

# Checks and Balances: Taking a Closer Look at Funding Acknowledgements Texts

Adèle Paul-Hus<sup>1</sup>, Nadine Desrochers<sup>1</sup>

<sup>1</sup>[adele.paul-hus@umontreal.ca](mailto:adele.paul-hus@umontreal.ca); [nadine.desrochers@umontreal.ca](mailto:nadine.desrochers@umontreal.ca)

Université de Montréal, École de bibliothéconomie et des sciences de l'information, C.P. 6128, Succ. Centre-Ville, H3C 3J7 Montreal, Qc, Canada

**Abstract:** Since 2008, funding acknowledgements have been indexed by Web of Science, allowing for large-scale analyses of acknowledgements. The objective of this study is to qualitatively analyse the content of acknowledgements texts in order to contextualize quantitatively obtained trends in acknowledgement practices. The preliminary results show that funding-related uses are more prevalent than expected, since even seemingly non-funding-related terms are in fact linked to funding statements.

**Keywords:** acknowledgements; qualitative content analysis

## 1. Introduction

Acknowledgements are paratextual traces (Genette, 1997) of authors' gratitude towards the colleagues, students, institutions, and funding bodies that contribute to research as a project moves from ideation to publication (Cronin, 1995). As a unique testimony to the many collaborators and support received by a given project, acknowledgements can therefore shed light on research funding and practices, collaboration, and infrastructure in science (Cronin, 2005); this has been well established by more than five decades of acknowledgement-related literature (Desrochers, Paul-Hus and Larivière, 2016).

Since 2008, funding acknowledgements have been indexed by Web of Science (WoS), which has allowed for large-scale analyses of research funding. Despite the name of the field in which acknowledgements are indexed ("Funding Text", FT) and the requirement for funding information to be mentioned in the text in order for the acknowledgements to be indexed (Paul-Hus, Desrochers and Costas, 2016), WoS does not limit the content of the FT to funding information; rather, it collects and indexes the full texts of acknowledgements that contain funding, including the mention of all other types of support acknowledged by authors.

In a recent large-scale analysis of more than one million acknowledgements from research articles and reviews published in 2014, we have shown that papers' acknowledgements provide much more than funding information (Paul-Hus et al., 2016). In order to identify and discriminate between the different types of contributions mentioned in acknowledgements, noun phrases were extracted using natural language processing techniques. The analysis of noun phrase patterns revealed the extent to which acknowledgement practices vary across disciplines, as well as which disciplines acknowledge similar tasks.

However, as stated in a previous phase of the research (Paul-Hus et al., 2016), one should be wary of drawing conclusions based on large-scale analyses without taking a closer look at the meaning of the extracted terms in their original context. Previous analysis has shown

that disclosures of potential conflicts of interest in the acknowledgements section of scientific papers may influence large-scale linguistic analyses because of the presence of pre-formulated statements recommended or required by funding bodies, ethical boards, or journals (Paul-Hus et al., 2016). A fine-grain qualitative content analysis of noun phrases in their full-text context can therefore contextualize the results obtained from the large-scale quantitative analysis, leading to a more robust interpretation of several noun phrases.

The objective of this ongoing analysis is thus to qualitatively analyse the content of acknowledgements through the NPs extracted. More specifically, this paper aims at answering the following research questions:

- What types of contribution and support are acknowledged in scientific papers?
- How do these types of contribution and support vary by discipline?
- Does WoS focus on funding have an impact on acknowledgements content in ways not revealed by the quantitative analysis?

## 2. Methodology

Data for this study were retrieved from WoS' Science Citation Index Expanded (SCI-E) and Social Sciences Citation Index (SSCI), which both include funding acknowledgement data. The full text of acknowledgements from all 2015 articles and reviews indexed in the SCI-E and the SSCI were extracted. The corpus includes a total of 1,009,411 acknowledgements for as many papers. The study focuses on the 214 most frequent noun phrases (NPs) extracted<sup>1</sup>.

For the qualitative phase of the analysis, a first codebook was designed to classify each of the 214 NPs (unit of analysis: individual NP). The initial version of the codebook was established inductively by one coder and revised by a second coder. All NPs were then coded by both coders and their work was reconciled through "negotiated agreement" (Campbell et al., 2013, p. 305; see also Hruschka et al., 2004; Schreier, 2012).

This first codebook serves as the foundation for an ongoing second round of coding, this time for the NPs in the context of the acknowledgement text, using the sentence as the unit of analysis. The first NP coded was the most prevalent in the corpus, the NP "work" (N = 467 689 occurrences across disciplines). Its analysis in context led to refining the codebook to include six categories: 1) The *Research* category includes terms referring to specific tasks related to investigation and analysis, supervision and management, peer communication, writing, the peer-review process, as well as non-specified contributions and overall gratitude; 2) The *People and organizations* category refers to the actual naming or designation of organizations and institutions, authors, awardees, reviewers, referees and editors, peers and colleagues, and outsiders; 3) The *Resources* category is used to code all terms referring to materials, resources, and infrastructure; 4) The *Money and ethics* category is applied to terms related to funding, as well as to conflicts of interest (or lack thereof), ethics, and responsibility disclaimers; 5) The *Dissemination* category groups all terms used to designate the research project at hand, the documents produced, including the paper, and other forms of research dissemination; 6) Finally, an *Ambiguous* code was added for cases in which a use is unclear. The codes are not mutually exclusive; on the contrary, their combination is what brings meaning to the analysis.

---

<sup>1</sup> For details on the noun phrases extraction procedure, please refer to Paul-Hus et al. (2016).

**Table 1. Acknowledgement Codebook**

Category	Code
1) Research	Investigation and analysis Supervision and management Writing Peer communication Contribution Gratitude
2) People and organizations	Organizations Authors Awardees Editors, referees and reviewers Peers and colleagues Outsiders
3) Resources	Materials, resources and infrastructures
4) Money and ethics	Funding Conflict of interest Ethics Disclaimer
5) Dissemination	Project Documents Other dissemination
6) N/A	Ambiguous

The NP “work” will provide an illustration of the coding process. The coding was done following the principles of saturation and qualitative sampling, whereby the sample is “conceptually representative of the set of all possible units” (Krippendorff, 2004, p. 84), in this case to establish the main tendencies. Acknowledgements full texts were stratified by discipline in order to reflect potentially different disciplinary uses of the NPs. Results are presented using rich (or thick) description (Patton, 2015; Krippendorff, 2004).

### 3. Preliminary findings

The findings related to the NP “work” already highlight the importance of analysing the full text of acknowledgements to correctly understand the meaning, in context, of an extracted term. The most prevalent combination of codes for the sentences in which the NP “work” was found is “Organizations” (named, identified) + “Funding” (acknowledged, whether a grant number is present or not) + “Project” (for which the NP “work” stands, in term of describing the whole endeavour). Typical examples include: “This work was supported, in whole or in part, by National Institutes of Health Grant R01CA168700” (Biomedical Research; Artinian et al., 2015); “This work was supported by the Swiss National Science Foundation” (Social Sciences; Purtschert, 2015). Table 2 shows the distribution of this dominant combination in comparison to the total number of sentences coded.

**Table 2. Coding of the NP “work”**  
**Dominant codes: Organizations + Funding + Project**

<b>Discipline</b>	<b>N dominant codes</b>	<b>N total</b>
Biology	8	16
Biomedical Research	8	21
Chemistry	11	17
Clinical Medicine	12	23
Earth and Space	14	20
Engineering and Technology	14	16
Health	12	19
Mathematics	25	36
Physics	13	19
Professional Fields	26	36
Psychology	18	36
Social Sciences	24	34
<b>Total</b>	<b>185</b>	<b>293</b>

Other combinations reinforce the presence of the NP “work” as linked to funding acknowledgements. To the aforementioned combination is sometimes added the code “Awardees”, where a grant or funding is associated with a person through names or initials. This was particularly visible in Biomedical Research and Biology, with an added 7 and 14 occurrences of this 4-code combination, respectively, each further supporting the use of “work” as pertaining to funding; examples include “This work was supported by an ANR grant to F.M. (ANR-12-BSV7-0013-01) and the French Ministry of Research and Education” (Psychology; Battesti, Moreno, Joly & Mery, 2015).

Another reinforcing variation is the relationship between the codes “Project” and “Documents”, both belonging to the “Dissemination” category. In certain cases, “work” is the broader unit in relationship to the paper, as in “This material is based upon work supported by the National Science Foundation under Award DMS-1104013” (Mathematics; Brown, 2015); in other cases, the reverse is seen, where “work” is the part and another term represents the whole, as in “This work is part of the project 'The Distribution and Dynamics of UK Citizens' Environmental Attitudes, Behaviours and Action' funded by the Economic and Social Research Council (ESRC-SDAI Grant no. ES/K002988/1)” (Social Sciences; Longhi, 2015). In either case, whether “work” designates the whole or the part, the coding still confirms the main tendency to use the NP “work” in a funding-related sentence.

Some other uses of the NP “work” can be found, presented here as counter-examples. Two sentences included the code “Peer communication”, as in “This work forms part of a research project submitted as a M.Sc. Dissertation to the Universidade Federal de Vicosa by E. Guatimosim, who would like to thank H. C. Evans for his suggestions” (Biology; Guatimosim et al., 2015; the other example was in the Social Sciences). Finally, two sentences were found where “work” had another meaning altogether; in the following example, the NP was used to designate a contribution spanning whole careers and

vocations: “This article is dedicated to all of the psychiatric-mental health nurses and other mental health clinicians who work diligently to help their patients learn how to manage their auditory hallucinations and stay safe” (Health; Trygstad, Buccheri, Buffum, Ju & Dowling, 2015).

#### **4. Discussion and conclusion**

This thick description demonstrates the inherent limitations of classifying acknowledgements content into clearly delimited categories without referring to their original contexts. It also highlights the limitations of quantitative, automated large-scale tools such as the natural language processing functions used to extract NPs from the FT field. Of course, this qualitative analysis is limited in its scope and coverage, but it should reveal trends that can contextualize meaning in an important way — at least enough to warrant this type of work alongside large-scale analyses.

The qualitative coding of the most frequent NPs will continue in order to assess their meaning in context, as well as the weight of the funding-related requirement for indexing acknowledgements in WoS. Indeed, the economic aspect of scientific research, while inherent to WoS FT, pervades the data in unexpected ways. While the word “work” may not seem funding-related at first glance, it proves to be used in funding-related ways when assessed in context. This type of finding supports the need to find a balance between generalizability, using large-scale datasets, and a deeper understanding of current disciplinary practices through qualitative methods.

*This work is part of a larger research project. Results of the previous phase can be found in Paul-Hus et al. (2016).*

#### **Acknowledgement**

The authors would like to thank the co-authors of the previous phase of this research project: Adrián A. Díaz-Faes, Rodrigo Costas, Maxime Sainte-Marie, Benoît Macaluso, and Vincent Larivière.

*This research was supported by the Social Sciences and Humanities Research Council of Canada.*



Social Sciences and Humanities  
Research Council of Canada

Conseil de recherches en  
sciences humaines du Canada

Canada

#### **Cited examples**

Artinian, N., Cloninger, C., Holmes, B., Benavides-Serrato, A., Bashir, T., & Gera, J. (2015). Phosphorylation of the Hippo pathway component AMOTL2 by the mTORC2 kinase Promotes YAP signaling, resulting in enhanced glioblastoma growth and invasiveness. *Journal of Biological Chemistry*, 290(32), 19387-19401.

<https://doi.org/10.1074/jbc.M115.656587>

Battesti, M., Moreno, C., Joly, D., & Mery, F. (2015). Biased social transmission in *Drosophila* oviposition choice. *Behavioral Ecology and Sociobiology*, 69(1), 83-87.

<https://doi.org/10.1007/s00265-014-1820-x>

Brown, A. W. (2015). Smooth stabilizers for measures on the torus. *Discrete and Continuous Dynamical Systems*, 35(1), 43-58.

Guatimosim, E., Pinto, H. J., Pereira, O. L., Fuga, C. A. G., Vieira, B. S., & Barreto, R. W. (2015). Pathogenic mycobiota of the weeds *Bidens pilosa* and *Bidens subalternans*. *Tropical Plant Pathology*, 40(5), 298-317. <https://doi.org/10.1007/s40858-015-0040-x>

Longhi, S. (2015). Residential energy expenditures and the relevance of changes in household circumstances. *Energy Economics*, 49, 440-450. <https://doi.org/10.1016/j.eneco.2015.03.018>

Purtschert, P. (2015). The return of the native: Racialised space, colonial debris and the human zoo. *Identities*, 22(4), 508-523. <https://doi.org/10.1080/1070289X.2014.944183>

Trygstad, L. N., Buccheri, R. K., Buffum, M. D., Ju, D., & Dowling, G. A. (2015). Auditory hallucinations interview guide promoting recovery with an interactive assessment tool. *Journal of Psychosocial Nursing and Mental Health Services*, 53(1), 20-28. <https://doi.org/10.3928/02793695-20141203-01>

## References

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294–320. <https://doi.org/10.1177/0049124113500475>

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

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

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*, 225-247. Boston, MA: De Gruyter.

Genette, G. (1997). *Paratexts: Thresholds of Interpretation*. (J. E. Lewin, Trad.). New York, NY: Cambridge University Press.

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>

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

Patton, M. Q. (2015). *Qualitative research & evaluation methods : Integrating theory and practice* (4<sup>th</sup> ed.). Los Angeles: SAGE.

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., Desrochers, N., Costas, R., Sainte-Marie, M., Macaluso, B. & Larivière, V. (2016). Beyond funding: What can acknowledgements reveal about credit distribution in science? In Ràfols, I., Molas-Gallart, J., Castro-Martínez, E. & Wolley, R. (Ed.), *Proceedings of the 21<sup>st</sup> International Conference on Science and Technology Indicators, València (Spain), September 14-16, 2016*, 598-606. <http://ocs.editorial.upv.es/index.php/STI2016/STI2016/paper/viewFile/4543/2327>

Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, California: SAGE Publications.