Franco Moretti
Stanford University

Under which King, Bezonian? Literary Studies between Hermeneutics and Quantification

What relationship is there between the quantitative literary history that has emerged in the past twenty years and the older hermeneutic-interpretive tradition? Answers have typically been of two kinds: for many in the interpretive camp, the two approaches are incompatible, and the newer one has little or no critical value; for most quantitative researchers, they are both valid, perfectly compatible, and in fact complementary. Writing these pages has convinced me of a third possibility – both valid, yet not complementary – that will emerge step by step from a comparison of how the two strategies work. How they work literally, in the conviction that, as Oleg Sobchuk and I have recently written, “practices – what we learn to do by doing, by professional habit, and often without being fully aware of what we are doing – have frequently larger theoretical implications than theoretical statements themselves.” (Moretti and Sobchuk 118). In that article, “practice” indicated the kind of visualization that has come to characterize the digital humanities; here, it refers to the chain of interconnected decisions that shape the research strategies of interpretation and quantification.

1 Moretti and Sobchuk (118) and the current essay are part of a series of reflections on the quantitative study of culture that includes Moretti (2013, 2016, 2017) – now all collected in Moretti et al. 2017.
But the aim is the same: understanding what a research paradigm actually does, rather than what it declares it wants to do.

There is a complication, however: since both the quantitative and – even more so – the hermeneutic approach are actually many approaches, often sharply at odds with each other (a Lacanian interpretation having nothing in common with a new historicist or an ecocritical one, and so on), I will restrict myself to work I have personally taken part in. This is of course a questionable decision – the exact opposite, in fact, of what we did in Hidden in plain sight, which examined over sixty articles by a multitude of authors – which I’m taking because, as one of the few who have used both approaches for over twenty years, I’ve been repeatedly taken aback by how different my work has ended up being in the two registers. To some extent it had to be different of course (that’s the point of using different methods to begin with), but there was something more here which felt slightly uncanny: I could no longer perceive an intelligible relationship – let alone a similarity – between the two sides of my work. In the case of a single person, this seemed strange, and worth thinking about. Maybe it’s just a case of personal inconsistency; maybe, the sign of something larger, with an objective significance.

1. Hermeneutics first. Nick Adams, the protagonist of Hemingway’s short story Big Two-Hearted River (190), is about to go fishing:

Nick took it from his hook book, sitting with the rod across his lap. He tested the knot and the spring of the rod by pulling the line taut. It was a good feeling. He was careful not to let the hook bite into his finger.

He started down to the stream, holding his rod; the bottle of grasshoppers hung from his neck by a thong tied in half hitches around the neck of the bottle. His landing net hung by a hook from his belt. Over his shoulder was a long flour sack tied at each corner into an ear. The cord went over his shoulder. The sack flapped against his legs.

Nick felt awkward and professionally happy with all his equipment hanging from him. The grasshopper bottle swung against his chest. In his shirt the breast pockets bulged against him with his lunch and his fly book. (190)

2 The analysis that follows but condenses the more detailed account I have given in Moretti (2019: 49-65).
First of all – does this passage even need an interpretation? Not really, if interpreting means dispelling the “obscurity” of a text: here, everything seems perfectly clear. Is it though? The idea “that understanding occurs as a matter of course”, wrote the founder of modern hermeneutics, is typical of the “less rigorous practice” of interpretation; for its “more rigorous” version it is however “misunderstanding [that] occurs as a matter of course, and so understanding must be willed and sought at every point” (Schleiermacher 109-110).

Willed at every point … Let’s start with this, then: that Hemingway’s few sentences include twenty-five prepositional phrases (those introduced by a preposition: “from his hook book”, “across his lap” and so on). Twenty-five, in 149 words: a lot. But they are there, because they’re doing something which is very important for the world of the story – gluing together all sorts of disparate elements: “the bottle of grasshoppers hung from his neck by a thong tied in half hitches around the neck of the bottle.” Like a Swiss army knife: an incredibly compressed and well-organized world, where every object has been dealt with – “professionally”. It’s a world of things:

With the ax he slit off a bright slab of pine from one of the stumps and split it into pegs for the tent. He wanted them long and solid to hold the ground […] He pegged the sides out taut and drove the pegs deep hitting them down into the ground with the flat of the ax until the rope loops were buried and the canvas was drum tight. (Hemingway 183)

A world of things, but not only: Nick wants pegs “for the tent”; they are “long and solid to hold the ground”, and he hits them “until the rope loops were buried”. It’s all so purposeful: done, in order to do something else – and having already in mind what that next thing will be. “Know how”, Gilbert Ryle (81)

3 See Kimmerle (109-110). “Obscurity”, Peter Szondi (27) rightly observed, is for Schleiermacher “hardly the only occasion for interpretation”.
4 Here they are in italics: “Nick took it from his hook book, sitting with the rod / across his lap. He tested the knot and the spring of the rod by pulling the line taut. It was a good feeling. He was careful not to let the hook bite into his finger. He started down to the stream, holding his rod, the bottle of grasshoppers hung from his neck / by a thong tied in half hitches / around the neck / of the bottle. His landing net hung by a hook / from his belt. / Over his shoulder was a long flour sack tied at each corner / into an ear. The cord went over his shoulder. The sack flapped against his legs. Nick felt awkward and professionally happy with all his equipment hanging from him. The grasshopper bottle swung against his chest. In his shirt the breast pockets bulged against him / with his lunch and his fly book.”
called these chains of silent interlinked movements. “Nick tied the rope [...] and pulled the tent up [...] and tied it to the other pine.” Always perfectly calm. But: this is a story. Isn’t calm the opposite of what a story ought to be?

It is. But Hemingway was writing against a very special historical background. “A generation that had gone to school on horse-drawn streetcars”, Walter Benjamin wrote about World War I, “now stood under the open sky in a landscape where nothing remained unchanged except for the clouds, and, beneath these clouds, in a force field of destructive torrents and explosions, was the tiny, fragile human body.” (Benjamin 143-4 in Eiland after Jennings 2002). We usually turn to stories because our life is not eventful enough; but what if the key experience of an entire generation has been the great war? Much too eventful: and so, the desire for a different kind of narrative arises, in which calm has a role to play. War literature, observed Eric Leed, is about “men who, as a rule, had little or no control over the events which threatened their lives.” (Leed 33). No control: this is the key. Hemingway’s style is all about control: of space, time, gestures, words. “Nick felt [...] happy with all his equipment hanging from him”: this is the snapshot of a young soldier – minus the war. And the same goes for his march through the woods, his reconnaissance, tent, camp – he even eats canned food on his trip. While not quite “playing” at war, Nick is replaying it; rewriting it. Life in the trenches had alternated between tedium or terror; nothing for days, then apocalypse. Hemingway’s prose is never boring, and never frightening; clean and cautious, it’s the perfect style for convalescence (three years later, the central episode – and happiest – of his first great success, A Farewell to Arms). This is war literature in the sense that it wants to recover from it: to resolve a dissonance of historical experience, to adapt Lukacs’ metaphor in Theory of the Novel. But on this, more later.

2.

From a few sentences of a single short story, to the most obvious novelty of the quantitative approach: the expansion of literary history well beyond a small canon of great works. Time was, a theorist would choose one text – Don Quijote, Robinson Crusoe, Crime and Punishment – and erect upon it a whole theory of the novel. “Type thinking”, Ernst Mayr has called it: “Tristram Shandy is the most typical novel in all of world literature”, as Shklovsky wrote in Theory of Prose (Schklovsky 170). In front of the swarm of nineteenth-century British novels of Figure 1, though, type thinking is useless: here, one must account for an entire “population” of novels. Not a very large one, in this case – 1,117, to be exact – but still irreducible to a single text.
The text: this is where the discord between old and new is at its sharpest. It used to be the object of literary study; here, it’s a dot. It has been made to be a dot. It’s a choice; fundamentally, the same that was made about historical “events” during the first great quantitative turn of seventy years ago. Before the advent of the Annales, events used to be as central to historiography as texts to literary study, and for the same reason: because of their uniqueness. “Historians resembled collectors”, wrote Krzysztof Pomian in a retrospective essay on that conceptual shift, “gather[ing] only rare and curious objects, and ignoring whatever looked banal, everyday, normal…” (Pomian 536, 543-4). Once events started to be studied “as elements of a series”, however, uniqueness lost all its prestige: “if one were to combine Braudel’s and Labrousse’s positions”, Pomian continued, “one would relegate events to the periphery of history, or perhaps disregard them altogether”. A history that disregards events: this is the core of the quantitative turn. A history of art “without names”, Wölfflin (66) once wrote. A history of literature without texts.

Without texts – in the sense that there are too many of them: the 250 canonical and 850 forgotten novels of Figure 1 (the orange and blue dots, respectively). We wanted to test whether linguistic richness contributed to a novel’s survival or oblivion, so all texts were sliced into 1,000-words segments, and their
type-token ratio was calculated (Algee-Hewitt et al. 2015)\(^5\). And it turned out that the segment with the highest score belonged to Edward Hawker’s *Arthur Montague, or an Only Son at Sea* (1850), and that with the lowest one to George Eliot’s *Adam Bede* (1859)\(^6\).

That a novel none of us had heard of should have so much more lexical variety than a canonical one like Eliot’s was the opposite of what we expected, so we of course read those segments, and many others that clustered around the two extremes of the spectrum. Hawker’s was a description, as were several high-scoring segments: which makes sense, because descriptions require details, and details increase lexical variety. Eliot’s was completely different: a young woman, confessing to having abandoned her child to die, and repeating the same words over and over again (thus uttering the most redundant passage of the entire century), as if she were chained – “but I couldn’t go away” – to that scene of death. Generalizing: high type-token ratio characterized the prose of the novelistic narrator: written, analytical, impersonal, and almost atemporal; low type-token ratio, the voice – the whisper – of a novelistic character: broken and desperate and haunted by a single terrible event.

\(^5\) See “Canon/Archive. Large-Scale Dynamics in the Literary Field”, now in *Canon/Archive*, cit. (Algee-Hewitt et al. 2015). Here, too, what follows is only a very abbreviated summary of the original study.

Type-token ratio is a standard measure of lexical variety that expresses the relationship between the number of different words used (types), and the number of actual words used (tokens). “Good morning, my good friend” has five types and five tokens, hence a type-token ratio of 5/5, or 1; “Good morning, Jim, good morning” has three types and five tokens, hence a type-token ratio of 3/5, or 0.6.

\(^6\) Here are two sections of those extreme segments where the hash sign indicates a word that had already appeared in the segment, and the asterisk a word that did not belong to the initial “dictionary” of novelistic English (about 230,000 words) created by Ryan Heuser who wrote the program for this part of the experiment.

*Arthur Montague:* “then cut through some acres of refreshing greensward, studded with the oak, walnut, and hawthorn, ascended a knoll, skirted an expansive sheet of water; afterwards entering an avenue of noble elms, always tenanted* by a countless host of cawing* rooks, whose clamorous conclaves* interrupted the stillness that reigned around, and whose visits to adjacent corn-fields* raised the ire and outcry of the yelling urchins employed to guard them from depredation.”

*Adam Bede:* “And I made haste out of the woods, but I could hear it crying all the while; and when I got out into the fields, it was as if I was held fast... I could n’t go away, for all I wanted so was to go. And I sat against the haystack to watch if anybody ’ud come: I was very hungry, and I d only a bit of bread left; but I could n’t go away.”
This is what reading texts as elements of a series is like: we didn’t go from the segment of *Adam Bede* to the whole novel and its role in mid-Victorian culture (as in Hemingway’s case); we linked that segment to *other segments* (from *Ennui*, *Tales of a Briefless Barrister*, *Marius the Epicurean*, *Lady Laura*…) in order to extract from them a set of abstract conceptual pairs – narrator/character, written/spoken, analytical/emotional and so on – that would define the space of narrative possibilities. Ideally, the *whole* space. Not easy, as we discovered – we tried to make sense of the segments lying at the center of the distribution, and failed – but the direction was clear: we were no longer studying texts, but *series* of texts. Different.

3.

Twelve sentences; 1,100 novels. A process of interpretation leading from a text to an aspect of contemporary history; a set of measurements leading from one extreme of a corpus to the opposite extreme. Different objects, and different research practices. And we are back to that initial question: what is the relationship between them?

First of all, they are both perfectly valid forms of knowledge: on this, the critiques coming from the interpretive camp are entirely groundless. Both valid – and with a moment of overlap, too. In the midst of measurement, there had been interpretation: we had taken Hawker’s “acres of refreshing greensward, studded with oak, walnut, and hawthorn” and had turned it into – which is to say: interpreted it as – “impersonal analytical prose of the novelistic narrator”.

Conversely, the interpretation of *Big Two-hearted River* had been triggered by a (very elementary) form of measurement: it was because prepositional phrases were so insanely frequent that I felt I had to make sense of them – to “will my understanding”, as Schleiermacher had put it. To some extent, each method had relied on the other one; there was a potential for complementarity, here, that seemed to support the “oscillatory” ideal advocated by some quantitative researchers: a type of work “moving back and forth between close and distant forms of reading in order to approach an imaginary conceptual center” (Piper 67–8).

7 We failed because extreme cases possess an epistemological clarity that average ones lack. For a similar interplay of the extreme and the average where the latter was analyzed a little more successfully (Algee-Hewitt, Heuser and Moretti 2016).

8 Similarly, Hoyt Long and Richard Jean So have urged “a method of reading that oscillates or pivots between human and machine interpretation, each providing feedback to the other in the critic’s effort to extract meaning from texts.” (267).
Moving back and forth … Was that what had happened? True, I had counted up to twenty-five while working on Hemingway; but the really instructive detail was that this kindergarten achievement had felt like all the measurement I needed. Same for *Adam Bede*: that confession was an extraordinary passage for interpretation to work on; we did almost nothing with it, and yet, again, it felt like all the interpretation we needed. And we weren’t wrong: it felt that way, because both studies entailed a very clear *hierarchy* between the two methods: measurement was the means and interpretation the end in the Hemingway study, and vice versa with the 1,100 novels. In one case, the aim was understanding how a story about trout-fishing could be so significant even for readers who didn’t care at all about its subject; Hemingway’s way of writing had to mean *something more* than just fishing, and when I saw all those prepositional phrases I focused on them thinking they might help explain that hidden “something”. But I was trying to understand the phrases: that there were twenty-five of them, or eighteen, or thirty, made no difference at all. Conversely with the interpretation of *Adam Bede*: we wanted to investigate variations in the lexical composition of the novelistic field, and those accursed hash signs were the perfect embodiment of our priorities; they showed repetitions right away, and that’s what mattered; if they also made the words illegible, so be it – reading was not the point here. (So much so, that the next step in the published essay was a series of correlations of type-token ratio with abstract grammatical categories like verb forms and adjectives that got rid of text, reading and interpretation at once).

This is why it’s important to see how a method concretely works: because it’s in the sequence of “tactical” decisions – not counting prepositions any more, inserting hash signs, turning towards a historical event, exploring another area of the corpus … – that “strategic” research priorities take form. Priorities, and exclusions: in the sense that, in both studies, once the critical pendulum had started swinging in one direction, it never came back. There had been a moment of overlap between the two methods – and then it passed: with each new step, interpretation became more interpretive, and quantification more quantitative. There was no moving back and forth here.

---

Better: it made a difference that there were enough of them to become visible, as it’s always been the case in stylistic criticism; exactly how many counted as “enough”, remained vague. See, by contrast, the accuracy with which Sarah Allison and Marissa Gemma (283) established the link between the register of conversation in the *Longman Grammar of Spoken and Written English* (mean type-token ratio of 30%), and the 500 lower-ranked segments, where type-token ratio oscillated between 27 and 33%. In their reflection, the pathos of research was inextricable from a precision that would be unimaginable in the hermeneutic tradition.
But perhaps there could be? Those studies had a one-sided aim from the start: that’s why they so drastically subordinated one method to the other. Couldn’t one design a study in which they would both have the same weight?

I know of no such study, neither in literary history, nor in history *tout court*: where in fact the quantitative dominant of the *Annales* and the interpretive one of microhistory have remained steadfastly unreconciled for half a century. In principle, though, there is no reason why a balanced study shouldn’t be possible. We can certainly interpret Eliot’s sentences as thoroughly as Hemingway’s, just as we can easily measure the frequency of prepositional phrases in all American short stories of the 1920s. Nothing prevents the two methods from working next to each other. Can they also work *for each other*? This is the real question. Can measurement lead to a kind of interpretation that would otherwise be unimaginable – and the other way around?

4.

Certain features, writes Georges Canguilhem in his study of nineteenth-century medical epistemology, *The Normal and the Pathological* (1943-66), are termed normal insofar as they designate average characteristics, which are most frequently practically observable. But they are also termed normal because they enter ideally into that normative activity called therapeutics […] the normal state designates both the habitual state of the organs, and their ideal (Canguilhem 122-3, 126).

The normal as frequent-habitual-average, and the normal as ideal-normative: a single signifier, that opens onto two distinct semantic pathways. What relationship, between quantification and hermeneutics? This: they rotate on the same hinge, but in different directions: one, towards the frequent-average aspects of literary phenomena, and the other towards the normative ones. Normative, not in the ethical sense – that is seldom the point, especially in modern literature – but in that Panofsky had in mind when he spoke of art as “an objectifying conflict, aiming at definitive results, between a forming power and a material to be overcome” in Panofsky (339). *Bewältigen*: dealing with, mastering, remolding historical materials by applying the power – or force: *Kraft* – of aesthetic form. Here is where interpretation comes in; it takes those “definitive results”, and tries as hard as it can to undo the work of form: proceeding backwards, from the text, through its technique, to the world and the “dissonance” that was being addressed. In this sense, interpretation is an understanding of literature that is always tempted to go beyond literature: like the essayist described in *Soul and Forms*, who pretends to be only discussing books but is actually “always talking about the ultimate questions of life”.

4.
Not that interpretation goes always that far. But it can: whereas the quantitative approach cannot. With the single text that is the typical object of interpretive criticism, reverse-engineering its structure may suggest which of the countless aspects of historical reality we should focus on; morphology, acting as a catalyst for historical intuition. With hundreds of texts – let alone more – this becomes implausible, and the “vertical” link between the text and the world is replaced by a “horizontal” one among texts that are all on the same plane. Hemingway’s sentences had led me to the war; Eliot’s segment, to other segments. It’s striking, how literature-bound the quantitative approach has turned out to be. There was so much to say, about Hetty’s confession in *Adam Bede*; we said almost nothing, because the logic of measurement had shifted our gaze from the world-shaping power of form to the inner structure of the literary system. The whole system, in principle: not just those “generals of literature” – in the ironic words of the Russian Formalists – who have always been the heroes of interpretive activity. One dimension is lost, in the shift, from hermeneutics to quantification, and one dimension is gained: we still know almost nothing about literary systems – which is absurd – and those 1,100 dots of Figure 1 had been a small but real step in the right direction: the link they had established between a morphological feature and the historical fate of novels had been like laying down a couple of pieces in the gigantic puzzle of the literary field. A puzzle, let it be said, that the hermeneutic tradition, with all its creative intelligence, has never been curious about.

5.

Near the end of the second part of *Henry IV*, one of Falstaff’s companions, Pistol, arrives with news about the king. Shallow, a country justice, questions him:

**SHALLOW:** Give me pardon, sir. If, sir, you come with news from the court, I take it there’s but two ways, either to utter them, or to conceal them. I am, sir, under the King in some authority.

10 The list of major phenomena (let alone minor ones) that could count as “the world” for any work of literature is virtually infinite. For an American in Europe in the 1920s, it would indeed have included the trenches of World War I, but also a socialist revolution, cars and airplanes, a decade of unprecedented sexual freedom, civil wars, rationalist architecture, the radio, incredible experiments in painting and music, the beginning of hyper-inflation … A single form usually reacts only to a few of these phenomena, which interpretation may succeed in isolating; with a large corpus, however, formal mechanisms multiply in every direction, and the threads that lead from the world to the work become hopelessly tangled.
PISTOL: Under which king, bezonian? Speak or die.

SHALLOW: Under king Harry.

PISTOL: Harry the Fourth, or Fifth?

There are no Harrys for now, and one doesn’t have to choose once and for all between the kingdom of hermeneutics and that of quantification. But they are different, and so are their principles. They have a common starting point in the reality of form – as it should be, because form demarcates an aesthetic discipline like literary history from other branches of the human sciences. And yet, even this common ground seems only to highlight their contrast. Interpretation moves between form and the world, pursuing the broad socio-historical significance of literary works; quantification moves between form and form, trying to chart a reliable atlas of literature. For the first, form is a process: a transformation of the given world which can only be understood when it is met with suspicion, countered, and ultimately reversed. For the other, form is a finished product, to be measured with a cool head, and placed within a many-sided system of relations. One leads towards history; the other, towards theory. Here, the pathos of struggle; there, that of discovery. Great passions, both. But too exclusive to join forces towards a common goal.

I had no idea I would reach this conclusion, when I started studying for this essay. Having worked now with one method and now with the other for quite a long time – but never together – I had been fantasizing about a book to come that would be a synthesis of sorts, and of which this essay could be a prefiguration. Then I started describing what I saw, and the fantasy vanished: no matter what the angle of observation was – the brief overlap between the two practices; the conceptual interplay of normative and frequent; the way of understanding form – the two methods were always drifting apart from each other.

So, let them drift. When paths diverge, some things become clearer; such, I hope, is the case for the logic behind interpretation and measurement described in these pages; a logic that needs neither mitigation nor diplomacy, but rather asks to be taken to its ultimate consequences. Needless to say, this is only my view of the matter; and someone else might find a way to achieve a good synthesis – tomorrow. Then, things will change. Right now, as Arnold Schönberg once said, we are in a situation in which the middle road is the only one that does not lead to Rome.
References


Moretti, Franco. “Operationalizing: or, the function of measurement in modern literary theory”, Literary Lab Pamphlet 6, 2013.


| Abstrakt |

Franco Moretti

Którego króla, ignorancie?11 Badania literackie pomiędzy hermeneutyką a kwantyfikacją

Jaka jest relacja między ilościową historią literatury ostatnich dwudziestu lat a starszą tradycją hermeneutyczną? Odpowiedzi na to pytanie wydają się dwojakie: dla wielu przedstawicieli obozu interpretacyjnego, nie da się pogodzić tych dwóch podejść, a nowsze podejście ma bardzo małą wartość krytyczną jeśli w ogóle można mówić o wartości w tym wypadku. Z kolei dla większości badaczy ilościowych są one całkowicie zgodne i w zasadzie wzajemnie się uzupełniają. W niniejszym tekście przedstawię trzecią możliwość, która krok po kroku wyprowadzona zostanie z porównania sposobu funkcjonowania obydwu strategii. Mam tu na myśli dosłownie sposób funkcjonowania; oparty na przekonaniu, że praktyki – czyli to, czego uczymy się poprzez wykonywanie, z zawodowego przyzwyczajenia i często bez pełnej świadomości tego, co robimy – często mają większe implikacje teoretyczne niż same stwierdzenia teoretyczne. Innymi słowy: ważniejsze jest zrozumienie tego, jak funkcjonuje paradygmat badawczy, niż określenie tego, jak powinien on funkcjonować. Taki jest plan niniejszego artykułu.

Słowa kluczowe: ilościowa historia literatury, tradycja hermeneutyczna, literatura porównawcza, paradygmaty badawcze

11 Przetłumaczone z Henryka IV przez Stanisława Barańczaka (Shakespeare 1299).
Abstract

Franco Moretti

Under which King, Bezonian? Literary Studies between Hermeneutics and Quantification

What is the relationship between the quantitative literary history of the past twenty years and the older hermeneutic tradition? Answers have typically been of two kinds: for many in the interpretive camp, the two approaches are incompatible, and the newer one has little or no critical value; for most quantitative researchers, they are instead perfectly compatible, and in fact complementary. Here, I will propose a third possibility, that will emerge step by step from a comparison of how the two strategies work. How they work, literally; in the conviction that practices – what we learn to do by doing, by professional habit, and often without being fully aware of what we are doing – have frequently larger theoretical implications than theoretical statements themselves. In other words: understanding what a research paradigm does, rather than what it declares it wants to do. This is the plan.

Keywords: quantitative literary history, hermeneutic tradition, comparative literature, research paradigms

Bio

Franco Moretti taught literature in Italy and the US from 1972 to 2016. He is the author of several books, among them The Way of the World, Maps Graphs Trees and Distant Reading. He is currently working on a collection of theoretical essays and a book on tragic form.

E-mail: moretti@stanford.edu

ORCID: 0000-0003-2866-9585