Some Reflections on Historiographical Uncertainty and Computational Modeling*

Prof. Dr.-Ing. Michael Piotrowski[1]

2022-05-12

[1] Faculty of Arts · Department of Language and Information Sciences · University of Lausanne

1 Introduction

As the title says, what I'd like to present are "some reflections on historiographical uncertainty and computational modeling." Some of them come from the one-year pilot project *An Agile Approach Towards Computational Modeling of Historiographical Uncertainty* conducted in 2020, which has motivated my new project *Towards Computational Historiographical Modeling: Corpora and Concepts*.[1]

This project is not primarily about uncertainty, but it is strongly motivated by insights from the pilot project, in particular the insight that uncertainty—and historiographical uncertainty in particular—is not something that can be handled as an afterthought, but constitutes a fundamental epistemological challenge.

Let me now take a step back and ask the dreaded but inevitable question: what do we mean by "digital humanities"? You're certainly aware that this question has been and continues to be a subject of much debate, to the extent that many people in the field have grown tired of it.

While this is to some extent understandable, I'm afraid it doesn't resolve the issue. What's funny, though, is that at the very same "digital humanists" who refuse to define what they mean by "digital humanities" do not seem to tire stressing the alienness of computational methods to the humanities and their immense epistemological impact. Even—or perhaps especially?—among proponents of digital humanities we find a fundamental unease with the "tools and methods" that presumably characterize digital humanities. Computational methods are essentially regarded as intrusion of a *useful but lesser mode of inquiry* into the humanities.

Slide 4

> While the research questions may be similar to those in the "analogue humanities," the constitution of what counts as an interesting answer may have changed so profoundly that it no longer is compatible with the traditional epistemology of the humanist disciplines.
> —Berry and Fagerjord (2017, 21)

And Dobson (2019) affirms that these conflicts between computation and "the traditional epistemology of the humanist disciplines" are not only still not resolved, but in fact "fundamentally unresolvable."

*Talk held at the CREATE Symposium on Historical Uncertainties

[1] Funded by the Swiss National Science Foundation, grant no. 105211_204305/1.

Slide 5

> [...] not only are these conflicts still not resolved, but they remain fundamentally unresolvable.
> —Dobson (2019, 6)

So it makes sense when Drucker (2011) exhorts:

Slide 6

> [D]igital humanities can no longer afford to take its tools and methods from disciplines whose fundamental epistemological assumptions are at odds with humanistic method.
> —Drucker (2011, para. 6)

What is remarkable, though, is that despite the seriousness of these issues, none of the authors seems to feel the need to tell us what "the traditional epistemology of the humanist disciplines" (Berry and Fagerjord 2017, 21) is, or how any cooperation may be possible with "disciplines whose epistemological foundations and fundamental values are," as (Drucker 2012, 85) claims, "at odds with, or even hostile to, the humanities."

In fact, there is very little research on the epistemological foundations of the digital humanities and on how they differ from the "traditional" humanities; in other words, research that tries to answer the question: what is the impact of computational methods on the production of knowledge in the humanities? Any answer would have to take into account the issue of historiographical uncertainty, or more generally, multi-interpretation.

There seem two main obstacles to even start thinking about these questions. First of all, as Abraham Moles ([1990] 1995) remarked, there actually is not much in terms of an epistemology of the humanities.

Slide 7

> [L]'épistémologie des sciences humaines n'existe pas encore. —Moles ([1990] 1995, 46)

Second, there is this striking refusal to define what one means by "digital humanities." Just to give you a very recent example, in an interview Valérie Schafer from the University of Luxembourg gave in March[2], after lengthily beating about the bush, she concluded:

Slide 8

> To summarize, it's a multi- and interdisciplinary field at the intersection of humanities and computer science, which is based on a shared passion for research, creativity, new ways of working, and values. —Valérie Schafer

This sounds rather like an artists' collective. But in research, this point of view raises a serious issue: if you cannot say what you are doing, and why you are doing it, it will be hard to say how you are producing knowledge.

2 Defining Digital Humanities

Don't worry, I will just give you my definition (Piotrowski 2018), because we need a definition if we want to move on.

[2] Source: Digital Humanities – A Science unto Itself? A Conversation with Valérie Schafer

Slide 10

Digital Humanities

1. The **applied digital humanities** are concerned with the construction of formal models of the phenomena studied by their "mother disciplines," as well as the methodology of their construction.
2. The **theoretical digital humanities** study the general properties of formal models in the humanities at a higher level of abstraction.

As you can see, I distinguish between applied and theoretical digital humanities. It is crucial to make this distinction, because their research objects and research objectives are very different: the theoretical digital humanities create and study the *meta*models whose concrete application to research questions in the disciplines of the humanities is the object of the applied digital humanities.

3 Models

Clearly the notion of *model* is central to my understanding of digital humanities. Let me briefly explain what I mean by "model." Even in the natural sciences, the explicit notion of *model* is surprisingly recent, and only began to be studied seriously in the 1950s. One of the first "modern" attempts at a *general* definition of models is the following by the Belgian philosopher Leo Apostel:

Slide 12

> [...] *any subject using a system A that is neither directly nor indirectly interacting with a system B to obtain information about the system B*, is using A as a model for B.
> —Apostel (1961, 36)

I think this is still an excellent, concise definition. And it does not just apply to science, but holds generally. The specific challenge of science, though, is to build *intersubjective models* that can be communicated and tested against some "reality." As Gilles-Gaston Granger ([1960] 1967) notes, the goal of any science (natural or other) is to build *coherent and effective models of the phenomena they study*.

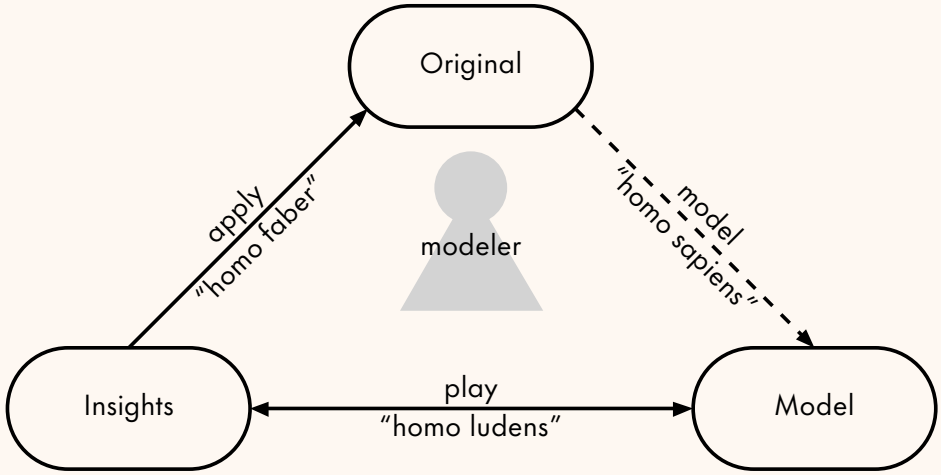
Slide 13

> Quant à l'intentionnalité scientifique, à la visée, nous l'avons déjà définie comme construction de *modèles cohérents et efficaces du phénomène*.
> —Granger ([1960] 1967, 215)

What we have then is a setting that I have recently started calling the *modeling triad*.[3]

[3] The terms *homo sapiens*, *homo ludens*, and *homo faber* on the edges are references to the *epistemological control loop* (erkenntnistheoretischer Regelkreis) by Georg Klaus (1966, 93).

Slide 14



First, there is the original, the phenomenon that we want to study and on which we aim to gain insights. The original is not directly and completely accessible to us—something which is obviously true for historical phenomena, and which is precisely why we *model*.

The second step is to *play* with the model. Playing can mean, for example, to change the value of a parameter or to add or remove some component, depending on the nature of the model. In this context, *playing* should be understood as a technical term for manipulating a model (as used, e.g., by Klaus 1966, 93ff), in contrast to *experimenting*, which refers to *strong interaction* (Moles [1990] 1995, 100), i.e., the manipulation of the original. Our hope is that playing with the model will yield new insights on the original.

Third, we *apply* our insights gained from our interaction with the model to the original. "Apply" clearly has to be understood in a very wide sense, as the nature of the application will be very different in different domains. In the natural sciences, a prediction made on the basis of a model can be compared to the actual outcome; in engineering, design decisions may be taken, but none of this is generally possible in the humanities, so the *viability* of the model will have to be evaluated in different ways. I say "viability," because, and let me stress this again, a model cannot be "compared" to the original. This would require an objective observer, and as Heinz von Foerster put it, objectivity is a subject's delusion that observing can be done without him.

4 Models in the Humanities

Some disciplines in the humanities have a long tradition of using explicit models (sometimes even semi-formal or formal models) under various names, for example *grammars* in linguistics. Other disciplines are openly hostile to this notion. With respect to historians, Gordon Leff (1972) observes:

Slide 16

> Historians as a profession are not given to constructing or employing models in any formal or explicit sense; where they do, it is mainly in areas bordering on other disciplines, especially economics and social studies. Most historians, if asked, would probably deny that models had anything to do with their subject.
> —Leff (1972, 148)

In fact, he points out, historians could hardly put pen to paper without having implicit models of what they were studying.

In digital humanities, the importance of modeling is frequently acknowledged, but—and this is probably due to the fact that modeling is widely considered foreign to the humanities—typically as a specific, clearly delimited activity, which is primarily needed for technical reasons. For example,

Ciula et al. (2018, 346) call modeling "one of the core research practices," which "is due to the fact that explicit models are extensively required in DH in order to operationalise research questions."

While I agree in principle, I would argue that this view is too limited and potentially misleading. As Joshua Epstein (2008) has remarked, the choice "is not whether to build models; it's whether to build explicit ones."

5 Uncertainty

With this in mind, let's talk about uncertainty. As Bertrand Russell has pointed out,

Slide 17

> All human knowledge is uncertain, inexact, and partial.
> —Russel (1948, 527)

In other words: "Uncertainty is a fundamental—and unavoidable—feature of daily life" (Halpern 2017, 1). However, it was also Russell who remarked that even if nothing is certain, it is not necessarily uncertain to the same degree:

Slide 18

> When one admits that nothing is certain one must, I think, also add that some things are more nearly certain than others.
> —Russell ([1949] 1997, 92)

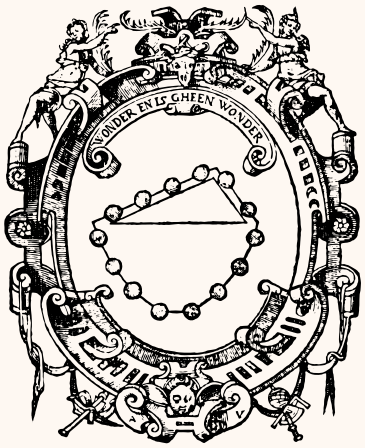
This is probably where the problem starts, especially when we want to represent "uncertain, inexact, and partial" knowledge formally and for computational processing.

When it comes to history, it is particularly evident that our knowledge of the past is always "uncertain, inexact, and partial." It is overwhelmingly not based on any kind of first-hand experience but it can only be gained indirectly.

From Roman historians we know about the Battle of the Teutoburg Forest in 9 AD. However, from the textual sources alone the site of the battle is uncertain. Archeological evidence seems to suggest that it took place near the modern village of Kalkriese rather than in what 19th-century historians believed to be the Teutoburg Forest. But how can we be *sure* that, unlike the hundreds of other locations that have been suggested over the centuries, this is the *real* site of the Battle of the Teutoburg Forest?

Another frequent example of uncertainty in historical research concerns dates. More often than not we do not have precise biographical data—date of birth, date of death, etc.—on historical persons. And this doesn't just concern "minor" figures; for example, we know hardly anything of Simon Stevin in this respect.

Slide 20



Slide 21

> L'historien de demain sera programmeur ou il ne sera pas.
> —Le Roy Ladurie (1968)[a]

[a] Photo: Claude Truong-Ngoc – ⓒⓘⓞ via Wikimedia Commons

When people started to use computers to process historical data, they realized that when you go beyond census data, for example, they'd need to represent this kind of information on the computer. Of course, if you build a database, you can just store "ca. 1500" or "early 18th century" or "between 1630 and 1659" or "1291 (?)" in a text field, and you're done. However, this has the downside that it cannot be processed automatically. A simple solution is to simply map this to something that a DBMS designed for exact information can handle. So, you replace "between 1630 and 1659" with the (rounded) average 1645. Now the data is now easily machine-processable—but we've lost a lot of information, in particular information about uncertainty.

So, people have devised more sophisticated representations. For example, the TEI Guidelines define ways for encoding uncertain dates using expressions like "not before," "not after," and so on.

Slide 22

The @notBefore and @notAfter attributes may be used to express a range of possibilities for a particular date (or time). For example the following element, extracted from an imaginary prosopographic entry for Anne Calthorpe, indicates that although the exact date of her death is not known, it can be narrowed down to a particular range: from 22 August 1579 to 28 March 1582, inclusive. Ostensibly the encoder has evidence that Anne Calthorpe was alive on the 22nd of August 1579 and evidence that she was already dead on the 28th of March 1582.

```
<death notBefore="1579-08-22" notAfter="1582-03-28"/>
```

(From the TEI Guidelines, chap. 13 "Names, Dates, People, and Places")

I guess that this is hardly news to you. In fact, I'm only telling you this to say that I'm actually not interested in this type of uncertainty—which I call historical uncertainty, but about another one, which is much less acknowledged, and which I have called *historiographical* uncertainty.

6 Historiographical Uncertainty

What do I mean by historiographical uncertainty?

Robin Winks's analogy of "the historian as detective" (Winks 1969) is a good way to explain the distinction. Historians are like detectives arriving at a crime scene. Take the prototypical case of a murder: if there is a dead body, there is little doubt that someone was killed. This is the historical

fact: something has happened in the past, and this is the result. The detective takes inventory of all the facts, everything found on the scene that may *potentially* serve as evidence.

The second and arguably the main task of the detective is to build a case, i.e., to create, on the basis of facts, which now serve as evidence, a *causal model* that explains how the crime came to be and who was involved.

In history, this is called the *narrative*. It is not merely a collection of facts, nor merely a chronology; as the legal principle states: *post hoc sed non propter hoc*. It is in the building of the case and the creation of a narrative that the facts become evidence that *supports* the proposition.

Historical uncertainty pertains to the *facts* of the past. We are pretty sure that Simon Stevin lived, but we do not know exactly when. We are pretty sure that the Battle of the Teutoburg Forest took place, but we don't know exactly where. We don't have absolut proofs, but we have circumstantial evidence, or as historians say: sources, in particular written documents.

Or, in the words of Carl L. Becker:

Slide 24

> No doubt throughout all past time there actually occurred a series of events which, whether we know what it was or not, constitutes history in some ultimate sense. Nevertheless, much the greater part of these events we can know nothing about, not even that they occurred; many of them we can know only imperfectly; and even the few events that we think we know for sure we can never be absolutely certain of, since we can never revive them, never observe or test them directly.
> —Becker ([1932] 1969, 6)

Thus, as Becker remarks, "[t]he event itself once occurred, but as an actual event it has disappeared; so that in dealing with it the only objective reality we can observe or test is some material trace which the event has left—usually a written document" (Becker [1932] 1969, 6).

Since there is no way we can directly compare the narrative to the reality of the past, there is always uncertainty, but I think it is useful to distinguish between these two types of uncertainty:

- *historical uncertainty*, which concerns historical facts, i.e., the past, and
- *historiographical uncertainty*, which concerns the interpretation of the facts, especially when used as evidence for a particular narrative.

I admit that the terms are not quite ideal and potentially misleading because "uncertainty" can be misunderstood to imply that there is some "truth," namely one that will emerge once all the documents have been "discovered, cleaned up, and put into the right order," as the historical positivists asserted.

Slide 25

> On peut penser qu'un jour viendra où, grâce à l'organisation du travail, tous les document auront été découverts, purifiés et mis en ordre, et tous les faits dont la trace n'a pas été effacée, établis. — Ce jour-là l'histoire sera constituée […].
> —Langlois and Seignobos (1898, 277)

I mention this here, because this 19th-century school, which considered itself *histoire scientifique* (Carbonell 1978), i.e., scientific history, is experiencing a revival (which one could perhaps call "digital positivism") with the increasing "datafication" of history, not only in cliodynamics (Turchin 2008), but also more generally in digital history.

The 19th-century French historian Fustel de Coulanges affirmed, "It is not I who speak, but history which speaks through me."

Slide 26

> Ne m'applaudissez pas, ce n'est pas moi qui vous parle ; c'est l'histoire qui parle par ma bouche.
> —Numa Denis Fustel de Coulanges, d'après Monod (1889), 278

But the causal links between the facts do *not* simply emerge from the facts; this would require some sort of Laplacian demon that not only "would know all forces that set nature in motion," but also all forces that determine human decisions and actions.

The narrative is *always* constructed, and more sources or a better knowledge do not necessarily imply a better approximation of the "truth." A nice example, for which I thank Floor Koeleman, is that of glossopetrae, fossilized shark teeth (see Koeleman 2022, 207f).

Glossopetrae are often seen in 17th-century depictions of *wunderkammern* (cabinets of curiosities), such as this one by Frans Francken the Younger (ca. 1620–1625).

Slide 27

The Swiss polymath Conrad Gessner had already noted the similarities between glossopetrae and shark teeth in his treatise *De rerum fossilium, lapidum et gemmarum maxime, figuris et similitudinibus liber*, published in Zurich in 1565 (Etter and Schmidt 2020). With his treatise *De glossopetris dissertatio*, published in Rome in 1616, Fabio Colonna was the first to convincingly demonstrate that glossopetrae are indeed shark teeth. In his 1667 publication *Canis carchariae dissectum caput*, Danish naturalist Nicolaus Steno discussed their composition and famously produced a depiction of a shark's head bearing such teeth.[4] Nevertheless, even beyond the end of the 17th century, they continued to be widely thought of as snake tongues or dragon teeth.

We don't know whether the depiction of the glossopetra in this painting and in this context has some special significance. Our modern knowledge that the depicted object is actually a shark's tooth doesn't help us to interpret the symbolic content of the painting and to discover its "true meaning."

Thus, what Neil Gershenfeld calls the "most common misunderstanding about science" is also the most common misunderstanding about history.

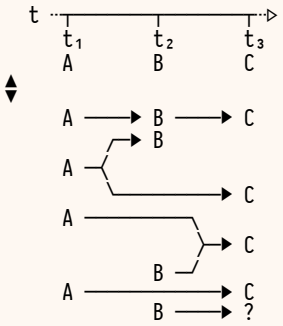
[4] See http://n2t.net/ark:/13960/t3mw8zm39.

Slide 28

> The most common misunderstanding about science is that scientists seek and find truth. They don't—they make and test models. […] Making sense of anything means making models that can predict outcomes and accommodate observations. Truth is a model.
> —Neil Gershenfeld

This is why I'm starting to think that *multi-interpretation*, the term used by Jean-Claude Gardin (1990), may actually be a better or more appropriate term. If we have established the chronology of events—which, due to historical uncertainty, may be easier said than done—there are still multiple possible interpretations regarding their causal relations.

Slide 29



The historical facts delimit the possible historiographical interpretations, but they *do not determine them completely*. One may say that historical uncertainty concerns the events A, B, C, whereas historiographical uncertainty concerns the conjectured causal links between them.

The causal connections on this slide are of course very much simplified for the sake of illustration. One may even wonder whether—unlike in the natural sciences—time and causality have to flow in the same direction: predictions, expectations, assumptions about future events (which may or may not actually take place) certainly have an impact on human decision-making.

A certain risk comes with this apparent dualism: one may believe that there is a neat separation between the two.

However, it's not that easy. The supposedly "objective" historical event of the Battle of the Teutoburg Forest turns out to be largely a construction as well, subject to what Granger has called the *découpage des phénomènes* (Granger [1960] 1967, chap. 4), the "carving out" of what we want to consider the phenomena.

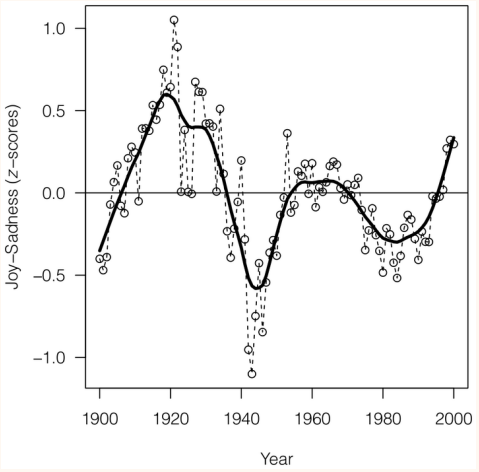
7 Découpage

Let me say a few more words about the notion of *découpage*.

While there is much discussion of methods and their appropriateness—and many common definitions as well as a large part of the criticism of DH are related to these methods—there is very little discussion of the underlying assumptions. For example,

Slide 31

According to Acerbi et al. (2013), moods in 20th-century US and UK literature have followed broad historical trends, including a "sad" peak corresponding to World War II, and two "happy" peaks, one in the 1920s and one in the 1960s. There is also a "sad" period starting in the 1970s, with an increase in "happiness" in the later years of the data set.



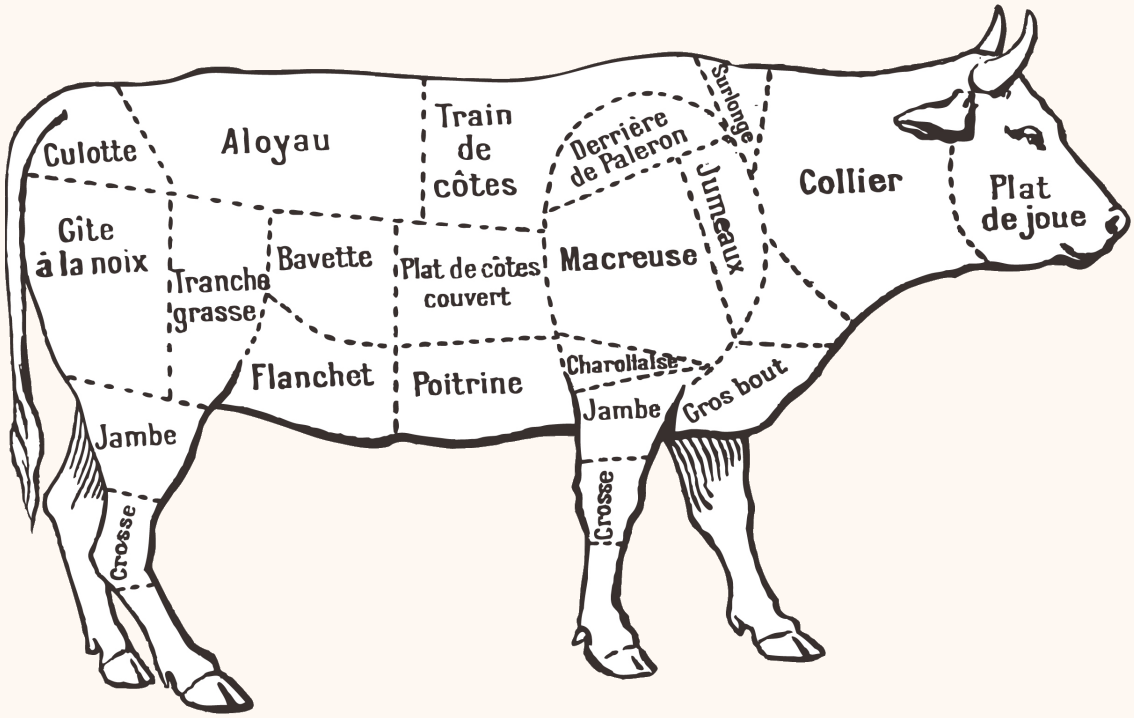
Hardly any of these terms are actually defined, and we don't know, for example, what "happiness" or "sadness" actually models. The authors do acknowledge that despite the assumed causal link between war and "sadness" in literature World War I does not seem to register any particular change in mood words. Thus, as you can see here, this causal link does not actually emerge naturally from the data: there is, in fact, a lot of historiographical uncertainty.

I would like to highlight two epistemological problems here, which *should* be obvious, but apparently are not.

The first one is that all these seemingly objective quantitative measures (n items of class X, m items of class Y, and so on) are crucially dependent on the preceding *qualitative* analysis that established the classes: the difference between an apple and an orange is not quantitative, but *qualitative*. What is more, it is in no way the only possible classification: classify apples, oranges, and pears as "fruit," or by color, size, or shape, or for that matter, acidity, and you may get completely different numbers.

Granger ([1960] 1967) calls this *le découpage des phénomènes*. As the English translation (Granger 1983) leaves the term *découpage* untranslated, I do so as well, but in my opinion, there is really nothing inherently untranslatable about it.

Slide 33



Granger notes that the *découpage* of human facts—i.e., those studied by the humanities—presents a special difficulty:

> Here the phenomena have an immediate *sense*, which means that they spontaneously take part in a universe of valorized and directed actions, either in the consciousness of an individual, or in the organization and functioning of a collectivity which is given as a whole, even when the relations of this whole escape us. This sense is transmitted by language for the speaking subject of each social group, and it is this that constitutes, for our consciousnesses as agents, the very essence of the given human fact.
> —Granger (1983, 49–50)

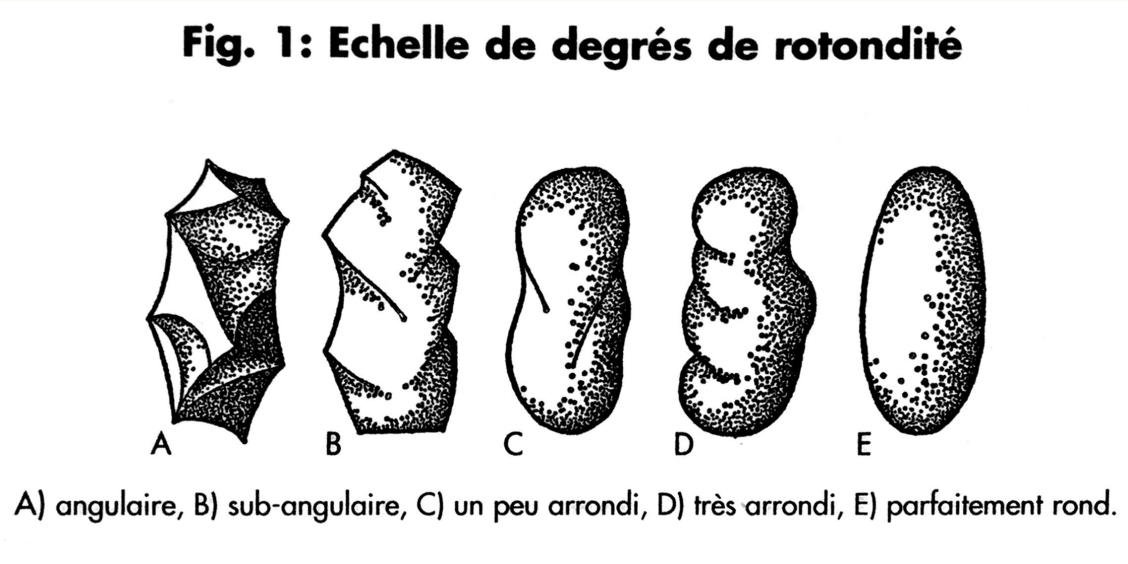
Slide 34

> C'est que le découpage des faits humains présente une difficulté spécifique. Les phénomènes ont ici un *sens* immédiat, ce qui veut dire qu'ils font spontanément partie d'un univers d'actions valorisées et orientées, soit dans la conscience d'un individu, soit dans l'organisation et le fonctionnement d'une collectivité qui se donne comme un tout, alors même que les liaisons de ce tout nous échappent. Ce sens est véhiculé par le langage pour le sujet parlant de chaque groupe social, et c'est lui qui constitue pour nos consciences d'acteurs l'essence même du fait humain donné.
> —Granger ([1960] 1967, 64)

As a naive instrument of this *découpage* of facts, Granger argues, language naturally transmits what he calls *myths*, as well as ideologies. One of the first challenges is thus to overcome this overdetermination of notions in order to reach scientific concepts.

Only then we can think about what Moles ([1990] 1995, 110) considers the first step in the transition from the vague towards the exact: measurement, as a system of comparison between the subjective human impression and external material data, as in the example of a scale of roundness (as a function of the radius of curvature of the various asperities [rough edges] and angles of the particle considered)—in a way establishing an intersubjective scale between a "rock" and a "pebble."

Slide 35



However, in general the humanities still have a long way to go in this respect. Granger ([1960] 1967, 66) notes that "the constitutive problem of the humanities can be described as the transmutation of lived meanings into a universe of objective meanings. Hence the fundamental importance, beyond the trials and tribulations, of new disciplines like information theory and cybernetics."

Slide 36

> Le problème constitutif des sciences de l'homme peut être dès lors décrit comme transmutation des significations vécues en un univers de significations objectives. De là vient l'importance fondamentale, par-delà les tâtonnements et les insuffisances, de disciplines nouvelles comme la théorie de l'information et la cybernétique.
> —Granger ([1960] 1967, 66)

The second issue is, again, a general one, but which is particularly tricky in our context. In statistical terms, a corpus can be considered a *sample* of some *population* that cannot be analyzed in its entirety. The core idea of statistics is that by analyzing the sample, we can obtain results that can be generalized to tell us something about the whole population.

As a matter of fact, this is merely a special case of the modeling triad I showed you earlier. Thus, and as I have argued before (Piotrowski 2019), a corpus should be considered a model in the sense of Apostel (1961). Creating a corpus thus means constructing a model, and modelers consequently have to answer questions such as: What is the original? In what respects is the model a reduction of it? And for whom and for what purpose am I creating the model?

The larger and the more "complete" a corpus appears, the greater the danger to succumb to an "implicit essentialism" (Mothon 2010, 19) and to mistake the model for the original, a fallacy that can frequently be observed in the field of cultoromics (Michel et al. 2011), when arguments are being made on the basis of the Google Books Ngram Corpus.

Slide 37

> A map is *not* the territory it represents, but, if correct, it has a *similar structure* to the territory, which accounts for its usefulness.
> —Korzybski (1933, 58)

However, an analysis of a corpus will *always* yield results; the crucial question is whether these can tell us anything about the original phenomenon it aims to model.

The crucial point is that corpora are not naturally occurring but intentionally constructed. A corpus is *already* a model and thus not epistemologically neutral. In my new project we aim to build on Gaston Bachelard's notion of *phénoménotechnique* (Bachelard [1934] 1968).

Slide 38

> La véritable phénoménologie scientifique est donc bien essentiellement une phénoménotechnique. […] Elle s'instruit par ce qu'elle construit.
> —Bachelard ([1934] 2020, 35)

Bachelard originally developed this notion, which treats scientific instruments as "materialized theories," as a way to study the epistemology of modern physics. As we have said before, the humanities have also always constructed the objects of their studies (to a large extent not directly observable, but only through artifacts, in particular texts) through, for example, the categorization and selection of sources and the (hermeneutic) postulation and affirmation of phenomena.

However, only the praxis has been codified to some extent as "best practices," such as source criticism. Theories—or perhaps better: models and metamodels, as the term "theory" has a somewhat different meaning in the humanities than in the sciences—are not formalized and are only suggested by the (natural language) narrative. What history (and the humanities in general) traditionally do not have is something that corresponds to the scientific instrument.

This changes with digitalization and datafication: phenomena are now constructed and modeled through data and code, and (like in the sciences), the computational model takes on the role of the instrument and "sits in the center of the epistemic ensemble" (Rheinberger 2005, 320). Corpora are then, methodologically speaking, phenomenotechnical devices and form the basis and influence how we build, understand, and research higher-level concepts.

Statistics provide us with means to formally describe and analyze a specific subclass of models that are able to represent originals that have particular properties. However, the phenomena studied by the humanities generally do not have these properties, and we thus lack adequate formal methods to describe them.

8 Formalization

Let me say a few more words about formalization. As I have mentioned before, when we consider the epistemology of the digital humanities, we are especially interested in a specific type of formal models, viz. computational models.

Concerning formalization in the humanities, we build in particular on the work of Gilles-Gaston Granger. As for the meaning of the word *formal*, he noted that to formalize is to reduce the language of a theory to primitive expressions and explicit rules of construction.

Slide 40

> Formaliser, c'est réduire le langage d'une théorie à des expressions primitives et à des règles explicites de construction.
> —Granger ([1960] 1967, 160)

This raises fears among some humanities scholars, because they associate formalization and computation with mathematization, and as Granger ([1960] 1967, 110) has observed, it is still true that "for many who work in the humanities, mathematization is equivalent to the introduction of quantity, indeed of number." This immediately raises the question: but isn't all formal modeling quantitative? Don't the digital humanities mean above all the treatment of large quantities of data, *big data*, and therefore the use of statistical methods, resulting in the "reduction of phenomena to calculations" (Granger [1960] 1967, 19)?

I don't think so, and also Granger stressed that formalization in the humanities consists in the transition from the unstructured to the structural, rather than in quantification (Granger [1960] 1967, 113).

Slide 41

The Challenge of Digital Humanities

> Il semble donc qu'une élaboration scientifique des notions qualitatives consiste dans le passage de l'astructuré au structural, bien plutôt que dans une quantification.
> —Granger ([1960] 1967, 113)

As Granger ([1960] 1967, 19) points out, there is a common confusion between formal thought and the work of mathematicians. While it is quite true in a sense that all effective scientific formalism tends towards a mathematical form, this still does not mean that the formalism is unavoidably reduced to the *usual* and *current* instruments of mathematics. And he asks: is it not just this supposed reduction that the opponents of all formalization in the humanities are criticizing? It will be one of the aspects of our task, Granger says, to show formal thought at work in the humanities, not only as a reduction of phenomena to calculations, but also as the invention of new structures, even, indeed, of an original mathematics.

Slide 42

> À cet égard, il convient de dissiper un préjugé qui nuit souvent à l'appréciation qu'on peut faire de la valeur du formalisme dans les sciences. C'est la confusion que l'on favorise d'ordinaire entre la pensée formelle et l'œuvre des mathématiciens. S'il est bien vrai en un sens que tout formalisme scientifique efficace tend vers un statut mathématique, ce n'est pas pour autant qu'il se réduise infailliblement aux instruments *usuels* et *actuels* des géomètres. N'est-ce pas cependant à cette réduction supposée que s'attaquent essentiellement les adversaires de toute formalisation dans les sciences de l'homme ? Ce sera l'un des aspects de notre tâche que de montrer la pensée formelle à l'œuvre dans les sciences humaines, non pas seulement comme réduction des phénomènes aux calculs, mais aussi comme invention de structures nouvelles, voire même d'une mathématique originale.
> —Granger ([1960] 1967, 19)

I believe that this could really be considered the main challenge of digital humanities.

It also highlights a limitation of current DH research, which is overwhelmingly quantitative, whereas humanities research questions are ultimately qualitative, *why?* questions. The "original mathematics" that Granger envisioned have thus to be understood as "mathematics without numbers" (Kemeny 1959), a general "science of patterns" (Resnik 1999).

9 Conclusion

What I have tried to outline in this talk is some of the background and the motivation for the project *Towards Computational Historiographical Modeling: Corpora and Concepts* and, in fact, my research program more generally.

So far, digital humanities has largely contented itself with borrowing methods from other fields and has developed little methodology of its own. The almost exclusive focus on methods and tools represents a major obstacle towards the construction of computational models that could help us to obtain new insights into *humanities* research questions rather than just automate primarily quantitative processing—which is, without doubt, useful, but inherently limited, given that the research questions are ultimately qualitative.

The outcome of the project is, of course, still open, but I believe that the combination of Bachelard's *phénoménotechnique*, Granger's theoretical work on formalization in the humanities, and the logicist approach of Gardin—which I couldn't go into today—constitute a solid epistemological basis for studying the epistemological impact of computational historiographical models, including the issue of uncertainty, as well as others that I could not discuss today.

References

Acerbi, Alberto, Vasileios Lampos, Philip Garnett, and R. Alexander Bentley. 2013. "The Expression of Emotions in 20th Century Books." *PLoS ONE* 8 (3): e59030. https://doi.org/10.1371/journal.pone.0059030.

Apostel, Leo. 1961. "Towards the Formal Study of Models in the Non-Formal Sciences." In *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, edited by Hans Freudenthal, 1–37. Dordrecht: D. Reidel. https://doi.org/10.1007/978-94-010-3667-2_1.

Bachelard, Gaston. (1934) 1968. *Le nouvel esprit scientifique*. 10th ed. Paris: Les Presses universitaires de France.

———. (1934) 2020. *Le nouvel esprit scientifique*. 1re édition critique. Paris: Les Presses universitaires de France.

Becker, Carl L. (1932) 1969. "Everyman His Own Historian." In *The Historian as Detective: Essays on Evidence*, edited by Robin W. Winks, 3–23. New York, NY: Harper & Row.

Berry, David M., and Anders Fagerjord. 2017. *Digital Humanities*. Cambridge: Polity.

Carbonell, Charles-Olivier. 1978. "L'histoire dite 'positiviste' en France." *Romantisme* 8 (21): 173–85. https://doi.org/10.3406/roman.1978.5216.

Ciula, Arianna, Øyvind Eide, Cristina Marras, and Patrick Sahle. 2018. "Models and Modelling Between Digital and Humanities: Remarks from a Multidisciplinary Perspective." *Historical Social Research* 43: 343–61. https://doi.org/10.12759/HSR.43.2018.4.343-361.

Dobson, James E. 2019. *Critical Digital Humanities: The Search for a Methodology*. University of Illinois Press. https://doi.org/10.5406/j.ctvfjdomf.

Drucker, Johanna. 2011. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5 (1). http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

———. 2012. "Humanistic Theory and Digital Scholarship." In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 85–95. Minneapolis, MN, USA: University of Minnesota Press. http://dhdebates.gc.cuny.edu/debates/text/34.

Epstein, Joshua M. 2008. "Why Model?" *Journal of Artificial Societies and Social Simulation* 11 (4). http://jasss.soc.surrey.ac.uk/11/4/12.html.

Etter, Walter, and Olivier Schmidt. 2020. "Gessner's Fossil Crab." *Journal of the History of Collections* 32 (3): 411–30. https://doi.org/10.1093/jhc/fhz039.

Gardin, Jean-Claude. 1990. "L'interprétation dans les humanités: réflexions sur la troisième voie." In *Interpretation in the Humanities: Perspectives from Artificial Intelligence*, edited by Richard Ennals and Jean-Claude Gardin, 22–59. Library and Information Research Report 71. London: British Library.

Granger, Gilles-Gaston. (1960) 1967. *Pensée formelle et sciences de l'homme*. Nouvelle éd. augmentée d'une préface. Paris: Aubier-Montaigne.

———. 1983. *Formal Thought and the Sciences of Man*. Dordrecht: Reidel.

Halpern, Joseph Y. 2017. *Reasoning about Uncertainty*. 2nd ed. Cambridge, MA, USA: MIT Press.

Kemeny, John G. 1959. "Mathematics Without Numbers." *Daedalus* 88 (4): 577–91. http://jstor.org/stable/20026529.

Klaus, Georg. 1966. *Kybernetik und Erkenntnistheorie*. Berlin: VEB Deutscher Verlag der Wissenschaften.

Koeleman, Floor. 2022. "Re-Viewing the *Constcamer*: A Digital Approach to Seventeenth-Century Pictures of Collections." In *Digital History and Hermeneutics: Between Theory and Practice*, edited by Andreas Fickers and Juliane Tatarinov, 201–18. Berlin: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110723991-010.

Korzybski, Alfred. 1933. *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. Lancaster, PA: International Non-Aristotelian Library Publishing Company. https://n2t.net/ark:/13960/t6c261n93.

Langlois, Charles-Victor, and Charles Seignobos. 1898. *Introduction aux études historiques*. Paris: Hachette.

Le Roy Ladurie, Emmanuel. 1968. "La fin des érudits." *Le Nouvel Observateur*, no. 182 (May): 38–39.

Leff, Gordon. 1972. "Models Inherent in History." In *The Rules of the Game: Cross-Disciplinary Essays on Models in Scholarly Thought*, edited by Teodor Shanin, 148–60. Abingdon: Tavistock.

Michel, Jean-Baptiste, Yuan K. Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331 (6014): 176–82. https://doi.org/10.1126/science.1199644.

Moles, Abraham A. (1990) 1995. *Les sciences de l'imprécis*. Edited by Élisabeth Rohmer-Moles. Paris: Seuil.
Monod, Gabriel. 1889. "M. Fustel de Coulanges." *Revue historique* 42 (2): 277–85. https://www.jstor.org/stable/40938008.
Mothon, Bernard. 2010. *Modélisation et vérité*. Paris: Archétype82.
Piotrowski, Michael. 2018. "Digital Humanities: An Explication." In *Proceedings of INF-DH 2018*, edited by Manuel Burghardt and Claudia Müller-Birn. Berlin: Gesellschaft für Informatik. https://doi.org/10.18420/infdh2018-07.
———. 2019. "Historical Models and Serial Sources." *Journal of European Periodical Studies* 4 (1): 8–18. https://doi.org/10.21825/jeps.v4i1.10226.
Resnik, Michael D. 1999. *Mathematics as a Science of Patterns*. Oxford: Oxford University Press. https://doi.org/10.1093/0198250142.001.0001.
Rheinberger, Hans-Jörg. 2005. "Gaston Bachelard and the Notion of 'Phenomenotechnique'." *Perspectives on Science* 13 (3): 313–28. https://doi.org/10.1162/1063614054774288026.
Russel, Bertrand. 1948. *Human Knowledge: Its Scope and Limits*. London: George Allen & Unwin. https://n2t.net/ark:/13960/t0dw01s44.
Russell, Bertrand. (1949) 1997. "Am I an Atheist or an Agnostic? A Plea for Tolerance in the Face of New Dogmas." In *The Collected Papers of Bertrand Russell: Last Philosophical Testament, 1943–68*, edited by John G. Slater, 11:89–92. London: Routledge.
Turchin, Peter. 2008. "Arise 'Cliodynamics'." *Nature* 454: 34--35. https://doi.org/10.1038/454034a.
Winks, Robin W., ed. 1969. *The Historian as Detective: Essays on Evidence*. New York, NY: Harper & Row.