Semiotic approaches to big data visualization

BY: **Pierluigi Basso Fossali, Maria Giulia Dondero, Lia Yoka**

1. Our project

With a few exceptions - focusing on journalism (Compagno 2017), on works of art (Chartier, Pulizzotto, Chartrand & Meunier 2019, Dondero 2020), on deep fake videos (Leone 2021, Dondero 2021) and on data epistemology (Bachimont 2018) - semiotics has been late to approach *big data*. In contrast to information and communication studies, that have been quick to recognize the big data revolution in medical, biological, political, urbanistic, and journalistic practices, and while digital humanities and digital art history continue to offer crucial insight with their analysis of large collections of artworks and heritage objects, semiotics has yet to live up to the big data challenge.

This tardiness is due to several factors: (i) the prevalence of the structuralist tradition's textualism, which encourages sophisticated analyses of either single objects of study, such as the production of a single author, or limited corpora of works; (ii) the belief that general semiotic theory is applicable to any cultural object, and the consequent assumption that there is no need to adapt semiotic models to new media or practices; (iii) the habit of working with objects endowed with *motivated* meaning (myths, paintings, novels, political discourse, etc.), whereas an archive of big data (a non-text in the traditional sense) is not considered a meaningful object, but rather a device that *opens up* to practices of meaning, meaning that is yet to be located, inferred and invested. Last, but not least, one of the most important reasons of the defiance of semiotics towards current big data phenomena is (iv) its outright rejection

ARTICLE INFO:

Volume: **08**

Issue: **01**

Summer **2022**

ISSN: **2459-2943**

DOI: **10.18680/hss.2022.0001**

Pages: **5-12**

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0001 | **punctum.gr**

of a theory of communication understood as a mathematical theory of information (Shannon & Weaver 1949), and, as a consequence, its distrust towards all theories for which *sign* and *signal* are interchangeable concepts.

A semiotic theory of meaning based on the close interlinking between the *expression plane* and *the content plane* can be useful in complexifying mechanistic approaches that either study the expression plane of artistic objects (see, for instance, Manovich's *Media Visualization*, Manovich 2020, and the *Replica* project, Seguin 2018), and are thereby indifferent to content, or privilege thematic and language-dependent analyses of works of art, which, conversely, subordinate expression to content. Furthermore, semiotics needs to contribute to the analysis of large collections in both diachronic and synchronic terms, and explore the extension or even the overall revision of some of its own key concepts such as *enunciation*, *isotopy*, *semi-symbolism*, etc.

In summary, this *Punctum* issue has attempted to:

1. contribute to the global discussion of possible semiotic approaches to big data theory and methodology;
2. explore specifically an epistemological approach to visualizations of big data, i.e. to their manipulation, design, display, and interpretation;
3. study, from a critical vantage point, the ideologies and epistemic perspectives that underlie the acts of collecting and visualizing big data, prior to their analysis.

Our goal is to trace the path that leads from visualization, regarded as a unifying representation of disparate data, to its full elevation to an interpretative device.

At the end of this introduction, we will take a short overview of the selected contributions. Ultimately, this issue has been an opportunity to assess the inscription of semiotic reflection in digital humanities, to value both the specific contributions of the discipline, and the demand to update epistemological frameworks and methodological approaches. Within the novel lines of argumentation and conceptual aspects these contributions seek to highlight, enunciative and rhetorical strategies of data management, forms of visualization and patterns of interpretation, the punctum of relevant circumstantial details and the panoramic *studium* of the mass of available data, all intersect and often collide with each other.

2. Historical perspective and complementary dimensions of Digital Humanities

In contrast to our initial intention when suggesting this issue, the contributions collected here focus more on the *present* and *future* of digital humanities, and less on the *past*, and thus on the historical perspective. For example, while art is very often being taken into account, (digital) art history hardly is. We thus take the liberty in this introduction

to mention a few points that can be considered complementary to what the reader will find in the articles. Notably, we would like to introduce some suggestions in the form of paired aspects, aspects potentially complementary, that often present themselves in the form of opposites or incompatibles.

(i) The first complementarity that resolves an old opposition is that between *quantitative* and *qualitative* analysis: it is increasingly evident that the analysis of vast corpora brings to the fore new objects of observation. Quantitative studies, when well conducted, create an important qualitative dimension for research. They do not only secure well-encoded results, but also allow a form of emancipation from the evidence of the already known, in other words, they mobilize active knowledge. At the end of the 1950s, Gilbert Simondon had already suggested the need to integrate computer technology in culture as a factor of emancipation, a factor that could release an organizing potential beyond finalities taken as justifications (Simondon, 1958: 105).

(ii) However, if we are to undertake a "comparativist visual semiotics", truly articulable with a "digital critical hermeneutics" (Rastier 2011), rooted in protocols of objectification and leading to practices of interpretation, it is necessary to resolve the classical opposition between the acceleration of computation, to the point of near "immediacy" of the results obtained, and the "studium", the slow analysis characterized by a truly interpretive approach, capable of elaborating on the difference, or rather on the "gap", the critical distance between the elements within a corpus. In short, the point is to render *enhanced computation* and *extended mediation* reconcilable with each other, so that the processing speed is not merely a direct performative condition of *felicity* for the results and furthermore, so that the multiplication of mediations, of "re-entries" into mediations already carried out, allows for the progressive revelation of new features in the objects under scrutiny according to a multiplicity of perspectives.

(iii) This struggle against the evidence of data in favor of a critical and situated hermeneutics leads us to look less at the latest technical breakthrough and more at the integration of informational and interpretive methods, an integration that remains largely incomplete. Innovations in image processing have less to do with the ambition of providing a new single technology. Rather, they consist of an unprecedented assembly of image-processing algorithms and machine learning techniques, such as reinforcement learning or incremental metric learning. These methods are interconnected and guided by an interpretative grid of cultural objects that relates their forms of composition on different levels of relevance. In this framework, the cultural object and its reproducibility no longer have to be thought of in terms of qualitative opposition, and always have a fixed starting point, namely an *origin*. What is at stake is a philological approach, an approach that will respect the identity of the cultural objects analyzed, and that is compatible with the recognition that the objects continue their own form of life in new ecological niches. For example, the museum of fine arts was not necessarily the pre-assigned place to host the aesthetic relationship with a work of art originally

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0001 | **punctum.gr**

located in a church and integrated within an iconographic program, itself composed within a vast network of other works or ritual objects. This, however, does not mean we should get rid of places of experience, but rather to reconstruct the variety and complexities in the reception and interpretation of cultural objects. In this sense, we can make compatible a *dematerialization* of documents (for example, the intertext) and a *reintegration* of in vivo *experience*.

(v) The complementarity between the search for the originality of a work, according to an open system of relevance, and the anchoring in a "site-specific experience" can coexist in the digital analysis of the life form of a cultural object, a life form changing over time. We might also speak of a relationship between the "virtual complementary spaces and a marked inter-contextuality", in the sense that the bringing together of multiple contexts does not fail to point to a *marked space*, a cultural gravitational space in which different contributions, once attracted, guarantee that the experience, though augmented, is not "without a sense of place."

3. Data visualization and augmented reality

The relationship between data visualization and augmented reality is rarely understood as crucial, as if the mass of available information could itself direct our "seeing in." However, an archive of data or even a well constituted and carefully selected corpus of data is never self-sufficient. Forms of "augmentation" are always necessary in order for us to liberate ourselves from the debilitating idea that a framework can never be fully reconstituted. This augmentation, relating an object to a cultural reality, will also serve to activate a "seeing as", if not a "seeing through", that is indispensable if we are to make fertile comparisons between domains of knowledge, as well as between constituted domains and a cross-domain that questions meaning itself.

Augmented reality is a symbolic circuit much more complex than the simple normative welding between two layers of reality that have pre-established and fixed roles. Far beyond a kind of amalgam, a work that is compatible and comprehensible must be accomplished. For example, it is not enough to recognize a standard model describing the reality of heritage, we would need a work of individuation that would renew the very model of heritage and thus allow for an optimal model of visualization.

Alternatives in data visualization are not simply options or tests unilaterally conducted by the analyst. The mass of data seems to release them from their cultural status, reducing them to information waiting to be exploited. In fact, visualization should be precisely a perspective that modulates its internal possibilities and then stabilizes itself where it encounters an interpretive question that can be associated

with a body of objects emerging from the bottom of the information archive. And it is left to artificial intelligence and its deep learning to offer *salient* visualizations. Meaningful results (namely *pregnant* results) occur when the expression plane of emergent forms can be associated with content, and the latter is rarely already immediately available. Instead, it requires "augmentation", collateral knowledge, other corpora already instituted. This does not indicate a relativization or a downgrading of the contribution of the digital moment, but more simply signals that the latter must have confidence in its own re-entries, i.e. in the circuit between visualization results and augmentation possibilities, which immediately entrusts the entries with interpretive depth. For example, visual Big Data do not constitute a totalizing, homogeneous, synchronic frame of History. Big Data Analyses allow us to discover a new nexus of cultural tensions (technical, conceptual, social, artistic *etc*.), but this does not prevent, indeed it fosters, the recognition that images are responses to other images that constitute themselves, sometimes unpredictably, as dialogic and polemical spaces.

Finally, there is the need to compose, in the analysis of visual Big Data, a diachronic perspective that can recognize genealogies of transformations of forms. If such a project is called "the life of forms", it is because forms are not conceived as following a linear history. As a matter of fact, several transversal dynamics have to be acknowledged: (a) influences coming from different genealogies of forms, (b) evolutions in techniques and in materials, (c) the contingency of reception. The intersection of autonomous factors (i-iii) with heteronomous ones (a-c) implies, for example, that Focillon's work (1934) in art history is both emancipated from formalism and determinism as well as deeply pragmatic, from top ("the tool awakens form from matter") to bottom (the impact of a form on its era is always controversial). Forms question, and are questioned by, a horizon of intentions (production) and of expectations (reception).[1]

[1] The life of forms is at the heart of the project ANR Augmented Artwork Analysis (https://anr.fr/Projet-ANR-20-CE38-0017). Regarding the epistemology of artwork interpretation, between analysis of large visual databases and augmentation of the object in its usual location, namely the museum, the *AAA* project stands in contrast to three major forms of reductionism. First, the reduction of the art image to a type or to general categories: the *AAA* tool engages the image as a nexus of cultural tensions (technical, conceptual, social, artistic, etc.) and thus as a specific historical solution. Second, the reduction of the art image to a "cultural singularity": the *AAA* considers the image to be inseparable from a system of other images, both congruent and deviant (every dimension of the image makes sense in the way it maintains or breaks linkages with the legacies of whole families of images). Third, the reduction of the art image to a historically derived product: the *AAA* represents the art image within families and traditions spanned by cultural tensions, naturally, but the image cannot be reduced to a totalizing, homogeneous frame that is standardized or stereotyped and called "History" (with a capital H) (Koselleck 1986). Within this perspective, the art image is always a response to other works; it is meant to be an alternative to other efforts, or a critique of them (Baxandall 1985; Damisch 1987).

Punctum. International Journal of Semiotics | **08**:01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0001 | **punctum.gr**

4. Contributions: open questions in designing and interpreting data visualisation

The issue opens with **Johanna Drucker's** acknowledgement and analysis of Jacques Bertin's 1967 *Sémiologie Graphique*, a work that is still indispensible for design in general and data visualisation in particular. His description of graphic variables (color, texture, value, pattern, shape, position, and orientation) offers concise categories that help us better grasp visual entities within a semiotic system, no matter what scale. Drucker devises new fundamental categories complementary to Bertin's, in accord with features of dynamic display of today's technology, and suggests their standardisation an instrumentalisation in Big Data visualisation and digital design.

From a different point of view, that of the user of data visualisation rather than the designer, **Federico Biggio** addresses the "framing of truthfulness" in data images. The "interpretive advantage" offered by visualisation software can turn into its opposite. In an exercise in applied semiotics, Biggio illustrates the need for a constant awareness that subjectivity is being constantly modulated in data images, and therefore also that, as Jacques Fontanille has noted, "every data image is the result of a computational enunciative practice."

A crucial part of this "framing of truthfulness" mentioned in the previous text is the management of "latent space" in a deep-learning system. **Cristina Voto** analyses the work of the artist-coder Jake Elwes, *Zizi-Queering the Dataset* (2019), in order to flesh out exactly how rhetorical and ideological premises of data visualisation are inherent in Artificial Intelligence software. Using Elwes' discussion of gender-parameter design bias, Voto points to the fact that the visualisation of latent space, i.e. mathematical map of what the Neural Network has learnt from the training dataset, helps us deconstruct and reveal these presumptions and prejudices.

Indeed, data visualisation is hardly transparent or objective, it is hardly a direct representation, as **Valeria Burgio** confirms. Elaborating on examples of scientific and journalistic images, Burgio employs the semiotic concept of enunciation in a way similar to that in critical design theory and digital humanities, and offers a reading of data visualisation as expression, and as interpretation. She discusses the uncertainty of the visualiser vis-a-vis the data observed, and suggests that the presence of this interpreter/visualiser be enhanced, so that the enunciative traces of visualisation become more evident.

Besides stressing the rhetorical and the interpretive moment, we need a clear process of analysing visualisation techniques one by one, in a reflexive informational framework that **Valentina Manchia** calls "discourse of data." This is true especially of media visualisation based on big cultural data, namely "visualization without reduction", as Lev Manovich has described it. With her two aptly chosen case studies, Manchia shows how data is the "chanelled result" of putting together a series of what Bastide has theorized as "devices of visualisation."

The issue ends on an interestingly confessional note. **Alon Friedman** and **Martin Thellefsen** discuss one of the most popular open-source applications for managing and analyzing social media data, the open-source R programming language, and test it against Peircean sign theory. While there is indeed a lot of space to further scrutinize, and expose, the techniques, prejudices, decisive interpretive moves and theoretical premises in data visualisation (its design and its study), as is shown in all contributions to this volume, it is much harder, as Friedman and Thellefsen state, to create algorithms that generate or suggest and interpretation of visual signs.

We deeply thank all the authors for their excellent texts on such a current, dynamic and fleeting subject, and for their willingness to cooperate with us all the way, for the creation of a coherent and challenging result. We are grateful to all the friends and colleagues who kindly offered their time and expertise: Anne Beyaert. Karin Boklund-Lagopoulou, Enzo D'Armenio, Rossana De Angelis, James-Peter Lancaster, Claudio Paolucci, Joao Queiroz, Alberto Romele, Marta Severo, Yannis Skarpelos, Matteo Treleani, Andrea Valle.

We extend our special thanks to the artist Paho Mann for kindly granting us permission to publish his work on the cover.

References

Bachimont, Bruno 2018. Between Data and Formats: when Communication becomes Recording. In: A. Romele & E. Terrone (eds.), *Towards a Philosophy of Digital Media* Basingstoke (UK): Palgrave Macmillan, 13-30.

Baxandall Michael 1985. *Patterns of intention. On the Historical Explanation of Pictures*. New Haven and London: Yale University Press.

Bertin Jacques 1967. *Sémiologie graphique*. Paris: Mouton et Gauthier-Villars.

Chartier, Francois, Davide Pulizzotto, Louis Chartrand and Jean-Guy Meunier 2019. A data-driven computational semiotics: The semantic vector space of Magritte's artworks, *Semiotica* 230: 19-69.

Compagno, Dario 2017. Signifiant et significatif. Réflexions épistémologiques sur la sémiotique et l'analyse des données. *Questions de communication* 31: 49-70.

Damisch, Hubert 1987. *L'origine de la perspective*. Paris: Flammarion.

Dondero, Maria Giulia 2020. *The Language of Images. The Forms and the Forces*. Dordrecht: Springer.

Dondero, Maria Giulia 2021. Composition and decomposition in artistic portraits, scientific photography and in deepfake videos. *Lexia* 37-38: 437-453.

Focillon, Henri 1943 [1934]. *Vie des formes suivi de "Éloge de la main."* Paris: Presses Universitaires de France.

Punctum. International Journal of Semiotics | **08**:01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0001 | **punctum.gr**

Koselleck, Reinhart 1986. Sozialgeschichte und Begriffsgeschichte. In: W. Schieder and V. Sellin (eds.), *Sozialgeschichte in Deutschland*, vol. 1, Göttingen: Vandenhoeck & Ruprecht.

Leone, Massimo (eds.) 2022. L'idéologie sémiotique des deepfake. *Interfaces numériques* 11/2: 1-16.

Manovich, Lev 2020. Computer vision, human senses, and language of art. *AI & Society*, November 22. http://manovich.net/content/04-projects/109-computer-vision-human-senses-and-language-of-art/manovich_computer_vision.pdf

Rastier, François 2011. *La mesure et le grain. Sémantique de corpus*. Paris : Champion.

Seguin, Benoit 2018. Making large art historical photo archives searchable, *EPFL Scientific Publications*. https://infoscience.epfl.ch/record/261212

Shannon C.E. and W. Weaver W. 1949. *The mathematical theory of communication*. Chicago: University of Illinois Press.

Simondon, Gilbert 1989 [1958]. *Du mode d'existence des objets techniques*. Paris: Aubier.

AUTHOR

Pierluigi Basso Full professor of Language Sciences at the Lyon 2 University and Director of the ICAR Laboratory at the ENS de Lyon.

AUTHOR

Maria Giulia Dondero F.R.S.-FNRS Research Director Professor University of Liège.

AUTHOR

Lia Yoka Associate professor of History and Theory of Art and Culture at the Department of Architecture, Aristotle University of Thessaloniki.