

## ARTICLE

# From the Archive to the Computer: Michel Foucault and the Digital Humanities

Henning Schmidgen<sup>1</sup>, Bernhard Dotzler<sup>2</sup>, Benno Stein<sup>1</sup>

<sup>1</sup> Bauhaus University, Weimar, <sup>2</sup> University of Regensburg

Keywords: DH theory, discourse analysis, computer science

<https://doi.org/10.22148/001c.55795>

---

Journal of Cultural Analytics

Vol. 7, Issue 4, 2022

---

Michel Foucault famously introduced the method of “discourse analysis” in the humanities, especially in historiography. In his *Archaeology of Knowledge*, originally published in 1969, in particular, Foucault argues for making the history of knowledge the object of discourse analyses. In the context of the current surge of interest in discourse analysis in the field of computer science, however, there are hardly any references to Foucault, partly because he never defined a methodological process that could be operationalized. Nonetheless we argue for re-reading the *Archaeology of Knowledge* in the context of computer science and the digital humanities. As a matter of fact, there are considerable affinities between Foucault’s search for the regularities of discourse and current projects dealing with the digitization of texts, their indexing, distributional features, stylometry, etc. We show that these projects were already quite prominent in Foucault’s day, to the point that historian Emmanuel Le Roy Ladurie could assert, in 1968, that “the future historian will be a programmer.” A year later, Foucault’s *Archaeology of Knowledge* actively responded and constructively took up the challenge – which, given the recent advances in machine learning and computational linguistics, strikes us as a crucial move today.

## Introduction<sup>1</sup>

At the height of student protests in May 1968, the French historian Emmanuel Le Roy Ladurie made a remarkable announcement in the pages of the weekly magazine *Le nouvel observateur*. Perhaps sympathizing with the students in the streets, Le Roy Ladurie postulated *la fin des érudits*, the “end of the scholars.”

The occasion and basis of this prophecy, however, was not the crisis of the university but the increasing use of computers in historical research. Pointing to examples from the United States and France, he explained that for historical projects employing information technology, promising perspectives were opening up: “One of the clearest directions is the analysis of vast *corpora* of documents whose data are of capital importance but whose scope has so far thwarted researchers’ efforts” (Le Roy Ladurie, “La fin des érudits: L’historien de demain sera programmeur ou ne sera pas” 3).

---

<sup>1</sup> This paper was written as part of the DFG research project “Process-oriented Discourse Analysis”. <https://gepris.dfg.de/gepris/projekt/326264959>. It is based on a comprehensive study on “Discourse Analysis in the Age of Intelligent Machines” by Bernhard Dotzler and Henning Schmidgen recently published in German (Dotzler and Schmidgen). While the three authors of the present paper share its general argument, they happily disagree on some of its details. We would like to thank Tim Gollub, Franziska Klemstein, and Johannes Hess for helpful suggestions and critical comments.

Le Roy Ladurie cites as his example the work of the US-American medievalist David Herlihy, who, starting in the 1950s, made use of computers to study Italian land registers. He could just as well have chosen his own work on the farmers in the Languedoc from 1966 or other work done in the orbit of the *Annales* school that had deployed electronic computing devices to process historical and sociological data. The conclusion Le Roy Ladurie draws from these projects was to make waves: “[t]he historian of tomorrow will be a programmer or he won’t be at all” (3; see also 1973).<sup>2</sup>

A year later, in 1969, Michel Foucault publishes the volume that to this day is being referred to as his book of method, *The Archaeology of Knowledge*. On the very first page, it evokes a central theme of the *Annales* school, the so-called *longue durée*. Foucault writes, for example: “For many years now historians have preferred to turn their attention to long periods” (Foucault, *Archaeology* 3).<sup>3</sup>

The allusions are clear later in the book as well, for example when Foucault explains that “the building-up of coherent and homogeneous corpora of documents” (10) is a decisive methodological problem of contemporary historical studies, or when, in specifying new methods in historical research, he lists “the quantitative treatment of data, the breaking-down of the material according to a number of assignable features whose correlations are then studied, interpretative decipherment, analysis of frequency and distribution” (11). Our argument is that Foucault here refers to the deployment of computers in the humanities and social sciences that Le Roy Ladurie had so prominently described a short time earlier.

This reference might come as a surprise, not only because it shifts the first flourishing of what we now call digital humanities fifty years into the past (Burdick et al. 121–36; Sterne 17–33). It also seems to jar with the popular image of Foucault frequenting the archives or at least regularly visiting the *Bibliothèque nationale*. But a different, more fitting image emerges when we take the historical, institutional, and intellectual context more fully into account. As early as the 1950s, Foucault was interested in “the statistical theory of information” (Foucault, “La psychologie” 136); in 1966, he speculated about bringing together “the analysis of languages” and “information processing” (Foucault, “Message ou bruit” 560); and a little later, he observed that “in the thickness [*épaisseur*] of natural processes . . . the structure of the message,” that is to say, ultimately, encoded information, could be discovered (Foucault, “Hyppolite” 784).

---

2 On the use of computing machinery in the *Annales* school, see Burke 53–64, and Lemny. For contemporary descriptions, see Price and Furet.

3 On this book, see, for example, Webb and 1999.

These and other remarks by Foucault become understandable against the backdrop of the flourishing of cybernetics and molecular biology in 1960s France (Erdur; Kay 275–76; Geoghegan). Prepared by the intense engagement with the relationship between life, language, and technology that shapes the work of Foucault’s most important academic mentors, Jean Hyppolite and Georges Canguilhem, Foucault sets out in *The Archaeology of Knowledge*, to sketch, retrospectively, the methodology that his analyses of the history of psychiatry, clinical medicine, biology, political economy, and linguistics had followed—or, rather, the methodology these studies had sketched over time.

As far as we know, Foucault never made use of a computer to conduct his discourse analyses. Personal computers for private use were not developed until the mid-1970s. His elaboration of the “archaeological” method, however, reflects the emerging automation of such analyses. At the same time, his reflections contain remarkable parallels with the so-called distributional hypothesis, which has been a central element in statistical semantics since the pioneering work of the linguist Zellig Harris and ever since it has been popularized by John Rupert Firth.

Computer science today summarizes this hypothesis by formulas such as “[w]ords which are similar in meaning occur in similar contexts” (Rubenstein and Goodenough 627; see also Harris). The Foucault of the *Archaeology* would not disagree. That is why the procedure he sketches—even if its methodological status remains unclear—can contribute to stimulating cooperation between computer scientists and scholars in the humanities and social sciences today.

In the following, we discuss some issues that could or – according to us – should be of interest in this cooperation. Essential elements here are the understanding of “discourses,” i.e. large aggregates of utterances, as central components of cultural and social life; the problem of their delimitation in recourse to the objects specific to them; and the question of the possibility of automating the delimitation and analysis of such discourses by means of computer technology.

Thus, we are not concerned with an application or operationalization of Foucault’s method, and we are not arguing that all practitioners in the digital humanities should read Foucault. Foucault’s concept of discourse is notoriously underdetermined and its meaning also changes over time. What the *Archaeology of Knowledge* does accomplish, however, is a sophisticated discussion of fundamental problems with the use of computer technology in the humanities, especially historiography. Our argument is that those digital humanities scholars can benefit from this discussion who are theoretically interested and who are willing to share that theoretical interest with other humanities scholars, even if the latter are not, or not yet, working in the digital field.

We proceed in three steps. In parts 1 and 2 we present the method that Foucault describes in the *Archaeology of Knowledge*. Special attention is given on the one hand to the discursive aspects of *object*, *style*, *concepts* and *themes* and on the other hand to the regularity of discourses. In parts 3 and 4 we present and discuss two historical attempts to automate individual aspects of discourse analysis. We focus on the General Inquirer developed by a team around the Harvard psychologist Philipp Stone around 1966 as well as the program developed by French linguist and philosopher Michel Pêcheux in the late 1960s to investigate the “deep structure” of discursive effects.

Against this background, thirdly, we discuss the perspectives of today’s automation of discourse analysis. Our conclusion is that, while Big Data and Machine Learning have significantly contributed to improve some aspects of automated discourse analysis, tasks such as the definition of research questions or the delimitation of research objects, as well as the interpretation of research results, still belong to the historian – or rather the archaeologist in Foucault’s sense.

## 1. Aspects of Foucault’s Method

The method that is spelled out in *Archaeology of Knowledge* can be presented in view of four aspects. First, as regards the starting point of the method, there is what Foucault repeatedly calls the “dispersion” of discursive events. Thus he declares discourse to be a “vast field” “made up of the totality of all effective utterances [*énoncés*] (whether spoken or written) in their dispersion as events and in the occurrence that is proper to them” (Foucault, *Archaeology* 26–27; translation amended).

The repeated evocation of the dispersion of discourse is due to Foucault adopting a statistical (in the broadest sense) perspective on discursive events (see also Herrmann 62–67). Instead of starting from individual historical actors (persons, authors), works, institutions, or disciplines, he places a mass of distributed discursive events (and in that sense, they are indeed strictly linguistic data) at the beginning. This perspectivization is joined by an epistemological motif still present in today’s debates about the digital humanities. When historians (or, to speak with Foucault: archaeologists) confront discourse as a set of data, they find themselves up against, according to Foucault, “linguistic sequences that . . . in sheer size, exceed the capacities of recording, memory, or reading” (Foucault, *Archaeology* 27).

At the same time, this avoids misperceptions and misjudgments that come with the position of the individual reader and his or her limited (if not in principle, then in practice) capacities. They dismiss or, as Foucault puts it, “eclipse . . . that form of history that was secretly, but entirely related to the synthetic activity of the subject” (14)—we might consider this to be Foucault’s version of the end of the scholar as we knew her, proclaimed by Le Roy Ladurie.

While the position sketched in *The Archaeology of Knowledge* is not anthropocentric, it remains a difficult question how order is to emerge again from the overwhelming quantity of discursive events. This is our second point: how, with the aid of which criteria and procedures, entities that can be studied at all are to be delineated in the sheer mass of discursive data.

In his televised debate with Noam Chomsky, Foucault is clear that the entire archaeological endeavor aims at investigating comparatively circumscribed discursive sets. He explains his interest in the discourse analysis of scientific knowledge by citing the history of medicine in the late eighteenth century:

[R]ead twenty medical works, it doesn't matter which, of the years 1770 to 1780, then twenty others from the years 1820 to 1830, and I would say, quite at random, that in forty or fifty years everything had changed; what one talked about, the way one talked about it, not just the remedies, of course, not just the maladies and their classifications, but the outlook itself. (Foucault and Chomsky 150)

The point where Foucauldian archaeology sets in, then, are breaks in discourse, fundamental changes in scientific utterances, that is, abrupt transformations in the schemata according to which words, parts of sentences, and finally entire texts are constructed in this domain—changes in “paradigms” in the linguistic, not in Thomas Kuhn’s sense of the term.

Foucault follows this remark up with the question: “Who was responsible for that? Who was the author of it? It is artificial, I think, to say Bichat, or even to expand a little and to say the first anatomical clinicians. It’s a matter of a *collective and complex transformation* of medical understanding in its practice and its rules” (Foucault and Chomsky 150; emphasis added).

At issue, then, are not individual and punctual discoveries, not individual scientists or authors but—not unlike in Kuhn—overarching changes in dominant forms of perception and procedures. Foucault’s *Archaeology of Knowledge* aims at describing, closely studying, and, as far as possible, explaining such collective and complex transformations on the level of discourses, that is, of actual utterances.

## 2. Discursive Regularities

The off-the-cuff remark on twenty medical books from different epochs is translated, in the *Archaeology*, in a complex schema that—our third point—includes, besides the *object* of a discursive formation, the questions of *style*, of *concepts*, and of the overarching *themes*. This schema we cannot discuss here in detail, but we can point out that two of its aspects—style and thematic—already played an important role in the digital humanities of the 1960s. The analysis of themes was the goal, for example, of the General Inquirer developed by a team around the Harvard psychologist Philipp Stone.

This computerized procedure for analyzing textual content, presented in 1966, soon garnered attention among the people then in Foucault's orbit (see Helsloot and Hak 78).

As far as "style," is concerned, it is often understood in the digital humanities at the time as "difference in frequency distribution and matrices of transition probability of a text's linguistic units from the corresponding [units] of language as a whole" (Müller 161). Foucault, as noted earlier, is not interested in the question of individual authorship, which is important, if not decisive, in stylometry to this day. In the *Archaeology of Knowledge*, though, he *is* interested in the "frequency and distribution" of historical data (Foucault, *Archaeology* 11), and what draws his attention is the "distribution" of objects in a discourse, "the interplay of their differences, . . . their proximity or distance" (46).

This brings us to his interest in the internal organization of discourse, in the rules that the utterances in a certain age and about a certain object follow. This is our fourth and final point. In *The Archaeology of Knowledge*, this interest takes the guise of the question whether the utterances of a discursive formation are organized *in* this formation or whether they might not be specifically organized *by* it, and whether they can be said to follow specific rules: "an order in their successive appearance, correlations in their simultaneity, assignable positions in a common space, a reciprocal functioning, linked and hierarchized transformations" (37).

Foucault accordingly sets out in search of the "intrinsic regularities of discourse." Rules, he never tires to emphasize, are not situated behind or above discourses but "at the most 'superficial' level (at the level of discourse)." They are not located in the consciousness of individuals, nor in a "mentality" of the kind the *Annales* school was working on, "but in discourse itself" (62–63).

In assuming such an immanentist position, Foucault is at the same time moving away from the separation between surface structure and deep structure operated in 1960s linguistics, most concisely by Chomsky. His position clearly is not far from the so-called distributional hypothesis, closely associated with the name of Chomsky's teacher Zellig Harris.<sup>4</sup>

And indeed, Foucault in the *Archaeology* stresses the proximity of "rule" and "regularity." For instance, he describes the entire set of rules of a given discursive practice as a "system of formation," which he wants to be understood as "a complex group of relations" that in turn function as rules for the four entities cited earlier—object, style, concept, thematic:

---

<sup>4</sup> Harris begins with the two entities that are central for Foucault as well: discourse (*discours*) on the one hand and utterance (*énoncé*) on the other. This and other similarities prompt Thomas Pavel (131) to draw a parallel between the conceptual apparatuses of Foucault and distributionalists such as Harris and his students. On this point, see also Dosse 241.

By system of formation, then, I mean a complex group of relations that function as a rule: it lays down *what must be related*, in a particular discursive practice, for such and such an utterance to be made, for such and such a concept to be used, for such and such a strategy to be organized. (Foucault, *Archaeology* 74; translation amended, emphasis added).

The question of the rules of discourse, it seems, thus dissolves in the question of the regularities of relationships between discursive elements. The normative aspect of discourse, we might say, is captured through distributions and relationships that can be determined statistically. Discursive regularity here follows from the frequency of discursive elements.

### 3. Automating Discourse Analysis

At the end of the 1960s, the philosopher and linguist Michel Pêcheux, a member of the *Cercle d'épistémologie* that was close to Foucault, presented a project for automating discourse analysis. Based on, in rough terms, a theory of discourse production as a “theory of the rule-governed variation of ‘deep structures,’” Pêcheux’s automated discourse analysis was concerned with going from a series of discursive “‘surface effects’” to a “‘deep structure,’” an “invisible structure which determines them” (Helsloot and Hak 96).

In the course of its implementation, this endeavor encountered a difficulty that rather resembles the one of Foucault’s *Archaeology* and yet differs fundamentally from it. Foucault’s procedure to determine the regularities of certain sets of discourses consists, it seems, in filtering these sets as such out of a mass of texts within an iterative process of pattern matching and membership recognition. Pêcheux’s analysis of discourse, in contrast, can operate only via corpora defined in advance. These corpora have already been constituted; the issue then is to determine their “deep structure” or “regularity.”

The dilemma that arises here, it seems to us, remains relevant—it belongs to the “theorytellings” of digital humanities today. Discourse analysis under the conditions of today’s technology is concerned with developing a system that would have to be, we might say, an algorithm for algorithm analysis, i.e., a meta-analysis algorithm. From Foucault’s perspective, this system would have to be capable of finding out not only the contents of certain sets of discourse but their regularities—beginning with the ability to filter these sets from an undefined mass of texts in a circular process of rule recognition and membership definition.

In more concrete terms, the four aspects of discursive formations Foucault brings out—object, style, concepts, and thematic—would have to be discovered in a largely automated way. While style and thematic are not uncommon problems in the digital humanities and work on them has clearly progressed in recent years, it is still largely unclear how a discourse can be defined starting from a given object. The definition of objects is a core domain

of scientific discourses, yet discourse analysis in Foucault's sense is far from willing to take conceptual definitions from the individual sciences and make them the basis of its own studies.

On the contrary: the productivity of Foucauldian archaeology very much derives from operating its own definitions of objects in order to open up new perspectives on the emergence of individual sciences—for example by showing how strongly the development of linguistics, biology, and economics in the seventeenth and eighteenth centuries depended on a specific yet largely implicit conception of the object 'human being' (see Foucault, *Order of Things*).

It seems equally unclear how a discourse's central concepts are to be identified if by "central concepts" we are not to mean simply the words most frequently used. There is in fact an essential difference between a scientific concept and a word—a point demonstrated not least of all by Georges Canguilhem, the philosopher and historian of science who shaped Foucault's thinking in important ways. Using as his example the concept of "reflex" in modern physiology, Canguilhem showed that the conception and the definition of the phenomenon designated by this term do not depend on its use. In the seventeenth century, Thomas Willis defined the phenomenon of the organism's reflexive reactions without evoking the optical analogy implied in the term, whereas Descartes spoke of "reflex" without having a precise physiological conception of the phenomenon. It is thus quite difficult to name and to find scientific concepts in a textual corpus in an automated way.

#### 4. Current Problems

The challenges of implementing Foucault's archaeological method can also be described from the perspective of current developments in machine learning, computational linguistics, and Big Data processing technology. We can see both chances for automating discourse analysis and some obvious limits. For tackling the task today, one option is to simplify the goal of discourse analysis altogether, thus making it a less moving target. Note that in today's computational linguistics (or language technology), discourse analysis is understood as "language processing beyond the sentence boundary" in order to "compute information about a text in order to supplement the results of sentence processing (e.g., when supplying a referent for a pronoun from context)" or "to combine sentence-level information to larger units (e.g., when inferring a causal relationship to hold between two sentences)" (Stede 11).<sup>5</sup>

In this context, the required computation is based on the identification of elementary discourse units, which are hierarchically organized and between which relations can be defined. While, on the one hand, discourse analysis becomes feasible under such an interpretation, it is, on the other hand, too

---

<sup>5</sup> Stede also says: "Discourse processing is the acquisition of information about a text, including assigning structural descriptions to it, so that the extraction of information from a text becomes more interesting, more fruitful, or more simple."



weak from a humanities perspective to tackle the scope of discourse-related research questions in the field, e.g., the development of discourses in specific historical periods or geographical areas, their internal changes as well as their relation to broader cultural and/or societal transformations.

Probably we have to adopt the immanentist position of Foucault, namely, to admit that discourse emerges as a function which *cannot be deductively inferred* from features, whatever their nature is. This is the move from linguistic structuralism to linguistic distributionalism (Biemann), which is prefigured in the *Archaeology of Knowledge*.

Today, the potential of this move is illustrated by the amazing progress in research of autoregressive language models. It brought forth the so-called Generative Pre-trained Transformer, GPT, which, in its latest generation is able to produce human-like text. The GPT architecture consists of two coupled recursive neural networks with millions of parameters, which can be effectively trained by the so-called attention mechanism to model human language.<sup>6</sup>

Whereas Pêcheux's model remained tied to a Chomskian quest for underlying grammars or structures,<sup>7</sup> GPT can be considered as one of the most effective and uncompromising manifestations of the distributionalism paradigm as defined by Harris. The technology has the potential to become another option to narrow the gap between the scientifically interesting and the technically feasible outlined above: Given some input sequence (prompt) in the form a clause, a sentence, or a paragraph of text, GPT outputs the most likely sequence of words, where likelihood is based on the associations (the distributional semantics) learned from the training data. When prompting GPT with a specific topic the generated text can be taken as a discourse on this topic, where the generation probability along with the topical distance—both can be quantified reasonably well—are criteria to halt the generation process and delineate a discourse.

Though the instances of the current GPT generation still not convince as discourse generation machines (shortcomings include: no long-distance consistency, no coherent argumentation, logically flawed), there is reasonable hope that future generations can do, this way becoming a new means to study discourse phenomena. However, currently we cannot learn (much) from GPT in terms of discourse *analysis*, and it is an open question whether we ever will. While symbolic text generation approaches apply schemes, heuristic rules, grammars, search strategies, or planning algorithms—a machinery that can similarly be used for text analysis purposes—the text generation principles of deep neural approaches remain implicit and hidden.

---

<sup>6</sup> The most recent version, GPT-3, was released in June 2020 (OpenAI). GPT-3's full version has a capacity of 175 billion machine learning parameters and was trained on 410 billion byte-pair-encoded tokens.

<sup>7</sup> Recall Chomsky's understanding of grammar, as a means to decide grammaticality, sufficiently specific to generate only sentences of the respective language, but also as a system whose rules can be identified by human introspection only (Chomsky 13–14).

## Discussion

What impulses, then, can Foucault's *Archaeology of Knowledge* provide in the further development of the digital humanities? The most important impulse today, perhaps, can be described as the transition from *digital* to *computational humanities*. We propose this term to distinguish between research focusing on digitization (digital humanities) and research focusing on algorithmic text processing with the goal of a semantic analysis (computational humanities), using machine learning and data mining methods among others. Though Foucault's engagement with the *Annales* school's use of computers remained implicit and he seems not to have seriously considered using computers for discourse analysis himself at the time, *The Archaeology of Knowledge* is rather close to this understanding of computational humanities: He envisaged a research agenda that neither concentrates on simply digitizing the cultural heritage (in cooperation with libraries, archives, and museums) nor focuses on analyzing bibliographic (meta) data as to their distribution in time and space—as is being done, in programmatically interesting ways, in certain forms of “macroanalysis” (Jockers).

In shifting attention away from discursive units that are identified by a common object or style, certain concepts, or overarching themes, toward the implicit regularity of discursive formations (and toward the notion that such formations are defined by such regularities alone), Foucault defined a problem that is very unlikely to be solved by means of the digital humanities' tools, which apply to digitized or “born digital” documents, but instead requires a means for the unsupervised learning of meta-learning strategies, tailored to the input data (computational humanities).

In our eyes, this would be one of the “mission-critical” conditions of an automatic discourse analysis: the system would have to be capable of finding out not only the contents of certain sets of discourse but their regularities—beginning with the ability to filter these sets from masses of texts within an iterative process of pattern matching and membership recognition.

Since we do not see a technical solution for this problem, it seems necessary and promising to keep the human in the loop, accepting her as an indispensable part of machine-based discourse analysis, and to provide technology to tighten and improving the connection to the machine.<sup>8</sup> In this regard, we are developing exploratory search technology to empower humanities scholars to deal with Big Data, enabling them to ask research questions that require the analysis of hundreds of books, papers, and other data—to repeat our second point—to bring order to the chaos of discursive dispersion (Gollub et al.). As

---

<sup>8</sup> This is a central goal of our project on “Process-oriented Discourse Analysis”, <https://gepris.dfg.de/gepris/projekt/326264959>.

we have seen, Foucault also highlighted the importance of scholarly expertise, despite his alleged ‘anti-humanism,’ in particular when it comes to define the object constitutive of a given discourse.

In addition to the recent deep learning technologies, which have brought great advances in text generation and semantic analysis, the classic computer science paradigms to problem solving are data parallelization and task parallelization.

The practical benefits and the feasibility of data parallelism in the humanities have been clearly demonstrated (Michel et al.; Kozłowski et al.), foremost by work done in the context of distant reading (Moretti, *Graphs*; Moretti, *Distant*; Underwood). From a problem reduction perspective, distant reading exploits data parallelism to significantly reduce the complexity of the resulting partial solutions so that subsequent analyses can be performed without computational support. Discursive regularities may be envisioned by projecting masses of text into low-dimensional “semantic spaces,” such as timelines, geographic maps, or topic networks, to become visually explored and analyzed by human experts, providing perspectives onto a corpus which is orthogonal to the reading direction of the documents.

Problem reduction via task parallelism becomes effective if subtasks can be executed independently of each other and, if the partial solutions for the subtasks can be combined to form a complete solution. Although the first condition is often only partially satisfied in humanities problems, we consider the second condition to be the more problematic. Task parallelization requires the outcomes of independent analyses to be combined, following a synthesis strategy, such as the celebrated map-reduce scheme known from Big Data processing. While such recombination schemes are successful for regularly structured problems, result aggregation usually fails for the type of research questions the humanities are dealing with. I.e., human expertise and interpretation capabilities are currently indispensable for putting together the solution pieces.

In the late 1960s, *Annales* historian Le Roy Ladurie spoke, with regard to the increasing use of the computer in the humanities, of the ‘end of the scholars.’ Some time later, Foucault, in his *Archaeology of Knowledge*, showed the perspectives and problems that this development entails for the humanities, especially in the realm of discourse analysis. As the question of the delimitation of discourses and the determination of their specific objects shows, the technical achievements of the last 50 years have shifted the problems in this field, but they have not solved them.

Against this background, one would like to confirm Le Roy Ladurie’s statement from 1968 in modified form: ‘The historian of tomorrow will be a computational humanist or she won’t be at all.’ At the same time, however, it should be stated that, for the time being, the historian—or rather, the

archaeologist in Foucault's sense—cannot delegate important tasks to machines, e.g., the definition of research questions, the delimitation of research objects, as well as the interpretation and evaluation of research results.

Submitted: June 14, 2022 EDT, Accepted: November 30, 2022 EDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

## REFERENCES

- Biemann, Chris. *Structure Discovery in Natural Language*. Springer Berlin Heidelberg, 2012, [doi:10.1007/978-3-642-25923-4](https://doi.org/10.1007/978-3-642-25923-4).
- Bublitz, Hannelore. *Foucaults Archäologie des kulturellen Unbewussten: Zum Wissensarchiv und Wissensbegehren moderner Gesellschaften*. Campus, 1999.
- Burdick, Anne, et al. *Digital Humanities*. MIT Press, 2012, [doi:10.7551/mitpress/9248.001.0001](https://doi.org/10.7551/mitpress/9248.001.0001).
- Burke, Peter. *The French Historical Revolution: The Annales School, 1929-1989*. Polity, 1990.
- Canguilhem, Georges. *La formation du concept de réflexe aux XVIIe et XVIIIe siècles*. 2nd rev., Vrin, 1977.
- Chomsky, Noam. *Syntactic Structures*. Mouton, 1957, [doi:10.1515/9783112316009](https://doi.org/10.1515/9783112316009).
- Dosse, François. *History of Structuralism*. Translated by Deborah Glassman, vol. 2, University of Minnesota Press, 1998.
- Dotzler, Bernhard, and Henning Schmidgen. "Foucault, digital." *Meson*, 2022.
- Erdur, Onur. *Die epistemologischen Jahre: Philosophie und Biologie in Frankreich, 1960-1980*. Chronos, 2018.
- Firth, John Rupert. "A Synopsis of Linguistic Theory 1930-1955." *Studies in Linguistic Analysis*, edited by The Philological Society, 2nd ed., Blackwell, 1962, pp. 1–32.
- Foucault, Michel. "Jean Hyppolite. 1907-1968." 1969. *Dits et écrits*, edited by Daniel Defert and François Ewald, vol. 1, Gallimard, 1994, pp. 779–85.
- . "La psychologie de 1850 à 1950." 1957. *Dits et écrits*, edited by Daniel Defert and François Ewald, vol. 1, Gallimard, 1994, pp. 120–37.
- . "Message ou bruit?" 1966. *Dits et écrits*, edited by Daniel Defert and François Ewald, vol. 1, Gallimard, 1994, pp. 557–60.
- . *The Archaeology of Knowledge and The Discourse on Language*. Translated by Alan Sheridan, Pantheon, 1982.
- . *The Order of Things: An Archaeology of the Human Sciences*. Routledge, 2002.
- Foucault, Michel, and Noam Chomsky. "Human Nature: Justice versus Power." *Reflexive Water: The Basic Concerns of Mankind*, edited by Fons Elders, Souvenir Press, 1974, pp. 133–97.
- Furet, François. "Le quantitatif en histoire." *Faire de l'histoire*, edited by Jacques Le Goff and Pierre Nora, vol. 1, Gallimard, 1974, pp. 42–61.
- Geoghegan, Bernard Dionysius. "Textocracy, or, the Cybernetic Logic of French Theory." *History of the Human Sciences*, vol. 33, no. 1, Feb. 2020, pp. 52–79, [doi:10.1177/0952695119864241](https://doi.org/10.1177/0952695119864241).
- Gollub, Tim, et al. "Exploratory Search Pipes with Scoped Facets." *2019 ACM SIGIR International Conference on Theory of Information Retrieval*, Sept. 2019, [doi:10.1145/3341981.3344247](https://doi.org/10.1145/3341981.3344247).
- Harris, Zellig S. "Discourse Analysis." *Language*, vol. 28, no. 1, Jan.–Mar. 1952, p. 1, [doi:10.2307/409987](https://doi.org/10.2307/409987).
- . "Distributional Structure." *Word*, vol. 10, no. 2–3, Aug. 1954, pp. 146–62, [doi:10.1080/00437956.1954.11659520](https://doi.org/10.1080/00437956.1954.11659520).
- . *Methods in Structural Linguistics*. The University of Chicago Press, 1951.
- Helsloot, Niels, and Tony Hak, editors. *Michel Pécheux: Automatic Discourse Analysis*, Rodopi, 1995.
- Herrmann, Hans-Christian von. "Totale Bibliothek und Schreibmaschine. Zum Begriff der Streuung in Foucaults Diskursanalyse." *Figurationen*, vol. 16, no. 2, Dec. 2015, pp. 62–72, [doi:10.7788/figurationen-2015-0207](https://doi.org/10.7788/figurationen-2015-0207).

- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013, [doi:10.5406/illinois/9780252037528.001.0001](https://doi.org/10.5406/illinois/9780252037528.001.0001).
- Kay, Lily E. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford University Press, 2000, [doi:10.1515/9781503617575](https://doi.org/10.1515/9781503617575).
- Kozlowski, Austin C., et al. “The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings.” *American Sociological Review*, vol. 84, no. 5, Sept. 2019, pp. 905–49, [doi:10.1177/0003122419877135](https://doi.org/10.1177/0003122419877135).
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. 4th ed., The University of Chicago Press, 2012.
- Le Roy Ladurie, Emmanuel. “La fin des érudits: L’historien de demain sera programmeur ou ne sera pas.” *Le Nouvel Observateur*, May 1968, pp. 2–3.
- . *Le territoire de l’historien*. Gallimard, 1973.
- Lemny, Stefan. “‘L’historien de demain sera programmeur’: Emmanuel Le Roy Ladurie et les défis de la science.” *L’histoire à la BnF*, Dec. 2017, <https://histoirebnf.hypotheses.org/1505>.
- Michel, Jean-Baptiste, et al. “Quantitative Analysis of Culture Using Millions of Digitized Books.” *Science*, vol. 331, no. 6014, Jan. 2011, pp. 176–82, [doi:10.1126/science.1199644](https://doi.org/10.1126/science.1199644).
- Moretti, Franco. *Distant Reading*. Verso, 2013.
- . *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005.
- Müller, Werner. “Textklassifikation und Stilanalyse: Gedanken zur automatischen Beschreibung eines Produktes und seines Produktionsprozesses.” *Literatur und Datenverarbeitung: Ein Tagungsbericht*, edited by Helmut Schanze, Niemeyer, 1972, pp. 160–87.
- OpenAI. “Language Models Are Few-Shot Learners.” *CoRR*, 2020, <https://arxiv.org/abs/2005.14165v4>.
- Pavel, Thomas. *Le Mirage linguistique*. Gallimard, 1988.
- Price, Jacob M. “Recent Quantitative Work in History: A Survey of the Main Trends.” *History and Theory*, vol. 9, 1969, p. 1, [doi:10.2307/2504167](https://doi.org/10.2307/2504167).
- Rubenstein, Herbert, and John B. Goodenough. “Contextual Correlates of Synonymy.” *Communications of the ACM*, vol. 8, no. 10, Oct. 1965, pp. 627–33, [doi:10.1145/365628.365657](https://doi.org/10.1145/365628.365657).
- Stede, Manfred. *Discourse Processing: Synthesis Lectures on Human Language Technologies*. Morgan and Claypool, 2011.
- Sterne, Jonathan. “The Example: Some Historical Considerations.” *Between the Humanities and the Digital*, edited by David Theo Goldberg and Patrik Svensson, MIT Press, 2015, pp. 17–33.
- Stone, Philip J., et al. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, 1966.
- Underwood, Ted. “A Genealogy of Distant Reading.” *Digital Humanities Quarterly*, vol. 11, no. 2, 2017.
- Webb, David. *Foucault’s Archaeology: Science and Transformation*. Edinburgh University Press, 2013.