Mapping the evolution of early modern natural philosophy: corpus collection and authority acknowledgement

Andrea Sangiacomo, Raluca Tanasescu, Silvia Donker & Hugo Hogenbirk


To link to this article: https://doi.org/10.1080/00033790.2021.1992502

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 15 Nov 2021.

Submit your article to this journal

Article views: 718

View related articles

View Crossmark data
Mapping the evolution of early modern natural philosophy: corpus collection and authority acknowledgement

Andrea Sangiacomo, Raluca Tanasescu, Silvia Donker and Hugo Hogenbirk

Faculty of Philosophy, University of Groningen, Groningen, Netherlands

ABSTRACT
Although natural philosophy underwent dramatic transformations during the seventeenth and eighteenth centuries, studying its evolution as a whole remains problematic. In this paper, we present a method that integrates traditional reading and computational tools in order to distil from different resources (the four existing Dictionaries of early modern philosophers and WorldCat) a representative corpus (consisting of 2,535 titles published in Latin, French, English, and German) for mapping the evolution of natural philosophy. In particular, we focus on gathering authors and works that were (directly or indirectly) engaged with the teaching of natural philosophy in the early modern academic milieu. We offer a preliminary assessment of the relevance of our corpus by investigating one aspect of this evolution, namely the trends in the acknowledgments of authorities linked with different and competing approaches to natural philosophy (scholastic, Cartesian, and Newtonian). The results not only corroborate existing knowledge, but they also show distinctive features and differences within these trends that were not observed previously, thus illustrating the heuristic potential of our computational method for corpus collection.

ARTICLE HISTORY
Received 4 May 2021
Accepted 8 October 2021

KEYWORDS
Early modern natural philosophy; corpus collection; distant reading; computational history of science; Scientific Revolution

Quivi ebbe Astolfo doppia meraviglia: che quel paese appresso era si grande, il quale a un picciol tondo rassimiglia a noi che lo miriam da queste bande; e ch’aguzzar convienigli ambe le ciglia, s’indi la terra e ’l mar ch’intorno spande, discernir vuol; che non avendo luce, l’imagin lor poco alta si conduce.
L. Ariosto, Orlando Furioso, xxxiv.71
1. Looking for early natural philosophy

Early modern natural philosophy is usually considered to be the ancestor of today’s natural sciences. At the beginning of the seventeenth century, natural philosophy was established in university curricula as a field of study strongly shaped around Aristotle’s *libri naturales*, including the *Physics*, although more directly influenced by late-scholastic, sixteenth-century authorities, and commentators. By the end of the eighteenth century, natural philosophy was profoundly changed. Studying this transformation, however, is not an easy task. The key difficulty is that it is far from obvious how to find ways of identifying and tracking what natural philosophy was at the time and how it evolved.

Three main approaches have been most popular among historians. The first consists in focusing on particular authors by studying, for instance, how Descartes or Aristotelians developed their own natural philosophy. The second approach consists in reconstructing a general narrative about the main shifts and transformations that the domain underwent during the early modern period. The most dominant narrative has been that of the Scientific Revolution, which has been contested from different perspectives, and even declared to be a ‘myth’, although it still enjoys some support. One of the major concerns raised against this second approach is that it builds a broad narrative by relying on a relatively small fraction of authors and works. The third approach seeks a middle way by focusing on particular episodes in the evolution of early modern natural philosophy or particular groups such as, for instance, the emergence of experimental philosophy in early seventeenth-century England. All these approaches share the same problem of representativeness: how can we extrapolate from the study of a particular case (or set of

---


3 Daniel Garber, ‘Galileo, Newton and all that: if it wasn’t a scientific revolution, what was it? (a manifesto)’, *Circumscribe*, *International Journal for the History of Science*, 7 (2009), 9–18; Helen Hattab, *Descartes on Forms and Mechanism* (Cambridge: Cambridge University Press, 2009).


cases) and offer a general account about the evolution of early modern natural philosophy as a whole?

An essential aspect of this difficulty consists in selecting an accessible and suitable set of sources (a corpus) for studying how natural philosophy was conceived and practiced at any given time in a certain historical context. Obviously enough, the corpus must be accessible, because we can do little with works that we cannot read in any way. But the corpus must also be suitable, in the sense that there must be some matching between the historian’s interests and questions and what the historical materials can provide. Traditional scholarship usually does case studies on small corpora (from one single work to a few works, perhaps some dozens), mostly because of any human reader’s limits in accessing and processing a large amount of information. It is clear that working with a small number of sources is no guarantee for establishing a suitable corpus. Small numbers can yield too little or something that is simply partial or not representative enough for studying broader trends, such as those that arguably shaped the evolution of early modern natural philosophy.

In this paper, we present a method for corpus collection that can be used by historians to inventory a larger corpus, potentially accessible, and whose suitability for the historic research can be directly tailored. This method would thus facilitate a leap into a larger-scale study of the history of natural philosophy and possibly provide the grounds for a greater integration of computational techniques in the toolkit of the historian of early modern philosophy and science. Even if the current paper is concerned mostly with introducing the method and illustrating its implementation, in doing so we shall provide some anticipations of how computational approaches might be integrated in the current discussion of the evolution of early modern natural philosophy.

In designing the method in question, we took into account three main desiderata. A method for corpus collection must be (i) customizable depending on the historian’s agenda and interests, (ii) scalable to different degrees of inclusiveness with respect to available sources, and (iii) meaningful, that is, yielding results that meet some of the initial expectations while also leading to new insights. In order to keep our discussion within the limits of a single paper, we shall present one specific instantiation of our method, in which we set certain departing assumptions (customization), make certain decisions concerning the sources to include or to filter out (scalability), and assess the heuristic potential of the obtained inventory (meaningfulness). However, as we document throughout our discussion, that does not mean that this method is exhausted by the final inventory that we present here, which is best understood as one instance of one implementation of the method itself. The purpose of this paper is not to justify any final inventory of early modern sources for natural philosophy and discard the method used to arrive at that inventory. Quite on the contrary, the purpose is to design a method that is simple, flexible, powerful, and insightful enough to help historians produce any
inventories that would fit best their desiderata and that they can actually use in their own research.

Our method for corpus collection combines human and computational tools to analyze available sources and make an inventory of authors and works representative of early modern natural philosophy (based on a customized understanding of it, as we shall explain below). In the current implementation, the final collected corpus includes 2,535 titles published in Latin, French, English, and German over a two-hundred-year period. We do not claim that this corpus provides the golden standard for the study of early modern natural philosophy, since such a standard cannot be defined apart from specific research questions that different historians might have. Nevertheless, we show that the corpus we collected is meaningful in the sense that it both corroborates existing scholarly knowledge about some major aspects of the evolution of natural philosophy and provides new insights about facets of this evolution that have not been observed previously.10

Given that this paper relies only on the inventory of a corpus (the list of titles included and the names of their authors), we will not offer an in-depth textual analysis of the content of the books we gathered. Nevertheless, we will show that even the inventory we produced alone might already offer a wealth of information to be mined further and interesting historical suggestions to be drawn. In particular, we focus on what we shall call ‘authority acknowledgement’ within the corpus we reconstructed. By ‘authorities’ we mean proper names of authors (or recognizable schools) that are mentioned in the titles of the works included in our collected corpus. It should be stressed from the very beginning that by ‘authorities’ we do not focus directly on the role that individual authors as such played in the shaping of natural philosophy. Rather, we are interested in tackling how proper names of certain authors were used as school flags for signalling the engagement with certain approaches, traditions or methods that were established, emerging or challenged in the debate on natural philosophy. We suggest that these references do not happen randomly, but they underscore major moves in the trends that shaped the evolution of the field, and different strategies adopted by the competing schools.11

---

10This combination of corroboration and new insights is usually recognized to be a valid criterion to support the soundness of methods aiming at combining computational and human analysis (or the so-called ‘distant’ and ‘close’ reading, respectively). See Franco Moretti, ‘Conjectures on World Literature’, New Left Review, 1 (2000), 54.

11Working with only the titles of the works might seem reductionist or inappropriate for historians of science. However, in other fields such as literary studies, studying large corpora by examining the titles they include has proved to be a very resourceful way of uncovering macro-features and patterns that shape the corpus itself. See for instance Franco Moretti, Graphs, Maps, Trees. Abstracts Models for a Literary History (London and New York: Verso, 2005); Matthew L. Jockers, Macroanalysis: Digital Methods and Literary History (Chicago, IL: Chicago University Press, 2013); Franco Moretti, F. Distant Reading (London and New York: Verso, 2013); Katherine Bode, A World of Fiction: Digital Collections and the Future of Literary History (Ann Arbor: University of Michigan Press, 2018). We suggest that this heuristic potential suits early modern philosophy and science well.
The implementation of computational approaches to the field of history of science is growing, and topics related to early modern natural philosophy have received some attention in recent years. Nonetheless, computational tools usually require large-scale corpora, which is precisely what historians of early modern natural philosophy still miss. The problem is not only a material lack of digitized books. This is in fact becoming less of a concern, given the growing mass of digitized materials more and more available even in the public domain. In order to deal with a large corpus, it is vital for this corpus to be gathered in such a way that historians can retain a degree of control over it, in order to check whether and to what extent it is not only accessible, but also suitable for the purposes of their investigation. Our corpus collection method provides the first attempt of its kind, endeavouring to create the ideal conditions for a more sustained and methodologically sound implementation of computational tools to the study of early modern natural philosophy as a whole. For this reason, our discussion is mostly exploratory in nature. It offers procedures and reflections that are potentially scalable or replicable by other historians facing similar problems but working on different subjects. Since there is very little established knowledge about how exactly to proceed in this respect, our paper charts the challenges, approaches, and limitations that the integration of computational and human analysis faces in this field, which we document throughout our discussion.

The paper proceeds as follows. Section 2 presents the leading intuition of our method, including our pragmatic working definition of early modern natural philosophy and our approach for integrating different kinds of sources currently available. Section 3 illustrates one of the many possible implementations of this methodology in order to inventory a corpus of early modern natural philosophy. Section 4 reflects on the meaningfulness of this inventory by discussing the insights that can be derived from studying authority acknowledgements in our corpus. Section 5 provides some concluding remarks about the potential of the presented method and findings for future research in computational history of philosophy and science.


14To mention some of the richer repositories: Google Books, Bibliothèque Nationale de France (BNF), Munich Digitization Centre (MDZ) of the Bavarian State Library, EEBO (Early English Books Online), EECO (Eighteenth Century Collections Online), e-rara, and archive.org.
2. Method

Any research project in early modern natural philosophy immediately confronts the difficulty of defining what its subject matter is supposed to be. For instance, mathematics and natural philosophy were quite sharply distinguished fields at the beginning of the seventeenth century, but they became increasingly more intertwined by the end of the eighteenth century. By contrast, crucial issues in early seventeenth-century natural philosophy included the relationship between bodily and non-bodily factors in living and non-living natural entities, whose study had an impact on both natural philosophy and medicine. However, by the end of the eighteenth century, they decisively grew into distinct and rather independent fields of study. The challenge, thus, consists in rightly dealing with the overlap between ‘natural philosophy’ broadly construed as a theoretical discipline concerned with the causes of natural phenomena and other fields, disciplines or debates, which include medicine, astronomy, alchemy, practical arts, theology, metaphysics, psychology, and more. Taking an overly inclusive attitude towards natural philosophy would make it a mere umbrella term for a variety of different practices, not necessarily always connected or engaged in the same enterprise. Taking an overly narrow focus would instead miss several important connections with other fields, which in their turn arguably influenced the evolution of natural philosophy.

In addressing this difficulty, we take a pragmatic approach. By ‘pragmatic’ we mean that we focus first on how natural philosophy was practiced at the time and how this practice can be feasibly reconstructed. Instead of taking today’s definitions and assumptions about what early modern natural philosophy was at face value, we attempt to reconstruct what was presented as ‘natural philosophy’ during the seventeenth and eighteenth centuries. In particular, we aim at inventorying all those authors and works that (directly or indirectly, actually or potentially) contributed to define how natural philosophy should have been taught during the period (both within and outside of universities). The advantage of this pragmatic approach is that it does not require today’s researchers to take for granted any substantive claims about the essence and scope of natural philosophy. Instead, one has to direct attention to the (changing) way in which natural philosophy was presented as a subject of teaching and learning during the period. Since the actual content of natural philosophy was contested, the

---

16See Gaukroger 2010 (n. 8).
expectation is that materials relevant for teaching purposes will also reflect how competing views opposed each other and tended to prevail over one another over time.

Our leading intuition is that by focusing on what was presented as ‘natural philosophy’ by those authors and works most directly committed to offer a systematic and comprehensive presentations of it, we can uncover a large number of sources that dealt with natural philosophy while simultaneously engaging with what today would be taken to be other fields of study. At the same time, we could also recognize (as documented in the following discussion) which works were clearly centred on problems other than providing a systematic account of natural philosophy. Hence, we could decide that they might be less relevant for the purposes of our current investigation. In other words, our approach is to study natural philosophy as an actor category (namely, as early modern scholars would have understood it, with all the diversity, fuzziness, and flux that this understanding historically entailed).

As we shall illustrate, this approach is consistent with customization, since we can deliberately play with different categories and keywords, depending on our research interests, to make our definition more or less inclusive or restrictive. In the actual implementation of the method illustrated below, we decide to focus on natural philosophy as a ‘speculative’ science concerned with studying natural phenomena for the sake of uncovering their causes. This is admittedly a backward-looking definition more attuned to late-scholastic positions. We chose to give it priority because we aim to explore to what extent (in time and space) this sort of understanding of natural philosophy would allow us to gather materials that their own authors would define as ‘natural philosophy’.18

As part of the customization of our method, we thus adopt the following working definition of ‘natural philosophy’ and the threefold classification that it entails:

Early modern natural philosophy is best understood as a field of study and debate, which is covered by teaching practices within an academic milieu at a given moment in time. In this field, three concentric spheres can be distinguished:

---

18 For instance, a large part of the books collected through our method does include discussions of astronomy and physiology as sub-fields of natural philosophy, in keeping with the relatively broad understanding of natural philosophy at the beginning of the seventeenth century. However, the inventory we discuss here does not include later works devoted entirely to astronomy, chemistry or medicine because of the systematicity criterion that guided our project: a work included in the inventory should present a systematic treatment of natural philosophy as such (which can also overlap with other fields). Works exclusively devoted to astronomy, chemistry and medicine do not fit this criterion. Furthermore, as we explain in what follows, making this decision was not a matter of automatic implementation of semantic rules, but was determined by the critical human annotation of the lists of potentially relevant works we initially put together.
Primary authors and works (core area): authors and works devoted to offer a systematic exposition of natural philosophy that could be used to teach it to new generations.19

Secondary authors and works (buffer area): authors and works devoted to expand, deepen and elaborate on some accepted systematic exposition linked with the core.20

Tertiary authors and works (peripheral area): authors and works that incorporate elements presented in the primary and secondary areas and bring them into dialogue with other disciplines, problems or debates.21

The fundamental intuition at the basis of this general working definition is that the best way of getting a sense of what ‘natural philosophy’ was at any given moment of time during the early modern period is to look at what was taught as ‘natural philosophy’ at that moment. Our definition offers an operational tool to direct attention towards appropriate sources for reconstructing the profile of the field of study. By focusing on ‘teaching,’ we do not necessarily refer to actual teaching practices (how the discipline was taught, using one pedagogical method or another, etc.), although these practices are included in the scope of this expression. One author might not have been an actual teacher of natural philosophy, but her or his works might have been written with the intention of contributing to the teaching of the discipline. In fact, there are a number of materials produced outside of university curricula that aim at illustrating (‘teaching’) new ways of doing and understanding natural philosophy.22 We do include all these materials in the scope of our research provided that they materialized in printed books (as they often did). Our focus on ‘teaching’ is thus meant to include all those accounts of the disciplines that are designed to provide an ideally complete presentation of what natural philosophy is, and address this presentation to someone (like a student, but also amateurs and other learned people) who want to learn about it. In this sense, we take ‘teaching’ in a normative rather than descriptive sense, as the author’s intention of writing a book that can be used to show what should be discussed by natural philosophers. We consider these kinds of works to be representative of the most ‘systematic’ accounts of natural philosophy. They usually provide definitions of the technical terminology used in the field, outline the main

19These include, for instance, Jean Crassot, Totius philosophiae peripateticae corpus absolutissimum (1619); Johannes Greydanus, Institutiones physicae libris undecim absolutae (1664); Émilie du Châtelet, Institutions de Physique (1740). Notice that we include more general multi-volume works (like Crassot’s) that systematically have one volume (at least) devoted to natural philosophy.

20Several examples are William Gilbert, De magnete magnetisque corporibus, et de magnomagnete tellure; physiologia nova, plurimis et aurgmentis, experimentis demonstrata (1600); Burchard de Volder, Oratio de conjugendis philosophicis et mathematicis disciplinis (1682); Bryan Robinson, Dissertation on the Aether of Sir Isaac Newton (1743).

21Among them Mary Astell, Letters concerning the Love of God (1695); Johannes Regius, Oratio de vera mentis cum corpore unione (1703); Benjamin Wilson, A Treatise on Electricity (1750).

topics, problems and issues addressed by authors involved with natural philosophy, and often advocate for specific solutions or methods used to solve them.

We stress the connection between this idea of systematicity and teaching practices by focusing on authors and works connected with an academic milieu. The expression ‘academic milieu’ aims at capturing something broader than what is included within the formalized university system for higher education. The academic milieu does include this university system, but it also encompasses all those other authors and works that were directly or indirectly linked with academic practices and were aiming at contributing to the debate and evolution of the field of natural philosophy. The three spheres of the academic milieu of natural philosophy define three broad groups of authors and works, which will be called ‘primary’ (core), ‘secondary’ (buffer) and ‘tertiary’ (periphery).

The primary works are those that are clearly systematic in nature, most comprehensive, most likely to be used as teaching materials and offer fewer doubts about the fact that they concern natural philosophy as a whole, for instance, textbooks, compendia, and overall systems of natural philosophy. Of course, primary works may include authors who are competing to impose different ‘systematic’ accounts of natural philosophy. Detecting these conflicting views will be a crucial element for future research. For now, it is enough to observe that primary works are simply the least controversial to identify.

Secondary works are similar to primary works except that they are not necessarily systematic. For instance, student disputations, academic orations, and other works on particular subjects within the domain of natural philosophy would be included in this category. Secondary works are often relevant to glance at more specific core issues that are debated in the discipline across time and space, even if they do not provide a complete and systematic picture of the discipline as a whole, as in the case of primary works.

Tertiary works are not necessarily or directly connected with natural philosophy as such but show a relevant use of key notions, debates or trends in natural philosophy for related topics and discussions. Tertiary works are not explicitly designed to provide a complete account of the discipline of natural philosophy as a whole. And yet, they can significantly contribute to study the domain of natural philosophy, especially by providing a sort of ‘view from the outside’ of the strict core of the discipline and by allowing the research to glance at how the internal debate in natural philosophy had any spillover in other fields. Identifying tertiary works is more difficult and arguably controversial. This is mostly based on the qualitative assessment of the information available about individual authors and works, and on explicit connections made with topics in natural philosophy in the titles of the works themselves.

---

In implementing the above-mentioned working definition of natural philosophy, we apply two main restrictions to it, in order to make our research more feasible. First, we filter out multiple editions or translations of the same book, since our primary goal is to reconstruct the degree of diversity of our corpus by considering how many unique titles and authors could be included. Our method can successfully be used to explore these further sources as well, but this task exceeds the limits of this present paper.24

By narrowing the focus of our research from the very beginning, we managed to actually test and check the viability of our approach in a way that would have proved unpractical if we had set overly ambitious goals from the start. As we demonstrate below, the method works and yields sizable and meaningful results. There is nothing in the method itself that prevents it from being applied to recover the information that will be filtered out in this current implementation (for instance, multiple editions or translations of the same work). Moreover, the impact of particular works due to multiple editions or translations can be best assessed by relating it to a background context that shows the actual diversity and multiplicity of unique authors and works published throughout the period. The latter is precisely the sort of corpus that we are seeking to reconstruct here, which can then also become propaedeutic for the study of how specific works and their multiple editions fared in the early modern period with respect to their many peers.

Second, we focus on printed books only by thus not including manuscript sources and new forms of publications like journal articles. Manuscripts would have been very difficult to trace and process with the technical means we use and the sources we have access to. Starting from the second half of the seventeenth century, new periodical journals like the Philosophical Transactions published by the Royal Society, or Acta Eruditorum become very popular venues for the dissemination of philosophical and scientific ideas. Nevertheless, we decided not to include these sorts of publications because of the technical difficulties in tracing all the papers that would be relevant for the subject of our study, and how to weight them against books. Moreover, journal articles are often used to disseminate new findings, hypotheses or theories, rather than presenting the current state of the art. We then assumed that if anything extremely important was circulated on a given paper, we should be able to see it reflected in more systematic works as well.25 Once again, this is part of our initial customization of the

---

24 The study of multiple editions and reprint of even just one single work is in itself a very demanding and complex task, as documented in Valleriani 2017 (see above n. 12) and De sphaera of Johannes de Sacrobosco in the Early Modern Period: The Authors of the Commentaries, ed. by Matteo Valleriani (Cham: Springer Nature, 2020).
25 Consider, for instance, the case of Antoine Le Grand’s An Entire Body of Philosophy (London, 1694). Le Grand writes what is basically a Cartesian textbook, but he is extremely keen on integrating experimental findings and trials discussed by all sorts of other authors, including Robert Boyle.
method. Nothing in the method itself theoretically prevents from including further sources.26

We use the threefold classification provided above as an operation tool for recognizing and classifying works and authors that will be included in our corpus. As we illustrate in Section 3, we implement this threefold classification in the context of human inventorying and annotation of lists of titles. The working assumption we make is that using the above definitions as guidelines, human annotators will have relatively small disagreements in identifying primary works, they might have more disagreement about secondary works, while tertiary works might offer even fewer clear-cut cases and more need for discussion. These categories are thus interpreted in terms of increasing uncertainty (or decreasing certainty) about their actual relevance for the purpose of defining natural philosophy as such. When discussing primary (and secondary, to an extent) works, one can be relatively certain that these works deal with natural philosophy proper just by looking at the title or the table of contents. In the case of tertiary works, this is much less certain and may be more open to interpretation and disagreement (and most often requires additional information about the work or its author). In this sense, this tripartite division does not create a sharp profile of the academic milieu of natural philosophy, but introduces a more nuanced picture, whose margins become fuzzier and fuzzier as one recedes from the core and moves towards the periphery.

Having defined natural philosophy as the tripartite topology of authors and works, we considered two sources for inventorying those authors and works: (a) the existing bio-bibliographical Dictionaries for seventeenth- and eighteenth-century philosophers of various nationalities (Dutch,27 British,28 French,29 and German30); and (b) the world-wide library catalogue (WorldCat).

(a) The Dictionaries are top-down sources, collected by teams of individual scholars based on their qualitative assessment and investigation of a vast array of authors and works. For each author included, the Dictionaries

---

26Hence, at this stage, we do not take into account the issue of time delay that might occur between the discovery or first discussion of new ideas in semi-public forms or even in private circles and their broader spreading. In other words, our focus is on tackling how ideas began to be normalized through circulating in printed books, not how they emerged.


provide (more or less extensive) biographical and intellectual sketches and some discussion of the contents of the authors’ works. This information is extremely helpful for a preliminary assessment of the relevance of particular authors or works for the field of natural philosophy. However, the *Dictionaries* have been assembled by different research teams, in different periods, using different criteria. Although inclusive, they are obviously not perfect and likely suffer from some omissions, which are impossible to quantify based on the *Dictionaries* themselves. They are not focused on natural philosophy per se, but rather on philosophy broadly understood, which entails that *Dictionaries* will likely include many natural philosophers, but not necessarily all of them. Moreover, the *Dictionaries* are incomplete from a geo-historical point of view, since they cover only four main national areas (British, Dutch, French, and German), and two of them cover only one century (seventeenth-century French philosophy and eighteenth-century German philosophy).

(b) *WorldCat* includes all authors and titles that are preserved in today’s library catalogues around the world. Obviously, *WorldCat* is not restricted to any particular domain of knowledge. This massive amount of data, however, is accessible only via keyword search. In order to decide which keywords to use and how relevant they may be, one needs to know beforehand how the field of natural philosophy was shaped and evolved across space, time, and different languages. One can surely begin with obvious keywords such as ‘natural philosophy’, but it is methodologically crucial to find criteria to decide which keywords should be taken into account, in which contexts, and when it is reasonable to assume that all the relevant keyword options have been investigated.

Before moving any further, an important disclaimer seems in order. Much of the following discussion will be shaped by our initial decision to rely on these two sources. They represent the best available sources we had access to at the moment we began this study. The *Dictionaries* are the only available tool to go beyond the narrow selection of authors and works canonized in the great majority of existing secondary literature. The *WorldCat* aggregator is the best tool to search simultaneously all library catalogues available online, without having to rely on what is included or excluded in any particular collection of any specific library. Of course, both sources have a number of limitations and implicit biases.

Take the *Dictionaries*, for instance. Their chronological covering is arbitrarily set in terms of those authors that were philosophically active within one or two hundred years, while we know that historical periods do not begin or end on any specific year. Considering the seventeenth or the eighteenth century as ‘the early modern’ period might then sound too rigid or unwarranted. This is right, of course. And yet, from a pragmatic point of view, it
makes sense to focus on a century and take it as the core of a period, admitting that the beginning and end of that period will have fuzzy boundaries, which are then more controversial to define and difficult to study. For the purposes of this paper, then, accepting the chronology entailed by the Dictionaries (i.e., a focus on whole centuries) does introduce some limitations (some sixteenth-century works and authors will be excluded from our study). However, we contend that these limitations are ultimately acceptable and even unavoidable. Using ad hoc criteria in order to try to define exact boundaries for our period of study would surely create more idiosyncratic and controversial results than simply accepting a somehow arbitrary chronological range based on the selection already established in the Dictionaries.

A perhaps more worrisome point concerns the focus on nationalities and languages in a period in which neither of them had the same profile we are familiar with today. Speaking of ‘British’, ‘Dutch’, ‘French’ or ‘German’ authors in the early modern period clearly needs a good deal of qualifications. Indeed, sometimes different Dictionaries include the same author, like Descartes who figures both in the Dutch and French Dictionaries (because he was French but lived in the Dutch Republic for a long time). Taking a pragmatic approach, however, it seems more fruitful to accept national labels for what they are, namely, labels used for the sake of broadly classifying authors working roughly in the same geo-political quarter of early modern Europe. In order to prevent historical inaccuracies, of course, one should avoid equating these national labels with the use of a certain language: Dutch authors most often wrote in Latin, sometimes in French, French authors both in Latin and French, British authors often in English, but also in Latin. The use of a language is not enough by itself to associate a certain author with a certain geo-political label.

From now on, wherever we use national labels to describe authors and works, it should be understood that we use the same labels introduced by the Dictionaries for the purely conventional and practical purpose of classifying them. In the most straightforward case, any mention of ‘nationality’ means ‘Dictionary-based nationality’. Hence, ‘Dutch’ means any author originally included in the Dictionary of Seventeenth and Eighteenth-Century Dutch Philosophers. In the case of authors that we added based on our search in WorldCat, we proceeded via human annotation in order to assign each new author to one of the four Dictionary-based nationalities. In order to do so, we relied on the Virtual International Authority File (Viaf.org) as a main source of bio-bibliographic information, alongside other secondary online databases, such as

---

31 Other examples include, for instance, Germans Christoph Scheibler, Johann Heinrich Schweitzer, and Johann Stier in the dictionary of seventeenth-century British philosophers and Daniel Duncan, who appears both in the dictionary of French philosophers and in the British one.

32 We then use today’s names for these four nationalities (and the corresponding contemporary abbreviations: UK, NL, FR, DE) since this is the terminology used in the titles of the Dictionaries themselves.
data.bnf.fr. When other references were not available, we assigned the nationality at birth. When it was not possible to assign any of the four Dictionary-based nationalities, we considered the author as belonging to ‘other (not-Dictionary-based) nationalities.’

Keeping this methodological disclaimer in mind, our discussion seeks to demonstrate the procedure that we designed in order to combine the distinctive strengths of our two main sources in a way that would compensate their respective shortcomings. The procedure is the following:

Step 1.a Human analysis of the information provided in the Dictionaries of British, Dutch and French philosophers,33 aimed at building a Dictionary-based corpus of primary, secondary and tertiary authors and works.

Step 1.b Consolidation of the Dictionary-based corpus via human annotation.

Step 2.a Generating relevant keywords from the Dictionary-based corpus in order to search for relevant titles in WorldCat.

Step 2.b Scraping and cleaning of the results obtained via WorldCat (both via automated cleaning34 and human annotation).

Step 3.a Back-checking the methodology by repeating steps 1 and 2 for the German Dictionary.

Step 3.b Back-checking whether the overall newly acquired corpus generated the same keywords used at the beginning (step 2a). If not, repeat the whole procedure until no new results are produced.

Step 3.c Comparing the final list of authors to the four Dictionaries indexes in order to check how many authors included in the corpus via the scraping of WorldCat were initially overlooked at step 1.a.

Figure 1 offers a visual representation of our workflow based on this procedure. The next section explains how these steps have been implemented and outlines the results that they generated.35

3. Implementation

(Step 1.a) The first step of the research consisted in inventorying the relevant primary, secondary and tertiary authors and works included in the British, Dutch and French Dictionaries. The principal investigator was responsible for creating the initial three lists of relevant authors and works for each

---

33We decided to start with the two Dictionaries that cover the whole period (British and Dutch) and then added the French one, because it covers the first half of our chronological span.

34We would like to thank the University of Groningen’s Center for Information Technology for providing access to the Peregrine high-performance computing cluster and to scientific programmer Cristian A. Marocico for building the customized scraping and deduplication code.

35A first version of the complete list of works we inventoried is available on Zenodo (10.5281/zenodo.5566681).
Dictionary. Each list was subsequently independently annotated by at least two other annotators. All problematic or conflicting cases have been discussed. The German corpus was dealt with separately (Step 3), after all the other Dictionary-based corpora were created, analyzed, and expanded, so that the team could verify whether the keyword generation method (Step 2.a) and the ensuing steps were actually viable.

(Step 1.b) The overall results of the annotation procedure are summarized in Table 1. Table 1 shows that the impact of annotation has been more significant on secondary and tertiary works. The number of titles that move from one category to another due to annotations correlates with the category itself. Annotations re-classified as secondary 22.3% of the Dutch tertiary works and 38.8% of the French ones. In the British corpus the movement is from secondary to tertiary works (10.7%) due to an initial classification of 11 electricity-related titles as secondary. These changes confirm our expectation that primary works are easier to identify than secondary or tertiary.

(Step 2.a) The lists of works generated in Step 1.b provided the starting point for the extraction of relevant keywords that would be used for the scraping of titles in WorldCat. The extraction of keywords was done by analysing the

---

36All annotators had a background knowledge in the history of early modern philosophy. They were instructed about our definitions of primary, secondary, and tertiary works. They were then presented with the list of works (including authors, titles, and publication details) and they were asked to indicate in which category each title should be placed.

37We decided to classify works on electricity as tertiary because they are most often focused on particular experiments and even when they offer theoretical reflections, they do not concern ‘natural philosophy’ as such, but rather attempt to define the nature of a particular phenomenon.

38Our choice of traditional word frequency lists was conditioned by the size of our study sample (i.e. the lists of titles), which was too small to employ more refined statistical tools, such as topic modelling or word2vec, which are better suited for whole-text searches (T. R. Tangherlini and P. Leonard, ‘Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research’, Poetics 41, no. 6 (2013), 725–49) or large corpora analysis (S. Graham, S. Weingart, and I. Milligan, ‘Getting Started with Topic Modeling and Mallet’, Programming Historian, web, 2012). Moreover, keywords represent the building blocks of bibliographic databases and keyword searches have been a common practice in large-scale bibliographic research across the humanities.
Table 1. Deviations in the dictionary-based corpus composition after annotation.

<table>
<thead>
<tr>
<th>Number of titles</th>
<th>NL 1</th>
<th>NL 2</th>
<th>NL 3</th>
<th>FR 1</th>
<th>FR 2</th>
<th>FR 3</th>
<th>UK 1</th>
<th>UK 2</th>
<th>UK 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before annotation</td>
<td>47</td>
<td>88</td>
<td>112</td>
<td>74</td>
<td>39</td>
<td>36</td>
<td>66</td>
<td>108</td>
<td>85</td>
</tr>
<tr>
<td>After annotation</td>
<td>51</td>
<td>103</td>
<td>92</td>
<td>69</td>
<td>56</td>
<td>24</td>
<td>65</td>
<td>102</td>
<td>88</td>
</tr>
<tr>
<td>Deviation</td>
<td>+8%</td>
<td>+11.7%</td>
<td>−17.85%</td>
<td>−6.76%</td>
<td>+43.5%</td>
<td>−33.3%</td>
<td>−1.16%</td>
<td>−5.56%</td>
<td>+3.52%</td>
</tr>
</tbody>
</table>
titles of both the annotated and non-annotated corpus derived from the Dictionary (we observed no differences between these two, inferring that annotation did not affect the keywords extracted). Words and two-word collocates that exceed a certain threshold of occurrences were considered keyword-candidates—a process executed with the AntConc concordancer. The terms connected with natural philosophy were retained as well as two-word collocations consisting of a more general term, such as the Latin *philosophia*, and a more specific one (e.g. *naturalis*). The shortness of the title strings provided us with a number of general terms, such as the Latin *philosophia* and *natura* or the English *body*, which would have generated scraped titles related to branches of philosophy that were not relevant to our study or titles concerned with literary texts, notably to lyric poetry. Therefore, we decided to use two-word collocates that reduced the degree of polysemy of our keywords and provided a more targeted search throughout WorldCat.

To make the extraction of keywords more robust we only used the works that have been annotated as primary. The considerations for not using the secondary and tertiary works were twofold. First, in our topology the primary works are more standard, relatively unproblematic and they more easily lead to identify works pertaining to natural philosophy. Second, the heterogeneity of secondary and tertiary works would have likely produced highly polysemous keywords. Even the primary corpus returned a large amount of highly polysemous words, such as ‘nature’ and ‘body’. Nonetheless, even using our restrictive keyword lists, we found that the expanded corpus includes more secondary and tertiary new titles than primary ones.

Hence, with the threshold set to three occurrences for singular keywords and two occurrences for two-word collocates (except for really small corpora containing only a handful of works, for which we lowered the frequency threshold for single keywords to two) we have extracted lists of keywords. The lists of one-string keywords and collocations obtained were used for focused web-crawling and scraping WorldCat by means of a customized Python script, which collected a total of 86,558 titles published in Latin, English, French, and German between 1600 and 1800. Table 2 illustrates the number of harvested titles per keyword.

---

39Laurence Anthony, AntConc (Version 3.5.8) [Computer Software] (Tokyo, Japan: Waseda University, 2019).
40Natural language processing (NLP) and web-scraping are two of the most popular methods to locate information online and generate corpora (e.g. Kyle L. Anglin, ‘Gather-Narrow-Extract: A Framework for Studying Local Policy Variation Using Web-Scraping and Natural Language Processing’, *Journal of Research on Educational Effectiveness*, 12 (4) (2019), 685–706), but they have been used only recently to generate and expand bibliographies (e.g. L. Lahti, J. Marjanen, H. Roivainen, and M. Tolonen, ‘Bibliographic Data Science and the History of the Book (c. 1500–1800)’, *Cataloging & Classification Quarterly*, 57 (1) (2019), 5–23.
41The searches performed using the keywords in Table 2 were language-restricted for efficiency reasons. A search with ‘natural history’, for instance, in an online world database that uses English meta-data would have returned results in any number of languages. Although language-restricted, our lists never contained titles in one language only. For instance, although we restricted the search for a keyword like *physique* to French, the scraped results included titles both in French and in Latin. Such occurrences clearly point to inconsistencies in library data indexing and management, which were beyond our control.
This large dataset was the starting point of a process of cleaning and distillation aimed at sorting out which of these titles were actually relevant for early modern natural philosophy. This cleaning phase required us to integrate automatic procedures and human annotation: relying on either of these approaches without the other would inevitably have proved to be unpractical (manually annotating 86,558 entries) or unreliable (allowing an automated script to decide which entries are relevant).

Before describing how we integrated this procedure, it is worth commenting on some specific features of the results generated by WorldCat. Since the WorldCat API does not return results by nationality, we had to proceed by language. We did so by considering the four languages (Latin, English, French and German) that make up most of the publications within the four Dictionary-based nationalities we took into account. More importantly, the testing of the script revealed that WorldCat returns the same first 5,000 results in an infinite loop. Because of this shortcoming, we were not able to scrape all the titles returned by searches using the following keywords: (Lat.) naturalis (8,854), (Lat.) physica (10,253), (Lat.) Aristotelis (8,223), (Fr.) physique (8,799), (En.) body (13,670), (En.) nature (27,751). However, that drawback was only apparent, as polysemous keywords like nature and body return an overwhelming number of poetry titles, for instance. Since it was very difficult to make sure that even the first 5,000 results using these keywords alone were related to our object of inquiry, we ran searches in association with a second keyword that promised to reduce the polysemy of the main one, as mentioned before. Thus, we used pairs such as ‘body & mind’, ‘history & nature’, and so on, for which the returned results were still in the thousands.

(Step 2.b) This large dataset was the starting point of a process of cleaning and distillation aimed at sorting out which of these titles were actually relevant for early modern natural philosophy. This cleaning phase required us to integrate automatic procedures and human annotation: relying on either of these approaches without the other would inevitably have proved to be unpractical (manually annotating 86,558 entries) or unreliable (allowing an automated script to decide which entries are relevant).

Before describing how we integrated this procedure, it is worth commenting on some specific features of the results generated by WorldCat. Since the WorldCat API does not return results by nationality, we had to proceed by language. We did so by considering the four languages (Latin, English, French and German) that make up most of the publications within the four Dictionary-based nationalities we took into account. More importantly, the testing of the script revealed that WorldCat returns the same first 5,000 results in an infinite loop. Because of this shortcoming, we were not able to scrape all the titles returned by searches using the following keywords: (Lat.) naturalis (8,854), (Lat.) physica (10,253), (Lat.) Aristotelis (8,223), (Fr.) physique (8,799), (En.) body (13,670), (En.) nature (27,751). However, that drawback was only apparent, as polysemous keywords like nature and body return an overwhelming number of poetry titles, for instance. Since it was very difficult to make sure that even the first 5,000 results using these keywords alone were related to our object of inquiry, we ran searches in association with a second keyword that promised to reduce the polysemy of the main one, as mentioned before. Thus, we used pairs such as ‘body & mind’, ‘history & nature’, and so on, for which the returned results were still in the thousands.

### Table 2. Number of harvested titles per keyword.

<table>
<thead>
<tr>
<th>Latin keywords</th>
<th>Scraped</th>
<th>French keywords</th>
<th>Scraped</th>
<th>English keywords (cont’d)</th>
<th>Scraped</th>
</tr>
</thead>
<tbody>
<tr>
<td>naturalis (&gt;5,000)</td>
<td>5,000</td>
<td>physique (&gt;5,000)</td>
<td>965</td>
<td>motion</td>
<td>5,090</td>
</tr>
<tr>
<td>philosophia naturalis</td>
<td>561</td>
<td>body (&gt;5,000)</td>
<td>450</td>
<td>body &amp; mind</td>
<td></td>
</tr>
<tr>
<td>philosophiae naturalis</td>
<td>1,553</td>
<td>body &amp; mind</td>
<td></td>
<td>body &amp; mind</td>
<td>450</td>
</tr>
<tr>
<td>institutions physicae</td>
<td>321</td>
<td>nature (&gt;5,000)</td>
<td></td>
<td>nature &amp; gravity</td>
<td>48</td>
</tr>
<tr>
<td>philosophia naturae</td>
<td>127</td>
<td>physique (&gt;5,000)</td>
<td>5,000</td>
<td>history &amp; nature</td>
<td>3,780</td>
</tr>
<tr>
<td>peripatetica</td>
<td></td>
<td>morale &amp; physique</td>
<td>946</td>
<td>law &amp; nature</td>
<td>2,950</td>
</tr>
<tr>
<td>physica (&gt;5,000)</td>
<td>5,000</td>
<td>nature &amp; physique</td>
<td>767</td>
<td>nature &amp; motion</td>
<td>373</td>
</tr>
<tr>
<td>physicanaturalis</td>
<td>493</td>
<td>logique &amp; physique</td>
<td>236</td>
<td>nature &amp; gravity</td>
<td>48</td>
</tr>
<tr>
<td>philosophia &amp; physica</td>
<td>627</td>
<td>Descartes</td>
<td>1,196</td>
<td>phenomena &amp; nature</td>
<td>110</td>
</tr>
<tr>
<td>physica &amp; rerum</td>
<td>499</td>
<td></td>
<td>8,145</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>physica&amp;generalis</td>
<td>336</td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Aristotelis (&gt;5000)</td>
<td>5,000</td>
<td>English Keywords</td>
<td>415</td>
<td>German Keywords</td>
<td></td>
</tr>
<tr>
<td>physica</td>
<td>4,991</td>
<td>mechanical philosophy</td>
<td></td>
<td>Naturgeschichte</td>
<td>4,014</td>
</tr>
<tr>
<td>mentis</td>
<td>2,271</td>
<td>laws of motion</td>
<td>523</td>
<td>Naturlehre</td>
<td>1,500</td>
</tr>
<tr>
<td>logicae</td>
<td>3,724</td>
<td>natural philosophy</td>
<td>2,052</td>
<td>Physik</td>
<td>1,743</td>
</tr>
<tr>
<td></td>
<td></td>
<td>experimental philosophy</td>
<td>643</td>
<td></td>
<td></td>
</tr>
<tr>
<td>metaphysicae</td>
<td>2,120</td>
<td>experiments</td>
<td>4,930</td>
<td>Total</td>
<td>11,271</td>
</tr>
<tr>
<td>physicarum</td>
<td>1,459</td>
<td>physics</td>
<td>4,989</td>
<td>TOTAL SCRAPED</td>
<td>86,558</td>
</tr>
<tr>
<td>scientiarum</td>
<td>4,990</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>metaphysicam</td>
<td>1,001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>40,789</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
but much more targeted. As far as the Latin naturalis, physica, and Aristotelis, and for the French physique, only the first 5,000 titles were harvested.

The overall 86,558-line scraped corpus contained a very large number of duplicates, as well as numerous instances of titles not pertaining to the field of natural philosophy due to the semantic loading of several keywords (e.g. motion, which can refer both to movement, and to a formal legislative proposal). The nature of the duplicates was manifold. First, we had to eliminate duplicates of the titles included in the Dictionary-based corpora. Second, duplication appeared in the same language corpus because the searches run in WorldCat do not concern only the titles per se, but also the associated metadata. For instance, the keyword physica returned titles like the Latin P. Ovidii Nasonis Opera Omnia In Tres Tomos Distributa or the German Gesammelte Werke. 40:4:Abteilung III: Materialien und Dokumente, which means not only that the searches return titles that do not contain the keyword of interest, but they also return results in languages other than the keyword’s language. Therefore, the same above-mentioned titles, although not relevant for our own corpus, were very likely to be found not only in the same Latin dataset, but also in the German one, which takes duplication to the next level—duplication between language corpora. Besides the latter example, there was also the problem of translations between pairs of the four languages, as well as issues related to the Latinization of names, of different spellings for the authors’ first and last name, different ways of entering titles (in full or abridged), and so on and so forth. As an aggregator of bibliographic resources across the globe, WorldCat returned a hyper-heterogeneous dataset that had to be deduplicated both computationally and using manual annotators.

Aside from the issue of data redundancy, there was also the problem of data irrelevance. For instance, the scraping process returned 26,353 titles (lines) in English, with a final list of unique titles amounting to only 6,886 lines following the automatic deduplication. Our further manual deduplication lowered the number to 3,345 lines. The latter were subsequently annotated via a three-person annotation process similar to the initial annotation process applied to the Dictionary-based corpora and led to a final corpus of only 336 unique new titles, accounting for only 7.84% of the scraped amount. This rather low figure is certainly not meant to be discouraging, but rather a guarantee for the inclusiveness of the scraping process. However, it is also a measure of the level of dataset redundancy, which would have been impossible to handle manually.

Furthermore, irrelevance was due in great part to the polysemous nature of the keywords, as well as to their relevance to fields other than natural philosophy: medicine, anatomy, and medicine (Physik), natural history and geography (physicall), husbandry (nature, physik), Alchemy/magic (Physik, elements), moral philosophy (natural laws in the moral sense of the term, motion in the political and legal sense), theology (experiences, natural), etc. In the eighteenth-century British corpus, for example, several titles concern
‘experiments’ on particular phenomena (heat, water, phlogiston, electricity). Although ‘experiments’ became an integral part of natural philosophy, the term was often used to describe any type of experimental findings, including those pertaining to medicine, chemistry, husbandry and so on. During the annotation process, three annotators went through the whole list of works (first individually, then discussing together points of disagreement), deciding case by case whether a particular work had potential for being directly linked with the theoretical foundations of natural philosophy and provided important insights on phenomena or new experimental trials (like a whole clutch of works dealing with electricity), or rather was more concerned with medicine, husbandry, various sort of remedies, strictly practical or technological implementations, or just curiosity and divertissement. Only the former kind of works have been retained as relevant for the present research. This type of filtering could only be handled manually.

In terms of actual workflow, we proceeded as follows. The large amount of scraped information was first cleaned, simplified, and tokenized via the pandas, numpy, and NLTK Python libraries. Subsequently, the .csv files created for each keyword and collocation were further consolidated, analyzed and processed to remove the data already present in the three existing Dictionaries. The differences between the cleaned corpora mirror the composition and the availability of these Dictionaries. For example, the small decrease in French titles is due to a relatively short list of known French authors extracted from the only Dictionary available, covering only the seventeenth century (unlike the Dutch and British Dictionaries, which cover the entire two-century time span). The drastic decrease of the corpus in German was due in part to the large number of early modern philosophy works translated into German, as well as to the large number of authors already discovered via the corpus of Latin titles.

The data deduplication was done by substring similarity matching via the FuzzyWuzzy Python library, which assigned similarity scores to authors and titles and removed those with values generally over 60%. The customized script was based on substring similarity calculated using the FuzzyWuzzy library and the Levenshtein distance, a metric that determines the number of edits required to change a one-word string into another. The higher the score, the more similar the strings. After sampling the data multiple times, it was decided that a high value of 60 applied both to the author and the

---

42 For instance, the annotation process dropped titles like: A. M., Rich closet physical secrets collected elaborate paines four several students physick digested together viz. childbearers cabinet preservative plague smallpox physical experiments presented late queen elizabeths hands certain approved medicines (London, 1652); Adolphus Speed, Adam out of Eden: or, an abstract of divers excellent experiments touching the advancement of husbandry (London, 1659); Allan Molines, An account of an experiment of the injection of Mercury into the blood, and its ill effects on the lungs (London, 1691); Arthur Lee, Experiments on the Peruvian bark (London, 1766).

43 Although it is a result of Step 3, explained later on in the paper, for the sake of conciseness the German material is incorporated in Table 3 alongside the other three corpora.
titles will eliminate only perfect or almost perfect duplicates and will keep a sufficiently high number of duplicate titles to prevent data loss during the second step of the automated deduplication process, targeted at repetitive new entries. Table 3 shows the decrease of the harvested corpus after the final automated cleaning.

The remaining data (17,463 lines) were then analyzed using close reading in order to eliminate:

a. any dictionary-based authors not removed during the automated deduplication because of variations in name spelling and Latinization;
b. any works not pertaining or not related to early modern natural philosophy;
c. manuscripts, letters, and notes;
d. false non-duplicates caused by incongruous indexing systems;
e. works originally published significantly before 1600 or after 1800;
f. translations of original works whenever the original was also present; and
g. multiple editions.

The cleaned corpus was broken down into the three working categories (primary, secondary, and tertiary) and, finally, each author entry was annotated for Dictionary-based nationality (namely, we assigned all authors to one of the four geographical areas covered by the existing Dictionaries, gathering the others in a different category of ‘other authors publishing in Latin’). Table 4 shows the impact of human annotation over the composition of the automatically deduplicated lists.

A comparison between Tables 3 and 4 shows that we have used a mild automated cleaning process, which very likely did not lead to data loss, but saved the team a significant amount of time and effort and prepared the data set for a more thorough human annotation stage. The latter stage saw a drastic decrease

| Table 3. Decrease of the corpus following the second automated cleaning stage. |
|-------------------------------|----------------|-------------|-------------|-------------|
|                              | Latin          | English      | French      | German      |
| Consolidated corpus          | 41,122         | 26,353       | 8,145       | 11,271      |
| Corpus after the 1st automated cleaning | 31,353         | 16,913       | 6,108       | 3,944       |
| Corpus after the 2nd automated cleaning | 6,183          | 8,458        | 1,529       | 1,293       |
| Decrease                     | 80.2%          | 49.9%        | 74.9%       | 67.2%       |

| Table 4. Progressive decrease of the corpus during the final manual data cleaning stage. |
|-----------------------------------------------|----------------|-------------|-------------|-------------|
|                                              | Latin          | English      | French      | German      |
| Corpus after the 2nd automated cleaning       | 6,183          | 8,458       | 1,529       | 1,291       | 17,461      |
| Corpus after manual annotation                | 1,067          | 436         | 484         | 212         | 2,199       |
| Decrease                                     | −82.7%         | −94.8%       | −68.3%      | −83.5%      | −83.54%     |
| Corpus after nationality assignment          | 814            | 336         | 288         | 163         | 1601        |
| Decrease                                     | −23.71%        | −22.9%       | −40.4%      | −21.1%      | −27.19%     |
in the number of unique lines, with only 1,175 titles pertaining to new Dutch, French, British, and German authors out of the total number of harvested WorldCat 2,199 single entries, which account for 6.15% of the initial scraped data set. The rest of the scraped corpus of 1,601 titles were 426 titles which were authored by authors of other nationalities, representing less than 2% of the harvested corpus.

(Step 3) The last stage of the new corpus acquisition was three-layered, aiming at three connected goals: (a.) verifying whether the methodology employed in the analysis of the Dutch, French, and British corpora could be successfully applied to the German case; (b.) checking whether the overall newly acquired corpus generated the same keywords used at the beginning (step 2a); (c.) comparing the final list of authors to the four Dictionaries and making sure none of the new authors were indexed (by thus backward checking whether, and to what extent, the initial human-generated lists were incomplete).

The expansion of the German corpus has thus been subject to an inverse investigation process. We extracted a list of authors and titles based on the Dictionary of eighteenth-century German philosophers, which turned out to contain an almost equal number of German and Latin primary titles (16 and 14, respectively). We then compared the manually cleaned results obtained from the scraping of WorldCat for Latin works and we separated all entries that could be associated with German authors. We observed that our reliance on Latin keywords extracted from other corpora is a valid method for corpus expansion because the keywords generated on the grounds of the primary German titles in Latin match the list we generated from the other three Dictionary-based corpora (with philosophiaenaturalis and physica as the most frequent ones). Nevertheless, the process revealed the need to run another search and scrape using German keywords Naturgeschichte, Naturlehre, and Physik, which returned 11,271 more lines (cleaned and reduced to 163 lines cf. Table 4).

The second sub-step (3b) was to repeat the whole procedure in order to verify whether further new relevant authors or titles could be discovered. We derived keyword lists from the whole new corpus generated in the first phase (steps 1.a to 2.b described above) and noticed that two new keyword colocations emerged compared to the departing list: (Lat.) cursus philosophicus and (Fr.) cours de philosophie, which we then used to scrape WorldCat once more, cleaned the results obtained (1,071 lines), and manually annotated the resulting lists. The two keywords returned only five new titles in the French corpus, three titles in the British corpus, and twelve in the German one.

The final sub-step (3c) consisted of back-checking the list of new authors against the existing Dictionaries. The idea of this final check was to ensure that authors and works listed as ‘new’ because they were introduced after the scraping and cleaning procedure (step 2) were not already included in the Dictionaries and accidentally omitted from the initial handmade lists (step 1.a). These omissions would indicate the limits of handmade lists, but in order to identify as ‘new’ the
unique authors and works gathered via the scraping procedure (by thus assessing the gaps in the existing Dictionaries), it would be necessary to back-check that those authors and works were absolutely absent from the Dictionaries themselves. This last annotation revealed that the initial dictionary-based selection process had been quite correct in the case of the Dutch and German dictionaries, as only six authors (five Dutch and one German) were misidentified via the scraping procedure as new results. However, the backchecking further removed 15 French and 51 British authors, originally included in the Dictionaries but not in the principal investigator’s initial lists.

We offer two reflections on this point. First, there seems to be no obvious systematic and overarching reason to explain these omissions, other than the fact that certain authors might not have looked relevant enough to the principal investigator based on the information provided in the Dictionaries (which sometimes do not include full bibliographical information about all the publications by each author). Second, our method proved useful to compensate for this more restrictive and selective human attitude with a more liberal inclusion of all works that show semantic affinity (in terms of titles’ keywords) with those already included. This proves that our method works in a balanced way and, while placing some constraints on what should be accepted or not, it is capable of compensating for the risk of overlooking data that were originally available but that were skipped over during close reading.

The final overall corpus consists of a total of 2,535 titles. Within this corpus, we obtained 1,601 titles from the scraping of WorldCat, which were added to the initial 934 dictionary-based titles. This means that our method expanded by 125% the starting dictionary-based corpus. Within the set of scraped titles, 1,175 can be divided into the four main nationalities covered by the Dictionaries (British, Dutch, French, German). The remaining 426 titles are published by authors who wrote in Latin and who could not be associated with regionally-specific Dictionaries.

4. Authority acknowledgments

A standard picture of the evolution of early modern natural philosophy shows how, at the beginning of the seventeenth century, the field was very much shaped around the authority of Aristotle, canonized in scholastic

44For a more detailed discussion of how our method extends the corpus in the existing Dictionaries and potential limitations that remain, see Andrea Sangiacomo, Raluca Tanasescu, Silvia Donker, and Hugo Hogenbirk, ‘Expanding the Corpus of Early Modern Natural Philosophy: Initial Results and a Review of Available Sources’, Journal of Early Modern Science 2021 forthcoming.

45From now on, we tend to discuss the corpus in terms of numbers of works, rather than authors, since sorting out authors will require a considerable amount of effort and time, given the noisy nature of the harvested data. Often, WorldCat lists the name of authors along other individuals and it makes necessary to check manually each and every person listed and clarify who is the (main) author. Moreover, many works are collective works and the same author may appear in all three categories we designed. For the purposes of this paper, however, the number of titles will be sufficient.
philosophy. During the seventeenth and eighteenth centuries Aristotle’s (scholastics) authority was challenged by a number of other authors. From a broad-brush perspective, two main competitors emerged: Descartes and Newton. Reducing the evolution of natural philosophy to the teleological progression scholastics-Descartes-Newton would surely be an oversimplification. Nonetheless, the attention (and further refinements) that this progression received in contemporary scholarship shows that it does capture some major shifts (at least) taking place in the period.

In this section, we use the collected corpus for two related purposes. First, we study the inventory from the point of view of current scholarly expectations in order to assess the meaningfulness of the derived results for mapping the evolution of early modern natural philosophy (as anticipated in the introduction). In particular, we show that the inventory of our corpus in itself can be studied in such a way as to offer some corroboration of the existing scholarly perception of the period. Second, we illustrate the heuristic potential of this corpus by showing that even the preliminary (computational) study of the corpus metadata leads to new insights and introduces further nuances in the scholarly picture sketched above.

It should be stressed that, for the purposes of this paper, the full corpus of books we inventoried was not available in digitized and readable formats. A significantly large part of these books is open-access and, through appropriate processing, can easily become available. However, working only with bibliographic information for some of the titles still provides valuable insight on their content. The computational interrogation of the available metadata (in our case, the titles of the works, their authors, and details about their publication places) showed that we can practice a valid form of ‘distant reading’ (see note 11).

More specifically, we focus on authority acknowledgements in the titles of the works included in our corpus. Defining a ‘tradition’ is surely a complex historiographical and philosophical problem. Less controversially, however, it can be maintained that, within traditions, certain authors become authorities, namely, acknowledged points of reference for other authors (who both aim at building on them, or rather intend to reject or engage in any other way

---

46See Shapin 1996 (n. 5) and Gaukroger 2006 (n. 8).
48We described the procedure for recovering titles and processing them in such a way to obtain high-quality OCR of the texts in a relatively feasible way in Andrea Sangiacomo, Antonia Karaisl, Nick White, Raluca Tanaseșcu, ‘Reading in the Mist: High-Quality Optical Character Recognition Based on Freely Available Early Modern Digitized Books’, manuscript under review.
with those authorities). As anticipated in the introduction of this paper, there is an important difference between an author and an authority: the authority is the name of an individual author (like ‘Descartes’ or ‘Newton’), or even of a school (like ‘scholastic’ or ‘Aristotelian’) that is used by other authors to engage (either pro, contra, or in any other way) with certain specific approaches, methods, problems, controversies or anything else. When the proper name of an individual author is used by others, and it is acknowledged as an authority, that proper name becomes a sort of flag and acquires a new range of meanings and implications, which depend on how others use that proper name for their purposes. Authorities are crucial in teaching practices, since they provide the primary sources on which the teaching of the discipline can be based and offer raw models and standards about its best instances. Our corpus is designed to capture how natural philosophy was taught and it can thus provide an ideal scenario to trace patterns in authority acknowledgement across time.

We thus studied the trends of authority acknowledgments by focusing on these three main categories: (i) scholastic, (ii) Cartesian, and (iii) Newtonian.\(^49\) Although none of these authorities are homogenous in themselves, for simplicity’s sake, we refer to them as individual labels. We focus on these three authorities because they are included in 23.5% of the titles in our overall corpus and thus constitute a phenomenon robust enough to deserve investigation.\(^50\)

It should be stressed that existing scholarship has not paid specific attention to authority acknowledgment as we just defined it. Our focus on authority acknowledgments is thus a way of tackling from a different (distant) perspective what existing scholarship has established by way of analysis of specific case studies and close reading of several individual texts. On the one hand, our investigation of authority acknowledgments aims to test the main results established by existing scholarship by providing new data in support of them (by thus also showing that the study of authority acknowledgments does pick up on a genuine phenomenon, which is expected to be encountered based on the findings of existing scholarship). On the other hand, we also expect that the study of authority acknowledgments will yield insights that were not discernible by using the traditional methods; this expectation is based on the different perspective and approach that investigating authority acknowledgments requires compared to the more

\(^{49}\)To discern these categories, we manually annotated all works that included in their titles the keywords: ‘Aristotle’, ‘Peripatetic’, ‘scholastic’, ‘Scotus’, ‘S. Thomas Aquinas’, ‘Octos Libros Physicae’, (all these keywords are included in the overall category ‘scholastic’); ‘Descartes’, ‘Cartesian’ (category ‘Cartesian’), ‘Newton’, ‘Newtonian’ (category ‘Newtonian’) and their immediately related declensions.

\(^{50}\)Besides authorities associated with the three that we study, we counted another 44 proper names of authors mentioned in titles. However, the largest majority of these is mentioned only in one title, while just a few are mentioned more often (Kant, Wolf, and Gassendi in 3 titles each, Galilei in 5 titles, Leibniz in 7), accounting for only 0.8% titles in our corpus.
traditional study of early modern materials by means of close reading and individual case studies.

In order to operationalize our investigation, we study the presence of authority acknowledgements from two perspectives. First, we focus on the way in which titles in our corpus include explicit mentions of any of the three above-mentioned authorities, by thus acknowledging that the work is going to engage with any of them as a primal focus or context for its discussion. Of course, authority acknowledgements do not have to be stated in titles. However, titles are relevant ways of introducing a book to a readership. We take this fact into account, by assuming that titles including authority acknowledgements represent a particularly explicit and emphatic way of acknowledging authorities. It must be emphasized that we do not study whether each title aims at engaging either pro, contra or in any other way with the authority mentioned. We simply observe whether titles include explicit mentions of authorities and, if they do, which authorities are mentioned. If there was absolutely nothing special about authorities mentioned in titles, we would expect to find a random distribution of authorities. However, as shown below, this is not what we observe in our corpus. The majority of titles do not include authority acknowledgements (we shall further reflect on this point in Section 5), but when they do, they most often fall into one of our three categories.

Second, we also take into account the way in which today’s scholars have used the same three kinds of authorities as holistic categories to identify the general orientation of certain authors or works. This implicit form of authority acknowledgement is the result of the interpretation provided in the Dictionaries (i.e. a scholar writing an entry on a particular author associates the author’s work with a core authority or school). Since the Dictionaries do not cover homogeneously the whole spatial and temporal landscape of our corpus, this information remains inevitably incomplete (with respect to seventeenth-century Germany and eighteenth-century France, for instance). Nonetheless, being aware of this limitation, we decided to take into account what it could add to our understanding of the dynamics discernible in our corpus.

We begin by noticing how authority acknowledgement is instantiated with respect to the threefold categorization of the corpus. Scholastic authority is present in all the three categories, with a preponderance of primary and secondary texts; a similar pattern can be discerned for the acknowledgements of Newtonian authority (although the number of works is less); Cartesian authority unfolds instead mostly via secondary and tertiary works. This fact might be interpreted as reflecting the relationship that these different trends had with the academic milieu. As it results from existing scholarship, scholastic philosophy was predominantly practiced within universities, which is then reflected in the kind of works that most often acknowledge it (primary and secondary works being the most certainly connected with teaching practices in academia). The Cartesian trend attempted to penetrate academic institutions and oppose
the scholastic trend, but without ever fully succeeding in this goal (at least, compared with the other trends). Most Cartesian authority acknowledgement remains more peripheral with respect to the teaching of natural philosophy (it rarely became the standard for textbooks and other systematic treatises), which might contribute to explain the relative modest success that it had in the long run. Newtonian authority acknowledgement followed a strategy more similar to the scholastic one, by penetrating significantly in primary and secondary works. Taking root at the centre of the academic milieu arguably contributed to the consolidation of the Newtonian trend over time.

These considerations can be further refined by considering how the trends in authority acknowledgement look like when plotted against time, as illustrated in Figures 2–5, which deal with each of the four national regions distinguished in our collected corpus.51

These figures invite two major remarks. The first remark is that our data somehow corroborate what was expected, namely, the presence of the three authorities discussed by existing scholarship as a matter of active authority acknowledgement within the corpus. To some extent, the chronological and geographical distribution of these works also corroborate the expected scenario (embedded in the canonical narrative about the scientific revolution) according to which the Cartesian trend in natural philosophy would remain mostly confined within the second half of the seventeenth century, spreading especially in France and the Dutch Republic, while the Newtonian trend would consolidate during the eighteenth century, enjoying a larger diffusion outside of the British Islands, including all the other major continental countries.

The second remark is that the same data show also something new concerning the patterns of authority acknowledgement in different regions. Let us first provide a synoptic description of the main situation revealed by our visualizations. Consider Figures 2 and 3, which are based on complete Dictionaries (and hence include complete information about implicit authority acknowledgement). In the Dutch context, scholastic works are well established but rarely explicit in their authority acknowledgement. This point is mirrored in the British context, although here the presence of scholastic works is visibly thinner. The Dutch context then shows a quite neat articulation of a Cartesian phase in the middle of the seventeenth century and then a Newtonian phase in the eighteenth century. These phases succeed one another and do not have much overlap, except for a certain resilience of the scholastic production up to the mid-eighteenth century. By comparison, the British context presents a fuzzier picture in the seventeenth century, but a similar trend. There is a shorter Cartesian phase, with some more attempts of combining scholastic

51In Figures 2 to 7, implicit authority acknowledgement based on the information included in the Dictionaries is always at .8 on the y axis and it is signalled in the legend by putting the relevant authority between brackets. Numbers on the y axis are conventional and used to discern the different acknowledgments. The titles of the following figures refer to Dictionary-based nationalities as defined above, not to languages.
and Cartesian authorities than in the Dutch context. This Cartesian trend, however, ends quickly when the Newtonian trend emerges. Unsurprisingly, in the British context the Newtonian trend seems to become the most conspicuous with respect to both implicit and explicit authority acknowledgements. Figure 4 shows a somehow different pattern in the French context. Here, the scholastic trend is even stronger than in the other two areas and it survives for almost the entire period. A short Newtonian trend emerges around the mid-eighteenth century (although the information here about implicit authority acknowledgement is incomplete), showing that Newton was arguably
perceived as an outsider. The Cartesian trend is instead very momentous, especially if one takes into account all the explicit and implicit hybridisations that it takes with both scholastic and Newtonian trends. It is due to these hybridisations that the Cartesian trend can survive almost an entire century (between the mid-seventeenth to mid-eighteenth century). Figure 5, concerning the German area, shows yet a different pattern. Here the scholastic phase is again the most consistent and long-lived (even considering the lack of

**Figure 4.** Explicit and implicit authority acknowledgement in the French expanded corpus.

**Figure 5.** Explicit and implicit authority acknowledgement in the German expanded corpus.
information about implicit authority acknowledgement for the seventeenth century). The Cartesian trend is almost absent as such, but implicit authority acknowledgement of Cartesian-Newtonian works seems to pave the way to the emergence of a more focused Newtonian trend in the early eighteenth century (more or less synchronically with the beginning of the same Newtonian trend in the Dutch context).

This situation suggests interesting aspects that distinguish the strategies that can be associated with Newtonian, Cartesian and scholastic authority acknowledgements.

a) The Newtonian trend is centred in the British context, where it becomes clearly the dominant one. It spreads in the Dutch and German contexts as well, almost at the same time, penetrating in France relatively later in the eighteenth century. This might reflect both political (French-English rivalries) and religious (Protestant-Catholic divide) circumstances. Be that as it may, the Newtonian trend is open for confrontation with the Cartesian, but it never engages directly with the scholastic trend, although our data also show that the scholastic trend was never quite extinguished even in the first half of the eighteenth century and ran in parallel with the Newtonian trend for some time. This might suggest that the Newtonian trend used a decisively different strategy in its authority acknowledgement, namely, avoiding any hybridization with its scholastic competitor. This remark is consistent with the tendencies (noted above) of both Newtonian and scholastic authority acknowledgements to be instantiated mostly in primary works, meaning that they could establish themselves at the centre of the academic milieu. Despite the presence of some Cartesian-Newtonian combinations of authorities, one can see that the real competition was between the surviving scholastic trend and the new emerging Newtonian trend.

b) The Cartesian trend spreads mostly in the French, Dutch, and British contexts, but only in the French context does it remain consistently present for a relatively long period. This might be due to the higher degree of hybridization of French Cartesian authority acknowledgement, which seems to be more open to involve (or enter in dialogue with) other authorities as well. Consistent with this remark, we noted above that Cartesian authority acknowledgement is most present in secondary and tertiary works. These works occupy the buffer and periphery of the academic milieu, are more flexible in their orientation, and can involve a different group of social actors, less established in the university institutional context (hence with a much smaller monopoly on the primary works). This greater power of hybridization, however, might suggest also a reason for the later disappearance of the Cartesian trend. Greater hybridization entails smaller homogeneity. Existing scholarship acknowledges that ‘Cartesianism’ was
indeed an umbrella term for a number of different approaches. In terms of authority acknowledgement, this has some payoff, since it allows Cartesian authors to be more flexible in their way of competing with other rivals, but it has the cost of making the whole trend less consistent and thus less effective in pushing its own distinctive brand (since it will become unclear at some point what it means to be a Cartesian). Of course, also Newtonians were far from being a homogeneous group, but in terms of authority acknowledgement we observe a tendency to use Newton’s name as a kind of flagship common denominator, which provided a clear-cut alternative to both Cartesians and scholastics. One can thus discern two very different authority acknowledgement strategies among Cartesians and Newtonians, which both have to do with how they position themselves with respect to other rival trends, and how they try to occupy the different regions of the academic milieu.

c) Our data show something remarkable (and most often overlooked by existing scholarship) concerning scholastic authority acknowledgement. Far from being confined within the first half of the seventeenth century (in the pre-Cartesian phase), scholastic authority acknowledgement spans almost throughout the whole early modern period and runs in parallel with Cartesian and Newtonian trends. This immediately suggests that scholastic authority acknowledgement had a reach that cannot be confined or explained in terms of religious or political divides (e.g. among Catholic and Protestant countries), but truly had a pan-European dimension.

This trend might be partially explained by the fact that, despite increasing polemics against scholastic approaches to the study of natural phenomena, other sections of the scholastic curriculum (like logic) did remain influential. In many textbooks, scholastic natural philosophy is presented together with other disciplines (including logic) in the same multi-volume work by the same author. Hence, the relative resilience of some parts of these works might justify a more inertial survival of the scholastic approach to natural philosophy proper. In general, this is an interesting hypothesis that would need to be explored in detail by comparing the contents of the books in our corpus and by examining the way in which discussions pertaining to potentially different fields influenced each other. However, this is beyond the scope of this current paper. Nevertheless, we can notice that even if there were an inertial component in the survival of scholastic natural philosophy proper, it would not necessarily rule out a more active and deliberate effort by several authors to advocate for the scholastic approach as valuable in its own right.

52See Ariew 2014 and Schmaltz 2016 (note 47).
The fact that scholastic authority acknowledgement is the most instantiated suggests that scholastic ideas were far from being established as a default, but they were rather in need of constant re-acknowledgment. A well-established tradition that is not challenged or under special pressure does not need to reaffirm its authorities very forcefully, since these authorities will be taken for granted. However, our data show that scholastics were not less involved in the business of authority acknowledgement than Cartesians or Newtonians. Keep in mind that our corpus is composed of unique titles published for the first time during the early modern period (no reprint, translations or multiple successive editions of the same work are included). The kind of scholastic authority acknowledgement we are dealing with, thus, reflect a whole wave of new scholastic titles that were produced ex novo and published at the time, explicitly championing for scholastic authorities.

One might wonder whether the long-lasting and flourishing production of scholastic works was not simply due to economic and intellectual reasons. It could be argued that in a number of relatively more traditional universities, natural philosophy courses remained shaped around the scholastic curriculum, which created a demand for new titles to be published. In response, we stress that shedding light on this issue would necessarily entail expanding the current research significantly to consider not only the actual content of our corpus but also its socio-economical context. But even so, the data presented in this paper suggest that it is at least plausible to surmise that something more than mere market demand drove scholastic authority acknowledgements. Market demands could have been satisfied by any new title in ‘natural philosophy’ and, in absence of serious competitors, it would have been redundant to acknowledge authority in such traditional titles. In fact, almost three quarters of our corpus does not make use of authority acknowledgments in titles, suggesting that some of these works might indeed be a response to an economic demand. But why then making the extra effort of introducing authority acknowledgment in otherwise traditional books? What is the added benefit of this strategy? Since authorities are potentially divisive, mere economic gains become very unlikely, since a reference to Aristotle may captivate scholastically-minded buyers, but may otherwise repel others.

If these scholastic authorities had not been under serious pressure to retain their role, there would have been no need for such a momentous production of titles explicitly acknowledging them (new editions of old and well-established titles would have been probably enough). Existing scholarship often considers scholastic philosophy as the default established tradition against which Cartesian and Newtonian trends reacted. It has been observed (based on case studies) that early modern scholastic texts often engage with contemporary sources and try to deal with problems raised by their non-scholastic

---

53See Valleriani et al. 2019 (n. 12).
opponents. Our data provide further context for this observation, by showing that, whatever the established tradition was, the scholastics were competitors for remaining established (at best), existing against and in parallel with new Cartesian and Newtonian trends.

This point is confirmed by Figure 6, which presents the way in which authority acknowledgement features in the Latin works published outside of the four main geographical areas considered so far.

It is apparent that scholastic authority acknowledgement retains almost a monopoly over the whole sub-corpus we collected. This does not entail that scholastic philosophy can be regarded here as the default and obvious philosophical outlook of natural philosophy. Nevertheless, we can see that scholastics were by far the most vocal approach (in terms of authority acknowledgement) outside of the Central-Northern areas (Britain, Dutch Republic, France, Germany) considered so far. The reasoning introduced above (a well-established tradition does not need to reinforce its authorities with new publications) is still valid, even if in the group of ‘other nationalities’ we do not observe any significant presence of Cartesian or Newtonian acknowledgments. This can be interpreted as a sign of the fact that, in the broad geographical domain outside of the four Central-Northern European areas considered above, scholastics were indeed successful in preserving a form of monopoly (at least insofar as Latin publications are concerned). Being successful in a competition does not entail no competition, but simply that the impact of competitors is significantly mitigated.

Having presented results for authority acknowledgments in each sub-corpus, Figure 7 now presents how they appear when the whole corpus is considered together.

Figure 7 confirms several points that have emerged above: scholastic authority acknowledgment tends to be continuous throughout the period, until the second half of the eighteenth century. Cartesian authority acknowledgment is more open to hybridization than Newtonian authority acknowledgment, and relatively more engaged with scholastic authorities than with Newtonian authorities. A strict Cartesian trend does not survive after the mid-eighteenth century, while the Newtonian trend outlives it.

Figure 7 also invites one further consideration: until the second half of the eighteenth century, the continuity and intensity of scholastic authority acknowledgment remains relatively unaffected by the emergence of the new trends of Cartesian and Newtonian authority acknowledgment. If we interpret

54 See Feingold 2003 ed. (n. 47).
55 One might hedge this remark by pointing out that our ‘scholastic’ categories are relatively heterogeneous since they include references to Aristotle, but also other authorities such as Aquinas and Scotus. However, this simply shows that Scholastics themselves were internally divided in many orientations (as is well acknowledged in existing scholarship), just like Cartesians and Newtonians were. The point of advocacy is not really about internal divisions within a certain ideological trend but rather in the competition with other trends that are perceived and presented as genuinely different (despite attempts of bringing them into dialogue).
authority acknowledgment as the symptom of the need of reaffirming the visibility of certain authorities connected with certain traditions or approaches, then the figure also suggests that in the second half of the eighteenth century, while the Newtonian trend is still consolidating itself, the scholastic trend dies out. We suggest that this observation should be interpreted in terms of relevance for the field of study (natural philosophy), rather than in terms of actual vitality of a certain tradition in itself. Let us explain.

Two waves of alternative approaches (Cartesian and Newtonian) did not manage to weaken the continuity and strength of the scholastic trend. Hence,
the fading away of the scholastic trend shown in Figure 7 can hardly be explained by appealing to (i) a ‘defeat’ of the scholastics or (ii) the fact that scholastic thought simply ended in the late eighteenth century. Rather, Figure 7 should be interpreted as a radar monitor, and authority acknowledgments as the signal that a certain tradition is actively engaged in the domain of natural philosophy. From this perspective, the fading away of the scholastic trend signals that scholastic natural philosophy ceases to be actively engaged in the debate on natural philosophy, and this suggests that (by consequence) it might have also ceased to be actually relevant for it. This latter point should, of course, be confirmed by an actual investigation of the works included in the corpus. Nonetheless, as a working hypothesis, it indicates that what ruled out scholastic approaches from natural philosophy was not necessarily any particular new approach or ideology (say Cartesian or Newtonian) that emerged during the seventeenth or even the eighteenth century, but rather a broader change of the discipline, such that engaging with scholastic authorities would no longer have been considered relevant for natural philosophy as such.

To put this hypothesis in more evocative (ecological) terms: the extinction of the scholastics was not due to the fact that their predators did away with them, but rather to the fact that the environment in which they operated changed so much that they had to leave. This hypothesis is worth considering because it broadens the focus of attention from the specific arguments and topics that have been the matter of the disagreement between scholastic and anti-scholastic voices, to the way in which the academic milieu itself changed meanwhile. Of course, particular debates and disagreements on methods and topics might well have contributed to this change, but arguably this was not the only factor involved. Hence, our hypothesis calls for a renewed effort in integrating historical, cultural, and political considerations within the study of the evolution of early modern natural philosophy.

The case of religious influences and policies that might have impacted the reshaping of early modern natural philosophy should be carefully handled. One might jump to the conclusion that since scholastic philosophy might be the most directly connected with religious agendas and authorities, it might also be the one that receives the greatest support. This, however, is neither obvious nor apparent from our data. To explore a little more this point, we decided to take into account the places of publication of all the works included in our collected corpus and check what are the trends that they reveal.

The map in Figure 8 shows the overall distribution of the publication places (the size of dots is proportional to the number of works published there).

Although not all publication places have been determined (10% are still unknown at the time of writing), over 75% of the works acknowledging the place were published in the author’s own country. The publication hubs can be found in the expected cities: Paris for France, Leiden for the Netherlands, London for the United Kingdom, and Leipzig for Germany. The frequency
distribution of places shows a typical long-tail: only a few cities appear many times, and most cities appear only once or twice. There are 191 publication places in total, of which 70 only appear once, while the top five cities cover half of the corpus.

Each region also shows some typical features. At one end of the spectrum, German publications appear to be the most scattered among the largest number of cities, which can partly be explained by the political division of the region. At the other end, works by British philosophers are hubbed in London, as shown by Figure 8, with little representation outside the capital. Also, at least three quarters of British, Dutch and German works (76% or more) are published in the home country of the authors. An exception are the works included in the French corpus: only half (55%) of them are published in French cities. The Dutch Republic shows an equally intense publishing activity. Leiden is by far the most important publication place for the Dutch authors, while Amsterdam is the preferred Dutch place of publication for authors of other nationalities. Dutch authors are the ones publishing abroad the least (11%), most often in German cities (4%).

This overview of publication places can be further enriched by taking into account the works that include authority acknowledgments (which make about a quarter of the whole corpus). Figure 9 provides a geo-localisation of the publication places of these works only.

In terms of places, Figure 9 confirms the trend already noticed with regards to the few active European publication hubs. There are however two very different trends concerning where works with authority acknowledgments are published. Works acknowledging Newtonian and Cartesian authorities are hosted mainly in capital cities, London and Paris respectively, with

---

**Figure 8.** Places of publication: complete Dutch, French, English, and German Corpus.
Newtonians reaching to the Dutch Republic (Leiden) and to German cities (Göttingen), and Cartesians heading to Dutch cities (Amsterdam and Franeker). This is not observed with respect to works including scholastic authorities, which appear significantly less geographically polarized and rather more scattered throughout European cities, without any clear publication hub emerging for them.

These two different publication trends raise a final consideration: while Newtonian and Cartesian trends might be somehow linked to more specific national dynamics or more clearly recognized as authorities representative of a national ethos (Newton as a flag of British ingenuity, Descartes as a flag of French sagacity), the scholastics belong more to a pan-European (pre-modern) structure, such as the university network which they monopolized for a long time. When it comes to natural philosophy, scholastics cannot rely on a central hub. As noticed above, scholastics might have direct connections with religious infrastructures and power dynamics. However, as these religious infrastructures are transformed and become increasingly more fragmented during early modern European history, this might have had an impact also on the survival of scholastic natural philosophy, which might have been increasingly less capable of being supported by unified, rampant and nationally based forces. This interplay between dispersal and polarization might be a further factor that contributed to the fading away of the scholastic trend, namely, the lack of a strong politically and nationally grounded power structure that could fight to support their dominance in a certain region. While this hypothesis would deserve further study and attention, which would go far beyond the scope of this present paper, it shows the validity of the method we used to generate our corpus and the heuristic value of investigating authority acknowledgments.
5. Conclusions

In this paper, we proposed a method to combine different resources available (the four existing Dictionaries of early modern philosophers and the WorldCat) and produce a sufficiently representative corpus for mapping the evolution of how natural philosophy was discussed and presented. Our customized method thus yielded an inventory of potentially relevant works, which we explored in greater detail in order to assess its meaningfulness. In particular, we investigated the trends in authority acknowledgments linked to three different and competing authorities that are the most instantiated in our corpus: scholastic, Cartesian and Newtonian. This process is much more complex than what it looks like when considering only isolated cases. Scholastic, Cartesian, and Newtonian authorities are acknowledged by using different strategies. We observe a remarkable endurance and spreading of scholastic authorities well beyond the chronological and geographical limits that are most commonly discussed.

Our data suggest two agenda points for future research. First, it would be surely most important to investigate the complex reasons (including political, religious, economic, and social) that might have shaped the fate of authority acknowledgements. At present, our data point for instance to a significant asymmetry in the way in which scholastic authorities fared in different European regions. In each case, they are a serious competitor, but outside of the Central-Northern European areas they seem also capable of preserving the monopoly of authority acknowledgement. Second, having discerned the trends in authority acknowledgement, one might further investigate what its impact was on the actual diffusion of those ideas associated with different authorities. Discerning this impact requires on to delve into the mass of the works that do not make any explicit authority acknowledgement connected with the three we discussed. This group of works is the place to find the existing established traditions in early modern philosophy (if any, and whichever they might be). One should expect to see some proportionality (direct or inverse) between the intensity and spread of acknowledgement of certain authorities and how they are received, integrated and dealt with in works that do not explicitly refer to them in their titles. Investigating this relationship, however, will require considerable efforts and falls outside the scope of the present study. Our results also call for integrating semantic analysis in order to fully explore the implications of what we discussed so far and test any further hypothesis or intuition. Our focus on authority acknowledgments has only landed on the tip of a textual iceberg, and yet it is already revealing some key features of that iceberg.

Disclosure statement

No potential conflict of interest was reported by the author(s).
Funding

This article is part of a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme [grant number 801653].

ORCID

Andrea Sangiacomo http://orcid.org/0000-0001-7529-9826
Ralouca Tanasescu http://orcid.org/0000-0002-2789-5145