the only exception of Health, where women account for the majority of authors, acknowledgees, and of acknowledgees who are also authors. Despite some disciplinary variations, proportions of female acknowledgees and female authors are quite similar (differences ranging from 4.5% between all acknowledgees and authors in Clinical Medicine to -3.1% between all acknowledgees and authors in Social Sciences and Mathematics). All disciplines taken together, women represent 28.4% of all authors, 29.7% of all acknowledgees, and 28.3% of the subset of acknowledgees who are also authors.

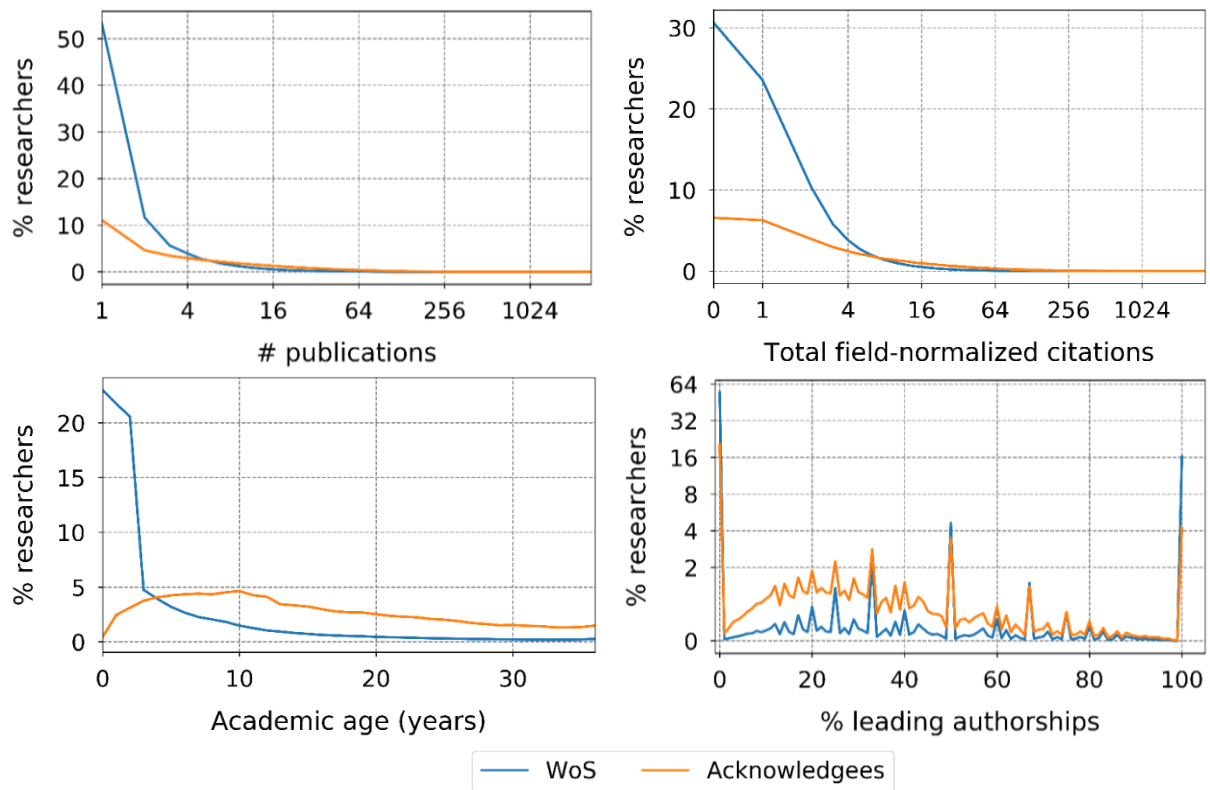
Table 3 presents the percentage of female acknowledgees as a function of leading author gender. For each discipline, the proportion of female acknowledgees is higher in female-led papers (women as first or corresponding authors) than in the mixed-led papers or male-led papers. The difference in the proportion of female acknowledgees between female-led papers and male-led papers ranges from 23.4% in Health to 11.0% in Biology with a difference of 18.0% when all disciplines are taken together.

Table 3. – Proportion of acknowledgees who are female as a function of the gender of the leading authors, by discipline

Discipline	Female-led papers	Mixed-led papers	Male-led papers	Diff. between Female-led and Male-led papers
Health	62.9%	45.1%	39.5%	23.4%
Professional Fields	48.1%	23.6%	24.8%	23.3%
Psychology	58.0%	46.9%	36.4%	21.7%
Social Sciences	42.4%	20.8%	22.5%	19.8%
Mathematics	28.7%	17.4%	10.4%	18.3%
Engineering and Technology	31.9%	19.6%	15.2%	16.8%
Clinical Medicine	49.3%	34.2%	33.0%	16.3%
Chemistry	36.3%	23.2%	21.5%	14.8%
Biomedical Research	42.4%	30.9%	30.1%	12.2%
Physics	23.7%	15.8%	11.6%	12.1%
Earth and Space	31.7%	23.7%	20.5%	11.2%
Biology	39.0%	27.5%	28.0%	11.0%
All disciplines	42.3%	27.1%	24.3%	18.0%

Academic status

Figure 2 compares the distributions of WoS authors to the subset of acknowledgees who are also authors, as a function of each of the four indicators.²³ It shows that, for all indicators, the distributions of the acknowledgees is less concentrated than that of all WoS authors.



Note: WoS = All authors who published at least one article or review in WoS between 2015-2017
Acknowledgees = Acknowledgees who published at least one article or review in WoS

Figure 2. – Distribution of researchers (all WoS authors, and acknowledgees who are also authors) by number of publications, total field-normalized citations, academic age, and percentage of leading authorships

²³ Descriptive statistics for the four indicators, detailed by disciplines, are available in the appendix.

Moreover, the acknowledgees' distributions spread on longer tails, with a smaller share of the distributions towards the lowest values for the four indicators. In terms of number of publications, 80% of all disambiguated WoS authors have less than 7 publications, while only 30% of acknowledgees have less than 7 publications (80% of the acknowledgees have less than 80 publications). When considering the total field-normalized citations, a very similar pattern is observed: 80% of WoS authors have less than 7 field-normalized citations, while only 27% of acknowledgees have less than 7 field-normalized citations (80% of acknowledgees have less than 145 field-normalized citations). As for academic age, 80% of WoS authors have an academic age of 6 years or less, while 22% of acknowledgees have an academic age of 6 years or less (80% of acknowledgees have an academic age of less than 24 years). Both distributions of leading authorships show similar patterns, with important peaks at 0% (no leading authorship), 50% (leading position in half of the authored publications) and at 100% (leading position in all authored publications), which is due to the high proportion of researchers having one or two papers. However, the distribution of acknowledgees is once again less concentrated than the WoS distribution, with 20% of the acknowledgees having less than 1% of leading authorships, while 54% of WoS authors have less than 1% of leading authorships. Taken altogether, these results show that acknowledgees who are also authors tend to have a higher position in the academic hierarchy when compared to all WoS authors.

Discussion

Our analyses of the academic status of acknowledgees, measured using numbers of publications, total field-normalized citations, academic age, and share of leading authorships, show that the subset of acknowledgees who are also authors tend to have a higher position in the academic hierarchy compared to all of WoS authors. These findings suggest that academic status may be associated with credit attribution practices, since acknowledgees appear to be rather senior researchers according to our four indicators, at least when considering the subset of acknowledgees who have already published (as defined by having at least one WoS-indexed publication).

Our findings demonstrate that acknowledgements are not limited to research assistants and less experienced researchers whose contributions to research (often

technical) cannot justify authorship, but also extend to researchers of higher academic status, according to our four indicators. Given the higher position in the academic hierarchy of the subset of acknowledgees who have already published, we may not only consider acknowledgements as a form of sub-authorship, as it has often been conceived (Costas & van Leeuwen, 2012; Diaz-Faes & Bordons, 2014; Patel, 1973), but also as genuine form of credit for informal collaboration with experienced colleagues. Our analysis does not allow us to associate academic status with the nature of the acknowledged contribution. However, it has been shown that senior researchers, when authors of a paper, are more frequently associated with conceptual tasks and resources contributions, while younger researchers are more likely to contribute to experimentation (Larivière et al., 2016). This division of labour might be mirrored in the acknowledgements. Our findings could thus be explained by the fact that manuscript revision and resource allocation (contributions frequently mentioned in the acknowledgements, Paul-Hus et al., 2017b) are reserved to researchers with higher seniority. In this sense, acknowledgements to researchers of higher academic status might reveal the “invisible college” of close-but-distant collaborators who contribute in informal ways to a research project (Price & Beaver, 1966). Furthermore, our results may be another manifestation of the Matthew Effect (Merton, 1968), as researchers of higher academic status who already have recognition and visibility in the scientific community tend to be overrepresented among the acknowledgees. When two researchers, one junior and one senior, contribute to a research project without meeting authorship criteria, we might be more inclined to acknowledge the contribution of a senior research, precisely because of its seniority.

Regarding the gender of individuals named in the acknowledgements of scientific papers, our analyses have shown that gender disparities generally found in authorship extend to acknowledgements. Globally, women are underrepresented in both authorships and acknowledgements of scientific papers. Furthermore, as previously found by Moore (1984) and by Sugimoto and Cronin (2012), our findings clearly confirm that women acknowledge proportionally more women than men do. Our results are in line with the gender homophily pattern in team composition and social networks, which refers to the tendency to associate more frequently with same-sex individuals (Ibarra, 1992; McPherson, Smith-Lovin & Cook, 2001). Women thus appear to be less homophilic than

men in their acknowledgement practices. This finding is consistent with previous analyses of gender homophily in scientific collaborations (Araújo et al., 2017; Bozeman & Corley, 2004) showing that men are more likely to collaborate with other men while women are more “egalitarian”.

However, our results also show important differences between disciplines. These differences in the percentages of women among acknowledgees could also be due to the gender composition in each discipline and the broad disciplines categorization used in our analyses. In fact, disciplines where we observed the highest levels of gender homophily, Health and Professional Fields, both contain research areas generally considered as highly feminized (Witz, 1992), such as Nursing and Education, as well as male-health oriented areas. In this context, observed gender homophily could be a second order effect of the gender composition of a discipline, if researchers of a given research area acknowledge individuals from their own area. Moreover, observed gender differences and more generally the greater proportion of male acknowledgees within our dataset could also be a second order effect, explained by the over-representation of researchers of higher academic status among the acknowledgees. Given the well-known over-representation of men in positions of higher rank in the scientific community (Charles, 2003; Etzkowitz, Kemelgor & Uzzi, 2000), the observed over-representation of acknowledgees with higher academic status implies a greater proportion of male among acknowledgees. The gender differences we found could thus be due, at least in part, to second order effects, without necessarily be a direct reflection of gender biased acknowledgement practices.

Limitations

Some limitations relating to our data source and methods must be considered when interpreting our results. First, acknowledgement data are limited to funded research since they are collected with the intended objective of tracking funded research (Web of Science, 2009). Acknowledgements are thus collected and indexed by WoS only if they include some kind of funding information. These indexation criteria could induce a bias toward funded research and funding-related aspects of acknowledgements. Second, the gender assignment algorithm we used, Wiki-gendersort (Bérubé et al., 2020) presents a limitation that is common to most gender assignment tools: lower reliability for names of Asian

origin, more specifically Chinese names. It is thus safe to suppose that Chinese names are overrepresented in the unknown gender category (Santamaría & Mihaljević, 2018). Finally, our results concerning the academic status of acknowledgees are restricted to individuals who have already published. It is thus reasonable to assume that this subset of acknowledgees might be characterized by a higher academic status than the rest of the acknowledgees who have not published, being either less experienced researchers or technicians and assistants. As a consequence, our conclusions might not apply to acknowledgees who have not published.

Conclusion

Scientific collaboration is often synonymous with co-authorship, despite the fact that it remains a partial indicator of collaboration (Katz & Martin, 1997). However, collaborators are not always authors of the papers to which they have contributed, and acknowledgements can help reveal not only the contribution of these non-author collaborators, but also their sociodemographic characteristics, making it possible to draw new insights on the social structure of science as well as on practices of collaboration, division of labor, and credit attribution.

Regarding the academic status of acknowledgees, our results show that acknowledgees who have already published tend to have a higher position in the academic hierarchy compared to all of WoS authors. These findings indicate that acknowledgements are not limited to less experienced researchers whose contributions cannot justify authorship, but also extend to more experienced researchers.

Our results also show that women are underrepresented in acknowledgements. In a broader context, academic stereotypes have been shown to act as gatekeepers by steering women away from certain fields (Cheryan, Master & Meltzoff, 2015). Perceived underrepresentation of women, whether when considering authorships or acknowledged contributions, can thus contribute to academic gendered stereotypes and exacerbate gender disparities in both local and global scientific communities (Dung et al., 2019; Larivière et al., 2013; West et al., 2013).

Acknowledgements

The authors would like to thank the two reviewers for their helpful comments and Gita Ghiasi for her help with Wiki-Gendersort.

Funding information

This research was supported by the Social Sciences and Humanities Research Council of Canada: Joseph-Armand Bombardier CGS Doctoral Scholarships (Paul-Hus); Insight Development Grant (Larivière).

References

- Araújo, E. B., Araújo, N. A. M., Moreira, A. A., Herrmann, H. J. & Andrade Jr, J.S. (2017). Gender differences in scientific collaborations : Women are more egalitarian than men. *PLOS ONE*, 12(5), e0176791. <https://doi.org/10.1371/journal.pone.0176791>
- Bérubé, N., Sainte-Marie, M., Ghiasi, G., & Larivière V. (2020). Wiki-Gendersort: Automatic gender detection using first names in Wikipedia. <https://osf.io/preprints/socarxiv/ezw7p>
- Bird, S., Klein, E. & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*. Sebastopol, CA: O'Reilly Media.
- Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, 14(6), 19-47.
- Bozeman, B. & Corley, E. (2004). Scientists' collaboration strategies : Implications for scientific and technical human capital. *Research Policy*, 33(4), 599-616. <https://doi.org/10.1016/j.respol.2004.01.008>
- Bozeman, B. & Youtie, J. (2016). Trouble in paradise : Problems in academic research co-authoring. *Science and Engineering Ethics*, 22(6), 1717-1743. <https://doi.org/10.1007/s11948-015-9722-5>
- Caron, E. & van Eck, N. J. (2014). Large-scale author name disambiguation using rule-based scoring and clustering. In E. Noyons (Ed.), *19th International Conference on*

Science and Technology Indicators “Context counts: Pathways to master big data and little data” (p. 79-86). Leiden University, The Netherlands.

Charles, M. (2003). Deciphering sex segregation : Vertical and horizontal inequalities in ten national labor markets. *Acta Sociologica*, 46(4), 267-287.

<https://doi.org/10.1177/0001699303464001>

Cheryan, S., Master, & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: increasing girls’ interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00049>

Coates, C. (1999). Interpreting academic acknowledgements in English studies: Professors, their partners, and peers. *English Studies in Canada*, 25(3-4), 253-276.

Cole, J. R. & Cole, S. (1973). *Social stratification in science*. Chicago: University of Chicago Press.

Costas, R. & van Leeuwen, T. N. (2012). Approaching the “reward triangle” : General analysis of the presence of funding acknowledgments and “peer interactive communication” in scientific publications. *Journal of the American Society for Information Science and Technology*, 63(8), 1647-1661.

Cronin, B. (1995). *The scholar’s courtesy : The role of acknowledgement in the primary communication process*. London: Taylor Graham.

Cronin, B., Martinson, A. & Davenport, E. (1997). Women’s studies : Bibliometric and content analysis of the formative years. *Journal of Documentation*, 53(2), 123-138.

<https://doi.org/10.1108/EUM0000000007196>

Diaz-Faes, A. A. & Bordons, M. (2014). Acknowledgments in scientific publications : Presence in Spanish science and text patterns across disciplines. *Journal of the Association for Information Science and Technology*, 65(9), 1834-1849.

<https://doi.org/10.1002/asi.23081>

Dung, S. K., López, A., Barragan, E. L., Reyes, R.-J., Thu, R., Castellanos, E., Catalan, F., Huerta-Sanchez, E. & Rohlf, R. V. (2019). Illuminating women’s hidden contribution

- to historical theoretical population genetics. *Genetics*, 211(2), 363-366.
<https://doi.org/10.1534/genetics.118.301277>
- Etzkowitz, H., Kemelgor, C. & Uzzi, B. (2000). *Athena unbound : The advancement of women in science and technology*. Cambridge: Cambridge University Press.
- Finkel, J.R., Grenager T. & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In K. Knight, H.T. Ng and K. Oflazer (Eds), *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (ACL 2005, pp. 363-370). Ann Arbor, MI, Association for Computational Linguistics.
- Haeussler, C. & Sauermann, H. (2013). Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship. *Research Policy*, 42(3), 688-703. <https://doi.org/10.1016/j.respol.2012.09.009>
- Heffner, A. G. (1979). Authorship recognition of subordinates in collaborative research. *Social Studies of Science*, 9(3), 377-384. <https://doi.org/10.1177/030631277900900305>
- Hoder-Salmon, M. (1978). Collecting scholar's sives. *Feminist Studies*, 4(3), 107-114.
<https://doi.org/10.2307/3177543>
- Ibarra, H. (1992). Homophily and differential returns : Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly*, 37(3), 422-447. <https://doi.org/10.2307/2393451>
- Jabbehdari, S. & Walsh, J. P. (2017). Authorship norms and project structures in science. *Science, Technology, & Human Values*, 42(5) 872-900.
<https://doi.org/10.1177/0162243917697192>
- Katz, J. S. & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1-18. [https://doi.org/10.1016/S0048-7333\(96\)00917-1](https://doi.org/10.1016/S0048-7333(96)00917-1)
- Laband, D. N. & Tollison, R. D. (2000). Intellectual collaboration. *Journal of Political Economy*, 108(3), 632-662. <https://doi.org/10.1086/262132>

- Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A. & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417-435. <https://doi.org/10.1177/0306312716650046>
- Larivière, V., Ni, C. C., Gingras, Y., Cronin, B. & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, 504(7479), 211-213. <https://doi.org/10.1038/504211a>
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1), 3-15. <https://doi.org/10.3152/147154402781776961>
- Lewis-Beck, J. A. (1980). The participation of men and women in educational research: Another look. *Sex Roles*, 6(4), 607-610.
- Lissoni, F., Montobbio, F., & Zirulia, L. (2013). Inventorship and authorship as attribution rights: An enquiry into the economics of scientific credit. *Journal of Economic Behavior & Organization*, 95, 49-69. <https://doi.org/10.1016/j.jebo.2013.08.016>
- Marušić, A., Bošnjak, L. & Jerončić, A. (2011). A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS ONE*, 6(9), e23477. <https://doi.org/10.1371/journal.pone.0023477>
- Mattsson, P., Sundberg, C. J. & Laget, P. (2011). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics*, 87(1), 99-105. <https://doi.org/10.1007/s11192-010-0310-9>
- McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather : Homophily in Social Networks. *Annual Review of Sociology*, 27(1), 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago and London: Chicago University Press.
- Moore, M. (1984). Sex and acknowledgements: A nonreactive study. *Sex Roles*, 10(11-12), 1021-1031.

- National Science Foundation. (2006). Science and engineering indicators. Chapter 5: Academic research and development. Data and terminology. Retrieved from <http://www.nsf.gov/statistics/seind06/c5/c5s3.htm#sb1>
- Patel, N. (1973). Collaboration in the professional growth of American sociology. *Social Science Information*, 12(6), 77-92.
- Paul-Hus, A., Desrochers, N. & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, 108(1), 167-182. <https://doi.org/10.1007/s11192-016-1953-y>
- Paul-Hus, A., Mongeon, P., Sainte-Marie, M. & Larivière, V. (2017a). The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements. *Journal of Informetrics*, 11(1), 80-87. <https://doi.org/10.1016/j.joi.2016.11.005>
- Paul-Hus, A., Díaz-Faes, A. A., Sainte-Marie, M., Desrochers, N., Costas, R. & Larivière, V. (2017b). Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences. *PLOS ONE*, 12(10), e0185578. <https://doi.org/10.1371/journal.pone.0185578>
- Ponomariov, B. & Boardman, C. (2016). What is co-authorship? *Scientometrics*, 109(3), 1939-1963. <https://doi.org/10.1007/s11192-016-2127-7>
- Price, D. J. de S. & Beaver, D. (1966). Collaboration in an Invisible College. *American Psychologist*, 21(11), 1011–1018.
- Santamaría, L. & Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, 4, e156. <https://doi.org/10.7717/peerj-cs.156>
- Shapin, S. (1989). The invisible technician. *American Scientist*, 77(6), 554-563.
- Sismondo, S. (2009). Ghosts in the machine publication planning in the medical sciences. *Social Studies of Science*, 39(2), 171-198. <https://doi.org/10.1177/0306312708101047>
- Smith, E., Williams-Jones, B., Master, Z., Larivière, V., Sugimoto, C. R., Paul-Hus, A., Shi, M. & Resnik, D. B. (2019). Misconduct and misbehavior related to authorship

- disagreements in collaborative science. *Science and Engineering Ethics*.
<https://doi.org/10.1007/s11948-019-00112-4>
- Sugimoto, C. R. & Cronin, B. (2012). Biobibliometric profiling: An examination of multifaceted approaches to scholarship. *Journal of the American Society for Information Science and Technology*, 63(3), 450-468. <https://doi.org/10.1002/asi.21695>
- Wager, E. (2009). Recognition, reward and responsibility: Why the authorship of scientific papers matters. *Maturitas*, 62(2), 109-112.
<https://doi.org/10.1016/j.maturitas.2008.12.001>
- Web of Science. (2009). Funding Acknowledgements.
http://wokinfo.com/products_tools/multidisciplinary/webofscience/fundingsearch/
- West, J. D., Jacquet, J., King, M. M., Correll, S. J. & Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PLoS ONE*, 8(7), e66212.
<https://doi.org/10.1371/journal.pone.0066212>
- Wislar, J. S., Flanagan, A., Fontanarosa, P. B. & DeAngelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ*, 343, d6128. <https://doi.org/10.1136/bmj.d6128>
- Witz, A. (1992). *Professions and patriarchy*. London: Routledge.
<https://doi.org/10.4324/9780203167786>
- Zuckerman, H. A. (1977). *Scientific elite: Nobel laureates in the United States*. New Brunswick and London: Transaction Publishers.

Appendix

Descriptive statistics for the number of publications, total field-normalized citations, academic age, and percentage of leading authorships, by discipline

Table 4. – Descriptive statistics for the number of publications, by discipline

	Number of publications					
	Mean		Median		Standard deviation	
	Ackn	WoS	Ackn	WoS	Ackn	WoS
Biology	37	8	16	2	59.2	20.3
Biomedical Research	57	7	25	1	85.2	22.5
Chemistry	68	9	28	1	105.1	32.1
Clinical Medicine	43	10	13	1	71.6	31.1
Earth and Space	56	10	31	1	76.3	25.8
Engineering and Technology	49	6	19	1	84.7	17.1
Health	42	4	13	1	75.1	11.1
Mathematics	44	10	28	2	56.5	20.4
Physics	83	19	46	1	122.3	76.8
Professional Fields	23	5	12	2	39.0	9.3
Psychology	41	8	15	2	68.2	20.4
Social Sciences	24	5	13	2	34.9	10.6

Note: Ackn = Acknowledgees who are also authors
 WoS = All authors who published at least one article or review in WoS between 2015-2017

Table 5. – Descriptive statistics for the total field-normalized citations, by discipline

	Total field-normalized citations					
	Mean		Median		Standard deviation	
	Ackn	WoS	Ackn	WoS	Ackn	WoS
Biology	61.4	10.0	17.5	1.5	137.7	35.0
Biomedical Research	126.5	10.9	37.2	1.2	271.7	51.4
Chemistry	119.1	11.3	34.9	1.2	256.3	56.5
Clinical Medicine	81.5	14.2	15.9	1.2	186.9	68.1
Earth and Space	118.9	15.0	44.4	1.5	240.4	62.5
Engineering and Technology	92.4	7.2	23.1	1.0	225.8	35.8
Health	84.6	5.1	15.4	1.1	211.9	20.6
Mathematics	80.6	11.4	38.6	1.6	163.4	42.3
Physics	217.5	37.2	81.9	1.4	432.1	205.7
Professional Fields	49.3	6.7	17.9	1.2	105.9	24.8
Psychology	75.5	11.3	18.1	1.5	174.5	42.5
Social Sciences	50.6	7.9	18.1	1.4	102.3	27.6

Note: Ackn = Acknowledgees who are also authors
 WoS = All authors who published at least one article or review in WoS between 2015-2017

Table 6. – Descriptive statistics for the academic age, by discipline

	Academic age					
	Mean		Median		Standard deviation	
	Ackn	WoS	Ackn	WoS	Ackn	WoS
Biology	14	5	12	2	9	7
Biomedical Research	15	4	14	2	10	6
Chemistry	15	4	12	2	10	6
Clinical Medicine	13	4	11	2	9	7
Earth and Space	16	5	14	2	9	7
Engineering and Technology	13	3	11	1	9	5
Health	13	3	10	2	9	5
Mathematics	18	6	17	2	9	8
Physics	16	5	15	2	9	7
Professional Fields	14	4	12	2	9	6
Psychology	13	5	10	2	9	7
Social Sciences	14	5	12	2	9	7

Note: Ackn = Acknowledgees who are also authors
 WoS = All authors who published at least one article or review in WoS between 2015-2017

Table 7. – Descriptive statistics for the percentages of leading authorships, by discipline

	Percentage of leading authorships					
	Mean		Median		Standard deviation	
	Ackn	WoS	Ackn	WoS	Ackn	WoS
Biology	29.0	29.4	25.0	0.0	26.7	37.3
Biomedical Research	26.4	23.5	23.9	0.0	23.0	35.0
Chemistry	28.7	29.1	25.0	0.0	24.9	38.3
Clinical Medicine	21.9	22.6	16.3	0.0	24.4	34.5
Earth and Space	27.6	28.4	23.5	0.0	23.5	37.3
Engineering and Technology	30.4	32.8	25.0	0.0	27.7	41.0
Health	27.9	27.6	20.3	0.0	29.1	39.0
Mathematics	56.1	52.6	56.3	55.6	28.9	40.8
Physics	34.0	26.2	28.8	0.0	26.0	36.8
Professional Fields	52.7	47.6	50.4	50.0	31.6	42.6
Psychology	32.4	34.7	29.1	22.2	28.4	38.6
Social Sciences	56.4	53.9	57.1	60.0	32.0	42.3

Note: Ackn = Acknowledgees who are also authors
 WoS = All authors who published at least one article or review in WoS between 2015-2017

3.3 Les types de contributions qui mènent aux remerciements

Article 3

Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences

Publié dans *PLoS ONE* (2017); 2(10): e0185578.

<https://doi.org/10.1371/journal.pone.0185578>

Cet article est inclus dans la présente thèse avec l'autorisation des co-auteurs.

© 2017 Paul-Hus et al. Cette version du manuscrit disponible en accès libre sous la licence CC-BY 4.0 <https://creativecommons.org/licenses/by/4.0/>

Contributions des auteurs (selon la taxonomie CRediT²⁴)

Adèle Paul-Hus: Conception de l'étude, méthodologie, investigation, analyse des données, visualisation, rédaction de l'ébauche initiale, révision critique du manuscrit et acquisition de soutien financier

Adrián A. Díaz-Faes : Conception de l'étude, méthodologie, calcul/computation, analyse des données, visualisation, rédaction de l'ébauche initiale et révision critique du manuscrit

Maxime Sainte-Marie: Gestion des données, méthodologie, calcul/computation et révision critique du manuscrit

Nadine Desrochers : Conception de l'étude, méthodologie, analyse des données, rédaction de l'ébauche initiale, révision critique du manuscrit et acquisition de soutien financier

Rodrigo Costas : Conception de l'étude, méthodologie, ressources, rédaction de l'ébauche initiale, révision critique du manuscrit et acquisition de soutien financier

Vincent Larivière: Conception de l'étude, méthodologie, investigation, analyse des données, supervision, ressources, rédaction de l'ébauche initiale, révision critique du manuscrit et acquisition de soutien financier

Contributions de l'auteure principale: Dans le cadre de ce projet, j'ai d'abord défini les objectifs et questions de recherche spécifiques à cet article. Sous la supervision de mes co-directeurs Nadine Desrochers et Vincent Larivière, j'ai défini le devis méthodologique initial de l'étude. Ainsi, après avoir réalisé la revue de la littérature pertinente, j'ai opté pour une analyse de correspondance des données de remerciements afin de répondre à mes questions de recherche. J'ai donc contacté le collègue Adrián A. Díaz-Faes afin d'obtenir son aide et d'obtenir de la formation et sa collaboration pour réaliser l'analyse de correspondance, puisque ce dernier avait déjà effectué ce type d'analyse sur un jeu de données de remerciements plus restreint (Díaz-Faes et Bordons, 2014). Une fois l'ensemble des co-auteurs réunis sur le projet, nous avons révisé le devis méthodologique et j'ai défini les paramètres qui permettrait d'aller plus loin que Díaz-Faes et Bordons (2014) et

²⁴ <https://casrai.org/credit/>

d'étudier d'autres aspects des remerciements à travers l'analyse de correspondance, tout en y ajoutant des outils et méthodes acquis lors des premières phases de mon projet. Suivant des critères d'inclusion et d'exclusion conjointement définis afin de répondre à mes questions de recherche, Vincent Larivière a élaboré des requêtes SQL pour récolter les données bibliométriques nécessaires (incluant les textes de remerciements). Maxime Sainte-Marie a ensuite implémenté un algorithme permettant d'extraire les syntagmes nominaux des chaînes de textes de remerciements. Sous ma supervision, Maxime Sainte-Marie a procédé à plusieurs itérations et corrections de l'algorithme. Une fois les syntagmes nominaux extraits, Adrián A. Díaz-Faes a produit une première itération de l'analyse de correspondance. Suite à l'examen de cette première itération, révisée par Nadine Desrochers, Rodrigo Costas et Vincent Larivière, j'ai défini les modifications à apporter aux paramètres de l'analyse de correspondance et Adrián A. Díaz-Faes a produit la version finale de l'analyse et les visualisations associées (Figures 1 et 2). J'ai ensuite fait les autres visualisations, l'analyse des données et l'interprétation des résultats; j'ai alors fait appel à mes collaborateurs Adrián A. Díaz-Faes, Nadine Desrochers, Rodrigo Costas et Vincent Larivière pour valider et étoffer mes analyses, selon une approche collaborative dont je demeurais la chercheuse principale et l'instigatrice. J'ai finalement rédigé l'ébauche originale de l'article, qui a ensuite été révisée, commentée et augmentée par l'ensemble des co-auteurs.

Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences

Adèle Paul-Hus¹, Adrián A. Díaz-Faes², Maxime Sainte-Marie¹, Nadine Desrochers¹, Rodrigo Costas³ and Vincent Larivière⁴

¹adele.paul-hus@umontreal.ca, maxime.sainte-marie@umontreal.ca,
nadine.desrochers@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal
PO Box 6128, Downtown Station, Montreal, Quebec, H3C 3J7 (Canada)

² diazfaes@ingenio.upv.es

INGENIO (CSIC-UPV), Universitat Politècnica de València, Camino de Vera s/n, 46022
Valencia (Spain)

³ rcostas@cwts.leidenuniv.nl

Centre for Science and Technology Studies (CWTS), Leiden University, P.O. box 905,
2300 AX Leiden (the Netherlands) and
Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch
University, Private Bag X1, Matieland 7602, (South Africa)

⁴ vincent.lariviere@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal
PO Box 6128, Downtown Station, Montreal, Quebec, H3C 3J7 (Canada) and
Université du Québec à Montréal, Centre Interuniversitaire de Recherche sur la Science et
la Technologie (CIRST), Observatoire des Sciences et des Technologies (OST), PO Box
8888, Downtown Station, H3C 3P8 Montréal, Qc. (Canada)

Abstract

For the past 50 years, acknowledgements have been studied as important paratextual traces of research practices, collaboration, and infrastructure in science. Since 2008, funding acknowledgements have been indexed by Web of Science, supporting large-scale analyses of research funding. Applying advanced linguistic methods as well as Correspondence Analysis to more than one million acknowledgements from research articles and reviews published in 2015, this paper aims to go beyond funding disclosure and study the main types of contributions found in acknowledgements on a large scale and through disciplinary comparisons. Our analysis shows that technical support is more frequently acknowledged

by scholars in Chemistry, Physics and Engineering. Earth and Space, Professional Fields, and Social Sciences are more likely to acknowledge contributions from colleagues, editors, and reviewers, while Biology acknowledgements put more emphasis on logistics and fieldwork-related tasks. Conflicts of interest disclosures (or lack of thereof) are more frequently found in acknowledgements from Clinical Medicine, Health and, to a lesser extent, Psychology. These results demonstrate that acknowledgement practices truly do vary across disciplines and that this can lead to important further research beyond the sole interest in funding.

Keywords: acknowledgment; contribution; funding; scientific credit; correspondence analysis

Introduction

For the past 50 years, acknowledgements have been studied as important paratextual traces of research practices, collaboration, infrastructure, and funding in science. In 1970, Crawford and Biderman first used acknowledgements found in footnotes to analyse papers' funding sources. Such analyses remained scarce for the following decades, mainly because of the lack of acknowledgement data on a large scale. This changed in 2008, when the Web of Science (WoS) started to index scientific papers' acknowledgements, with a special focus on the funding information they contain. This has led to several analyses of trends in funding sources, as well as the relationship between funding, productivity, and the impact of publications (e.g., Costas & van Leeuwen, 2012; Lewison & Markusova, 2010; Rigby, 2013; Tang, Hu & Liu, 2016; Zhou & Tian, 2014). However, papers' acknowledgements provide much more than funding information: they convey the indebtedness of authors to people and institutions whose contributions deserve to be publicly noted. As such, acknowledgements can be seen as expressions of scientific debt and have even been conceptualized as “super-citations” (Edge, 1979, p. 106).

Although the concept of collaboration is typically measured through co-authorship, some of its dimensions can also be captured by examining the number and role of people acknowledged for their contributions (Paul-Hus et al., 2017). Contributions and assistance acknowledged in scholarly papers have been described as a form of “subauthorship collaboration” (Patel, 1973, p. 86). In that sense, acknowledgements help to reveal the—

otherwise invisible—infrastructure that supports research. They show how colleagues, tools, equipment, materials, and grants are mobilized in the context of a research project (Cronin, 2005). Patel goes further in stating that “some illuminating details of group research are recorded nowhere else but in footnotes” (1973, p. 85). Investigating the forms of rewards associated with scientific collaboration, Laudel (2002) found that certain types of scientific contributions are only rewarded through acknowledgements, while a significant portion of contributions remained completely invisible due to a lack of formal reward either in the byline or acknowledgement of a paper. This may be explained by both the lack of clear criteria for authorship and the fact that authors may not be familiar enough with such criteria (Marušić, Bošnjak, & Jerončić, 2011).

Cronin (2005) also put forward the value of acknowledgements, positioning them as data to be used alongside authorship and citation in the “reward triangle”, and thereby allowing “a more nuanced understanding of scholarly communication and interaction” (Cronin & Weaver, 1995, p. 7). Acknowledgements can be perceived as credit for contributions that complement authorship and could therefore be used to better understand collaboration and the division of labour in research. In short, acknowledgements have been studied as important paratextual traces (Genette, 1997) of collaboration, disciplinary research practices, and infrastructure in science for the past 50 years (Desrochers, Paul-Hus & Larivière, 2016; Desrochers, Paul-Hus & Pecoskie, 2017); yet the lack of standardization in both their form and indexation, as well as an incomplete understanding of their function in the scholarly communication process, has led them to remain peripheral in the reward system of science.

In the 1990s, Cronin and McCain proposed a classification of acknowledgements by contribution types. The “peer interactive communication” (PIC) category was first described by McCain (1991, p. 512) and later adapted by Cronin (Cronin et al., 1993) to refer to conceptual and cognitive contributions. Other categories included financial support, access to data and materials, technical assistance and manuscript preparation (Cronin et al., 1993; McCain, 1991). In the following decades, acknowledgement analyses showed differences between disciplines in the prevalence of acknowledgements (Cronin,

Shaw & La Barre, 2003; Salager-Meyer et al., 2010) and in the nature of support acknowledged (Cronin, Shaw & La Barre, 2004).

Giles and Council (2004) was the first large-scale analysis of acknowledgements using natural language processing to extract named entities. Based on 188,052 acknowledgements automatically extracted from 335,000 research publications from the field of computer science, the authors proposed a measure of the impact of the most frequently acknowledged entities (AEs), the citation to acknowledgement (C/A) metric (Giles & Council, 2004). Giles and his colleagues pursued and presented this pioneering work in two conferences in 2012. First, Khabza, Treeratpituk and Giles (2012) extracted and disambiguated AEs from 526,930 acknowledgements from publications in computer science, mathematics, and statistics and found a correlation between the number of acknowledgements received by individuals and their h-index. Second, Khabza, Koppman and Giles (2012) furthered the analysis of AEs, looking at the social networks between named entities extracted from acknowledgements in computer science and mathematics. These studies were the first ones to use natural language processing on large datasets of acknowledgements. In the present study, the focus is moved from named entities to the tasks and contributions acknowledged in scientific paper and a multidisciplinary perspective, using publications from biomedical, natural and social sciences, is presented in order to compare acknowledgements practices in different fields.

Two analyses of PIC using WoS acknowledgement data have also been performed recently (Costas & van Leeuwen, 2012; Díaz-Faes & Bordons, 2014). Costas and van Leeuwen (2012) operationalized PIC acknowledgements as the presence of terms relating to peer review and informal communication, such as ‘referee’, ‘comment’, ‘discussion’, ‘reading’ and ‘advice’. Using that definition, PIC acknowledgements were associated with a lower level of co-authorship. Costas and van Leeuwen (2012) also found that papers that included funding acknowledgements presented a higher impact in terms of citations when compared to papers without funding acknowledgements. Focusing on papers published by Spanish researchers, Díaz-Faes and Bordons (2014) compared acknowledgements patterns in four disciplines—cardiac and cardiovascular systems, economics, evolutionary biology,

statistics and probability—and found that PIC acknowledgements were mostly associated with theoretical and social research.

The objective of this paper is to go beyond the funding disclosure inherent to the acknowledgements index in the WoS and study the main types of contributions found in acknowledgements across various disciplines and in the differences they exhibit. More specifically, this study aims at answering the following research questions:

- What types of contributions are acknowledged?
- How do the types of acknowledged contributions vary by discipline?

Given the well-known disciplinary difference in authorship practices (Birnholtz, 2006; Larivière et al., 2016; Pontille, 2004), one might expect to observe similar disciplinary differences in the types of contributions acknowledged.

Data and Methods

Data for this study were retrieved from the Web of Science’s Science Citation Index Expanded (SCI-E) and Social Sciences Citation Index (SSCI), which both include funding acknowledgement data. WoS started to collect funding acknowledgement data in August 2008; however, until 2015, only articles and reviews covered in the SCI-E were included for funding acknowledgement. In 2015, WoS began to collect and index funding acknowledgement from articles and reviews covered in the SSCI (Paul-Hus, Desrochers & Costas, 2016).

These data are structured in three fields: the “Funding Text” (FT), “Funding Agency” (FO) and “Grant Number” (FG). FT is the full text of the acknowledgements and therefore contains funding information, but also all other types of support acknowledged by the authors. As noted above, however, acknowledgements are only collected and indexed by WoS if they include funding source information; they must also be written in English. In order to be able to compare acknowledgements patterns from the disciplines covered by the SCI-E (natural and biomedical sciences) to the ones covered by the SSCI (social sciences), 2015 was chosen as the target year for data collection and analysis.

Acknowledgements from 2015 articles and reviews (hereafter referred to as papers) indexed in the SCI-E and the SSCI were extracted. The corpus includes a total of 1,009,411 acknowledgements for as many papers. Discipline assignment was done using the National Science Foundation (NSF) field classification of journals which assigns only one discipline specialty to each journal, preventing the possible double counts of papers (National Science Foundation, 2006).

In order to identify and discriminate between the different types of contribution mentioned in the acknowledgements we collected, we performed a linguistic analysis focusing on noun phrase patterns. Broadly speaking, nouns phrases (or nominal phrases) are groups of words centered on a given noun and which, together with zero or more constituents of various syntactical categories (or adnouns), perform the same grammatical function as single nouns (for example, ‘financial support’, compared to simply ‘support’).

Many linguistic pre-processing stages were necessary to efficiently extract noun phrases from the 1,009,411 acknowledgements of our dataset. First, the acknowledgements retrieved from the FT field were segmented into words using the Penn TreeBank tokenizer (Marcus, Marcinkiewicz & Santorini, 1993; Marcus et al., 1994) of the Natural Language ToolKit (Bird, 2009). In order to identify common nouns present in the corpus, each tokenized acknowledgement was then morphologically and syntactically analyzed using the Stanford Log-Linear Part-of-Speech (POS) Tagger (Toutanova et Manning, 2000; Toutanova et al., 2003), based on the Penn TreeBank English POS tagset (Marcus, Marcinkiewicz & Santorini, 1993; Marcus et al., 1994).

With respect to the identification and extraction of noun phrases, various implementation alternatives are available, since the characterization of noun phrases is still an active field of research in linguistics. Noun phrase chunking was implemented using a modified version of the grammatical ruleset designed by Kim et al. (2010) for keyword extraction. This ruleset consists of two rules. The first rule aims to detect the nominal components of a noun phrase, from a single noun such as ‘funding’ to a sequence of nouns and/or adjectives ending with a noun, as in ‘internal funding program’ (Lukas et al., 2015). The objective of the second rule is to merge any pair of consecutive nominal components

identified by the first rule and separated by any preposition or subordinating conjunction (of, for, with, in...), such as in ‘technical assistance in bacterial challenge experiment degeneration’ (Zhang et al., 2015). However, as formulated by Kim et al. (2010), this second rule only allows for one such merging, which prevents the full identification of noun phrases such as ‘strategy for gene discovery in schizophrenia’ (Pouget et al., 2015) or ‘coating of functionalized polysaccharide with embedded nanoparticles’ (Sarbova et al., 2015). In order to properly extract such noun phrases, the second rule by Kim et al. (2010) was modified in its application to our dataset in order to allow for multiple mergings of this sort. Following these linguistic processing stages, the noun phrase extraction procedure yielded initial totals of 408,608 distinct noun phrases and 5,646,656 occurrences.

A minimal transformation and standardization of the data was then performed by removing all noun phrases (NPs) containing only references to grant numbers. This was done by removing all NPs containing actual numbers and one of the following substrings: ‘no.’, ‘no’, ‘number’, ‘ref.’). NPs corresponding to any single letter of the alphabet were also removed, along with any punctuation sign at either extremity of the NP.

A frequency score was generated for each NP extracted, providing the number of occurrences for each NP in the dataset. A threshold of at least two occurrences was applied, meaning that all NPs appearing only once (*hapax legomena*) were removed. Fifteen NPs were also removed because they were deemed to have no meaning for the purpose of the analysis (i.e., ‘Auspex’, ‘and/or publication’, ‘Da’, ‘der’, ‘du’, ‘fur’, ‘herein’, ‘http’, ‘la’, ‘none’, ‘part’, ‘portion’, ‘section’, ‘van’ and ‘year’). Finally, eight NPs were merged in order to standardize distinct versions of the same NP (e.g., ‘field work’ and ‘fieldwork’, ‘lab’ and ‘laboratory’). As a result of these operations, the final dataset was reduced to 97,766 distinct NPs and a total of 4,875,216 occurrences (see Supporting Information for the frequency distribution of all NPs in the corpus).

For the purposes of the analysis, the dataset was partitioned by discipline and a Correspondence Analysis (CA, Beh & Lombardo, 2014; Benzécri, 1992; Greenacre, 2007) was applied to these subsets following the procedure described in Díaz-Faes and Bordons (2014) and using a MATLAB program (Vicente-Villardón, 2014). This multivariate

exploratory method allows for the displaying of the associations between rows and columns of a contingency table in a low-dimensional space in such a way that the underlying lexical patterns can be brought to the foreground. The closer the points (NPs, but also disciplines) are on the map, the more related they are. In other words, if two disciplines are very close in terms of projection (the direction from the origin), they have similar acknowledgement patterns. It should be noted that the proximity between NPs and disciplines is measured by χ^2 distances, meaning that the weight assigned to each element of the matrix is inversely proportional to the row and column marginal totals. Since WoS only indexes acknowledgements that include funding information, NPs related to funding are expected to be very frequent, whereas NPs related to credit for peer interactive communication, intellectual indebtedness, or any other forms of contribution will have lower frequencies. These lower-frequencies NPs will be highlighted by the CA method, revealing the hidden infrastructure that supports scientific research.

To perform the CA, NPs were grouped using k-means clustering with cosine similarity. A threshold of 2,000 occurrences was established (i.e., NPs that appear at least 2,000 times in total were included). Thresholds of 2,500 and 3,000 occurrences were also tested, and the disciplines that strongly contribute to each factor were very similar for the various threshold values. The 2,000-occurrences threshold corresponds to the 214 most frequent NPs of the dataset, which account for 74% of the total NP occurrences for the whole corpus (see Supporting Information for the frequency table and plot of the 214 NPs included in the analysis). The resulting matrix of 214 NPs x 12 disciplines was then analyzed. As shown in Table 1, by retaining five axes, 88% of the variance is explained. As factors obtained are uncorrelated, the explained variance for a NP or discipline on a particular plane is the sum of the contributions to the axes forming this plane, and this specific value is known as ‘quality of representation’ (QLR). For the five axes shown in Table 2, there is a good QLR for most NPs and disciplines (see Supporting Information for detailed QLR by NP and by discipline). In order to test the robustness of our results, an alternative analysis was performed using relative thresholds normalized by discipline, instead of an absolute threshold of 2,000 occurrences. The CA was rerun using the 100 most frequent NPs by discipline (i.e., 287 NPs in total). Using this alternative dataset, extremely similar clusters were obtained. However, the explained variance of the resulting

axes were weaker, which led us to keep the original method of selecting NPs by threshold of occurrences.

Table 1. – Explained and cumulative variance for each axis

	Explained Variance (%)	Cumulative Variance (%)
Axis 1	36.56	36.56
Axis 2	19.00	55.56
Axis 3	15.79	71.35
Axis 4	10.48	81.83
Axis 5	6.20	88.03

Table 2 presents the percentage of the variance explained in each axis for the 12 disciplines. QLR is shown on a scale of 0-100 points and only disciplines and NPs with at least 40% of the information accounted for by a particular plane ($QLR \geq 40$) were selected for the visualization. Such a threshold ensures that the lexical patterns observed are representative of the discipline. In the cases of Biomedical Research, Chemistry, Clinical Medicine, Engineering and Technology, Physics, Professional Fields and Social Sciences, the first plane (formed by Axes 1 and 2) explains approximately 60% of the variance, while for Earth and Space and Psychology, Axes 1 and 2 explain less than 50% of the variance. However, for Biology (which is the most dispersed discipline) and Mathematics, the variance is mostly explained in a residual plane (formed by Axes 3 and 4), suggesting a different lexical pattern than the one found in the first plane. In the case of Biology, none of the axes explain particularly well the acknowledgement patterns of the discipline, which could be explained by its heterogeneous distribution and by the specific types of assistance acknowledged in Biology.

Table 2. – Relative contributions of the factor to the element for disciplines (expressed as a percentage)

Column	Discipline	Axis1	Axis2	Axis3	Axis4	Axis5
1	Biology	8.2%	12.9%	11.9%	28.6%	0.9%
2	Biomedical Research	57.8%	16.6%	24.4%	0.2%	0.8%
3	Chemistry	26.8%	36.7%	14.4%	1.8%	3.9%
4	Clinical Medicine	55.8%	10.0%	30.6%	0.0%	2.8%
5	Earth and Space	30.4%	17.6%	17.4%	15.7%	7.1%
6	Engineering and Technology	58.1%	10.4%	8.3%	1.8%	1.1%
7	Health	10.3%	30.9%	9.5%	1.7%	19.7%
8	Mathematics	12.9%	2.4%	6.5%	63.2%	11.1%
9	Physics	39.4%	26.0%	3.8%	0.3%	1.8%
10	Professional Fields	16.3%	42.2%	2.6%	9.8%	23.7%
11	Psychology	2.5%	42.9%	0.1%	4.7%	10.9%
12	Social Sciences	14.9%	40.2%	4.9%	3.5%	26.7%

Results

Table 3 details the presence of acknowledgements in the 2015 papers indexed in WoS and which, it bears repeating, must contain funding information in order to be included. It shows that acknowledgements are not evenly distributed across disciplines. While the large majority of papers from the natural and medical sciences indexed in WoS contain funding acknowledgements—with a percentage above 80% for Biomedical Research and Chemistry—less than half of papers in the social sciences present acknowledgements with the required funding information.

Table 3. – Number of papers indexed in WoS (all and with funding acknowledgements) and percentage of papers with funding acknowledgements, by discipline (2015)

Discipline	All Papers	Papers with Acknow.	% with Acknow.
Biomedical Research	189,066	158,067	83.6%
Chemistry	151,947	123,806	81.5%
Earth and Space	92,238	72,922	79.1%
Physics	124,556	95,676	76.8%
Biology	105,279	76,281	72.5%
Mathematics	49,997	35,390	70.8%
Engineering and Technology	241,124	165,590	68.7%
Clinical Medicine	389,311	218,367	56.1%
Health	37,309	18,703	50.1%
Psychology	31,286	15,085	48.2%
Social Sciences	50,420	16,972	33.7%
Professional Fields	41,015	12,552	30.6%
Total	1,503,548	1,009,411	67.1%

Figure 1 and 2 present the bidimensional CA of acknowledgement patterns by discipline. Six clusters can be distinguished. The first cluster, formed by Chemistry, Physics and Engineering and Technology, is marked by the acknowledgement of technical support with NPs referring to specific technical tasks (e.g., *image, equipment, computational resource, measurement, code, calculation*). NPs related to peer discussion (e.g., *fruitful discussion, helpful discussion, valuable discussion*) and to funding (e.g., *financial assistance, financial support, partial financial support*) also appear prominently in these natural sciences disciplines. Three main themes for these disciplines are therefore established as: technical tasks, PIC and funding.

In the second cluster, comprised of Clinical Medicine, Health and Psychology, NPs referring to funding and disclosure of potential conflict of interest (or lack of thereof) are dominant (e.g., *fee, honorarium, conflict of interest, consultant, employee, financial conflict, financial interest, financial involvement, sponsor, other relevant affiliation*). NPs specifically related to the experimental work involving the participation of human subjects

also appear in the acknowledgement patterns of that cluster (e.g., *family, patient, participant, participation, trial*). Here, the main themes are funding-related ethics, and human subjects, which define the fields. Interestingly, NPs associated with Psychology, a discipline often difficult to categorize, are located halfway between the applied health cluster (cluster 2) and the Social Sciences and Professional Fields (see cluster 4).

The third cluster consists of Biomedical Research and is mainly characterized by NPs referring to funding (e.g., *funder, grant sponsor, fund, fellowship, studentship, recipient*). There is also a prevalence of terms related to performing the research (e.g., *analysis, data collection, preparation, technical assistance, excellent technical assistance*). This cluster therefore sits at the juncture of finance and analysis.

Earth and Space, Professional Fields and Social Sciences are all in the fourth cluster. It is the cluster where peer communication and intellectual debt are clearly the dominant types of contributions acknowledged, with numerous NPs referring to input by reviewers, editors, and colleagues (e.g., *guidance, feedback, valuable suggestion, useful comment, helpful comment, valuable comment, insightful comment, editor, reviewer, anonymous reviewer, anonymous referee*). The presence of Earth and Space in that cluster illustrates its more heterogeneous profile in terms of acknowledgements, with information scattered on different axes, in a manner similar to what is found in Biology (see Table 3). This PIC- and peer-review-centric cluster is thus one that reaches across very different fields.

The fifth cluster consists of Biology (Figure 2) and is characterized by NPs related to the specific nature of experimental work in that field (e.g., *field, fieldwork, access, sample, laboratory, logistical support, technical support*). Furthermore, collaboration appears as an important aspect of acknowledgements in Biology with NPs such as *cooperation, crew, team, and staff*. Finally, Mathematics is found as a peripheral cluster (cluster 6) and is mainly characterized with NPs referring to authorship (e.g., *first author, second author, third author*) and the presence of the NP 'referee'. These two fields hence stand alone, but Biology shows a potential for more specific comparisons with Clusters 1 and 3 in terms of technical work or analysis, and clusters 1 and 4 in terms of collaboration-

related terms. Mathematics is the eccentric field, but shows the importance of collaboration, assigning individual responsibilities or funding to specific authors.

While we could have expected that NPs referring to the peer review process would be scattered across all disciplines (since peer review is a common denominator in all fields), acknowledgements directly thanking referees, reviewers and editors appear as a distinctive feature of the disciplines in the fourth cluster (Earth and Space, Professional Fields and Social Sciences) and in Mathematics, which stands alone as cluster 6. In the same way, NPs referring to funding take a different form in the second cluster—which regroups applied health disciplines (Clinical Medicine, Health and Psychology)—where ethical considerations and disclosure of conflict of interest appear as a dominant trend, while they did not appear as prevalent in the other clusters. In this language-based analysis, form and wording appear to have a capital influence. Indeed, the absence of given themes or topics in a cluster does not mean that said topics and themes are absent from a discipline altogether; it does, however, mean that the patterns formed by the NPs included in the analysis favour other, stronger trends in terms of disciplinary practices. Furthermore, variations in the wording of similar themes across disciplines, such as funding and PIC, could certainly be analysed further to see where the parallels between disciplinary cultures begin and end, as well as the weight of formulaic or required statements.

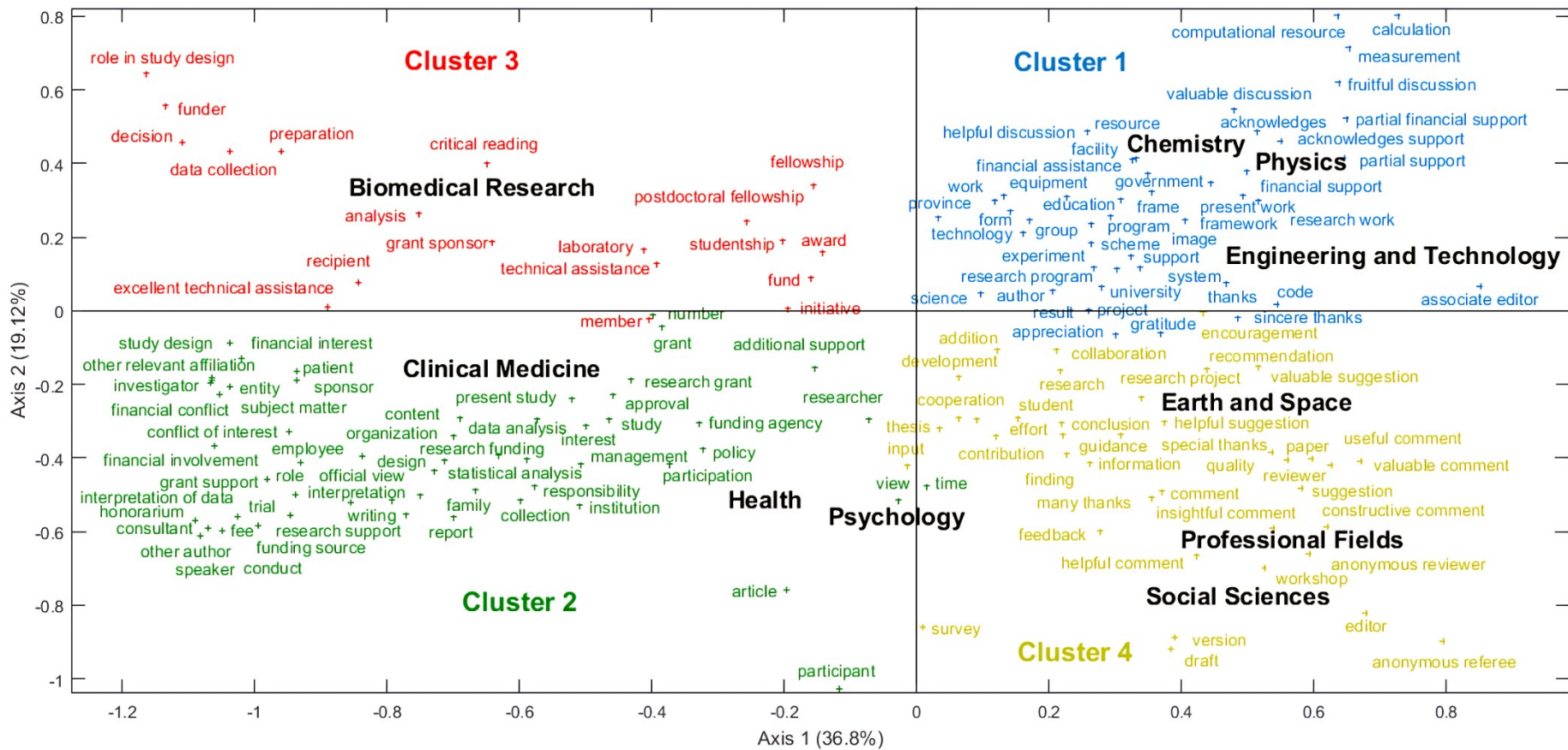


Figure 1. – Bidimensional Correspondence Analysis for acknowledgements patterns by discipline (plane 1-2)

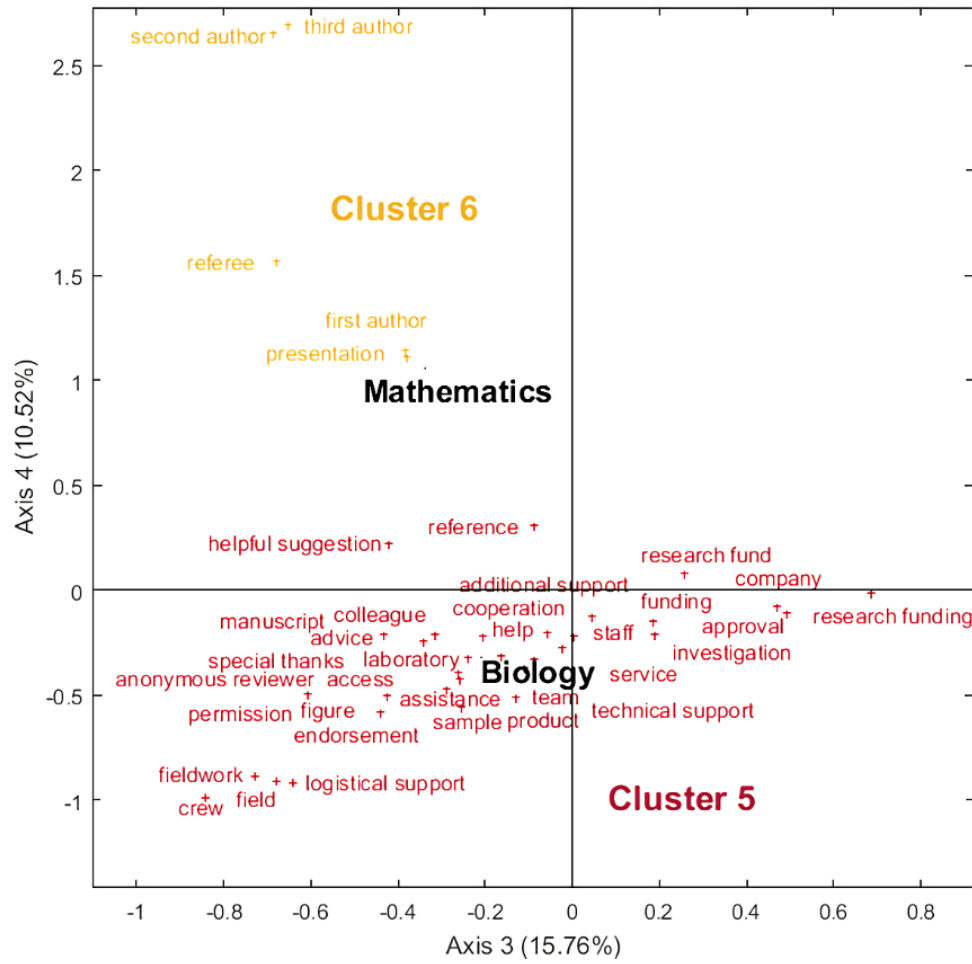


Figure 2. – Bidimensional Correspondence Analysis for acknowledgements patterns by discipline (plane 3-4)

Discussion and Conclusion

As one might expect from WoS acknowledgement indexation—which, again, only includes acknowledgements if funding information is provided and if they are written in English—our results show that the proportion of papers with acknowledgements varies across disciplines, with a higher proportion of papers containing funding acknowledgements in the biomedical sciences (>80%), followed by the natural sciences (from 70% to 80%), clinical medicine (\cong 50%) and the social sciences (>30%). These results are in line with previous findings (Costas & van Leeuwen, 2012; Díaz-Faes & Bordons, 2014; Paul-Hus, Desrochers & Costas, 2016; Xu, Tan & Zao, 2015).

As often discussed in the literature, acknowledgements found in scholarly papers provide a window on the otherwise invisible contributions made to research by individuals and organisations. Previous research has shown that these contributions—which are not considered sufficient to grant authorship, key to the accumulation of capital in modern science (Bourdieu, 1975; Cronin, 2005)—have traditionally been grouped into similar types of categories since Cronin et al. (1993) and McCain (1991) proposed their taxonomies: conceptual and cognitive, financial support, access to data and materials, technical assistance, and manuscript preparation. A large body of research on acknowledgements has been published since these foundational models were proposed (see Desrochers, Paul-Hus & Pecoskie, 2017, for a meta-synthesis of the literature), and while there exists a sizeable amount of large-scale analyses of authorship and collaboration, their evolution over time, and the variations they exhibit across disciplines (e.g. Henriksen, 2016; Larivière et al., 2015; Ponomariov & Boardman, 2016; Wuchty, Jones & Uzzy, 2007), no multidisciplinary large-scale study had previously analysed how acknowledged contributions vary across the fields of biomedical, natural and social sciences. Our quantitative, data-driven analysis offers a large-scale overview of the main disciplinary patterns found in terms of types of contributions acknowledged; this provides insight on the tasks and work valued in each field, the practices in acknowledging them, the trends they create with regard to how things are expressed, the importance given to certain aspects of research such as the peer review process or ethics-related statements, as well as on the division of labor, in certain cases. Our results are of particular interest in the case of the

social sciences, since previous WoS-based studies, including our own (Paul-Hus et al., 2016), were mainly limited to SCI-E, as funding acknowledgement data for SSCI have only been collected since 2015. By applying advanced linguistic methods as well as Correspondence Analysis to 1,009,411 acknowledgements from papers published in 2015, this study contributes to provide a better understanding of acknowledgement practices by reaffirming but also expanding the main types of contributions found in the traditional taxonomies of acknowledgements proposed in 1990s (Cronin et al., 1993; McCain, 1991). More importantly, our analysis highlights important disciplinary variations in the practices, trends and etiquette of acknowledging, all of which are direct reflections of different disciplinary cultures in research, collaboration, and scholarly communication itself.

Technical support was more frequently acknowledged by scholars in the natural sciences (Chemistry, Physics and Engineering). Earth and Space, Professional Fields, and Social Sciences were more likely to acknowledge contributions from colleagues, editors, and reviewers, and Biology acknowledgements put more emphasis on logistics and other fieldwork-related tasks. While Biomedical Research mostly acknowledged funding—which might be a reflection of the larger spectrum of contributions that lead to authorship in that specific discipline, and thus to funding disclosures from more authors (Pontille, 2004)—Clinical Medicine, Health and, to a lesser extent, Psychology, were much more likely to include statements related to various forms of conflicts of interest. Conflicts of interest constitute important issues in clinical studies, given the potential consequences of fraud and unethical behaviour in these fields.

This suggests that acknowledgements are not confined to credit attribution, such as the traditional acknowledgement taxonomies seem to indicate (Cronin et al., 1993; McCain, 1991). In fact, disclosures of potential conflicts of interest show the presence of pre-formulated statements recommended by funding bodies, ethical boards, or editorial requirements may influence language-based analyses of acknowledgements and reveal important disciplinary requirements and practices. Expressions associated with conflict of interest—or lack thereof—were mostly found in the clinical and applied health fields, in which disclosure of potential conflicts is made mandatory by most journals' guidelines, as established by the largest consortium of medical journal editors, the International

Committee of Medical Journal Editors (ICMJE, 2015). These (self-)imposed statements obviously protect third parties, including the journals themselves, as well as the researchers. Interestingly, these traces of accountability add to the use of acknowledgements alongside authorship lists to attribute—or remove—responsibility in terms of the results published.

This leads to the importance of interpreting NPs in light of their original context, which the preliminary findings from ongoing qualitative analyses already show. These reveal interesting aspects to these NPs: keeping the example of conflict of interest disclosure, the NP ‘role in study design’ may not be the reflection of a task performed, but rather an indication that it was not, as in the sentence, “The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript” (Pasiarski et al., 2015). Indeed, that NP is often associated with conflict of interest disclosure statements recommended by PLOS journals, PeerJ, and other scholarly communication venues. The NP ‘speaker’ is also linked to conflict of interest issues in Clinical Medicine, as in this typical example: “JMH has acted as a consultant, received grants, and acted as a speaker in activities sponsored by Astra-Zeneca, Eli Lilly and Company, Glaxo-SmithKline, and Lundbeck” (Novick et al., 2015); here, the ‘speaker’ is the researcher and the use of the NP is related to disclosure. However, ‘speaker’ can also be used to describe a person other than the researcher in order to acknowledge their contribution, as in this example: “We would like to thank the following student research assistants: [...] native speaker of English Stewart Campbell for proofreading the manuscript” (Hefter et al., 2015); in such cases, the use of the NP shifts back to collaboration and support. The qualitative coding of NPs in the context of their use constitutes the next phase of the project and will complement the findings presented here by adding various layers to our understanding of scientific acknowledgements. As this qualitative analysis calls for a reporting style based on thick description, it is best suited for another paper due to the time and space it requires, but it constitutes further research worth pursuing, as it will allow us to measure the relative weight of funding-related and non-funding-related uses of the NPs found in acknowledgements. This will in turn provide us with a better understanding of how acknowledgements indexed in WoS can support the analysis of scientific practices beyond the core concern with funding.

Papers' acknowledgements provide much more than funding information. They shed light on otherwise invisible contributions that complement authorship and provide insight on researchers' collaboration patterns, division of labour, and credit attribution practices. Furthermore, acknowledgement structures, lexicons, and uses do vary by discipline, and so this paratext can reveal much about researchers' practices, the context in which they conduct research, and the specific aspects of the scholarly work they value or deem worth mentioning to protect themselves or those who support them. In many instances, acknowledgements remain the only space where such revealing details about specific and disciplinary research contexts are recorded. More forays into the role played by acknowledgements as testimonies to these practices, beyond funding, are needed to better understand how they can be used to illustrate disciplinary issues, differences, and similarities.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insightful suggestions. This research was supported by the Social Sciences and Humanities Research Council of Canada, Joseph-Armand Bombardier CGS Doctoral Scholarships, Paul-Hus; Insight Development [grant number 430-2014-0617], Larivière and Desrochers; as well as by funding from the South African DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), Rodrigo Costas.

References

- Beh E.H. & Lombardo R. (2014) *Correspondence analysis: Theory, practice and new strategies*. Chichester, UK: John Wiley & Sons.
- Benzécri, J.P. (1992). *Correspondance analysis handbook*. New York: Marcel Dekker.
- Bird, S., Klein, E. & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*. Sebastopol, CA: O'Reilly Media.
- Birnholtz, J.P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, 57(13), 1758-1770.

<https://doi.org/10.1002/asi.20380>

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, 14(6), 19-47.

Costas, R. & van Leeuwen, T.N. (2012). Approaching the « reward triangle »: General analysis of the presence of funding acknowledgments and « peer interactive communication » in scientific publications. *Journal of the American Society for Information Science and Technology*, 63(8), 1647 -1661.

<https://doi.org/10.1002/asi.22692>

Crawford, E.T. & Biderman, A.D. (1970). Paper money: Trends of research sponsorship in American sociology journals. *Social Science Information*, 9(1), 50-77.

<https://doi.org/10.1177/053901847000900103>

Cronin, B. (2005). *The hand of science: Academic writing and its rewards*. Lanham, Maryland: Scarecrow Press.

Cronin, B., McKenzie, G., Rubio, L. & Weaver-Wozniak, S. (1993). Accounting for influence: Acknowledgments in contemporary sociology. *Journal of the American Society for Information Science*, 44(7), 406-412. [https://doi.org/10.1002/\(SICI\)1097-4571\(199308\)44:7<406::AID-ASI6>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4571(199308)44:7<406::AID-ASI6>3.0.CO;2-8)

Cronin, B., Shaw, D. & La Barre, K. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, 54(9), 855-871.

<https://doi.org/10.1002/asi.10278>

Cronin, B., Shaw, D. & La Barre, K. (2003). Visible, less visible, and invisible work: Patterns of collaboration in 20th century chemistry. *Journal of the American Society for Information Science and Technology*, 55(2), 160-168.

<https://doi.org/10.1002/asi.10353>

Cronin, B. & Weaver, S. (1995). The praxis of acknowledgement: From bibliometrics to influmetrics. *Revista Española de Documentación Científica*, 18(2), 172–177.

- Desrochers, N., Paul-Hus, A. & Larivière, V. (2016). The angle sum theory: Exploring the literature on acknowledgments in scholarly communication. In C. R. Sugimoto (ed), *Theories of Informetrics and Scholarly Communication*. Berlin: de Gruyter Mouton (p. 225-247). <https://doi.org/10.1515/9783110308464>
- Desrochers, N., Paul-Hus, A. & Pecoskie, J. (2017). Five decades of gratitude: A meta-synthesis of acknowledgments research. *Journal of the Association for Information Science and Technology*, 68(12), 2821–2833. <https://doi.org/10.1002/asi.23903>
- Díaz-Faes, A.A. & Bordons, M. (2014). Acknowledgments in scientific publications: Presence in Spanish science and text patterns across disciplines. *Journal of the Association for Information Science and Technology*, 65(9), 1834 -1849. <https://doi.org/10.1002/asi.23081>
- Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of Science*, 17(2), 102-134.
- Genette, G. (1997). *Paratexts: Thresholds of interpretation* (Jane E. Lewin, trans.). Cambridge: Cambridge University Press.
- Giles, C.L. & Councill, I.G. (2004). Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17599-17604. <https://doi.org/10.1073/pnas.0407743101>
- Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). London: Chapman & Hall/CRC.
- Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S. & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction*, 39, 11-22. <https://doi.org/10.1016/j.learninstruc.2015.05.002>
- Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455-476. <http://doi.org/10.1007/s11192-016-1849-x>

- Hirsch, W.W. & Singleton, F. (1965). Research support, multiple authorship, and publications in sociological journals, 1932-1964. Unpublished preprint.
- Khabsa, M., Koppman, S. & Giles, C.L. (2012). Towards building and analyzing a social network of acknowledgments in scientific and academic documents. In S.J. Yang, A.M. Greenberg & M. Endsley (eds), *Social Computing, Behavioral - Cultural Modeling and Prediction* (p. 357-364). Berlin: Springer.
https://link.springer.com/chapter/10.1007/978-3-642-29047-3_43
- Khabsa, M., Treeratpituk, P. & Giles, C.L. (2012). AckSeer: A repository and search engine for automatically extracted acknowledgments from digital libraries. In K. B. Boughida & B. Howard (eds), *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (p. 185-194). New York, NY, ACM.
<http://doi.acm.org/10.1145/2232817.2232852>
- Kim, S.N., Baldwin, T. & Kan, M.Y. (2010). Evaluation N-gram based evaluation metrics for automatic keyphrase extraction. In C. Huang & D. Jurafsky (eds), *Proceedings of the 23rd International Conference on Computational Linguistics* (p. 572-580). Beijing: Tsinghua University Press.
- Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A. & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417-435. <https://doi.org/10.1177/0306312716650046>
- Larivière, V., Gingras, Y., Sugimoto, C.R. & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7), 1323-32.
<https://doi.org/10.1002/asi.23266>
- Lewison, G. & Markusova, V. (2010). The evaluation of Russian cancer research. *Research Evaluation*, 19(2), 129-144. <https://doi.org/10.3152/095820210X510098>
- Lukas, J., Pockrandt, A. M., Seennann, S., Sharif, M., Runge, F., Pohlers, S., Zheng, C, Gläser, A., Beller, M., Rolfs, A. & Giese, A.K. (2015). Enzyme enhancers for the

treatment of Fabry and Pompe Disease. *Molecular Therapy*, 23(3), 456-64.
<https://doi.org/10.1038/mt.2014.224>

Marcus, M.P., Marcinkiewicz, M.A. & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313-330.

Marcus, M.P., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K. & Schasberger, B. (1994). The Penn Treebank: Annotating predicate structure. In *Proceedings of the workshop on Human Language Technology* (p. 114-119). Plainsboro NJ; Association for Computational Linguistics.

McCain, K. W. (1991). Communication, competition, and secrecy: The production and dissemination of research-related information in genetics. *Science, Technology & Human Values*, 16(4), 491–516. <http://doi.org/10.1177/016224399101600404>

Marušić, A., Bošnjak, L. & Jerončić, A. (2011). A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS ONE*, 6(9), e23477. <http://doi.org/10.1371/journal.pone.0023477>

Novick, D., Hong, J., Montgomery, W., Duenas, H., Gado, M. & Haro, J.M. (2015). Predictors of remission in the treatment of major depressive disorder: Real-world evidence from a 6-month prospective observational study. *Neuropsychiatry Disease and Treatment*, 11, 197-205. <http://doi.org/10.2147/NDT.S75498>

Pasiarski, M., Grywalska, E., Kosmaczewska, A., Gozdz, S., Steckiewicz, P., Garus, B., Bilski, M., Hymos, A. & Rolinski, J. (2015). Assessment of peripheral blood and bone marrow T, NK, NKT and dendritic cells in patients with multiple myeloma. *Postepy Higieny I Medycyny Doswiadczalnej*, 69, 1435-42.

Patel, N. (1973). Collaboration in the professional growth of American sociology. *Social Science Information*, 12(6), 77 -92.

Paul-Hus, A., Mongeon, P., Sainte-Marie, M. & Larivière, V. (2017). The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements. *Journal of Informetrics*, 11(1), 80-87. <http://doi.org/10.1016/j.joi.2016.11.005>

- Paul-Hus, A., Díaz-Faes, A.A., Desrochers, N., Costas, R., Sainte-Marie, M., Macaluso, B. & Larivière, V. (2016). Beyond funding: What can acknowledgements reveal about credit distribution in science? In I. Rafols, J. Molas-Gallart, E. Castro-Martinez & R. Woolley (eds), *Proceedings of the 21st International Conference on Science and Technology Indicators*. València: Universitat Politècnica de València.
- Paul-Hus, A., Desrochers, N. & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, 108(1), 167-182. <https://doi.org/10.1007/s11192-016-1953-y>
- Ponomariov, B. & Boardman, C. (2016). What is co-authorship? *Scientometrics*, 109(3), 1939-1963. <https://doi.org/10.1007/s11192-016-2127-7>
- Pontille, D. (2004). *La signature scientifique: Une sociologie pragmatique de l'attribution*. Paris: CNRS Éditions.
- Pouget, J.G., Goncalves, V.F., Nurmi, E.L., Laughlin, C.P., Mallya, K.S., McCracken, J.T., Aman, M. G., McDougale, C.J., Scahill, L., Misener, V.L., Tiwari, A.K., Brandl, E.J., Felsky, D., Leung, A.Q., Lieberman, J.A., Meltzer, H.Y., Potkin, S.G., Nielding, C., Steimer, W.... Kennedy, J.L. (2015). Investigation of TSPO variants in schizophrenia and antipsychotic treatment outcomes. *Pharmacogenomics*, 16(1), 5-22. <https://doi.org/10.2217/pgs.14.158>
- Rigby, J. (2013). Looking for the impact of peer review: does count of funding acknowledgements really predict research impact? *Scientometrics*, 94(1), 57-73. <https://doi.org/10.1007/s11192-012-0779-5>
- Salager-Meyer, F., Alcaraz-Ariza, M.Á., Luzardo-Briceño, M. & Jabbour, G. (2010). Scholarly gratitude in five geographical contexts: a diachronic and cross-generic approach of the acknowledgment paratext in medical discourse (1950–2010). *Scientometrics*, 86(3), 763 -784. <https://doi.org/10.1007/s11192-010-0329-y>
- Sarbova, V., Koschella, A., Cheng, F., Kelly, S.M. & Heinze, T. (2015). Studies on the sulfation of cellulose alpha-lipoate and ability of the sulfated product to stabilize

- colloidal suspensions of gold nanoparticles. *Carbohydrate Polymers*, 124, 117-23.
<https://doi.org/10.1016/j.carbpol.2015.01.080>
- Tang, L., Hu, G. & Liu, W. (2017). Funding acknowledgment analysis: Queries and caveats. *Journal of the Association for Information Science and Technology*, 68(3), 790-794. <https://doi.org/10.1002/asi.23713>
- Toutatnova, K., Klein, D., Manning, C. & Singer, N. (2003). Feature-rich Part-Of-Speech Tagging with a Cyclic Dependency Network. In M. Hearst & M. Ostendorf (eds) *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology* (p. 252-259). Stroudsburg, PA: Association for Computational Linguistics.
- Toutanova, K. & Manning, C. (2000). Enriching the Knowledge Sources Used in a Maximum Entropy Part-Of-Speech Tagger. In H. Schiitze & K-Y Su (eds) *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora* (p. 63-70). Stroudsburg, PA: Association for Computational Linguistics.
- Vicente-Villardón, J.L. (2014). MultBiplot: A package for multivariate analysis using biplots. Departamento de Estadística, Universidad de Salamanca [Software].
<http://biplot.usal.es/ClassicalBiplot/index.html>
- Wuchty, S., Jones, B.F. & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039.
<https://doi.org/10.1126/science.1136099>
- Xu, X., Tan, A.M. & Zhao, S.X. (2015). Funding ratios in social science: the perspective of countries/territories level and comparison with natural sciences. *Scientometrics*, 104(3), 673-684. <https://doi.org/10.1007/s11192-015-1633-3>
- Zhang, R., Li, R., Wang, J., Wang, S., Zhang, M., Hu, X., Zhang, L., Wang, S., Wang, R. & Bao, Z. (2015). Identification, characterization and expression profiling of the Tollip gene in Yesso scallop (*Patinopecten yessoensis*). *Genes & Genetic Systems*, 90(2), 99-108. <https://doi.org/10.1266/ggs.90.99>

Zhou, P. & Tian, H. (2014). Funded collaboration research in mathematics in China. *Scientometrics*, 99(3), 695 -715. <https://doi.org/10.1007/s11192-013-1212-4>

Supporting Information

Figure 3. – Frequency distribution of noun phrases found in acknowledgements

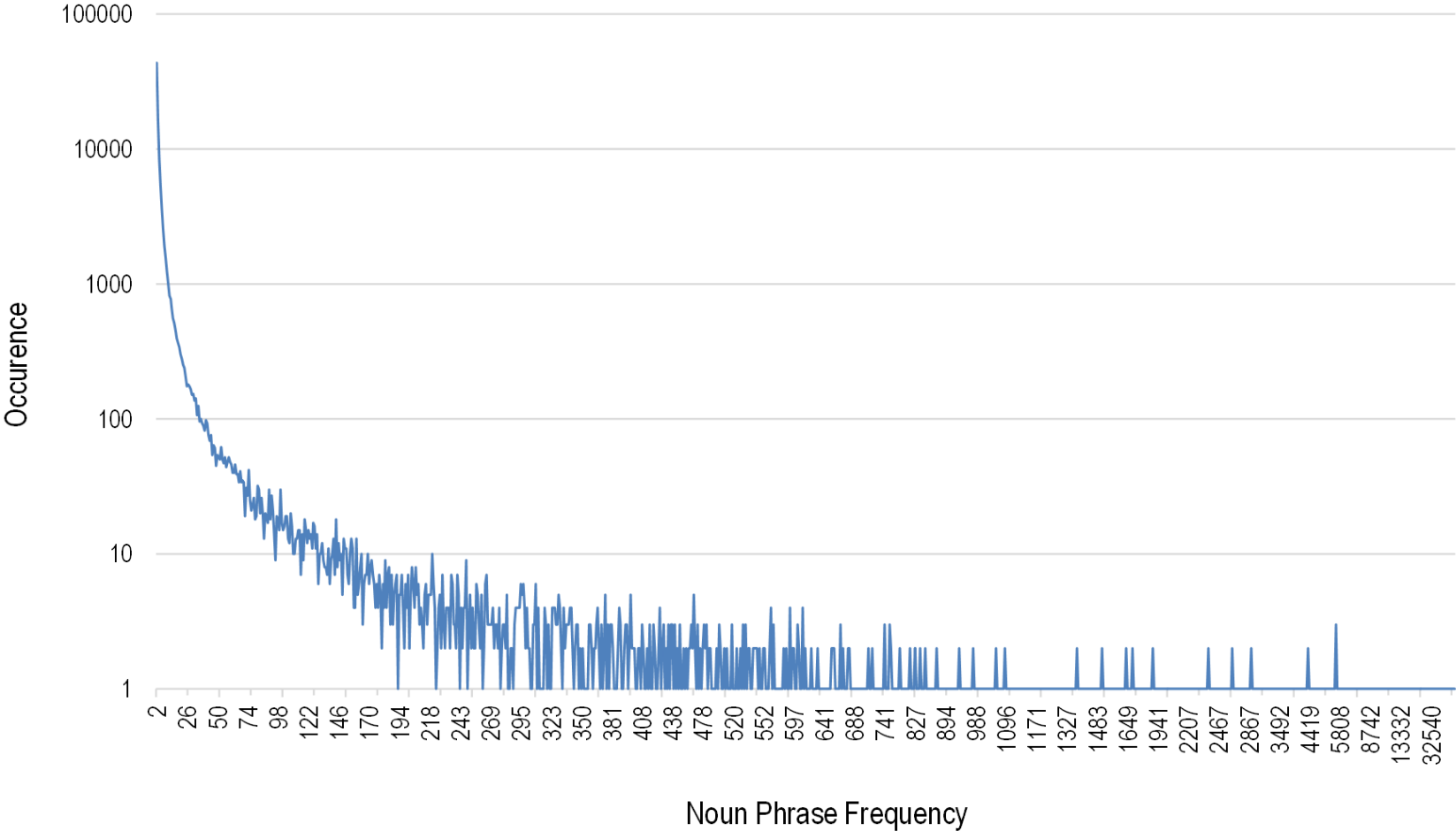


Tableau 5. – Frequency of the 214 most frequent noun phrases, by discipline

Rank	Noun Phrases	Biomedical Biology	Research Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total	
1	work	27038	83705	57640	79759	32782	94048	4609	14445	62636	3181	4202	3644	467689
2	author	23900	35134	39828	58117	27356	63986	7951	13460	26270	6351	4858	7308	314519
3	grant	14399	52953	12223	69030	16819	16372	3837	8079	9172	1756	4648	2123	211411
4	study	19381	32009	7624	73724	16214	10583	8271	470	3790	1962	4350	1992	180370
5	research	17981	20857	14714	25729	19783	28703	5506	9724	12642	5195	6791	7244	174869
6	project	17342	19255	17017	19069	19593	28145	2799	4858	14901	2243	1936	3520	150678
7	financial support	10617	10738	38990	11388	9460	31349	2119	1929	13712	2441	1037	4212	137992
8	support	8256	14795	15017	14852	17204	21579	2078	3488	16727	1768	1376	2711	119851
9	manuscript	16271	36332	2727	18837	15135	4266	1548	1647	3064	654	2013	1022	103516
10	paper	3895	2889	1355	5658	12175	13837	1214	4601	3795	2622	849	4446	57336
11	program	3947	7625	7036	6155	6699	8806	760	1112	6676	379	422	677	50294
12	analysis	2119	25144	2380	10872	2788	1751	1439	669	606	230	630	435	49063
13	funding	4448	7499	4758	8614	4203	4958	1271	460	2734	506	590	946	40987
14	preparation	1018	23015	673	6575	809	599	536	797	387	137	622	134	35302
15	assistance	6159	5438	3774	5829	4985	3711	583	75	1559	320	744	638	33815
16	anonymous reviewer	7538	1720	331	679	13006	3838	294	670	765	1382	608	2261	33092
17	data collection	1067	22639	307	5760	444	189	719	590	64	179	706	195	32859
18	help	5872	5516	3793	3914	5243	4147	258	290	2212	273	756	497	32771
19	data	2399	3657	1022	6986	11931	1926	1030	192	843	553	915	1086	32540
20	decision	500	22613	150	7125	210	201	616	617	69	63	319	56	32539
21	thanks	4728	2515	6049	2125	4906	3871	370	640	3034	461	413	1057	30169
22	article	1316	3497	632	8433	1720	1772	2976	741	474	1780	1845	3583	28769
23	funder	470	21115	75	4264	111	60	632	590	10	70	163	66	27626

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
24	role in study design	355	19967	54	3275	61	44	220	579	6	17	115	9	24702
25	comment	4003	2741	426	1583	5672	1818	290	1000	1533	862	532	1979	22439
26	fellowship	2237	5690	4355	4052	1115	1496	157	315	1256	41	219	171	21104
27	fund	2142	4250	2086	4700	1438	2299	330	387	1124	160	234	205	19355
28	suggestion	2234	1456	609	828	4866	2615	174	1921	1256	1004	250	1743	18956
29	contract	816	2228	1777	2581	4756	2972	264	405	1951	73	100	107	18030
30	number	1358	4306	1149	6272	886	1882	308	327	787	171	232	281	17959
31	publication	1075	1963	537	4050	1617	5018	711	141	1062	164	283	199	16820
32	helpful discussion	702	3282	2770	1390	1766	1691	26	494	3338	77	112	168	15816
33	research grant	1105	2660	992	6892	802	1288	331	311	588	247	327	256	15799
34	technical assistance	2687	4706	1355	3981	820	1001	52	10	594	14	161	43	15424
35	member	1387	4379	626	3949	1079	1078	390	242	719	201	362	275	14687
36	contribution	1690	1711	838	2913	3277	1694	556	108	748	257	431	348	14571
37	discussion	956	2639	1076	908	2754	1366	45	435	3670	135	117	423	14524
38	helpful comment	2272	1399	316	988	2921	1087	231	924	810	991	546	1875	14360
39	staff	3225	2004	553	3211	2102	813	523	36	674	135	375	304	13955
40	material	1382	1868	1389	2694	1174	2765	274	295	1118	243	216	328	13746
41	view	784	1312	337	3528	1278	1929	1101	171	706	731	601	1050	13528
42	conflict of interest	834	3081	111	7995	139	92	646	11	25	39	327	32	13332
43	content	378	3035	307	5801	530	513	1113	177	151	234	586	219	13044
44	scholarship	2529	2257	1812	2197	987	1621	209	172	575	61	194	147	12761
45	laboratory	2118	4735	833	2420	1087	771	39	26	370	25	127	51	12602
46	facility	1322	1953	2785	1073	1547	2081	46	52	1504	30	51	65	12509
47	experiment	1904	2089	1722	1580	889	2044	30	29	1507	65	356	147	12362
48	use	999	1598	1826	1010	3619	1755	83	58	1086	69	86	111	12300
49	result	599	1337	1128	2011	2027	2168	242	421	1295	158	146	337	11869

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
50	first author	1490	431	181	810	1292	1964	299	3324	396	454	727	310	11678
51	award	774	2548	1875	2712	1078	1115	228	216	607	63	213	108	11537
52	sponsor	482	3271	46	6568	252	398	163	9	56	30	79	49	11403
53	access	2509	1523	1395	719	2264	1099	126	42	666	114	103	328	10888
54	framework	784	988	1479	779	1376	2602	54	279	1758	93	60	214	10466
55	review	659	960	209	2734	1071	3160	270	22	1103	48	116	61	10413
56	valuable comment	1898	705	269	468	2124	1731	96	992	728	475	178	671	10335
57	technical support	1415	2091	1187	1883	1003	1547	62	25	879	23	111	59	10285
58	anonymous referee	1128	235	75	84	2453	954	81	1663	554	883	149	1964	10223
59	editor	1128	352	99	281	2600	1355	166	844	233	866	162	1414	9500
60	constructive comment	1423	513	141	278	4071	1137	55	451	301	342	110	472	9294
61	reviewer	1089	486	434	447	2791	1803	128	632	425	370	195	483	9283
62	education	745	1188	1598	1144	775	1977	75	353	920	100	54	79	9008
63	collection	1338	1178	191	4459	572	110	473	13	25	75	305	155	8894
64	responsibility	168	1448	291	3968	298	476	866	93	161	234	510	366	8879
65	useful discussion	196	712	892	207	1555	875	6	282	4012	21	20	58	8836
66	research project	1026	763	666	1149	1064	1857	222	286	537	402	190	580	8742
67	special thanks	1833	842	560	909	1791	1207	198	107	374	226	248	404	8699
68	measurement	298	554	3000	332	817	2018	5	1	1216	1	8	17	8267
69	version	1942	579	50	585	1633	425	221	355	162	594	471	1170	8187
70	recipient	561	3113	383	3380	198	178	84	27	77	17	56	19	8093
71	grant sponsor	206	2846	1186	2867	192	105	1	247	26	39	149	126	7990
72	gratitude	912	627	846	978	1019	1704	183	337	568	172	138	259	7743
73	advice	1376	1785	408	1118	1061	630	108	124	386	128	143	294	7561
74	role	152	1674	50	4287	186	60	788	15	24	52	214	56	7558
75	writing	260	1210	81	4171	210	121	747	60	72	91	271	163	7457

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
76	participant	240	686	33	2147	451	226	699	78	143	763	545	1368	7379
77	official view	113	1288	208	3311	188	184	741	65	108	100	454	113	6873
78	fruitful discussion	235	678	1084	218	812	1093	4	200	2374	10	27	34	6769
79	opinion	404	580	274	1077	717	1455	373	111	556	476	299	443	6765
80	study design	132	2605	32	3119	77	30	329	5	2	40	190	38	6599
81	design	247	1026	136	3369	192	308	719	12	178	61	187	45	6480
82	foundation	398	1248	674	1690	460	737	102	191	634	79	90	120	6423
83	sample	1172	1331	708	713	1218	639	11	2	506	7	31	50	6388
84	collaboration	722	709	386	1004	1023	859	180	107	768	112	118	181	6169
85	resource	204	774	1146	793	789	876	75	70	1260	22	55	31	6095
86	development	524	714	456	1394	963	885	237	64	421	146	112	160	6076
87	government	321	746	661	876	408	1913	91	179	597	54	70	82	5998
88	field	2842	571	84	242	1676	154	18	20	80	15	120	108	5930
89	report	166	585	89	3747	259	157	351	33	37	75	235	74	5808
90	interpretation	138	821	83	3387	351	131	408	10	70	60	197	121	5777
91	colleague	878	896	300	826	967	578	100	47	496	154	100	274	5616
92	conclusion	446	511	282	896	657	1293	302	117	235	268	181	306	5494
93	financial assistance	716	622	1505	604	366	935	48	39	490	65	17	87	5494
94	receipt	101	249	275	1590	122	377	1143	3	17	427	515	675	5494
95	second author	235	54	29	93	231	689	52	3203	186	244	305	142	5463
96	partial support	185	368	911	288	753	874	42	377	1421	42	20	49	5330
97	consultant	26	299	8	4730	16	8	105	0	3	10	77	8	5290
98	critical reading	596	2614	297	1066	150	155	7	13	353	3	10	11	5275
99	time	445	381	406	1262	632	351	522	90	387	179	242	330	5227
100	team	530	662	205	775	1747	549	134	29	329	56	100	106	5222
101	referee	173	95	103	60	1235	409	20	2061	555	151	15	336	5213

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
102	addition	454	636	377	992	650	846	152	93	432	143	117	141	5033
103	useful comment	731	322	130	150	1375	378	59	351	781	173	59	493	5002
104	present study	485	958	127	2258	529	273	72	6	156	24	91	19	4998
105	group	389	858	727	736	640	637	77	120	629	50	55	61	4979
106	hospitality	246	119	60	47	815	124	12	879	2433	41	14	182	4972
107	research work	425	447	723	428	448	1808	22	149	399	61	14	33	4957
108	quality	606	190	91	180	1872	1019	25	353	239	164	45	122	4906
109	authorship	8	124	115	1521	24	318	1162	1	2	419	521	677	4892
110	fieldwork	2202	395	9	202	1548	72	49	1	11	25	60	290	4864
111	student	820	562	377	733	645	635	185	76	228	249	217	132	4859
112	permission	1604	535	212	354	1046	421	70	24	106	44	150	194	4760
113	finding	391	413	225	681	581	1035	276	88	125	259	194	286	4554
114	organization	223	925	150	2336	224	185	193	26	65	57	62	83	4529
115	associate editor	204	56	5	92	454	2318	5	365	528	226	16	150	4419
116	draft	1209	270	18	302	640	143	178	110	222	322	254	750	4418
117	research support	131	278	146	3081	91	131	77	45	63	103	95	177	4418
118	postdoctoral fellowship	480	1316	735	881	241	264	52	123	166	12	91	38	4399
119	patient	21	916	20	3052	18	20	225	5	8	4	99	8	4396
120	valuable discussion	232	598	672	246	473	665	4	124	1198	10	17	22	4261
121	grant agreement	265	848	345	778	648	694	59	96	328	57	35	86	4239
122	institution	246	288	68	2430	417	138	149	83	155	44	45	118	4181
123	information	718	359	158	578	1081	504	176	39	217	81	73	172	4156
124	honorarium	3	261	13	3575	0	3	82	0	0	3	46	2	3988
125	financial interest	40	928	164	2429	41	56	133	1	65	2	54	12	3925
126	recommendation	524	389	259	249	565	1009	48	108	196	184	111	178	3820
127	acknowledges	79	592	721	91	912	466	4	31	897	2	6	7	3808

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
128	data analysis	291	760	118	1654	319	146	202	4	113	37	138	25	3807
129	science	351	640	480	665	448	524	70	132	279	49	65	103	3806
130	acknowledges support	42	554	518	76	993	468	0	21	1050	1	2	6	3731
131	employee	153	574	66	2375	97	127	213	4	38	24	35	20	3726
132	family	271	820	35	1474	171	42	212	8	27	59	522	77	3718
133	province	350	543	637	765	181	713	21	108	349	16	7	15	3705
134	cooperation	581	402	214	775	639	444	118	16	169	62	107	97	3624
135	conduct	51	365	10	2680	22	19	308	0	11	19	92	9	3586
136	valuable suggestion	547	321	262	225	590	565	12	486	315	102	32	110	3567
137	studentship	287	955	545	740	378	224	56	31	207	6	76	42	3547
138	company	272	422	205	1670	209	470	53	14	111	35	42	21	3524
139	insightful comment	434	261	88	179	920	493	48	200	175	254	89	362	3503
140	interpretation of data	64	442	22	2414	46	28	288	2	2	34	120	26	3488
141	helpful suggestion	493	464	227	259	675	340	23	382	210	112	79	218	3482
142	effort	435	312	118	729	821	387	129	22	262	46	117	67	3445
143	equipment	436	463	695	491	328	572	21	4	320	5	24	15	3374
144	participation	265	554	90	1198	384	181	199	21	111	72	203	89	3367
145	guidance	507	381	210	463	503	454	134	96	174	143	71	203	3339
146	approval	352	394	265	1521	102	148	184	6	197	13	65	26	3273
147	research funding	137	294	166	2040	129	198	78	10	51	36	38	68	3245
148	fee	9	120	6	2994	6	8	63	1	0	1	24	0	3232
149	investigation	305	369	383	708	405	682	47	31	189	22	55	25	3221
150	research fellowship	297	468	1096	495	200	309	22	51	227	11	20	16	3212
151	figure	398	491	137	423	1207	159	15	25	204	9	41	75	3184
152	observation	110	87	118	26	2542	151	0	13	77	20	7	17	3168
153	appreciation	408	252	335	467	359	798	87	66	144	68	67	80	3131

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
154	excellent technical assistance	341	1115	121	1473	25	35	1	2	7	0	8	0	3128
155	encouragement	412	351	391	233	461	423	27	199	351	64	17	152	3081
156	grant support	44	321	46	2317	30	33	65	6	21	9	31	10	2933
157	code	154	232	240	243	844	422	11	97	557	36	13	81	2930
158	university	183	337	445	431	261	522	46	155	330	71	40	109	2930
159	management	314	336	119	1415	211	144	152	3	52	23	76	42	2887
160	statistical analysis	597	509	34	1317	142	48	64	5	13	19	90	29	2867
161	funding source	88	410	32	1757	72	47	216	2	10	26	151	30	2841
162	other author	26	188	13	2434	16	19	53	12	12	2	57	4	2836
163	software	199	364	259	325	900	340	17	42	243	18	31	24	2762
164	research fund	191	405	236	816	145	476	77	71	181	59	36	51	2744
165	frame	230	296	459	258	312	683	6	70	378	10	7	33	2742
166	form	242	506	540	377	315	353	33	50	222	16	20	40	2714
167	calculation	25	178	867	59	301	390	4	16	847	4	4	8	2703
168	present work	206	298	390	364	224	732	17	96	338	16	14	7	2702
169	investigator	31	887	30	1503	33	35	76	16	5	10	33	16	2675
170	subject matter	3	639	91	1821	4	6	88	4	3	2	0	4	2665
171	scheme	164	292	411	428	248	527	52	58	305	41	38	56	2620
172	specimen	1639	326	32	109	352	114	0	0	20	0	5	23	2620
173	computational resource	33	284	770	51	230	496	5	29	702	4	2	5	2611
174	additional support	308	512	163	621	411	202	107	20	142	23	51	46	2606
175	trial	204	326	5	1840	8	25	133	4	1	1	45	3	2595
176	endorsement	859	264	81	382	481	213	101	9	116	26	30	30	2592
177	entity	7	641	93	1686	7	16	92	0	1	4	6	3	2556
178	image	314	303	496	209	444	495	0	8	209	4	7	36	2525
179	initiative	210	577	183	588	237	249	94	21	195	34	34	60	2482

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
180	presentation	86	115	29	214	485	465	34	683	166	98	19	76	2470
181	interest	145	364	54	1175	124	110	112	100	159	21	71	32	2467
182	speaker	21	104	10	2205	15	8	37	0	5	3	36	2	2446
183	financial involvement	0	619	91	1632	0	2	85	0	0	0	0	0	2429
184	financial conflict	0	627	89	1597	0	4	85	0	0	1	3	0	2406
185	simulation	40	181	228	48	895	368	8	39	555	5	9	13	2389
186	system	183	305	220	282	549	412	14	32	341	18	21	11	2388
187	generous support	120	328	550	424	190	220	60	64	135	89	38	141	2359
188	crew	615	233	24	28	1351	30	0	0	56	4	8	9	2358
189	researcher	335	345	146	550	206	279	136	24	92	75	57	107	2352
190	third author	84	16	8	44	74	362	21	1385	80	111	116	51	2352
191	partial financial support	152	147	525	120	170	535	9	90	534	24	6	34	2346
192	technology	197	405	305	513	132	496	20	24	205	21	7	15	2340
193	agreement	179	312	154	522	354	342	42	65	248	28	33	48	2327
194	survey	483	132	7	583	436	100	196	9	4	115	55	201	2321
195	corresponding author	97	134	54	365	752	607	42	125	60	41	17	26	2320
196	other relevant affiliation	0	607	89	1533	0	2	80	0	0	0	0	0	2311
197	many thanks	531	192	99	191	535	224	59	46	92	59	76	165	2269
198	feedback	283	272	48	186	350	284	115	31	120	188	137	249	2263
199	funding agency	115	369	71	756	180	228	189	8	133	59	104	49	2261
200	service	142	312	214	734	384	176	69	7	132	12	24	28	2234
201	consortium	119	454	156	478	426	253	30	17	218	13	24	31	2219
202	input	339	299	77	476	371	197	112	18	84	69	58	105	2205
203	sincere thanks	376	142	276	186	298	468	34	122	168	34	9	62	2175
204	research program	230	302	213	274	237	485	24	47	222	35	29	53	2151

Rank	Noun Phrases	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Mathematics	Physics	Professional Fields	Psychology	Social Sciences	Total
205	conference	63	90	15	279	103	270	79	65	82	313	36	739	2134
206	logistical support	1081	189	14	138	536	38	20	0	9	7	49	45	2126
207	visit	357	109	84	38	342	195	2	493	418	21	6	48	2113
208	database	203	219	25	423	945	90	36	13	31	30	12	49	2076
209	thesis	346	266	121	436	320	152	82	135	89	22	46	56	2071
210	course	267	291	138	272	252	286	41	37	230	66	28	143	2051
211	reference	173	257	167	320	266	291	29	210	212	44	29	49	2047
212	product	470	189	64	519	491	170	32	5	30	19	9	30	2028
213	workshop	302	130	40	152	340	85	45	83	281	153	19	393	2023
214	policy	87	425	65	563	204	167	178	14	53	81	80	99	2016

Table 5. – Quality of representation of the rows (cumulative contribution for each NP)

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
1	work	11.70	77.90	88.80	92.00	92.90
2	author	45.40	48.70	76.80	87.80	89.10
3	grant	82.20	83.30	84.10	87.50	97.10
4	study	56.30	78.60	92.10	97.20	98.80
5	research	39.50	61.20	62.30	79.30	85.40
6	project	85.20	89.70	89.80	93.20	93.30
7	financial support	38.70	61.30	73.40	74.20	86.30
8	support	69.90	84.80	86.10	86.50	86.90
9	manuscript	33.80	34.00	83.20	94.90	97.10
10	paper	45.10	68.40	73.30	83.40	83.50
11	program	48.50	87.70	88.60	93.30	93.90
12	analysis	69.00	77.50	96.50	96.80	98.40
13	funding	4.50	5.80	12.90	61.50	73.90
14	preparation	59.10	71.20	96.70	98.50	99.60
15	assistance	1.80	4.90	9.40	74.40	74.50
16	anonymous reviewer	24.50	54.90	80.50	97.60	98.30
17	data collection	59.20	69.40	96.10	97.60	99.70
18	help	14.10	14.30	28.20	80.10	80.80
19	data	1.20	35.20	42.10	60.80	66.30
20	decision	64.60	75.60	96.60	98.40	99.40
21	thanks	58.90	60.50	60.90	77.80	78.90
22	article	3.20	52.00	53.00	60.90	95.40
23	funder	54.80	68.10	95.70	97.60	99.30
24	role in study design	50.50	66.10	96.40	98.20	99.50
25	comment	20.40	56.10	91.60	94.40	94.60

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
26	fellowship	8.20	47.50	47.50	52.80	54.50
27	fund	42.60	56.70	59.80	68.40	69.60
28	suggestion	38.60	65.20	89.50	94.60	94.70
29	contract	26.80	27.30	28.30	39.70	56.60
30	number	79.60	79.60	89.80	90.00	91.30
31	publication	1.10	1.50	21.30	21.30	21.40
32	helpful discussion	14.60	66.50	66.60	66.60	67.50
33	research grant	51.40	61.00	90.70	90.90	93.80
34	technical assistance	45.00	49.90	53.40	70.50	71.40
35	member	90.50	90.80	96.20	96.30	96.50
36	contribution	3.90	45.90	47.90	80.30	82.00
37	discussion	21.40	31.20	38.10	38.70	43.30
38	helpful comment	17.20	59.30	80.80	82.10	90.00
39	staff	1.80	31.10	34.20	71.40	73.10
40	material	26.60	29.70	56.10	56.60	60.60
41	view	0.10	50.40	55.50	59.10	89.70
42	conflict of interest	72.30	80.90	97.70	97.70	99.50
43	content	63.20	77.90	88.90	91.10	92.50
44	scholarship	0.30	2.90	3.00	31.10	31.30
45	laboratory	39.70	46.20	73.40	87.70	87.90
46	facility	28.20	73.50	74.50	90.80	91.20
47	experiment	13.50	41.80	42.20	61.10	62.20
48	use	30.60	31.90	37.30	67.60	72.50
49	result	56.50	56.50	58.80	58.80	63.40
50	first author	12.10	21.50	27.20	77.60	86.90
51	award	17.60	40.00	46.70	50.70	51.00
52	sponsor	77.80	80.30	90.20	90.20	94.10
53	access	20.60	21.90	39.50	86.20	86.30
54	framework	67.70	90.60	92.60	92.90	93.10

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
55	review	2.90	3.00	27.90	28.40	30.70
56	valuable comment	44.50	67.80	90.40	93.00	94.10
57	technical support	0.00	32.90	32.90	78.70	80.30
58	anonymous referee	24.10	54.80	70.70	85.90	88.00
59	editor	26.70	65.80	82.20	85.60	90.60
60	constructive comment	27.00	51.20	76.00	84.50	90.00
61	reviewer	49.20	71.50	88.20	88.30	90.00
62	education	42.50	83.10	90.60	90.90	90.90
63	collection	42.90	74.40	88.90	92.60	94.10
64	responsibility	37.40	62.70	81.50	85.00	92.00
65	useful discussion	22.90	37.20	37.40	37.40	42.80
66	research project	43.50	64.40	64.80	65.80	85.40
67	special thanks	23.50	51.30	65.40	91.50	92.00
68	measurement	32.70	71.70	82.90	85.80	88.20
69	version	10.10	62.10	79.40	79.60	90.40
70	recipient	93.40	94.20	94.20	94.50	95.40
71	grant sponsor	63.70	69.20	69.50	71.50	71.50
72	gratitude	80.60	82.80	83.30	83.40	84.10
73	advice	1.00	8.90	59.20	81.50	81.90
74	role	62.80	75.10	92.50	93.00	93.00
75	writing	50.40	71.60	93.90	94.70	94.80
76	participant	0.60	46.70	46.80	52.50	91.40
77	official view	46.90	63.60	81.50	83.60	85.80
78	fruitful discussion	30.60	59.50	60.90	60.90	63.20
79	opinion	12.50	36.10	37.30	42.20	75.20
80	study design	96.70	97.40	98.40	99.10	99.10
81	design	46.00	60.80	87.80	88.10	88.30
82	foundation	21.70	38.30	68.70	74.40	77.40
83	sample	4.60	7.80	29.90	89.30	94.90

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
84	collaboration	37.70	47.10	47.60	63.60	65.20
85	resource	24.20	61.80	70.10	72.20	72.90
86	development	5.10	44.80	55.30	73.20	73.60
87	government	25.20	44.90	57.40	61.10	61.80
88	field	4.50	17.00	36.30	70.80	75.10
89	report	42.30	64.10	97.40	97.50	99.00
90	interpretation	48.70	70.60	96.90	97.00	97.40
91	colleague	10.80	32.60	55.80	83.40	84.60
92	conclusion	14.80	43.10	44.20	46.40	74.40
93	financial assistance	23.00	49.50	57.40	65.80	71.50
94	receipt	1.60	34.80	40.70	44.80	84.00
95	second author	6.70	9.70	14.00	78.90	91.10
96	partial support	48.00	67.80	69.10	72.30	77.80
97	consultant	38.50	50.20	88.10	88.10	95.00
98	critical reading	53.20	73.50	93.20	94.30	94.30
99	time	0.00	45.50	52.00	52.00	71.10
100	team	7.90	24.40	36.70	64.70	73.50
101	referee	15.60	20.40	29.50	77.60	94.10
102	addition	28.40	49.40	57.60	60.60	65.60
103	useful comment	40.30	58.50	79.80	79.80	80.30
104	present study	56.00	67.80	78.10	84.30	96.70
105	group	25.50	69.70	70.10	76.10	77.50
106	hospitality	22.50	23.90	24.90	37.40	51.30
107	research work	41.60	58.50	63.60	63.60	63.90
108	quality	39.90	54.70	71.30	73.10	84.50
109	authorship	1.80	35.50	41.40	46.40	83.60
110	fieldwork	5.70	26.70	48.60	81.40	83.20
111	student	8.90	41.10	43.30	50.60	61.20
112	permission	7.20	24.20	43.10	75.90	76.30

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
113	finding	11.80	45.80	45.90	47.40	80.00
114	organization	65.40	77.10	96.30	96.30	97.20
115	associate editor	40.70	40.90	41.20	53.40	53.50
116	draft	7.80	52.20	63.50	64.00	81.90
117	research support	30.80	50.50	87.20	87.70	89.50
118	postdoctoral fellowship	25.70	48.90	51.90	52.90	53.80
119	patient	64.20	71.90	96.10	96.50	98.70
120	valuable discussion	28.80	66.40	67.10	67.10	69.30
121	grant agreement	3.50	11.70	18.70	23.70	29.10
122	institution	27.00	56.20	85.80	85.80	93.90
123	information	15.70	54.80	64.70	93.10	93.70
124	honorarium	39.60	50.40	88.00	88.10	94.90
125	financial interest	68.60	71.50	94.40	94.50	96.80
126	recommendation	47.90	54.00	59.50	59.60	68.50
127	acknowledges	29.40	56.00	56.60	62.90	65.50
128	data analysis	64.50	81.30	94.20	95.70	96.70
129	science	41.80	52.00	52.00	52.90	53.50
130	acknowledges support	26.70	45.30	46.40	51.40	56.50
131	employee	53.40	65.10	95.60	95.60	98.00
132	family	32.20	49.30	51.40	51.60	54.20
133	province	7.00	52.60	80.00	80.10	80.80
134	cooperation	2.30	49.00	49.30	91.90	92.30
135	conduct	45.70	61.50	97.00	97.20	98.70
136	valuable suggestion	42.10	45.70	60.80	77.50	91.20
137	studentship	24.90	46.90	48.20	61.50	62.10
138	company	30.50	37.70	82.60	83.80	89.90
139	insightful comment	30.90	67.60	87.90	88.10	91.70
140	interpretation of data	49.20	66.30	98.00	98.30	99.00
141	helpful suggestion	27.60	45.40	80.10	89.70	92.50

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
142	effort	4.90	44.00	44.60	73.70	80.90
143	equipment	18.10	52.10	59.90	84.00	84.20
144	participation	34.00	75.80	81.60	82.20	83.00
145	guidance	20.70	69.00	77.80	79.50	93.20
146	approval	36.40	45.50	87.80	89.70	90.80
147	research funding	35.00	48.50	89.90	89.90	93.20
148	fee	36.10	47.60	87.00	87.10	94.70
149	investigation	12.50	14.60	43.00	61.70	63.50
150	research fellowship	4.90	32.70	41.60	45.80	48.10
151	figure	5.10	15.70	37.60	68.20	82.10
152	observation	12.20	20.90	33.40	51.10	62.70
153	appreciation	41.30	43.10	48.40	51.30	55.80
154	excellent technical assistance	84.50	84.50	85.30	86.50	89.40
155	encouragement	72.30	72.30	83.10	83.70	84.10
156	grant support	44.30	54.00	88.80	88.80	95.00
157	code	46.90	47.00	50.00	53.70	64.70
158	university	58.20	66.70	74.00	90.50	93.30
159	management	39.70	66.20	93.80	96.70	98.50
160	statistical analysis	43.10	63.40	67.40	75.20	79.30
161	funding source	51.00	70.00	97.30	97.60	97.70
162	other author	39.10	50.70	88.10	88.20	95.30
163	software	19.30	20.10	29.50	52.20	66.70
164	research fund	9.80	10.30	74.50	81.00	81.30
165	frame	56.40	91.30	95.80	97.00	97.20
166	form	11.50	53.90	54.30	65.00	66.20
167	calculation	30.80	68.40	75.50	76.50	76.50
168	present work	40.70	74.80	87.00	87.50	88.60
169	investigator	86.30	87.60	94.70	95.30	96.70
170	subject matter	66.30	69.40	91.60	91.70	94.80

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
171	scheme	45.90	68.60	95.70	95.70	98.90
172	specimen	1.30	6.30	18.20	44.00	45.90
173	computational resource	30.70	79.70	85.80	86.10	86.50
174	additional support	21.10	42.30	45.10	83.40	85.40
175	trial	50.10	64.30	94.10	94.20	98.40
176	endorsement	2.20	18.50	25.80	61.80	64.80
177	entity	69.30	72.00	93.00	93.10	95.50
178	image	43.90	60.20	60.20	88.70	88.70
179	initiative	62.90	63.00	63.00	70.50	76.70
180	presentation	16.70	22.40	28.30	79.80	96.40
181	interest	43.50	60.40	85.00	88.90	96.10
182	speaker	36.30	48.10	87.00	87.00	94.50
183	financial involvement	69.10	71.50	92.30	92.50	95.10
184	financial conflict	70.50	72.70	92.70	92.90	95.30
185	simulation	36.90	38.90	41.60	48.90	59.60
186	system	38.00	42.50	44.00	60.90	74.90
187	generous support	6.50	7.20	13.30	14.10	44.40
188	crew	9.00	20.80	46.20	81.00	94.30
189	researcher	3.00	51.10	52.30	55.80	80.30
190	third author	6.90	9.50	13.40	79.40	91.60
191	partial financial support	51.90	85.80	93.60	94.00	94.00
192	technology	0.70	46.10	74.80	76.80	76.90
193	agreement	13.70	18.00	24.10	27.40	59.20
194	survey	0.00	77.80	78.60	85.70	92.20
195	corresponding author	21.60	33.00	34.70	35.10	48.80
196	other relevant affiliation	70.50	72.60	92.40	92.60	95.10
197	many thanks	18.80	57.10	77.00	93.40	94.20
198	feedback	8.40	47.60	56.60	56.80	91.40
199	funding agency	27.60	50.90	67.30	67.60	74.10

Row	NP	Axis1	Axis2	Axis3	Axis4	Axis5
200	service	12.80	22.90	41.30	63.20	73.70
201	consortium	0.90	1.80	4.30	26.60	42.50
202	input	0.10	71.40	75.30	91.00	94.90
203	sincere thanks	72.20	72.30	73.00	73.70	74.70
204	research program	58.10	69.60	69.80	71.70	72.60
205	conference	4.50	28.90	30.00	36.90	74.30
206	logistical support	2.80	17.20	33.00	65.80	69.00
207	visit	26.30	26.90	36.40	61.30	86.70
208	database	2.00	26.90	37.70	60.40	73.40
209	thesis	0.60	48.10	55.40	55.50	69.00
210	course	28.90	39.00	46.40	48.10	64.30
211	reference	29.90	31.10	34.50	76.00	97.20
212	product	0.10	31.60	34.40	79.00	93.00
213	workshop	15.30	42.20	50.60	51.30	65.70
214	policy	23.90	56.00	56.60	57.90	82.00

Table 6. – Quality of representation of the columns (cumulative contribution for each discipline)

Column	Discipline	Axis1	Axis2	Axis3	Axis4	Axis5
1	Biology	8.20	21.20	33.10	61.60	62.60
2	Biomedical Research	57.80	74.40	98.80	99.10	99.90
3	Chemistry	26.80	63.50	77.80	79.60	83.60
4	Clinical Medicine	55.80	65.80	96.50	96.50	99.30
5	Earth and Space	30.40	48.00	65.40	81.10	88.30
6	Engineering and Technology	58.10	68.40	76.80	78.50	79.70
7	Health	10.30	41.20	50.80	52.50	72.20
8	Mathematics	12.90	15.30	21.80	85.00	96.10
9	Physics	39.40	65.40	69.20	69.50	71.30
10	Professional Fields	16.30	58.50	61.10	70.90	94.60
11	Psychology	2.50	45.40	45.50	50.20	61.20
12	Social Sciences	14.90	55.10	59.90	63.50	90.10

3.4 Le contexte d'utilisation des termes les plus fréquents dans les remerciements

Article 4

Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015

Publié dans PLoS ONE (2019); 14(12): e0226727.

<https://doi.org/10.1371/journal.pone.0226727>

Cet article est inclus dans la présente thèse avec l'autorisation de la co-auteure.

© 2019 Paul-Hus et Desrochers

Contributions des auteurs (selon la taxonomie CRediT²⁵)

Adèle Paul-Hus: Conception de l'étude, méthodologie, gestion des données, investigation, analyse des données, visualisation, rédaction de l'ébauche initiale, révision critique du manuscrit et acquisition de soutien financier

Nadine Desrochers : Conception de l'étude, méthodologie, investigation, analyse des données, révision critique du manuscrit et acquisition de soutien financier

Contributions de l'auteure principale: Dans le cadre de ce projet, qui découlait directement de l'article 3 et en formait le complément, j'ai d'abord défini les objectifs et questions de recherche spécifiques à cet article, sous la supervision de ma co-directrice et co-auteure Nadine Desrochers. J'ai ensuite procédé à la revue de la littérature spécifique, notamment en ce qui a trait à l'analyse qualitative de ce type de données. En collaboration avec ma co-auteure, j'ai élaboré le devis méthodologique de l'étude. Dans un processus consultatif et itératif, nous avons défini un arbre de codage. Un échantillon test a été codé en parallèle avec ma co-auteure pour effectuer la réconciliation nécessaire à la validité et la fiabilité du codage. Une fois tous les cas de l'échantillon test réconciliés, j'ai fait l'échantillonnage des textes de remerciements à coder et procédé au codage de l'ensemble de l'échantillon. Les cas ambigus ont été validés au cours du codage auprès de ma co-auteure. J'ai ensuite fait l'analyse des données, les visualisations et l'interprétation des résultats, qui ont été révisés par ma co-auteure. J'ai finalement rédigé l'ébauche originale de l'article, qui a ensuite été révisée et commentée par ma co-auteure.

²⁵ <https://casrai.org/credit/>

Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015

Adèle Paul-Hus¹, Nadine Desrochers¹

¹ adele.paul-hus@umontreal.ca, nadine.desrochers@umontreal.ca

École de bibliothéconomie et des sciences de l'information, Université de Montréal
PO Box 6128, Downtown Station, Montreal, Quebec, H3C 3J7 (Canada)

Abstract

Acknowledgements in scientific articles can be described as miscellaneous, their content ranging from pre-formulated financial disclosure statements to personal testimonies of gratitude. To improve understanding of the context and various uses of expressions found in acknowledgements, this study analyses their content qualitatively. The most frequent noun phrases from a Web of Science acknowledgements corpus were analysed to generate 13 categories. When 3,754 acknowledgement sentences were manually coded into the categories, three distinct axes emerged: the contributions, the disclaimers, and the authorial voice. Acknowledgements constitute a space where authors can detail the division of labour within collaborators of a research project. Results also show the importance of disclaimers as part of the current scholarly communication apparatus, an aspect which was not highlighted by previous analyses and typologies of acknowledgements. Alongside formal disclaimers and acknowledgements of various contributions, there seems to remain a need for a more personal space where the authors can speak for themselves, in their own name, on matters they judge worth mentioning.

Keywords: acknowledgement, authorial voice, contribution, disclaimer, qualitative content analysis

Introduction

The idea of using acknowledgements as a source for bibliometric indicators has been surrounding their study since the 1990s. In 1991, Cronin was already asking, “why are acknowledgement counts excluded from formal assessments of individual merit or influence, such as tenure review?” (Cronin, 1991, p. 236). In 1995, Cronin and Weaver were encouraging

the development of an Acknowledgement Index, based on the model of the Science Citation Index. Almost two decades later, Costas and van Leeuwen (2012) suggested that it was perhaps time “to employ this sort of tool to facilitate development of the so-called ‘inflometrics’” (p. 1659). For their part, Diaz-Faes and Bordons (2014) highlighted that the inclusion of acknowledgement information in the Web of Science (WoS) was offering new avenues to study collaboration in science, going beyond traditional bibliometric indicators. McCain (2018) went further and assessed the feasibility of a formal Personal Acknowledgements Index. And yet, despite decades of studies positioning acknowledgements alongside citations and authorship in what Cronin called the “reward triangle” (Cronin, 1995), the consideration of acknowledgements as an indicator of scientific credit has not really materialized and, at best, remains a proposal at the exploratory stage, or even simply a rhetorical idea (see Desrochers, Paul-Hus & Pecoskie, 2017 for a meta-synthesis of this literature).

At the same time, many studies have used funding-related indicators based on acknowledgement data (e.g. Gausia et al., 2015; Harter & Hooten, 1992; Lewison, Grant & Jansen, 2001; Shah et al., 2014). In fact, acknowledgement studies can no longer be separated from the financial aspect of scientific research. In 2008, WoS started to collect and index funding sources found in the acknowledgements of scientific papers. These new data were added by WoS in response to many funding bodies’ requirement to acknowledge the sources supporting research. Since then, large-scale acknowledgement data have been used as a bibliometric tool to follow the money trail of research and funding-related analyses have become a dominant trend in recent acknowledgement literature (Desrochers, Paul-Hus & Pecoskie, 2017). To this day, acknowledgements have been more closely related to funding indicators than to any other kind of scientific credit indicators.

The literature also underlines the elusive nature of acknowledgements, pointing to their form and tone, which have been described as sometimes flowery, personal, and even manipulative:

Acknowledgements are permeated by hyperbole, effusiveness, overstatement, and exaggeration. (Hollander, 2001, p. 64)

Acknowledgements have been discussed as a form of patronage in scholarly communication, where the reality of the past may be purposefully glossed over and where the author could be looking toward the possibility of receiving future favours. (Forzetting, 2010, p. 4)

Furthermore, several studies mention the lack of standardization of acknowledgements as one important limitation hindering their analyses:

The format of acknowledgement varies from field to field and from journal to journal. As noted, persons and institutional sources may be listed in the methods and materials section of an article or explicitly thanked in an acknowledgement section. (McCain, 1991, p. 506)

Since there are no established formats for acknowledgements in papers, as there are for citations, expressions of gratitude vary greatly and sometimes it was difficult to identify the correct type of support, and even more difficult, the correct funding organization. (Jeschin, Lewison and Anderson, 1995, p. 238)

The first source of simple error may arise through the misspelling of the names of funding bodies and potentially the names of grants and grant codes [...]. A second difficulty will be that researchers will not correctly remember the funding bodies and grants that they used to support the research. (Rigby, 2011, p. 368-369)

Acknowledgements may thus contain formally required statements of gratitude but have also been used as personal spaces of authorial expression, and as such, acknowledgement texts have been analysed as a genre per se. Several discourse and linguistic analyses have studied on acknowledgements found in dissertations, theses, monographies, and research articles (e.g. Bing & Ruhl, 2008; Gesuato, 2004; Giannoni, 1998; 2002; 2006a; 2006b).

Acknowledgements analyses have also led to numerous typologies or classifications of the contributions acknowledged in scientific publications. In 1972, Mackintosh (1972) proposed the first qualitative content analysis of acknowledgements based on a typology of the three main types of “services” acknowledged in scientific papers: *facilities*, *access to data*, and *help of individuals*. Twenty years later, McCain (1991) offered a finer typology of acknowledgements, using five categories: *access to research-related information*, *access to unpublished results and data*, *peer interactive communication*, *technical assistance*, and *manuscript preparation*. The same year, Cronin introduced his first version of a six-part typology of acknowledgements (*paymaster*, *moral support*, *dogsbody*, *technical*, *prime mover*, and *trusted assessor*) which was created before encountering Mackintosh’s 1972 and McCain’s 1991 work (Cronin, 1991; McCain, 1991). Subsequent versions of this typology—developed with different collaborators through the years (namely McKenzie, Rubio and Weaver(-Wozniak))—include the *peer interactive communication* category borrowed from McCain (1991) alongside *moral support*, *access* (to resources, materials and infrastructure), *clerical support*, *technical support*, and *financial support* (Cronin, McKenzie & Rubio, 1993; Cronin, McKenzie & Stiffler, 1992; Cronin et al., 1993; Cronin & Weaver, 1995). Cronin’s model has since been adopted, adapted, and augmented in several studies (e.g. Innes, 2006; Rattan, 2013;

Salager-Meyer et al., 2006; Salager-Meyer, Alcaraz-Ariza & Berbesi, 2009; Tiew & Sen, 2002; Weber & Thomer, 2014).

More recently, Giles and Council (2004) used natural language processing to extract named entities from more than 180,000 acknowledgements published in computer science research papers. In their content analysis, the most frequently acknowledged entities are classified into four categories: *funding agencies*, *corporations*, *universities* and *individuals*. Other studies have analysed the content of acknowledgements focusing on funding bodies and classifying them by sectors and subsectors (e.g. Crawford & Biderman, 1970; Dawson et al., 1998; Lewison, 1998; Lewison & Devey, 1999; Lewison, Grant & Jansen, 2001).

Finally, linguistic studies have also used classifications of acknowledgements, focusing on the structure and patterns of dissertation acknowledgement texts (e.g. Al-Ali, 2010; Cheng, 2012; Gesuato, 2004; Hyland, 2003; 2004; Hyland & Tse, 2004) and on the socio-pragmatic construction of acknowledgements found in research articles and academic books (Giannoni, 1998; 2002; 2006a; 2006b).

Typologies and classifications aim to describe and categorize the content of acknowledgements in a synthetic manner. However, these taxonomies are based on small-scale samples of acknowledgements, the only exception being the work of Giles and Council (2004) which focused solely on named entities. More recently, a large-scale multidisciplinary analysis of acknowledgement texts was published by the authors and collaborators in PLOS One (Paul-Hus et al., 2017). This analysis of acknowledgements from more than one million articles and reviews published in 2015, highlighted important variations in the practices of acknowledging. Focusing on the 214 most frequent noun phrases of that corpus, the study showed that acknowledgement practices truly do vary across disciplines. Noun phrases referring to technical support appeared more frequently in natural sciences while noun phrases related to peers (colleagues, editors and reviewers) were more frequent in earth and space, professional fields, and social sciences. Noun phrases referring to logistics and fieldwork-related tasks appeared prominently in biology. Pre-formulated statements used in the context of conflict of interest or responsibility disclosures were more frequently found in acknowledgements from clinical medicine, health, and psychology. However, this analysis also led to further questions concerning the interpretation of these noun phrases in their original context. Findings from this study showed that acknowledgements are not limited to credit attribution and that the numerous taxonomies and classifications found in the literature do not account for the current

acknowledgement practices where pre-formulated statements of financial assistance and conflict of interest disclosures appear to be frequent (Paul-Hus et al., 2017). Conclusions from this study raise further questions because these pre-formulated statements could have an influence on large-scale analyses that use automated linguistic methods, thus calling for a qualitative analysis of acknowledgements in the context of their use.

Objective and Research Questions

To improve understanding of the context and various uses of expressions found in acknowledgements, this study proposes to analyse their content qualitatively. More specifically, this study aims at answering the following research questions:

- In which contexts are specific expressions used?
- Do the contexts and meanings vary by discipline?
- What does a qualitative analysis reveal in terms of offering avenues for a more contextualized use of acknowledgements in large-scale studies?

Data and Methods

Data for this study were retrieved from WoS's Science Citation Index Expanded (SCI-E) and Social Sciences Citation Index (SSCI), which both include funding acknowledgement data. It bears repeating that acknowledgments are collected and indexed by WoS only if they include funding source information (Paul-Hus, Desrochers & Costas, 2016). Access to WoS data in a relational database format was provided by the *Observatoire des sciences et des technologies* (<http://www.ost.uqam.ca>). The full text of acknowledgements from all 2015 articles and reviews indexed in the SCI-E and the SSCI were extracted. The original corpus includes a total of 1,009,411 acknowledgements for as many papers.

In a previous analysis, we identified the 214 most frequent noun phrases of that corpus of acknowledgement using natural language processing (Paul-Hus et al., 2017). For the purpose of the present qualitative analysis, these 214 noun phrases were reduced to single words (e.g. “technical assistance” was reduced to “technical” and “assistance”) and redundant words were excluded, for a final corpus of 154 single words. Each single word could therefore be found in context, no matter its proximity to other single words; this offered us the possibility to code various types of occurrences of each word, whether it was part of a noun phrase or not.

The coding was done in two steps. First, an initial codebook was established inductively by one researcher to classify each of the 154 words and revised by a second researcher. All words were then coded by both researchers and their work was reconciled through “negotiated agreement” (Campbell et al., 2013, p. 305, see also Hruschka et al., 2004; Schreir, 2012). Second, 20 words were selected from the corpus of 154 words by purposeful sampling, where cases for study are selected because “they offer useful manifestations of the phenomenon of interest” (Patton, 2002, p. 40). Selection of the words included in the final sample was based on the quantitative analysis findings (Paul-Hus et al., 2017), which highlighted the potential importance of pre-formulated statements such as “The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript” (ut 000367510900041). Special attention was given to the words frequently used in those statements (e.g. analysis, collection, design, preparation). Sampling decisions were also oriented towards potential polysemous words which could lead to different contextual meanings (e.g. “assistance”). The 20 words of the final sample were coded within the context of their original sentences, extracted from acknowledgements. Words were thus used as a seed to refer back to full acknowledgement sentences.

The coding process entails data reduction where the many meanings of a sentence must be reduced or summarized under one main category (Miles & Huberman, 1984) in order to reflect a practice or a phenomenon on a humanly manageable scale. The principles of saturation and qualitative sampling, whereby the sample is “conceptually representative of the set of all possible units” (Krippendorff, 2004, p. 84), ensures that the phenomenon is reflected in its full complexity. Therefore, acknowledgements were stratified by discipline to reflect potentially different disciplinary uses of a word. Coding was then performed on this sample of 20 words within their original acknowledgement contexts, using the sentence as the unit of analysis and adapting the codebook in an iterative manner as finer meanings emerged.

The final codebook is composed of 13 categories, presented in Table 1. The coding was done by one researcher and guided by the question, “in which context is this word used?” One category was selected for each sentence coded, aiming at qualifying the context in which a word is used. Each word of the sample was coded in a minimum of 15 original sentences per discipline, for all 12 disciplines, resulting in a total of 3,754 sentences coded. Results are reported in “thick description” using sufficient descriptions and quotations to allow “thick interpretation”, which means connecting individual cases to the larger context without going into trivial details (Patton, 2004, p. 503).

Table 1. – Codebook: categories of acknowledgement content and their definitions

Category	Definition	Example
Financial disclosure	Includes all types of funding and financial support or assistance.	“The financial assistance of the National Research Foundation (NRF grant: Unlocking the future-FA2007043000003) towards this research is hereby acknowledged.” (ut 000350024900008)
Conflict of interest	Refers to potential or actual conflict of interest or the absence of conflict of interest, which can be financial or otherwise.	“P.A.P. has an equity interest in Digital Proteomics, LLC, a company that may potentially benefit from the research results.” (ut 000356625700007)
Disclaimer	Responsibility disclaimer that content/opinions/conclusions are those of the author(s) solely and not of the funder or of another organization.	“The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.” (ut_code 000366223600042)
Ethics	Refers to ethical review, ethical approval of the research; can include some form of "seal of approval" by agencies.	“Institutional review board approval was granted under University IRB PRO12110345.” (ut 000364165000006)
Peer communication	Refers to intellectual contribution and communication with colleagues and peers, trusted assessors. Includes the process of comments, feedback, suggestions and peer review.	“Tim Birt, David Anderson, Anna Tigano, Rebecca Taylor, Nathaniel Clark, Catherine Dale and Raphael Lavoie provided insightful discussions.” (ut 000367457200004)
Investigation and Analysis	Refers to specific tasks such as the collection, treatment and analysis of data; the cycle of pre-writing work.	“Thanks Dr. Dongliang Li and Dr. Jianjun Cao from Nanjing Xiaozhuang University, for their help on field work and data analysis.” (ut 000350479600001)
Supervision and Management	Tasks and roles related to supervision, leadership and management responsibilities.	“Research included in this review was partly completed at the University of Newcastle, Australia, under the supervision of Dr John Clulow and Dr Micheal Mahony.” (ut 000346218400001)
Materials and Resources	Refers to all kinds of study materials, samples, computing resources, infrastructure, physical installations and instrumentation and reagents. People as objects of study (such as patients or population/sample) are also included.	“We thank Calcul Quebec and Compute Canada for access to the Mammoth supercomputer.” (ut 000363365000021)

Writing	Includes creation and/or presentation of the published work: original draft preparation, contribution to the writing itself; can include the creation of visualizations, maps, figures, tables, and illustrations.	“We thank Donald Cochrane, University of Saskatchewan, for his writing assistance” (ut 00035342600019)
Dissemination	Includes project, documents, and other forms of dissemination, such as conference presentations. Includes issues linked to cost of publication and open-access models.	“Data and supporting materials necessary to reproduce the numerical results will be available at www.hobolt.com upon publication.” (ut 00035033760013)
Organization	Refers to institutions or organizations, research centres, research groups, research chairs (can include funding organizations).	“The second author would like to thank Guangxi Experiment Center of Information Science.” (ut 000353065700007)
Combination	Two or more clear categories combined.	“The authors are grateful to the two referees and the editor for comments and suggestions and to Alfio Viola (University of Catania) for SEM assistance.” (ut 000353204900002)
Vague or other	Meaning cannot be inferred or is not covered by any other categories.	“We thank Jonas Klevas and Dainius Prapakavicius for their contribution during various stages of the paper preparation.” (ut 000357274600063)

Results

The results of the coding process are summarized in Table 2 which presents, for each word of the sample, the percentage of all the occurrences attributed to a specific category. The analysis reveals the importance of three distinct axes: the contributions, the disclaimers, and the authorial voice. Moreover, disciplinary patterns bring another layer of analysis as divergent uses of the coded words emerge.

Table 2. – Acknowledgement words coding results

	Investigation & analysis	Financial disclosure	Disclaimer	Peer comm.	Materials & resources	Dissemin.	Writing	Conflict of interest	Org.	Ethics	Other *
work	1%	89%	0%	1%	3%	1%	1%	1%	1%	0%	2%
author	10%	41%	9%	16%	8%	0%	1%	10%	1%	0%	6%
analysis	47%	8%	32%	7%	6%	0%	0%	0%	0%	0%	0%
preparation	21%	17%	35%	12%	4%	0%	7%	0%	0%	0%	4%
assistance	62%	19%	0%	1%	2%	0%	7%	0%	0%	0%	8%
help	65%	2%	0%	4%	1%	0%	10%	0%	0%	0%	18%
data	29%	3%	24%	1%	40%	3%	0%	0%	1%	0%	0%
decision	2%	28%	65%	0%	0%	1%	0%	0%	4%	0%	1%
contribution	28%	18%	0%	7%	0%	8%	1%	0%	1%	0%	37%
discussion	0%	1%	0%	98%	1%	1%	0%	0%	0%	0%	0%
experiment	62%	5%	2%	5%	15%	1%	0%	0%	9%	1%	0%
results	1%	52%	15%	14%	5%	9%	0%	1%	0%	1%	2%
access	0%	2%	3%	0%	70%	23%	0%	1%	0%	1%	0%
review	2%	1%	7%	58%	0%	25%	6%	0%	0%	2%	0%
collection	54%	5%	38%	0%	3%	0%	0%	0%	0%	0%	0%
measurement	64%	13%	1%	2%	14%	3%	0%	1%	2%	1%	1%
writing	5%	20%	36%	18%	4%	0%	16%	0%	0%	0%	1%
design	22%	2%	55%	10%	1%	2%	2%	0%	6%	0%	1%
interpretation	25%	1%	61%	13%	0%	0%	0%	0%	0%	0%	0%
code	45%	53%	0%	0%	0%	2%	0%	0%	0%	1%	0%
Total	26%	22%	18%	13%	9%	4%	2%	1%	1%	0%	4%

Note: Words are presented in the table in descending order of their frequency in the corpus.

* “Other” regroups the following categories: Supervision and Management, Combination, and Vague or other

Contributions

Acknowledgements constitute a space where authors can detail “who has done what” during the research process. Most often, authors use this space to thank colleagues that contributed to the research, as in the following example: “The authors thank Colleen Dalton and four anonymous reviewers for their helpful comments that improved the manuscript. We thank Fan-Chi Lin for providing FTAN measurements for comparison, and Anna Foster, Jiayi Xie and Goran Ekstrom for informative discussion.” (ut 000355321800013; earth and space). However, in some cases acknowledgements can also include contributorship statements from the authors in order to reflect the distribution of labour: “A.P., V.M. and V.P were involved in writing the manuscript. A.B.G and Y.A.K. were responsible for conception of the idea” (ut 000365808000014; clinical medicine).

The categories peer communication, investigation and analysis, materials and resources, and writing refer to specific types of contribution to research. These categories, taken together, represent half of the sample coded (50%), confirming the importance of the contributions axis within the acknowledgements’ context. Moreover, some words are used most often to refer to specific categories of contribution, such as “access” which is used mainly in the category materials and resources (70% of the occurrences coded), “discussion” which is almost exclusively associated to the peer communication category (98% of the occurrences coded), and “assistance”, “experiment”, “help”, and “measurement” which are all mainly associated to the category investigation and analysis (more than 60% of the occurrences coded).

Disclaimers

Acknowledgements are not necessarily thank-you notes or recognition of responsibility. Financial disclosure, conflict of interest, disclaimer, and ethics account for more than 40% of the sample coded. In fact, the categories financial disclosure and disclaimer are among the most frequent in the sample, accounting respectively for 22% and 18% of all occurrences coded. The words “analysis”, “collection”, “decision”, “design”, “interpretation”, “preparation”, and “writing”, which could all seemingly refer to types of contributions, were in fact used in the context of responsibility statements in a substantial share of the cases analysed. Moreover, the words “decision”, “design” and “interpretation” are also mostly found in those kinds of

responsibility disclaimers (in respectively 65%, 55% and 61% of the occurrences coded for these specific words).

Non-responsibility statements of funding bodies are the most frequent disclaimers. The following example presents a typical statement: “The funding source had no role in the design of the study, the analysis and interpretation of the data or the writing of, nor the decision to publish the manuscript.” (ut 000352854700010). However, we found declarations of non-responsibility for other types of contributors over some part of a research project, as in the following sentence: “The data collectors have no responsibility over the analysis and interpretations presented in this study.” (ut 000349266800011). Furthermore, disclaimers are not always non-responsibility statements and can, on the contrary, disclose the specific responsibility of an organization, such as: “This study was funded by Xi'an Janssen Pharmaceutical Ltd (Beijing, People's Republic of China) who was responsible for study design and data collection, analysis, and interpretation.” (ut 00035659490000).

Contributions and Disclaimers Crossovers

In many cases, the disciplinary stratification provided a further level of analysis. The words “analysis”, “assistance”, and “code” present clear disciplinary patterns where the coding highlights the distinction between the two main contextual uses: the contributions axis and the disclaimers axis. For instance, the word “analysis” is used primarily in the sample to describe an investigation and analysis type of contribution: “We are grateful to Nahoko Adachi for her help in conducting the statistical analysis” (ut 000353959400005; psychology). However, for biomedical research, clinical medicine, and health, “analysis” is used mainly within the category disclaimer (example: “The funding agencies did not have any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript” [ut 000346498800018; clinical medicine]). Mathematics is a divergent discipline, where the dominant category for “analysis” is financial disclosure, as exemplified by the following sentence: “This work was supported by the International Max-Planck Research School, 'Analysis, Design and Optimization in Chemical and Bio-chemical Process Engineering', Otto-von-Guericke-Universitat Magdeburg” (ut 000362588800005; mathematics).

Similarly, the word “assistance” is generally used across disciplines to describe a contribution pertaining to the category investigation and analysis (example: “The authors thank S. Watmough and K. Finder for assistance with field sampling at Dorset, and A. McDonough for assistance with the classification of plant species” [ut 000347756900044; earth and space]), except in engineering and technology and in mathematics where “assistance” is used to disclose financial help (financial disclosure) in the majority of the cases examined, as in this sentence: “The financial assistance of the National Research Foundation (NRF grant: Unlocking the future-FA2007043000003) towards this research is hereby acknowledged” (ut 000350024900008; mathematics).

Two distinct contextual uses emerge for the word “code”: it is found most often within the disclaimers axis (financial disclosure category) in biology, biomedical research, chemistry, health, psychology and social sciences (example: “The research (project code: TSY-11-3820) was supported by the Research Fund of Erciyes University” [ut 000363704000011; biology]) while it is used to describe a specific contribution (investigation and analysis category) in the majority of the cases studied in earth and space, engineering and technology, mathematics, physics and professional fields (example: “We thank Prof. D. Karaboga and Dr. B. Basturk for providing their excellent ABC MATLAB codes to implement this research” [ut 000361400900022; earth and space]).

In the case of the word “review”, the coding process also highlights two dominant uses, varying with the discipline: in biology, biomedical research, earth and space, mathematics, physics, and in the professional fields, “review” is used primarily to describe some part of the peer communication process (peer communication category), as in the following example: “We would like to express our gratitude to the anonymous referee for his or her careful review and insightful comments, in particular, for pointing out a simple proof of Lemma 1.8.” (ut 000347714700003; engineering and technology). However, in clinical medicine, a different use is made of the word “review” mainly to refer to the document per se (dissemination category), as in this example: “We are grateful to Dr. Mozzetta for critically reading the manuscript and all members of the lab for stimulating discussions during the preparation of this review” (ut 000352374400001; clinical medicine). For all the remaining disciplines (chemistry, health,

psychology, and social sciences), both categories (peer communication and dissemination) appear frequently.

The word “data” also presents distinct disciplinary patterns in the sample coded. “Data” is used mainly within the contributions axis (materials and resources category) in biology, clinical medicine, earth and space, engineering and technology, and social sciences (example: “The authors thank Chesapeake Energy for providing access to the VSP data we used” [ut 000364362900035; earth and space]). Moreover, the word “data” refers to a task within the investigation and analysis category in an important share of the cases coded in chemistry, physics, professional fields, and psychology (example: “We thank all graduate research assistants who helped with data collection” [ut 000348882900009; psychology]). However, “data” is mainly found within the disclaimers axis in clinical medicine and health (disclaimer category) as in the following example: “The funding agencies had no role in the study design, data collection and analysis, the decision to publish or preparation of the manuscript” [ut 000345586900003; clinical medicine].

Authorial Voice

Although details of contributions and various disclaimers represent a substantive share of their content, acknowledgements also constitute a space for personal testimony. Notwithstanding the expectations of funders and ethical considerations, acknowledgements remain the subjective presentation of researchers’ practices and of research contexts. The authors are the voice of the acknowledgements and as such, the word “author” is one of the most frequent with more than 339,000 occurrences in our dataset. Moreover, even when the word “author” is absent, the concept is not. In fact, the authorial voice cannot be reduced to a single category, because it pervades the acknowledgements whether the authors speak in the first or third persons, as shown in the following sentences:

“I would like to thank Iliana Flores, Amy Harrison, and Shannon Kahlden for their help with data collection.” (ut 000361977300090)

“We would also like to thank two anonymous reviewers for the contributions to this manuscript.” (ut 000364777400031)

“Also, our thanks go to Mr Vit Hanousek who designed an original computer tool suitable for making all the above-discussed measurements.” (ut 000346267600010)

“The authors declare that they have no competing interests.” (ut 000369908800022)

“The authors wish to express their appreciation to the National Iranian Copper Industry Company (NICICO) for funding this work.” (ut 000344595900005)

“Schuster is profoundly grateful to all the families who hosted her but especially Hasidullah, his wife, son and grandson who were unfailingly patient and kind with the strange cuckoo in their nest and to the Leverhulme Trust for funding her time in Afghanistan.” (ut 000350285300006)

“This review is dedicated to the memory of my father who was a source of inspiration.” (ut 000349637500005)

Furthermore, as exemplified by the cases presented above, the varied nature of the testimonies found in acknowledgements underlines a need for a “free space” within research publications. Alongside formal disclaimers and acknowledgements of various contributions, authors seem to require a more personal space where they can speak for themselves, in their own name, on matters they judge worth mentioning.

Discussion and Conclusion

In the last decades, acknowledgements have become a “constitutive element of academic writing” (Cronin, Shaw & La Barre, 2004, p. 160). However, the acknowledgement section is not a mandatory part of a scientific article and its content could certainly be described as miscellaneous, ranging from pre-formulated financial disclosure statements to personal testimonies of gratitude. Moreover, acknowledgements’ content and practices have evolved over time, just as citations and authorship attribution practices have changed following the transformations that are affecting the whole reward system of science (Desrochers et al., 2018).

Typologies and classifications of acknowledgements have been a consistent topic in the acknowledgement literature (Desrochers, Paul-Hus & Pecoskie, 2017). Most of these typologies and classifications revolve around the contributions axis of acknowledgements, focusing on “who gets thanked for what” and “what types of contributions are acknowledged”. This qualitative analysis of acknowledgements content confirms the importance of the contributions axis: acknowledgements are indeed still a space where authors can detail the division of labour within all collaborators of a research project. However, our findings also reveal the importance of disclaimers as part of the current scholarly communication apparatus, an aspect which was not highlighted by previous analyses and typologies.

It should be noted that, our analysis was restricted to a corpus of single words, sampled from noun phrases identified by correspondence analysis (Paul-Hus et al., 2017). Further research could now seek to recombine those single words into noun phrases that present variations in meaning around a common concept, such as “assistance” (e.g. “technical assistance” and “financial assistance”). Furthermore, our coding of acknowledgement sentences was done using mutually exclusive categories, an epistemological choice. Given the fact that sentences can perform more than one kind of action, another avenue would be to use open coding and place occurrences in non-exclusive, mutually complementary categories.

Our qualitative results show that caution should be used when working with acknowledgement data. Large-scale acknowledgement data are limited to funded research, given that in the two main bibliographic databases, Web of Science and Scopus, acknowledgements are collected with the intended objective of identifying funding sponsors and tracking funded research (Clarivate Analytics, 2019; Scopus, 2019). The indexation of acknowledgements are thus limited to acknowledgements that contain some kind of funding information; this could in turn induce a potential bias toward funding-related aspects within acknowledgements’ content (Paul-Hus, Desrochers & Costas, 2016). This indexation bias could then, at least in part, explain the importance of funding disclosures in the dataset analysed here, but also elsewhere in large-scale studies.

Yet, our findings show that acknowledgements cannot be described as having one single and homogeneous purpose; they can include expected, if not imposed, acknowledgement of financial resources as well as infrastructure alongside very personal testimonies of gratitude, all at the same time, as the following excerpt exemplifies: “Data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. [...]. The authors wish to extend special thanks to those of Hawaiian ancestry, on whose sacred mountain we are privileged to be guests. Without their generous hospitality, the observations would not have been possible” (ut 000363471600015). On rare occasions, personal matters discussed in the acknowledgements become the center of attention, such as when an author proposed to his girlfriend in the acknowledgement of a paper: “C.M.B. would

specifically like to highlight the ongoing and unwavering support of Lorna O'Brien. Lorna, will you marry me?" (Brown & Henderson, 2015). This particular paper was covered by many news outlets and online media sites when it was published, ranking in the 20th position of the Altmetrics Top100 ranking for the year 2015. Such a case highlights the potential unexpected effect an acknowledgement can have on the visibility of a paper.

Clearly delimited and dedicated spaces for funding information, conflict of interest disclosures and contributorship statements are already implemented in some scientific journals (e.g. PLOS One, The Lancet, Science). Nonetheless, such examples are far from being the norm at the moment. In light of our findings, if an effort of standardization of acknowledgements is to be made, acknowledgements should at least include three main sections: ethics of research (financial disclosure, conflict of interest and responsibility disclaimers), contributions made to research, and personal testimony. These three indexation fields would, in turn, allow large-scale analysis of acknowledgements without the equivocality that currently characterizes these texts, yet without narrowing the space left for the authorial voice. The question remains as to whether there is a real wish within the scientific community to delineate such acknowledgement sections; if not, acknowledgement data are likely destined to remain simple tracking devices for science funding, the contributions and the authorial voices lost in large-scale analyses of scientific credit.

Acknowledgement

The authors would like to thank Vincent Larivière for his comments and the three anonymous reviewers for their insightful suggestions and careful reading of the manuscript. This research was supported by the Social Sciences and Humanities Research Council of Canada: Joseph-Armand Bombardier CGS Doctoral Scholarships (Paul-Hus) and, Insight Development [grant number 430-2014-0617] (Desrochers).

References

- Al-Ali, M. N. (2010). Generic patterns and socio-cultural resources in acknowledgements accompanying Arabic Ph.D. dissertations. *Pragmatics*, 20(1), 1–26.
<https://doi.org/10.1075/prag.20.1.01ali>

- Ben-Ari, E. (1987). On acknowledgments in ethnographies. *Journal of Anthropological Research*, 43(1), 63–84.
- Bing, J. & Ruhl, C. (2008). It's all my fault! The pragmatics of responsibility statements. *Journal of Pragmatics*, 40(3), 537–558. <https://doi.org/10.1016/j.pragma.2007.04.010>
- Brown, C. M. & Henderson, D. M. (2015). A new horned dinosaur reveals convergent evolution in cranial ornamentation in ceratopsidae. *Current Biology*, 25(12), 1641-1648. <https://doi.org/10.1016/j.cub.2015.04.041>
- Campbell, J. L., Quincy, C., Osserman, J. & Pederson, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294–320.
- Cheng, S. W. (2012). A contrastive study of master thesis acknowledgements by Taiwanese and North American students. *Open Journal of Modern Linguistics*, 02(01), 8–17. <https://doi.org/10.4236/ojml.2012.21002>
- Clarivate Analytics. (2019). LibGuides: Web of Science Core Collection: Who's funding MY research? Retrieved April 25, 2019, from [//clarivate.libguides.com/woscc/funding](http://clarivate.libguides.com/woscc/funding)
- Crawford, E. T. & Biderman, A. D. (1970). Paper money: Trends of research sponsorship in American sociology journals. *Social Science Information*, 9(1), 50–77. <https://doi.org/10.1177/053901847000900103>
- Cronin, B. (1991). Let the credits roll: A preliminary examination of the role played by mentors and trusted assessors in disciplinary formation. *Journal of Documentation*, 47(3), 227–239. <https://doi.org/10.1108/eb026878>
- Cronin, B., McKenzie, G. & Rubio, L. (1993). The norms of acknowledgement in four social sciences disciplines. *Journal of Documentation*, 49(1), 29–43. <https://doi.org/10.1108/eb026909>
- Cronin, B., McKenzie, G., Rubio, L. & Weaver-Wozniak, S. (1993). Accounting for influence: Acknowledgments in contemporary sociology. *Journal of the American Society for Information Science*, 44(7), 406–412.

- Cronin, B., McKenzie, G. & Stiffler, M. (1992). Patterns of acknowledgement. *Journal of Documentation*, 48(2), 107–122.
- Cronin, B., & Weaver, S. (1995). The praxis of acknowledgement: From bibliometrics to inflometrics. *Revista Española de Documentación Científica*, 18(2), 172–177.
- Dawson, G., Lucocq, B., Cottrell, R. & Lewison, G. (1998). *Mapping the landscape: National biomedical research outputs 1988-95* (No. Policy Report number 9). London, England: The Wellcome Trust.
- Desrochers, N., Paul-Hus, A., Haustein, S., Costas, R., Mongeon, P., Quan-Haase, A., Bowman, T.D., Pecoskie, J., Tsou, A. & Larivière, V. (2018). Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social Science Information*, 57(2), 223–248.
<https://doi.org/10.1177/0539018417752089>
- Desrochers, N., Paul-Hus, A. & Larivière, V. (2016). The angle sum theory: Exploring the literature on acknowledgments in scholarly communication. In C. R. Sugimoto (ed), *Theories of Informetrics and Scholarly Communication* (pp. 225–247). Berlin : de Gruyter Mouton.
- Desrochers, N., Paul-Hus, A. & Pecoskie, J. (2017). Five decades of gratitude: A meta-synthesis of acknowledgments research. *Journal of the Association for Information Science and Technology*, 68(12), 2821–2833. <https://doi.org/10.1002/asi.23903>
- Forzetting, S. A. (2010). *Personal, peer, patron: Scholarly interactions with librarians and archivists in the acknowledgements of women's history books*. (Master Dissertation). University of North Carolina at Chapel Hill.
- Gesuato, S. (2004). Acknowledgments in PhD dissertations: The complexity of thanking. In C. Taylor Torsello, M. Grazia Bùsa, & S. Gesuato (eds), *Lingua inglese e mediazione linguistica. Ricerca e didattica con supporto telematico* (pp. 273–318). Padova: Unipress.

- Giannoni, D. S. (1998). The genre of journal acknowledgments: Findings of a cross-disciplinary investigation. *Linguistica e Filologia*, 6, 61–84.
- Giannoni, D. S. (2002). Worlds of gratitude: A contrastive study of acknowledgement texts in English and Italian research articles. *Applied Linguistics*, 23(1), 1–31.
<https://doi.org/10.1093/applin/23.1.1>
- Giannoni, D. S. (2006a). Book acknowledgements across disciplines and texts. In K. Hyland & M. Bondi (eds), *Academic discourse across disciplines* (pp. 151–176). New York, NY: Peter Lang.
- Giannoni, D. S. (2006b). Evidence of generic tension in academic book acknowledgements. In V. K. Bhatia & M. Gotti (eds), *Explorations in specialized genres* (pp. 21–42). Bern: Peter Lang.
- Giles, C. L. & Councill, I. G. (2004). Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17599–17604.
<https://doi.org/10.1073/pnas.0407743101>
- Heffner, A. G. (1981). Funded research, multiple authorship, and subauthorship collaboration in four disciplines. *Scientometrics*, 3(1), 5–12. <https://doi.org/10.1007/BF02021860>
- Hollander, P. (2001). Acknowledgments: An academic ritual. *Academic Questions*, 15(1), 63–76. <https://doi.org/10.1007/s12129-001-1056-x>
- Hruschka, D. J., Schwartz, D., St.John, D. C., Picone-Decaro, E., Jenkins, R. A. & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, 16(3), 307–331. <https://doi.org/10.1177/1525822X04266540>
- Hyland, K. (2004). Graduates' gratitude: The generic structure of dissertation acknowledgements. *English for Specific Purposes*, 23(3), 303–324.
[https://doi.org/10.1016/S0889-4906\(03\)00051-6](https://doi.org/10.1016/S0889-4906(03)00051-6)
- Hyland, K. (2003). Dissertation acknowledgements: The anatomy of a Cinderella genre. *Written Communication*, 20(3), 242–268. <https://doi.org/10.1177/0741088303257276>

- Hyland, K. & Tse, P. (2004). "I would like to thank my supervisor". Acknowledgements in graduate dissertations. *International Journal of Applied Linguistics*, 14(2), 259–275. <https://doi.org/10.1111/j.1473-4192.2004.00062.x>
- Innes, J. (2006). *Scholarly communication and knowledge management in American zoos* (Doctoral Dissertation). Nova Southeastern University.
- Jeschin, D., Lewison, G. & Anderson, J. (1995). A bibliometric database for tracking acknowledgements of research funding. In M.E.D. Koenig & A. Bookstein (eds) *Proceedings of the Fifth International Conference of the International Society for Scientometrics and Informetrics* (p. 235–244). Medford, NJ: Learned Information.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Lewison, G. (1998). Gastroenterology research in the United Kingdom: Funding sources and impact. *Gut*, 43(2), 288–293. <https://doi.org/10.1136/gut.43.2.288>
- Lewison, G. & Devey, M. E. (1999). Bibliometric methods for the evaluation of arthritis research. *Rheumatology*, 38(1), 13–20. <https://doi.org/10.1093/rheumatology/38.1.13>
- Lewison, G., Grant, J. & Jansen, P. (2001). International gastroenterology research: Subject areas, impact, and funding. *Gut*, 49(2), 295–302. <https://doi.org/10.1136/gut.49.2.295>
- Mackintosh, S. H. (1972). *Acknowledgment patterns in sociology* (Ph.D. Thesis). University of Oregon, Michigan.
- McCain, K. W. (1991). Communication, competition, and secrecy: The production and dissemination of research-related information in genetics. *Science, Technology & Human Values*, 16(4), 491–516. <http://doi.org/10.1177/016224399101600404>
- McCain, K. W. (2018). Beyond Garfield's Citation Index: An assessment of some issues in building a personal name Acknowledgments Index. *Scientometrics*, 114(2), 605–631. <https://doi.org/10.1007/s11192-017-2598-1>

- Miles, M. B. & Huberman, A. M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher*, 13(5), 20–30.
<https://doi.org/10.3102/0013189X013005020>
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Paul-Hus, A., Desrochers, N. & Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, 108(1), 167–182. <https://doi.org/10.1007/s11192-016-1953-y>
- Paul-Hus, A., Díaz-Faes, A. A., Sainte-Marie, M., Desrochers, N., Costas, R. & Larivière, V. (2017). Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences. *PLOS ONE*, 12(10), e0185578. <https://doi.org/10.1371/journal.pone.0185578>
- Rattan, G. K. M. (2013). Acknowledgement patterns in Annals of Library and Information Studies 1999-2012. *Library Philosophy and Practice, e-journal* (Paper 989).
<http://digitalcommons.unl.edu/libphilprac/989/>
- Rigby, J. (2011). Systematic grant and funding body acknowledgement data for publications: New dimensions and new controversies for research policy and evaluation. *Research Evaluation*, 20(5), 365–375. <https://doi.org/10.3152/095820211X13164389670392>
- Salager-Meyer, F., Alcaraz-Ariza, M. A. & Berbesi, M. (2009). “Backstage Solidarity” in Spanish- and English-written medical research papers: Publication context and the acknowledgment paratext. *Journal of the American Society for Information Science and Technology*, 60(2), 307–317. <https://doi.org/10.1002/asi.20981>
- Salager-Meyer, F., Alcaraz-Ariza, M. Á., Berbesí, M. P. & Zambrano, N. (2006). Paying one’s intellectual debt: Acknowledgments in conventional vs complementary/alternative medical research. In M. Gotti & F. Salager-Meyer (eds), *Advances in Medical Discourse Analysis: Oral and Written Contexts* (pp. 407–430). Bern: Peter Lang.
- Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: Sage Publications.

Scopus. (2019). Funding acknowledgement information available for 9 million articles.

Elsevier Scopus Blog. Retrieved April 25, 2019, from

<https://blog.scopus.com/posts/funding-acknowledgement-information-available-for-9-million-articles>

Shah, K., Sussex, J., Hernandez-Villafuerte, K., Garau, M., Rotolo, D., Hopkins, M. M., Grassoni, N., Crane, P., Lang, F., Hutton, J., Pateman, C., Mawer, A., Farrell, C. & Sharp, T. (2014). *Exploring the interdependencies of research funders in the UK | OHE*.

<https://www.ohe.org/publications/exploring-interdependencies-research-funders-uk>

Tiew, W. S. & Sen, B. K. (2002). Acknowledgement patterns in research articles: A bibliometric study based on Journal of Natural Rubber Research 1986-1997. *Malaysian Journal of Library & Information Science*, 7(1), 43–56.

Weber, N. & Thomer, A. (2014). Paratexts and documentary practices: Text mining authorship and acknowledgment from a bioinformatics corpus. In N. Desrochers & D. Apollon (eds), *Examining paratextual theory and its applications in digital culture* (pp. 84–109). Hershey, PA: IGI Global.

Appendix

Table 3. – References of the acknowledgement excerpts cited

UT codes	References
000367510900041	Arciero, E., Biagini, S. A., Chen, Y., Xue, Y., Luiselli, D., Tyler-Smith, C., ... Ayub, Q. (2015). Genes Regulated by Vitamin D in Bone Cells Are Positively Selected in East Asians. <i>PLOS ONE</i> , 10(12), e0146072. https://doi.org/10.1371/journal.pone.0146072
000350024900008	Arashi, M., Bekker, A., Loots, M. T., & Roux, J. J. J. (2015). Integral Representation of Quaternion Elliptical Density and its Applications. <i>Communications in Statistics-Theory and Methods</i> , 44(4), 778-789. https://doi.org/10.1080/03610926.2012.753089
000356625700007	Safonova, Y., Bonissone, S., Kurpilyansky, E., Starostina, E., Lapidus, A., Stinson, J., ... Pevzner, P. A. (2015). IgRepertoireConstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis. <i>Bioinformatics</i> , 31(12), 53-61. https://doi.org/10.1093/bioinformatics/btv238
000366223600042	Pyun, S. Y., Jeong, J.-H., & Bae, J. S. (2015). Recurrent Guillain-Barre syndrome presenting stereotypic manifestations, positive antiganglioside antibodies, and rapid recovery. <i>Clinical Neurology and Neurosurgery</i> , 139, 230-233. https://doi.org/10.1016/j.clineuro.2015.10.022
000364165000006	Lamberton, C. M., Leana, C. R., & Williams, J. M. (2015). Measuring Empathetic Care: Development and Validation of a Self-Report Scale. <i>Journal of Applied Gerontology</i> , 34(8), 1028-1053. https://doi.org/10.1177/0733464813507131
000367457200004	Friesen, V. L. (2015). Speciation in seabirds: why are there so many species...and why aren't there more? <i>Journal of Ornithology</i> , 156, S27-S39. https://doi.org/10.1007/s10336-015-1235-0
000350479600001	Gu, Z., Sanchez-Azofeifa, G. A., Feng, J., & Cao, S. (2015). Predictability of leaf area index using vegetation indices from multiangular CHRIS/PROBA data over eastern China. <i>Journal of Applied Remote Sensing</i> , 9, 096085. https://doi.org/10.1117/1.JRS.9.096085
000346218400001	Browne, R. K., Kaurova, S. A., Uteshev, V. K., Shishova, N. V., McGinnity, D., Figiel, C. R., ... Cosson, J. (2015). Sperm motility of externally fertilizing fish and amphibians. <i>Theriogenology</i> , 83(1), 1-13. https://doi.org/10.1016/j.theriogenology.2014.09.018

000363365000021	Vapaavuori, J., Grosrenaud, J., Pellerin, C., & Bazuin, C. G. (2015). In Situ Photocontrol of Block Copolymer Morphology During Dip-Coating of Thin Films. <i>Acs Macro Letters</i> , 4(10), 1158-1162. https://doi.org/10.1021/acsmacrolett.5b00483
000353426000019	Bisol, C. A., Valentini, C. B., & Rech Braun, K. C. (2015). Teacher education for inclusion: Can a virtual learning object help? <i>Computers & Education</i> , 85, 203-210. https://doi.org/10.1016/j.compedu.2015.02.017
000350337600013	Wilson, T. L., & Hobolt, S. B. (2015). Allocating Responsibility in Multilevel Government Systems: Voter and Expert Attributions in the European Union. <i>Journal of Politics</i> , 77(1), 102-113. https://doi.org/10.1086/678309
000353065700007	Zhang, B., & Zhu, Z. (2015). A modified quasi-Newton diagonal update algorithm for total variation denoising problems and nonlinear monotone equations with applications in compressive sensing. <i>Numerical Linear Algebra with Applications</i> , 22(3), 500-522. https://doi.org/10.1002/nla.1968
000353204900002	Sciuto, F., Rosso, A., Sanfilippo, R., & Maniscalco, R. (2015). New faunistic data on the Pleistocene environmental evolution of the south-western edge of the Hyblean Plateau (SE Sicily). <i>Carnets De Geologie</i> , 15(4-5), 41-57.
000357274600063	Dobrovolskas, V., Kučinskas, A., Bonifacio, P., Caffau, E., Ludwig, H.-G., Steffen, M., & Spite, M. (2015). Three-dimensional hydrodynamical CO5BOLD model atmospheres of red giant stars - IV. Oxygen diagnostics in extremely metal-poor red giants with infrared OH lines. <i>Astronomy & Astrophysics</i> , 576, A128. https://doi.org/10.1051/0004-6361/201424885
000355321800013	Jin, G., & Gaherty, J. B. (2015). Surface wave phase-velocity tomography based on multichannel cross-correlation. <i>Geophysical Journal International</i> , 201(3), 1383-1398. https://doi.org/10.1093/gji/ggv079
000365808000014	Pandey, A., Malek, V., Prabhakar, V., Kulkarni, Y. A., & Gaikwad, A. B. (2015). Nanoparticles: A Neurotoxicological Perspective. <i>Cns & Neurological Disorders-Drug Targets</i> , 14(10), 1317-1327. https://doi.org/10.2174/1871527314666150821112411
000352854700010	Nilson, F., Bonander, C., & Jonsson, A. (2015). Differences in Determinants Amongst Individuals Reporting Residential Fires in Sweden: Results from a Cross-Sectional Study. <i>Fire Technology</i> , 51(3), 615-626. https://doi.org/10.1007/s10694-015-0459-0
000349266800011	Garcia-Gomez, P., Hernandez-Quevedo, C., Jimenez-Rubio, D., & Oliva-Moreno, J. (2015). Inequity in long-term care use and unmet need: Two sides of the same coin. <i>Journal of Health</i>

	<i>Economics</i> , 39, 147-158. https://doi.org/10.1016/j.jhealeco.2014.11.004
000356594900001	Si, T., Zhang, K., Tang, J., Fang, M., Li, K., Zhuo, J., & Feng, Y. (2015). Efficacy and safety of flexibly dosed paliperidone palmitate in Chinese patients with acute schizophrenia: an open-label, single-arm, prospective, interventional study. <i>Neuropsychiatric Disease and Treatment</i> , 11, 1483-1492. https://doi.org/10.2147/NDT.S81760
000353959400005	Ando, K., Yorifuji, K., Ohnuma, S., Matthies, E., & Kanbara, A. (2015). Transmitting pro-environmental behaviours to the next generation: A comparison between Germany and Japan. <i>Asian Journal of Social Psychology</i> , 18(2), 134-144. https://doi.org/10.1111/ajsp.12078
000346498800018	Bischoff, A. L., Folsgaard, N. V., Vissing, N. H., Birch, S., Brix, S., & Bisgaard, H. (2015). Airway Mucosal Immune-suppression in Neonates of Mothers Receiving A(H1N1)pnd09 Vaccination During Pregnancy. <i>Pediatric Infectious Disease Journal</i> , 34(1), 84-90. https://doi.org/10.1097/INF.0000000000000529
000362588800005	Camejo, C. C., & Warnecke, G. (2015). The singular kernel coagulation equation with multifragmentation. <i>Mathematical Methods in the Applied Sciences</i> , 38(14), 2953-2973. https://doi.org/10.1002/mma.3272
000347756900044	Cuss, C. W., & Gueguen, C. (2015). Relationships between molecular weight and fluorescence properties for size-fractionated dissolved organic matter from fresh and aged sources. <i>Water Research</i> , 68, 487-497. https://doi.org/10.1016/j.watres.2014.10.013
000350024900008	Arashi, M., Bekker, A., Loots, M. T., & Roux, J. J. J. (2015). Integral Representation of Quaternion Elliptical Density and its Applications. <i>Communications in Statistics-Theory and Methods</i> , 44(4), 778-789. https://doi.org/10.1080/03610926.2012.753089
000363704000011	Gogebakan, T., & Eraslan, G. (2015). Single-dose toxicokinetics of permethrin in broiler chickens. <i>British Poultry Science</i> , 56(5), 605-611. https://doi.org/10.1080/00071668.2015.1085957
000361400900022	Song, X., Gu, H., Tang, L., Zhao, S., Zhang, X., Li, L., & Huang, J. (2015). Application of artificial bee colony algorithm on surface wave data. <i>Computers & Geosciences</i> , 83, 219-230. https://doi.org/10.1016/j.cageo.2015.07.010
000347714700003	Couceiro, M., Lehtonen, E., & Schölzel, K. (2015). A complete classification of equational classes of threshold functions included in clones. <i>RAIRO - Operations Research</i> , 49(1), 39-66. https://doi.org/10.1051/ro/2014034

000352374400001	Brancaccio, A., & Palacios, D. (2015). Chromatin signaling in muscle stem cells: interpreting the regenerative microenvironment. <i>Frontiers in Aging Neuroscience</i> , 7, 36. https://doi.org/10.3389/fnagi.2015.00036
000364362900035	Li, Y., & Hardage, B. A. (2015). SV-P extraction and imaging for far-offset vertical seismic profile data. <i>Interpretation-a Journal of Subsurface Characterization</i> , 3(3), SW27-SW35. https://doi.org/10.1190/INT-2015-0002.1
000348882900009	Chen, C., Liu, C., Chen, C., Moyzis, R., Chen, W., & Dong, Q. (2015). Genetic variations in the serotonergic system and environmental factors contribute to aggressive behavior in Chinese adolescents. <i>Physiology & Behavior</i> , 138, 62-68. https://doi.org/10.1016/j.physbeh.2014.09.005
000345586900003	Chong, H., Qiu, Z., Su, Y., Yang, L., & He, Y. (2015). Design of a highly potent HIV-1 fusion inhibitor targeting the gp41 pocket. <i>Aids</i> , 29(1), 13-21. https://doi.org/10.1097/QAD.0000000000000498
000361977300090	Atkinson, L. (2015). Locating the Politics in Political Consumption: A Conceptual Map of Four Types of Political Consumer Identities. <i>International Journal of Communication</i> , 9, 2047-2066.
000364777400031	Rowland, L., Lobo-do-Vale, R. L., Christoffersen, B. O., Melem, E. A., Kruijt, B., Vasconcelos, S. S., ... Meir, P. (2015). After more than a decade of soil moisture deficit, tropical rainforest trees maintain photosynthetic capacity, despite increased leaf respiration. <i>Global Change Biology</i> , 21(12), 4662-4672. https://doi.org/10.1111/gcb.13035
000346267600010	Krbalek, M., & Sleis, J. (2015). Vehicular headways on signalized intersections: theory, models, and reality. <i>Journal of Physics A-Mathematical and Theoretical</i> , 48(1), 015101. https://doi.org/10.1088/1751-8113/48/1/015101
000369908800022	Zhou, J., Wang, J., Zeng, Y., Zhang, X., Hu, Q., Zheng, J., ... Zhang, W.-M. (2015). Implication of epithelial-mesenchymal transition in IGF1R-induced resistance to EGFR-TKIs in advanced non-small cell lung cancer. <i>Oncotarget</i> , 6(42), 44332-44345. https://doi.org/10.18632/oncotarget.6293
000344595900005	Lotfalian, M., Ranjbar, M., Fazaelpoor, M. H., Schaffie, M., & Manafi, Z. (2015). Continuous Bioleaching of Chalcopyritic Concentrate at High Pulp Density. <i>Geomicrobiology Journal</i> , 32(1), 42-49. https://doi.org/10.1080/01490451.2014.921256
000350285300006	Schuster, L., & Majidi, N. (2015). Deportation Stigma and Re-migration. <i>Journal of Ethnic and Migration Studies</i> , 41(4), 635-652. https://doi.org/10.1080/1369183X.2014.957174

000349637500005	Naim, H. Y. (2015). Measles virus A pathogen, vaccine, and a vector. <i>Human Vaccines & Immunotherapeutics</i> , 11(1), 21-26. https://doi.org/10.4161/hv.34298
000363471600015	McGurk, R. C., Max, C. E., Medling, A. M., Shields, G. A., & Comerford, J. M. (2015). Spatially Resolved Imaging and Spectroscopy of Candidate Dual Active Galactic Nuclei. <i>Astrophysical Journal</i> , 811(1), 14. https://doi.org/10.1088/0004-637X/811/1/14

Note: References are presented in ascending order of UT codes.

Chapitre 4. Discussion

Dans ce chapitre, nous discutons des résultats présentés dans le précédent chapitre et de la façon dont ils apportent des éléments de réponses à nos questions de recherche. Puis, notre discussion fait la synthèse des résultats de notre recherche à la lumière de la littérature existante et de notre cadre théorique. Cette synthèse s'articule autour de quatre thèmes : 1) les limites du co-autorat comme mesure de la collaboration, 2) l'effet Saint-Matthieu et l'effet Matilda, 3) les différences disciplinaires, et 4) les impacts de l'indexation à grande échelle des remerciements. Finalement, nous présentons les principales limites de notre recherche.

4.1 Les fonctions des remerciements dans le système de la reconnaissance scientifique

L'objectif de notre recherche était de décrire les fonctions des remerciements dans le système de la reconnaissance scientifique pour mieux comprendre leur valeur du point de vue du crédit scientifique. Nous avons également exploré les facteurs qui peuvent influencer les caractéristiques des remerciements dans le contexte des pratiques d'attribution du crédit scientifique.

Pour ce faire, nous avons extrait le nombre d'individus mentionnés dans les remerciements et analysé la façon dont le crédit scientifique était distribué au sein de l'ensemble des contributeurs d'un article scientifique (QR1), dans le cadre de l'article 1. Nos résultats montrent une grande variabilité disciplinaire dans la proportion des remerciés et celle des auteurs par rapport à l'ensemble des contributeurs (auteurs et remerciés) d'un article. C'est en physique qu'on observe le plus grand nombre moyen de contributeurs par article, dont une très forte proportion sont des auteurs. En médecine clinique et en sciences biomédicales, on trouve des ratios d'auteurs et de remerciés similaires à ce qui est observé en physique. De manière générale, les disciplines des sciences sociales (champs professionnels, psychologie, sciences de la santé et sciences sociales) sont parmi celles qui montrent les proportions de remerciés les plus élevées par rapport à l'ensemble des contributeurs (auteurs et remerciés) d'un article. Dans la discipline des sciences sociales, le nombre moyen de remerciés est même légèrement supérieur au nombre moyen d'auteurs. Dans les sciences naturelles, la biologie et les sciences de la terre

se démarquent particulièrement par leurs proportions de remerciés qui représentent en moyenne un peu plus du tiers des contributeurs d'un article. Dans l'ensemble, ces résultats montrent que les importantes différences disciplinaires traditionnellement observées en termes de collaboration, telle que mesurée par le co-atorat, sont réduites lorsque l'ensemble des contributeurs d'un article sont pris en compte. Par ailleurs, nos analyses montrent que le nombre moyen de remerciés par article a tendance à diminuer, ou à tout le moins demeurer stable, à mesure que le nombre d'auteurs par article augmente, et ce dans l'ensemble des disciplines. Cette tendance est particulièrement marquée en biologie, dans les champs professionnels et en sciences sociales. Ces résultats appuient l'idée que des pratiques d'atorat moins inclusives pourraient expliquer un nombre moyen d'auteurs par article moins élevé dans ces disciplines, plutôt que de réelles différences dans les niveaux de collaboration de ces disciplines.

Dans le cadre de l'article 2, notre analyse s'est concentrée sur les caractéristiques des individus remerciés dans les publications savantes (QR2). Notre analyse du genre des individus remerciés montre qu'un peu moins du tiers des remerciés de l'ensemble de notre jeu de données sont des femmes. Cette proportion varie selon la discipline étudiée, et en fonction du genre des auteurs principaux d'un article. Toutefois, dans presque tous les cas de figure, les femmes sont sous-représentées chez les remerciés. Les seules exceptions sont en sciences de la santé, où les femmes représentent un peu plus de la moitié des remerciés, et en psychologie, dans le sous-ensemble des articles où les femmes sont auteures principales, une majorité de remerciés sont des femmes.

Pour l'analyse du statut académique des individus remerciés, afin de pouvoir interpréter les résultats obtenus pour les quatre indicateurs bibliométriques compilés (nombre de publications signées en tant qu'auteur, proportion de publications signées en tant qu'auteur principal, nombre de citations reçues comme auteur et âge académique) nous avons comparé la distribution des remerciés (qui sont également auteurs de publications indexées dans WoS) à la distribution de l'ensemble des auteurs de WoS. Une même tendance se dégage pour les quatre indicateurs, les distributions des remerciés sont moins concentrées que celles de l'ensemble des auteurs du WoS. Il en résulte qu'une plus faible proportion des remerciés (également auteurs) sont caractérisés par des valeurs peu élevées pour les quatre incateurs bibliométriques. Ce

résultat se confirme également en termes disciplinaires²⁶. En effet, bien que des différences disciplinaires puissent être observées dans les scores obtenus par les remerciés pour les quatre indicateurs, ces derniers sont caractérisés par des scores plus élevés que l'ensemble des auteurs du WoS dans toutes les disciplines, sans exception, et ce pour les quatre indicateurs. De manière générale, selon les quatre indicateurs analysés, les remerciés également auteurs sont caractérisés dans une plus grande proportion par un statut académique plus élevé que l'ensemble des auteurs du WoS, et ce dans l'ensemble des disciplines.

Afin de mieux décrire les fonctions des remerciements, nous avons également analysé les types de contributions qui mènent aux remerciements (QR3) dans le cadre de l'article 3. Pour y arriver, nous avons extrait les syntagmes nominaux des textes de remerciements et réalisé une analyse de correspondance permettant de mettre en lumière les syntagmes nominaux les plus significatifs pour chaque discipline. La chimie, le génie et la technologie et la physique se trouvent ainsi regroupés autour de mêmes syntagmes nominaux désignant des contributions liées aux tâches techniques, à la communication avec les pairs et au financement. Un second regroupement de disciplines, composé de la médecine clinique, de la psychologie et des sciences de la santé, est caractérisé par des syntagmes nominaux référant principalement au financement, aux déclarations de conflit d'intérêts ainsi qu'au travail expérimental qui implique la participation de sujets humains. La recherche biomédicale est pour sa part caractérisée par des syntagmes nominaux relatifs au financement ainsi qu'à des tâches liées au processus de recherche (notamment à la collecte et à l'analyse des données et à l'assistance technique). Les champs professionnels, les sciences de la terre et les sciences sociales sont quant à elles regroupées autour de syntagmes nominaux référant à la communication avec les pairs et au processus de révision par les pairs. La biologie est caractérisée par des syntagmes nominaux qui réfèrent à la nature spécifique du travail expérimental de cette discipline ainsi qu'à la nature collaborative de celle-ci. Les mathématiques sont pour leur part principalement caractérisées par des syntagmes nominaux relatifs à l'autorat (désignant un ou des auteurs spécifiques). En somme, notre analyse des syntagmes nominaux les plus fréquents dresse un portrait global des principales tendances disciplinaires en termes de types de contributions mentionnées dans les

²⁶ Les résultats disciplinaires pour les quatre indicateurs sont présentés en annexe de l'article 2.

remerciements, mais aussi des aspects de la recherche qui sont valorisés dans certaines disciplines et qui apparaissent comme moins importants dans d'autres.

Finalement, dans l'article 4, l'analyse qualitative de contenu nous a permis d'approfondir les résultats obtenus par l'analyse à grande échelle réalisée dans l'article 3 en nous intéressant aux contextes d'utilisation des termes les plus fréquents des remerciements (QR4). En effet, l'analyse de correspondance des syntagmes nominaux les plus fréquents a révélé que les remerciements n'étaient pas limités à l'attribution de crédit scientifique et que les déclarations préformulées relatives au financement et aux conflits d'intérêts étaient chose courante, voire même prédominantes dans certaines disciplines. Notre analyse qualitative du contexte d'utilisation des termes les plus fréquemment employés dans les remerciements a permis de mettre en lumière trois fonctions principales des remerciements dans le contexte actuel de la communication savante : les remerciements permettent de témoigner de la contribution de collaborateurs non-auteurs, de rendre compte des aspects éthiques liés à la recherche, et d'offrir un espace d'expression personnelle aux auteurs scientifiques. Par ailleurs, pour plusieurs termes fréquents, la stratification disciplinaire permet de révéler des utilisations différentes en fonction de la discipline analysée. Ainsi, dans plusieurs cas, un même terme est utilisé de façon majoritaire pour témoigner de la contribution de collaborateurs dans certaines disciplines, alors qu'il est plutôt utilisé dans le contexte de déclarations liées à l'éthique de la recherche dans d'autres disciplines. Les résultats de notre analyse qualitative suggèrent qu'une utilisation éclairée des remerciements dans les analyses à grande échelle nécessiterait un certain effort de standardisation du champ des remerciements afin de minimalement y délimiter des sections distinctes pour les aspects liés à l'éthique de la recherche, aux contributions et aux témoignages personnels.

4.2 Les limites du co-autorat comme mesure de la collaboration

Les résultats de notre recherche doctorale mettent en lumière les limites de la mesure du co-autorat comme indicateur de la collaboration scientifique et appuient ainsi les études qui ont par le passé souligné ces limites (Katz et Martin, 1997; Laudel, 2002; Lundberg et al., 2006). Notre analyse du nombre de contributeurs par article suggère que les différences traditionnellement

observées entre les disciplines en matière de co-autorat seraient en partie amplifiées par l'indicateur utilisé dans la majorité des études bibliométriques consacrées à la collaboration.

La nature des objets de recherche et le mode d'organisation du travail scientifique qui en découle ont évidemment une influence sur les pratiques de collaboration des chercheurs. En effet, dans les disciplines qui nécessitent l'accès à des infrastructures et équipements coûteux, comme c'est le cas dans le domaine de la recherche biomédicale, la collaboration peut devenir une nécessité afin de partager les coûts et les expertises (Birnholtz, 2006; Katz et Martin, 1997; Pontille, 2016). À l'opposé, dans plusieurs disciplines des sciences humaines et sociales, la recherche est souvent menée de façon individuelle (Ossenblok, Verleysen et Engels, 2014; Pontille, 2004). Toutefois, nos résultats laissent croire qu'au-delà des pratiques de collaboration, il existe, entre les disciplines, des différences dans la façon d'attribuer le crédit scientifique au sein de l'ensemble des contributeurs prenant part à un projet de recherche.

Par ailleurs, dans plusieurs disciplines des sciences naturelles et biomédicales, des auteurs « périphériques », dont les contributions sont jugées comme relativement mineures, occupent une position de moins grande visibilité dans la liste d'auteurs d'un article (Baerlocher et al., 2007; Mongeon et al., 2017; Perneger et al., 2017). De manière parallèle, plusieurs études ont mis de l'avant la fonction de « sous-autorat » (« *sub-authorship* », Patel, 1973) des remerciements qui constitueraient alors une forme de crédit pour le travail de contributeurs périphériques (p. ex. techniciens, étudiants et chercheurs juniors) (Costas et van Leeuwen, 2012; Cronin, 1995; Cronin, Shaw et La Barre, 2003; Cronin, Shaw et La Barre, 2004; Salager-Meyer et al., 2011). Contrairement à cette prémisse positionnant les remerciements comme forme de crédit scientifique de seconde zone pour les contributeurs périphériques, nos analyses du statut académique montrent que les remerciements permettent de souligner les contributions de chercheurs expérimentés.

Notre analyse du statut académique des individus remerciés qui sont également auteurs ne nous permet pas d'associer le statut académique à la contribution pour laquelle les individus sont mentionnés. Toutefois, il a été montré que les auteurs plus expérimentés sont plus fréquemment associés aux tâches conceptuelles et à la contribution de ressources, alors que la contribution des plus jeunes auteurs est plus souvent expérimentale et technique (Larivière et

al., 2016). Une telle division du travail pourrait également être reflétée par les contributions remerciées. En effet, plusieurs études, de même que nos propres analyses des types de contributions remerciées, ont montré que les contributions liées à la révision critique de manuscrits et la contribution de ressources (équipements, matériels et infrastructures) sont fréquemment mentionnées dans les remerciements (Cronin, Shaw et La Barre, 2003; Cronin, Shaw et La Barre, 2004; Díaz-Faes et Bordons, 2014; Paul-Hus et al., 2017b). La surreprésentation des chercheurs expérimentés chez les individus remerciés (ayant déjà publié) pourrait alors être en partie expliquées par le fait que les tâches conceptuelles liées à la révision critique de manuscrits et la contribution de ressources sont plus souvent associées aux chercheurs dont le statut académique est plus élevé. En ce sens, les remerciements adressés aux chercheurs expérimentés nous offrent une fenêtre sur le « collègue invisible » d'un auteur, que Price (1963; 1986) a décrit comme un réseau de chercheurs centré autour d'un même objet de recherche, interagissant de façon informelle pour se partager des idées et articles en préparation.

Selon la perspective de Bourdieu, les pratiques d'attribution du crédit scientifique refléteraient la structure des champs disciplinaires où les individus partagent un même *habitus*, spécifique à une discipline (Bourdieu, 1975). Les actions et pratiques des chercheurs étant orientés vers l'acquisition de capital symbolique, c'est-à-dire de reconnaissance de la part de leurs pairs, elles doivent pour ce faire se conformer à la structure des forces dominantes du champ disciplinaire et aux façons de faire propre à ce champ. Les comportements attendus et façons de faire étant définis par l'intérêt des dominants dans un champ, la position hiérarchique d'un chercheur dans son champ disciplinaire pourrait alors avoir un effet sur ses pratiques d'attribution du crédit scientifique. Selon cette hypothèse, un chercheur en position d'autorité pourrait surestimer sa contribution et sous-estimer la contribution de ses collaborateurs, de manière consciente ou non, et ainsi s'approprier une plus grande part de reconnaissance en excluant certains collaborateurs de la liste d'auteurs, en les reléguant aux remerciements. Une telle situation constituerait une manifestation de l'effet Saint-Matthieu et de l'accumulation des avantages décrits par Merton (1968), et discutés plus avant dans la section suivante.

Nos résultats laissent croire que les pratiques d'attribution du crédit scientifique ne sont pas uniquement le fait de cultures épistémiques différentes, c'est-à-dire de pratiques de création

des connaissances particulières à une discipline (Knorr-Cetina, 2009), mais peuvent également être le reflet de la structure du champ disciplinaire et du rapport de forces entre les contributeurs d'une même équipe de recherche. Nos résultats vont dans le sens des conclusions de Bozeman et Youtie (2016) qui ont montré que plusieurs facteurs peuvent expliquer l'exclusion d'un contributeur de la liste d'auteurs d'un article, notamment des différences d'expérience et de pouvoir entre les individus, ainsi que les dynamiques de genre (qui interagiraient avec les dynamiques de pouvoir).

En somme, nos résultats suggèrent que l'utilisation des remerciements au côté de l'aurat permet d'obtenir un portrait plus complet de la collaboration scientifique à l'heure actuelle. Toutefois, étant donné le caractère peu encadré du contenu des remerciements, il ne peut être exclu que certains individus soient remerciés pour des contributions de nature non scientifiques.

4.3 L'effet Saint-Matthieu et l'effet Matilda

Nos analyses des caractéristiques des individus remerciés constituent des manifestations empiriques de l'effet Saint-Matthieu (Merton, 1968), dans le cas du statut académique des remerciés, et de l'effet Matilda (Rossiter, 1993), dans le cas du genre des remerciés.

Notre analyse du statut académique, qui se limite aux remerciés ayant déjà publié, montre que les chercheurs expérimentés sont surreprésentés chez les individus remerciés. En effet, une même tendance se dégage pour les quatre indicateurs bibliométriques considérés (le nombre de publications signées en tant qu'auteur, le nombre de citations normalisées par discipline, l'âge académique ainsi que la proportion des articles pour laquelle un remercié occupe une position d'auteur principal) : une plus grande proportion des individus remerciés est caractérisée par un statut académique élevé, comparativement à la population des auteurs du WoS. De plus, cette tendance est présente pour toutes les disciplines analysées.

L'effet Saint-Matthieu, qui consiste à attribuer une part de reconnaissance disproportionnée aux chercheurs renommés et, en contrepartie, priver d'une telle reconnaissance les chercheurs moins renommés, amplifierait l'accumulation d'avantages pour les uns et de

désavantages pour les autres (Cole et Cole, 1973; Merton, 1968). Ainsi, un remerciement adressé à un chercheur expérimenté pourrait en partie être motivé par la renommée déjà existante de ce chercheur. En retour, le remerciement participerait à amplifier la visibilité de ce dernier dans la communauté scientifique. Selon cette prémisse, à contribution égale, un chercheur moins expérimenté ne serait sans doute pas mentionné dans les remerciements. Nos résultats, qui montrent une présence accrue de chercheurs expérimentés chez les remerciés, tendent à appuyer cette hypothèse.

Par ailleurs, nos résultats montrent que les disparités de genre maintes fois démontrées en matière d'autorat sont également présentes dans les remerciements. En moyenne, les femmes représentent environ le tiers des auteurs et des remerciés du corpus analysé. La moins grande visibilité des femmes au sein des remerciements s'ajoute aux nombreuses démonstrations des inégalités entre les hommes et les femmes en recherche, qui ont été mesurées en matière d'autorat, de citations, de types de contributions, de financement et de salaire (Kaatz et al., 2016; Larivière et al., 2013; Ley et Hamilton, 2008; Macaluso et al., 2016; Shen, 2013). Nos résultats appuient donc la thèse de l'effet Matilda (Rossiter, 1993) qui désigne la tendance à minimiser les contributions scientifiques faites par des femmes ou à les attribuer à des hommes.

Nos résultats montrent également que les femmes ont tendance à remercier une proportion plus élevée de femmes. En effet, la proportion de femmes remerciées est plus élevée lorsqu'une ou des femmes sont auteures principales de l'article que lorsque ce sont un ou des hommes qui sont auteurs principaux, et ce pour toutes les disciplines. Bien que notre devis de recherche ne nous permette pas d'affirmer ou d'infirmer le rôle du genre des individus sur les pratiques d'attribution du crédit scientifique, nos résultats indiquent que dans les remerciements, les femmes soulignent la contribution d'une plus grande proportion de femmes. Ce résultat laisse supposer que les femmes ont tendance à s'entourer d'une plus grande proportion de femmes dans leurs équipes de recherche et vient donc appuyer l'hypothèse de l'homophilie dans la formation des liens et réseaux d'individus, c'est-à-dire la tendance à s'associer à des pairs présentant des caractéristiques semblables (McPherson, Smith-Lovin et Cook, 2001). Toutefois, dans une étude s'intéressant au rôle de l'homophilie dans la collaboration scientifique, Araújo et al. (2017) ont trouvé que les hommes collaborent dans une proportion plus importante avec

des hommes alors que les femmes collaborent avec une proportion égalitaire d'hommes et de femmes, en ce sens, les femmes seraient peut-être plus égalitaires dans leurs pratiques de remerciements que les hommes.

Bien que Merton n'aborde pas directement la question du genre lorsqu'il présente l'effet Saint-Matthieu (Merton, 1968), les biais liés au genre des chercheurs participent au même processus d'accumulation des avantages (Rossiter, 1993). Dès lors, qu'un biais soit dû au statut académique d'un individu ou à son genre, il aura le même effet de concentration de la visibilité et la reconnaissance pour les chercheurs possédant déjà visibilité et reconnaissance. Pour Bourdieu, « parler d'*habitus*, c'est poser que l'individuel, et même le personnel, le subjectif, est social, collectif. L'*habitus* est une subjectivité socialisée » (1992, p.101). En ce sens, l'*habitus* d'un chercheur est structuré par le champ disciplinaire dans lequel il évolue, mais également par son statut dans ce champ ainsi que par son genre. La discipline, le statut académique et le genre de l'individu agissent et interagissent donc comme un système de schèmes qui guide ses actions. Les rapports de forces et les stratégies mises en place pour acquérir du capital symbolique sont donc forcément influencés par les *habitus* d'un individu. Pour Merton comme pour Bourdieu, le statut et le genre d'un individu interagissent avec la structure d'une discipline et influencent les relations entre les individus et leur capacité à acquérir la reconnaissance de leurs pairs.

En somme, nos résultats montrent que les biais de genre et les effets liés au statut académique des individus, généralement observés en matière d'atorat, se révèlent également dans les remerciements.

4.4 Les différences disciplinaires

L'ensemble de nos résultats souligne l'importance du contexte disciplinaire pour appréhender les pratiques d'attribution du crédit scientifique, et plus précisément les pratiques de remerciements. Nos analyses du nombre d'auteurs par article mettent en lumière d'importantes différences disciplinaires en matière d'atorat. Dans un premier temps, nos résultats corroborent ce qui est démontré dans la littérature depuis plus de dix ans (notamment Larivière, Gingras et Archambault, 2006; Wuchty, Jones et Uzzi, 2007), à savoir que le champ disciplinaire des

sciences sociales²⁷ est caractérisé par une plus faible collaboration, lorsque mesurée par le nombre d'auteurs par article. Toutefois, l'originalité de nos résultats tient à ce que nous considérons le nombre total de contributeurs d'un article, c'est-à-dire les auteurs et les remerciés. Lorsque l'ensemble des contributeurs sont pris en compte, les disciplines des sciences sociales montrent des niveaux de collaboration comparables à ceux observés dans les disciplines des sciences naturelles (chimie, et génie et technologie). De manière générale, nos résultats montrent que les importantes différences disciplinaires traditionnellement observées en termes de taille d'équipe (et mesurée par le co-autorat) sont grandement réduites lorsque les remerciés sont pris en compte.

Les différences disciplinaires en termes de nombre d'auteurs par article semblent donc être le reflet de différences dans les pratiques d'attribution du crédit scientifique, plutôt que strictement dans les pratiques de collaboration. Cette hypothèse est appuyée par les tendances observées en ce qui a trait au nombre de remerciés par article en fonction du nombre d'auteurs. En effet, nos résultats montrent que pour toutes les disciplines analysées, le nombre moyen de remerciés par article diminue ou demeure stable à mesure que le nombre moyen d'auteurs augmente. Ainsi, le nombre d'auteurs plus faible caractérisant certaines disciplines, notamment dans le champ disciplinaire des sciences sociales, pourrait être le reflet de pratiques d'attribution de l'autorat moins inclusives dans ces disciplines.

Par ailleurs, notre analyse comparative des syntagmes nominaux contenus dans les remerciements souligne également des différences disciplinaires, cette fois en matière de types de contributions remerciées. Nos résultats fournissent la première démonstration empirique à grande échelle des variations qui existent entre les disciplines dans le contenu des remerciements. Ces variations révèlent l'importance accordée à certains aspects de la recherche, ou à tout le moins dans la façon de les exprimer dans les remerciements. Sans pouvoir conclure quant aux types de contributions qui mènent à l'autorat et aux contributions qui mènent aux

²⁷ Selon la classification de la NSF (2006), le champ disciplinaire des sciences sociales inclut la psychologie, les sciences de la santé, les sciences sociales et les champs professionnels.

remerciements, nos résultats mettent en lumière les différences entre les disciplines dans les aspects du travail de recherche qui sont valorisés et ainsi jugés « dignes de mention ».

Finalement, notre analyse qualitative des remerciements montre également des différences disciplinaires dans l'utilisation de certains termes. Ces différences viennent encore une fois souligner l'importance du contexte disciplinaire dans l'analyse des pratiques des auteurs et nous renvoient encore une fois au concept d'*habitus* de Bourdieu (1984). Ainsi, le contenu des remerciements refléterait l'*habitus* propre à un champ disciplinaire, c'est-à-dire une partie de ses enjeux, de ses façons de faire, de ses objets de recherche, de ce qui y est valorisé. Pour Merton, ce sont plutôt les objectifs communs à une discipline qui permettent d'expliquer la façon dont le travail est organisé, divisé et valorisé. Pour Merton comme pour Bourdieu, le champ disciplinaire agit comme force structurant l'organisation du travail scientifique et les pratiques qui en découlent. Le fait que certains termes soient employés dans des contextes qui varient selon les disciplines tend à confirmer que les champs disciplinaires possèdent des façons de faire qui leur sont propres.

En somme, l'ensemble de nos résultats montre qu'une utilisation éclairée des remerciements, que ce soit dans le cas d'analyses de la collaboration ou des types de contributions, doit obligatoirement prendre en compte le contexte disciplinaire.

4.5 Les impacts de l'indexation à grande échelle des remerciements

Durant le cours de notre recherche doctorale, en plus des résultats présentés au chapitre 3, nous avons effectué une synthèse de la littérature sur les remerciements (Desrochers, Paul-Hus et Pecoskie, 2017) ainsi qu'une analyse descriptive et une caractérisation des données de remerciements disponibles dans la base de données WoS (Paul-Hus, Desrochers et Costas, 2016). Ces deux publications ne sont pas incluses dans les résultats de notre thèse, mais font partie de nos travaux sur les remerciements qui ont permis d'éclairer et d'orienter la méthodologie utilisée dans notre thèse.

Les résultats présentés dans notre thèse, de même que les conclusions de notre synthèse de la littérature et de notre analyse descriptive des données de remerciements vont dans le même sens et participent tous, selon différentes perspectives, à mettre en lumière une des conséquences

de l'indexation à grande échelle des remerciements dans le WoS, soit la création de deux corpus distincts de remerciements : les remerciements rédigés en anglais qui incluent une mention de financement et qui sont donc disponibles à l'échelle agrégée dans WoS, et les autres remerciements, ceux rédigés dans des langues autres que l'anglais, qui n'incluent pas de mention de financement et auxquels WoS ne donne pas accès.

Notre analyse descriptive des données indexées par WoS souligne les nombreuses limites liées à l'utilisation de ce corpus et pose des balises précises en ce qui a trait à ce qui est inclus dans ce corpus et ce qui en est exclu, notamment les disciplines des arts et des humanités, les disciplines des sciences sociales avant l'année 2015, tous les remerciements qui n'incluent pas clairement une mention de financement, les remerciements rédigés dans une langue autre que l'anglais, ainsi que les remerciements provenant de publications autres que les articles originaux et les articles de synthèse (p. ex. les éditoriaux, les notes et les articles d'actes de conférences).

Notre synthèse de la littérature (Desrochers, Paul-Hus et Pecoskie, 2017) révèle quant à elle une bifurcation importante vers l'étude de l'aspect financier de la recherche et des impacts du financement dès la création du nouveau corpus de remerciements dans WoS en 2008. Notre analyse des syntagmes nominaux contenus dans les remerciements met également de l'avant l'importance de l'aspect financier dans les remerciements. Notre analyse qualitative du contenu des remerciements montre quant à elle que les typologies des remerciements développées au cours des dernières décennies (Cronin, McKenzie et Stiffler, 1992; Cronin et al., 1993; Cronin, McKenzie et Rubio, 1993; Cronin et Weaver, 1995; Mackintosh, 1972; McCain, 1991) ne reflètent plus le contenu actuel des remerciements, du moins pour le corpus indexé par WoS. En effet, nos analyses soulignent non seulement l'importance de l'aspect financier de la recherche, mais également des aspects liés à l'éthique de la recherche (notamment les déclarations de responsabilité et de conflits d'intérêts).

Nos analyses des remerciés (nombre de remerciés par article et caractérisation des individus remerciés) témoignent quant à elles des limites identifiées par notre description du corpus indexé par WoS. En effet, nos analyses des remerciés sont limitées aux publications parues à partir de 2015 dans les disciplines des sciences naturelles, biomédicales et sociales. Nos résultats témoignent donc du fait que la recherche sur les remerciements et les questions

qui peuvent y être abordées sont, depuis 2008, en partie déterminées par la source de données utilisée.

En somme, les politiques d'indexation des remerciements de WoS ont eu un effet de schisme dans l'étude des remerciements. À l'heure actuelle, au-delà des choix épistémologiques et méthodologiques qui peuvent être faits, certains aspects et certaines questions (p. ex. en ce qui a trait aux langues ou à la recherche non financée) ne peuvent tout simplement pas être étudiés à l'échelle agrégée à l'aide des données de WoS et se voient relégués aux analyses à une autre échelle, faute de corpus à grande échelle.

4.6 Limites de la recherche

Une limite générale de notre recherche est liée aux méthodes quantitatives employées qui visent l'identification de tendances générales par des analyses de grands jeux de données. La contrepartie de l'analyse à grande échelle est évidemment le besoin de simplifier un phénomène complexe. Notre recherche fournit un aperçu des variables qui entrent en jeu dans les pratiques d'attribution du crédit au sein des contributeurs d'un projet de recherche, mais ne permet pas de comprendre ces pratiques de façon approfondie. Nos analyses quantitatives nous fournissent des tendances généralisables quant au contenu des remerciements, mais ne nous donnent pas accès à une compréhension approfondie des décisions entourant l'attribution du crédit scientifique dans une équipe de recherche et de la façon dont les remerciements sont rédigés dans ce contexte, comme pourraient le faire des entrevues avec des chercheurs. Notre analyse qualitative des termes fréquents dans les remerciements vise justement à utiliser le contexte pour approfondir notre compréhension du contenu des remerciements, mais cette contextualisation est limitée à l'analyse des termes fréquemment employés dans les remerciements. De plus, nos analyses quantitatives portent sur des données bibliométriques qui, bien que disponibles à grande échelle, ne nous donnent accès qu'à un reflet indirect du processus d'attribution du crédit scientifique à travers les remerciements et les listes d'auteurs des articles scientifiques. Nous n'observons pas directement les pratiques des chercheurs, mais plutôt leurs résultats sous forme de données bibliométriques (MacRoberts et MacRoberts, 2018).

Des limites plus spécifiques s'appliquent également à notre recherche. Premièrement, tel que discuté précédemment, la source de données utilisée (WoS) présente certaines limites. Les remerciements sont indexés et disponibles dans la base de données WoS seulement s'ils sont rédigés dans la *lingua franca* de la science, l'anglais et qu'ils incluent une mention de financement. Ces critères d'indexation induisent donc un biais potentiel envers la recherche financée et les analyses du contenu des remerciements peuvent donc, en partie, refléter ce biais vers l'aspect financier de la recherche. La limite linguistique entraîne aujourd'hui un manque de connaissance quant à l'impact potentiel de la langue sur les pratiques de remerciements dans les publications savantes. De plus, les remerciements des articles en arts et humanités ne sont pas indexés par WoS à l'heure actuelle. Ces éléments limitent donc la généralisabilité de nos résultats aux remerciements rédigés en anglais, qui découlent de la recherche financée dans les disciplines des sciences naturelles, biomédicales et sociales.

En ce qui a trait à l'analyse des individus remerciés, notre unité d'analyse est le nom des individus. Identifier et dénombrer le nombre d'individus mentionnés dans les remerciements permet d'obtenir des tendances globales du nombre de remerciés par article et par discipline. En relation avec le nombre d'auteurs qui signent ces mêmes articles, nous obtenons un portrait approximatif du nombre d'individus qui ont contribué d'une façon ou d'une autre au projet de recherche faisant l'objet de la publication. Toutefois, nos analyses ne nous permettent pas de mettre en relation les individus remerciés avec les raisons pour lesquelles ils sont mentionnés. Ainsi, les individus peuvent être mentionnés dans les remerciements pour des raisons de natures très diverses, qui peuvent dans certains cas être assez éloignées du travail de recherche. Un exemple de remerciements ayant fait les manchettes pour sa nature inusitée incluait notamment une demande en mariage destinée à la compagne d'un des auteurs de l'article : « C.M.B. would specifically like to highlight the ongoing and unwavering support of Lorna O'Brien. Lorna, will you marry me? » (Brown et Henderson, 2015). Dans cet exemple précis, avec nos méthodes, le nom de la compagne aurait été compté comme contributrice au projet de recherche. Les résultats de nos analyses demeurent donc une approximation du nombre de contributeurs d'un projet de recherche, mais ne nous permettent pas d'évaluer la nature des contributions des individus mentionnés. De plus, nos méthodes ne permettent pas de comptabiliser les groupes ou individus

qui ne sont pas remerciés par leur nom (notamment dans le cas de remerciements adressés aux « auxiliaires de recherche » et aux « techniciens » qui ont participé à un projet).

L'identification du statut d'auteur des individus remerciés présente également certaines limites. La désambiguïsation des individus et de leur profil d'auteur n'est jamais une opération parfaite lorsque réalisée à très grande échelle de manière automatisée, et la méthode que nous avons privilégiée ne fait pas exception (Caron et van Eck, 2014). Afin d'éviter les faux positifs dans l'identification des paires remerciés-auteurs, nous avons adopté des critères plus restrictifs qui ont pour effet de privilégier la précision plutôt que le rappel. Ce choix a pour conséquence de restreindre la taille de notre échantillon en ce qui concerne les individus remerciés qui sont également auteurs, mais qui nous permet de caractériser cet échantillon avec plus de certitude. En contrepartie, notre méthode ne nous permet pas d'avoir une estimation juste de la proportion totale des remerciés qui sont également auteurs dans l'ensemble de notre jeu de données puisque nos critères visaient plutôt l'identification précise des remerciés également auteur.

Conclusion

Dans ce dernier chapitre, nous présentons d'abord un résumé chronologique de notre recherche doctorale. Puis, nous présentons de manière synthétique les principaux apports de notre thèse. Finalement, nous proposons quelques pistes pour des recherches futures.

Résumé de la recherche

Comme dans tous les parcours doctoraux, notre projet de recherche a évolué et nos objectifs se sont précisés au cours des cinq dernières années. Notre recherche doctorale a débuté par un projet de revue systématique de la littérature sur les remerciements qui a finalement pris la forme d'une synthèse analytique de ce corpus de littérature (Desrochers, Paul-Hus et Pecoskie, 2017). Cette synthèse nous a permis de situer notre projet au sein du champ de recherche sur les remerciements, de constater les questions qui avaient été abordées dans ce champ, et surtout les aspects qui demeuraient à explorer. Cette synthèse a notamment permis de mettre en lumière les impacts de l'indexation à grande échelle des remerciements sur l'objet de recherche lui-même et sur les méthodes d'investigation employées pour analyser les remerciements. Plusieurs études mentionnaient le flou entourant les critères d'indexation des remerciements dans la base de données WoS. En effet, une couverture incomplète ou partielle des remerciements dans WoS aurait nécessairement des impacts sur les analyses qui pouvaient être faites avec ces données (Costas et van Leewen, 2012; Díaz-Faes et Bordons, 2014; Rigby, 2011). Nous avons donc effectué une analyse systématique des données de remerciements indexées par WoS afin d'en comprendre les limites et de pouvoir en faire une utilisation éclairée (Paul-Hus, Desrochers et Costas, 2016). Cette caractérisation des données de remerciements nous a permis de mieux définir les questions de recherche qu'il était possible d'aborder grâce à ce jeu de données et de préciser notre méthodologie.

Notre recherche doctorale visait d'abord à décrire les fonctions des remerciements dans le système de la reconnaissance scientifique pour mieux comprendre leur valeur en termes de crédit scientifique, ainsi qu'à explorer les facteurs qui peuvent influencer leurs caractéristiques. Nous nous sommes d'abord intéressés à la façon dont le crédit scientifique était distribué parmi

l'ensemble des contributeurs d'un projet de recherche, soit les auteurs et les remerciés (Paul-Hus et al., 2017a). Cette analyse du nombre d'auteurs et de remerciés par article a permis d'élargir la notion de collaboration en recherche en y intégrant les individus remerciés, et de souligner l'importance du contexte disciplinaire dans les pratiques d'attribution du crédit scientifique.

Nous avons ensuite entrepris l'analyse du contenu des remerciements afin d'identifier les types de contributions qui mènent aux remerciements. Plus précisément, nous avons opérationnalisé la notion de contributions remerciées par les syntagmes nominaux extraits des remerciements. Au moyen d'une analyse de correspondance, nous avons mis en lumière d'importantes variations disciplinaires dans l'emploi des syntagmes nominaux les plus fréquents, et donc des types de contributions fréquemment mentionnées (Paul-Hus et al., 2017b). Cette analyse quantitative des syntagmes nominaux a également révélé la présence de déclarations préformulées, souvent suggérées, parfois imposées par les organismes de financement et les politiques éditoriales de certaines revues. Les résultats de cette analyse quantitative, et plus particulièrement la présence de phrases préformulées, sont venus appuyer la nécessité d'une analyse qualitative du contenu des remerciements afin de contextualiser les usages des termes les plus fréquemment employés.

Notre analyse qualitative des termes extraits des remerciements a non seulement apporté une compréhension approfondie des usages de ces termes dans leur contexte d'origine, mais a également mené à une nouvelle typologie des remerciements (Paul-Hus et Desrochers, 2019). Notre typologie va au-delà des types de contributions traditionnellement mises de l'avant par les typologies et classifications de remerciements (notamment Mackintosh, 1972; McCain, 1991; Cronin et Weaver, 1995). Ainsi, notre typologie reflète les diverses fonctions que remplissent aujourd'hui les remerciements dans les publications scientifiques, c'est-à-dire de rendre compte des aspects éthiques liés à la recherche, de témoigner de la contribution de collaborateurs non-auteurs et d'offrir un espace où la voix des auteurs peut s'exprimer librement.

En parallèle avec notre analyse qualitative du contenu des remerciements, nous avons effectué une analyse des caractéristiques des individus remerciés afin d'explorer les effets du genre et du statut académique sur les pratiques d'attribution du crédit scientifique (Paul-Hus et

al., 2020). Nos résultats suggèrent les pratiques de remerciements sont associées au genre et au statut académique des individus remerciés.

Apports de la recherche

Apports empiriques

Les principaux apports de notre recherche sont des démonstrations empiriques qui participent à la validation de concepts théoriques ou de phénomènes qui n'avaient jusqu'alors pas été démontrés à grande échelle, telles que des démonstrations empiriques des différences disciplinaires dans les pratiques d'attribution du crédit scientifique qui permettent d'élargir la notion de collaboration. Notre analyse des caractéristiques des individus remerciés révèle la sous-représentation des femmes chez les individus remerciés et constitue une manifestation empirique supplémentaire appuyant la thèse de l'effet Matilda (Rossiter, 1993). Notre analyse du statut académique des individus remerciés révèle quant à elle une surreprésentation des chercheurs expérimentés au sein des individus remerciés qui ont déjà publié, comparativement à l'ensemble des auteurs du WoS. Ce résultat pourrait constituer une manifestation empirique de l'effet Saint-Matthieu (Merton, 1968) dans laquelle un chercheur expérimenté reçoit du crédit sous forme de remerciement en raison même de sa renommée dans la communauté scientifique.

Apports méthodologiques

Notre recherche apporte également des contributions originales aux méthodes bibliométriques et aux approches méthodologiques employées pour l'étude des remerciements. Nous proposons ainsi l'utilisation du nombre total de contributeurs d'une publication, auteurs et individus remerciés²⁸, comme indicateur bibliométrique de la collaboration en recherche. Cette méthode permet, en partie, de pallier les limites du co-autorat comme indicateur de la collaboration. En effet, toutes les collaborations ne mènent pas à l'autorat et l'inclusion des individus remerciés

²⁸ Notre méthode se limite aux individus remerciés désignés par leur nom puisque les groupes ou individus désignés uniquement par leur titre ou leur fonction ne peuvent être comptés (p. ex. les auxiliaires de recherche, les techniciens et les évaluateurs).

permet d'obtenir un portrait plus complet du nombre d'individus ayant contribué, d'une façon ou d'une autre, à un projet de recherche.

Par ailleurs, les importantes variations disciplinaires dans les pratiques des remerciements mettent en lumière les limites liées à l'utilisation des remerciements comme indicateur bibliométrique dans le contexte de l'évaluation de la recherche au niveau individuel. En effet, plusieurs études ont suggéré l'utilisation des remerciements pour évaluer les contributions individuelles des chercheurs (Costas et van Leewen, 2012; Cronin, 1991; Cronin et Weaver, 1995; McCain, 2018). Les différences disciplinaires révélées par nos analyses soulignent toutefois l'importance qui devrait être accordée au contexte disciplinaire dans la mise en place d'un indicateur centré sur les remerciements, afin qu'un tel indicateur ne participe pas à amplifier les biais disciplinaires qui caractérisent déjà les indicateurs bibliométriques centrés sur l'aurat.

Notre recherche contribue au champ d'études des remerciements, et plus largement aux études bibliométriques de par le devis mixte employé. En effet, les études quantitatives ont souvent appelé à la contextualisation des résultats à l'aide de méthodes qualitatives (dans le champ d'études des remerciements : Costas et van Leewen, 2012; Cronin, 2001; Liu et al., 2015), mais, à notre connaissance, aucune étude récente n'avait combiné les méthodes quantitatives et qualitatives pour l'étude des remerciements. Ainsi, notre devis de recherche mixte permet d'allier la force des méthodes quantitatives pour établir des tendances générales à l'analyse qualitative de contenu, qui permet d'approfondir et de contextualiser les résultats obtenus à une autre échelle.

Apports pratiques

Notre analyse du contenu des remerciements nous a permis d'établir des recommandations pour la mise en place de champs d'indexation distincts pour les remerciements, si un effort de standardisation était entrepris. Ces champs devraient minimalement inclure une section dédiée à l'éthique de la recherche (déclarations de financement, de responsabilités, et de conflit d'intérêts), aux contributions faites à la recherche par l'ensemble des contributeurs, et au témoignage personnel des auteurs. L'instauration de tels champs favoriserait les analyses à grande échelle d'aspects plus précis et mieux balisés des remerciements.

Recherches futures

De nombreuses avenues de recherche demeurent à explorer pour mieux comprendre les pratiques d'attribution du crédit scientifique. Dans un premier temps, nous avons étudié la taille des équipes de recherche en combinant le nombre d'auteurs et le nombre de remerciés par article. Dans un deuxième temps, nous avons étudié les caractéristiques des individus remerciés (genre et statut académique). Il pourrait maintenant être intéressant de croiser ces deux éléments et d'explorer la relation entre les caractéristiques des auteurs et le nombre de contributeurs crédités dans un article. Le genre et le statut académique des auteurs ont-ils un effet sur le nombre d'individus remerciés?

D'autres méthodes d'investigation et d'analyse pourraient permettre d'approfondir notre compréhension des fonctions des remerciements dans le système de la reconnaissance scientifique ainsi que des facteurs qui en influencent le contenu. Par exemple, des études qualitatives qui utiliseraient des questionnaires ou des entrevues auprès d'équipes de recherche pourraient permettre de mieux comprendre les décisions concernant l'attribution du statut d'auteur et de remercié, leur contexte, et leur justification. Par ailleurs, étant donné l'importance du contexte disciplinaire dans les pratiques d'attribution du crédit scientifique, et plus précisément dans les pratiques de remerciements, les équipes interdisciplinaires fournissent un cas intéressant à analyser puisqu'une collaboration entre des chercheurs provenant de diverses disciplines laisse supposer un croisement des *habitus* de division et d'organisation du travail, ainsi que des pratiques distinctes d'attribution du crédit scientifique.

Le projet de créer un index des remerciements pour évaluer la contribution individuelle des chercheurs, similaire aux index de citations, comme l'avait envisagé Cronin dans les années 1990, nous semble désormais révolu. L'index des remerciements prend aujourd'hui la forme d'une indexation des sources de financement mentionnées dans les remerciements, qui inclut au passage les individus remerciés. Notre analyse qualitative du contenu des remerciements a notamment mené à la recommandation pour l'instauration de champs distincts dans les remerciements si l'on désire continuer à utiliser ces données pour analyser les pratiques des chercheurs. Toutefois, notre recherche doctorale nous amène aujourd'hui à nous interroger sur

la volonté réelle de la communauté scientifique à standardiser l'espace peu balisé que constituent les remerciements dans les publications scientifiques. Ainsi, peut-être que le flou qui caractérise ce champ sied bien à l'ambiguïté et la subjectivité qui entourent l'application des critères d'autorat, et offre une latitude qui permet d'accommoder les rôles très divers et parfois mal définis des différents contributeurs d'un projet de recherche. Par ailleurs, certaines initiatives récentes contribuent certainement à mieux définir le rôle de l'ensemble des contributeurs d'un projet de recherche. Mentionnons notamment l'adoption par plus de 120 revues scientifiques de la taxonomie détaillée Contributor Role Taxonomy (CRediT) depuis son introduction en 2014 (Allen, O'Connell et Kiermer, 2019).

Cela dit, une attention particulière doit être accordée à la nature des données de remerciements analysées, à leur source, ainsi qu'aux agents que l'on sert lorsque l'on analyse de telles données. En effet, la recherche à grande échelle sur les remerciements demeure aujourd'hui restreinte aux données propriétaires détenues par WoS et Scopus, et présente de ce fait les mêmes biais disciplinaires et linguistiques qui caractérisent ces bases de données, en plus du biais financier lié aux critères d'indexation des remerciements. Par ailleurs, nos analyses montrent que si l'on désire continuer à utiliser les remerciements pour tirer des conclusions sur les pratiques académiques, les recherches devront prendre en considération les effets liés à la hiérarchie académique et aux relations de pouvoir au sein des équipes de recherche, aux biais liés au genre des individus, ainsi qu'à l'influence du financement sur le contenu des remerciements.

Bibliographie

- Al-Ali, M. N. (2010). Generic patterns and socio-cultural resources in acknowledgements accompanying Arabic Ph.D. dissertations. *Pragmatics*, 20(1), 1-26. <https://doi.org/10.1075/prag.20.1.01ali>
- Allen, L., O'Connell, A., et Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, 32(1), 71-74. <https://doi.org/10.1002/leap.1210>
- Alwi, I. (2007). Tips and tricks to make case report. *Acta medica Indonesiana*, 39(2), 96-98.
- Araújo, E. B., Araújo, N. A. M., Moreira, A. A., Herrmann, H. J. et Andrade Jr, J.S. (2017). Gender differences in scientific collaborations : Women are more egalitarian than men. *PLoS ONE*, 12(5), e0176791. <https://doi.org/10.1371/journal.pone.0176791>
- Atkins, H. (2016). Author credit : PLOS and CRediT update. The Official PLOS Blog website: <https://blogs.plos.org/plos/2016/07/author-credit-plos-and-credit-update/>
- Averna, M. (2015). The effect of ezetimibe on NAFLD. *Atherosclerosis Supplements*, 17, 27-34. [https://doi.org/10.1016/S1567-5688\(15\)50007-X](https://doi.org/10.1016/S1567-5688(15)50007-X)
- Baerlocher, M. O., Newton, M., Gautam, T., Tomlinson, G. et Detsky, A. S. (2007). The meaning of author order in medical research. *Journal of Investigative Medicine: The Official Publication of the American Federation for Clinical Research*, 55(4), 174-180. <https://doi.org/10.2310/6650.2007.06044>
- Baggs, J. G. (2008). Issues and rules for authors concerning authorship versus acknowledgements, dual publication, self-plagiarism, and salami publishing. *Research in Nursing & Health*, 31(4), 295-297. <https://doi.org/10.1002/nur.20280>
- Bates T, Anić A, Marušić M. et Marušić A. (2004). Authorship criteria and disclosure of contributions : Comparison of 3 general medical journals with different author contribution forms. *JAMA*, 292(1), 86-88. <https://doi.org/10.1001/jama.292.1.86>

- Beaver, D. et Rosen, R. (1978). Studies in scientific collaboration Part I. *Scientometrics*, 1(1), 65-84. <https://doi.org/10.1007/BF02016840>
- Beaver, D. et Rosen, R. (1979a). Studies in scientific collaboration Part II. *Scientometrics*, 1(2), 133-149. <https://doi.org/10.1007/BF02016966>
- Beaver, D. et Rosen, R. (1979b). Studies in scientific collaboration Part III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics*, 1(3), 231-245. <https://doi.org/10.1007/BF02016308>
- Beh, E. J. et Lombardo, R. (2014). *Correspondence analysis : Theory, practice and new strategies*. Hoboken, NJ: John Wiley & Sons.
- Ben-Ari, E. (1987). On acknowledgments in ethnographies. *Journal of Anthropological Research*, 43(1), 63-84.
- Benzécri, J.-P. (1980). *Pratique de l'analyse des données*. Paris: Dunod.
- Benzécri, J.-P. (1992). *Correspondance analysis handbook*. New York: Marcel Dekker.
- Bérubé, N., Sainte-Marie, M., Ghiasi, G. et Larivière V. (2020). Wiki-Gendersort: Automatic gender detection using first names in Wikipedia. <https://osf.io/preprints/socarxiv/ezw7p>
- Biagioli, M. (1998). The instability of authorship : Credit and responsibility in contemporary biomedicine. *The FASEB Journal*, 12(1), 3-16.
- Biagioli, M. (2003). Rights or rewards ? Changing frameworks of scientific authorship. Dans P. Galison et M. Biagioli (Dir.), *Scientific authorship : Credit and intellectual property in science* (p. 255-279). New York, NY: Routledge.
- Bing, J. et Ruhl, C. (2008). It's all my fault ! The pragmatics of responsibility statements. *Journal of Pragmatics*, 40(3), 537-558. <https://doi.org/10.1016/j.pragma.2007.04.010>
- Bird, S., Klein, E. et Loper, E. (2009). *Natural language processing with Python : Analyzing text with the Natural Language Toolkit*. Sebastopol, CA: O'Reilly Media.

- Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, 57(13), 1758-1770.
<https://doi.org/10.1002/asi.20380>
- Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing : Experimental evidence from The American Economic Review. *American Economic Review*, 81(5), 1041-1067.
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45(1), 197-245. <https://doi.org/10.1002/aris.2011.1440450112>
- Bosch, X., Pericas, J. M., Hernández, C. et Torrents, A. (2012). A comparison of authorship policies at top-ranked peer-reviewed biomedical journals. *Archives of Internal Medicine*, 172(1), 70-72. <https://doi.org/10.1001/archinternmed.2011.600>
- Bourdieu, P. (1975). La spécificité du champ scientifique et les conditions sociales du progrès de la raison. *Sociologie et sociétés*, 7(1), 91-118.
- Bourdieu, P. (1979). *La distinction : Critique sociale du jugement*. Paris : Les éditions de Minuit.
- Bourdieu, P. (1984). *Homo academicus*. Paris: Les éditions de Minuit.
- Bourdieu, P. (1986). Habitus, code et codification. *Actes de la recherche en sciences sociales*, 64(1), 40-44. <https://doi.org/10.3406/arss.1986.2335>
- Bourdieu, P. (1992). *Réponses*. Paris: Éditions du Seuil.
- Bourdieu, P. (2001). *Science de la science et réflexivité : Cours du Collège de France, 2000-2001*. Paris: Raisons d’agir.
- Bourdieu, P. (2003). *Méditations pascaliennes* (Éd. rev. et corr.). Paris: Éditions du Seuil.

- Boyack, K. W. et Jordan, P. (2011). Metrics associated with NIH funding : A high-level view. *Journal of the American Medical Informatics Association*, 18(4), 423-431.
<https://doi.org/10.1136/amiajnl-2011-000213>
- Bozeman, B. et Boardman, C. (2014). Assessing research collaboration studies : A framework for analysis. Dans *Research Collaboration and Team Science* (p. 1-11). Berlin : Springer.
- Bozeman, B. et Corley, E. (2004). Scientists' collaboration strategies : Implications for scientific and technical human capital. *Research Policy*, 33(4), 599-616.
<https://doi.org/10.1016/j.respol.2004.01.008>
- Bozeman, B. et Youtie, J. (2016). Trouble in paradise : Problems in academic research co-authoring. *Science and Engineering Ethics*, 22(6), 1717-1743.
<https://doi.org/10.1007/s11948-015-9722-5>
- Bridgwater, C. A., Bornstein, P. H. et Walkenbach, J. (1981). Ethical issues and the assignment of publication credit. *American Psychologist*, 36(5), 524-525.
<https://doi.org/10.1037/0003-066X.36.5.524>
- Brown, C. M. et Henderson, D. M. (2015). A new horned dinosaur reveals convergent evolution in cranial ornamentation in ceratopsidae. *Current Biology*, 25(12), 1641-1648.
<https://doi.org/10.1016/j.cub.2015.04.041>
- Budapest Open Access Initiative. (2002). The Budapest Open Access Initiative.
<http://www.budapestopenaccessinitiative.org/read>
- Butler, L. et Ginn, D. (1998). Canadian nurses' views on assignment of publication credit for scholarly and scientific work. *The Canadian Journal of Nursing Research/Revue canadienne de recherche en sciences infirmieres*, 30(1), 171-183.
- Campbell J. L., Quincy, C., Osserman, J. et Pederson, O. K. (2013). Coding in-depth semistructured interviews : Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294-320.

- Caron, E. et van Eck, N. J. (2014). Large-scale author name disambiguation using rule-based scoring and clustering. Dans E. Noyons (Dir.), *19th International Conference on Science and Technology Indicators "Context counts: Pathways to master big data and little data"* (p. 79-86). Leiden University, The Netherlands.
- Castelvecchi, D. (2015). Physics paper sets record with more than 5,000 authors. *Nature News*. <https://doi.org/10.1038/nature.2015.17567>
- Chang, Y.-W. (2019). Definition of authorship in social science journals. *Scientometrics*, 118, 563-585. <https://doi.org/10.1007/s11192-018-2986-1>
- Charles, M. (2003). Deciphering sex segregation : Vertical and horizontal inequalities in ten national labor markets. *Acta Sociologica*, 46(4), 267-287. <https://doi.org/10.1177/0001699303464001>
- Chartier, R. (2003). Foucault's chiasmus : Authorship between science and literature in the seventeenth and eighteenth centuries. Dans M. Biagioli et P. Galison (Dir.), *Scientific authorship : Credit and intellectual property in science* (p. 13-31). New York: Routledge.
- Cheng, S. W. (2012). A contrastive study of master thesis acknowledgements by Taiwanese and North American students. *Open Journal of Modern Linguistics*, 02(01), 8-17. <https://doi.org/10.4236/ojml.2012.21002>
- Cheryan, S., Master, A. et Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers : Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00049>
- Chevallier, S. et Chauviré, C. (2010). *Dictionnaire Bourdieu*. Paris: Ellipses.
- Clarivate Analytics. (2019). LibGuides : Web of Science Core Collection: Who's funding MY research? <https://clarivate.libguides.com/woscc/funding>
- Clausen, S.-E. (1998). *Applied correspondence analysis : An introduction*. Thousand Oaks, CA: Sage Publications.

- Coates, C. (1999). Interpreting academic acknowledgements in English studies : Professors, their partners, and peers. *English Studies in Canada*, 25(3-4), 253-276.
- Cole, J. R. et Cole, S. (1973). *Social stratification in science*. Chicago: University of Chicago Press.
- Costas, R. et van Leeuwen, T. N. (2012). Approaching the “reward triangle” : General analysis of the presence of funding acknowledgments and “peer interactive communication” in scientific publications. *Journal of the American Society for Information Science and Technology*, 63(8), 1647-1661.
- Costas, R. et Yegros-Yegros, A. (2013). Possibilities of funding acknowledgement analysis for the bibliometric study of research funding organizations : Case study of the Austrian Science Fund (FWF). Dans J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger et H. Moed (Dir.), *Proceedings of the 14th International Conference of the International Society for Scientometrics and Informetrics, II* (p. 1401-1408). Vienne, ISSI.
- Crane, D. (1972). *Invisible college: diffusion of knowledge in scientific communities*. Chicago, IL: University of Chicago Press.
- Crawford, E.T. et Biderman, A. D. (1970). Paper money : Trends of research sponsorship in American sociology journals. *Social Science Information*, 9(1), 50-77. <https://doi.org/10.1177/053901847000900103>
- Creswell, J. W. (2009). *Research design : Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Cronin, B. (1991). Let the credits roll : A preliminary examination of the role played by mentors and trusted assessors in disciplinary formation. *Journal of Documentation*, 47(3), 227-239. <https://doi.org/10.1108/eb026878>
- Cronin, B. (1995). *The scholar's courtesy: The role of acknowledgement in the primary communication process*. Taylor Graham.

- Cronin, B. (2001). Hyperauthorship : A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, 52(7), 558-569. <https://doi.org/10.1002/asi.1097>
- Cronin, B. (2004). Bowling alone together : Academic writing as distributed cognition. *Journal of the American Society for Information Science and Technology*, 55(6), 557-560. <https://doi.org/10.1002/asi.10406>
- Cronin, B. (2005). *The hand of science : Academic writing and its rewards*. Lanham, Maryland: Scarecrow Press.
- Cronin, B. (2014). Foreword : The penumbral world of the paratext. Dans N. Desrochers et D. Apollon (Dir.), *Examining paratextual theory and its applications in digital culture*. Hershey, Penn.: IGI Global.
- Cronin, B. et Weaver-Wozniak, S. (1993). Online access to acknowledgements. Dans M. E. Williams (Dir.), *Proceedings of the fourteenth National Online Meeting* (p. 93-98). Medford: NY, Learned Information, Inc.
- Cronin, B., Martinson, A. et Davenport, E. (1997). Women's studies : Bibliometric and content analysis of the formative years. *Journal of Documentation*, 53(2), 123-138. <https://doi.org/10.1108/EUM0000000007196>
- Cronin, B., McKenzie, G. et Rubio, L. (1993). The norms of acknowledgement in four social sciences disciplines. *Journal of Documentation*, 49(1), 29-43. <https://doi.org/10.1108/eb026909>
- Cronin, B., McKenzie, G., Rubio, L. et Weaver-Wozniak, S. (1993). Accounting for influence : Acknowledgments in contemporary sociology. *Journal of the American Society for Information Science*, 44(7), 406-412.
- Cronin, B., McKenzie, G. et Stiffler, M. (1992). Patterns of acknowledgement. *Journal of Documentation*, 48(2), 107-122.
- Cronin, B. et Overfelt, K. (1994). The scholar's courtesy : A survey of acknowledgement behaviour. *Journal of Documentation*, 50(3), 165-196.

- Cronin, B., Shaw, D. et La Barre, K. (2003). A cast of thousands : Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, 54(9), 855-871.
- Cronin, B., Shaw, D. et La Barre, K. (2004). Visible, less visible, and invisible work : Patterns of collaboration in 20th century chemistry. *Journal of the American Society for Information Science and Technology*, 55(2), 160-168. <https://doi.org/10.1002/asi.10353>
- Cronin, B. et Weaver, S. (1995). The praxis of acknowledgement : From bibliometrics to inflometrics. *Revista española de documentación científica*, 18(2), 172-177.
- Dawson, G., Lucocq, B., Cottrell, R. et Lewinson, G. (1998). *Mapping the landscape : National biomedical research outputs 1988-95* (N° Policy Report number 9). London, England: The Wellcome Trust.
- Desrochers, N., Paul-Hus, A., Haustein, S., Costas, R., Mongeon, P., Quan-Haase, A., Bowman, T.D., Pecoskie, J., Tsou, A. et Larivière, V. (2018). Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social Science Information*, 57(2), 223–248. <https://doi.org/10.1177/0539018417752089>
- Desrochers, N., Paul-Hus, A. et Larivière, V. (2016). The angle sum theory : Exploring the literature on acknowledgments in scholarly communication. Dans C. R. Sugimoto (Dir.), *Theories of Informetrics and Scholarly Communication* (p. 225-247). Berlin: De Gruyter Mouton.
- Desrochers, N., Paul-Hus, A. et Pecoskie, J. (2017). Five decades of gratitude : A meta-synthesis of acknowledgments research. *Journal of the Association for Information Science and Technology*, 68(12), 2821-2833. <https://doi.org/10.1002/asi.23903>
- Díaz-Faes, A. A. et Bordons, M. (2014). Acknowledgments in scientific publications : Presence in Spanish science and text patterns across disciplines. *Journal of the Association for Information Science and Technology*, 65(9), 1834-1849. <https://doi.org/10.1002/asi.23081>

- Dung, S. K., López, A., Barragan, E. L., Reyes, R.-J., Thu, R., Castellanos, E., Catalan, F., Huerta-Sanchez, E. et Rohlf, R. V. (2019). Illuminating women's hidden contribution to historical theoretical population genetics. *Genetics*, 211(2), 363-366. <https://doi.org/10.1534/genetics.118.301277>
- Edge, D. (1979). Quantitative measures of communication in science : A critical review. *History of Science*, 17(2), 102-134.
- Elsevier. (2017). *Scopus content coverage guide*. https://www.elsevier.com/_data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf
- Escofier, B. et Pagès, J. (2008). *Analyses factorielles simples et multiples : Objectifs, méthodes et interprétation*. Paris: Dunod.
- Etzkowitz, H., Kemelgor, C. et Uzzi, B. (2000). *Athena unbound : The advancement of women in science and technology*. Cambridge: Cambridge University Press.
- Finkel, J.R., Grenager T. et Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. Dans K. Knight, H.T. Ng and K. Oflazer (Dir.), *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pp. 363-370. Ann Arbor, MI: Association for Computational Linguistics.
- Fisher, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics*, 10(1), 422-429. <https://doi.org/10.1111/j.1469-1809.1940.tb02264.x>
- Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D. et Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *JAMA*, 280(3), 222-224. <https://doi.org/10.1001/jama.280.3.222>
- Fortin, F. et Gagnon, J. (2016). *Fondements et étapes du processus de recherche : Méthodes quantitatives et qualitatives* (3e éd.). Chenelière éducation.

- Forzetting, S. A. (2010). *Personal, peer, patron: Scholarly interactions with librarians and archivists in the acknowledgements of women's history books*. (Mémoire de maîtrise). University of North Carolina at Chapel Hill.
- Franzoni, C. et Sauermann, H. (2014). Crowd science : The organization of scientific research in open collaborative projects. *Research Policy*, 43(1), 1-20. <https://doi.org/10.1016/j.respol.2013.07.005>
- Galison, P. (2003). The collective author. Dans P. Galison et M. Biagioli (Dir.), *Scientific authorship : Credit and intellectual property in science* (p. 325-355). New York, NY: Routledge.
- Gastón, M. S., Schiöth, H. B., De Barioglio, S. R. et Salvatierra, N. A. (2015). Gabaergic control of anxiety-like behavior, but not food intake, induced by ghrelin in the intermediate medial mesopallium of the neonatal chick. *Hormones and Behavior*, 67, 66-72. <https://doi.org/10.1016/j.yhbeh.2014.11.015>
- Gausia, K., Thompson, S. C., Lindeman, M. A., Brown, L. J. et Perkins, D. (2015). Contribution of university departments of rural health to rural health research : An analysis of outputs. *Australian Journal of Rural Health*, 23(2), 101-106. <https://doi.org/10.1111/ajr.12142>
- Genette, G. (1997). *Paratexts : Thresholds of interpretation* (J. E. Lewin, Trad.). <http://ebooks.cambridge.org/ebook.jsf?bid=CBO9780511549373>
- Gesuato, S. (2004). Acknowledgments in PhD dissertations : The complexity of thanking. Dans C. Taylor Torsello, M. Grazia Bùsa et S. Gesuato (Dir.), *Lingua inglese e mediazione linguistica. Ricerca e didattica con supporto telematico* (p. 273-318). Padova: Unipress.
- Ghiasi, G., Larivière, V. et Sugimoto, C. R. (2015). On the compliance of women engineers with a gendered scientific system. *PLOS ONE*, 10(12), e0145931. <https://doi.org/10.1371/journal.pone.0145931>
- Giannoni, D. S. (1998). The genre of journal acknowledgments : Findings of a cross-disciplinary investigation. *Linguistica e Filologia*, 6, 61-84.

- Giannoni, D. S. (2002). Worlds of gratitude : A contrastive study of acknowledgement texts in English and Italian research articles. *Applied Linguistics*, 23(1), 1-31. <https://doi.org/10.1093/applin/23.1.1>
- Giannoni, D. S. (2006a). Book acknowledgements across disciplines and texts. Dans K. Hyland et M. Bondi (Dir.), *Academic Discourse Across Disciplines* (p. 151-176). New York, NY: Peter Lang.
- Giannoni, D. S. (2006b). Evidence of generic tension in academic book acknowledgements. Dans V. K. Bhatia et M. Gotti (Dir.), *Explorations in specialized genres* (p. 21-42). Bern: Peter Lang.
- Giles, C. L. et Councill, I. G. (2004). Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17599–17604. <https://doi.org/10.1073/pnas.0407743101>
- Ginn, D. et Butler, L. (1998). Moving toward policy development on assigning publication credit for oncology nurses' contributions to scholarly and scientific work. *Canadian Oncology Nursing Journal / Revue canadienne de soins infirmiers en oncologie*, 8(2), 108-113.
- Glass, R. M. (1992). New information for authors and readers : Group authorship, acknowledgements, and rejected manuscripts. *JAMA*, 268(1), 99.
- Gök, A., Rigby, J. et Shapira, P. (2016). The impact of research funding on scientific outputs : Evidence from six smaller European countries. *Journal of the Association for Information Science and Technology*, 67(3), 715-730. <https://doi.org/10.1002/asi.23406>
- Graf, C., Wager, E., Bowman, A., Fiack, S., Scott-Lichter, D. et Robinson, A. (2007). Best practice guidelines on publication ethics : A publisher's perspective. *International Journal of Clinical Practice*, 61(s152), 1-26. <https://doi.org/10.1111/j.1742-1241.2006.01230.x>

- Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). London: Chapman & Hall/CRC.
- Greener, I. (2007). The politics of gender in the NHS: Impression management and getting things done. *Gender, Work & Organization*, 14(3), 281-299.
<https://doi.org/10.1111/j.1468-0432.2007.00343.x>
- Guttman, L. (1941). The quantification of a class of attributes : A theory and method of scale construction. Dans P. Horst (Dir.), *The prediction of personal adjustment* (p. 251-364). New York, NY: Social Science Research Council.
- Haeussler, C. et Sauermann, H. (2013). Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship. *Research Policy*, 42(3), 688-703. <https://doi.org/10.1016/j.respol.2012.09.009>
- Harter, S. P. et Hooten, P. A. (1992). Information science and scientists : JASIS, 1972–1990. *Journal of the American Society for Information Science*, 43(9), 583-593.
[https://doi.org/10.1002/\(SICI\)1097-4571\(199210\)43:9<583::AID-ASII>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<583::AID-ASII>3.0.CO;2-O)
- Heffner, A. G. (1979). Authorship recognition of subordinates in collaborative research. *Social Studies of Science*, 9(3), 377-384. <https://doi.org/10.1177/030631277900900305>
- Heffner, A. G. (1981). Funded research, multiple authorship, and subauthorship collaboration in four disciplines. *Scientometrics*, 3(1), 5-12. <https://doi.org/10.1007/BF02021860>
- Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S. et Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction*, 39, 11-22.
<https://doi.org/10.1016/j.learninstruc.2015.05.002>
- Henderson, R., Jaffe, A. B. et Trajtenberg, M. (1998). University patenting amid changing incentives for commercialization. Dans P. G. B. Navaretti, P. P. Dasgupta, P. K.-G. Mäler, et P. D. Siniscalco (Dir.), *Creation and transfer of knowledge* (p. 87-114).
https://doi.org/10.1007/978-3-662-03738-6_6

- Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455-476. <https://doi.org/10.1007/s11192-016-1849-x>
- Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4), 520-524. <https://doi.org/10.1017/S0305004100013517>
- Hoder-Salmon, M. (1978). Collecting scholar's wives. *Feminist Studies*, 4(3), 107-114. <https://doi.org/10.2307/3177543>
- Hollander, P. (2001). Acknowledgments : An academic ritual. *Academic Questions*, 15(1), 63-76. <https://doi.org/10.1007/s12129-001-1056-x>
- Hruschka, D. J., Schwartz, D., St.John, D. C., Picone-Decaro, E., Jenkins, R. A. et Carey, J. W. (2004). Reliability in coding open-ended data : Lessons learned from HIV behavioral research. *Field Methods*, 16(3), 307-331. <https://doi.org/10.1177/1525822X04266540>
- Hyland, K. (2003). Dissertation acknowledgements : The anatomy of a Cinderella genre. *Written Communication*, 20(3), 242-268. <https://doi.org/10.1177/0741088303257276>
- Hyland, K. (2004). Graduates' gratitude : The generic structure of dissertation acknowledgements. *English for Specific Purposes*, 23(3), 303-324. [https://doi.org/10.1016/S0889-4906\(03\)00051-6](https://doi.org/10.1016/S0889-4906(03)00051-6)
- Hyland, K. et Tse, P. (2004). "I would like to thank my supervisor". Acknowledgements in graduate dissertations. *International Journal of Applied Linguistics*, 14(2), 259-275. <https://doi.org/10.1111/j.1473-4192.2004.00062.x>
- Ibarra, H. (1992). Homophily and differential returns : Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly*, 37(3), 422-447. <https://doi.org/10.2307/2393451>
- Ilakovac, V., Fister, K., Marusic, M. et Marusic, A. (2007). Reliability of disclosure forms of authors' contributions. *CMAJ*, 176(1), 41-46. <https://doi.org/10.1503/cmaj.060687>
- Innes, J. (2006). *Scholarly communication and knowledge management in American zoos* (Thèse de Doctorat). Nova Southeastern University.

- International Committee of Medical Journal Editors. (1997). Uniform requirements for manuscripts submitted to biomedical journals. *The New England Journal of Medicine*, 336(4), 309-315. <https://doi.org/10.1056/NEJM199701233360422>
- International Committee of Medical Journal Editors. (2015, décembre). Recommendations. http://www.icmje.org/recommendations/archives/2015_dec_urm.pdf
- International Committee of Medical Journal Editors. (2018, décembre). Recommendations. <http://www.icmje.org/icmje-recommendations.pdf>
- Ivaniš, A., Hren, D., Marušić, M. et Marušić, A. (2011). Less work, less respect : Authors' perceived importance of research contributions and their declared contributions to research articles. *PLOS ONE*, 6(6), e20206. <https://doi.org/10.1371/journal.pone.0020206>
- Ivaniš, A., Hren, D., Sambunjak, D., Marušić, M. et Marušić, A. (2008). Quantification of authors' contributions and eligibility for authorship : Randomized study in a general medical journal. *Journal of General Internal Medicine*, 23(9), 1303-1310. <https://doi.org/10.1007/s11606-008-0599-8>
- Jabbehdari, S. et Walsh, J. P. (2017). Authorship norms and project structures in science. *Science, Technology, & Human Values*. <https://doi.org/10.1177/0162243917697192>
- Jeschin, D., Lewison, G. et Anderson, J. (1995). A bibliometric database for tracking acknowledgements of research funding. Dans M.E.D. Koenig & A. Bookstein (Dir.), *Proceedings of the Fifth International Conference of the International Society for Scientometrics and Informetrics* (p. 235–244). Medford, NJ: Learned Information.
- Jurafsky, D. et Martin, J. H. (2009). *Speech and language processing* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Kaatz, A., Lee, Y. G., Potvien, A., Magua, W., Filut, A., Bhattacharya, A., Leatherberry, R., Zhu, X. et Carnes, M. (2016). Analysis of National Institutes of Health R01 application critiques, impact, and criteria scores : Does the sex of the principal investigator make a difference? *Academic Medicine*, 91(8), 1080-1088.

<https://doi.org/10.1097/ACM.0000000000001272>

Katz, J. S. et Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1-18.

[https://doi.org/10.1016/S0048-7333\(96\)00917-1](https://doi.org/10.1016/S0048-7333(96)00917-1)

Khabsa, M., Koppman, S. et Giles, C.L. (2012). Towards building and analyzing a social network of acknowledgments in scientific and academic documents. Dans S.J. Yang, A.M. Greenberg & M. Endsley (Dir.), *Social Computing, Behavioral - Cultural Modeling and Prediction* (p. 357-364). Berlin: Springer.

https://link.springer.com/chapter/10.1007/978-3-642-29047-3_43

Khabsa, M., Treeratpituk, P. et Giles, C.L. (2012). AckSeer: A repository and search engine for automatically extracted acknowledgments from digital libraries. Dans K. B. Boughida & B. Howard (Dir.), *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (p. 185-194). New York, NY, ACM.

<http://doi.acm.org/10.1145/2232817.2232852>

Kim, S.N., Baldwin, T. et Kan, M.Y. (2010). Evaluation N-gram based evaluation metrics for automatic keyphrase extraction. Dans C. Huang & D. Jurafsky (Dir.), *Proceedings of the 23rd International Conference on Computational Linguistics* (p. 572-580). Beijing: Tsinghua University Press.

Knorr-Cetina, K. D. (1991). Epistemic cultures : Forms of reason in science. *History of Political Economy*, 23(1), 105-122.

Knorr-Cetina, K. D. (2009). *Epistemic cultures : How the sciences make knowledge*. Cambridge, Mass: Harvard University Press.

Koley, S. et Sen, B. K. (2013). Acknowledgements in research papers in electronics and related fields : 2008-2012. *SRELS Journal of Information Management*, 50(5), 619-627.

Krippendorff, K. (2004). *Content analysis : An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

- Kuhn, T. S. (2012). *The structure of scientific revolutions* (4th edition). Chicago: University of Chicago Press.
- Laband, D. N. et Tollison, R. D. (2000). Intellectual collaboration. *Journal of Political Economy*, 108(3), 632-662. <https://doi.org/10.1086/262132>
- Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A. et Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417-435. <https://doi.org/10.1177/0306312716650046>
- Larivière, V., Gingras, Y. et Archambault, É. (2006). Canadian collaboration networks : A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*, 68(3), 519-533. <https://doi.org/10.1007/s11192-006-0127-8>
- Larivière, V., Haustein, S. et Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLOS ONE*, 10(6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>
- Larivière, V., Ni, C. C., Gingras, Y., Cronin, B. et Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, 504(7479), 211-213.
- Larivière, V. et Sugimoto, C. R. (2018). *Mesurer la science*. Montréal : Presses de l'Université de Montréal. <https://pum.umontreal.ca/catalogue/mesurer-la-science>
- Lasaky, F. G. (2011). A contrastive study of generic organization of doctoral dissertation acknowledgements written by native and non-native (Iranian) students in applied linguistics. *The Modern Journal of Applied Linguistics*, 3(2), 175-199.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1), 3-15. <https://doi.org/10.3152/147154402781776961>
- Lebart, L. et Salem, A. (1994). *Statistique textuelle*. Malakoff : Dunod.
- Lewis-Beck, J. A. (1980). The participation of men and women in educational research : Another look. *Sex Roles*, 6(4), 607-610.
- Lewisson, G. (1998). Gastroenterology research in the United Kingdom : Funding sources and impact. *Gut*, 43(2), 288-293. <https://doi.org/10.1136/gut.43.2.288>

- Lewison, G. et Devey, M. E. (1999). Bibliometric methods for the evaluation of arthritis research. *Rheumatology*, 38(1), 13-20. <https://doi.org/10.1093/rheumatology/38.1.13>
- Lewison, G., Grant, J. et Jansen, P. (2001). International gastroenterology research : Subject areas, impact, and funding. *Gut*, 49(2), 295-302. <https://doi.org/10.1136/gut.49.2.295>
- Lewison, G. (2009). Financial acknowledgements on the Web of Science : A new resource for bibliometric analysis. Dans B. Larsen et J. Leta (Dir.), *Proceedings of ISSI 2009—12th International Conference of the International Society for Scientometrics and Informetrics* (p. 968-969). Rio de Janeiro: BIREME/PAHO/WHO et Federal University of Rio de Janeiro.
- Lewison, G. et Markusova, V. (2010). The evaluation of Russian cancer research. *Research Evaluation*, 19(2), 129-144. <https://doi.org/10.3152/095820210X510098>
- Ley, T. J. et Hamilton, B. H. (2008). The gender gap in NIH grant applications. *Science*, 322(5907), 1472-1474. <https://doi.org/10.1126/science.1165878>
- Lincoln, A. E., Pincus, S., Koster, J. B. et Leboy, P. S. (2012). The Matilda Effect in science : Awards and prizes in the US, 1990s and 2000s. *Social Studies of Science*, 42(2), 307-320. <https://doi.org/10.1177/0306312711435830>
- Lissoni, F., Montobbio, F. et Zirulia, L. (2013). Inventorship and authorship as attribution rights : An enquiry into the economics of scientific credit. *Journal of Economic Behavior & Organization*, 95, 49-69. <https://doi.org/10.1016/j.jebo.2013.08.016>
- Liu, W., Hu, G., Tang, L. et Wang, Y. (2015). China's global growth in social science research : Uncovering evidence from bibliometric analyses of SSCI publications (1978–2013). *Journal of Informetrics*, 9(3), 555-569. <https://doi.org/10.1016/j.joi.2015.05.007>
- Loan, N. T. T. (2017). Generic structures and linguistic features of TESOL master's thesis acknowledgements written by Vietnamese postgraduates. *3L: Language, Linguistics, Literature*, 23(2), 27-40. <https://doi.org/10.17576/3L-2017-2302-03>
- Lukas, J., Pockrandt, A. M., Seennann, S., Sharif, M., Runge, F., Pohlrs, S., Zheng, C, Gläser, A., Beller, M., Rolfs, A. et Giese, A.K. (2015). Enzyme enhancers for the treatment of

- Fabry and Pompe Disease. *Molecular Therapy*, 23(3), 456-64.
<https://doi.org/10.1038/mt.2014.224>
- Lundberg, J., Tomson, G., Lundkvist, I., Skar, J. et Brommels, M. (2006). Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics*, 69(3), 575-589.
<https://doi.org/10.1007/s11192-006-0170-5>
- Lyubarova, R., Itagaki, B. K. et Itagaki, M. W. (2009). The impact of National Institutes of Health funding on U.S. cardiovascular disease research. *PLoS ONE*, 4(7), e6425.
<https://doi.org/10.1371/journal.pone.0006425>
- Macaluso, B., Larivière, V., Sugimoto, T. et Sugimoto, C. (2016). Is science built on the shoulders of women? A study of gender differences in contributorship. *Academic Medicine*, 91(8), 1136-1142. <https://doi.org/10.1097/ACM.0000000000001261>
- Mackintosh, K. H. (1972). *Acknowledgment patterns in sociology* (Thèse de doctorat). University of Oregon.
- MacRoberts, M. H. et MacRoberts, B. R. (2018). The mismeasure of science : Citation analysis. *Journal of the Association for Information Science and Technology*, 69(3), 474-482.
<https://doi.org/10.1002/asi.23970>
- Marcus, M.P., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K. et Schasberger, B. (1994). The Penn Treebank: Annotating predicate structure. Dans *Proceedings of the workshop on Human Language Technology* (p. 114-119). Plainsboro NJ; Association for Computational Linguistics.
- Marcus, M.P., Marcinkiewicz, M.A. et Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313-330.
- Markusova, V. A., Libkind, A. N. et Aversa, E. (2012). Impact of competitive funding on research output in Russia. *Collnet Journal of Scientometrics and Information Management*, 6(1), 61-69. <https://doi.org/10.1080/09737766.2012.10700924>

- Martin, B. R. (2016). Editors' JIF-boosting stratagems – Which are appropriate and which not? *Research Policy*, 45(1), 1-7. <https://doi.org/10.1016/j.respol.2015.09.001>
- Marušić, A., Bates, T., Anić, A. et Marušić, M. (2006). How the structure of contribution disclosure statements affects validity of authorship : A randomized study in a general medical journal. *Current Medical Research and Opinion*, 22(6), 1035-1044. <https://doi.org/10.1185/030079906X104885>
- Marušić, A., Bošnjak, L. et Jerončić, A. (2011). A Systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS ONE*, 6(9), e23477. <https://doi.org/10.1371/journal.pone.0023477>
- Mattsson, P., Sundberg, C. J. et Laget, P. (2011). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics*, 87(1), 99-105. <https://doi.org/10.1007/s11192-010-0310-9>
- McCain, K. W. (1991). Communication, competition, and secrecy: The production and dissemination of research-related information in genetics. *Science, Technology & Human Values*, 16(4), 491-516. <https://doi.org/10.1177/016224399101600404>
- McCain, K. W. (2018). Beyond Garfield's Citation Index : An assessment of some issues in building a personal name Acknowledgments Index. *Scientometrics*, 114(2), 605-631. <https://doi.org/10.1007/s11192-017-2598-1>
- McPherson, M., Smith-Lovin, L. et Cook, J. M. (2001). Birds of a Feather : Homophily in Social Networks. *Annual Review of Sociology*, 27(1), 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology*, (1-2), 115-126.
- Merton, R. K. (1957). Priorities in scientific discovery : A chapter in the sociology of science. *American Sociological Review*, 22(6), 635-659. <https://doi.org/10.2307/2089193>
- Merton, R. K. (1968). The Matthew Effect in science. *Science*, 159(3810), 56-63.

- Merton, R. K. (1973). *The sociology of science : Theoretical and empirical investigations*. Chicago et London: Chicago Univeristy Press.
- Miles, M. B. et Huberman, A. M. (1984). Drawing valid meaning from qualitative data : Toward a shared craft. *Educational Researcher*, 13(5), 20-30. <https://doi.org/10.3102/0013189X013005020>
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, Springer.
- Mongeon, P., Smith, E., Joyal, B. et Larivière, V. (2017). The rise of the middle author : Investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. *PLOS ONE*, 12(9), e0184601. <https://doi.org/10.1371/journal.pone.0184601>
- Moore, M. (1984). Sex and acknowledgements : A nonreactive study. *Sex Roles*, 10(11-12), 1021-1031.
- Mowery, D. C., Nelson, R. R., Sampat, B. N. et Ziedonis, A. A. (2001). The growth of patenting and licensing by U.S. universities : An assessment of the effects of the Bayh–Dole act of 1980. *Research Policy*, 30(1), 99-119. [https://doi.org/10.1016/S0048-7333\(99\)00100-6](https://doi.org/10.1016/S0048-7333(99)00100-6)
- Murray, S. O. (1982). The dissolution of “classical ethnoscience”. *Journal of the History of the Behavioral Sciences*, 18(2), 163-175. [https://doi.org/10.1002/1520-6696\(198204\)18:2<163::AID-JHBS2300180206>3.0.CO;2-E](https://doi.org/10.1002/1520-6696(198204)18:2<163::AID-JHBS2300180206>3.0.CO;2-E)
- Mwendera, C. A., de Jager, C., Longwe, H., Hongoro, C., Mutero, C. M. et Phiri, K. S. (2017). Malaria research in Malawi from 1984 to 2016 : A literature review and bibliometric analysis. *Malaria Journal*, 16, 246. <https://doi.org/10.1186/s12936-017-1895-8>
- National Science Foundation. (2006). Science and engineering indicators. Chapter 5 : Academic research and development. Data and terminology. <http://www.nsf.gov/statistics/seind06/c5/c5s3.htm#sb1>
- Novick, D., Hong, J., Montgomery, W., Duenas, H., Gado, M. et Haro, J. M. (2015). Predictors of remission in the treatment of major depressive disorder : Real-world evidence from a

- 6-month prospective observational study. *Neuropsychiatric Disease and Treatment*, 11, 197-205. <https://doi.org/10.2147/NDT.S75498>
- Ossenblok, T. L. B., Verleynsen, F. T. et Engels, T. C. E. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000–2010). *Journal of the Association for Information Science and Technology*, 65(5), 882-897. <https://doi.org/10.1002/asi.23015>
- Parker, E. B., Paisley, W. J. et Garrett, R. L. (1967). Bibliographic citations as unobtrusive measures of scientific communication. Dans *National Science Foundation Grant Gn-434*. Stanford University. Institute for Communication Research.
- Pasiarski, M., Grywalska, E., Kosmaczewska, A., Gozdz, S., Steckiewicz, P., Garus, B., Bilski, M., Hymos, A. et Rolinski, J. (2015). Assessment of peripheral blood and bone marrow T, NK, NKT and dendritic cells in patients with multiple myeloma. *Postepy Higieny I Medycyny Doswiadczonej*, 69, 1435-42.
- Patel, N. (1973). Collaboration in the professional growth of American sociology. *Social Science Information*, 12(6), 77-92.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3e ed.). Thousand Oaks, CA: Sage Publications.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods : Integrating theory and practice* (4e éd.). Thousand Oaks, CA: SAGE Publications.
- Paul-Hus, A., Desrochers, N. et Costas, R. (2016). Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science. *Scientometrics*, 108(1), 167-182. <https://doi.org/10.1007/s11192-016-1953-y>
- Paul-Hus, A., Díaz-Faes, A.A., Desrochers, N., Costas, R., Sainte-Marie, M., Macaluso, B. & Larivière, V. (2016). Beyond funding: What can acknowledgements reveal about credit distribution in science? In I. Rafols, J. Molas-Gallart, E. Castro-Martinez & R. Woolley (eds), *Proceedings of the 21st International Conference on Science and Technology Indicators*. València: Universitat Politècnica de València.

- Paul-Hus, A., Mongeon, P., Sainte-Marie, M., et Larivière, V. (2017a). The sum of it all : Revealing collaboration patterns by combining authorship and acknowledgements. *Journal of Informetrics*, 11(1), 80-87. <https://doi.org/10.1016/j.joi.2016.11.005>
- Paul-Hus, A., Díaz-Faes, A. A., Sainte-Marie, M., Desrochers, N., Costas, R. et Larivière, V. (2017). Beyond funding : Acknowledgement patterns in biomedical, natural and social sciences. *PLOS ONE*, 12(10), e0185578. <https://doi.org/10.1371/journal.pone.0185578>
- Paul-Hus, A. et Desrochers, N. (2019). Acknowledgements are not just thank you notes : A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015. *PLOS ONE*, 14(12), e0226727. <https://doi.org/10.1371/journal.pone.0226727>
- Paul-Hus, A., Mongeon, P., Sainte-Marie, M. et Larivière, V. (2020). Who are the acknowledgees ? An analysis of gender and academic status. *Quantitative Science Studies*. https://doi.org/10.1162/qss_a_00036
- Perneger, T. V., Poncet, A., Carpentier, M., Agoritsas, T., Combescure, C. et Gayet-Ageron, A. (2017). Thinker, soldier, scribe : Cross-sectional study of researchers' roles and author order in the Annals of Internal Medicine. *BMJ Open*, 7(6), e013898. <https://doi.org/10.1136/bmjopen-2016-013898>
- Pickard, A. J. (2007). *Research methods in information*. London: Facet.
- Pluye, P., Nadeau, L., Gagnon, M., Grad, R., Johnson-Lafleur, J. et Griffiths, F. (2012). Les méthodes mixtes. Dans *Approches et pratiques en évaluation de programme* (2^e éd.). Montréal: Presses de l'Université de Montréal.
- Ponomariov, B. et Boardman, C. (2016). What is co-authorship? *Scientometrics*, 109(3), 1939-1963. <https://doi.org/10.1007/s11192-016-2127-7>
- Pontille, D. (2004). *La signature scientifique : Une sociologie pragmatique de l'attribution*. Paris: CNRS.
- Pouget, J.G., Goncalves, V.F., Nurmi, E.L., Laughlin, C.P., Mallya, K.S., McCracken, J.T., Aman, M. G., McDougale, C.J., Scahill, L., Misener, V.L., Tiwari, A.K., Brandl, E.J.,

- Felsky, D., Leung, A.Q., Lieberman, J.A., Meltzer, H.Y., Potkin, S.G., Niending, C., Steimer, W.... Kennedy, J.L. (2015). Investigation of TSPO variants in schizophrenia and antipsychotic treatment outcomes. *Pharmacogenomics*, 16(1), 5-22. <https://doi.org/10.2217/pgs.14.158>
- Price, D. J. de S. (1963). *Little science, big science*. New York: Columbia University Press.
- Price, D. J. de S. (1986). *Little science, big science and beyond*. New York: Columbia University Press.
- Price, D. J. de S. et Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist*, 21(11), 1011-1018.
- Radicchi, F. et Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, 97(3), 627-637. <https://doi.org/10.1007/s11192-013-1027-3>
- Ratinov, L. et Roth, D. (2009). Design challenges and misconceptions in named entity recognition. Dans S. Stevenson et X. Carreras (Dir.), *Proceedings of the Thirteenth Conference on Computational Natural Language Learning* (p. 147–155). Boulder, CO: Association for Computational Linguistics. <https://www.aclweb.org/anthology/W09-1119>
- Rattan, G. K. M. (2013). Acknowledgement patterns in Annals of Library and Information Studies 1999-2012. *Library Philosophy and Practice, e-journal* (Paper 989). <http://digitalcommons.unl.edu/libphilprac/989/>
- Rennie, D., Yank, V. et Emanuel, L. (1997). When authorship fails : A proposal to make contributors accountable. *JAMA*, 278(7), 579-585. <https://doi.org/10.1001/jama.1997.03550070071041>
- Resnik, D. B. (1997). A proposal for a new system of credit allocation in science. *Science and Engineering Ethics*, 3(3), 237-243. <https://doi.org/10.1007/s11948-997-0023-5>

- Rigby, J. (2011). Systematic grant and funding body acknowledgment data for publications : New dimensions and new controversies for research policy and evaluation. *Research Evaluation*, 20(5), 355-365. <https://doi.org/10.3152/095820211X13164389670392>
- Rigby, J. (2013). Looking for the impact of peer review: Does count of funding acknowledgments really predict research impact? *Scientometrics*, 94, 57-73. <https://doi.org/10.1007/s11192-012-0779-5>
- Rong, X. L., Grant, L. et Ward, K. B. (1989). Productivity of women scholars and gender researchers : Is funding a factor? *The American Sociologist*, 20(1), 95-100.
- Rossiter, M. W. (1993). The Matthew Matilda Effect in science. *Social Studies of Science*, 23(2), 325-341.
- Saint-Martin, A. (2013). *La sociologie de Robert K. Merton*. Paris : La Découverte.
- Salager-Meyer, F., Alcaraz-Ariza, M. Á. et Berbesí, M. P. (2009). “Backstage solidarity” in Spanish- and English-written medical research papers : Publication context and the acknowledgment paratext. *Journal of the American Society for Information Science and Technology*, 60(2), 307-317. <https://doi.org/10.1002/asi.20981>
- Salager-Meyer, F., Alcaraz-Ariza, M. Á., Berbesí, M. P. et Zambrano, N. (2006). Paying one’s intellectual debt : Acknowledgments in conventional vs. Complementary/alternative medical research. Dans M. Gotti et F. Salager-Meyer (Dir.), *Advances in Medical Discourse Analysis : Oral and Written Contexts* (p. 407-430). Bern: Peter Lang.
- Salager-Meyer, F., Alcaraz-Ariza, M. Á., Luzardo Briceño, M. et Jabbour, G. (2010). Scholarly gratitude in five geographical contexts : A diachronic and cross-generic approach of the acknowledgment paratext in medical discourse (1950–2010). *Scientometrics*, 86(3), 763-784. <https://doi.org/10.1007/s11192-010-0329-y>
- Santamaría, L. et Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, 4, e156. <https://doi.org/10.7717/peerj-cs.156>

- Sarbova, V., Koschella, A., Cheng, F., Kelly, S. M. et Heinze, T. (2015). Studies on the sulfation of cellulose alpha-lipoate and ability of the sulfated product to stabilize colloidal suspensions of gold nanoparticles. *Carbohydrate Polymers*, 124, 117-123. <https://doi.org/10.1016/j.carbpol.2015.01.080>
- Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: Sage Publications.
- Scopus. (2019). Funding acknowledgement information available for 9 million articles. Elsevier Scopus Blog. <https://blog.scopus.com/posts/funding-acknowledgement-information-available-for-9-million-articles>
- Seeman, J. I. et House, M. C. (2010a). Influences on authorship issues : An evaluation of giving credit. *Accountability in research*, 17(3), 146-169. <https://doi.org/10.1080/08989621.2010.493094>
- Seeman, J. I. et House, M. C. (2010b). Influences on authorship issues : An evaluation of receiving, not receiving, and rejecting credit. *Accountability in research*, 17(4), 176-197. <https://doi.org/10.1080/08989621.2010.493094>
- Seeman, J. I. et House, M. C. (2015). Authorship Issues and Conflict in the U.S. Academic Chemical Community. *Accountability in Research*, 22(6), 346-383. <https://doi.org/10.1080/08989621.2015.1047707>
- Shah, K., Sussex, J., Hernandez-Villafuerte, K., Garau, M., Rotolo, D., Hopkins, M. M., Grassoni, N., Crane, P., Lang, F., Hutton, J., Pateman, C., Mawer, A., Farrell, C. et Sharp, T. (2014). *Exploring the interdependencies of research funders in the UK | OHE*. <https://www.ohe.org/publications/exploring-interdependencies-research-funders-uk>
- Shapin, S. (1989). The invisible technician. *American Scientist*, 77(6), 554-563.
- Shapira, P. et Wang, J. (2010). Follow the money. *Nature*, 468(7324), 627-628. <https://doi.org/10.1038/468627a>
- Shen, H. (2013). Inequality quantified : Mind the gender gap. *Nature News*, 495(7439), 22. <https://doi.org/10.1038/495022a>

- Sismondo, S. (2009). Ghosts in the machine publication planning in the medical sciences. *Social Studies of Science*, 39(2), 171-198. <https://doi.org/10.1177/0306312708101047>
- Smith, E. et Williams-Jones, B. (2012). Authorship and responsibility in health sciences research : A review of procedures for fairly allocating authorship in multi-author studies. *Science and Engineering Ethics*, 18(2), 199-212. <https://doi.org/10.1007/s11948-011-9263-5>
- Smith, E., Williams-Jones, B., Master, Z., Larivière, V., Sugimoto, C. R., Paul-Hus, A., Shi, M. et Resnik, D. B. (2019). Misconduct and misbehavior related to authorship disagreements in collaborative science. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-019-00112-4>
- Spiegel, D. et Keith-Spiegel, P. (1970). Assignment of publication credits : Ethics and practices of psychologists. *American Psychologist*, 25(8), 738-747. <https://doi.org/10.1037/h0029769>
- Stehr, F. et Forkel, M. (2013). Funding resources for rare disease research. *Biochimica et biophysica acta*, 1832(11), 1910-1912. <https://doi.org/10.1016/j.bbadis.2013.04.016>
- Sugimoto, C. R. et Cronin, B. (2012). Biobibliometric profiling: An examination of multifaceted approaches to scholarship. *Journal of the American Society for Information Science and Technology*, 63(3), 450-468. <https://doi.org/10.1002/asi.21695>
- Swales, J. M. et Feak, C. B. (2000). *English in today's research world : A writing guide*. Ann Arbor: University of Michigan Press.
- Tang, L., Hu, G. et Liu, W. (2016). Funding acknowledgment analysis : Queries and caveats. *Journal of the Association for Information Science and Technology*, 68(3), 790-794. <https://doi.org/10.1002/asi.23713>
- Teixeira da Silva, J. A. et Dobránszki, J. (2016). How authorship is defined by multiple publishing organizations and STM publishers. *Accountability in Research*, 23(2), 97-122. <https://doi.org/10.1080/08989621.2015.1047927>

- Templeton Gay, J., Griffith Lavender, M. et McCard, N. (1987). Nurse education views of assignment of authorship credit. *IMAGE: Journal of Nursing Scholarship*, 19(3), 134-137.
- Thomson Reuters. (2015). *Thomson Reuters Bibliographic Policy Funding Acknowledgements*. Document interne.
- Tiew, W. S. (1998). Journal of Natural Rubber Research 1987-1996 : A ten-year bibliometric study. *IASLIC Bulletin*, 43(2), 49-57.
- Tiew, W. S. et Sen, B. K. (2002). Acknowledgement patterns in research articles : A bibliometric study based on Journal of Natural Rubber Research 1986-1997. *Malaysian Journal of Library & Information Science*, 7(1), 43-56.
- Toutatnova, K., Klein, D., Manning, C. et Singer, N. (2003). Feature-rich Part-Of-Speech Tagging with a Cyclic Dependency Network. Dans M. Hearst & M. Ostendorf (Dir.) *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology* (p. 252-259). Stroudsburg, PA: Association for Computational Linguistics.
- Toutanova, K. et Manning, C. (2000). Enriching the Knowledge Sources Used in a Maximum Entropy Part-Of-Speech Tagger. Dans H. Schiitze & K-Y Su (Dir.) *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora* (p. 63-70). Stroudsburg, PA: Association for Computational Linguistics.
- Turner, S. (2007). Merton's 'norms' in political and intellectual context. *Journal of Classical Sociology*, 7(2), 161-178. <https://doi.org/10.1177/1468795X07078034>
- Vicente-Villardón, J. (2015). *MultBiplot : A package for multivariate analysis using Biplots*. <http://biplot.usal.es/ClassicalBiplot/index.html>
- Wager, E. (2009). Recognition, reward and responsibility : Why the authorship of scientific papers matters. *Maturitas*, 62(2), 109-112. <https://doi.org/10.1016/j.maturitas.2008.12.001>

- Waltz, C. F., Nelson, B. et Chambers, S. (1985). Assigning publication credits. *Nursing Outlook*, 33(5), 233-238.
- Wang, J. et Shapira, P. (2011). Funding acknowledgement analysis : An enhanced tool to investigate research sponsorship impacts : The case of nanotechnology. *Scientometrics*, 87(3), 563–586. <https://doi.org/10.1007/s11192-011-0362-5>
- Waquet, F. (2005). Acknowledgments : Instructions for use. *Modern Intellectual History*, 2(03), 361–385. <https://doi.org/10.1017/S147924430500048X>
- Web of Science. (2009). Funding Acknowledgements.
http://wokinfo.com/products_tools/multidisciplinary/webofscience/fundingsearch/
- Weber, N. et Thomer, A. (2014). Paratexts and documentary practices : Text mining authorship and acknowledgment from a bioinformatics corpus. Dans N. Desrochers et D. Apollon (Dir.), *Examining paratextual theory and its applications in digital culture* (p. 84-109). Hershey, Penn.: IGI Global.
- Werley, H. H., Murphy, P. A., Gosch, S. M., Gottesmann, H. et Newcomb, B. J. (1981). Research publication credit assignment : Nurses' views. *Research in Nursing & Health*, 4(2), 261-279. <https://doi.org/10.1002/nur.4770040208>
- Werner, O. (1998). Short take 25 : Referencing native consultants. *CAM: The Cultural Anthropology Methods Journal*, 10(2), 29-30.
<https://doi.org/10.1177/1525822X980100020201>
- West, J. D., Jacquet, J., King, M. M., Correll, S. J. et Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PLoS ONE*, 8(7), e66212.
<https://doi.org/10.1371/journal.pone.0066212>
- Wislar, J. S., Flanagan, A., Fontanarosa, P. B. et DeAngelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals : A cross sectional survey. *BMJ*, 343, d6128. <https://doi.org/10.1136/bmj.d6128>
- Witz, A. (1992). *Professions and patriarchy*. London: Routledge.
<https://doi.org/10.4324/9780203167786>

- Wuchty, S., Jones, B. F. et Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036-1039. <https://doi.org/10.1126/science.1136099>
- Xu, X., Tan, A. M. et Zhao, S. X. (2015). Funding ratios in social science : The perspective of countries/territories level and comparison with natural sciences. *Scientometrics*, 104(3), 673-684. <https://doi.org/10.1007/s11192-015-1633-3>
- Yang, L. Y., Yue, T., Ding, J. L. et Han, T. (2012). A comparison of disciplinary structure in science between the G7 and the BRIC countries by bibliometric methods. *Scientometrics*, 93(2), 497-516. <https://doi.org/10.1007/s11192-012-0695-8>
- Zhang, R., Li, R., Wang, J., Wang, S., Zhang, M., Hu, X., Zhang, L., Wang, S., Wang, R. et Bao, Z. (2015). Identification, characterization and expression profiling of the Tollip gene in Yesso scallop (*Patinopecten yessoensis*). *Genes & Genetic Systems*, 90(2), 99-108. <https://doi.org/10.1266/ggs.90.99>
- Zhao, M. et Jiang, Y. (2010). Dissertation acknowledgement : Generic structure and linguistic features. *Chinese Journal of Applied Linguistics*, 33(1), 94-109.
- Zhou, P. et Tian, H. (2014). Funded collaboration research in mathematics in China. *Scientometrics*, 99(3), 695-715. <https://doi.org/10.1007/s11192-013-1212-4>
- Ziman, J. (2002). *Real science : What it is and what it means*. Cambridge: Cambridge University Press.
- Zuckerman, H. A. (1968). Patterns of name ordering among authors of scientific papers : A study of social symbolism and its ambiguity. *American Journal of Sociology*, 74(3), 276-291.
- Zuckerman, H. A. (1977). *Scientific elite : Nobel laureates in the United States*. Piscataway, NJ: Transaction Publishers.

Annexe 1. Données de l'article 1

The sum of it all: revealing collaboration patterns by combining authorship and acknowledgements

Données de la Figure 1. Cumulative distribution of papers (with acknowledgements) (%), as a function of numbers of authors

Nb authors	Biology	Biomedical Research	Chemistry	Clinical Medicine	Earth and Space	Eng. and Tech.	Health	Math.	Physics	Prof. Fields	Psychology	Social Sciences
1	2,7%	1,5%	1,3%	1,5%	3,4%	2,6%	4,6%	18,3%	5,3%	14,0%	4,8%	26%
2	12,4%	7,6%	9,6%	6,0%	13,7%	13,8%	11,1%	37,4%	16,5%	26,1%	18,8%	29%
3	17,1%	10,9%	15,4%	8,9%	18,0%	20,4%	16,0%	27,4%	19,4%	29,0%	22,3%	22%
4	18,1%	12,7%	17,9%	11,2%	18,0%	20,6%	16,9%	11,0%	17,2%	16,7%	19,2%	11%
5	15,7%	12,9%	17,2%	12,5%	14,8%	16,2%	15,4%	3,5%	13,1%	7,3%	13,6%	5%
6	11,9%	12,3%	14,3%	12,4%	10,5%	10,8%	12,4%	1,3%	9,3%	3,1%	8,8%	3%
7	8,4%	10,2%	9,8%	10,7%	6,9%	6,3%	7,8%	0,5%	6,2%	1,5%	4,8%	2%
8	5,2%	8,1%	6,1%	8,8%	4,4%	3,8%	5,5%	0,3%	4,2%	0,9%	3,0%	1%
9	3,0%	6,2%	3,8%	6,7%	2,8%	2,2%	3,5%	0,1%	2,7%	0,5%	1,8%	0%
10	2,0%	4,7%	2,2%	5,5%	1,8%	1,4%	2,2%	0,1%	1,7%	0,3%	1,0%	0%
11	1,1%	3,3%	1,0%	3,9%	1,2%	0,7%	1,5%	0,0%	1,1%	0,1%	0,6%	0%
12	0,7%	2,5%	0,6%	2,9%	0,8%	0,4%	0,9%	0,0%	0,7%	0,1%	0,4%	0%
13	0,4%	1,7%	0,3%	2,0%	0,5%	0,3%	0,7%	0,0%	0,4%	0,1%	0,2%	0%
14	0,3%	1,3%	0,2%	1,5%	0,4%	0,2%	0,4%	0,0%	0,3%	0,0%	0,2%	0%
15	0,2%	0,9%	0,1%	1,2%	0,3%	0,1%	0,3%	0,0%	0,2%	0,0%	0,1%	0%
16	0,1%	0,6%	0,1%	0,9%	0,3%	0,1%	0,2%	0,0%	0,2%	0,0%	0,0%	0%
17	0,1%	0,5%	0,0%	0,6%	0,2%	0,0%	0,1%	0,0%	0,1%	0,0%	0,1%	0%
18	0,1%	0,4%	0,0%	0,5%	0,2%	0,0%	0,2%	0,0%	0,1%	0,0%	0,0%	0%
19	0,0%	0,3%	0,0%	0,4%	0,1%	0,0%	0,1%	0,0%	0,1%	0,0%	0,1%	0%
20	0,0%	0,2%	0,0%	0,3%	0,1%	0,0%	0,1%	0,0%	0,1%	0,0%	0,0%	0%
> 20	0,2%	1,3%	0,0%	1,5%	1,4%	0,1%	0,2%	0,0%	1,1%	0,1%	0,1%	0%
Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100%

Données de la Figure 2. Distribution of papers by number of authors (a)

Nb authors	Nb papers	Nb authors	Nb papers	Nb authors	Nb papers	Nb authors	Nb papers	Nb authors	Nb papers	Nb authors	Nb papers	Nb authors	Nb papers
1	35353	31	180	61	12	91	0	121	0	151	2	181	0
2	121570	32	153	62	15	92	8	122	0	152	2	182	0
3	158049	33	124	63	14	93	3	123	0	153	2	183	0
4	158662	34	129	64	17	94	8	124	4	154	3	184	0
5	139453	35	100	65	13	95	8	125	2	155	2	185	1
6	112542	36	114	66	19	96	4	126	3	156	2	186	0
7	82021	37	78	67	15	97	3	127	5	157	0	187	0
8	58784	38	90	68	13	98	7	128	4	158	0	188	1
9	40758	39	73	69	6	99	0	129	1	159	2	189	0
10	29373	40	76	70	7	100	2	130	3	160	1	190	0
11	19348	41	70	71	16	101	5	131	0	161	2	191	0
12	13715	42	61	72	7	102	9	132	3	162	0	192	0
13	9241	43	49	73	7	103	2	133	1	163	1	193	0
14	6612	44	44	74	18	104	2	134	0	164	0	194	0
15	4990	45	45	75	9	105	3	135	0	165	3	195	2
16	3540	46	57	76	6	106	4	136	3	166	0	196	0
17	2599	47	46	77	0	107	0	137	0	167	1	197	0
18	2091	48	44	78	5	108	2	138	0	168	0	198	0
19	1574	49	29	79	6	109	1	139	1	169	4	199	2
20	1325	50	38	80	4	110	2	140	1	170	0	200	0
21	994	51	24	81	12	111	4	141	2	171	4		
22	836	52	27	82	9	112	3	142	0	172	0		
23	671	53	25	83	5	113	1	143	0	173	2		
24	573	54	30	84	5	114	3	144	5	174	1		
25	469	55	34	85	7	115	2	145	4	175	0		
26	376	56	15	86	9	116	2	146	3	176	0		
27	327	57	21	87	1	117	0	147	4	177	0		
28	282	58	18	88	9	118	4	148	0	178	0		
29	245	59	16	89	4	119	2	149	1	179	0		
30	229	60	20	90	4	120	10	150	6	180	1		

Données de la Figure 2. Distribution of papers by number of acknowledgees (b)

Nb acknow.	Nb papers	Nb acknow.	Nb papers	Nb acknow.	Nb papers	Nb acknow.	Nb papers	Nb acknow.	Nb papers	Nb acknow.	Nb papers	Nb acknow.	Nb papers
0	646644	30	98	60	4	90	1	120	0	150	1	180	0
1	133031	31	68	61	5	91	1	121	0	151	0	181	0
2	76997	32	64	62	10	92	2	122	0	152	0	182	0
3	46330	33	48	63	4	93	2	123	0	153	0	183	0
4	30540	34	52	64	3	94	0	124	0	154	1	184	0
5	20392	35	36	65	3	95	1	125	0	155	0	185	0
6	14227	36	27	66	8	96	0	126	2	156	0	186	0
7	9750	37	30	67	3	97	1	127	0	157	0	187	0
8	7247	38	29	68	4	98	0	128	0	158	0	188	0
9	5426	39	27	69	2	99	0	129	0	159	0	189	0
10	3933	40	27	70	2	100	0	130	0	160	0	190	0
11	2960	41	16	71	0	101	0	131	0	161	0	191	0
12	2248	42	18	72	3	102	2	132	1	162	0	192	0
13	1730	43	12	73	3	103	0	133	0	163	0	193	0
14	1363	44	21	74	1	104	0	134	0	164	0	194	0
15	1125	45	19	75	1	105	1	135	0	165	0	195	0
16	844	46	28	76	3	106	1	136	1	166	1	196	0
17	709	47	12	77	2	107	1	137	0	167	0	197	0
18	592	48	9	78	2	108	2	138	0	168	0	198	0
19	491	49	17	79	1	109	1	139	0	169	0	199	0
20	387	50	4	80	2	110	0	140	1	170	0	200	1
21	280	51	12	81	0	111	1	141	0	171	0		
22	280	52	6	82	2	112	0	142	0	172	0		
23	204	53	15	83	2	113	0	143	0	173	0		
24	203	54	10	84	2	114	1	144	1	174	0		
25	169	55	2	85	1	115	0	145	0	175	0		
26	148	56	8	86	1	116	0	146	0	176	0		
27	119	57	10	87	1	117	1	147	0	177	0		
28	98	58	6	88	1	118	0	148	0	178	0		
29	102	59	6	89	1	119	1	149	0	179	0		

Données de la Figure 3. Mean number of authors and acknowledgees, by discipline

Discipline	Mean Nb Ackn.	%	Mean Nb Authors	%	Total
Mathematics	0,553	18%	2,544	82%	3,098
Professional Fields	1,962	39%	3,088	61%	5,050
Eng. & Tech.	0,588	11%	4,609	89%	5,197
Social Sciences	2,822	51%	2,714	49%	5,536
Chemistry	0,662	11%	5,141	89%	5,804
Psychology	1,871	31%	4,163	69%	6,035
Health	1,010	17%	5,096	83%	6,106
Biology	2,594	34%	4,969	66%	7,563
Earth & Space	2,334	30%	5,460	70%	7,794
Clinical Medicine	0,912	11%	7,200	89%	8,112
Biomedical Research	1,330	16%	6,785	84%	8,115
Physics	1,027	9%	10,670	91%	11,697
All Disciplines	1,472	22%	5,203	77%	6,676

Données de la Figure 4. Mean number of acknowledgees by number of authors

Nb authors	1	2	3	4	5	6	7	8	9	10
Biology	3,854	1,670	0,978	0,653	0,468	0,360	0,288	0,261	0,231	0,190
Social Sciences	3,419	1,492	0,881	0,524	0,391	0,319	0,348	0,282	0,318	0,232
Professional Fields	3,141	0,997	0,690	0,357	0,212	0,168	0,140	0,221	0,185	0,046
Earth & Space	2,727	1,215	0,733	0,549	0,435	0,378	0,332	0,301	0,301	0,273
Psychology	2,062	1,009	0,617	0,459	0,343	0,311	0,276	0,188	0,198	0,152
Physics	1,698	0,608	0,357	0,236	0,163	0,129	0,114	0,115	0,108	0,100
Health	1,600	0,551	0,298	0,213	0,174	0,159	0,133	0,129	0,119	0,185
Biomedical Research	1,364	0,659	0,396	0,301	0,230	0,193	0,169	0,154	0,147	0,134
Chemistry	1,234	0,364	0,222	0,157	0,117	0,096	0,090	0,084	0,080	0,085
Clinical Medicine	0,790	0,434	0,278	0,205	0,164	0,132	0,116	0,108	0,100	0,101
Eng. & Tech.	0,757	0,290	0,183	0,134	0,107	0,094	0,095	0,087	0,082	0,082
Mathematics	0,609	0,246	0,184	0,148	0,130	0,113	0,101	0,102	0,126	0,257

Annexe 2. Données de l'article 2

Who are the acknowledgees? An analysis of gender and academic status

Données de la Figure 1. Percentage of female authors, acknowledgees, and acknowledgees who are also authors by discipline

Discipline	Percentage of female acknowledgees who are also authors	Percentage of female acknowledgees	Percentage of female authors
All disciplines	28,3%	29,7%	28,4%
Physics	11,5%	13,7%	13,7%
Mathematics	11,2%	13,8%	16,9%
Eng.and Tech.	18,3%	18,8%	18,7%
Earth and Space	22,5%	23,1%	24,8%
Chemistry	24,3%	25,7%	24,8%
Social Sciences	23,9%	29,0%	32,1%
Biology	30,6%	31,6%	31,2%
Professional Fields	27,4%	32,9%	34,9%
Biomedical Research	35,1%	34,4%	32,2%
Clinical Medicine	44,7%	39,7%	35,1%
Psychology	47,8%	48,1%	47,6%
Health	52,7%	53,7%	55,7%

Données de la Figure 2. Distribution of researchers (all WoS authors, and acknowledgees who are also authors) by number of publications, total field-normalized citations, academic age, and percentage of leading authorships

Note : La figure 2 présente les distributions d'auteurs (WoS) et de remerciés (également auteurs) pour les quatre indicateurs bibliométriques. Étant donnée la taille du jeu de données analysé, il n'est pas possible de présenter les données brutes sous formes agrégées dans la thèse.

Annexe 3. Données de l'article 3

Beyond funding: Acknowledgement patterns in biomedical, natural and social sciences

Données de la Figure 1. Bidimensional Correspondence Analysis for acknowledgements patterns by discipline (plane 1-2)

Note : Les données pour la figure 1 sont présentées en annexe de l'article 3 (Supporting Information) aux pages 143-160 de la thèse.

Données de la Figure 2. Bidimensional Correspondence Analysis for acknowledgements patterns by discipline (plane 3-4)

Note : Les données pour la figure 2 sont présentées en annexe de l'article 3 (Supporting Information) aux pages 143-160 de la thèse.

Annexe 4. Données de l'article 4

Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015

Données du Tableau 2. Acknowledgement words coding results

ID	NP	Financial disclosure	Conflict of interest	Discl.	Ethics	Mat. & resources	Peer comm.	Invest. & analysis	Writing	Dissemin.	Org.	Combination	Vague/ other	Outsiders	Supervision	Total
1	work	261	3	0	1	9	2	3	2	2	4	0	6	0	0	293
2	author	124	31	26	0	24	47	29	2	0	2	9	9	0	0	303
12	analysis	11	0	44	0	8	10	65	0	0	0	0	0	0	0	138
14	preparation	24	0	48	0	6	17	29	9	0	0	2	3	0	0	138
63	collection	8	0	65	0	6	0	94	0	0	0	0	0	0	0	173
68	measurement	21	1	1	1	24	3	108	0	5	3	0	0	1	0	168
15	assistance	36	0	0	0	3	2	118	14	0	0	0	16	0	0	189
18	help	3	0	0	0	2	8	118	19	0	0	0	32	0	0	182
36	contribution	33	0	0	0	0	13	51	2	14	1	3	64	0	0	181
90	interpretation	1	0	109	0	0	23	45	0	0	0	0	0	0	0	178
81	design	4	0	102	0	2	18	40	3	4	12	1	0	0	0	186
157	code	92	0	0	1	0	0	78	0	4	0	0	0	0	0	175
53	access	4	2	5	2	127	0	0	0	42	0	0	0	0	0	182
75	writing	36	0	66	0	7	32	10	29	0	0	0	2	0	0	182
19	data	6	0	43	0	73	1	53	0	5	1	0	0	0	0	182
47	experiment	9	0	3	2	27	9	111	0	2	17	0	0	0	0	180
20	decision	52	0	120	0	0	0	3	0	1	7	0	0	0	1	184
49	results	94	2	28	1	9	26	2	0	16	0	0	4	0	0	182
37	discussion	1	0	0	0	1	176	0	0	2	0	0	0	0	0	180
55	review	2	0	12	3	0	104	3	10	44	0	0	0	0	0	178
Total		822	39	672	11	328	491	960	90	141	47	15	136	1	1	3754

