Introduction to Meaningful data

Dario Compagno, Matteo Treleani

Editorial

Dario Compagno* and Matteo Treleani

Introduction to Meaningful Data/Données signifiantes

1 Are data signs?

The human sciences are interested in objects of study characterized by the fact of having *meaning*. Whether they are artefacts or behaviors, they can be interpreted by someone and this is what makes them relevant (Geertz 1973; Eco 1976 [1975]). Until recently, the widely accepted general opinion was that meaning could not be defined using quantitative methods of research because measurements may only describe the material and contingent component of meaningful activities (their vehicle or "expression," in the sense of Hjelmslev 1961 [1943]) but are unable to account for their truly semiotic core (their intentionality or "content"). This is the reason why qualitative methods are often preferred to quantitative ones in the humanities – *data are never meaningful*, according to this widespread opinion and they cannot account for meaning-forging practices. In fact, how could meaning be objectified in data? Does data not exclude the observer, thus making the observed meaningless, as we could say adapting Dilthey's words (see Dilthey 2010)? In this perspective, data cannot account for the living essence of signs and actions, which is eminently intersubjective. This seems especially true when we consider that quantitative methods do not directly apply to signs but instead to measurements taken of signs. Studying the chemistry of film photography or the optics of projection is clearly insufficient to understand cinema and similarly any other kind of measurement must necessarily miss the spirit and meaning of cultural life.

In recent years, however, more and more researchers have explored the possibility of working on culture with quantitative and especially computational tools, founding what is referred to as the *digital humanities*. For these researchers,

***Corresponding author: Dario Compagno,** STID - IUT de Metz, CREM, Universite de Lorraine, Ile du Saulcy, 57000, France, E-mail: dario.compagno@univ-lorraine.fr
https://orcid.org/0000-0001-8295-792X
Matteo Treleani, SicLab, Universite Nice Sophia Antipolis 98, bd Edouard Herriot BP 3209 06 204, Nice Cedex 3, France, E-mail: matteo.treleani@univ-cotedazur.fr

data can be used to study human activities and products. Data are used to transform a cultural object into a model, an organized representation. The object or phenomenon under investigation becomes an entity that researchers can manipulate. From this perspective, it is harder to neatly disentangle signs and data – signs and data are both ways to approach an object and give access to it through interpretation. In Peircean terms, we could say that research turns the "dynamic object" under analysis into an "immediate object," depicted in the light of signs (see EP 1). Interpreted data are also traces of their object (Merzeau 2009). Furthermore, data analysis is *immanent* to the collected traces, as traces become the only manipulable link to the object under investigation. These are interesting affinities between semiotic and statistical procedures of analysis – cultural data analysis is a way to *relogicize* culture as Roland Barthes (1966) said of narrative analysis.

The way in which data and signs are interwoven can be seen also by looking at how research is conducted. Scientific modelling is a semiotic activity that depends on the production of signs like sentences, pictures, and diagrams, which provide context for measurements and "data alone" are difficult to imagine in practice. This is so true that some researchers suggest that numbers alone cannot entail inferences of any sort because they lack the power of argumentation (Carel 2011). Numbers become meaningful, producing interpretants potentially capable of improving knowledge, only when they are combined with linguistic and, more generally, semiotic elements. Data are never neutral, in the sense of being unaffected by the observer's procedures. Data are already produced (gathered, processed) by researchers within a setting and are thus always the result of human activity in which subject and object are both present. Data bear the traces of their production. But then all these considerations imply that *data are meaningful*. Within the framework of a research question and of certain habits of collection and analysis, data may play a heuristic role and authentic discoveries can be made about culture with the appropriate use of quantitative tools. If data lose their intentionality towards the object they are used to represent, then they are not even data any more. The matter becomes then how to use data to deal with signs that already have an intentionality on their own.

A model is never the object, as a map is not the territory it depicts. There is a complex referential relationship between model and object and this opens up the possibility of mistakes and approximation (Eco 1976 [1975]). How models represent objects is a major question for contemporary epistemology and also for semiotics. Modelling features and their manipulations may be of different kinds (iconic, symbolic) and follow different logics. For example, some scholars have highlighted the inevitably metaphorical nature of models (Hesse 1966; Black 1962) while others, closely following Peirce, today study the nature of diagrams

and of the operations that can be performed on them (Stjernfelt 2007; Chapman et al. 2018). In other words, clearly models are *construed* by researchers working within defined social practices. In fact, the lenses of culture enable certain aspects of phenomena to be seen while hiding other traits (Eco 2000 [1997]). Nonetheless, this does not mean that modelling is arbitrary and unrelated to the object it is used to approach. On the contrary, the point is to understand how to perform *motivated* measurements and formalizations (Meunier 2017). Which kinds of operations can be performed on data to learn about the meaning of the activities and products from which they come?

It goes without saying that scientists are well aware of the impact of research practices, ethics, and rhetoric on the production and interpretation of results. In fact, an entire discipline – statistics – deals with the improvement of quantitative research procedures and their correct interpretation. However, the semiotic nature of models, within which measures make sense, implies that semiotics also has an important role to play alongside statistics to ensure researchers get a complete, fully-informed grasp of data analysis including and going beyond the communication of statistical results (Huff 1954; Tufte 1983) and the production of diagrams (Bertin 1983 [1967]; Wilkinson 2005). Until today, semiotics has been an almost entirely qualitative discipline dedicated to making procedures for the interpretation of cultural objects and practices sound and explicit. The question now is *how can semiotic and statistical skills be combined to study culture?*

2 The quantitative study of culture beyond explanation

Can the study of culture be purely based on quantitative research as if data were not signs or signs could be encoded numerically without residuals? Can statistics do without the insights provided by the qualitative disciplines in the definition of good practices for cultural research? These are not rhetorical questions, as, in the natural sciences, statistics provides practically the only interdisciplinary methodological support to accompany disciplinary theories and practices. The concept of significance, which is at the very foundations of modern statistics (Fisher 1947), is *the* criterion to help disciplines define which results are meaningful (relevant, important) to them, although obviously there are some drawbacks (Colquhoun 2014, Colquhoun 2017). Therefore, one possibility is that statistics really is the only *organon*, in the Aristotelian sense, needed for research even in the human and cultural sciences, as if the only theoretical approach to signification needed for human research was the

statistical approach. In fact, quantitative cultural studies have already started to rely on statistical significance, looking for emergent properties in the evolution of culture (Sperber 1996; Baronchelli et al. 2015) and this is also true also for semiotic experimental research (for example Fusaroli et al. 2012, Fusaroli et al. 2015). If this perspective is wholeheartedly endorsed, the traditional threshold between natural and human sciences, formalized by hermeneutics as the opposition between explanation and understanding (see Dilthey 2010), tends to simply be removed as every truly scientific activity becomes an explanation, possibly even a *causal* explanation. This occurs because there has never been a *finalistic* interpretation of statistical correlation and any accomplished interpretation of correlation inevitably leads to cause-effect relationships within an explanatory model. Statistics taken alone "push" towards causal explanations, so to speak, and the specificity of the human sciences seems destined to be overcome by the development of tools that allow researchers to see the "causes of meaning" as if intentions and reasons could be reduced to causes (as, for example, in the view of Davidson 1980).

Another possibility is that, instead, the recent adoption of quantitative tools has made everyone aware of the need for a second *organon* in the human and social sciences that is complementary to statistics and capable of identifying general procedures for cultural analysis and interpretation. And above all which is capable of giving a more comprehensive definition to *significance* and adequate for talking about human communication practices and meaning. Next to causes there are signs, which also bring about effects but through the mediation of interpretation and intentionality, which cannot be reduced to causality (in the perspective defended by Anscombe 1957; von Wright 1971; Bratman 1999). For this second approach, understanding is not expunged in favor of explanation – the quantitative detour is aimed to better understanding. Without being contemptuous of studies that aim to explain cultural regularities and evolution, is it possible to also design a *statistically-aided form of understanding*? To give a *finalistic interpretation to statistical correlation*? Currently, the most important thing may not be to reaffirm the distinction between *Erklären* and *Verstehen*, thus between positivism and hermeneutics, or to submit one approach to the rule of the other but rather to look for ways to mediate between the two. This has appeared to occur in several recent semiotic experimental studies on the evolution of language games and social coordination (for a presentation and discussion, see Galantucci and Garrod 2011; Tylén et al. 2013).

The relationship between explanation and understanding has notably been addressed twice in semiotics; once by Greimas and Ricoeur (1989). For the French scholars, semiotics is useful to provide a nomothetic explanation of narrative structures that is a prerequisite to idiographic hermeneutic understanding. This

two-step procedure presupposes the possibility of pure (notably transcendental) explanation in culture and assigns the role of performing it to semiotics while at the same time reaffirming that full understanding has necessarily to go beyond signs and their study to reach actual, lived experience. In this perspective, meaning is the result of an activity of the mind which goes beyond semiosis (and famously including, for Ricoeur, the lived experience of time). If we think carefully about these assertions, they imply that semiotics could ideally become a purely quantitative discipline *but* (or *because*) it cannot reach the human essence based on lived experience, beyond signs. If we consider the matter in more depth, to say that there is something in man that is not sign is both a way to reduce the legitimacy of semiotics in favor of phenomenology and a means to state that something fundamental in human nature necessarily resists explanation (and potentially quantification). This finally saves Dilthey's threshold as it is.

However, the structuralist and phenomenological perspective presented above is not the only one within which semiotics can exist. A different take on explanation and understanding was promoted first by Barthes (1974 [1970]) and then in a more developed way by Eco (1979, 1994 [1990]). For Eco, pure semiotic explanation is not possible because analysis necessarily depends on interpretation and cultural explanation cannot exist outside of understanding. In the human and social sciences, to explain is always to explain interpretation – a scientific reading is still a reading. This means that transcendental constitution, upon which Greimas and Ricoeur based their *dividi et impera*, cannot be obtained because pure meanings remain out of reach (see Derrida 2011 [1967]) and therefore so does objective analysis. Also, for Eco as well as for Peirce, there is no hermeneutics without and beyond signs and no mundane understanding (residual to transcendental constitution) or authentic *proprium* and "secret" of humankind that can never be explained. Human behavior is entirely based on sign processes. Still these processes cannot be fully accounted for by causal explanations because of the space of choice given to individuals, mediating (through intentionality) between stimuli and responses (Eco 1994 [1990]). Thus, from this perspective, understanding finality and meaning strategies becomes one of the main aims of the discipline.

We have seen why the adoption of quantitative methods in the human and social sciences involves important epistemological consequences (see also Doueihi 2011) and forces us to rethink about the importance of signs and semiotics for data analysis today. If we aim to improve the understanding of man without simply replacing finality with causality, then semiotics could become a bridge between the two cultures, helping to redefine a concept of significance valuable for understanding as well as for explaining. On a methodological level, then, it is important to estimate how semiotics, as a discipline with its proper historical tradition, can enter into a practical dialogue with quantitative research practices.

3 What do we gain from the quantitative detour?

3.1 A critical look at data

One of the merits of the recent widespread media-based, political and scientific discourses about data collection and usage is that they have called for a critical awakening. The most visible results of such critical studies has been to highlight the ideological dimension of certain uses of data – and this critical analysis of ideologies resembles what was carried out in semiotics in the seventies, especially by Eco. The *datafication* of cultural practices and of human activities can be seen as a form of domestication in the sense given to this term by the Marxist *critique* of cultural industries (the administered world of Adorno and Horkheimer 2002 [1944]). This kind of *critique* was later used in certain seminal texts about the philosophy of technique. In Bernard Stiegler's approach, for instance, media techniques are seen as the major instrument for the domestication of time and memory (Stiegler 2010). The widespread resonance of certain media events helped the critical discourse to reach a larger public. The 2013 Snowden case, in particular, determined a visible shift in how public opinion depicts data collection and analysis (Casilli and Cardon 2015). Before this event, technique was often seen as an instrument of emancipation in a perspective linked to seventies counterculture from which the study of digital culture was born (Turner 2006). After 2013, technique can be observed to have been increasingly presented as a device of mass control. An ideological dimension was even found behind the concept of transparency linked to certain cultural practices of digitization (Chul-Han 2015). The discourse on technique – both in its critical and its optimistic declinations, of which Morozov (2013) and Jenkins (2006) are two examples – aims to react to naive forms of technological determinism. The collection of data performed by national agencies, for example, is not in itself a phenomenon that is exclusively related to digital tools and practices; these just make data collection more efficient and ubiquitous. The problem is then obviously not technology *per se* but domination and the drive for social control.

Technology is also the primary place where power relations are condensed and thus made visible to analysts' eyes. It is precisely this *revealing* aspect of technique that is at issue for semiotics. Technology makes certain aspects of social reality visible. Again, there is nothing really new in this perspective. In 1939 Walter Benjamin had already observed how in cinematographic technique, the use of slow motion made, for example, elements of matter and movement perceivable that would not otherwise be observable (2008 [1939]). Moreover, slow motion invents a new reality that is inaccessible to the naked eye, opening

up and making visible some of the processes responsible for the creation of meaning. Today, working on data is similarly a matter of exploiting the observability potential of digital methods (Rogers 2013) and finding their semiotic relevance. At the same time, the reflexive use of digital tools is also a way of appropriating them and somehow exploiting their potential "against" their original intended aims. This is a potentially groundbreaking innovation. Semiotics could in fact appropriate tools derived from digital methods within a critical perspective, going beyond the purely qualitative criticism of ideological and technological discourse. Semiotics could learn from such cultural baggage and try to derive heuristic tools from it, within the aims and procedures of the discipline's framework. To do this, semiotics should integrate the spirit of critical considerations, while at the same learning how to make use of data for its analyses. Semiotics could aim to highlight what *makes sense* in digital methods, practicing the digital *pharmakon* (Stiegler 1998) to turn it into a heuristic tool at a turning point in the history of the communication landscape.

3.2 A shift in scale and in methodology

Critical discourse is one of the conceptual frameworks within which semiotics may work on data but it is the use of quantitative digital tools that will be determinant from a methodological point of view. In fact, quantitative tools operate a shift in scale for semiotic analysis, which means moving from analyses of individual texts and case studies to that of larger corpora. Franco Moretti called this type of approach *distant reading* (Moretti 2015, 2013), allowing the researcher to work on large textual archives without losing a grasp of their meaning. This change of perspective implies rethinking some of the discipline's foundations. To begin with, a corpus in itself is not intentionally significant unlike the objects usually analyzed by semiotics such as novels, pictures or films. And yet an explicit corpus enables a better delimitation of the meaning of its elements. For Rastier (2011) only a *well-construed corpus* can give the conditions of interpretability of the texts it includes. To work with explicit corpora also has a further heuristic consequence, namely, the extension to *larger and smaller scales* of observation (Alge-Hewitt et al. 2015). In literature, for instance, the emergence of style is often invisible at the level of individual words, while a shift of scale from words to units smaller or larger than words can pave the way for a new form of stylistics. Another shift of scale, from sentences to paragraphs, can show the hidden presence of themes instead. Now, to obtain such shift in scale implies making use of *statistical methods*, which leads to more major advantages of quantitative analysis for semiotics. The phenomena observed on a larger scale present a *representativeness*

that goes beyond anecdotal examples and allows researchers to generalize and test hypotheses, which are crucial aspects of research in experimental sciences but are often underestimated by their interpretative counterparts (Piper 2016a). In fact, statistical tools permit to perform regulated *generalizations by induction* from a large number of units. Moreover, experimental research grounded on data collection and analysis can work under formal *hypothesis testing*, therefore permitting one to look for empirical evidence to bring new perspectives on old questions, as has, for example, been done by Andrew Piper for fictionality (Piper 2016b). It should be said, however, that even without recurring to quantitative tools, semiotics has always looked for cultural regularities, which means the digital turn we are living appears highly compatible with the discipline's traditional means and ends.

3.3 Repetition, a new observable

One of the most fascinating aspects of applying quantitative methods to cultural objects is the possibility of identifying *new observables*, that is, significant patterns that would not otherwise be visible (Rastier 2011). This involves taking structures hidden by a mass of small elements into account, making some traits related to the entire corpus visible (frequency, repetition, redundancy, etc.). As an example, within an archive of audiovisual documents, specifically news programs, the opening titles are a recurring and redundant element (Cagé et al. 2017). The role of the opening titles is understandable only by considering the recurrence of many identical or similar occurrences. In other words, as explained above, there are regularities of meaning that are visible only on a scale larger than that of a single text (Algee-Hewitt et al. 2015).

An effect of repetition is also at work in many digital cultural productions. Creative practices such as remixes, mash-ups or virality – which are obviously not exclusively linked to digital culture but are very visible in digital forms – have been observed from an aesthetic or semiotic point of view as practices of recontextualization and as bottom-up reappropriations from a sociological point of view, questioning the dissemination principles of traditional cultural industries (Navas et al. 2014). However, a radical questioning about the meaning of repetition in itself is still missing. Quantitative methods allow us to deal with these practices in order to better understand the intersection between an aesthetic or semiotic analysis, usually based on the singularity of an object and of its context and the larger dimension of its repurposing. A media object such as a meme, for example, derives its social meaning from its circulation and from the multiplicity of the different contexts in which it appears. The study of its symbolic force

cannot ignore this repetitive trait. A quantitative approach to semiotics therefore needs to learn how to take repetition in itself into account as a logic of meaning production. This is just an example showing that the quantitative detour is more than just a new methodological instrument for traditional analysis. The new observables created by technological tools, or what Benjamin called new realities, invite a re-foundation of certain semiotic theoretical and methodological principles. The issues at stake seem to be close to those of a new or digital materialism (in the sense of Doueihi and Louzeau 2017).

4 The limits of quantification for the study of meaning

To talk about the limits of science and technology is always a bad idea given that progress amuses itself in finding ways of turning down pessimistic expectations. And yet to see an unlimited source of scientificity for the human disciplines in data analysis would probably be too optimistic and forgetful of the past waves of positivism. There are actual risks in the adoption of cultural data analysis that are visible to researchers. First and foremost, the risk of reductionism, that is the idea that the questions asked by the human sciences can be answered by another contiguous and more scientific domain. To read literature is definitely a cognitive activity but the *Divine Comedy* is not in any of our brains and was not even in Dante's. We do not study Dante to find more cognitive biases but for the value of his work. The human sciences attack complex objects deploying a complex research apparatus that is hybrid in nature and often uncontrolled for in all its details. So to reduce complexity for the sake of explanation may end up *killing meaning by dissection*, so to speak, and entirely missing the meaning of literature and art. The specificity of semiotics is meaning in its richest sense, including when it is reached through aesthetic activity. It is increasingly probable that quantitative research can help us to know more about meaning but there is still no theoretical new dawn that has turned meaning into a computable object and therefore any tentative approach still risks counting words instead of reading them. What is observable and what is interpretable are interwoven (at least from a semiotic perspective) but this does not also mean that we can state everything that is important for meaning beforehand and thus yet produce an adequate model of human comprehension.

All of this comes hand in hand with the ongoing delegation of semiotic questions and answers to computer scientists. Today we may have the impression that many of the questions asked by the human sciences over past centuries

are very current and important in a fully digitalized and semiotized world but they are mostly studied by computer scientists who have a very different way of answering them. The specificity of the semiotic dimension, clearly identified since the very beginning of Western philosophy, has never been reduced to mathematical logic and it will probably haunt objectivity for some more time (see Compagno 2018). Our discipline's *savoir faire* is based on this residual specificity. It is not the same to train a literature student by discussing his or her individual interpretation of poems or instead to carry out experimental research into other students' reading. Semioticians have different tastes and skills to sociologists or psychologists and the reason for this could be that the deep networks in our brains need time and repetition, supervised learning, and direct experience of texts and interpretations to work at their best. Enthusiasm for data risks delegating research intelligence to instruments (see Rebillard 2011) diverting our attention to contingent matters but with the promise of control.

At the opposite side of the spectrum lies another risk – that of intentional or unintentional intellectual imposture. Recurrently using quantitative representations and statistical inference has an undeniable rhetorical appeal for many publics. It is true that this occurs for a reason, given that it becomes possible to evaluate otherwise fuzzy affirmations, but one should not take numbers *per se* as a hint of stronger argumentation. A scholar who is not trained in quantitative methods may be unable to detect fallacies and just fall under the charm of diagrams. Specialist communication already needs skilled readers to be controlled and this risk grows exponentially with scientific popularization because data and diagrams are widely used to tell stories for which they provide only a limited rational support but a large rhetorical push. Data seems especially versatile for the production of sublime, baroque representations in which the richness of details hides the bigger picture, intentionally or not.

5 Contributions

Chartier, Pulizzotto, Chartrand, and Meunier carried out a computational semiotic analysis of a corpus of annotated images. The corpus includes most of Réné Magritte's artworks, manually annotated with descriptors whose purpose is "to encode or categorize with lexemes and syntagms the iconic visual content of artworks. Therefore, artwork iconic analysis is achieved through the analysis of these semantic annotations." The authors built a semantic vector space (SVS) trained on the annotations. A SVS is "both a computational model and algorithmic based methods for the inductive discovery from combinatorial patterns of

signs in a corpus (usually textual, but not exclusively) of meaning structures." Chartier et al. describe some formal operations to automatically derive semantic inferences applied by the authors in three experiments. In the first, the reliability of the corpus's semantic indexing was evaluated by predicting one annotation (the syntagm 'man') given the co-occurrence of other syntagms strongly associated to it in the corpus. A similar but reverse approach was used for the second experiment. Instead of looking for one target absent syntagm on the basis of the others present, the authors started with one syntagm present to identify which other syntagms are often associated with it. "The aim of the second experiment is to develop a method of componential analysis that decomposes the syntagmatic signature of a descriptor into its different combinatorial sub-patterns called its semantic component." More specifically, they found that: "there is no artwork in the studied corpus in which the woman is represented without a syntagmatic signature correlated to 'nudity' or correlated to 'hair'." This means that Magritte's iconic idiolect codifies the syntagm 'woman' very strongly. Thirdly, Chartier et al. made a topic analysis of the corpus: "There are groups of descriptors and groups of artwork annotations that, because they share similar combinatorial patterns in the corpus, they are projected into the same regions of the SVS. Therefore, the SVS is structured by various high density regions." For the authors, these regions can be interpreted as isotopies.

In their work, Reyes and Sonesson looked for computational tools capable of improving the analysis of the plastic layer of images as defined and explored by Greimas, the Groupe μ, and Floch. The authors begin with a discussion of plastic analysis and its current limitations. Its first limitation is a lack of methodological rigor. If we look at Floch's analyses, "it seems that there are other intuitive divisions of the picture which may be at least as fully supported by the position of binary oppositions as the one proposed by Floch." Another more systematic approach would require greater effort and this is how computation comes in:

> A more open-ended analytical procedure would be to take into account all imaginable divisions of a picture, whether they are binary, ternary or whatever, and to investigate which of these divisions can best be supported by a great quantity of plastic dimensions ... No doubt this is too big a task for a human researcher once it is applied to a series of pictures, as token analysis requires. If we want to implement the tall task thus setup, we need to develop computer algorithms for the analysis of the plastic layer.

In this way, the authors found a striking similarity between the features identified by the Groupe μ and those used by computer scientists working on image compression and analysis. Of course, the relevance of these features for plastic analysis cannot be taken for granted and its recognition would be a major theoretical advancement in itself. For the moment, Reyes and Sonesson have

developed a "proof of concept" to show the heuristic value of data analysis for pictorial semiotics. A corpus of digital pictures of Rothko's paintings was studied using different visualization techniques, through which "it is possible to gain insight about the colour signature specific to Rothko." The authors conclude by imagining a sort of Turing test based on the automatic production of pictures following the plastic "grammar" of late Rothko's paintings. This could be a way to validate the results of plastic analysis which currently lacks shared validation procedures as is indeed stated by the authors at the beginning of their paper.

Gefen and Reboul analyzed what is called culturonomics, namely, the use of big databases for cultural history. The authors worked on the history of the idea of literature as a word, as a field of study, and as a concept. They aimed to create an empirical history of literature based on data mining. In fact, data make it possible to verify hypotheses that are advanced by qualitative research but are difficult to prove because they are based on an intuitive synthesis performed by individual researchers. Gefen and Reboul state that "the essential thing is to see to what extent [the uses of data] modify our traditional approach to the administration of evidence in the human sciences: with the possibility of 'operationalizing' (Moretti 2015), i.e. verifying theoretical or historical hypotheses, the proposals of the human sciences become falsifiable or, more simply put, verifiable." In any case, the emergent field of the quantitative history of ideas demands the what Gefen and Reboul call finesse in thinking. The interpretation of data, the conception of a theoretical hypothesis and the constitution of corpora are largely dependent on interpretative choices and on theoretical frames that cannot just be found in the data alone.

To evaluate the interest of quantitative approaches for semiotic analysis, Dondero explores the relationship between the semiotic analysis of images and the computational analysis of image corpora. In particular, she studies the works produced by the Cultural Analytics Lab of Lev Manovich (see Manovich 2013). Through the visualization of vast archives of images, Lev Manovich produces diachronic analyses. The work of Manovich is a kind of extension of the visual case studies produced by Franco Moretti on literature. Dondero's study mixes the prisms of post-Greimassian semiotics and Peircian semiotics. The former is used with respect to the issue of images-within-an-image and metavisual visualization and the latter to explain the notion of diagram. The work of Manovich is a "visual analysis" obtained by classifying and distributing images according to their metadata or plastic features within an aggregate visualization. There is a clear metavisual dimension in the procedure – that of using images to study images. This is produced by what Dondero calls diagrams of images and montages; heuristic tools making visible contrasting areas and superpositions. The quantitative analysis of visual categories in the work of

Manovich should further ground qualitative significance analyses. In other words, a visual map of image corpora is heuristic for semiotics whether it guides interpretation through the selection of reading paths or not. What is interesting is that the paths identified by Manovich are found thanks to computation exclusively performed on the images' expressive plane and so somehow independently of any interpretation of their content.

The dichotomy between information and uncertainty is the important issue in Burgio's contribution. Burgio analyzes the representation of uncertainty in infographics and in data visualizations with the aim of understanding whether infographics are linked to a rhetorical use of data, in the form of a practice of enunciation that hides the enunciator's role. The use of data in journalism and communication can thus be seen as a new kind of rhetorical procedure to create objectivity. The camouflage of the enunciator is a "degree zero of writing" that can be questioned with the qualitative analysis of case studies. Burgio thus explores the topic of uncertainty: how do we represent the vagueness of certain data in infographics? Data journalism demands precision, accuracy, and clarity. So, whenever data are used to reinforce a sentiment of belief, it becomes interesting to identify situations where statistic irrelevance is turned to meaning. Burgio affirms that, in some circumstances, data visualization "embeds a feeling of skepticism and frustration and even detachment from data that cannot be trusted." Therefore, to identify the means of showing and not hiding uncertainties is an important challenge for data visualization. A visualization by Amnesty International about the death penalty worldwide in 2014, for instance, includes a text at the bottom stating that the reported numbers do not include China, "which alone has carried out more executions than all the other countries combined." This sentence makes the visualized data less interesting and invites the reader to investigate the Chinese situation in more detail. Burgio concludes affirming that uncertainty is actually what focuses the reader's attention with a pathemic effect. This shows that any data visualization expresses the position of the author rather than being a truthful mirror of reality.

Jégou's contribution discusses the quantitative approaches that can be used to study the visual complexity of cartographic images. His work focuses on the development of Bertin's semiology of graphics within statistical cartography, which aims to understand how "the types of relations between elements of a statistical variable ... will correspond to some means of representation, variations of graphical forms transmitting them in a quick and effective way" (our translation). A first set of experiments conducted by Jégou aims to identify methods for evaluating the reading complexity of a cartographic image. Complexity can be defined from several perspectives, namely, information theory, visual perception and semantics. Computational tools already enable

the visualization of the informational complexity of an image and of its regions: "a good indicator of an image's informational complexity is its compression rate. The less a portion of an image can be compressed, the more information it contains" (our translation). But a purely informational approach is insufficient from a semiotic perspective because it highlights a too specific kind of visual complexity "that can reveal itself to be far from more general complexities in perception and interpretation." Instead, by taking into account the "development of the psychology of perception, it is possible to concentrate on the reader's perspective, in order to estimate his or her capacity to understand and become engaged by an image" (our translation). Computational models are used by psychologists to simulate perception and estimate visual complexity. The concept of visual saliency is the basis of such approaches: "the estimation of how quickly an element in an image is perceived and of how long attention is paid to it" (our translation). Jégou suggests two biases in this psychological approach. Firstly, it is used to estimate only overall, undirected attention (called preattentional perception); and secondly, it is tailored to "natural" images, "that is, visual scenes that the human eye encounters ordinarily" while cartographic images "are not natural, they are bidimensional graphical representations created artificially which have a certain reading mode (titles, legends and other elements) to which we have developed habits via education" (our translation). Still, for the author visual saliency can be practically used to evaluate if a map responds to its communicative objectives, such as whether information considered important by the map's author is quickly spotted and retains the reader's attention long enough. Jégou then moves on to discuss details regarding the colors of maps and presents a tool he developed for choosing color scales.

Crémier, Bonenfant and Lafrance St-Martin question the notion of raw data in the digital humanities. They examine the meaning-making procedures based on automated data analysis in order to better understand the parameters and shortfalls of data-driven research methods in the human and social sciences. The authors propose a semiotic model of data production and circulation to problematize the idea that data have ceased to stand for a formalization tool and have instead become a direct presentation of the world. Using Peircian semiotics, the authors affirm that digitization is thus an hypersymbolic semiotic process, which brings about a naturalization of meaning, the illusion of iconicity and of rhetorical efficiency.

Sarti, Citti and Piotrowski propose a mathematical framework for the notion of "differential heterogenesis" proposed by Deleuze. The important issue is to understand what lays beneath signs–how to model the presemiotic apparatus embodied in the brain that enables the emergence of meaningful forms. For the authors, existing structuralist and morphogenetic approaches insisted on a

static, homogeneous and globally well-defined set of constraints. Instead, Deleuze's insight suggests that the field of constraints itself should be defined as the dynamic result of higher-order operations. These new fields lack any fixed reference–individual positions within them are "floating twice," so to speak because they are variables into already variable spaces of reference. For the authors, this constitutes a radical innovation compared to any approach based on the way in which contemporary physics operate:

> In opposition to the kind of differential calculus that is usually adopted in mathematical-physical modelling, which tends to assume a homogeneous differential equation applied to an entire homogeneous region, heterogenesis allows differential constraints of qualitatively different kinds in different points of space and time. These constraints can then change in time, opening the possibility for new kinds of differential dynamics and the emergence of distinct entities and forms. (Sarti et al. forthcoming)

From Sarti et al.'s standpoint, each brain constitutes its own unique world derived from an ontogenetic history of space-formation – the higher-order morphology the authors call heterogenesis. The mathematical implementation proposed by the authors is based on the three concepts of lift operators, assemblages of operators and commutators. A lift operator applied to an empty Cartesian space returns a homogeneous vector field. But what happens if we lift a lifted space or, in other words, if we intersect lifted spaces? We obtain an assemblage of operators, whose result are vector spaces which are now truly heterogeneous (they do not follow any regularity proper to all spaces independently of how they have been produced). This happens because the intersection of lifted fields "contains commutators that did not exist in each of the lifted operators separately and the interaction is much more than the simple union of the collected vector fields. Commutators interpret in a formal way the differences of differences which are so important in the Deleuzian construction." In a morphological perspective, "if the assemblage of operators is considered in turn a new differential operator, heterogenesis can be viewed as a morphogenesis of the assemblage operator."

References

Adorno, T. & M. Horkheimer. 2002 [1944]. *Dialectic of enlightenment*. Stanford, CA: Stanford University Press.

Alge-Hewitt, M., R. Heuser & F. Moretti. 2015. On paragraphs: Scale, themes, and narrative forms. *Stanford Literary Lab* 10.

Anscombe, G. E. M. 1957. *Intention*. London: Blackwell.

Baronchelli, A., V. Loreto & A. Puglisi. 2015. Individual biases, cultural evolution, and the statistical nature of language universals: The case of color naming systems. *PLoS ONE* 10. 5.

Barthes, R. 1966. An introduction to the structural analysis of narrative. *New Literary History* 6(2). 237–272.
Barthes, R. 1974 [1970]. *S/Z: An essay*. New York: Farrar, Straus and Giroux.
Benjamin, W. 2008 [1939]. *The work of art in the age of mechanical reproduction*. London: Penguin.
Bertin, J. 1983 [1967]. *Semiology of graphics*. Madison, WI: Wisconsin University Press.
Black, M. 1962. *Models and metaphors: Studies in language and philosophy*. Ithaca, NY: Cornell University Press.
Bratman, M. 1999. *Faces of intention: Selected essays on intention and agency*. Cambridge: Cambridge University Press.
Cagé, J., N. Hervé & M-L. Viaud. 2017. *L'information à tout prix*. Paris: Ina Editions.
Carel, M. 2011. *L'entrelacement argumentative: Lexique, discours et blocs sémantiques*. Paris: Honoré Champion.
Casilli, A. & D. Cardon. 2015. *Qu'est-ce que le digital labor?* Paris: Ina Editions.
Chapman, P., G. Stapleton, A. Moktefi, S. Perez-Kriz & F. Bellucci (eds.). 2018. *Diagrammatic representation and inference*. Berlin: Springer.
Chul-Han, B. 2015. *The transparency society*. Stanford, CA: Stanford University Press.
Colquhoun, D. 2014. An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science* 1.
Colquhoun, D. 2017. The reproducibility of research and the misinterpretation of p-values. *Royal Society Open Science* 4.
Compagno, D. (ed.). 2018. *Quantitative semiotic analysis*. Berlin: Springer.
Davidson, D. 1980. *Essays on actions and events*. Oxford: Oxford University Press.
Derrida, J. 2011 [1967]. *Voice and phenomenon: Introduction to the problem of the sign in Husserl's phenomenology*. Evanston, IL: Northwestern University Press.
Dilthey, W. 2010. *Understanding the human world: Selected works of Wilhelm Dilthey*, vol. 2. R. A. Makkreel & F. Rodi (eds.). Princeton, NJ: Princeton University Press.
Doueihi, M. 2011. *Digital cultures*. Cambridge, MA: Harvard University Press.
Doueihi, M. & F. Louzau. 2017. *Du matérialisme numérique*. Paris: Hermann.
Eco, U. 1976 [1975]. *A theory of semiotics*. Bloomington, IN: Indiana University Press.
Eco, U. 1979. *The role of the reader: Explorations in the semiotics of texts*. Bloomington, IN: Indiana University Press.
Eco, U. 1994 [1990]. *The limits of interpretation*. Bloomington, IN: Indiana University Press.
Eco, U. 2000 [1997]. *Kant and the platypus: Essays on language and cognition*. New York: Harcourt.
Fisher, R. A. 1947. *The design of experiments*. Edinburgh: Oliver and Boyd.
Fusaroli, R., B. Bahrami, K. Olsen, A. Roepstorff, G. Rees, C. D. Frith & K. Tylén. 2012. Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science* 23. 931–939.
Fusaroli, R., M. Perlman, A. Mislove, A. Paxton, T. Matlock & R. Dale. 2015. Timescales of massive human entrainment. *PLoS ONE* 10. 4.
Galantucci, B. & S. Garrod. 2011. Experimental semiotics: A review. *Frontiers in Human Neuroscience* 5. 11.
Geertz, C. 1973. *The interpretation of cultures*. New York: Basic.
Greimas, A. J. & P. Ricoeur. 1989. On narrativity. *New Literary History* 20(3). 551–562.
Hesse, M. 1966. *Models and analogies in science*. Notre Dame: Notre Dame University Press.

Hjelmslev, L. 1961 [1943]. *Prolegomena to the theory of language*. Madison, WI: Wisconsin University Press.
Huff, D. 1954. *How to lie with statistics*. New York: Norton.
Jenkins, H. 2006. *Convergence culture*. New York: New York University Press.
Manovich, L. 2013. *Software takes command*. New York: Bloomsbury Academic.
Merzeau, L. 2009. Du signe à la trace: L'information sur mesure. *Hermès* 53(1). 21–29.
Meunier, J.-G. 2017. Humanités numériques et modélisation scientifique. *Questions De Communication* 31(1). 19–48.
Moretti, F. 2005. *Graphs, Maps, Trees: Abstract Models for Literary History*. London & New York: Verso.
Moretti, F. 2013. *Distant reading*. London & New York: Verso.
Moretti, F. 2015. "L'opérationnalisation" ou, du role de la mesure dans la théorie littéraire moderne. *Critique* 819–820. 712–734.
Morozov, E. 2013. *To save everything click here*. New York: Public Affairs.
Navas, E., O. Gallagher & X. Burrough (eds.). 2014. *The Routledge companion to remix studies*. London: Routledge.
Peirce, Charles S. 1992. *Essential Peirce: Selected philosophical writings*, vol. 1, 1867–1893, N. Houser & C. Kloesel (eds.). Bloomington: Indiana University Press. [Reference to vol. 1 of *Essential Peirce* will be designated EP 1.].
Piper, A. 2016a. There will be numbers. *Journal of Cultural Analytics* http://culturalanalytics.org/2016/05/there-will-be-numbers/ (accessed 20 June 2019).
Piper, A. 2016b. Fictionality. *Journal of Cultural Analytics* http://culturalanalytics.org/*2016*/12/*fictionality*/ (accessed 20 June 2019).
Rastier, F. 2011. *La mesure et le grain: Sémantique de corpus*. Paris: Champion.
Rebillard, F. 2011. L'étude des médias est-elle soluble dans l'informatique et la physique? *Questions de Communication* 20. 353–376.
Rogers, R. 2013. *Digital methods*. Cambridge, MA: MIT Press.
Sperber, D. 1996. *Explaining culture: A naturalistic approach*. New York: Wiley.
Stiegler, B. 1998. *The fault of Epimetheus* (Technics and time 1). Stanford, CA: Stanford University Press.
Stiegler, B. 2010. *Cinematic time and the question of malaise* (Technics and time 3). Stanford, CA: Stanford University Press.
Stjernfelt, F. 2007. *Diagrammatology: An investigation on the borderlines of phenomenology, ontology, and semiotics*. Berlin: Springer.
Tufte, E. 1983. *The visual display of quantitative information*. Cheshire, CA: Graphics.
Turner, F. 2006. *From counterculture to cyberculture*. Chicago: Chicago University Press.
Tylén, K., R. Fusaroli, P. Bundgaard & S. Østergaard. 2013. Making sense together: A dynamical account of linguistic meaning making. *Semiotica* 194(1/4). 1–24.
von Wright, G. H. 1971. *Explanation and understanding*. Ithaca, NY: Cornell University Press.
Wilkinson, L. 2005. *The grammar of graphics*. Berlin: Springer.