The HPS “Long View” and the Case of the Journal Article

Charles H. Pence • @pence@scholar.social

CEFISES HPS, 2025-12-05

UCLouvain

LIBST
LOUVAIN INSTITUTE OF BIOMOLECULAR SCIENCE AND TECHNOLOGY

Outline

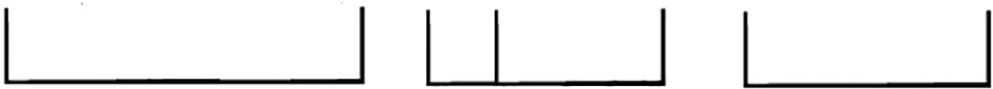
1. A contemporary problem: text and practice
2. A shifting historical target
3. Analyzing these shifts computationally

The take-home: Longitudinal, historical studies of "journal articles" need to be sensitive to the changing historical status and function of journal publication.

Journal articles are often the paradigmatically “stable” data source

Anatomy of a Digital Object Identifier (DOI)

https://doi.org/ 10.7935 / K5H41PBP



Resolver service

Directory indicator

Assigned by DOI Foundation

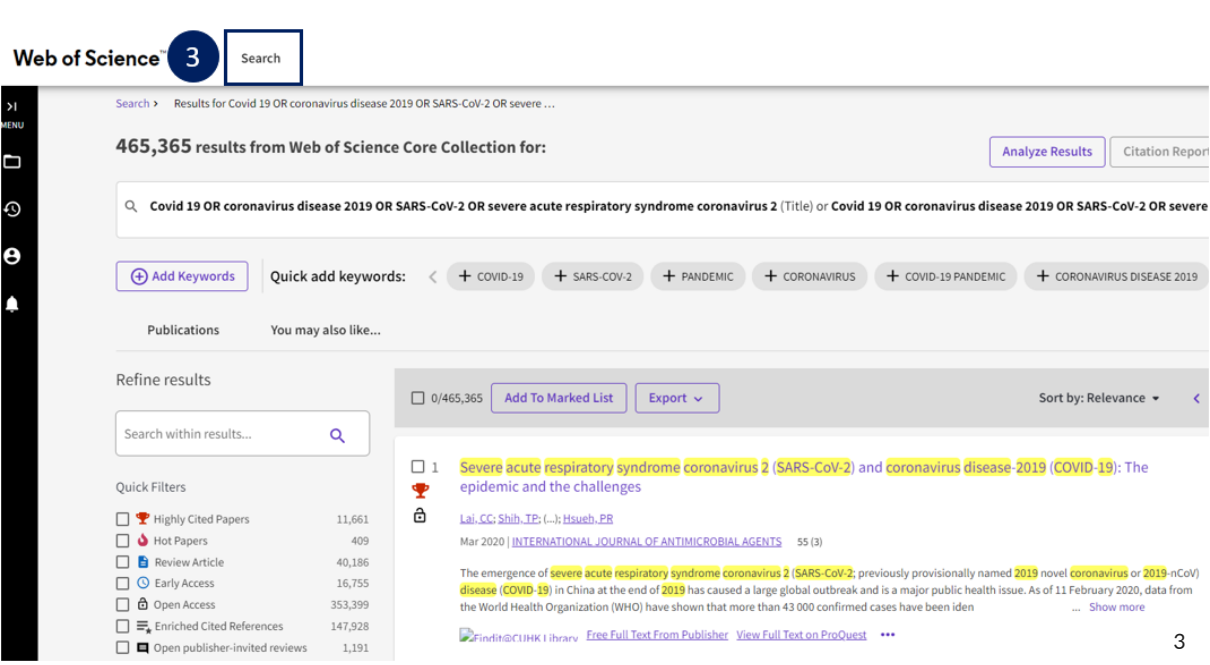
Prefix

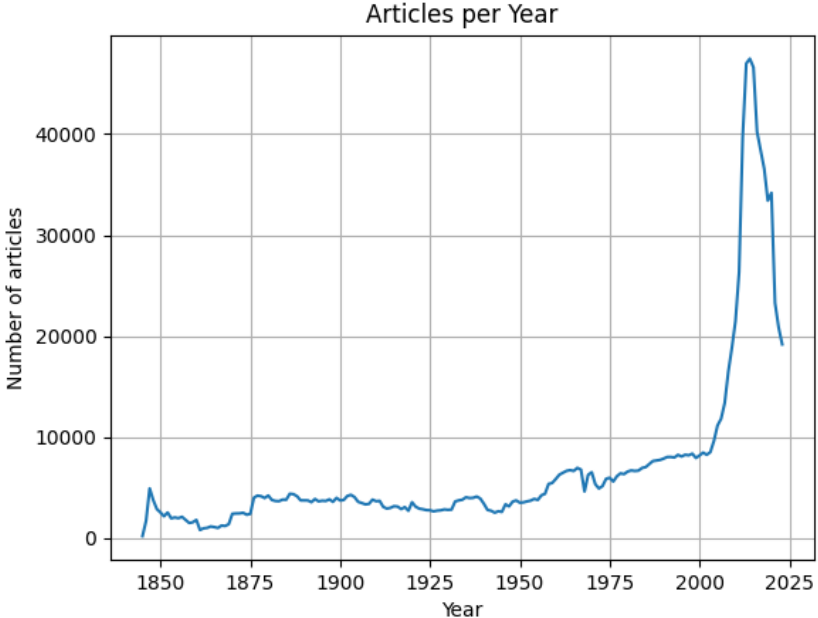
Assigned by DOI Registration Agencies (RAs) e.g. CrossRef, DataCite, etc.

Suffix

Assigned by RA Members e.g., F1000 Research, California Digital Library (EZID)

[**Adapted from Austalian National Data Service (ANDS) Digital Object Identifier System and DOI Names (DOIs) Guide, http://www.ands.org.au/guides/doi]**


Web of Science 3
Search
Search > Results for Covid 19 OR coronavirus disease 2019 OR SARS-CoV-2 OR severe ...
465,365 results from Web of Science Core Collection for:
Analyze Results
Citation Report
Covid 19 OR coronavirus disease 2019 OR SARS-CoV-2 OR severe acute respiratory syndrome coronavirus 2 (Title) or Covid 19 OR coronavirus disease 2019 OR SARS-CoV-2 OR severe
Add Keywords
Quick add keywords:
+ COVID-19
+ SARS-COV-2
+ PANDEMIC
+ CORONAVIRUS
+ COVID-19 PANDEMIC
+ CORONAVIRUS DISEASE 2019
Publications
You may also like...
Refine results
Search within results...
Quick Filters
Highly Cited Papers 11,661
Hot Papers 409
Review Article 40,186
Early Access 16,755
Open Access 353,399
Enriched Cited References 147,928
Open publisher-invited reviews 1,191
0/465,365
Add To Marked List
Export
Sort by: Relevance
1
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges
Lai, CC; Shih, TP; (...); Hsueh, PR
Mar 2020 | INTERNATIONAL JOURNAL OF ANTIMICROBIAL AGENTS 55 (3)
The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; previously provisionally named 2019 novel coronavirus or 2019-nCoV) disease (COVID-19) in China at the end of 2019 has caused a large global outbreak and is a major public health issue. As of 11 February 2020, data from the World Health Organization (WHO) have shown that more than 43 000 confirmed cases have been iden
... Show more
Findit@CUHK Library
Free Full Text From Publisher
View Full Text on ProQuest


Articles per Year
Number of articles
0
10000
20000
30000
40000
1850
1875
1900
1925
1950
1975
2000
2025
Year

Text and Practice

> [A “professionalization”-based approach to the history of science] ignores, or at best treats as peripheral, the forms in which knowledge appeared, **assuming that publication in specialist periodicals was already established** as the only legitimate means for announcing new discoveries, thus **downplaying other methods such as conversation, books, letters, and museum displays.** (Secord 2009, p. 444)

Digital Literature Analysis for Empirical Philosophy of Science

Oliver M. Lean, Luca Rivelli, and Charles H. Pence

Empirical philosophers of science aim to base their philosophical theories on observations of scientific practice. But since there is far too much science to observe it all, how can we form and test hypotheses about science that are sufficiently rigorous and broad in scope, while avoiding the pitfalls of bias and subjectivity in our methods? Part of the answer, we claim, lies in the computational tools of the digital humanities, which allow us to analyse large volumes of scientific literature. Here we advocate for the use of these methods by addressing a number of large-scale, justificatory concerns—specifically, about the epistemic value of journal articles as evidence for what happens elsewhere in science, and about the ability of digital humanities tools to extract this evidence. Far from ignoring the gap between scientific literature and the rest of scientific practice, effective use of digital humanities tools requires critical reflection about

A Shifting Historical Target

> ...historians of science have tended to read specialist journals as historical sources rather than as historical phenomena in their own right, often taking the existence of journals for granted rather than viewing them as objects whose existence requires explanation. (Baldwin 2015, p. 11)

An assumption: the “modern” journal article

> [Before the early twentieth century,] journals by no means constituted a uniquely appropriate medium for announcing a new claim to discovery, and if the category “scientific journal” existed at all, it did not correspond to [the modern conception]. When learned journals first emerged in the seventeenth century, they took after newspapers and gazettes, by design ephemeral and often in disrepute. (Csiszar 2018, p. 5)

> One problem with these ideas is that they suppose that the format and uses of journals and papers have remained more or less constant throughout their existence. But this is not true. The *Philosophical Transactions* itself exemplifies the problem: few publications have taken on as diverse a set of formats and meanings as this publication has over its long history. (Csiszar 2018, p. 12)

The Objects which it is proposed to attain by this Periodical are, first, to place before the general public the results of Scientific Work and Scientific Discovery, and to urge the claims of Science to a more general recognition in Education and Daily Life; and secondly, to aid Scientific Men themselves, by giving early information of all advances made in any branch of natural knowledge throughout the world, and by affording them an opportunity of discussing the various scientific questions which arise from time to time. (advertisement for *Nature*, 1869)

In the nineteenth century, contributors began using *Nature* and its weekly turnaround time to debate scientific questions and to give abstracts of longer forthcoming papers in monthly or quarterly journals. In the early twentieth century, some contributors began employing a new strategy and used *Nature* for the immediate publication of interesting results before a paper was prepared or submitted elsewhere. (Baldwin 2015, p. 14)

En troisième lieu, on fera sçavoir les expériences de Physique et de Chimie, qui peuvent servir à expliquer les effets de la Nature, les nouvelles découvertes qui se font dans les Arts et dans les Sciences, comme les machines et les inventions utiles ou curieuses que peuvent fournir les mathématiciens ; les observations du Ciel, celles des Météores, et ce que l'anatomie pourra trouver de nouveau dans les animaux. ("L'imprimeur au lecteur," *Journal des Sçavans*, 1665)

1. “un catalogue exact des principaux livres qui s’imprimeront dans l’Europe”
2. “quand il viendra à mourir quelque personne célèbre… on en fera l’éloge”
3. (sciences)
4. “les principales décisions des Tribunaux séculiers et ecclésiastiques”
5. “...on taschera de faire en sorte qu’il ne se passe rien dans l’Europe digne de la curiosité des gens de Lettres, qu’on ne puisse apprendre par ce journal.”

> Les articles étaient très courts, parfois de quelques lignes seulement ; il était rare qu'ils occupassent plus de deux pages. (Morgan 1928, 73)

> Au XVII^{e} siècle, le succès d'un journal dépendait encore plus qu'aujourd'hui, de la personnalité de son rédacteur. Ni la hardiesse de Sallo, ni la négligence de Gallois n'avaient pu détourner l'intérêt du monde lettré de la nouvelle critique. (Morgan 1928, 175)

How can we make our readings (whether close or distant) of these texts **sensitive** to these shifts in use and broader context?

Ways Out

1. Classifying articles by “type” in advance?
2. “Multi-modal” analyses, just for text?
3. How to balance close reading with digital work?

> As journals became not only purveyors of scientific news but also archives of discovery, it became more common to conceive of science as a series of discrete discovery events localized in time and connected with an individual author. This raised other tricky questions about the status of collective knowledge. (Csiszar 2018, p. 8)

How can we analyze “scientific articles” with **different kinds of propositional content** – claims to new discoveries, full justifications of scientific knowledge claims, summaries of other articles, debates among scientists, editorials, obituaries, prize notices…

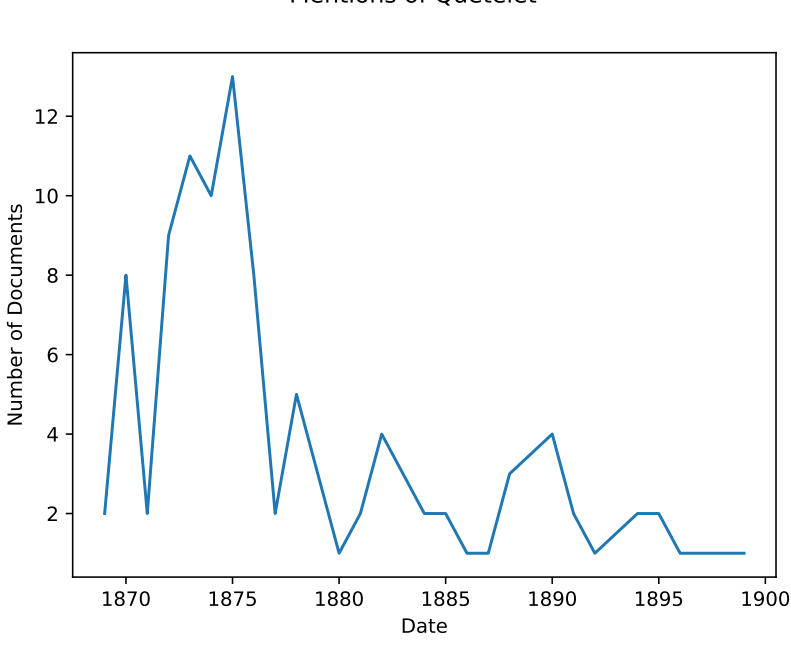
Case Study

Let's take a case study of this issue that's (1) manageable in size, (2) known to have lots of different kinds of articles, and (3) about which I'm expert enough in the ground truth to confirm analyses by hand.

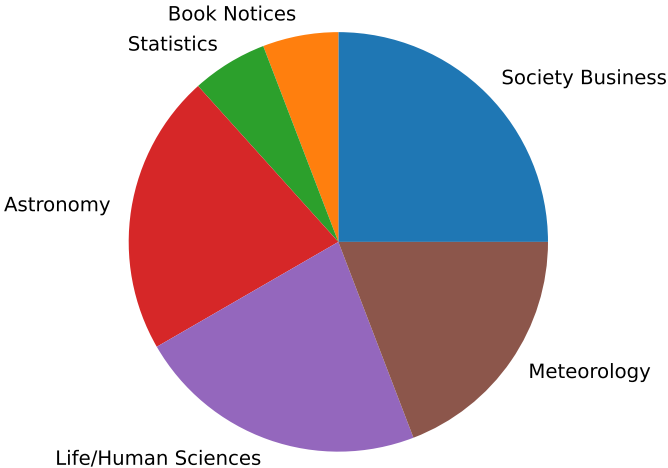
Case Study

Let's take a case study of this issue that's (1) manageable in size, (2) known to have lots of different kinds of articles, and (3) about which I'm expert enough in the ground truth to confirm analyses by hand.

Mentions of Quetelet in *Nature*, prior to 1900.


Mentions of Quetelet
Number of Documents
Date
12
10
8
6
4
2
1870
1875
1880
1885
1890
1895
1900

Type of Mention



Ways Out

1. **Classifying articles by “type” in advance?**
2. “Multi-modal” analyses, just for text?
3. How to balance close reading with digital work?

Automated Classification

What if we hand-coded the papers that I pulled from searching Quetelet ($n = 104$, very manageable for manual analysis), and tried to use them to build an automated classifier?

Four classes:

1. normal “scientific” articles
2. book/paper notices
3. notes
4. society business

Classifier Results

Applied to all articles pre-1900 in *Nature*:

Unclassified scientific pieces	28,456	79%
Book/article notices	1,489	4%
Notes	4,853	13%
Society business	1,372	4%

Book and Article Notices

[Book Reviews]
Our Astronomical Column
Notes
Societies and Academies
Scientific Serials
Our Book Shelf
University and Educational Intelligence
Geographical Notes
Books, Pamphlets, and Serials Received
Books Received

Notes

Societies and Academies
Notes
Scientific Serials
University and Educational Intelligence
Our Astronomical Column
[Book Reviews]
Books, Pamphlets, and Serials Received
Our Book Shelf
Geographical Notes
Books Received

Society Business

Notes
Societies and Academies
Our Astronomical Column
Scientific Serials
University and Educational Intelligence
Geographical Notes
Books, Pamphlets, and Serials Received
[Book Reviews]
Diary of Societies
Books Received

Results

Perfect? Nope.

But promising enough, with lots of tuning left to be done.

Ways Out

1. Classifying articles by “type” in advance?
2. **“Multi-modal” analyses, just for text?**
3. **How to balance close reading with digital work?**

Questions?

charles@charlespence.net • https://pencelab.be • @pence@scholar.social

PENCE LAB

fnrs
LA LIBERTÉ DE CHERCHER

Scientific Text and Scientific Practice

To use scientific texts for HPS:

scientific practice ↔ scientific literature ↔ useful generalizations ↔ empirically informed HPS

From Literature to Generalizations

This is the task of **digital humanities,** with a healthy assist from **corpus linguistics** – though that latter connection often goes unnoticed!

Corpus Linguistics

- **sociolinguistics** — meaning is defined by past use within the corpus; “meaning does not concern the world outside the discourse” (Teubert 2005)
- **cognitive linguistics** — corpora let us test hypotheses about the use and meaning of concepts in language (the “entrenchment” of some uses of concepts as “grammatically acceptable”; Glynn 2014)

Corpus Linguistics

This **won't** be my topic today – but it's important to note that we probably *should* be paying more attention to it than we are…

Why Write Journal Articles?

From Practice to Text

How might we theorize about **the reasons that scientists write journal articles?**

From Practice to Text

How might we theorize about **the reasons that scientists write journal articles?**

There is a **significant** literature on this topic from the perspective of contemporary science.

A Low Bar

Journal articles have to be good for **something,** or scientists wouldn't spend time reading them together in journal clubs!

The “Logical” View

Articles are the vehicle by which scientists intend to **justify their findings and conclusions to each other**

The Structure of a Scientific Paper*

Frederick Suppe†‡
Department of Philosophy, Committee on the History and Philosophy of Science,
University of Maryland

Scientific articles exemplify standard functional units constraining argumentative structures. Severe space limitations demand every paragraph and illustration contribute to establishing the paper's claims. Philosophical testing and confirmation models should take into account each paragraph, table, and illustration. Hypothetico-Deductive, Bayesian Inductive, and Inference-to-the-Best-Explanation models do not, garbling the logic of papers. Micro-analysis of the fundamental paper in plate tectonics reveals an argumentative structure commonplace in science but ignored by standard philosophical

The “Sociological” View

Articles are intended to **gain converts to the research program,** the recorded history of **a power struggle**

The Manufacture of Knowledge

An Essay on the Constructivist and Contextual Nature of Science

by

KARIN D. KNORR-CETINA

Department of Sociology
University of Pennsylvania, Philadelphia

The “Fraud” View

Articles are **fake, inductive presentations** of the “real” scientific reasoning, which is actually Popperian or falsificationist

SCIENCE IN BOOKS

IS THE SCIENTIFIC PAPER FRAUDULENT?

Yes; It Misrepresents Scientific Thought

I HAVE chosen for my title a question: Is the scientific paper a fraud?

I ought to explain that a scientific "paper" is a printed communication to a learned journal, and scientists make their work known almost wholly through papers and not through books, so papers are very important in scientific communication. As to what I mean by asking "is the scientific paper a fraud?" Now, John Stuart Mill's deeper motive in working out what he conceived to be the essential method of science was to apply that method to the solution of sociological problems: He wanted to apply to sociology the methods which the practice of science had shown to be immensely powerful and exact. It is ironical that the application to sociology of the inductive method, more or less in the form in which Mill himself conceived it, should have been an almost entirely fruitless one.

The simplest application of the Millsian process of induction to sociology came in a rather strange movement called Mass Observation. The belief underlying Mass Observation was apparently this: that if one could only record and set down the actual raw facts about what people do and what people say in pubs, in trains, when they make love to each other, when they are playing games, and so on, then somehow, from this wealth of information, a great generalization would

I ought to explain that a scientific "paper" is a printed communication to a learned journal, and scientists make their work known almost wholly through papers and not through books, so papers are very important in scientific communication. As to what I mean by asking "is the scientific paper a fraud?" evidence until the "discussion" section, and in the discussion you adopt the ludicrous pretense of asking yourself if the information you have collected actually means anything.

Of course, what I am saying is rather an exaggeration, but there is more than a mere element of truth in it.

The conception underlying this style of scientific writing is that scientific discovery is an inductive process. What