

Maurice Erb · Universität Zürich · maurice.erb@uzh.ch

Simon Ganahl · Universität Zürich · simon.ganahl@uzh.ch

Patrick Kilian · Universität Zürich · patrickandreas.kilian@uzh.ch

Distant Reading and Discourse Analysis

With the publication of this special issue, *Le foucaldien* continues its experiment of updating the thought of Michel Foucault. Can historical discourse analyses be carried out with the aid of computers? In order to examine this question, we compare Franco Moretti's Distant Reading with Foucault's archaeological method. Despite their common origins in the French Annales School, the two approaches differ fundamentally. While Moretti interprets literary data by means of social history, Foucault seeks the immanent meaning of discourses. Our preliminary conclusion: digital archaeology appears to founder on the operationalization of the complex concept of the statement (*énoncé*).

KEYWORDS

algorithm, archaeology, big data, digital humanities, discourse analysis, distant reading, foucault, moretti

HOW TO CITE

Maurice Erb, Simon Ganahl, Patrick Kilian (eds.): "Distant Reading and Discourse Analysis", in: *Le foucaldien*, 2/1 (2016), DOI: 10.16995/lefou.16

Contents

1. Digital Order of Things	2
2. Annales and the Consequences	3
3. Data Hermeneutics vs. Archaeology	4
4. Structural Analysis of the Signified	6

1. Digital Order of Things

With the publication of this issue, *Le foualdien* continues its experiment of updating the thought of Michel Foucault. Since Foucault's death in 1984, the humanities — the emergence of which he had described in his book *The Order of Things* (1966) — have been through fundamental changes. These changes apply not only to the blurring of boundaries between the disciplines. Rather, change has also taken place in the way in which this knowledge of the human is created: Articles and books are now written using computers as a matter of course. This is by no means a trivial fact, if we follow Nietzsche and Friedrich Kittler in the belief that our writing utensils also work on our thoughts.¹ This text, for example, was drafted in Los Angeles, Zurich, Mannheim, Vienna, and on trains and planes, typed on MacBooks and ThinkPads, stored in the 'cloud' and also jointly edited there. Without doubt, such media conditions affect the style and structure of a text, its argumentation, vocabulary, its use of references, and so on.

In this edited volume, however, we do not intend to explore the influence of digital media on the writing process, but instead their effect on thinking in the humanities. More specifically: our subject is the techniques and practices of the so-called Digital Humanities, a research field aiming to connect humanities and information technology. And because this field is far too broad, we shall focus on one specific question: Can historical discourse analyses as practiced by Michel Foucault be carried out with the aid of computers? It would be a sign of ignorance to evade this question on the basis of general skepticism towards quantitative procedures. Foucault's emphatically anti-hermeneutic 'archaeology of knowledge,' which seeks rules within the field of statements, appears to tend towards scientific methodology of its own accord. A similar tendency is displayed in the currently much-discussed work of the Italian literary scholar Franco Moretti, who founded the Literary Lab at Stanford University. *Nomen est omen*: Moretti and his laboratory assistants do not want to interpret individual works of art but to measure textual corpora, evaluate them through the use of algorithms, and visualize them.

This special collection began with the intuitive impression that the procedure that Moretti called *Distant Reading* — with a good instinct for academic marketing — had certain factors in common with Foucault's archaeology. If this suspicion were to be confirmed, we considered, then computers and algorithms might also be used for discourse analyses.² To test out this hypothesis, the *foucaultblog* organized a [workshop](#) in Vienna in November 2015, attended by historians and scholars of literature and media studies from Austria, Germany, and Switzerland.

¹ "Our writing tools are also working on our thoughts" (Friedrich Nietzsche 1882 in a letter to the German writer and composer Heinrich Köselitz), for this quotation and Friedrich Kittler's media-archaeological adaptation, see Friedrich A. Kittler: *Gramophone, Film, Typewriter*, trans. Geoffrey Winthrop-Young and Michael Wutz, Stanford: Stanford University Press 1999 [German 1986], esp. p. 203; see also Friedrich A. Kittler: *Discourse Networks 1800 / 1900*, trans. Michael Metteer and Chris Cullens, Stanford: Stanford University Press 1990 [German 1985].

² We are aware that statistical analysis of data is already part of social science routine, for example by means of SPSS (Statistical Package for the Social Sciences). However, these research practices are very different from Michel Foucault's concept of historical discourse analysis, see for instance Philipp Sarasin: *Geschichtswissenschaft und Diskursanalyse*, Frankfurt a. M.: Suhrkamp 2003.

[Maurice Erb](#), [Patrick Kilian](#), [Peer Trilcke](#) and [Frank Fischer](#) then expanded their presentations into substantial papers, which were published in *Le foucauldien*, and are collected in this special issue. The collection also contains a long [interview with Franco Moretti](#), conducted in Zurich by Ruben Hackler and Guido Kirsten in March 2016.

2. Annales and the Consequences

The lines of the intellectual origins behind Foucault's discourse analysis and Moretti's Distant Reading cross in the Annales School. "In the introduction to his 1969 book *Archaeology of Knowledge*," Patrick Kilian recalls in his article, "Foucault not only directly referred to the Annales School, but also sympathized with quantitative tools like sampling, statistics, series, and the analysis of frequency that later would become Distant Reading's weapons of choice."³ In his own work, Moretti also refers to the quantitative methods established in historical scholarship by Marc Bloch, Lucien Febvre, and above all Fernand Braudel.⁴ Braudel's analyses are not concerned with the dynamics of individual historical events, but with medium-term cycles and ultimately with the deep structures referred to as *longue durée*, which extend across centuries.⁵ In his interview with Hackler and Kirsten, Moretti emphasizes the influence of the Annales historians on his thinking and mentions that he initially spoke of "serial reading" in their sense: "They talked about a serial history of the third level. That was my frame of reference and then — you know how these things happen in your brain — 'distant reading' occurred and I decided to get rid of 'serial reading.'"⁶

Serial reading would have been a methodologically more precise term, yet Distant Reading makes it clearer what the process is directed against. Moretti aims to oppose the precise study of selected texts, taught at humanities departments as 'close reading,' with the algorithmic analysis of series of texts. In Maurice Erb's words, he rejects a philology, "in which one bourgeois *exegesis* individual sends a friendly greeting across time and space to another bourgeois *author* individual, but always in the medium of the 'intellect' or of 'understanding'"⁷. The opponent, in other words, is hermeneutics as a method for understanding the deeper meaning of a text, the mind behind the letters. What was the author trying to tell us? This standard question in philological seminars is obsolete for both Moretti and Foucault. "We do not seek below what is manifest the half silent murmur of another discourse," Foucault tells us in *The Archaeology of*

³ Patrick Kilian: Of Trees and Genealogies. A Foucauldian Commentary on Franco Moretti, in: *Le foucauldien*, 2/1 (2016), p. 3, DOI: 10.16995/lefou.21. See also Michel Foucault: *The Archaeology of Knowledge and the Discourse on Language*, trans. A.M. Sheridan Smith, New York: Pantheon Books 1972 [French 1969], p. 10.

⁴ See Franco Moretti: *Graphs, Maps, Trees: Abstract Models for a Literary History*, London: Verso 2005, p. 13.

⁵ See Fernand Braudel: *Histoire et Sciences sociales: La longue durée*, in: *Annales. Économies, Sociétés, Civilisations*, 13/4 (1958), pp. 725–753 (for an English translation, see Fernand Braudel: "History and the Social Sciences: The 'Longue Durée,'" in: Fernand Braudel: *On History*, trans. Sarah Matthews, Chicago: University of Chicago Press 1980, pp. 25–54).

⁶ Franco Moretti in Ruben Hackler and Guido Kirsten: Distant Reading, Computational Criticism, and Social Critique. An Interview with Franco Moretti, in: *Le foucauldien*, 2/1 (2016), p. 6, DOI: 10.16995/lefou.22.

⁷ Maurice Erb: Alles oder gar nichts lesen? Foucault, Moretti und die Verheißungen des Algorithmus, in: *Le foucauldien*, 2/1 (2016), p. 2, DOI: 10.16995/lefou.14 [our translation].

Knowledge.⁸ Instead, the aim is to provide a sober description of what is positively present, which statements appear repeatedly and which regularities form thereby.

This practical and programmatic refusal to read between the lines is grounded in a theoretical rejection of essential parts of Hegelian philosophy. In perfect accordance with new social history, Foucault and Moretti position themselves against the idea that "the course of history might be described as a developmental course of the 'spirit,' which in research practice would boil down to the hermeneutical reconstruction of the intentional statements of 'great' men or minds"⁹. The outstanding works of intellectual history correspond to the heroic deeds of world history, be they works of philosophy or literature. Instead of interpreting this authoritative canon, Foucault and Moretti are interested in a large number of lesser or barely known texts, the serial reading of which is designed to bring structures to light — repeating elements that may be described as ordering patterns. Maurice Erb refers to this process as cultural historical "pattern recognition"¹⁰, which subverts categories such as 'work' or 'author' and does not ask after the 'meaning,' at least not a metaphysical meaning: "a corpus is not written by anyone," Moretti says in the interview, "it has no message and, in a sense, no meaning."¹¹

3. Data Hermeneutics vs. Archaeology

Up to this point, we have described Foucault and Moretti as social historians. In the case of the former, this is certainly too narrow a definition; we will return to this question below. In Moretti's case, however, one might refer to the continuation of a social history of literature by digital means. Essentially, he remains loyal to the concepts of Marxist literary theory, although he does not interpret individual works, but rather evaluates large quantities of literary data with the aid of computers. That is at least the reputation that proceeds Moretti, the digital humanist from Stanford. According to Frank Fischer and Peer Trilcke, his work does not confirm this reputation. Firstly, they contend, Moretti's Distant Reading has nothing to do with computer-based analysis in as far as the information technologies and practices play not the slightest role in his writing; and secondly, the more fitting term would be "mid-distance reading," because none of his text corpora are actually *big data*.¹² "I don't program," says Moretti in the interview, "I could never completely understand how long it takes these students or younger colleagues to do the programming."¹³ This admission is not surprising, seeing as the Digital Humanities are concerned with useful cooperation between scholars of culture and computer scientists. Philologists and

⁸ Foucault: *The Archaeology of Knowledge*, p. 28.

⁹ Philipp Sarasin: Sozialgeschichte vs. Foucault im Google Books Ngram Viewer: Ein alter Streitfall in einem neuen Tool, in: Pascal Maeder et al. (eds.): *Wozu noch Sozialgeschichte? Eine Disziplin im Umbruch*, Göttingen: Vandenhoeck & Ruprecht 2012, pp. 151–174, here p. 168 [our translation].

¹⁰ Erb: *Alles oder gar nichts lesen?*, pp. 5–6.

¹¹ Moretti in Hackler and Kirsten: *Distant Reading, Computational Criticism, and Social Critique*, p. 16.

¹² Peer Trilcke and Frank Fischer: Fernlesen mit Foucault? Überlegungen zur Praxis des *distant reading* und zur Operationalisierung von Foucaults Diskursanalyse, in: *Le foucaldien*, 2/1 (2016), p. 10, DOI: 10.16995/lefou.15 [our translation].

¹³ Moretti in Hackler and Kirsten: *Distant Reading, Computational Criticism, and Social Critique*, p. 7.

historians ought to understand information technology, but they do not have to master software development. The real problem is that Moretti's studies do not report on the digital work undertaken in each case. What kind of algorithms were developed to order the data? And what data is it exactly? Far from answering such questions, Moretti barely discusses them in his books and essays.

That means that the abstract models Moretti uses to interpret literary history cannot be reconstructed in the literal sense by his readers. We cannot recreate the graphs, maps, and evolutionary trees because their realization remains a mystery. Thus, Moretti's key achievement ultimately lies in the 'close reading' of these diagrams — an interpretation that follows the patterns of social history. He understands cultural forms as the result of a combination of societal forces,¹⁴ and above all of political and economic forces: the literary genre of the village story is shaped by the evolution of the nation state; the techniques of the detective story succeed on the literary market (or fail); etc. In principle, this "data hermeneutics"¹⁵ follows the classic Marxist scheme, whereby the factual societal conditions determine the ideological superstructure. Trilcke and Fischer distinguish this approach, which they describe with Tom Scheinfeldt as "framing knowledge in a theoretical or ideological construct," from the practice of Digital Humanities: While 20th-century scholars were interested in grand theories, meta-narratives, the current focus is on methodological procedures and on operationalizing concepts.¹⁶

Foucault's archaeology, as described above, is inspired by social history in its French form. His approach breaks with this tradition, however, by assigning no lesser reality to statements than to other events. On the contrary: discourses are not only "representations" or "groups of signs," but rather actual "practices that systematically form the objects of which they speak"¹⁷. By the repeated issuing of statements, rules form that determine what can be said, what counts as true in a discourse. These formations have a performative power, they create "objects" as "epistemic things" (Hans-Jörg Rheinberger).¹⁸ Discourses are thus not only different representations of a stable reality made of 'real' objects. Instead, Foucault analyzes historically shifting realities: forms of knowledge consisting of both say-able and see-able elements. According to *The Archaeology of Knowledge*, these heterogeneous ensembles can be described as discursive. As Maurice Erb explains in his article, however, Foucault understood the ordering of knowledge both in his early work and in the 1970s above all as spatial structures.¹⁹ The decisive point here is that discourse analysis is an immanent process. Its text corpora may not be cohesive works created by authors. Against Moretti's claim, however, they do have "meaning" — meaningful

¹⁴ See Moretti: *Graphs, Maps, Trees*, p. 64.

¹⁵ Trilcke and Fischer: *Fernlesen mit Foucault?*, p. 11 [our translation].

¹⁶ See Tom Scheinfeldt: *Sunset for Ideology, Sunrise for Methodology?*, in: *Debates in the Digital Humanities* (2012), URL: <http://dhdebates.gc.cuny.edu/debates/text/39>.

¹⁷ Foucault: *The Archaeology of Knowledge*, p. 49.

¹⁸ See Sarasin: *Sozialgeschichte vs. Foucault*, p. 170; see also Hans-Jörg Rheinberger: *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*, Stanford: Stanford University Press 1997.

¹⁹ See Erb: *Alles oder gar nichts lesen?*, p. 4 and 6.

content resulting from the regularity of the statements. Whereas Foucault aims to recognize this link in the quantity of discursive material, Moretti sees himself confronted with meaningless data sets, which must be given meaning via social historical interpretation.

4. Structural Analysis of the Signified

There is reasonable doubt that the research practice of Distant Reading might help us to analyze discourses in Foucault's sense by using computers. Firstly, Moretti's "reverse engineering"²⁰ from cultural forms to societal forces cleaves to the tradition of social history; secondly, his books and articles tell us too little about the information technology used in the studies. That does not mean, however, that an algorithmic discourse analysis is not possible in principle. It is a question, as outlined above, of a kind of *pattern recognition*, which has been practiced for some time in the field of *data mining*. In order to find out whether these techniques might be used to evaluate large data quantities for an archaeology of knowledge, we must concentrate on the core, the atom of Foucault's procedure: the statement (*énoncé*). The pattern that a discourse analysis sets out to recognize is after all a regularity of discursive events in the form of a series of statements.

So what is a statement in accordance with Foucault? Frustratingly, the *Archaeology of Knowledge* does not provide a clear answer to this question. Foucault's readers learn above all what a statement is *not*. It is not a logical proposition, nor a grammatical sentence, nor a linguistic act of speech, nor even a regular verbal combination.²¹ Foucault gives an example of a random series of letters on a keyboard, which does not make up a meaningful word but can be a statement in the context of a typewriting textbook. What algorithms might be capable of finding such a verbal entity? "Foucault's concept of the statement seems deliberately to block its operationalization through digital analysis procedures," Peer Trilcke and Frank Fischer write in their article.²² This problem stems from the archaeological method, which aims at a structural analysis not of the *signifier* but of the *signified*, thus that which is referred to and not the chains of signs with which both structuralism and text mining are concerned.

Foucault's semantic *pattern recognition* therefore, as Maurice Erb establishes, "requires at least a basal-hermeneutic understanding"²³. In order to recognize a statement and its function in a discourse, its superficial meaning must be understood (though not its deeper meaning). To conclude from the outcome of the *foucaultblog's* workshop in Vienna in November 2015, the present state of information technology does not yet allow us to automate this process. It is unlikely that, for example, the development of the *semantic web* will change this technological starting point. Without doubt, the algorithms of the communication industry as used by Google

²⁰ See Kilian: Of Trees and Genealogies, p. 10.

²¹ See Foucault: Defining the statement, in: *The Archaeology of Knowledge*, pp. 79–87.

²² Trilcke and Fischer: Fernlesen mit Foucault?, p. 17 [our translation].

²³ Erb: Alles oder gar nichts lesen?, p. 6 [our translation].

or Facebook, for instance, are far more advanced than the approaches of Digital Humanities. However, subsequent to Erb's article we would like to remind our readers that the computer is not a miraculous machine, but at its core nothing more than a very fast clerk who reads, transcribes, and deposits files (data) according to instructions (algorithms).²⁴

The main problem is thus that the digital possibilities do not match up to the complexities of our methodological concepts. To analyze discourses by computer, we would need to simplify the difficult concept of Foucault's statement in such a way that it could be 'understood' by these fast but simpleminded clerks. Instead, we would like to call upon those clerks to educate themselves along with us, to become not only faster but also more flexible in carrying out their tasks. These are our hopes for the digital world. And when it comes to the humanities, the following quote from Moretti, discussed in Trilcke and Fischer's article, sets the direction:

*Forget programs and visions; the operational approach refers specifically to concepts, and in a very specific way: it describes the process whereby concepts are transformed into a series of operations — which, in their turn, allow to measure all sorts of objects. Operationalizing means building a bridge from concepts to measurement, and then to the world.*²⁵

For *Le foucauldien*, updating Foucault's thinking also means operationalizing his concepts. Can we translate terms such as 'statement' or 'dispositif' into a series of operations without losing their depth and complexity? The point of this bridge-building exercise would lie, on one hand, in the possible automation of the analysis. What appears more important, on the other hand, is the prospect of a future archaeology of knowledge or genealogy of dispositifs, which might be achieved on a collaborative or participatory basis by means of digital media. Such a collaboration, however, requires precise work on the concepts, to which this special collection on *Distant Reading and Discourse Analysis* hopes to contribute.

²⁴ See Erb: *Alles oder gar nichts lesen?*, p. 8.

²⁵ Franco Moretti: "Operationalizing": or, the function of measurement in modern literary theory, in: *Pamphlets of the Stanford Literary Lab*, 6 (2013), p. 1, URL: <http://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>.