Distant reading in literary studies: a methodology in quest of theory

Fabio Ciotti

Università degli Studi di Roma Tor Vergata
(fabio.ciotti@uniroma2.it)

Abstract

Since Franco Moretti coined the successful term distant reading, quantitative/computational text analysis methods have gained wide circulation in literary studies. The diffusion of distant reading approaches has raised a lively debate and has attracted various criticisms, both from "traditional literary scholars" and from self-critical adopters. One important reason underlying these critical positions is the fact that it lacks sound and coherent rationales from the point of view of the theory: distant reading is the first methodology in literary studies that does not come with a theory of literature embedded in it. Consequently, all distant reading studies derive their theoretical frameworks and terms from literary theories that mostly rely on the notion that literary texts can be explained only by the way of interpretation. On what grounds, then, can we construct a theory of literature amenable to distant reading methods? I think that the better theoretical frameworks are the cognitive and bio-evolutionistic approaches to literature and cultural evolution studies. These theoretical approaches require a change in the level of description of the literary domain and justify the move from "interpretation" to "explanation" as the real aim of the scholarly inquiry.

1. Introduction

Since Franco Moretti coined the widely successful term of *distant reading* (Moretti 2000) quantitative computational text analysis in literary studies has escaped from its historical fate of being a "niche activity" to gain a wide circulation in the literary debate. We can even speak of a "distant reading critical school", nowadays, although the more generic label of "computational literary criticism" is also used. This success has been supported and fueled, so to say, by the recent developments in the theories and techniques of Data/Text Mining and Machine Learning, which have brought into the field new methods and computational tools like probabilistic topic modeling, support vector machines, naïve Bayes classifiers, word embedding, and machine learning. Needless to say, the more general techno-cultural hype around Big Data and Artificial Intelligence, more recently, has contributed to the background conditions in which all this happened.

Nowadays, we can find many literary scholars engaged in research projects based on quantitative and statistical computational processing of textual collections, in network analysis, and other related techniques and methods. On the other hand, if we come to the actual outcomes of distant reading in literary criticism, after almost two decades of experimentations the landscape is still unsatisfactory. I am not saying that there are no results at all; but, in general, we must admit that they have a limited impact in the mainstream literary-critical or historical discourse (especially if we compare the momentum of other approaches in the past, say Structuralism, Deconstruction,

Cultural studies or New Historicism). My view is that this is due to a lack of theory, or better, to a wrong positioning in the space of theories available in the literary field. The rest of this paper will try to argue this position.

2. Definitions and theoretical underpinnings

It is important to premise some preliminary definitions and theoretical points: first, in this article, I am referring exclusively to the role and impact of quantitative/computational methods in literary and cultural studies; in other areas of the humanities, like synchronic and diachronic corpus linguistics or quantitative historiography, as in quantitative social sciences, computational tools and statistical methods have always been acknowledged an important role.

Second, I will adopt the scholarly label "distant reading" in a rather specific sense[1], encompassing *quantitative* and *large-scale data-driven* analysis of literary phenomena executed by the way of complex computational methods. I want to stress the fact that I refer to **large-scale** data-driven analysis: I know very well that adopting the term "large" in reference to literary data can be seen as a misuse of the term if confronted with the dimensions of the data sets in hard and social sciences. But all in all, as Moretti pointed out since the very beginning, the scale here is defined by the human limits of what can be *really* read in a given timespan (even if it is the whole lifetime of a scholar). Under this respect, even some thousands of books can work as an example of big data.

As it is well known, the tradition of quantitative and statistical methods in literary scholarship is not new at all: in fact, it is coeval with the foundation of the scholarly field that today we call Digital Humanities (Hoover 2013; Underwood 2017a). However, I think that some distinctive features differentiate traditional computer-aided text analysis and stylometry and contemporary distant reading. They share the fundamental tenet that literary phenomena, to a certain extent, are or can be reduced to a set of measurable features and hence be subject to statistical and numerical analysis. The output resulted from such analysis constitutes the basis for the *understanding* of those same phenomena. But the change in scale and dimensionality of the analyzed document sets does determine a radical change both in the kind of observable phenomena and in the methodology, as Matt Jockers (one of the first collaborators of Moretti at the Stanford Literary Lab, and author of one of the foundational books of the distant reading school) expresses neatly:

> massive digital corpora offer us unprecedented access to the literary record and invite, even demand, a new type of evidence gathering and meaning making. The literary scholar of the twenty-first century can no longer be content with anecdotal evidence, with random "things" gathered from a few, even "representative," texts. We must strive to understand these things we find interesting in the context of everything else, including a mass of possibly "uninteresting" texts. (Jockers, 2013: 8)

[1] Ted Underwood in its article "The genealogy of Distant reading" (2017a) and later in its book *Distant Horizons* (2019) adopts the term with a similar scope, stressing how "Distant reading is simply a new scale of description. It doesn't conflict with close reading any more than an anatomical diagram of your hand would conflict with the chemical reactions going on inside your cells. Instead of displacing previous scales of literary description, distant reading has the potential to expand the discipline — rather as biochemistry expanded chemistry toward a larger scale of analysis."(2019: xvii).

Similar ideas are expressed by another important exponent of the distant reading school Andrew Piper, in the introductory article to the first issue of the "Journal of Cultural Analytics", where he identifies what he calls an "evidence gap" in the generalization's efforts of "traditional" criticism, taking as an example one of the capital attempts of such generalizations in Western history of literature, *Mimesis* by Eric Auerbach:

> And yet even Mimesis' insistence on the representedness of reality was still blind to the representativeness of its own examples. Whether it was the passages that stood for the works from which they were excerpted or the works that stood for the culture from which they were drawn, there was no way for Auerbach to ad-dress the fraught relationship between part and whole. (Piper 2016: 4)

All these contributions, we can say, are delineating a real paradigmatic shift in literary studies. The traditional paradigm oscillates between two poles: textual close reading, the deep critical analysis/interpretation of singular (or limited groups of) texts; speculative and theoretical approaches that *use* the texts (or other derived literary phenomena) as anecdotal and symptomatic evidence of their doctrines (literary scholars, especially theorists, will forgive my oversimplification here). Quantitative distant reading is based, instead, on observations and measurements of quantifiable features in/of the literary texts, and the successive algorithmic elaboration that provides evidence for literary arguments.

The problem and the challenge for computational criticism are therefore that of finding interesting literary objects (and theories) that are amenable to be properly quantified and adapted to the formal requirements of computational processes. "Interesting" here means to be able to license valuable (and possibly new) critical or historical interpretations. Moretti has proposed to borrow the notion of *operational definition* introduced in the epistemology of physics by P. W. Bridgman in 1927 (Bridgman 1927) to characterize this process. Operationalism's main tenet was that "theoretical terms" used in scientific theories could be substituted by the specification of the procedures and instruments required to measure its observable effects, thereby eliminating the apparent references to unobservable entities. Without considering all the philosophical implications and complications of Bridgman's concept (Chang 2019), in its widely cited article "'Operationalizing' or, the Function of Measurement in Literary Theory", Moretti suggests that this methodological approach "means building a bridge from concepts to measurement, and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts." (Moretti 2013b).

3.   The theoretical problems of distant reading

Historically, computational (and quantitative) text analysis has had little impact on the mainstream literary discourse. In the late 80s, after the publication of a book like *Computation into Criticism* (Burrows 1987) written by one of the most important scholars of the early phase of the field, Rosanne Potter opened a critical review essay on the relation between Literary Criticism and Literary computing observing that "Most traditional literary critics remain blissfully ignorant of literary computing" (Potter 1988: 91). Twenty years later (hence well into the "distant reading turn") Willard McCarty reiterates this complaint in a talk entitled "Neglected, not rejected: Is there a future for

literary computing?"[2], where he individuates in a lack of theory the main deficiency of computational literary analysis.

In only a decade this neglect has been abundantly overcome, as the impact of distant reading approaches in the literary theoretical debate has been undeniably relevant. Nonetheless, if we consider the specific critical and historiographical results of computational literary criticism, we still find an unsatisfactory situation. As I have already said, it is not that there are no results at all, but, in general, these results (except for authorship attributions studies, that is nonetheless a rather specific and somewhat limited subfield) have little impact in the mainstream literary-critical or historical discourse, especially if they are compared to the impact and expectations raised by the correspondent theoretical, methodological and even ideological debate.

Many reasons can be found for these difficulties, not last the fact that distant reading is an approach that for the first time aims at introducing strictly experimental and mathematical based methods in the core of the humanities, resulting for most literary scholars as the evidence of an attempt of invasion of their sacred territories by the barbarians coming from the land of the hard sciences (and bearing with them the capitalistic/market-oriented/neo-liberal subsumption of knowledge), against which they stand stating that "Literature is not data" (Marche 2012; Golumbia 2014; Allington, Brouillette and Golumbia 2016).

I think that one of the main reasons is instead the fact that distant reading lacks a solid foundation from the point of view of the theory: in fact, we can say that distant reading is the first methodology in literary studies that does not come with a theory of literature embedded in it, as it was for all its predecessors. Let's take for instance the Semiotic/Structuralist approach: it worked as a method for analyzing literature (*inter alia*), in as much as it was based on the theory (and ontology) of the literary text as a semiotic object, based on the distinction between expression, meaning and reference, on the notions of oppositional structures, of the conventionality of the code, on the non-psychological nature of semiotic processes, and so on.

Actually, Moretti when he started his exploration into the new methodological realm had in mind a well-defined theoretical background. Namely, his theoretical points of reference were the theory of biological evolution (especially in the interpretation given by Ernst Mayr) and the *world system theory*, introduced in social science by Immanuel Wallerstein, that he explored in *Graphs, Map, Trees* (Moretti 2005) and a later short article entitled "Evolution, World-Systems, Weltliteratur", appeared in various venues before being included in the *Distant Reading* book (Moretti 2013a, chap. 6). But Moretti himself recognizes that after the initial phase he has been progressively attracted by the explorative and experimental side of his interaction with computational/quantitative approaches, at the expense of a theoretical systematization of the inquiry, as we can read in the short introduction to the version of the essay republished in his *Distant Reading*:

> the need for a theoretical framework was for a few years forgotten in the heady mood of permanent exploration. As I write, the results of the explorations are finally beginning to settle, and the un-theoretical interlude is ending; in fact, a desire for a general theory of the new literary archive is slowly emerging in the world of digital humanities. It

---

[2] The talk was presented, in a very similar form, in three scholarly events in 2008. The transcriptions texts are available on the personal Web page of McCarty: https://www.mccarty.org.uk.

is on this new empirical terrain that the next encounter of evolutionary theory and historical materialism is likely to take place.

Regarding the intellectual relationship between Moretti and the theory of evolution (and in general the biological epistemology) as a model for literary history, much should be said[3]. My impression is that this relationship has been more on the side of a *mindful metaphorical adoption* than on that of a *deep theoretical convergence*. And it's noticeable that in the mainstream distant reading community these theoretical suggestions in Moretti's works have been almost entirely neglected.

Consequently, all distant reading studies derive their theoretical frameworks and terms from theories in the literary domain that generally relies on the fundamental idea that literary texts can be explained only by the way of interpretation or if we prefer *hermeneutics*. Computational literary analysis, in this sense, would fall inside the border of the critical/semiotic interpretation of the text(s) that, according to Umberto Eco, has the objective to discover the textual and contextual structures that generates all the possible semantic interpretations of the actual readers and the critics when they act as readers (Eco 1990: 29):

> L'interpretazione semantica o semiosica è il risultato del processo per cui il destinatario, di fronte alla manifestazione lineare del testo, la riempie del significato. L'interpretazione critica o semiotica è invece quella per cui si cerca di spiegare per quali ragioni strutturali il testo possa produrre quelle (o altre alternative) interpretazioni semantiche.

Or, if we prefer, it would be perfectly compatible with the theory of interpretation developed by Ricoeur with his notion of structural explanation: an explanation directed toward the analysis of the internal relations of the text (the parts) (Ricoeur and Thompson 1981).

The problem is that any literary interpretation based on a quantitative, immanent, and purely formalist approach is subject to the theoretical criticism that was expressed by Stanley Fish in his harsh and seemingly ultimate criticism of stylistics in "What Is Stylistics, and Why Are They Saying Such Terrible Things About It?" (a talk given in 1973 at the English Institute, later published in his famous book *Is there A text In This Class?* (Fish 1980)). Basically Fish, advanced two radical criticisms to stylistic studies, that apply perfectly to contemporary computational criticism:

1. stylistics (and similar formalist approaches) presupposes an "arbitrary relationship between the specification of formal patterns and their subsequent interpretation"(Fish 1979: 129) so that there is a *non sequitur* between the descriptive account of the text and its interpretation;
2. more radically, Fish argued in a later essay that "formal patterns are themselves the products of interpretation and that therefore there is no such thing as a formal pattern, at least in the sense necessary for the practice of stylistics: that is, no pattern that one can observe before interpretation is hazarded, and which therefore can be used to prefer one interpretation to another"(Fish 197: 130).

---

[3] Michele Cometa in his last book *Letteratura e darwinismo* dedicates a thoughtful chapter to this (Cometa 2018)

The point for Fish was not to criticize the methods *per se*, but the possibility to extract meaningful literary interpretations directly from the simple linguistic facts, the idea of an "algorithmic interpretation" (Fish's words!), since interpretation always starts from a contextual and situated point of view that pre-defines the very objects of its actuation. It is noticeable that to avoid the accusation of relativism and solipsism, Fish introduced the notion of *community of interpreters* that guarantees to the interpretation a socially based inter-subjectivity.

About the same years another great scholar, Cesare Segre, in an essay devoted to the concept of style, expressed his doubts that a purely stylistic approach could be of any use for the interpretation of a text as an esthetic object (Segre 1985: 322):

> Se invece si mira all'interpretazione del testo come prodotto artistico, si deve considerarne la lingua come un sistema autonomo e autotelico. Sacrosanto perciò quanto dichiarano le Thèses di Praga del '29, che «l'opera poetica è una struttura funzionale, e i vari elementi non possono essere compresi al di fuori della loro connessione con l'insieme».

The literary text, adopting Jury Lotman's words (Lotman 1970), is a secondary modeling semiotic system that builds on the linguistic level, but it is not at all exhausted by it.

It is out of scope following here Fish and Segre in the details of their arguments. The fact is that these arguments are still a strong methodological criticism of the efficacy of quantitative methods in literary criticism and interpretation. It is not surprising that this rather old essay of the eminent North American critic has recently attracted the attention of Ted Underwood, one of the more theoretically aware exponents of the distant reading movement (Underwood 2017b). I want to stress that in revamping Fish's criticism, I am not neglecting here the possibility that some good and interesting hermeneutical and critical work can be carried out also by adopting computational analysis, or that, in some specific and specialized areas, computational methods proved to be effective (for example in empirical authorship attributions studies or linguistic reuse detection, but again, outside the scope of interpretation).

My position is that as long as distant reading methods - especially data-driven text analytics - are adopted inside the traditional hermeneutical framework typical of literary studies they are deemed to produce unsatisfying results. After all, it's worth remembering that for Ricoeur the *real* interpretation of a literary work is what he calls the *understanding*, that is a process directed at grasping the holistic meaning of a text, what the text talks about, in order to include it into the subjectivity of the interpreter.

4.   An example: topic modeling[4]

As an example, we can take one of the most widespread methods adopted in distant reading analysis, *topic modeling*. Is it possible to assign a clear theoretical status in 'classical' literary theory and criticism to the notion of topic model(ing)? Clearly, this quest has nothing to do with the mathematical and statistical foundations of the method, which are firmly based on Bayesian

---

[4] This paragraph is a summary and translation into English of (Ciotti, 2016).

statistics and probability theory, or with the technicalities of one or another of its algorithmic implementations.

From a technical point of view topic models are unsupervised text mining techniques whose properties are so summarized by David Blei (2012): "Topic models are algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents. Topic models can organize the collection according to the discovered themes". Many different algorithms are going under the label of topic model, but the most known and widespread is by far the *Latent Dirichlet Allocation* (LDA) TM, based on Bayesian probability theory (Blei, Ng and Jordan 2003).

Intuitively we can say that the LDA topic model is based on a naïve generative model of the text. It specifies a simple probabilistic process by which documents can be produced, based on two main assumptions: texts are composed of a set of underlying (latent) subjects; subjects, on their part, are manifested inside the text as sets of words and there is a way to measure the relevance (and hence the presence rate) of each word inside a specific subject[5]. In technical terms, we say that:

1. a text is a distribution of probability over a set of topics;
2. a topic is a distribution of probability over a set of words.

This idealized generative model can be reversed adopting the methods of Bayesian probabilistic inference. If we conceive the actual textual data distribution as the evidence and the topic structure as the hidden variables that determine it, we can compute the conditional posterior distribution of those hidden variables[6]. This amounts to extracting an underlying set of topics from a set of documents, their probability distribution over documents, and the distributions of single word types over the topics. Notice that in the inferential version of topic modeling, the number of topics to be extracted cannot be determined by the algorithm itself and must be provided as an a priori parameter. Topics are distributions of probability over a subset of the dictionary D of the corpus, they are, in other terms, lists of word types ordered on the probability to belong to that topic. There is no other structure or property that characterizes the topic. Consequently, topics are to be interpreted and labeled to be useful, and there is no certainty that a topic word list is semantically coherent (whatever sense of semantic coherence we can adopt). Keep in mind that the assignment of a single token to a topic is based on the fact that it simply co-occurs in the same document with other words of the topic: in other terms, topics are collocations whose context is the document. Additionally, topic modeling assumes that the distribution of the words inside the text is random and that there is no motivation for the word w^i+1 to appear after the word w^i; that is the probability of the former to appear is not determined by the probability of the latter. This is a very strong assumption for language in general, but it is particularly problematic for literary text where, as

---

[5] When an author wants to write a text, in the first place, he makes a choice on the set of subjects he wants to write about, the topics of his text, and subsequently he defines the proportion by which any single topic is presented in the document. Now our writer will randomly draw from each bag/topic a number of words proportional to the relevance that she has assigned to any topic. In the end, she will lay down in a casual sequence the words she has extracted from the bags, and there it is her masterpiece! Obviously, the frequency distribution of each word type is determined by the relevance of the topics in the text and by the relevance of the words inside the topics.

[6] This is achieved assuming a prior casual distribution of the word tokens over the topics and applying recursively an algorithm (*Gibbs sampling*) that updates this distribution on the base of the current frequency of the word type in the topics and of the number of word tokens in the document belonging to the topic (plus a couple of hyper-parameter). After *n* iterations, the algorithm converges in producing a stable assignment of each word token in each document to one or more topics.

Jakobson observed "the poetic function projects the principle of equivalence from the axis of selection into the axis of combination" (Jakobson 1960: 368).

Given this mathematical characterization of topic modeling, we can go back to the theoretical question we posited: what kind of literary theoretical concepts can be operationalized by the probabilistic computational concept of a topic model? What are the literary phenomena of which the topic model could act as a proxy? It is rather hard to systematically individuate the adequate candidates in the vast theoretic literature about the semiotic structures of the literary text. One driving criterion is the fact that a topic model must be related to some semantic features of the text that emerge from its linear verbal manifestation (or discursive level). These features, furthermore, must be semantically distinctive; they must deal with the overall aboutness of the text. Building on these two preliminary assumptions, I propose the following candidates.

1. theme and motif
2. isotopy
3. discourse

The concepts of theme and motif are two of the most ancient and controversial theoretical terms in literary studies. It is not possible here even trying to hint at the vast number of definitions and interpretations (not rarely contradictory) they have received in the literature. In general, we can say that themes and motifs are diachronically and synchronically recurring semantic entities of a text that differ for the level of abstraction and generalization: the themes are macro semantic entities (the voyage, the treason, the personal identity, the double...), related to a whole text, or at least to large sections of it; the motifs are smaller stereotypical semantic units; they can be the sub-component of a theme, but they can also have autonomous manifestation and can become highly codified to the point of becoming topoi, like the *locus amoenus*, the foolishness of the hero and so on. They are identifiable as themes and motifs and not generic semantic contents, by virtue of their nature of intertextual cultural contents that reside in the collective cultural memory to which both authors and readers resort.

Both themes and especially motifs (and even more *topoi*, that can be also stylistically codified) can have linguistic manifestations by the way of specific phrases, sentences or textual components (and even the whole text), but for themselves, they are meta-textual semantic units that do not have an immediate linguistic correlate. It is worth noting, moreover, that even if a motif is lexicalized it is not necessary that its verbal units have a notable frequency, and that themes and motifs can have complex structural internal articulations. Consequently, the concepts of literary themes and motifs are rather far from the kind of homogeneous word list produced by LDA topic modeling.

The notion of isotopy is not less problematic. We owe to Greimas the first formulation of the concept (Greimas 1966): "a redundant set of semantic categories which make possible the uniform reading of the story". The concept has been revised more times by the French semiotician himself, and by several other scholars, like the *Group μ*, Orecchioni and Eco. Therefore, we are in face of an *umbrella term*, where we can find references to many diverse kinds of semiotic phenomena like the phonetic, prosodic, stylistic, rhetoric, morphosyntactic, semantic and narrative isotopies.

In its original formulation, an isotopy is a paradigmatic class of semantic traits (semes) that are homogeneous and that recur in a text by the way of lexemes, morphemes, phrases or other kinds

of linguistic and rhetorical devices. The isotopy is the condition for the coherence of a text. Eco has extended the idea of isotopy identifying a complex hierarchy of isotopic levels conceived as intensional semantic objects co-produced by the reader interpretative cooperation. One relevant aspect of the isotopy concept is that a literary text is usually pluri-isotopic and that isotopies occur at different levels of abstraction. This explicates why literary texts are devices that can produce multiple semantic interpretations.

Isotopy as a general phenomenon cannot be reduced to a lexical aspect and its definition implies the relevance of the linear nature of the textuality, and of the direction or order relation that connects its components. That said, it seems that under some restricted conditions of applications and adequate segmentation of the texts, the results of a topic modeling algorithm could approximate some of the possible levels of isotopies with lexical dominance that innervate a literary text.

Conscious of the difficulties in assigning an adequate semantic interpretation to the notion of "topic" produced by LDA and similar statistical algorithms when applied to the literary text, Ted Underwood has proposed to consider them as "discourses", that is "kinds of language that tend to occur in the same discursive contexts" (Underwood 2012). This solution is either a substantial weakening of the analytical relevance of the topic model methods, or a terminological trick that simply moves the problem, in that now we are required to give a theoretical explanation of the term discourse. Some have been tempted to apply in this context the notion of "formation discursive" (discursive formation) devised by Foucault. For instance (Roe, Gladstone and Morrissey 2016) affirm:

> It is not unreasonable, for instance, to posit that Foucault's concept of archeology, in fact, justifies the "bag of words" analytical model used by topic modeling and other machine-learning algorithms; a model that has often come under scrutiny (for good reason) by humanists.[…] From this perspective, topic modeling, and the "bag of words" model that underlies it, can be used to identify multiple discourses in text collections based on the probabilistic co-occurrence of words in the same discursive context. Computer scientists call these clusters of co-occurring words "topics," we prefer to think of them as "discourses."

The problem is that the French philosopher has used the concept of discourse in different and not always consistent ways and, more importantly, that he has never limited it to a purely linguistic notion (Foucault 1969: 52–53):

> Une telle analyse n'essaierait pas d'isoler, pour en décrire la structure interne, des îlots de cohérence; elle ne se donnerait pas pour tâche de soupçonner et de porter en pleine lumière les conflits latents; elle étudierait des formes de répartition (…) elle décrirait des systèmes de dispersion.
> Dans le cas où on pourrait décrire, entre un certain nombre d'énoncés, un pareil système de dispersion, dans le cas où entrer les objets, les types d'énonciation, les concepts, les choix thématiques, on pourrait définir une régularité (un ordre, des corrélations, des positions et des fonctionnements, des transformations), on dira, par convention, qu'on a affaire à une formation discursive, – évitant ainsi des mots trop lourds de conditions et de

> conséquences, inadéquats d'ailleurs pour désigner une pareille dispersion, comme «
> science » ou « idéologie », ou « théorie » ou « domaine d'objectivité ».

A discourse is somehow co-extensive with the notion of episteme or cultural semiotic system (and we must add that for Foucault even the concept of "statement" must not be conceived linguistically, but as an act that brings into existence the enunciative chain). On the other hand, Underwood's proposal to describe the "topics" resulting from topic modeling as "discourse" seems to limit the notion to its strict linguistic sense that is a synonym of "diction"; in this sense topic models are methods to individuate notably relevant portion of the single authors' idiolects and their collective sociolects.

4. Distant reading, literary theories and the role of interpretation

The discussion about the possible literary interpretations of topic modeling outputs and the consideration of the theoretical difficulties we have identified, lead us to affirm that it is not easy to identify a univocal and satisfying literary theoretic correlate of the machine learning concepts and techniques. But this acknowledgment does not imply that literary criticism must abandon the quantitative/computational methods or consider them as irrelevant for our understanding of literary texts. This is what ultimately suggests Nan Z. Da in *The Computational Case against Computational Literary Studies*, one of the deeper and methodologically sound criticisms of the field, published in the prestigious journal "Critical Inquiry". Da examines some of the most appreciated literary analysis based on distant reading and computational approaches[7] and, on the basis of the presumed pitfalls and statistical inconsistencies she identifies[8], she affirms:

> The nature of my critique is very simple: the papers I study divide into no-result papers—those that haven't statistically shown us anything—and papers that do produce results but that are wrong. I discuss what it is about the nature of the data and the statistical tools that leads to such outcomes.

The final consequence she derives is even far more general: "It may be the case that computational textual analysis has a threshold of optimal utility, and literature—in particular, reading literature well—is that cut-off point". Da's article raised a lively debate and many rebuttals, and especially this ultimate conclusion is surely not entailed by her analysis, as Fotis Jannidis has brilliantly pointed out (2020). But I don't think that this sort of position can be easily dismissed since as we have seen in the topic modeling discussion above, the problem of finding a good theoretical framework under which computational methods can give interesting results is a "difficult problem" indeed. Many important scholars active in the field believe that is possible to reconcile traditional

---

[7] But her selection is done only among the articles published in English and coming from the Anglo-American context, leaving aside all the European and non-Western scientific production (Jannidis 2020).

[8] Many of the scholars whose work has been criticized in the article have immediately responded pointing out on their part technical errors in Da's work, partially acknowledged by the author herself. *Critical Inquiry* blog section (https://critinq.wordpress.com/2019/03/31/computational-literary-studies-a-critical-inquiry-online-forum/) has hosted a number of those responses (as well as support statements, one of which by Stanely Fish), and others have been hosted by the *Journal of Cultural Analytics,* in particular a long article by Andrew Piper (Piper 2020) and one by Fotis Jannidis (Jannidis 2020).

theories (and problems) of literature with computational/quantitative methods. Just to make a couple of examples of this "consilience" theses, we can cite Andrew Piper and Michael Gavin.

In his methodological works, Piper proposes a sort of computational hermeneutics, that integrates distant and close reading, quantitative and qualitative analysis (Piper 2015):

> My aim in this essay is to offer a methodological polemic against the either/or camps of close versus distant reading or shallow versus deep that have metastasized within our critical discourse today. I want us to see how impossible it is not to move between these poles when trying to construct literary arguments that operate at a certain level of scale (although when this shift occurs remains unclear). In particular, I want us to see the necessary integration of qualitative and quantitative reasoning, which, as I will try to show, has a fundamentally circular and therefore hermeneutic nature. As we move out from a small sample of texts toward larger, more representative populations and back into small, but now crucially different samples, such circularity serves as the condition of new knowledge, of insight per se. It puts into practice a form of conversional reading, one whose telos is not a single, radical insight, but instead an iterative and circular process that can serve as a vehicle for conceptual change.

In a similar vein, Michael Gavin proposes to understand the affordances of vector representation of texts and distributional semantic hypothesis adopting William Empson's theory of ambiguity (Gavin 2018):

> In this regard, I'll argue, vector semantics share a set of assumptions with literary critic William Empson, who devoted his career to explaining how poets played with words' many meanings. Words were, in his view, "compacted doctrines" that always carried their various senses as latent semantic potential. Empson's method of close reading broke words into their putatively constituent units of connotation, and vector semantics pushes this conceit to an extreme he would have found as baffling and exhilarating as his first reviewers found him.

The problem with these approaches is that their outcomes are either unsatisfactory from a literary-critical point of view (or to put it simply, they yield to the "so-what" reaction suggested by Da) or are self-contradictory, in that the hermeneutical and critical parts of the discourse are self-standing, the critical arguments are logically independent from the results of the computational analysis. See for instance the following analysis proposed by Gavin of one Milton's passage (Gavin 2018: 672) :

It's possible to scale out further, taking a composite vector of the entire passage (fig. 14). When taken as a whole, the lines describe a kind of evil thinking, suddenly confronted by an unexpected pleasure of sensation. The words that orbit thinking form rings of association that descend toward evil on one side and connect loosely, impotently toward solace on the other. We might say that the space Satan briefly occupies is outside history, or that it provides history's conditions of possibility and is therefore conceivable only by abstracting history away. We might say something similar about vector space, where variety permeates across many dimensions that are irreconcilable to the human experience of actuality in time.

Most of these considerations (irrespectively of their correctness) can easily derive from a traditional close reading of the poem[9].

Based on these considerations, I think that distant reading should not be considered a methodological innovation to be applied to pre-existing theories of literary texts (in all their rhizomatic variants): it is necessary to find a suitable theory or framework where these methods can yield to interesting and innovative results. To make a substantial step in this direction we should first take seriously the original notion of distant reading as devised by Moretti and overcome the idea of literature as made of singular special individual entities (the great or the important texts), subjects of the interpretative work of literary critics and historians, or reduced to abstract ideal types, under the scrutiny of literary theoreticians with no clue with empirical evidence. This means that we should take seriously the move from *interpretation* to *explanation* as the real aim of the scholarly inquiry, a move that Moretti had already identified as central in his early methodological reflections on distant reading (Moretti 2005: 106):

---

[9] Similar observations could be advanced for some of Moretti's critical essays, where the methods and the critical inquiry just looks as if they are juxtaposed: the article on the network analysis of Hamlet is epitomical in this respect, in my view (Moretti, 2011).

> i modelli che ho discusso (e altri, analoghi, che si potrebbero aggiungere) condividono tutti una chiara preferenza per la spiegazione rispetto all'interpretazione. Questa, naturalmente, è una distinzione di grande portata, e che merita un ragionamento a sé; ma per ora, e per capirsi, diciamo almeno che qui non ho cercato di stabilire il significato specifico di questa o quell'opera individuale, bensì di ricostruire le strutture astratte, di validità generale

Needless to say, this discourse would make sense only if we have a thorough argument on what is meant by "interpretation" and "explanation", and in what they diverge. This analysis is out of the scope of the current article, and future work will be dedicated to that[10]. For the moment it should suffice to adopt a sort of common-sense explication of the two concepts: interpretation are the results of thought processes applied to individual entities (texts, persons, facts); are based on a set of subjective assumptions and presuppositions, although they are public discourses that can be intersubjectively shared by communities; are value driven and hard to distinguish from esthetic judgment; are not necessarily mutually incompatible. Explanations are arguments that connects observations (data) independent from the individual observer (this is the most controversial part of my argument) and theories, by the way of formal hypothesis or models; under certain condition they are not dependent on subjective judgments; they are usually applicable to all the objects of the same kind in a domain; they are mutually incompatible, given a layer of description of the domain.

On what theoretical basis, then, can we build a theoretical account of literature amenable to distant reading methods and oriented toward an explanation-based methodology? One possible direction to be explored, as some scholars like Underwood suggest (English and Underwood 2016) is that distant reading should fall inside the wider field of social sciences and the *Longue durée* history, according to the Annales School tradition (Underwood 2017a):

> In particular, I want to emphasize that distant reading is not a new trend, defined by digital technology or by contemporary obsession with the word data. The questions posed by distant readers were originally framed by scholars (like Raymond Williams and Janice Radway) who worked on the boundary between literary history and social science. Of course, computer science has also been a crucial influence. But the central practice that distinguishes distant reading from other forms of literary criticism is not at bottom a technology. It is, I will argue, the practice of framing historical inquiry as an experiment, using hypotheses and samples (of texts or other social evidence) that are defined before the writer settles on a conclusion.

Although there are many reasons to lean toward a sociological theory of literature as an optimal theoretical framework for distant reading, I think that cognitive and bio-evolutionistic approaches to literature and cultural evolution studies are equally if not more relevant frameworks.

---

[10] A very good introduction to the epistemological debate on this issue (that anyway does not solve all the problems and requirements of our analysis) are the two articles devoted to this theme in the two authoritative online philosophical encyclopedias, the *Stanford Encyclopedia of Philosophy* (Woodward and Ross, 2021), and the *Internet Encyclopedia of Philosophy* (Mayes, 2021).

5.   Cognitive and evolutionary approaches to literature and distant reading

Cognitive poetics/narratology, and bio/evolutionary literary studies have been two of the most interesting waves of innovation in the literary field of the last 30 years and are now established fields of inquiry. With different graduation, depending on the authors, they have advocated the introduction of a scientific methodology in the study of literature, looking for methodological and theoretical insights in cognitive science and evolutionary psychology. We cannot delve into the details of the various area of research that have been characterizing this field and of the different approaches proposed, and see how the specific kinds of problems studied in cognitive poetics and narratology could be analyzed by the way of distant reading methods.

What is more interesting from the point of view of my thesis is that, not surprisingly, the debate around the legitimacy and acceptability of the cognitive approaches in literary studies has determined a discussion on the problem of literary interpretation that has many similarities with the arguments I have proposed in this article. The problem was posited clearly already since early 2000, by Tony E. Jackson, who observed the radical irreducibility of the literary interpretation to the explanation and justification of ordinary sciences (Jackson 2003):

> To conclude, we must understand the most common use of the term literary interpretation dialectically, which is to say in two ways at once. Literary refers to the kind of text being interpreted. But it also refers to the kind of interpretation being performed. An interpretation is literary if it conjoins with the literary text so as to bring out in a determinate way the text's indeterminacy by revealing the critic's own account of literary meaning. An outcome of all this is that literary interpretation falls somewhere in between (''inter'') straightforward logical explanation and literature itself.

The outcome of Jackson's paper was a radical skepticism in the productivity of cognitive literary studies. In the following years, anyway, this field has grown in impact and recognition, somehow ignoring or understating Jackson's criticism. And recently Marco Caracciolo, has re-opened the debate, turning Jackson's criticism into a positive stimulus for a radical paradigmatic change: "In order to contribute to cognitive science, literary scholarship has to complement—and in some cases even supplant—interpretation with a different set of goals and methods". (Caracciolo 2016: 193). Later in the same essay, Caracciolo argues that the real critical point for cognitive literary studies has been the incapacity to detach from the venerable practice of close reading and the necessity to deliver interpretative accounts of individual texts, which *per se* cannot provide a better self-justification than any other kind of literary interpretation:

> Cognitive approaches to literature have followed in this tradition, largely out of deference to the accepted conventions of literary scholarship. Yet cognitive-scientific models sit less comfortably with interpretive practices than the poststructuralist agendas that preceded them: on the one hand, interpretation evacuates the scientificity of cognitive scientific models, insofar as any reading using cognitive science as ground (RICS) stands on an equal footing with other readings (AORs). On the other hand, some of the most promising lines of research in cognitive literary studies—what I have called the "processual" and the "functional" approaches—appear to give up interpretation as the main focus of research".

The other scientific field where literary studies can find a theoretical framework that takes great advantage of the distant reading methodology is that of cultural evolution. This field of study that as of now has no application in literary studies, aims at providing a naturalist and empirical explanation of the nature and evolution of culture, adopting widely mathematical/statistical and computational modeling. One of the theoretical underpinnings of cultural evolution is the adoption of the *population thinking* framework, taken from evolutionary biology (after Ernest Mayr interpretation of Darwin's theory) and population genetics, and its application to cultural phenomena as pointed out recently by Dan Sperber and his collaborators (Claidière, Scott-Phillips and Sperber 2014):

> Population thinking involves looking at a system (such as culture) as a population of relatively autonomous items of different types with the frequency of types changing over time. The types themselves are not defined by their 'essence' but as historical subpopulations, features of which may change over time.

Literature is part of the cultural sphere, so it can be considered a population of individual items (the texts) whose members are defined by sets of measurable features. The description of the population at a given state (synchronic) and its evolution (diachronic) is feasible by the way of statistical and data-driven analysis: exactly, distant reading. Another interesting theoretical construct of cultural evolution theories is that of *cultural attractors*, again a thinking tool, to cite Dennett, devised by Dan Sperber. This concept is useful to explain why, notwithstanding in the vast majority of cases cultural traits are neither properly copied nor selected, but reconstructed each time an instance of transmission happens, some cultural traits show a strong permanence (Sperber, 1996):

> Causal chains of mental representations and public productions can be described as moving, with each transformation, over a space of possibilities. In this space, there are attractors such that, in their vicinity, transformations tend to be of limited amplitude and to cancel one another out, mimicking replication. Attractors themselves have quite diverse etiologies. Some of these attractors are constant across cultures and times; others are culture specific and precarious. The main force driving cultural evolution is the selective stabilization brought about by these attractors.

For example, in the oral transmission of a fairy tale, e.g. *Little Red Hood*, it is highly unlikely that the story will be repeated verbatim at each passage. Still, some defining features, say the walk in the wood, or the wolf, perhaps because they are particularly memorable, will act as attractors, and will be repeated (reconstructed) each time by different narrators. Cultural transmission here has relatively low fidelity, and non-random distortions and reconstructions play an important role in maintaining cultural diversity and stability.

My suggestion is that this notion strictly relates to phenomena like genres, themes and plot intertextuality and evolution. In addition, in a sense, also the role of the noticeable individuals is

retained in the framework: it is the individual that introduces a substantial mutation that becomes a new attractor[11].

6. Conclusions

To conclude, I think that computational literary and cultural studies must find proper theoretical frameworks to take full advantage of the most advanced methods and analytical techniques, like text mining and machine learning. The analytical results produced in the context of traditional literary theories, concepts and methodologies, in fact, are easily amenable to the 'so what' criticism, however sophisticated they can be. More fundamentally, the traditional field of literary studies relies on the hermeneutic paradigm, while distant reding leans toward a methodology based on hypothesis formulations and testing, observations and data-oriented inquiry and explanations. It is time to change the framework and depart from this hermeneutical tradition.

In the space of the theories, I think that the bio-cognitive theory of literature and cultural evolution can offer the best option to leverage distant reading methods, and to understand a whole set of phenomena related to literature and culture that are beyond the reach of traditional approaches. Of course, at this level of analysis, it is not possible to derive an interpretation of single works, but explanations of large-scale distributions and variations of the population. These will continue to be the object of hermetical and close reding oriented literary studies, that as Underwood has rightly observed will continue to be a legitimate way to study literary texts (Underwood 2019: XVII):

> Distant reading is simply a new scale of description. It doesn't conflict with close reading any more than an anatomical diagram of your hand would conflict with the chemical reactions going on inside your cells. Instead of displacing previous scales of literary description, distant reading has the potential to expand the discipline—rather as biochemistry expanded chemistry toward a larger scale of analysis.

Yes, this will require that we really forget how to read, but maybe we will come up with a deeper understanding of one of the most complex and multi-layered outcomes of human evolution: literature.

Bibliography

Acerbi, A. et al, "The Expression of Emotions in 20th Century Books", *PLOS ONE*, 8(3), 2013. doi: 10.1371/journal.pone.0059030.

Allington, D., Brouillette, S. and Golumbia, D. , "Neoliberal Tools (and Archives): A Political History of Digital Humanities", *Los Angeles Review of Books (LARB)*, 1.5.2016. https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/.

Blei, D. M., "Probabilistic Topic Models", *Communications of the ACM*, 55(4), pp. 77–84, 2012. doi: 10.1145/2133806.2133826.

---

[11] Interesting examples of the productivity of the "interbreeding" between cultural evolution and computational literary studies are three recent papers authored by some of the most interesting researchers in the former field (Acerbi *et al.*, 2013; Morin and Acerbi, 2017; Morin, Acerbi and Sobchuk, 2019).

Blei, D. M., Ng, A. Y. and Jordan, M. I., "Latent Dirichlet Allocation", *J. Mach. Learn. Res.*, 3, pp. 993–1022, 2003.

Bridgman, P. W., *The Logic of Modern Physics*, New York, Macmillan, 1927.

Burrows, J., *Computation into Criticism: A Study of Jane Austens Novels and an Experiment in Method*, Oxford, Clarendon Press, 1987.

Caracciolo, M., "Cognitive Literary Studies and the Status of Interpretation: An Attempt at Conceptual Mapping", *New Literary History*, 47(1), pp. 187–207, 2016. doi: 10.1353/nlh.2016.0003.

Chang, H., "Operationalism", in Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. *Winter 2019*. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/win2019/entries/operationalism/.

Ciotti, F., "What's in a Topic Model? Critica teorica di un metodo computazionale per l'analisi del testo", *Testo & Senso*, 18, 2017.

Claidière, N., Scott-Phillips, T. C. and Sperber, D., "How Darwinian is cultural evolution?", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642), p. 20130368, 2014. doi: 10.1098/rstb.2013.0368.

Cometa, M., *Letteratura e darwinismo: introduzione alla biopoetica*. Roma, Carocci, 2018.

Eco, U., *I limiti dell"interpretazione*. Milano, Bompiani, 1990.

English, J. F. and Underwood, T., "Shifting Scales: Between Literature and Social Science", *Modern Language Quarterly*, 77(3), pp. 277–295, 2016. doi: 10.1215/00267929-3570612.

Fish, S., *Is There a Text in This Class? The Authority of Interpretive Communities*, Cambridge, Harvard Univ. Press, 1980.

Fish, S. E., "What is Stylistics and Why Are They Saying Such Terrible Things about It? - Part II", *Boundary 2*, 8(1), 1979, pp. 129–146. doi: 10.2307/303144.

Foucault, M., *L'archeologie Du Savoir*, Paris, Gallimard, 1969.

Gavin, M., "Vector Semantics, William Empson, and the Study of Ambiguity", *Critical Inquiry*, 44(4), 2018, pp. 641–673. doi: 10.1086/698174.

Golumbia, D., "Death of a Discipline", *differences*, 25(1), 2014, pp. 156–176. doi: 10.1215/10407391-2420033.

Greimas, A. J., *Sémantique structurale. Recherche de méthode*, Paris, Larousse, 1966.

Hoover, D. L., "Textual Analysis", in Price, K. M. and Siemens, R. (eds) *Literary Studies in the Digital Age*. Modern Language Association of America, 2013. doi: 10.1632/lsda.2013.3.

Jackson, T. E., "'Literary Interpretation' and Cognitive Literary Studies", *Poetics Today*, 24(2), 2003, pp. 191–205. doi: 10.1215/03335372-24-2-191.

Jakobson, R., "Linguistics and poetics", in *Style in language*. MA: MIT Press, 1960, pp. 350–377.

Jannidis, F., "On the perceived complexity of literature. A response to Nan Z. Da", *Journal of Cultural Analytics*, 2020. doi: 10.22148/001c.11829.

Jockers, M. L., *Macroanalysis: Digital Methods and Literary History*. Urbana, University of Illinois Press, 2013.

Lotman, Ju. M., *Struktura xudozhestvennogo teksta*, Mosh'a, Iskusstvo, 1970, Eng. tr. *The Structure of the Artistic Text*, Providence, Brown University Press, 1977.

Marche, S., "Literature is not Data: Against Digital Humanities", *Los Angeles Review of Books*, 28 October, 2012. http://lareviewofbooks.org/essay/literature-is-not-data-against-digital-humanities#.

Mayes, R., "Theories of Explanation", *Internet Encyclopedia of Philosophy*, 2021. https://iep.utm.edu/explanat/ (Accessed: 25 June 2021).

Moretti, F., "Conjectures on World Literature", *The New Left Review*, 1, 2000, pp. 54-68. https://newleftreview.org/issues/ii1/articles/franco-moretti-conjectures-on-world-literature.

– *Graphs, maps, tree : abstract models for a literary history*, London-New York, Verso, 2005.

– *La letteratura vista da lontano*, Torino, Einaudi, 2005.

– "Network Theory, Plot Analysis", *The New Left Review*, 68, 2011, pp. 80–102.

– *Distant Reading*, London, Verso, 2013.

– "'Operationalizing'. Or, the Function of Measurement in Literary Theory", *New Left Review*, 84, 2013.

Morin, O. and Acerbi, A., "Birth of the cool: a two-centuries decline in emotional expression in Anglophone fiction", *Cognition and emotion*, 31(8), 2017, pp. 1663–1675.

Morin, O., Acerbi, A. and Sobchuk, O., "Why people die in novels: testing the ordeal simulation hypothesis", *Palgrave Communications*, 5, 2019, p. 62.

Piper, A., "Novel Devotions: Conversional Reading, Computational Modeling, and the Modern Novel", *New Literary History*, 46(1), 2015, pp. 63–98. doi: 10.1353/nlh.2015.0008.

– "There Will Be Numbers", *Journal of Cultural Analytics*, 2016. doi: 10.22148/16.006.

– "Do we know what we are doing?", *Journal of Cultural Analytics*, 2020. doi: 10.22148/001c.11826.

Potter, R. G., "Literary Criticism and Literary Computing: The Difficulties of a Synthesis", *Computers and the Humanities*, (22), 1988, pp. 91–97.

Ricoeur, P., *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*, Cambridge, Cambridge University Press, 1981. doi: 10.1017/CBO9781316534984.

Roe, G., Gladstone, C. and Morrissey, R., "Discourses and Disciplines in the Enlightenment: Topic Modeling the French Encyclopédie", *Frontiers in Digital Humanities*, 2, 2016. doi: 10.3389/fdigh.2015.00008.

Segre, C, *Avviamento all"analisi del testo letterario*, Torino, Einaudi, 1985.

Sperber, D., *Explaining Culture: A Naturalistic Approach*, Oxford, Basil Blackwell, 1996.

Underwood, T., "What kinds of 'topics' does topic modeling actually produce?", *The Stone and the Shell*, 2012. http://tedunderwood.com/2012/04/01/what-kinds-of-topics-does-topic-modeling-actually-produce/.

– "A Genealogy of Distant Reading", *Digital Humanities Quarterly*, 11(2), 2017. http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html.

– "We're probably due for another discussion of Stanley Fish", *The Stone and the Shell*, 13 July, 2017. https://tedunderwood.com/2017/07/13/were-probably-due-for-another-discussion-of-stanley-fish.

– *Distant horizons: digital evidence and literary change*, Chicago, The University of Chicago Press, 2019.

Woodward, J. and Ross, L., "Scientific Explanation", in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy. Summer 2021*. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/.