

TYPE: Research in Progress

On the citation gap of articles naming countries

Philippe Mongeon¹, Adèle Paul-Hus², Antoine Archambaut³, Sarah Cameron-Pesant⁴,
Maxime Sainte-Marie⁵, Vincent Larivière⁶

¹*philippe.mongeon@umontreal.ca* ²*adele.paulhus@umontreal.ca* ³*antoine.archambault@umontreal.ca*
⁴*sarah.cameron-pesant@umontreal.ca* ⁵*maxime.sainte-marie@umontreal.ca* ⁶*vincent.lariviere@umontreal.ca*

¹⁻⁶ École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal (Canada)
⁶ Observatoire des sciences et des Technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Montréal (Canada)

Abstract

Titles of scholarly articles are generally a reflection of their content. They inform the reader about the methods, design, results or conclusion of the study, as well as on the context of the research. The mention of the name of a country, for example, provides a geographical contextualization of the article. In order to better understand the effect of these signaling devices on the reception of a study, this research in progress paper investigates the difference in citation rates of articles that mention a country in their title or abstract and articles that do not. It shows, using WoS-indexed papers published between 1996 and 2013, that mentioning a country in either the title or the abstract is associated with lower citation rates, and that this is observed for every country when all disciplines are combined. The gap in citation rates is also greater in Social Sciences than in other disciplines, which is likely due to their stronger focus on national issues.

Conference Topic

Research on local topics and issues, citation analysis, country-level studies

Introduction

The title of a research article play a central role in the scholarly communication process: it is usually the first point of contact with its potential readers (Diener, 1984; Gross, Harmon & Reidy, 2002; Salager-Meyer & Alcaraz Ariza, 2013). According to Yitzhaki (1994, p. 321), the “primary functions [of a title] are to draw a reader’s attention to a paper and to indicate its content from a short glimpse, thus contributing to its initial selection or rejection”. The title can be considered as the most important summary of a research article since it is “generally the first (and sometimes the only) information obtained from the published article” (Paiva, Lima, & Paiva, 2012, p. 509). It can also reflect the content of a paper by describing its methods, design, results or conclusion, or by revealing important contextual attributes of a piece of research. Similarly, we can posit that mentioning a country name in the title of an article constitutes a geographical contextualization of the article in question.

Several bibliometric studies have investigated the relationship between title characteristics and scholarly impact. Many studies found that articles with shorter titles were more cited (Gnewuch & Wohlrabe, 2017; Paiva et al., 2012; Subotic & Mukherjee, 2014), while others found the opposite (Habibzadeh & Yadollahie, 2010; Jacques & Sebire, 2010). Other studies found that the direction of the correlation between title length and citations differs between disciplines (van Wesel, Wyatt, & ten Haaf, 2014) or found no significant correlation (Rostami, Mohammadpoorasl, & Hajizadeh, 2014). Buter and van Raan (2011) found that the occurrence of at least one non-alphanumeric character in the title was associated with a higher impact, but the correlation strength varied by discipline and was in some cases non-significant or negative.

Titles containing a hyphen or colon (Jacques & Sebire, 2010; Rostami et al., 2014), and titles containing acronyms have also been associated with higher citations rates. According to Subotic and Mukherjee (2014), we should remain cautious when interpreting the results of these studies since “we are only beginning to understand how relevant title characteristics relate to each other in an integrative context” (p. 123). Indeed, “what is said probably matters more than how it is said” (p. 121).

We found five studies investigating the relationship between citations and the presence of a specific geographical region in articles’ title, which is the focus of the present study. Nair and Gibbert (2016) showed that context attributes of titles, such as reference to a specific country, company or industry name, had no significant influence on citations, yet the results of the four other studies point in the opposite direction. Indeed, titles referring to geographical feature were associated with fewer citations by Paiva et al. (2012), Rostami et al. (2014), Jacques and Sebire (2010) and Abramo, D’Angelo and Di Costa (2016).

While most of these studies obtained a negative correlation between the geographical contextualization of a research and its scientific impact, they either used small samples (a few hundred publications) or focused on specific disciplines. The study by Abramo et al. (2016) attempted to overcome these limits by analyzing all WoS-indexed Italian publications published between 2004 and 2011. The authors divided articles in a “Country” group containing all publications mentioning “Italy” in their title, abstract or keywords (n=40,024), and a “No country” group including all the remaining publications (n=416,686). Despite being the largest study on the topic performed so far, its methodological design has serious shortcomings. For instance, factors other than the discipline that might affect the number of citations received by a paper were not controlled. Moreover, the “No country” group contains all publications that do not mention Italy, and therefore likely contains articles that mention *other* countries. Consequently, it is unclear how well the study isolates the effect of country mentions on citations. Another limitation of the research relates to its small geographical scope, both at the level of the researchers (the analysis is restricted to publications by authors affiliated to Italian institutions) and at the level of the research object (the analysis is restricted to publications mentioning Italy). This raises questions regarding the generalizability of their findings to other geographical and thematic contexts.

The present research aims to overcome these shortcomings by analyzing a large dataset of worldwide publication and controlling for article similarity. We also investigate how the amplitude of the citation (dis)advantage of articles naming countries in their title or abstract varies according to the country mentioned. More specifically, we provide answers to the following research questions:

1. What is the relation between the mention of a country in the title or abstract of articles and their relative number of citations?
2. How does this relation differ by discipline?
3. How does this relation differ by mentioned country?

Methods

We used all articles, notes and reviews published in Clarivate Analytics’ Web of Science (WoS) between 1996 and 2013, each assigned to one of four broad categories—Arts and Humanities (AH), Biomedical Research (BM), Natural Sciences and Engineering (NSE) and Social Sciences (SS)—based on the journal’s NSF classification. The twenty countries with the highest number of papers in the WoS over that period were chosen for this analysis (see Table 1). We searched for these countries’ names as well as their associated adjectives and demonyms to

create our “Country” dataset. For each article included in the “Country” dataset, we searched for the most similar article published in the same journal and year, but without any country mention (i.e., not limited to the twenty countries included in the analysis) in the title or abstract¹. To measure the similarity between two articles, we computed the cosine similarity of their respective title and abstract (reduced to their constitutive noun phrases) as well as the Jaccard similarity of their references list (references to WoS source items only). We then ranked candidate pairs per these two criteria and, to ensure maximal proximity between the two articles, only cases where the same article was ranked first per both criteria were considered in the comparative analysis. Table 1 reports the resulting number of article pairs for each country and discipline. The group “Title” includes articles for which the country was mentioned in the title, and the “Abstract” group refers to articles for which the country was mentioned in the abstract, and did not appear in the title. Also, we decided to exclude Arts and Humanities from the analysis since the number of papers in the dataset was too small (a total of 1,456 and 2,161 articles in the “Title” and “Abstract” groups, respectively).

Table 1. Number of articles with a country name in the title or abstract by country and discipline

Country	BM		NSE		SS	
	Title	Abstract	Title	Abstract	Title	Abstract
Australia	5,307	10,579	9,118	12,524	2,884	5,971
Belgium	770	1,837	577	1,551	327	977
Brazil	6,353	7,555	7,002	9,261	1,539	2,033
Canada	4,233	9,721	6,696	11,166	2,282	5,247
China	18,645	21,766	19,608	20,914	6,434	6,845
France	4,469	9,430	2,931	7,927	1,433	3,164
Germany	4,470	12,213	2,306	8,337	2,486	5,638
India	7,134	8,498	8,747	10,746	2,148	2,458
Italie	5,180	8,339	3,818	9,075	1,555	2,726
Japan	13,913	15,283	7,977	10,399	2,588	2,984
Netherlands	3,170	7,383	975	2,659	1,870	4,408
Poland	1,117	1,719	1,251	2,408	364	676
Russia	785	1,495	1,500	3,019	639	995
South Korea	5,772	4,332	3,039	2,914	2,076	1,340
Spain	4,328	7,889	3,880	9,273	2,385	4,406
Sweden	3,901	6,841	1,792	3,248	1,623	3,042
Switzerland	1,501	4,165	1,404	3,152	452	1,059
Taiwan	3,258	3,621	2,217	2,888	1,873	3,001
United Kingdom	8,658	26,984	8,238	13,686	7,355	16,827
United States	11,226	56,393	12,538	80,999	8,153	24,387
Total	114,190	226,043	105,614	226,146	50,466	98,184

Results

Figure 1 shows the average difference between the field-normalized citations rates (to enable the comparison of effect size between disciplines) of articles that mention a country in their abstract and their associated control article (i.e., the most similar article with no mention of a country). Four observations can be made. First, the results show that, on average, there is a

¹ To identify mentions of other countries, we used a list of 593 country names, adjectives and demonyms, retrieved from:

https://en.wikipedia.org/wiki/List_of_adjectival_and_demonymic_forms_for_countries_and_nations

citation disadvantage for articles who mention a country either in their title or abstract. This is true for all countries when all disciplines are combined. Second, the citation disadvantage appears to be smaller when a country is mentioned in the abstract only, except for articles in BM that mention the United States or Germany or Sweden, and articles in NSE mentioning China or Sweden. Third, the citation disadvantage is generally higher in SS than in BM and NSE. Fourth, the citation disadvantage appears to be much higher for some countries than others. While this varies by discipline, we observe that the citation disadvantage is generally smaller for countries who have traditionally dominated the scientific front. There are however notable exceptions, such as papers in NSE mentioning China that appear to have the smallest citation disadvantage. In BM, articles mentioning the United States or Germany in their title also stand out, as they exhibit the opposite results than other countries and fields: they receive on average more citations than the control articles, and they are also more cited on average than articles who mentioned the country only in the abstract.

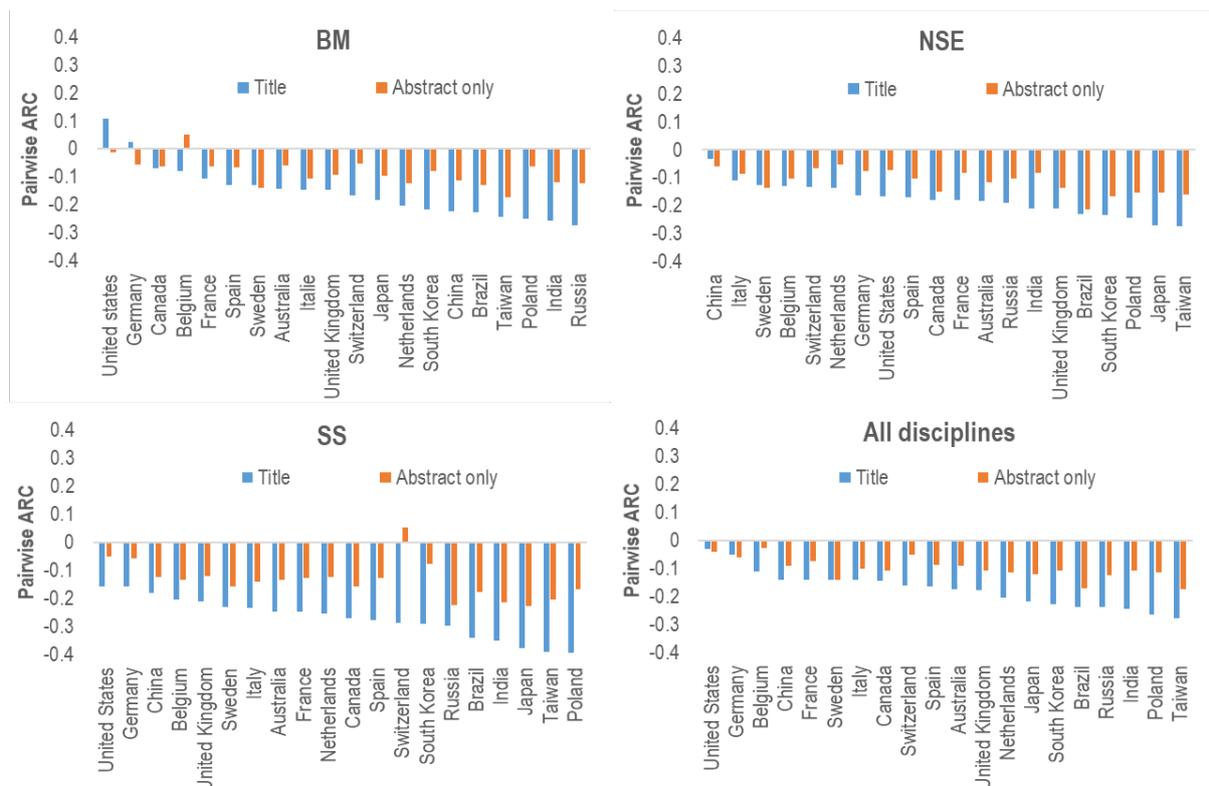


Figure 1. Citation (dis)advantage when mentioning a country in the title or abstract by country and discipline.

Discussion and conclusion

Our results show that articles mentioning a country in their title or abstract receive on average less citations than similar articles that do not mention a country and, thus, are coherent with those of Paiva et al. (2012), Rostami et al. (2014), Jacques and Sebire (2010) and Abramo et al. (2016). However, since we compared articles from the first group with their closest neighbour (based on the title, abstract and reference similarity), our method allows a better control for the research topic, and thus better isolates the “country effect” than previously used methods. Despite the unequivocal nature of the results, important nuances should be considered in their interpretation. For instance, while the appearance of a country in the title or abstract is likely to reflect the *local* relevance of an article—or at least its research setting—it is unclear to what extent the absence of geographical contextualization reflects its *global* relevance.

Consequently, it remains unclear whether the observed citation disadvantage is caused by a *perceived* lack of global relevance in the eyes of other researchers or by an *actual* lack of relevance of the articles for research in other geographical or global contexts. The greater citation gap—observed in Social Sciences—can be explained by the fact that research in these disciplines is more often contextualized within a specific region or nation, compared with Natural Sciences where research objects are often by definition international (Gingras & Mosbah-Natanson, 2010). Thus, it could be argued that in Social Sciences, mentioning a country in the title or abstract does act as the marker of a geographical or national context of the research, and that the results may in fact be difficult to transfer to other contexts. This raises a few questions. Is the citation gap caused by the title of the article, or by its content? Would removing the geographical location in title of articles with a local or national focus result in higher citations? Given that the title of a paper serves as a signalling device for those to whom it may be most useful, and that research on local topics is particularly relevant for local stakeholders (Stremersch, Verniers, & Verhoef, 2007), removing markers of geographical context from titles in an attempt to increase citations might come at the expense of other forms of impact.

This research in progress paper also shows that the amplitude of the citation disadvantage varies by mentioned country. Since some countries have a larger research output than others, this may be partially explained by the idea that researchers in a given country are perhaps more likely to do research on local topics, but also more likely to perceive articles mentioning their country as relevant. Moreover, given the importance of citations in research evaluation processes, even researchers in smaller and/or non-English speaking countries might be incentivized to work on research of local issues in larger English-speaking countries like the United States to get more citations (Larivière, 2014). This potential effect on citations might also be amplified by the overrepresentation of research published in the English language in the WoS database (Mongeon & Paul-Hus, 2016).

Further steps of this research will be to 1) extend our list of geographical markers to ensure that the “No country” group does not mention any location. It will also include the country of the authors in the analysis since research on local issues performed by local scientists may be of more local (and less global) relevance than topics that are studied by the global scientific community, which might also influence citation rates. Finally, there is empirical evidence that research focusing on local issues tends to be more interdisciplinary (Chavarro, Tang, & Rafols, 2014), and that interdisciplinarity is often associated with higher citation rates (Larivière, Haustein, & Börner, 2015). Thus, further research should also look at how geographical contextualisation of research is related to interdisciplinarity, and how this may affect citation rates.

Acknowledgements

This study was funded by the Social Sciences and Humanities Research Council of Canada and the Canada Research Chair on the Transformation of Scholarly Communications.

References

- Abramo, G., D’Angelo, C. A., & Di Costa, F. (2016). The effect of a country’s name in the title of a publication on its visibility and citability. *Scientometrics*, 109(3), 1895-1909.
- Biacchi, A. (2003). Relation complexity of titles and texts: A semiotic taxonomy. In L. Merlini Barbaresi (Ed.), *Complexity in language and text* (pp. 319-341). Pisa, Italy: PLUS.
- Buter, R. K., & van Raan, A. F. J. (2011). Non-alphanumeric characters in titles of scientific publications: An analysis of their occurrence and correlation with citation impact. *Journal*

- of Informetrics*, 5(4), 608-617.
- Chavarro, D., Tang, P., & Rafols, I. (2014). Interdisciplinarity and research on local issues: Evidence from a developing country. *Research Evaluation*, 23(3), 195-209.
- Diener, R. A. V. (1984). Informational dynamics of journal article titles. *Journal of the American Society for Information Science*, 35(4), 222-227.
- Gingras, Y., & Mosbah-Natanson, S. (2010). Les sciences sociales françaises entre ancrage local et visibilité internationale. *European Journal of Sociology*, 51(2), 305-321.
- Gnewuch, M., & Wohlrabe, K. (2017). Title characteristics and citations in economics. *Scientometrics*, 110(3), 1573-1578.
- Gross, A. G., Harmon, J. E., & Reidy, M. S. (2002). *Communicating Science: The Scientific Article from the 17th Century to the Present*. Oxford University Press.
- Habibzadeh, F., & Yadollahie, M. (2010). Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals. *Croatian Medical Journal*, 51(2), 165-170.
- Jacques, T. S., & Sebire, N. J. (2010). The impact of article titles on citation hits: An analysis of general and specialist medical journals. *JRSM Short Reports*, 1(1), 1-5.
- Larivière, V. (2014). The importance of national journals. *University Affairs*. <http://www.universityaffairs.ca/opinion/in-my-opinion/importance-national-journals/>
- Larivière, V., Haustein, S., & Börner, K. (2015). Long-distance interdisciplinarity leads to higher scientific impact. *PLOS ONE*, 10(3), e0122565.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213-228.
- Nair, L. B., & Gibbert, M. (2016). What makes a “good” title and (how) does it matter for citations? A review and general model of article title attributes in management science. *Scientometrics*, 107(3), 1331-1359.
- Paiva, C., Lima, J., & Paiva, B. (2012). Articles with short titles describing the results are cited more often. *Clinics*, 67(5), 509-513. [https://doi.org/10.6061/clinics/2012\(05\)17](https://doi.org/10.6061/clinics/2012(05)17)
- Rostami, F., Mohammadpoorasl, A., & Hajizadeh, M. (2014). The effect of characteristics of title on citation rates of articles. *Scientometrics*, 98(3), 2007-2010.
- Salager-Meyer, F., & Alcaraz Ariza, M. A. (2013). Titles are “serious stuff”: a historical study of academic titles. *JAHHR - European Journal of Bioethics*, 4(7), 257-271.
- Stremersch, S., Verniers, I., & Verhoef, P. C. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing*, 71(3), 171-193.
- Subotic, S., & Mukherjee, B. (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science*, 40(1), 115-124.
- van Wesel, M., Wyatt, S., & ten Haaf, J. (2014). What a difference a colon makes: How superficial factors influence subsequent citation. *Scientometrics*, 98(3), 1601-1615.
- Yitzhaki, M. (1994). Relation of title length of journal articles to number of authors. *Scientometrics*, 30(1), 321-332.