FORMALISM AND DIGITAL RESEARCH OF LITERATURE

Marek Debnár[1]
Constantine the Philosopher University in Nitra, Slovakia
mdebnar@ukf.sk

Abstract

The relation between the humanities and information technologies has become so strong in recent decades that it is no longer possible to see this relationship as a mere temporary phenomenon. Together with massive digitalization of books, journals and other texts, collected into extensive electronic libraries and hypertextual databases, it is now necessary to rethink and redefine not only the concept of reading, but to specify new possibilities for analysing literary and specialized texts. The aim of this study is to point at new approaches to reading large text collections in the light of Moretti's method of distant reading. This paper uses the methodological issues of relation between distant reading and Russian formalism as background for this consideration.

Keywords: distant reading, humanities, Russian formalism

Distant reading

The idea of "world literature" as a set of shared cosmopolitan values has been known since the time of J. W. von Goethe. The proposed methodology of literary science transcends the horizon of a complete work of one author, period or genre, appearing in Russian formalism. The idea of *distant reading,* i.e. reading mediated by IT tools, that was introduced by Franco Moretti in the essay *Conjectures on World Literature* in 2000[2], refers to both of these perspectives.

Moretti's approach to the idea of "world literature", as was understood by Goethe in conversations with Eckerman (1827), and twenty years later by Marx and Engels in the *Communist Party Manifesto*[3] (1848), is critical. He describes it as a contemporary intellectual gesture limited by the area of Western Europe, more precisely, by "German philologists working on French literature" (Moretti 2000, 54), which does not mean that he turns away of this concept. "World literature" is the main area of his interest, but not as a label for a set of literary works but rather as a methodological problem. In order to define the term "world literature", we need to find such a research method that corresponds to the current state of the problem; a method that exceeds a simple but unviable notion that it is enough to read more.

In *Conjectures on World Literature*, Moretti does not openly avow the thoughts and procedures of Russian formalism. However, he is so strongly inspired by formalism that for his theory of distant reading that he employed the title "new formalism without close reading" (Arac 2002, 38). Moretti then describes his method as "a little pact with the devil: we know how to read texts, now let's learn how *not to read* them. Distant reading, where distance is a *condition of knowledge*, it allows us to focus on units that are much smaller or much larger than the text: literary formation, themes, tropes – or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can

---

[2] Essay was originally published in the magazine *New Left Review* No. 1/2000 and was later included by Moretti to the book *Distant Reading*, Verso 2013.

[3] With reference to a text: "In place of the old local and national seclusion and self-sufficiency, we have intercourse in every direction, universal interdependence of nations. And as in material, so also in intellectual production. The intellectual creations of individual nations become common property. National one-sidedness and narrow-mindedness become more and more impossible, and from the numerous national and local literatures, there arises a world literature." (Marx – Engels 1977, 151 – 152).

justifiably say, that less is more. If we want to understand the system in its entirety, we must accept losing something." (Moretti 2000, 57).

The enthusiasm for a new methodological approach is indisputable, as is the latent inspiration for formalism that can teach literary history the regularities of a literary field, as written in the study *The Slaughterhouse of Literature* (*The Slaughterhouse of Literature* 2000). In this study, published in the same year as *Conjectures on World Literature*, he adds: "Form is a repeatable element of literature – what returns fundamentally unchanged over many cases and many years" (Moretti 2013, 86). In this period, Moretti speaks very broadly about formalism and openly admits the inspiration by Viktor Shklovsky and his *Theory of Prose* (1929) five years later in his book *Graphs, Maps, Trees – Abstract Models for Literary History* (*Graphs, Maps, Trees – Abstract Models for Literary History* 2005). The work belongs to the founder's work in the field of literary science on textual analysis using IT tools, whereby Moretti represents three models of quantitative analysis. Using *graphs,* he analyses the history of changes in the complexity of the novel genre; through *maps*, he illustrates geographical changes in aspects of English "rural" prose; and genealogical *trees* enable him to analyse various mutations in the detective genre. A trinity of quantitative analyses represent three approaches to what was described by Eichenbaum as a goal of formal method in the text devoted to work of the Petrograd *Society for the Study of Poetic Language (Opoyaz)*, called *The Theory of the Formal Method* earlier in 1925: "We do not incorporate into our work issues involving biography or the psychology of creativity, assuming that those problems, very serious and complex on their own, ought to have their place in other disciplines. We are concerned with finding in evolution the features of immanent historical laws – that is why we ignore all that seemed, from this point of view, *circumstantial*, not concerned with literary history. We are interested in the very process of evolution, in the very dynamics of literary form, insofar as it is possible to observe them in the facts of the past. For us, the central problem of the history of literature is the problem of evolution outside individual personality – the study of literature as a self-formed social phenomenon" (Eichenbaum 1971, 50).

Besides historical patterns of the form, stressed by Moretti in both texts in 2000, we cannot even define a biographic author mentioned by Eichenbaum in his approach. Formalism thus reacts to an approach that reduced literary history to the psychological interpretation of works of "great authors". This movement did not perceive the literary work and its value framework as an isolated product; however, they perceived it on the background of other works and in connection with them. From the point

of view of literary form development, the most important aspect is the impact of work on work, which was the reason the formalists broadened the sphere of research from "high literature" to the popular and folk literature that played a significant role in the creation and development of literary genres and forms. This quantitative expansion of literary research field forms the basis not only for formalistic research but also the procedures in the field of textual analysis that are based on information technologies.

Moretti extended this deflection to a whole set of authors whose works are in a close dialogue with literary history: "Trouble with close reading in all of its incarnations (from the new criticism to deconstruction) depends on extremely limited criteria... we invest so much in individual texts only because we think that very few of them really matter."(Moretti 2000, 57). Moretti's version of "formalism without close reading" is primarily about creating abstract models for the development of literary forms on the ground of specialized text corpora, therefore shifting the attention from privileged authors and works to the level of individual changes in the whole genre.

False clues

An example of such an exploration is the use of indicia in relation to the development of English detective fiction from the time of their first occurrence, that is, from the end of the nineteenth century. The topic is discussed in the last chapter of *Graphs, Maps, Trees*, which is an extended version of the study *The Slaughterhouse of Literature* and, at the same time, best illustrates Moretti's relationship to Russian formalism. While in *Slaughterhouse* he does not acknowledge his inspiration by Shklovsky, in the study *Trees* he directly refers to him. In the *Theory of Prose*, specifically in his essay *Mystery novella*, Shklovsky analyses (along with other works) Arthur Conan Doyle's Sherlock Holmes stories from the perspective of genre construction, form and time sequence of syuzhet. At first, there is nothing extraordinary, though he is attracted by a special way Doyle works with clues: "Secondary data are the most important, integrated in a way, a reader will not notice them" (Shklovsky 1971, 142). From a schematic point of view, these are common short stories with a secret: "this scheme was not created by Conan Doyle, but he did not even steal it. It results from the subject matter" (Shklovsky 1971, 143). Speaking about clues (Shklovsky calls them "hints"), he writes: "Everyone intent on engaging in the creation of Russian plot-based literature should pay close attention to Conan Doyle's use of clues and the way the denouement emerges out of them"(Shklovsky 1971, 143).

In creating the genealogical tree of detective genre development, outlined in Moretti's *Trees*, it is clear that the functional use of clues (present, visible, decidable, etc.) appearing in Conan Doyle's and some of his contemporaries' work is gradually becoming the main developmental branch of detective genre fiction. The outcome of these investigations is that only works of that period that contain the right variants of clues – and Moretti proves that it is a really small percentage–are brought into the literary awareness of the present time. In the case of Doyle, Moretti suggests one possible hypothesis as to the success of Sherlock Holmes stories could be based on the fact that they had been posted in the prestigious *Strand Magazine*: "But why should form be the decisive reason for survival? Why not social privilege instead – the fact that Doyle was writing for a well-established *Strand Magazine* and his rivals were not?... So, I went to the library, where I discovered that, in the course of the 1890s, over one hundred detective stories by twenty-five different authors had been published in the *Strand Magazine* alongside Sherlock Holmes. Since so many authors had access to the same venue as Doyle, the social privilege objection lost its force." (Moretti 2005, 74).

In this case, Moretti proceeded with data examination in the manner of Vladimir J. Propp in his pioneering work *Morphology of the Folktale* (1928), wherein he studied one particular variant of a tale based on one file, and without the use of IT tools. This enabled him to identify thirty-one functions that make up the basic building blocks of every "magic folktale" (Propp, 1971, s. 35-72), or a fairy tale which contains a magical object (for example "self-setting tablecloths"). Propp worked with a precisely defined corpus of texts, with the "material" that contained folktales listed in the index of Aarne – Thompson, a folklorist and a collector of folk literature, marked by numbers 300 – 749. Propp then subsequently narrowed down this collection, analysing 175 total texts. Obviously, Propp had read all of these texts and extracted the individual variants of the whole collection himself. At first, Moretti carried out his research on much larger set of texts using digital tools, but the results' verification principle remained the same as in Propp's case.

Quantitative results reveal much more about the development of form, but by reading all the short stories of one particular collection of texts Moretti also revealed something more important than just confirming Conan Doyle's *status quo*. "Secondary" and forgotten detective stories, that no one paid attention to, create two new branches of the genealogical tree of detective fiction with regards to plot and storyline construction. The enthusiasm for this discovery was so great that Moretti summarized this ex-

perience in words: "The more one looked in the archive, in other words, the more complex and "Darwinian" became the genre's morphospace."(Moretti 2005, 74). This language is not coincidental. In the preface to the second edition of *Modern European Literature: A Geographical Sketch* (Modern European Literature: A Geographical Sketch) included at the beginning of the book *Distant Reading* (Distant Reading 2013), Moretti admits that one of his most powerful inspirations was the evolutionary theory of the origin of species by Ernst Mayr which he wanted to apply to the evolution of morphological (genre) transformations of literary field (see Moretti 2013, 1 – 3). Similar to Propp, Shklovsky and other formalists, also in Moretti's "quantitative formalism", in the pamphlet of *Stanford Literary Lab* he called morphological (genre) category a presumption of quantification analysis. This makes it a necessary aspect of distant reading: at the beginning of the research, it is used to set up a typological "spectrum of variations" (Mayr), and at the end, it is used to set up the genealogy of morphospace with all its developmental branches.

   Another necessary aspect of distant reading is *a distance* that makes it possible to "read" quantitative data from the point of view of form: "we must step back from quantitative field and look for solution in a completely different one, and it is a field of form (morphology)" (Moretti 2005, 24). Distance from data is what enables their "reading", or interpretation. This aspect is also observed by Sean McCann in the article *A Few Quibbles about Moretti's Graphs, Maps, Trees* (A Few Quibbles about Moretti's Graphs, Maps, Trees 2011): "It is wonderful to see the graphs and to get a sense of the vast range of material that still needs to be discovered and understood ...but as Moretti notes, all this is just data until hypotheses are generated (produced, by the way, via "interpretation") (McCann 2011, 109).

   After clarifying Moretti's formalistic resources, we are getting back to the distant reading, metaphorical description of quantitative formalism that Moretti on mentioned in *Graphs, Maps, Trees* once. He describes it very briefly as a reading that does not put emphasis on details but on the distance as a specific form of knowledge that "reveals broader interconnections. Shapes, relations, structures. Forms. Models" (Moretti 2005, 1). In the final texts, he even introduces a distinction between "explanation"–that falls more within cognizance of quantitative formalism–and the interpretation of a text, as if he was already aware of the risk in working with a great deal of material, and that the hypothesis is likely to prove true (concerning the size of the text archive). In other words, the outcome of quantitative formalism, for which the morphological category is a presumption of analysis, is that which had been expected at the beginning. Moretti's study of

clue occurrence in detective genre was also based upon the same assumption. He explicitly focused on clues as found in Doyle's work, as he considered them a dominant evolutionary shift within the genre. He focused beforehand on the apologetics of the Western European Detective canon in order to create a model for the whole genre tree. Such a procedure resulted in tautology, and tautologies seem true in every possible interpretation.

Conclusion

It has been thirteen years since the first publication of *Conjectures on World Literature* (2000) until its re-publishing in the collection of all Moretti's texts devoted to *Distant Reading* (2013). The range of digital archives has grown enormously during that time, eliminating the old barrier to literary research represented by the unavailability of texts. However, new barriers have emerged, such as the impossibility of identifying a great deal of forgotten, secondary texts, so called *The Great Unread,*[4] problems that have also occurred in other areas of textual analysis. Nevertheless, the use of IT tools has its future in literary and broader humanities research. The oft-mentioned crisis of the human sciences is also caused by their procrastination, and is accompanied by a sort of a priori mistrust towards new technologies. In defence of our approach, it can be said that if humanities are losing their attractions at present, they should not hesitate to seek new approaches to find meaning and use.

---

[4] A term *The Great Unread* was introduced to literary science by a narratologist Margaret Cohen (see Cohen 2009).

References

Arac, Jonathan. 2002. Anglo-globalism? *New Left Review* 16: 35 – 45.

Cohen, Margaret. 2009. Narratology in the Archive of Literature. *Representations* 108 (1): 51 – 75.

Eichenbaum, Boris. 1971. Teória formálnej metódy. In *Teória literatúry – výber z „formálnej metódy"*, ed. Mikuláš Bakoš. Bratislava: Tatran.

Marx, Karol – Fridrich Engels. 1977. Manifest komunistickej strany. In *Antológia z diel filozofov. Marxisticko-leninská filozofia*, ed. Ladislav Holata. Bratislava: Pravda.

Moretti, Franco. 2000. Conjectures on World Literature. *New Left Review* 1.

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

Moretti, Franco. 2013. *Distant Reading*. London: Verso.

Propp, J. Vladimír. 1971. *Morfológia rozprávky*. Bratislava: Tatran.

Šklovskij, Viktor. 1971. *Teória prózy*. Bratislava: Tatran.