This version of the contribution has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/978-3-032-08697-6_2. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms.

Model, Corpus, Interpretation: Elements of Computational Hermeneutics

Michael Piotrowski[0000-0003-3307-5386]

Faculty of Arts, Department of Language and Information Sciences, University of Lausanne, Bâtiment Anthropole, 1015 Lausanne, Switzerland
michael.piotrowski@unil.ch

Abstract This paper introduces a conceptual framework for *computational hermeneutics* that rethinks the relationship between interpretation and computation. Rather than treating interpretation as a purely humanistic or subjective act, the framework reconceptualizes it as the construction and manipulation of models—a process amenable to formalization and computational representation. Drawing on epistemology, systems theory, and knowledge representation, the paper highlights how corpora and interpretative models function as phenomenotechnical devices, shaping the very phenomena they aim to analyze. It argues that computational hermeneutics requires neither automation of interpretation nor superficial application of digital tools, but instead demands explicit, structured representations of interpretive processes. This approach enables the integration of reasoning, data modeling, and symbolic computation into traditionally qualitative domains, offering new avenues for human–machine collaboration in knowledge construction.

1 Introduction

Large-scale digitization of sources and research objects has made more information accessible, which can potentially aid in the interpretation of cultural objects, and it has enabled the use of computational approaches—under a variety of labels, such as digital humanities, computational humanities, cultural analytics, cultoromics, cliodynamics, etc.—that can provide additional information. However, the availability of more sources, more data, more analyses renders the task of interpretation also more complex and raises new questions: a larger corpus is not necessarily more representative, and the epistemological impact of computational methods is still open.

This contribution has argued that interpretation, long considered a characteristic of the humanities, can usefully be reframed as the construction and fusion of models. In this conceptual framework of computational hermeneutics, models are not passive representations but phenomenotechnical devices, active components of epistemic systems: they shape the very phenomena they aim to explain, and are in turn shaped by the interpretive contexts in which they appear.

2 On Terminology

Neither "digital hermeneutics," the term used in the title of this volume, nor "computational hermeneutics," as used in the title of this contribution are established designations with a fixed, precise meaning. I will therefore briefly review the main usages of these terms. The term "digital hermeneutics" can be—and has been—interpreted in various ways: as hermeneutics in the Digital Age, as hermeneutics of "the digital" (media, networks, etc.), and as hermeneutics using digital methods and tools. In the last sense, "digital" is closest to its original technical notion, but also ambiguous: does it refer to the underlying representation and storage in binary format? In this case, using a word processor to write an interpretation would count as "digital hermeneutics." Or does it rather refer to the use of *computational* methods?

These interpretations are often related: digital media, social networks, etc., are phenomena of the Digital Age. Their digital representation enables the use of computational methods for their analysis; these can reduce the effort required for traditional methods and enable new types of analysis, such as topic modeling, stylometry, named entity recognition, or sentiment analysis. Mohr et al. [1] point out that "[i]nstead of just content, we are now able to focus on style, and on the ways in which texts are embedded in broader literary conversations." These types of analysis require "larger and more comprehensive datasets," which are now available, but their sheer size effectively makes the use of computational tools imperative if the data is to be analyzed on a corresponding scale.

Regarding the development of (notions of) digital hermeneutics in recent decades, Capurro [2], Mohr et al. [1], and Romele et al. [3] present interesting "snapshots" that document also evolving sensibilities. The title of the last-mentioned paper, *Digital hermeneutics: From interpreting with machines to interpretational machines*, foreshadowed the most recent developments, which point towards an ever important role of machines in the interpretation, and not only of texts, but also of other artifacts. This was already suggested by Busa when he wrote that "[h]umanities computing is precisely the automation of every possible analysis of human expression [. . .], in the widest sense of the word, from music to the theater, from design and painting to phonetics" [4]. Today this sounds no longer far-fetched: computers can not only be used to analyze data and crunch numbers, but also to generate texts that—to put it carefully—resemble interpretations of human expression (to use Busa's term). This raises important epistemological questions; or else, it restates them, since they have already been raised, for example by Klaus [5] or Weizenbaum [6], or as Capurro [2] put it: "Who are we in the digital age?" For example, Stork [7] notes that "in the past few years, rigorous automated image analysis has assisted some art historians, critics, and connoisseurs in their scholarly studies of fine-art paintings and drawings." He points out that this type of image analysis is "rather different from traditional 'digital humanities,' which has generally concentrated on digital methods of capture and display but where the fundamental analyses and interpretations are still performed by human scholars and connoisseurs." He stresses that "artworks pose several profound problems that require sophisticated methods beyond those in traditional digital humanities," in fact, they "represent a grand challenge to AI, well beyond what is generally addressed in research in digital humanities programs and even mainstream artificial intelligence" [7].

Whatever one may think of this, it is indeed true that in "traditional digital humanities" the computer analysis is thought to precede the actual interpretation: the computer merely takes over the mechanical work and provides statistics and analyses of various kinds, visualizations, and so on. An example of this is *Voyant* [8], even though the authors' thoughts on "computer-assisted interpretation in the humanities" (the subtitle of the book) are actually more profound.

Such a "division of labor" intuitively makes sense, but it is inherently limited and there is, as van Zundert [9] warns, a danger of reducing "the hermeneutical act to a post-processing of what remains of data after the processes of curation, analysis, and visualization." He stresses that "those processes of curation, analysis, and visualization have a hermeneutics of their own" and that one cannot "simply posit a dichotomy between the quantitative and qualitative, and relegate hermeneutics to a qualitative aspect of interpretation of given data as if this data would not be value-laden and interpreted already." Indeed, as Meunier [10] points out, already the digitization of a text implements interpretative decisions.

To avoid this false dichotomy, we need to conceptualize hermeneutics differently: to adequately handle computer-assisted interpretation with all its quantitative and qualitative, formal and informal, human and machine aspects and stages, we need to *formalize* the hermeneutic process. We will go into more detail in the next section, but I hasten to underline that formalization of the hermeneutic process does *not* imply the automation of the interpretation itself.

3 Computational Hermeneutics

The term "digital hermeneutics" often implies the use of computational methods at some point. Nevertheless, it is much less ambiguous to use the term *computational hermeneutics* for approaches that aim to formalize the hermeneutic process. This reasoning is consistent with the arguments that have been made for computational rather than digital humanities [e.g., 11, 12].

Historically, computational hermeneutics (as I understand it) can in part be traced back to research on natural-language understanding in computational linguistics, cognitive science, and (symbolic) AI; this includes Turing's famous paper *Computing machinery and intelligence* [13], which introduced what is today known as the "Turing test," Weizenbaum's ELIZA [14] and Winograd's SHRDLU [15]. The connection to hermeneutics, however, seems to have been realized only in the 1980s; Romele et al. [3] assert that "digital hermeneutics began with Mallery, Hurwitz, and Duffy's 'Hermeneutics: From Textual Explication to Computer Understanding' (1986)" [16]. Some of the authors cited in this report, in particular Dreyfus [17] and Winograd [18], draw on hermeneutics as a philosophical framework to reflect on the possibility of machine understanding; the work most directly related to concrete issues of interpretation as understood in the humanities is by Lehnert et al. [19], who already used the term "computational hermeneutics" to describe their work.

Today, computational linguistics has to a large extent been displaced by engineering-style natural language processing (NLP), and symbolic approaches have been replaced by large language models (LLMs). The question of machine understanding is more relevant than ever, but the current discussion prompted by LLMs seems to focus mostly on "understanding" as an emergent property, revisiting Searles's Chinese Room [20].[1] There is, however, recent work in philosophical logic under the name of "computational hermeneutics" by Fuenmayor and Benzmüller [21–23] on logical analysis of natural-language arguments based on automated theorem proving and argumentation theory. This type of computational hermeneutics may seem unrelated at first, and Elkind [24] comments that the name is "slightly unfortunate" because it is used to describe "hermeneutic practices involving very different computational tools"; he cites Mohr et al. [1], Rockwell and Sinclair [8], and Piotrowski and Neuwirth [25] to argue that "this alternative usage seems to be well-established among digital humanists."

This is a misunderstanding, though.[2] In fact, Fuenmayor and Benzmüller's computational hermeneutics is part of a long tradition of formal approaches that one may ultimately trace back to Leibniz's *characteristica universalis*. In digital humanities, comparable approaches have so far been found primarily in knowledge representation; in particular, $\mathrm{CRM}_{\mathrm{inf}}$ [28, see also 29], an extension of CIDOC CRM [30],[3] which aims to make explicit the processes of argument-making and the states of belief at a particular point in time in a composite inference, and to explicitly connect them to a domain ontology. One could perhaps say that Fuenmayor and Benzmüller's computational hermeneutics is more formal than what is typically found in the digital humanities, but this is primarily because its scope is much more strictly delineated.

The "missing link," as one might say, are formal approaches in the humanities, which are largely forgotten today. My thinking is particularly inspired by Granger [31] and by Gardin's *logicist approach* [see, e.g., 32, 33], including attempts to implement it using symbolic AI techniques [34, 35].

Formalization is often met with hostility in the humanities: many scholars associate formalization and computation with mathematization, and as Granger [31] observes, "for many who work in the humanities, mathematization is equivalent

[1] The Chinese Room is a thought experiment presented by Searle to argue that a computer executing a program cannot understand or think, regardless of how human-like its behavior may seem.

[2] A misunderstanding compounded by Elkind [24] associating the usage of "computational hermeneutics" in digital humanities with that of Harnad [26]; in fact, Harnad's *Against Computational Hermeneutics*—which didn't have the catchy title when it was originally published [27]—uses "hermeneutics" as a quasi-derogative term to criticize the computationalist view in cognitive science as merely projecting an interpretation (this is what he means by "hermeneutics") onto otherwise intrinsically meaningless symbols.

[3] The CIDOC Conceptual Reference Model (CRM) provides an extensible ontology in the field of cultural heritage and museum documentation.

to the introduction of quantity, indeed of number." Concerning hermeneutics in particular, many humanities scholars "subscribe to the notion that interpretation can in principle never come to a conclusion, and indeed the fascination of hermeneutics seems to lie in its inherent incompleteness" [25]. Gadamer famously describes interpretation as "fusion of horizons" [36, transl.: 37]; this description is possibly more useful than it seems, but is less than ideal as a starting point for making the notion of hermeneutics more explicit.

My conceptual view of hermeneutics is as follows. In order to make sense of the world, we build models—in fact, everything we do is based on models: decisions, actions, etc. [38]. This goes for everyday life as for science; I have already quoted Granger's definition of science as "the construction of coherent and effective models of the phenomenon" [31, transl.: 39]. This includes the production of what we today consider cultural artifacts, the main object of interpretation in the humanities. "Understanding" these artifacts involves the *reconstruction* of the models underlying its original construction. But understanding itself is the construction of a model, i.e., *interpretation is model construction*.

A naïve interpretation (e.g., by a child) of a historical painting such as Rembrandt's *Night Watch* just aims to make sense of the perception by using the viewer's world model. So, a child may have a notion of pictures, because they have learned about pictures and they are drawing pictures themselves. They recognize human figures, animals, and natural surroundings. The clothing is strange, but a child familiar with carnival may interpret it as fancy dresses.

A scholarly interpretation, however, aims for more, notably for an understanding of the meaning of the artwork in its *original context*, its cultural significance, the creator's intentions, etc. This requires a model of "the past" in general, of the specific period in particular, etc. The model that is being constructed for the interpretation of some historical artifact builds on the general model, e.g., to relate an individual's social position to society. However, it is important to note that any model of the past is highly deficient, and a model of a specific artifact is similarly limited. Any model of the past is crucially dependent on our models of the present: we fill in the gaps by assuming a shared *conditio humana* and extrapolating from our own experience. In this framework, Gadamer's "fusion of horizons" can be thought of as a fusion of models: our model of the present (essentially what Gadamer calls "predjudices" [36, transl.: 37]), our model of the past, our model of a specific artifact (or creator, phenomenon, etc.)

3.1 Model

The notion of *model* is central to my conception of hermeneutics. It is often affirmed that "[o]ne of the core practices of DH research is indeed modelling" [40], typically with a reference to McCarty [41]. There are, however, two problems. First, due to "the indisputable Anglo-American hegemony" [42] in mainstream DH, earlier approaches tend to be largely ignored, in particular Gardin's logicist approach [43], and approaches influenced by him [e.g., 44, 45]. A case in point, Ciula et al. [40] credit the "growing interest" in modeling in digital humanities to McCarty [41].

Second, the dominant view is that modeling is "one of the core practices," in the sense that it is a neatly delimited activity, which is primarily necessary for the computer. I, however, maintain that modeling is a *fundamental* activity, not something that is specific to digital humanities. As Epstein [46] has remarked, the choice "is not whether to build models; it's whether to build explicit ones." Even if it is rarely acknowledged, the construction of models is not new in the humanities; what is new, is the construction of explicit, formal, and specifically computational models.

Let me explain in some more detail 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, who wrote that "*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" [47, emph. orig.].

Again, this does not just apply to science. Science, however, aims to build *intersubjective models* that can be communicated and tested against "reality." Dupuy [48] remarks that despite much debate in history and philosophy of science about the nature and foundations of scientific knowledge, "the fact that science as an *activity* consists essentially in constructing objects in the form of models is incontestably true, even if it is not as well known among nonscientists as it should be."

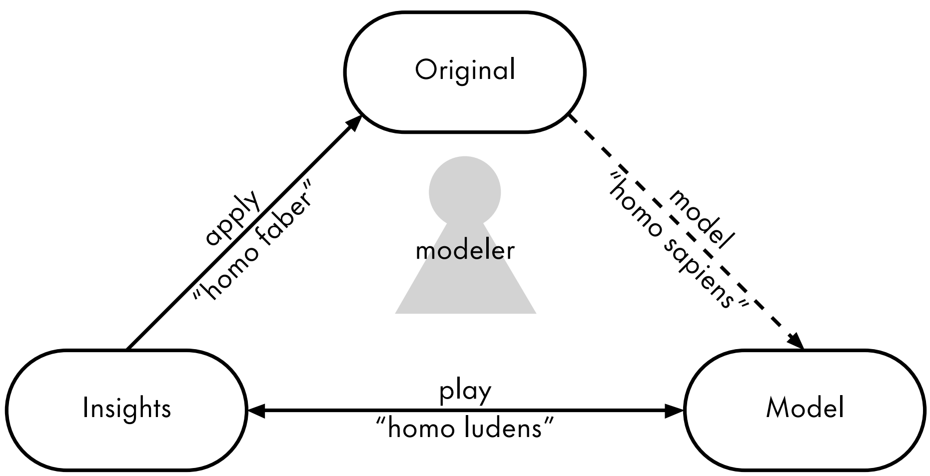

Fig. 1. Modeling triad. The terms *homo sapiens*, *homo ludens*, and *homo faber* are references to the *epistemological control loop* (*erkenntnistheoretischer Regelkreis*) [5].

What we have then is a setting that I call the *modeling triad* (see Fig. 1). First, there is the *original* (in the sense of Stachowiak [49]), i.e., 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 trivially true for historical phenomena. 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 [5], in contrast to *experimenting*, which refers to *strong interaction* [50], 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 has to be evaluated in different ways.

The big question is now: what are models in the humanities, and computational models in particular? 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. For example, Leff [51] observes that "[h]istorians as a profession are not given to constructing or employing models in any formal or explicit sense [...]. Most historians, if asked, would probably deny that models had anything to do with their subject" [51]. 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 needed for technical reasons. In actual research practice in DH, modeling is explicit and formal, as required by computational methods, but often lacking reflection. This becomes especially clear when we consider corpora.

3.2 Corpus

When we look for epistemological differences between "traditional" and digital humanities, *machine-readable corpora* stand out. That they are machine-readable is today usually implied, so people just talk about corpora, and I will use the term in this sense as well, which has been likely inspired by its usage in computational linguistics. Of course, scholars have always created and studied collections of documents and artifacts; indeed, this could be said to be a significant aspect of scholarship.

Rastier [52] asserts that "all texts must be related to a corpus in order to be interpreted" ("tout texte doit être rapporté à un corpus pour être interprété"), whether the corpus is implicit, as it is traditionally in literary studies, or explicit, as in corpus linguistics.

What is clearly new, however, is that corpora are now machine-readable. This has obvious practical advantages: they are easier to store, and they can be processed automatically. What tends to be overlooked, though, is the epistemological impact. After all, it is not some menial task that is being automated: the main results of many DH projects is based on the automatic analysis of some corpus. The availability of a machine-readable corpus is the precondition for the use of computational methods, at least as they are commonly understood, i.e., quantitative methods.

Mayaffre [53] notes that, on the one hand, digitization dematerializes the text in that it is lifted from its previous support, but on the other hand, the construction of a digital corpus materializes it more rigorously than before. He points out that, traditionally, in the humanities corpora tend to be potentialities rather than realities: one *could* go and consult a certain document in some archive, but this may only be rarely done, and the corpus may have never been anything but an "intellectual object" [53].

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 surprisingly little discussion of corpora in DH. In a typical DH paper or project proposal, just a few words are said about the corpus that was used, and most of it tends to concern its size and composition (n items of class X, m items of class Y, and so on) and the technical aspects of its construction (e.g., how it was scraped), if the authors do not simply use an existing corpus. The methods used and the results achieved (and their interpretation and visualization) are typically discussed extensively, though. I would like to just highlight two epistemological problems here.

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 a quantitative, but a *qualitative* difference. 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.

So this all depends on how we analyze our reality—in the literal sense of the Ancient Greek *analysis* "breakup," and there is more than one way to do it. Granger [31] calls this the *découpage*. Granger notes that the *découpage* of human facts—i.e., those studied by the humanities—presents a special difficulty because "the phenomena have an immediate *sense*, which means that they spontaneously take part in a universe of valorized and directed actions," and 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" [39].

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, say, all French-language novels published in the 19$^{\text{th}}$ century. 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 (Fig. 1). Thus, and as I have argued before [54], a corpus should be considered a model in the sense of Apostel [47]. 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?

These are not new questions: every time historians select sources, they construct models, even before any detailed analysis. However, machine-readable corpora are not only potentially much larger than any material collection of sources—which is already not inconsequential—but also have important epistemological consequences. The larger and the more "complete" a corpus is, the greater the danger to succumb to an "implicit essentialism" [55] and to mistake the model for the original, a fallacy that can frequently be observed in the field of cultoromics [56], when arguments are being made on the basis of the Google Books Ngram Corpus.

The same then goes for any analysis of a corpus: if the corpus is "true," so must be the results of the analysis; if there is no evidence of something in the corpus, it did not exist. This allure is even greater when the analysis is done automatically and in particular using opaque quantitative methods: as the computational analysis is assumed to be completely objective, there seems to be no reason to question the results—they merely need to be interpreted.

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 typical case of a corpus-based digital humanities project thus resembles an *experimental system* [57]—an appearance supported by the laboratory metaphor typical for the DH. However, this appearance is deceptive, as ultimately research objects, research questions, and research methods differ substantially from the natural sciences. The crucial point is that corpora are not naturally occurring but intentionally constructed. A corpus is *already* a model and thus not epistemologically neutral.

Mellet [58] highlights how far linguistics, the discipline that arguably pioneered the use of machine-readable corpora, have come in its understanding of corpora: while in 1992, Leech [59] still claimed that the "data of a corpus [...] are independent of the tenets of the theory they are required to test," now (i.e., in 2002), "the corpus has become a conscious mediation between the researcher and the linguistic fact; its construction is thus based on explicit procedures that are an integral part of the hermeneutics implemented" (my translation). I think digital humanities is still lagging behind in this respect.

Here we build on Gaston Bachelard's notion of *phénoménotechnique* [60]. 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" [61]. Corpora are then, methodologically speaking, phenomenotechnical devices and form the basis and influence how we build, understand, and research higher-level concepts.

Thus, a corpus *produces* the phenomenon to be studied. As a model, it has Stachowiak's three characteristics of models: the *characteristic of mapping*, the *characteristic of shortening*, and the *characteristic of pragmatical model-function* [49].[4] Note also that while a model does not have all properties of its corresponding original, it has *abundant attributes* [49], i.e., attributes that are not present in the original. In corpora, this includes, for example, annotation—but also the mutual relationship of the elements that constitute the corpus.

We still do not fully understand how corpora function as models when they are *not* statistical samples: statistics provide us with means of formally describing and analyzing a specific subclass of models capable of representing originals with particular properties. However, the phenomena studied by the humanities do not generally have these properties, and the corpora used in the digital humanities are often not statistical samples; consequently, we do not have adequate formal methods to describe them.

3.3 Interpretation

What is interpretation then? I already suggested above to consider interpretation as the construction and fusion of models. The objective of computational hermeneutics is then to formalize the models and their processing.

This is of course a high-level conceptual view and should not be understood to mean that interpretation can be reduced to a purely mechanical, "objective" process. Formalization can render certain aspects in some sense more "objective." In particular, what could be called "objective" is the *execution* of code by machines: the results do not depend on who launched the execution, and they will be the same on any machine that correctly implements the required interpreter. However, the code is *not* objective: it is a human interpretation of the world written in a formal language. Weizenbaum [6] points out:

> Computers make possible an entirely new relationship between theories and models. I have already said that theories are texts. Texts are written

[4] The English terms are from Stachowiak [62], but there are a number of issues with the translation, and only *Allgemeine Modelltheorie* [49] should be considered authoritative.

> in a language. Computer languages are languages too, and theories may be written in them. [...] Theories written in the form of computer programs are ordinary theories as seen from one point of view. [...] But the computer program has the advantage not only that it may be understood by anyone suitably trained in its language, just as a mathematical formulation can be readily understood by a physicist, but that it may also be run on a computer. [...] A theory written in the form of a computer program is thus both a theory and, when placed on a computer and run, a model to which the theory applies.

The ideas can be readily generalized. In physics, the term "theory" has a fairly precise meaning; a more neutral term like "metamodel" would also emphasizes that we are really dealing with a hierarchy of models that build onto each other and together form what could be called "prejudice" in Gadamer's terms. In any case, a theory in this sense does not necessarily need to be mathematized. For example, a historian's understanding of the Middle Ages could be an example of such a "theory," and this theory is put to work to interpret a *specific* research question, resulting in a model in the framework of the theory that explains the studied phenomenon and leads to a better understanding of both the phenomenon, but also generally, when it is integrated into the theory. As Ricœur [63] put it, "expliquer plus, c'est comprendre mieux," i.e., to explain more is to understand better.

The goal of interpretation is to create meaning, but as Cilliers [64] points out, "[m]eaning is conferred not by a one-to-one correspondence of a symbol with some external concept or object, but by the relationships between the structural components of the system itself." In our framework, the system is constituted—in part—by the set of models, as well as the model creators and model users.

Cilliers stresses that meaning is the result of a dialectical process that involves elements from inside and outside the system, "as well as historical, in the sense that previous states of the system are vitally important." The process of interpretation, even when partially formalized, takes place in an active, open, and complex system.

4 Relation to the IVIS4DH Reference Model

What is the relationship between the framework outlined above and the IVIS4DH Reference Model [65]? First of all, it is important to realize that they are at different levels of abstraction: the framework described here is purely abstract, whereas the IVIS4DH Reference Model aims to describe real research workflows at a much less abstract level, where specific tasks and activities, such as reading, annotation, classification, etc., can be identified. The Reference Model is therefore useful for the design of concrete systems that support "hermeneutic" research workflows: it focuses on the processing stages that the research material undergoes. As such, it is structurally typical of models in computing, and indeed the authors remark that it was developed on the basis of a generic model from computer science.

The IVIS4DH Reference Model should thus not be taken as describing interpretation *per se*. It is concerned with the observable workflow that ultimately leads up to some interpretation, but it does not claim that the various steps *causally* produce an interpretation.

The IVIS4DH Reference Model is concerned with the material aspects of interpretation, whereas the conceptual framework outlined in this contribution is situated on a much higher level of abstraction. In this sense, they could be said to be complementary. However, the IVIS4DH Reference Model focuses on actions, and the models created and manipulate during the workflow remain largely implicit, or are reduced to operations on data structures. This is perhaps the most obvious limitation of the IVIS4DH Reference Model in its current state.

5 Conclusion

In this contribution, I have argued that interpretation, long considered a characteristic of the humanities, can usefully be reframed as the construction and fusion of models. This conceptual framework of computational hermeneutics considers models not passive representations but as phenomenotechnical devices, active components of epistemic systems: they shape the very phenomena they aim to explain, and are in turn shaped by the interpretive contexts in which they appear.

Viewed through a systems-theoretical lens, interpretation becomes a recursive, mediated process embedded within a larger network of observers, representations, and instruments. Just as cybernetics emphasizes the feedback loop between system and observer, so too does hermeneutics—when formalized computationally—reveal its circularity: the models we build inform our understanding, which then reshapes the models themselves.

This systems perspective underscores that interpretation is always situated, contingent, and iterative. Rather than seeking universal methods or deterministic outputs, this view of computational hermeneutics invites us to engage with interpretation as a dynamic process unfolding within open, complex systems. Formalization in this context is not a reductive gesture, but a way of making the hermeneutic process more transparent, communicable, and amenable to interdisciplinary collaboration.

By integrating formal, specifically computational modeling with the reflexivity of the humanities, computational hermeneutics offers not only a new, more rigorous approach to interpretation, but a deeper understanding of how meaning is generated through recursive, context-sensitive modeling processes.

Acknowledgments. This paper is based on work supported by the Swiss National Science Foundation (SNSF) in the project *Towards Computational Historiographical Modeling: Corpora and Concepts* (grant no. 204305). I thank the anonymous reviewers for their constructive feedback.

Disclosure of Interests. The author declares no competing interests.

References

1. Mohr, J.W., Wagner-Pacifici, R., Breiger, R.L.: Toward a computational hermeneutics. Big Data & Society. 2, 1–8 (2015). https://doi.org/10.1177/2053951715613809.
2. Capurro, R.: Digital hermeneutics: An outline. AI & Society. 25, 35–42 (2009). https://doi.org/10.1007/s00146-009-0255-9.
3. Romele, A., Severo, M., Furia, P.: Digital hermeneutics: From interpreting with machines to interpretational machines. AI & Society. 35, 73–86 (2020). https://doi.org/10.1007/s00146-018-0856-2.
4. Busa, R.A.: Foreword: Perspectives on the digital humanities. In: Schreibman, S., Siemens, R., and Unsworth, J. (eds.) A companion to digital humanities. pp. xvi–xxi. Blackwell, Oxford (2004). https://doi.org/10.1002/9780470999875.fmatter.
5. Klaus, G.: Kybernetik und Erkenntnistheorie. VEB Deutscher Verlag der Wissenschaften, Berlin (1966).
6. Weizenbaum, J.: Computer power and human reason. W.H. Freeman, San Francisco, CA, USA (1976).
7. Stork, D.G.: Computer vision, ML, and AI in the study of fine art. Communications of the ACM. 67, 68–75 (2024). https://doi.org/10.1145/3633454.
8. Rockwell, G., Sinclair, S.: Hermeneutica: Computer-assisted interpretation in the humanities. MIT Press, Cambridge, MA (2016).
9. van Zundert, J.J.: Screwmeneutics and hermenumericals. In: Schreibman, S., Siemens, R., and Unsworth, J. (eds.) A new companion to digital humanities. pp. 331–347. Wiley, Hoboken, NJ (2016). https://doi.org/10.1002/9781118680605.ch23.
10. Meunier, J.-G.: Le texte numérique: Enjeux herméneutiques. Digital Humanities Quarterly. 12, (2018).
11. Meunier, J.-G.: Humanités numériques ou computationnelles: Enjeux herméneutiques. Sens public. (2014).
12. Piotrowski, M.: Schism or renaissance? On the relationship between computational humanities and digital humanities. In: Nunn, C.A. and Oorschot, F. van (eds.) Compendium computational theology: Introducing digital humanities to theology. pp. 31–49. heiBOOKS, Heidelberg (2024). https://doi.org/10.11588/HEIBOOKS.1521.C21938.
13. Turing, A.M.: Computing machinery and intelligence. Mind. LIX, 433–460 (1950). https://doi.org/10.1093/mind/lix.236.433.
14. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM. 9, 36–45 (1966). https://doi.org/10.1145/365153.365168.
15. Winograd, T.: Understanding natural language. Cognitive Psychology. 3, 1–191 (1972). https://doi.org/10.1016/0010-0285%2872%2990002-3.
16. Mallery, J.C., Hurwitz, R., Duffy, G.: Hermeneutics: From textual explication to computer understanding? MIT AI Lab, Cambridge, MA (1986).
17. Dreyfus, H.: What computers can't do: The limits of artificial intelligence. Harper & Row, New York, NY (1979).
18. Winograd, T.: What does it mean to understand language. In: Norman, D.A. (ed.) Perspectives on cognitive science. pp. 231–263. ABLEX, Norwood, NJ, USA (1981).

19. Lehnert, W.C., Alker, H.R., Schneider, D.K.: The heroic Jesus: The affective plot structure of Toynbee's *Christus Patiens*. In: Burton, S.K. and Short, D.D. (eds.) Sixth international conference on computers and the humanities. pp. 358–367. Computer Science Press, Rockville, MD (1983).
20. Searle, J.R.: Minds, brains, and programs. Behavioral and Brain Sciences. 3, 417–424 (1980). https://doi.org/10.1017/s0140525x00005756.
21. Fuenmayor, D., Benzmüller, C.: A case study on computational hermeneutics: E. J. Lowe's modal ontological argument. In: Sousa Silvestre, R., Göcke, P.B., Béziau, J.-Y., and Bilimoria, P. (eds.) Beyond faith and rationality: Essays on logic, religion and philosophy. pp. 195–228. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-43535-6_12.
22. Fuenmayor, D., Benzmüller, C.: Computational hermeneutics: An integrated approach for the logical analysis of natural-language arguments. In: Liao, B., Ågotnes, T., and Wang, Y.N. (eds.) Dynamics, uncertainty and reasoning: The second chinese conference on logic and argumentation. pp. 187–207. Springer Singapore, Singapore (2019).
23. Fuenmayor, D., Benzmüller, C.: A computational-hermeneutic approach for conceptual explicitation. In: Nepomuceno-Fernández, Á., Magnani, L., Salguero-Lamillar, F.J., Barés-Gómez, C., and Fontaine, M. (eds.) Model-based reasoning in science and technology. MBR 2018: Inferential models for logic, language, cognition and computation. pp. 441–469. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32722-4_25.
24. Elkind, L.D.C.: Computer verification for historians of philosophy. Synthese. 200, 1–28 (2022). https://doi.org/10.1007/s11229-022-03678-y.
25. Piotrowski, M., Neuwirth, M.: Prospects for computational hermeneutics. In: Marras, C., Passarotti, M., Franzini, G., and Litta, E. (eds.) Atti del IX Convegno Annuale AIUCD: La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica. pp. 204–209. Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD) (2020). https://doi.org/10.6092/UNIBO/AMSACTA/6316.
26. Harnad, S.: Against computational hermeneutics. Invited commentary on Eric Dietrich's Computationalism. Social Epistemology. 4, 167–172 (1990).
27. Balogh, I., Beakley, B., Churchland, P., Gorman, M., Harnad, S., Mertz, D., Pattee, H.H., Ramsey, W., Ringen, J., Schwarz, G., Slator, B., Strudler, A., Wallis, C.: Responses to "Computationalism." Social Epistemology. 4, 155–199 (1990). https://doi.org/10.1080/02691729008578567.
28. Doerr, M., Fafalios, P., Kritsotaki, A., Stead, S., Ore, C.-E.: Definition of the CRMinf: An extension of CIDOC CRM to support argumentation. Paverprime Ltd.; FORTH (2023).
29. Doerr, M., Kritsotaki, A., Boutsika, K.: Factual argumentation—a core model for assertions making. Journal on Computing and Cultural Heritage. 3, 8:1–8:34 (2011). https://doi.org/10.1145/1921614.1921615.
30. CIDOC CRM Special Interest Group: Definition of the CIDOC Conceptual Reference Model. The International Committee for Documentation of the International Council of Museums (ICOM–CIDOC), Paris (2024).
31. Granger, G.-G.: Pensée formelle et sciences de l'homme. Aubier-Montaigne, Paris (1967).
32. Gardin, J.-C.: Archaeological constructs: An aspect of theoretical archaeology. Cambridge University Press, Cambridge (1980).

33. Gardin, J.-C.: Le questionnement logiciste et les conflits d'interprétation. Enquête. 5, 35–54 (1997). https://doi.org/10.4000/enquete.1043.
34. Ennals, R.: Artificial intelligence: Applications to logical reasoning and historical research. Ellis Horwood, Chichester (1985).
35. Ennals, R., Gardin, J.-C. eds: Interpretation in the Humanities: Perspectives from Artificial Intelligence. British Library, London (1990).
36. Gadamer, H.-G.: Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik. In: Gesammelte Werke. Mohr Siebeck, Tübingen (1990).
37. Gadamer, H.-G.: Truth and method. Continuum, London/New York (1989).
38. Forrester, J.W.: Counterintuitive behavior of social systems. (1995).
39. Granger, G.-G.: Formal thought and the sciences of man. Reidel, Dordrecht (1983).
40. Ciula, A., Eide, Ø., Marras, C., Sahle, P.: Modelling between digital and humanities: Thinking in practice. Open Book Publishers (2023). https://doi.org/10.11647/obp.0369.
41. McCarty, W.: Humanities computing. Palgrave Macmillan, Basingstoke (2005).
42. Fiormonte, D.: Toward a cultural critique of digital humanities. In: Debates in the digital humanities 2016. pp. 438–458. University of Minnesota Press, Minneapolis, MN (2016). https://doi.org/10.5749/j.ctt1cn6thb.38.
43. Gardin, J.-C.: Le calcul et la raison: essais sur la formalisation du discours savant. Éd. de l'École des hautes études en sciences sociales, Paris (1991).
44. Borillo, M.: Informatique pour les sciences de l'homme: Limites de la formalisation du raisonnement. Mardaga, Bruxelles (1985).
45. Orlandi, T.: Informatica umanistica. La Nuova Italia Scientifica, Rome (1990).
46. Epstein, J.M.: Why model? Journal of Artificial Societies and Social Simulation. 11, (2008).
47. Apostel, L.: Towards the formal study of models in the non-formal sciences. In: Freudenthal, H. (ed.) The concept and the role of the model in mathematics and natural and social sciences. pp. 1–37. Reidel, Dordrecht (1961). https://doi.org/10.1007/978-94-010-3667-2_1.
48. Dupuy, J.-P.: On the origins of cognitive science: The mechanization of the mind. MIT Press, Cambridge, MA (2009).
49. Stachowiak, H.: Allgemeine Modelltheorie. Springer, Wien, New York (1973).
50. Moles, A.A.: Les sciences de l'imprécis. Seuil, Paris (1995).
51. Leff, G.: Models inherent in history. In: Shanin, T. (ed.) The rules of the game: Cross-disciplinary essays on models in scholarly thought. pp. 148–160. Tavistock, Abingdon (1972).
52. Rastier, F.: La mesure et le grain: Sémantique de corpus. Champion, Paris (2011).

53. Mayaffre, D.: Philologie et/ou herméneutique numérique: nouveaux concepts pour de nouvelles pratiques? In: Rastier, F. and Ballabriga, M. (eds.) Corpus en lettres et sciences sociales: des documents numériques à l'interprétation. Actes du XXVII^{e} Colloque d'Albi "Langages et signification." pp. 15–25. CALS-CPST (2006).
54. Piotrowski, M.: Historical models and serial sources. Journal of European Periodical Studies. 4, 8–18 (2019). https://doi.org/10.21825/jeps.v4i1.10226.
55. Mothon, B.: Modélisation et vérité. Archétype82, Paris (2010).
56. Michel, J.-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., The Google Books Team, Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M.A., Aiden, E.L.: Quantitative analysis of culture using millions of digitized books. Science. 331, 176–182 (2011). https://doi.org/10.1126/science.1199644.
57. Rheinberger, H.-J.: Experimentalsysteme und epistemische Dinge: Eine Geschichte der Proteinsynthese im Reagenzglas. Wallstein, Göttingen (2001).
58. Mellet, S.: Corpus et recherches linguistiques: Introduction. Corpus. (2002). https://doi.org/10.4000/corpus.7.
59. Leech, G.: Corpora and theories of linguistic performance. In: Svartvik, J. (ed.) Directions in corpus linguistics. Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991. pp. 105–122. De Gruyter Mouton, Berlin; New York (1992). https://doi.org/10.1515/9783110867275.
60. Bachelard, G.: Le nouvel esprit scientifique. Les Presses universitaires de France, Paris (1968).
61. Rheinberger, H.-J.: Gaston Bachelard and the notion of "phenomenotechnique." Perspectives on Science. 13, 313–328 (2005). https://doi.org/10.1162/106361405774288026.
62. Stachowiak, H.: Models. In: Unesco (ed.) Scientific thought: Some underlying concepts, methods and procedures. pp. 145–166. Unesco/Mouton, Paris; The Hague (1972).
63. Ricœur, P.: De l'interprétation. In: Du texte à l'action. Essais d'herméneutique, II. pp. 11–35. Seuil, Paris (1986).
64. Cilliers, P.: Complexity and postmodernism: Understanding complex systems. Routledge, London; New York (1998).
65. Mix, B., Bruchhaus, S., Möbus, D., Leh, A., Nawroth, C., Tamla, P., Störl, U., Hemmje, M.: IVIS4DH: Towards a reference model for digital hermeneutics. In: Digital hermeneutics: Sources, analysis, interpretation, annotation, and curation. Springer, Cham (2025).