Towards a Computational History of Ideas

Betti, A.; van den Berg, H.

Published in:
CEUR Workshop Proceedings

Towards a Computational History of Ideas

Arianna Betti & Hein van den Berg

University of Amsterdam & VU Amsterdam

ariannabetti@gmail.com, hein.van.den.berg2@gmail.com

Abstract

The History of Ideas is presently enjoying a certain renaissance after a long period of disrepute. Increasing quantities of digitally available historical texts and the availability of computational tools for the exploration of such masses of sources, it is suggested, can be of invaluable help to historians of ideas. The question is: how exactly? In this paper, we argue that a computational history of ideas is possible if the following two conditions are satisfied: (i) *Sound Method*. A computational history of ideas must be built upon a sound theoretical foundation for its methodology, and the only such foundation is given by the use of *models*, i.e., fully explicit and revisable interpretive frameworks or networks of concepts developed by the historians of ideas themselves. (ii) *Data Organisation.* Interpretive models in our sense must be seen as topic-specific *k*nowledge *o*rganisation *s*ystems (KOS) implementable (i.e. formalisable) as e.g. computer science ontologies. We thus require historians of ideas to provide *explicitly structured* semantic framing of domain knowledge *before* investigating texts computationally, and to constantly re-input findings from the interpretive point of view. In this way, a computational history of ideas maximally profits from computer methods while also keeping humanities experts in the loop. We elucidate our proposal with reference to a model of the notion of axiomatic science in 18^{th}-19^{th} century Europe.

1 The Problem(s)

There is a conceptual, interdisciplinary approach to intellectual history, according to which it is the historian's task to trace the history of "certain 'unit-ideas' as they find expression in a wide range of cultural fields from philosophic systems to literature, the other arts, the sciences, and social thought" (Macksey 2003: 1084). This approach, the 'History of Ideas', largely founded by Arthur O. Lovejoy in the early twentieth century, and associated with his *The Great Chain of Being* (1936), is presently enjoying a certain renaissance after a long period of disrepute. To be more precise, one central presupposition of the History of Ideas is enjoying such a renaissance, namely the tenet that it is legitimate to conduct historical investigations spanning long stretches of time of "central concepts in our political, ethical and scientific vocabularies" (Armitage 2012: 497, cf. the so-called 'longue durée history'; cf. also McMahon 2013).

There is more than a passing indication that one of the reasons why 'big history' studies are witnessing a new legitimacy is the growing availability of massive amounts of digital textual sources, as well as the growing availability of computational tools and best practices for the exploration of such masses of sources (cf. e.g. again Armitage 2012: 506-7; Armitage & Guldi 2014; Armitage & Guldi 2015: 291). Important examples of the latter are Franco Moretti's powerful experiments with, and reflections on, 'distant reading' (see e.g. Moretti 2013), and recent contributions somewhat nearer to intellectual history (e.g. Michel et al. 2011). Increasing quantities of digitally available historical texts, it is suggested, can be of invaluable help to historians of ideas. The question is: how exactly?

As practitioners in the field ourselves, we consider these new developments extremely valuable and exciting. But to what extent are they suitable to enact a real computational turn in the history of ideas and related fields? We say that there is an important sense in which the particular kind of computational practices mentioned above, however valuable, contrasts strikingly with the reality of research in the history of ideas and related fields such as history of science. Studies such as Michel et al. 2011 apply generic computer methods, simple n-gram analyses (i.e. detection of sequences of a number n of words), and shallow Natural Language Processing (NLP) tools to historical textual material. Yet researchers in the history of (philosophical or scientific) ideas typically apply painstakingly fine-grained analyses to diverse textual material of extremely high conceptual density. The contrast between computational practice and scholarly reality in the humanities forms thus an easy target for skeptics towards digital humanities. Skeptics will see two problems.

> *Unorganised Data.* Having generic, simple and shallow bottom-up analysis of a massive amount of diverse, 'long' and complex data cannot help in reconstructing the history of an idea such as *science* through centuries. Feeding a computer masses of diverse and complex texts can only yield masses of unorganized details, at least in this field.
>
> *Faulty Method*. Even if it were possible to make sense of such data within a computational enterprise, there are fundamental problems with the very method of the history of ideas. The notion of *one idea* traceable through centuries of thought is illusory: for ideas cannot be studied in isolation from their context, and their meaning is in constant flux (Skinner 2002).

Can these two problems be overcome? In other words, is creating the new field of the computational history of ideas possible? Most of all, what should its *method* be? Our question is not whether it is valuable to apply distant reading in order to trace 'big ideas' (it is), but whether there is a way to scale up the mix of close reading, conceptual analysis and historical contextualisation typical of the historian of ideas. We maintain that what is needed to create a computational history of ideas is certainly more than distant reading, and in this paper we lay out what we take that more to be.

In a nutshell, we argue that a computational history of ideas is possible on condition that the two problems above are tackled as follows.

> *Sound Method*. A computational history of ideas must be built upon a sound theoretical foundation for its methodology, and the only such foundation is given by the use of *models*, i.e., fully explicit and revisable interpretive frameworks or networks of concepts developed by the historians of ideas themselves. This requires construing ideas as complex relational frameworks (models) that combine both stable parts (*continuities*) and variable parts (*discontinuities*) (Betti 2014: 14; Betti and van den Berg 2014.)
>
> *Data Organisation.* Interpretive models in our sense must be seen as topic-specific *k*nowledge *o*rganisation *s*ystems (KOS) implementable (i.e. formalisable) as e.g. computer science ontologies. We thus require historians of ideas to provide *explicitly structured* semantic framing of domain knowledge *before* investigating texts computationally, and to constantly re-input findings from the interpretive point of view. In this way, a computational history of ideas maximally profits from computer methods while also keeping humanities experts in the loop.

Here below we elucidate our proposal with reference to a model (in our sense) of the notion of axiomatic science in 18th-19th century Europe. In the first two sections we discuss and give a solution to the problem of *Faulty Method*. Here we draw on our previous Betti and van den Berg 2014 to make the present paper self-contained for a digital humanities audience.[1] In the rest of the paper we tackle the problem of *Unorganised Data* and propose a way-out.

Three remarks before we start are in order. First, the two problems just mentioned are both methodological, and importantly related, for without solving the second (Faulty Method) there can be no hope of solving the first (Unorganised Data);[2] still, while the second problem is more of an ideological one, and requires no reflection on new computational tools to be tackled, the first one does require such reflection. Second, our proposal as to the second problem comes with a neutral, non-ideological philosophical stance on what ideas are: we are neutral on the *ontological* or *metaphysical* nature of ideas, and we take this neutrality to be a fundamental presupposition for a successful and workable operational proposal. Third, we should say right away that the word 'model' might be confusing to the reader, since in computational contexts what is meant by 'model' is normally a *computational* model. What we mean by 'models' are instead *non-computational* conceptualisations within a (pure, non-computational) humanities domain; in this sense, models are technical artefacts. The reader might most fruitfully identify our models with very complex, articulated working definitions of a concept, conceptual tools that are provided by historians of ideas themselves in order to, say, 'chart their landscape' to their own interpretive goals. In principle, models as we intend them are (to be) used and constructed within history of ideas without any reference to a computational enterprise.[3] This said, in this paper our point is specifically the following. The only kind of history of ideas that can successfully take a computational turn must rely on models in our sense for its own methodology; if historians of ideas wish to take a computational turn, they must turn the knowledge they possess about the ideas they want to study into models of those ideas in our sense *before* any computational exploration begins.[4]

---

1 Portions of section 2 and 3 have been previously published in Betti, Arianna & van den Berg, Hein 2014, Modelling The History of Ideas, *British Journal for the History of Philosophy* 22 (4), 812-835. Published online 28 Aug 2014, copyright © BSHP. Reprinted by permission of the publisher (Taylor & Francis Ltd, www.tandfonline.com, on behalf of BSHP). The authors paid 64.80 USD to reuse material of their original article.

2 This point in particular distinguishes our approach from the one assumed in Kenter et al. 2015 (which nevertheless mentions Betti & van den Berg 2014 among previous work in history of ideas); another important point of difference is the use of non-technical corpus (newspapers) and a bottom-up approach, while the present paper regards technical corpora and avows an ontology-approach. See also section 4, point II below.

3 McCarty 2008c's personification characteristics (a theoretical model of personification), or Armitage 2012's analysis of three definitions of civil war might be seen as approximating raw material to build a model in our sense. Very few reflections are available on the use of models in this sense in history. We follow Leff in maintaining that this scarcity should not be taken as signifying that models have no place in history, but rather as a sign that the use of models by historians is largely *implicit*: 'a historian could hardly put pen to paper without having an implicit model of what he was studying', Leff 1972: 148. Cf. also Finley 1986's discussion of Max Weber's notion of *ideal types* in conjunction with modeling in history. We owe the reference to Leff and Finley to McCarty 2004. In turn, we owe reference to the work of Willard McCarty to a helpful anonymous referee.

4 It is true that, as McCarty (2008a: 15) points out, "[b]y inducing us to model our heretofore largely tacit methods, [the gift of computing to the humanities, ab&hvdb] invites us to look backwards to what we have done and forwards to what we can imagine with it". We fully agree that if our enterprise is a computational one, we cannot (anymore) afford to leave our methods implicit. However, our tenet is the (possibly stronger) one that explicit modeling in the history of ideas in our sense is a precondition for doing history of ideas properly *tout court*, i.e. also independently of a computational enterprise (cf. Betti & van den Berg 2014).

2 A Faulty Method? Lovejoyan History of Ideas

Lovejoy characterized the history of ideas as being concerned with *unit-ideas*, simple entities that retain their meaning through time. The historian must isolate certain unit-ideas (e.g. the idea of *truth*) and trace their history in various contexts, that is, periods and intellectual settings (Lovejoy 1936: 3-7, 15). Lovejoy's approach has been subject to a number of points of criticism, mainly directed against the notion of a unit-idea. Critics of Lovejoy have denied the existence of unchanging unit-ideas; they deny, in particular, that ideas retain their meaning over time and across contexts. This kind of criticism has been taken to undermine the tenet that Lovejoy's method is a proper historical method. In the literature, we can distinguish four main points of criticism (note that we use 'ideas' and 'concepts' interchangeably):

a) *Holism.* The meaning of concepts is contextual and always in flux. Hence, the notion of *one* unit-idea traceable through time is illusory.

b) *Conceptual Change*. If the meaning of concepts is fixed, how can unit-ideas have a *history*?

c) *Scope*. It is difficult to define the scope of unit-ideas. If, for example, we want to write a history of the concept of force in the modern period, we must define it in a way that allows us to attribute this concept to both Descartes and Newton. However, a narrowly Newtonian concept of force cannot be attributed to Descartes, and if we define the concept too broadly we will miss important differences between historical actors.

d) *Arbitrariness and biases*. Lovejoy's method is (i) arbitrary, because it abstracts from the intentions of authors and hence does not yield a proper understanding of historical texts, and (ii) biased, because it interprets these texts by necessarily applying preconceived schemas.

Defenders of Lovejoy have formulated promising counterarguments against (b) and (c). Defenders, in particular Jouni-Matti Kuukkanen, give up the tenet that ideas are *simple*: ideas can change exactly because they are *complex*. Ideas have parts (or subconcepts), and what changes are an idea's parts. Some parts of an idea belong to its stable *core* (the features of the concepts that persist over some amount of time), whereas some other parts belong to its variable *margins*, its context-specific features (Kuukkanen 2008: 367-68). The notion of identity of an idea is reduced to that of identity of the *core* of ideas. In this way, we say e.g. that an idea is adopted by a variety of thinkers when its core is, whereas we can highlight the peculiarities of an author's use of an idea by referring to the idea's varying margins in that author. A concept (core) is – we say – thus *contextualized* by reference to its varying margins in different authors.

However, Kuukkanen's talk of *core* and *margins* of a concept cannot counter (a) and (d) (these objections are due in particular to holist theorists and to Quentin Skinner). It is also unclear how the core/margin terminology should be implemented. In the next section, we expound an implementable methodology for studying the history of ideas which is able to also counter adequately both the holists' (a) and Skinner's objection (d).

3 A Sound Method: Neolovejoyanism, or The Model Approach to the History of Ideas

How should we specify the core and the margins of a concept? To do this, we say, we must represent many (shifting) relations among the parts of complex ideas, which will often form a network. And we say that such representations are best accomplished by using (what we call) *models*: explicitly stated conceptual frameworks or perspectives developed with the overt intention to achieve insightful interpretive goals. Models are abstract relational structures or networks of (sub)concepts.[5] In this section we propose a method for the history of ideas based on models. In our 'model approach to the history of ideas' concepts or ideas are construed as (parts of) models.

To illustrate our model approach we will discuss de Jong and Betti's Classical Model of Science (CMS). The CMS is articulated in seven conditions that capture a concept of *proper science* (according to an axiomatic view of science) adopted throughout history by different philosophers and scientists. These conditions, for any system *S* of propositions and concepts (terms), are the following:

(1) All *propositions* and all *concepts* (or *terms*) of *S* concern a *specific set of objects* or are **about** a *certain domain of being*(*s*).

(2a) There are in *S* a number of so-called **fundamental** *concepts* (or *terms*).

(2b) All other *concepts* (or *terms*) occurring in *S* are **composed of** (or are **definable from**) these fundamental *concepts* (or *terms*).

(3a) There are in *S* a number of so-called **fundamental** *propositions*.

(3b) All other *propositions* of *S* **follow from** or **are grounded in** (or **are provable** or **demonstrable from**) these fundamental *propositions*.

(4) All *propositions* of *S* are **true**.

(5) All *propositions* of *S* are **universal** and **necessary** in some sense or another.

(6) All *propositions* of *S* are **known to be true**. A non-fundamental *proposition* is **known to be true through its proof** in *S*.

(7) All *concepts* or *terms* of *S* are **adequately known**. A non-fundamental *concept* is **adequately known through its composition** (or **definition**).

(de Jong and Betti 2010: 186, our emphasis)

Conditions (1)-(7) are very abstract. Together they are meant to capture the *core* of the concept of proper science according to the axiomatic ideal, the basics of an idea held by multiple thinkers in various periods. The concept of proper science articulated here is also highly complex: its core is described as a relational structure of seven conditions containing, in turn, many (sub)concepts. In terms of the core/margins distinction, De Jong & Betti's claim is that whereas many thinkers adhered to the core of the concept of proper science as fixed by (1-7), this concept may differ in different thinkers in its variable margins (terms in *italics*), that is, authors may apply or

[5] 'Any well-articulated idea would qualify as a model of its subject' (McCarty 2004: Background): our models are a concrete example of what 'well-articulated' (said of an idea) means for research in the history of ideas: a complex relational structure of subconcepts of both stable (core) and variable (margins) elements the historian ascribes to that idea. These constructs are a precise conceptual systematisation of what we know about a certain idea, and points towards a smoothing of the 'radical difference' between what McCarty calls 'what we know' and 'what we can specify computationally' (ibid.), as a middle course between the two. See also Floridi 1999: 105 'concepts and ideas are very often expressed by semantic relations and structures, not just by a specific lexicon'.

interpret those conditions differently. We take the core of the concept of proper science to be formed by the terms in **bold**.

We said that in applying the CMS to individual thinkers by specifying conditions (1-7) as to their margins, we are able to *contextualize* it. In this way we can maintain that various authors, say Bolzano and Kant, have adhered to *the same* conception of proper science and still allow for variation. To make this clearer, consider the following fragment of CMS in its most stripped down version, one that reveals only its core:

(1) *S* has items *x*s and items *y*s that are **about** *a*.

[…]

(3a) Some *x*s in S are **fundamental** (call them *Fx*s). […]

Whereas in bold you see features of the *core*, which is stable, the variables *x*, *y*, which are placeholders, indicate the *margins* of the idea of axiomatic science, i.e. elements that can be determined or specified differently. Take, for example, condition (3a). Whereas Kant thought that a science should contain fundamental *judgments*, (so Kant's '*xs*' are *Urtheile*) Bolzano thought that a science should contain fundamental *propositions-in-themselves* (so Bolzano's '*xs*' are *Sätze an sich*). This difference concerns the margins of the concept of proper science. Insofar as both Kant and Bolzano still adhered to conditions (1)-(7), which specify the core of the concept of proper science, they can nevertheless be said to adhere to the same concept of proper science. In this way, we can trace both continuity and differences in the history of thought.

The model approach can counter Skinner's arbitrariness & biases objections (objection (d) in the previous section) in the following way. De Jong and Betti maintain that the concept of science according to (1)-(7) is accepted by the historical actors they study, and that no *other* condition plays a more or equally significant role in shaping the notion of proper science. In addition, they claim that when historians investigate how the model is determined by historical authors, they describe the author's intentions. Models thus allow one to formulate empirical research hypotheses that are open to falsification: Skinner's arbitrariness objection is thus met. Skinner's biases objection is met instead by maintaining that the best way to avoid biases in the history of ideas is to make them *explicit*. Models allow just that. Models such as CMS simply *are* interpretative biases made explicit: whoever employs models like the CMS to reveal (dis)continuities in a certain concept, makes her interpretive biases fully explicit and revisable. By explicating presuppositions in models, and by allowing models to be revised if empirical research shows them to be falsely applied, we minimize the risk of interpretative biases.

The model approach can counter the holist objection (a) in the following way. Recall that the holist maintains that concepts change meaning across contexts. We agree, and think this accords with our method. When we work with a model such as the CMS, we reconstruct *specifications* of some abstract framework that are at most *similar* to each other, similar with respect to, indeed (their function as fixed by) the CMS. Such is the case with Bolzano specification of the CMS, who, as we have seen, took proper sciences to consists of fundamental *propositions*, *versus* Kant's specification of the CMS, who took proper sciences to consist of fundamental *judgments*. Hence, Kant and Bolzano ultimately had a different concept of proper sciences. But as to the latter point in particular we can say that both Bolzano's fundamental propositions and Kant's fundamental judgments have the same *function* (or *purpose*) and the same place in a complex network of concepts that is explicated by the CMS. There is thus a significant amount of continuity in their views on proper science, and this continuity can be adequately traced by means of CMS.[6]

We have come to a solution of the problem of *Faulty Method*: to do research in the history of ideas in a sound way, you should employ models. If no such model is at hand for the concept(s) you wish to investigate, you should build your own.[7] So far, we have not ventured beyond the inner methodological workings of the discipline of history of ideas itself. In the next section we approach the question of how historians of ideas can best take the computational turn in their discipline.

4 Unorganised Data

To our mind, the model approach as described above is not only the best defense of Lovejoy's method known to us, it is also the most appropriate foundation to do *computational* research in the history of ideas. We substantiate our claim here below. First we review the current state of affairs in the methodology of text-based digital humanities and detect problems; then we elucidate how the use of models solves them.

To the best of our knowledge, text mining projects in (text-based) digital humanities tend to typically follow this procedure: (1) humanities experts indicate text corpus; (2) computational experts apply tools to the corpus bottom-up; (3) humanities experts interpret the results. Now, it is far from our intentions to say that this procedure is *always* fruitless – especially because our group currently runs experiments according to that procedure, too;[8] what we intend to say is this. Unless certain interpretive choices that are in fact - and inevitably so – at play in (1-3), and which are kept implicit in the process of applying the tools, are either made explicit, or else critically assessed, and built in the computational tools themselves, this procedure, as it stands, will deliver unsatisfactory results from the point of view of research in the history of ideas. That is, the procedure will not be able to scale up the historian of ideas' traditional method – namely a combination of elements of close reading, historical contextualisation and conceptual analysis in an interdisciplinary, cooperative and non-departmental setting (cf. Lovejoy 1940).

The core risk is, so to speak, one of *interpretive self-deception*, for steps (1)-(3) are too dissimilar from how humans read and interpret texts to be apt for massive data enterprises in the field we are concerned with. Research in cognitive psychology and psychology of teaching since the seventies has shown that comprehending texts involve an interaction between sensory input (such as recognition of words) and *prior* contextual knowledge. Such contextual knowledge is a prerequisite for comprehending prose passages: it is an interpretive key without which texts are

---

[6] One might want to push the point further, and end up saying that neither the core of the concept of proper science is strictly speaking the same in two different authors. We intentionally disregard this point. We also intentionally bypass all sorts of philosophical discussions of what meaning, ideas, the relation between abstract objects and their specifications 'really' are, whether ideas or concepts are literally composed of parts, and so on. As stated in our introduction, our model approach is a theory-free methodology that is deliberately neutral on such issues. As a result, it is compatible with e.g. virtually all positions on concepts. The literature on such topics in the philosophy of language is immensely rich, and a variety of choices is on offer in case one wishes to defend any particular doctrine in this respect. Philosophers of nominalistic persuasion might e.g. find it useful to associate our idea of 'serving the same purpose' with a broadly Quinean view on language, meaning and translation, cf. Quine 1960: 196, and might thus understand identity of concept cores as similarity (cf. e.g. Decock and Douven 2011) .

[7] Models for other concepts are available e. g. the model of fact as fixed in chapter 1 of Betti 2015, expressly built to trace the concept of fact in two millennia of philosophy.

[8] See van Wierst, Vrijenhoek, Schlobach & Betti (2016), this volume.

meaningless.[9] There is no such thing as *first* bottom-up-reading *then* top-down-interpreting. *All* reading is interpretation-laden. Applying these findings to computational text mining for humanities research in the field we are concerned with, and to the aims specified in this paper, we can say: a text mining procedure that *first* turns a massive corpus into data with computational tools, and *then* lets humanities scholars interpret the results, is bound to be either unsatisfactory, or illusory.

In particular, the procedure shifts the traditional problem of interpretation (the humanities´ question: how to interpret these *texts*?) from texts to (derivative) *data* (how to interpret this *data*?). Far from helping with addressing the traditional problem of interpretation, the shift from texts to data in the three-step procedure above might *magnify* it in at least three ways: by giving us more bias, more complexity, and making corpus selection more challenging.

I. More bias. Humanities experts tend to have little knowledge of the technology applied, and to be little involved in tool development or calibration. As a consequence, in such cases humanities experts are unaware of the extent to which the resulting data extraction might be technologically biased, even when shallow text mining tools are applied. That is, the tools applied might have been designed or calibrated in ways that might risk to jeopardize certain interpretations (think, for a trivial case, of a text mining tool designed to standardly remove spelling variants applied in the context of a humanities investigation in which such variants do, in fact, matter). Biases in data extraction are new in the humanities, and as such they are bound to be exponentially more difficult to spot and handle than the biases in text interpretation humanities scholars are familiar with.[10]

II. More complexity. The humanities are concerned with complex texts, i. e. texts licensing a large variety of interpretations. In the field of the authors of this paper, the history and philosophy of sciences such as biology, logic and mathematics, the texts involved are of exceeding conceptual complexity. Generation after generation of philosophy freshmen struggle to understand texts of such complexity, coming back years after years to the same passages again and again, to only discover to their amazement yet a new layer of interpretation. The reason why researchers in a field such as this tends to stick to the study of one period, one author, and even a small cluster of the author's texts all their academic life, is that not only the *answers* to one's interpretive questions are unknown, but that even the right interpretive *questions* to be asked are unknown. In fact, in these fields, choosing the right interpretive question in a certain context forms a fundamental part of the research. The reason is that the interpretive questions to be asked depend on vast and complex contextual information, often of a highly technical nature, such as

---

[9] See the literature quoted in Betti 2014, section 2 and n. 3.

[10] An anonymous referee has pointed out to us that work such as Moretti's distant reading approach is supposed to reduce biases. Let us, to keep things simple, disregard the point that computational tools might themselves introduce biases: even so, one should realise that work such as Moretti's takes place under interpretive conditions which are implicit but heavily infused with theory. Take Moretti's work on the British novel 1740-1850 (Moretti 2013: 179-210): the meaning of 'gothic novel' (Moretti 2013: 207) is implicitly at work in the interpretation of the results of the application of computational tools to the corpus. The meaning of 'gothic' in 'gothic novel' is what guides the selection of certain novels among all novels: the gothic ones. That is: the corpus to which computational tools are applied comes already conceptualised in certain categories given by a pre-existing, implicit modelling (in our sense) of literary genre. That is a bias. A middle course between the latter case (Moretti's gothic novel) as implicit modeling and our proposal as explicit modeling is Overton 2013. In the latter, the NLP analysis is kept shallow, but an extremely rich interpretive step is at play at two junctures, consisting in a sophisticated and well-articulated conceptual systematisation (near to a model in our sense) of a (variety of) concepts of explanation (Overton 2013: 1396) and of the concept of inference to the best explanation (Overton 2013: 1398), through which the results of the experiments are interpreted. In both the implicit (Moretti) and the middle-course case (Overton), certain controlled conditions are in place, i.e. certain interpretive choices are made upfront which in fact control and guide the application of the tools. Our proposal is to take these interpretive choices - the conceptualisations implicit at play in the, so to say, 'external' conditions of application of the tools -, make them as explicit as possible, and make them 'internal', build them in the computational tools from the very start. The advantage is to enable experiments to be applied to truly massive corpora free from selection biases. See also point III below.

theories of plant nutrition in the 18th (cf. e.g. van den Berg 2013), abstraction principles in the mathematics of the 19th (cf. e.g. Mancosu 2014) or implicit quantification practices in formal languages at the beginning of the 20th century (cf. e.g. Betti and Loeb 2012). If computational tools are to help deliver real research results here, then the typical procedure described above must be inverted: interpretive questions have to be clarified before we turn texts into data, for the data must be data *for* an interpretation. The usual three-step procedure adds instead yet another layer of complexity *to the texts themselves*, because it first transforms them into a new, rather specific kind of (derivative) data that requires to be interpreted anew.

III. More challenging corpus selection. Suppose that a historian of ideas wants to investigate the transformations in the concept of axiomatic science in two-hundred years of the history of human thought as represented in texts. In principle, researchers should consider *everything* relevant to their research question. What is this 'everything relevant'? WorldCat counts 17,843,437 books published between 1700 and 1900. Imagine to have solved all sorts of digitisation problems and actually *have* this corpus: a large-scale interdisciplinary, multi-script, multi-language corpus of extreme conceptual density *and* diversity.[11] This is a historian of ideas' dream: it's like a demographer having maximally complete datasets for all populations, the ideal body of evidence to argue for or against interpretive hypotheses. But the task of doing anything sensible with it is daunting. If we *first* apply computational tools to this corpus in order to extract data, and *then* set out to interpret the data, we will be faced with an impossible task, for this mass of data so extracted will be well-nigh impossible to interpret – it'll be pure noise. One solution seems to be to restrict the corpus. Researchers working with traditional methods have developed a quite effective selection method, one that enables them to focus on a corpus of only one to several works by one author – exactly the amount a skilled researcher can study in three-four years. But traditional corpus selection is strongly biased: peer criteria of an author's 'importance' are implicitly endorsed which are infused with – again, implicit – interpretation. Progress and knowledge discovery is often achieved by researchers that *challenge* corpus selection biases by including little known or forgotten (albeit demonstrably influential) sources in the corpus for new and deeper insight. The computational turn in the humanities gives us finally the opportunity of getting rid of biased corpus selection and to operate with a universal corpus: in our thought-experiment we would consider *everything*. But applying the three-step procedure above won't work in this situation. What we would need is a new principle of selection for universal corpora. How shall we proceed?

5 Organised Data

We maintain that by employing our model approach as a foundation to text mining in humanities fields concerned with tracing ideas through texts, all the problems mentioned in the previous section are solved (or kept to a minimum). We propose to ground a computational history of ideas on an ontology-based information extraction procedure where the construction of initial ontologies is guided by and reflects the stable parts of models developed by humanities experts. The (looping) procedure we recommend will roughly look as follows: (1) humanities experts provide computational experts with a model, i.e. *explicitly* structured, *shared* (or shareable) - though *not* formal - semantic framing of domain knowledge about a certain concept; (2) the computational experts turn the core (stable parts) of the model into an ontology (initial ontology), and adapt techniques of ontology extraction to the domain and the corpus, all of this in close collaboration with the humanities experts ('co-makers', cf. McCarty 2008b: 255); (3) ontology extraction is applied to the corpus. This procedure is first tested by building tools to be initially applied to the corpus to address research questions with answers that are already known or highly corroborated by research

[11] Data coming from a WorldCat search from November 11, 2014. As known, WorldCat has a high number of duplicate records (Calhoun and Patton 2011). We ignore the issue here.

results obtained by traditional research. The latter results will form a gold standard to evaluate the tools. When the tools deliver satisfactory results on known findings, they are applied to research questions that are still open.

This procedure avoids the problem of adding an extra layer of complexity to the sources as it builds a specific interpretation in the computational tools explicitly and right away. The problem of interpretation remains thus at the level of the texts themselves, instead of being pushed to a new level. Or better: the procedure is a novel computational version of the traditional method: it scales it up but keeps as many as possible of its advantages. The procedure also solves the problem of corpus selection insofar as the selection is afforded by the framing of knowledge given by models: if computational tools are applied that are built upon models, the universal corpus is filtered in a thematically relevant way which is automatic, explicit and thus revisable, as well as, crucially, guided by domain experts. Finally, the procedure minimizes the problem of added technological biases, because the tools are built by humanities experts in close cooperation with computational experts, and constantly refined and monitored against a gold standard obtained with a traditional, non-computational method. By proceeding in the way sketched, a computational history of ideas maximally profits from computer methods, reduces biases to a minimum by making them *explicit*, and makes a case for keeping humanities experts in the loop.

It should be stressed that from a computational point of view this procedure looks rather natural, because models in our sense can be seen as topic-specific *k*nowledge *o*rganisation *s*ystems (KOS) similar to ontologies. However, our proposal differs from ontology-approaches in the practice in computer science in at least two ways: first, the nearest computational equivalent of models, ontologies, are usually built by computer engineers; second, they are explicit and shared, like our models, but unlike the latter, they are *formal* constructs (Guarino, Oberle, and Staab 2009: 1). It should also be said that since models have stable and variable parts, they can be used to frame the *dynamics* of concepts in an ontology-based approach, an aspect for which the knowledge engineering community has shown considerable interest (cf. e.g. Wang, Schlobach, and Klein 2011).

6 Conclusion

Our Neolovejoyan model approach as described above improves on Lovejoy's method and later improvements in three ways: it abandons Lovejoy's view of *simple* unit-ideas, addresses the problem of biases effectively, and it integrates the objection made against Lovejoy by rival views (holism, Cambridge School) that the *context* in which ideas appear is important. It is also a more general approach than Lovejoy's because it is able to detect continuity of ideas even when ideas have nothing else in common than their *place* in a conceptual structure serving as context (cf. Betti & van den Berg 2014: 14).

We maintain that Neolovejoyanism can be used as a foundation for a computational history of ideas for the following reasons. Models are complex and revisable frameworks that are able to capture shifting perspectives on the development of concepts. They can guide the construction of appropriate computer science ontologies because models frame knowledge in a similar way, that is, in an explicit and shared (or shareable), though non-formal, way. Models can thus inform the construction of computer tools that, when applied to huge masses of texts, are able to aid the extraction of relevant information in a way that approximates what a human researcher in the domain would do. Since they are not formal, humanities experts *themselves* can be expected to possess the skills to build models to frame adequately the domain knowledge they possess.

References

ARMITAGE, David and GULDI, Jo, 2014, The History Manifesto. Cambridge University Press [online]. 2014. [Accessed 24 November 2015]. Available from: http://www.cambridge.org/nl/academic/subjects/history/history-ideas-and-intellectual-history/history-manifesto

ARMITAGE, David and GULDI, Jo, 2015, Le retour de la longue durée: une perspective anglo-américaine. Annales. Histoire, Sciences sociales. 2015. Vol. 70, no. 2, p. 289–318.

ARMITAGE, David, 2012, What's the Big Idea? Intellectual History and the Longue Durée. History of European Ideas. 2012. Vol. 38, no. 4, p. 493–507.

BETTI, Arianna and DE JONG, Willem R, 2010, The Classical Model of Science I – A Millennia-Old Model of Scientific Rationality. Synthese [online]. 2010. Vol. 174, no. 2, p. 185-203. Available from: http://www.springerlink.com/content/0039-7857/174/2/

BETTI, Arianna and LOEB, Iris, 2012, On Tarski's Foundations of the Geometry of Solids. Bulletin of Symbolic Logic. 2012. Vol. 18, no. 2, p. 230–260.

BETTI, Arianna and VAN DEN BERG, Hein, 2014, Modelling the History of Ideas. British Journal for the History of Philosophy. 2014. Vol. 22, no. 4, p. 812–835.

BETTI, Arianna, 2014, On Haslanger's Focal Analysis of Race and Gender in *Resisting Reality* as an Interpretive Model. Krisis. 2014. Vol. 1, p. 13–18.

BETTI, Arianna, 2015, Against facts. Cambridge, MA : MIT Press.

CALHOUN, Karen S. and PATTON, Glenn, 2011, WorldCat quality: An OCLC report [online]. OCLC. [Accessed 12 December 2014]. Available from: http://www.oclc.org/content/dam/oclc/reports/worldcatquality/214660usb_WorldCat_Quality.pdf

DECOCK, Lieven and DOUVEN, Igor, 2011, Similarity After Goodman. Review of Philosophy and Psychology. 2011. Vol. 2, no. 1, p. 61–75.

FLORIDI, Luciano, 1999, Philosophy and Computing: An Introduction. 1 edition. London; New York: Routledge.

GUARINO, Nicola, OBERLE, Daniel and STAAB, Steffen, 2009, What is an Ontology? In Staab, S. & Studer, R. (eds.): Handbook on Ontologies. Berlin: Springer. p. 1–17.

KENTER, Tom, WEVERS, Melvin, HUIJNEN, Pim and DE RIJKE, Maarten, 2015, Ad Hoc Monitoring of Vocabulary Shifts over Time. In: CIKM2015 - 24th ACM Conference on Information and Knowledge Management. Melbourne. 2015.

KUUKKANEN, Jouni-Matti, 2008, Making Sense of Conceptual Change. History and Theory. 2008. Vol. 47, no. 3, p. 351–372.

LEFF, Gordon, 1972, Models inherent in history. In Shanin, T. (ed.): The Rules of the Game: Cross-disciplinary Essays on Models in Scholarly Thought. 1972. London: Tavistock Publications, p. 148–160.

LOVEJOY, A. O., 1936, The Great Chain of Being. Cambridge (MA): Harvard University Press.

LOVEJOY, Arthur O., 1940, Reflections on the History of Ideas. Journal of the History of Ideas. 1940. Vol. 1, no. 1/4, p. 3–23.

MACKSEY, Richard, 2003, The history of ideas at 80. MLN. 2003. Vol. 117, no. 5, p. 1083–1097.

MANCOSU, Paolo, 2014, Grundlagen, Section 64: Frege's Discussion of Definitions by Abstraction in Historical Context. History and Philosophy of Logic. 2014. Vol. 36, no. 1, p. 62–89.

MCCARTY, Willard, 2004, Modeling: a study in words and meanings. In Schreibman, S., Siemens, R. & Unsworth, J. (eds.): A companion to digital humanities. Oxford: Blackwell. p. 254–272.

MCCARTY, Willard, 2008a, Being reborn: the humanities, computing and styles of scientific reasoning. In Bowen, W.R. & Siemens, R. (eds.): New Technologies and Renaissance studies. Tempe (Arizona): Iter. p. 1–23.

MCCARTY, Willard, 2008b, Knowing … : Modeling in Literary Studies. In Schreibman, S. & Siemens, R. (eds.): A Companion to Digital Literary Studies [online]. Oxford: Blackwell. Available from: http://www.digitalhumanities.org/companionDLS/

MCCARTY, Willard, 2008c, What's going on? Literary and linguistic computing. 2008. Vol. 23, no. 3, p. 253–261.

MCMAHON, Darrin M., 2014, The Return of the History of Ideas? In McMahon, D.M. & Moyn, S. (eds.): Rethinking modern European intellectual history. 2014. Oxford: Oxford University Press. p. 13–31.

MICHEL, Jean-Baptiste, SHEN, Yuan Kui, AIDEN, Aviva Presser, VERES, Adrian, GRAY, Matthew K, PICKETT, Joseph P, HOIBERG, Dale, CLANCY, Dan, NORVIG, Peter, ORWANT, Jon, PINKER, Steven, NOWAK, Martin a and AIDEN, Erez Lieberman, 2011, Quantitative analysis of culture using millions of digitized books. Science (New York, N.Y.). January 2011. Vol. 331, no. 6014, p. 176–82. DOI 10.1126/science.1199644.

MORETTI, Franco, 2013, Distant Reading. Verso Books.

OVERTON, James A, 2013, "Explain" in scientific discourse. Synthese. 2013. Vol. 190, no. 8, p. 1383–1405.

QUINE, Willard Van Orman, 1960, Word and Object. 3rd Edition. Cambridge, MA : MIT Press.

SKINNER, Quentin, 2002, Visions of Politics - Vol. 1: Regarding Method. Cambridge : Cambridge University Press.

VAN DEN BERG, Hein, 2013, The Wolffian roots of Kant's teleology. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences. December 2013. Vol. 44, no. 4, Part B, p. 724–734. DOI 10.1016/j.shpsc.2013.07.003.

VAN WIERST, Pauline, VRIJENHOEK, Sanne, SCHLOBACH, Stefan and BETTI, Arianna, 2016, Phil@Scale: Computational Methods within Philosophy [in press]. Transactions in Digital Humanities Luxembourg. 2016. (this volume)

WANG, Shenghui, SCHLOBACH, Stefan and KLEIN, Michel, 2011, Concept drift and how to identify it. Web Semantics Science Services and Agents on the World Wide Web. 2011. Vol. 9, no. 3, p. 247–265.

Biographies of the authors

Arianna Betti is Professor of Philosophy of Language at the University of Amsterdam. After studying historical and systematic aspects of ideas such as *axiom*, *truth* and *fact* (*Against facts*, MIT Press, 2015), she is now trying to trace the development of ideas such as these with computational techniques.

Hein van den Berg is Assistant Professor at the University of Amsterdam and VU Amsterdam. He studies axiomatic ideals of science in the 18th century, the history and philosophy of biology, and the application of computational tools to philosophical and scientific ideas.