



**Conference Reader**  
4<sup>th</sup> Annual Conference of  
Computational Literary Studies  
**CCLS2025 Kraków**  
July 3-4, 2025

---

<b>Venue</b>	Jagiellonian University, Faculty of Philology Al. Mickiewicza 9, 31-120 Kraków
<b>Local Organizer</b>	Jagiellonian Centre for Digital Humanities Jan Rybicki
<b>Web</b>	<a href="https://jcls.io/site/ccls2025/">https://jcls.io/site/ccls2025/</a>
<b>Contact</b>	<a href="mailto:jchc@uj.edu.pl">jchc@uj.edu.pl</a>
<b>Hashtag</b>	#CCLS2025
<b>JCLS Editors</b>	Evelyn Gius, Christof Schöch, Peer Trilcke
<b>JCLS Editorial Assistants</b>	Svenja Guhr, Julian Häußler, Élodie Ripoll, Henny Sluyter-Gäthje
<b>Contact JCLS</b>	<a href="mailto:info@jcls.io">info@jcls.io</a>   <a href="https://jcls.io/site/contact/">https://jcls.io/site/contact/</a>

# Conference Programme

## Thursday | July 3, 2025

9:15 a.m. to 9:30 a.m. | Opening

9:30 a.m. to 10:30 a.m. | Session 1

- Fotis Jannidis, Rabea Kleymann, Julian Schröter, Heike Zinsmeister: **Do Large Language Models Understand Literature? Case Studies and Probing Experiments on German Poetry**
- Keli Du, Uygur Navruz, Nazan Sınır, Julian Valline, Christof Schöch, Sarah Ackerschewski: **Reconstructing Shuffled Text. Bad Results for NLP, but Good News for Using In-Copyright Text**

11:00 a.m. to 12:30 a.m. | Session 2

- Maria Levchenko: **Computational Analysis of Literary Communities: Event-Based Social Network Study of St. Petersburg 1999-2019**
- Gilad Aviel Jacobson, Yael Dekel, Itay Marienberg-Milikowsky: **From Readers to Data. Uncertainty in Computational Literary Citizen Science**
- Julia Neugarten: **A Powerful Hades is an Unpopular Dude: Dynamics of Power and Agency in Hades/Persephone Fanfiction**

1:30 p.m. to 3:00 p.m. | Session 3

- Daniil Skorinkin, Boris Orekhov: **The Outward Turn: Geocoding the Expansion of Fictional Space in Russian 19th Century Literature**
- Svenja Guhr, Jessica Monaco, Alexander J. Sherman, Matt Warner, Mark Algee-Hewitt: **Making BERT Feel at Home. Modelling Domestic Space in 19th-Century British and Irish Fiction**
- Eva Eglāja-Kristsons, Anda Baklāne, Valdis Saulespurēns: **Urban Transportation in the Latvian Early Novels or “Why do you use a 19th-century horse-drawn cab when you have a 20th-century taxi?”**

### **3:30 p.m to 4:30 p.m. | Session 4**

- Rongqian Ma, Keli Du, Yiwen Zheng: **Verse within Prose. Annotating and Classifying Narrative Functions of Embedded Poems in Chinese Qing (1644–1912) Vernacular Fiction**
- Natalie M. Houston: **Rhymefindr: An Historical Poetics Method for Identifying Rhymes in Nineteenth-Century English Poetry**

### **5:00 p.m. to 6:00 p.m. | Keynote**

- Maciej Eder: **Text Analysis Made Simple (Kind of), or Ten Years of Stylo**

### **7:00 p.m. | Conference Dinner**

## **Friday | July 4, 2025**

### **9:15 a.m. to 10:45 a.m. | Session 5**



- Katrin Rohrbacher: **Opening Worlds: Narrative Beginnings and the Role of Setting**
- Noa Visser Solissa, Andreas van Cranenburgh, Federico Pianzola: **Event Detection between Literary Studies and NLP. A Survey, a Narratological Reflection, and a Case Study**
- Andrew Piper: **Towards a Moral History of the Novel Using Large Language Models**




### **11:15 a.m. to 12:45 p.m. | Session 6**

- Julia Havrylash, Christof Schöch: **Exploring Measures of Distinctiveness. An Evaluation Using Synthetic Texts**
- Allison Keith, Antonio Rojas Castro, Hanno Ehrlicher, Kerstin Jung, Sebastian Padó: **A Computation Analysis of Character Archetypes in the Works of Calderón de la Barca**
- Yuri Bizzoni, Pascale Feldkamp, Kristoffer L. Nielbo: **Encoding Imagism? Measuring Literary Imageability, Visuality and Concreteness via Multimodal Word Embeddings**

### **12:45 p.m. to 1:00 p.m. | Closing**

# The Outward Turn: Geocoding the Expansion of Fictional Space in Russian 19th Century Literature

Daniil Skorinkin <sup>1</sup>   
Boris Orekhov <sup>2, 3</sup> 

1. Digital Humanities Network, University of Potsdam , Potsdam, Germany.
2. Laboratory for Digital Research of Literature and Folklore, Institute of Russian Literature , Saint Petersburg, Russia.
3. School of Linguistics, HSE University , Moscow, Russia.

## Citation

Daniil Skorinkin and Boris Orekhov (2025). "The Outward Turn: Geocoding the Expansion of Fictional Space in Russian 19th Century Literature". In: *CCLS2025 Conference Preprints 4* (1). [10.26083/tuprints-00030144](https://doi.org/10.26083/tuprints-00030144)

Date published 2025-06-17

Date accepted 2025-04-17

Date received 2025-02-10

## Keywords

geocoding, maps, Russia, 19th century, prose, novel

## License

CC BY 4.0 

## Reviewers

## Note

This paper has been submitted to the conference track of JCLS. It has been peer reviewed and accepted for presentation and discussion at the 4th Annual Conference of Computational Literary Studies at Krakow, Poland, in July 2025. Please check [jcls.io](https://jcls.io) for the final journal version.

## Abstract.

We examine the large-scale geospatial dynamics of Russian prose literature in the 19th century. Specifically, we analyze how the distribution of location mentions shifts from the early 19th-century romantic era to the late 19th-century realist period. We demonstrate how realist literature, with its emphasis on portraying 'typical characters in typical settings', moves away from the historical (and often heavily mythologized) landscapes of Russia, Poland, Ukraine, and the Baltics. Instead, it increasingly focuses on the then-new capital, Saint Petersburg, as well as Western Europe and the expanding eastern and southern peripheries of Russia, reflecting the country's ongoing military and economic expansion.

## 1. Introduction

Of all 'distant reading' methods, geocoding is the one that most tangibly embodies the 'distance' metaphor. With maps, one can literally zoom in and out of vast research material, possibly consisting of thousands of texts, all laid out on a geographic projection of the Earth, and produce conclusions, generalizations, and interpretations on a grand scale.

This does not mean that every geocoding of literature is always meaningful - as Döring (Döring 2013) put it, 'the benefit of any map of literature has to be that it visualizes things that would otherwise remain invisible' and for some literary maps, "[t]here seems to be hardly any analytical value in" them. Literature reduced to dots, lines, and polygons (the basic units of any map) loses most of its inner complexity, and there is always the danger of throwing the baby out with the bathwater. But at the same time, reduction is exactly what gives strength to any modelling attempt in research: only by reducing the complexity and detail, we can see the large drifts of literary movements and the long dynamics of cultural development that is not inferable from close reading of a selection of 'significant' texts.

In our work, we apply mapping and geocoding to study the large-scale geospatial dynamics of Russian prosaic literature over the course of the 19th century, a time when a Russian novel became a global cultural phenomenon through the works of Gogol, Tolstoy,

Turgenev, Dostoevsky and other authors. We analyse the changes in the distribution of mentions of geotaggable toponyms between the two extremely important periods of Russian literature: the early 19th-century romantic era and the late 19th-century realist period. We show how realist literature with its tendency to depict 'typical' characters in 'typical' settings (Fridlender 1971, 105) and not shy away from 'ordinary' and 'average', turns from the mythologized landscapes of historical Russia, Ukraine, Poland and the Baltics, to Western Europe, the then-new northern capital and trading outpost of Saint-Petersburg, and the 'new' eastern and southern peripheries of Russia as the country continues its military, cultural and economic expansion in all directions.

## 2. Corpus and research design

As members of the PyZeta team put it in the description of their project, "[t]he methodological and epistemological paradigm of comparison is deeply rooted in the Humanities" (Du et al. 2025). In our experience, a research endeavour in computational literary studies typically benefits from having a clear two-sided comparison. Even if such comparison comes at a price of some simplification. Therefore we chose to structure our research around the comparison between the prosaic works of Russian 19th-century realism and the Russian romantic prose that preceded it.

The problem of defining realism in literature is a long-standing one. As Fanger (Fanger 1998, 3) put it, "few literary terms have suggested more and signaled less than 'realism'". Realism often seems too broad a term, combining too many things that lack a common denominator. To quote Molly Brunson (Brunson 2016, 2), "this monolithic presence of realism more often than not splinters into equivocation or endless classification. It is little wonder, given the dizzying array of objects that must crowd beneath this singular term". And for scholars of Russian literature, this was additionally complicated by the ideologically charged understanding of realism in the Soviet era, which led to a strong aversion to the term in the post-Soviet times (see e.g. Vdovin et al. 2020). However, Vdovin et al. (Vdovin et al. 2020) also show that removing the term completely, while continuing to talk about romanticism, classicism and other traditionally labeled literary movements, does not seem feasible either. It is therefore reasonable to keep using it, acknowledging the ambiguity and inner contradictions of the term.

Luckily, in this particular academic endeavour we neither intended nor needed to answer the 'what is realism' question. For us, it was enough to adopt a functional definition that would allow us to make a split in a collection of Russian prose (without any prior genre or literary movement markup) and obtain a 'realistic enough' corpus for computational analysis. We therefore followed the chronological approach. In many cases Russian literary realism is defined as something that started in the 1840es with the projects of the so-called Natural School (Brunson 2016) or more specifically mid-1840es (Bowers 2022, 2). 1845 was the year of the publication of the 'The Physiology of Saint-Petersburg' (Физиология Петербурга), the first artistic manifesto of the Natural School, compiled by Nikolai Nekrasov. In 1846 the second one, 'The Petersburg Collection' (Петербургский сборник) was published by Nekrasov. The first one was a compendium of short 'physiological sketches' by Vissarion Belinskiy, Nikolai Nekrasov, Dmitry Grigorovich, Vladimir Dal, and the Ukrainian writer Yevhen Hrebinka. The

second, bigger volume contained the first large novel by Fyodor Dostoevsky ('The Poor Folk'), as well as texts by Ivan Turgenev, Alexander Herzen, Ivan Panayev, Apollon Maikov and Vladimir Odoevsky. Also in 1846 Belinsky, the most prominent Russian critic of the era, called for new literature that "dealt with life and reality in their true light". We chose to adopt the year 1845 as the 'starting date' of the realist period in our corpus.

As the end of the clearly realist period we selected 1890. This year is frequently named as the starting point of modernism in Russian literature (Douglas Clayton 2016, Ioffe 2009). Of course, there were many realistic works created after that date too (realism never really 'stopped' the same way e.g. classicism did), but without any reliable metadata we had to rely on temporal borders and chose to stop at 1890 to keep modernism out of our corpora.

In the end, having consulted with a number of specialists in Russian 19th-century prose, we received their blessing to consider for the purposes of our quantitative investigation everything written between the years 1845–1890 to belong to realism and everything that fell into the period between 1800 and 1840 to belong to romanticism. We are aware, of course, of how imprecise this division is. However, as Algee-Hewitt et al. (Algee-Hewitt et al. 2018) put it, "[d]irty hands are better than empty," so we continued our research, hoping that the size of the corpus would rectify the lack of quality in our crude criterion for the split.

To compose a corpus of Russian 19th-century prose for our study, we used two main sources of texts. One of them was the 'Corpus of Russian narrative prose' by Oleg Sobchuk, published in the Open Repository on Russian Literature and Folklore (Sobchuk and Lekarevich 2025). Another source, also published in the same repository, was the corpus of the 'Forgotten novels of Russian writers from the collections of the Pushkin House (1857–1917)' by Elena Kazakova (Kazakova 2024). We then filtered out everything that was written outside of the periods we were interested in (1800–1841 and 1845–1890). Our resulting corpus consists of 506 texts between the years 1800 and 1890, of which 96 belong to the romantic subcorpus and 421 – to the realist subcorpus. The list of all texts and their dates of publication is available here.<sup>1</sup>

While this corpus is far from being a comprehensive source of Russian 19th-century literary heritage, it contains prosaic work by all the well-known authors of the period (Pushkin, Lermontov, Gogol, Turgenev, Goncharov, Dostoevsky, Tolstoy), as well as a large number of lesser known writers. The total number of word tokens in the corpus is 46.4 mln.

### 3. Methods

Geocoding literary texts to explore the relationship between literature and geography has a long tradition that spans more than a century. As early as 1910s (Bartholomew 1914) one can find numerous literary maps based on the works of Balzac, Dickens, Dumas, George Eliot, and other authors. In the field of Digital Humanities, the application of geocoding in literary studies has been notably championed by Franco Moretti in his

1. [https://github.com/DanilSko/mapping\\_russian\\_prose](https://github.com/DanilSko/mapping_russian_prose)

book 'The Atlas of the European Novel' (Moretti 1999). He stated that "geography is not an inert container, is not a box where cultural history 'happens', but an active force that pervades the literary field and shapes it in depth." Mapping literature, according to Moretti, makes visible "the connection between geography and literature" and reveals "significant relationships that have so far escaped" scholarly attention. Through a series of case studies, he examined the geographical dimensions of 19th-century European literature, highlighting the prominence of Paris in French novels, contrasting depictions of urban and rural environments in English literature, and analyzing representations of the Russian landscape in the works of authors such as Dostoevsky and Tolstoy. Since Moretti's work, there has been, as Döring described it, "a small boom in maps of literature" (Döring 2013). In Bodenhamer et al. 2010 the emergence of the Spatial Humanities was proclaimed, stating that by 2010, there had been "wide application" of Geographic Information Systems (GIS) to "historical and cultural questions." Multiple scholars have contributed to this growing field. To cite just a few examples, Döring (Döring 2013) examined the toponymy of Berlin in German literature after 1989; Kuzmenko and Orekhov (Kuzmenko and Orekhov 2016) geocoded the Russian national poetic corpus and analyzed the frequency of references to countries and cities; and Barbaresi (Barbaresi 2018) mapped the satirical literary magazine *Die Fackel*. More recent example is the paper by Wilkins et al. (Wilkins et al. 2024), who mapped the geographies of American fictional books and compared them to those found in non-fiction texts.

In all those recent cases, NER tools were used to extract toponyms. We followed these footsteps and for the initial location extraction we utilized *Natasha* (natasha/natasha 2024), a natural language processing library for the Russian language with NER toolkit. We extracted approximately 12,000 unique locations from our corpus, which were then manually filtered to eliminate evident homonyms. Specifically, we excluded toponyms frequently used as surnames in our corpus (e.g., 'Rostov', which in 90% of the cases was the surname of one of the member of the Rostov family in *War and peace*) and those typically employed metaphorically (e.g., 'Babylonian').

The filtered toponyms were subsequently geocoded using the 'wikipedia' Python package. Geocoding helped us remove duplicates: different spellings and word forms of the same city (e.g. Saint-Petersburg can be spelled in at least 6 different variants in our corpus), country, or river were merged based on the matching coordinates. Thus, the pair of latitude and longitude became the primary ID of each location that we analysed. Extracted and geocoded locations are available here.<sup>2</sup>

We then produced symbol maps that overlay frequencies in the texts onto geographical locations. We analysed the raw frequencies of locations and their relative increase or decrease in frequency between the periods of romanticism and realism.

It is important to note that our analysis of geographical material focused not on where events take place but on all mentions of place names. This highlights the writers' and Russian society's attention to these parts of the world.

Additionally, to compare contexts of the same toponyms in the two periods we used word2vec (Mikolov et al. 2013). With this we attempted to detect the contextual change for some of the most frequent locations that were found in both corpora.

2. [https://github.com/DanilSko/mapping\\_russian\\_prose](https://github.com/DanilSko/mapping_russian_prose)



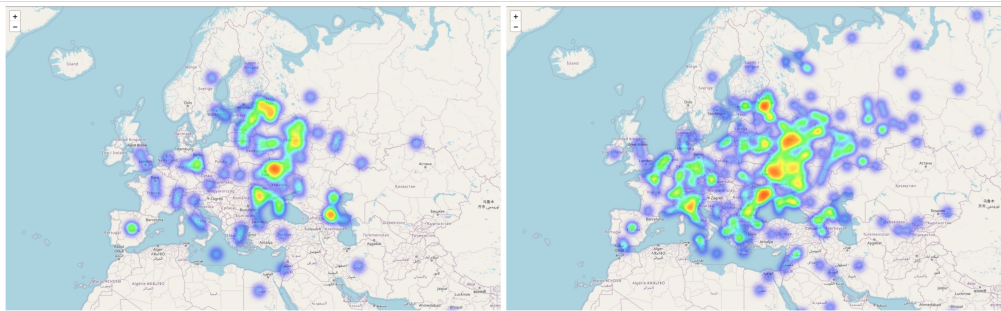


Figure 1: A heatmap depicting location frequencies in romantic (left) and realist (right) texts, visualized through surface occupied and color intensity. Focused on Eurasia.

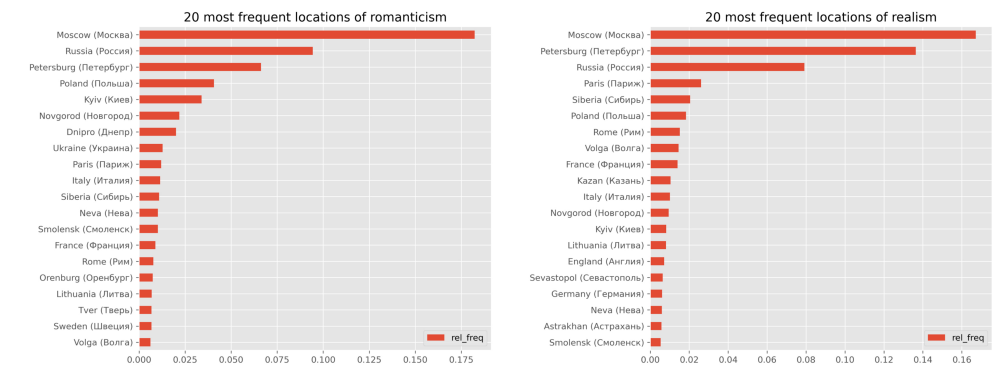


Figure 2: Top 20 locations by relative frequency in romantic (left) and realist (right) texts.

#### 4. Results

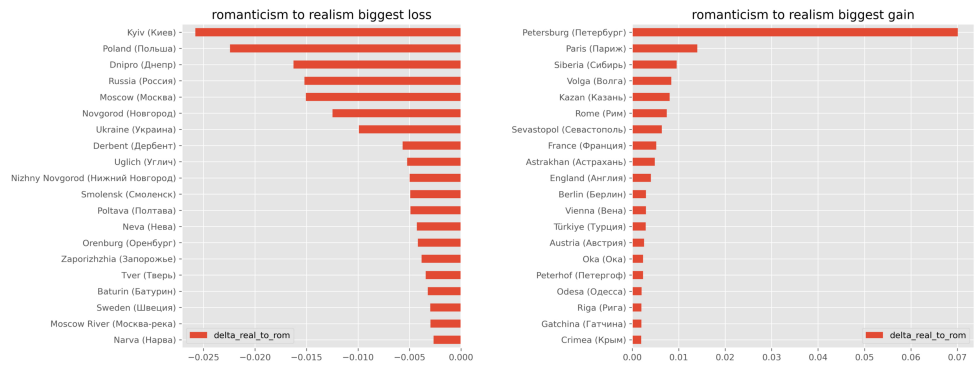
A comparison of the geographical distribution of locations in the romanticist and realist corpora reveals a discernible shift. Figure 1 shows two heatmaps reflecting the frequencies of geotagged locations in both corpora.

This visualisation already demonstrates certain key differences, such as relatively more attention to Western Europe in the realist period, as well as a bigger relative presence of Saint Petersburg. However, it is hard to analyze such heatmaps in detail. Figure 2 contains a more traditional bar plot diagram with the top 20 locations by relative frequency in each of the two subcorpora, providing a more detailed zoom into their differences.

Although Moscow is the most frequently mentioned location in both corpora, its dominance significantly diminishes in the realist texts. In the romanticist corpus, Moscow's mentions surpass those of Saint Petersburg by a factor of 2.5, whereas in the realist corpus, Saint Petersburg's mentions are only 20% fewer than Moscow's.

The emergence of Saint Petersburg as a prominent location is unsurprising; it serves as a primary setting for many significant Russian novels of the realist period. Dostoevsky's 'Crime and Punishment' and other works, Goncharov's 'An Ordinary Story' and 'Obломov,' as well as Tolstoy's 'Anna Karenina' and 'War and Peace', are set in Saint Petersburg. Additionally, many lesser-known works of Russian realism are set there. Russian literary tradition often attributes to realism a focus on depicting 'typical' characters in 'typical' settings, and these characters were frequently situated in the then-capital and administrative hub of Saint Petersburg.





**Figure 3:** Top 20 biggest relative loss (left) and relative gain (right) from romanticism to realism.

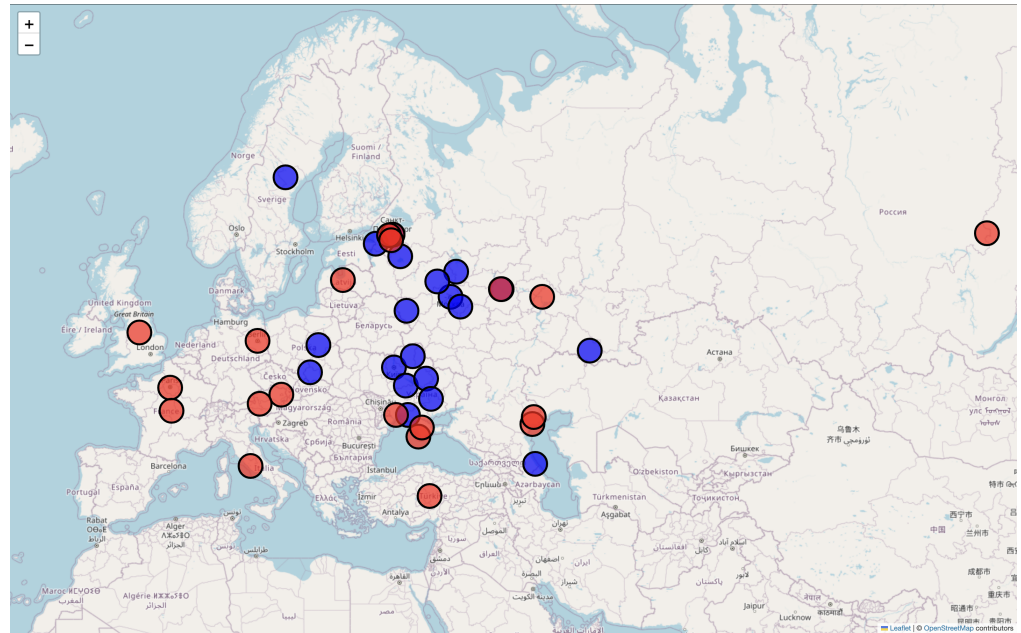
A second significant shift in the realist corpus is the diminished prominence of Kyiv and other Ukrainian locations. While Kyiv ranked fifth in the romantic corpus, it declines to twelfth place in the realist texts, being surpassed not only by Western capitals such as Paris and Rome but also by peripheral Russian locations, including Kazan, the Volga River, and Siberia. Similarly, other Ukrainian locations, such as the Dnipro River and Poltava, exhibit a noticeable decrease in relative frequency.

To systematically capture these changes and emphasize the locations that underwent the most substantial shifts, we calculated the relative overall change in location frequencies and ranked them accordingly. This approach enables a clear visualization of the locations that experienced the most pronounced relative increase or decrease in the realist subcorpus compared to the romantic corpus. The corresponding ranking is presented in Figure 3.

Among the locations that experienced the most significant decline in frequency during the transition from the romantic to the realist period (Figure 3, right), a distinct group comprises Ukrainian toponyms, including Kyiv, Dnipro, Poltava, Baturin, and Zaporizhzhia. This category can be further extended to include neighboring Polish and Baltic locations (Poland, Narva). These Ukrainian, Polish, and Baltic territories — historically contested regions of Eastern Europe — played a crucial role in the literary landscape of Russian historical fiction.

Key historical events, such as the Time of Troubles (Smuta), the Polish–Russian War of 1605–1618, the Cossack uprisings against Polish and Russian rule, and the Great Northern War of 1700–1721 (which accounts for the inclusion of Poltava and Narva), unfolded largely within the territories of present-day Ukraine and the Baltic states. These events provided a rich source of inspiration for numerous Russian-language authors of the romantic era, including Alexander Pushkin, Nikolai Polevoy, Mikhail Zagoskin, Faddey Bulgarin, and Fyodor Glinka. Within their works, these contested lands of Eastern Europe function similarly to Scotland in the novels of Walter Scott, serving as a backdrop for narratives of conflict, heroism, and national identity.

Another significant group of locations prominent during the romantic era but less favored during the realist period includes historical cities of Central Russia, such as Novgorod (the capital of the Novgorod Republic and a popular ‘unrealized alternative’ to monarchical centralized Moscow), Uglich (known for the death of Tsarevich Dmitry,



**Figure 4:** Top-20 locations with the biggest loss (blue) and the top 20 locations with the biggest gain (red) in the realist subcorpus as compared to the romanticist one.

a pivotal event in Russian history), and Moscow itself. Both Moscow and Kyiv, which were among the most frequently depicted locations during the romantic period, lost their literary prominence to Saint Petersburg as Russian literature shifted its focus from a romanticized past to the contemporary present.

Realist literature, oriented towards the present, also shifted its geographical focus westwards — away from the historically contested lands of Eastern Europe and the Baltics, towards the Western European capitals (Paris, Rome, Berlin, Vienna) and countries (France, England, Austria, Switzerland). The characters of realist novels no longer engage in battles in Poland, Lithuania, or Ukraine; instead, they travel to and from France, Italy, or Switzerland, often by train, much like the protagonist of Dostoyevsky's 'The Idiot' or characters of Tolstoy's 'Anna Karenina'.

Concurrently with the westward shift, there was an eastward expansion in literature. In the 19th-century, Russia was actively colonising territories in the Volga region, the Urals, and beyond into Siberia. Notably, "Siberia" exhibits the third largest relative increase in frequency within the realist subcorpus, following only Paris and Saint Petersburg. Other prominent locations in this context include the Volga, the Urals, Kazan, Ufa, and Saratov. While some of these locations possess historical significance, during the realist period they were primarily associated with new economic development. At the same time, these remote places play a bigger role in the ever-growing wave of literature dealing with the topics of prison, penal labour system (katorga) and penal exile of political prisoners, typically members of revolutionary movements.

In Figure 4 we mapped the 20 locations that saw the biggest loss (blue) and biggest gain (red) in their relative frequency in transition from romanticism to realism.

Figure 4 demonstrates that the overall picture is clearly that of an expansion. With the advent of realism, Russian literature transitions from its historical roots in East

	Romanticism	Realism
Saint Petersburg	Moscow (Москва) a village (деревня) a city/town (город) Simbirsk (Симбирск) Kabarda (Кабарда) service (as in army service, government service) (служба) a capital (столица) Kursk (курск) Siberia (Сибирь) to practice (упражняться)	Moscow (Москва) a university (университет) Paris (Париж) a province (провинция) a village (деревня) a grammar school (гимназия) a city/town (город) a capital (столица) Germany (Германия) Kyiv (Киев)
Moscow	Petersburg (Петербург) a city/town (город) a village (деревня) Simbirsk (Симбирск) empty (obsolete) (порожний) an army (армия) kursk (Курск) Kabarda (Кабарда) to practice (упражняться) a tavern (трактир)	Petersburg (Петербург) a city/town (город) a village (деревня) Paris (Париж) Kyiv (Киев) a capital (столица) Petersburg (colloquial) (питер) a monastery (монастырь) Russia (Россия) a university (университет)
Kyiv	an army (армия) Paris (Париж) Smolensk (Смоленск) a fortress (крепость) a monastery (монастырь) a neighborhood (соседство) resurrection (воскресение) a tavern (корчма) a gang (шайка) a province (губерния)	Astrakhan (Астрахань) Kazan (Казань) Berlin (Берлин) Paris (Париж) Vienna (Вена) a horde (typically the Golden Horde) (орда) Germany (Германия) Petersburg (colloquial) (Питер) Ryazan (Рязань) Siberia (Сибирь)

**Figure 5:** Top 10 contextual neighbours for Saint Petersburg, Moscow, and Kyiv in romanticist (left) and the realist (right) corpora.

Slavic civilization (Novgorod, Kyiv, Moscow) to a focus on contemporary life in Saint Petersburg. This shift also facilitates connections with Western Europe (England, France, Italy, Germany) and provides insights into developments at new trading outposts and ports of the Empire, such as Astrakhan, Kazan, Crimea, and Siberia. As the nation undergoes economic and military expansion, new territories are also being explored by its literature.

Of course, there are limitations to what one can find out looking at frequency changes only. Not only do frequencies of toponyms change, but also the contexts in which they are used. To look into the differences we trained two word2vec models on our corpora. We then compared the contextual semantic neighbours (i.e. words with the closest vectors in the model) for the three most prominent capital cities in our corpus: Kyiv, Moscow, and Saint Petersburg. Figure 5 lists the top 10 most similar words for each of the three cities in both the romanticist and the realist corpora.

Comparing the sets of contextual neighbors for these three cities in two models, we can see that in the case of Saint Petersburg there is a very obvious modernisation of contexts.

While the closest word is Moscow in both models (which is totally understandable given the nature of word2vec mechanics: the two capitals appear in very similar functional positions in texts), the second is a village in the romanticist corpus and a university in the realist corpus. Other realist connotation in the list of contextual neighbors is a grammar school (гимназия) — none of those modern education-related words are present in the romanticist contexts. Notable is also the generally more ‘western’ selection of locations that appear most similar: Germany, Paris.

In the case of Moscow, such modernisation of contexts is much less visible. A village remains as the third closest contextual neighbour, while a university is only on the 10th position, below both the word monastery as well. This highlights Moscow’s more traditional and non-modern connotations, which likely contribute to its relative decline in frequency that we reported in Figure 3.

As for Kyiv, we likely see the total change of its function in the texts. Its contextual neighbors in the romanticist corpus suggest Kyiv being the centre of historical action: the mentions of armies, taverns, gangs... In the realist corpus (where, as we remember, there is much less Kyiv, so this should be taken with a grain of salt), on the other hand, Kyiv becomes just one item in the list of many locations, which in our view is an indicator of the city losing its function as the setting of literary plots.

## 5. Conclusion and discussion

Our research is an early attempt at modeling the spatial component of Russian 19th-century prose through geocoding and mapping. Our approach obviously lacks many important nuances. For one, we do not differentiate between different functions of toponyms inside the texts, be it a random mention of a place or a location important for the development of the plot. But what we were interested in was primarily the expansion of ‘mental’ geography of Russian writers and readers. Regardless of whether a certain city or country was just ‘mentioned’ or actually was part of the plot, its appearance in the text is a clear sign that it entered the mental map of Russian literate society. Secondly, and maybe more importantly, we did not normalize locations in proportion to the size of work. A lengthy novel set in Moscow can contain hundreds of mentions of the city and will inevitably skew the whole map towards it. We intend to handle this issue in the next iterations of this work.

Despite these and other caveats, we believe that our results demonstrate the utility of the method as a tool to track large scale literary changes on relatively big corpora in the paradigm of distant reading. Most of the novels we worked with belong to the ‘great unread’ of Russian 19th-century literature. The ability to derive conclusions regarding the evolution of literature in relation to the economic, political, and cultural developments of the Russian Empire — without the necessity of reading hundreds of individual texts — presents a promising research perspective. The detection of an expansion in geographic boundaries during the second half of the 19th century through quantitative analysis further demonstrates that methods of distant reading can yield meaningful insights into literary corpora.

## 6. Data Availability 281

Data can be found here: [https://github.com/DanilSko/mapping\\_russian\\_prose/tree/main/geodata](https://github.com/DanilSko/mapping_russian_prose/tree/main/geodata) 282  
283

## 7. Software Availability 284

Software can be found here: [https://github.com/DanilSko/mapping\\_russian\\_prose/tree/main/code](https://github.com/DanilSko/mapping_russian_prose/tree/main/code) 285  
286

## 8. Author Contributions 287

**Daniil Skorinkin:** Project administration, Writing – original draft, Writing – review editing, Visualization 288  
289

**Boris Orekhov:** Conceptualization, Formal Analysis, Data curation, Investigation 290

## References 291

- Algee-Hewitt, Mark, Sarah Allison, Marissa Gemma, Ryan Heuser, Franco Moretti, and Hannah Walser (2018). *Canon/archive : large-scale dynamics in the literary field*. ISSN: 2164-1757 Issue: 11 Pages: 14 Publication Title: Stanford Literary Lab: Pamphlets ; 11 Series: Pamphlets of the Stanford Literary Lab Type: workingpaper Volume: Stanford Literary Lab. <https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf>. 292  
293  
294  
295  
296
- Barbaresi, Adrien (Mar. 15, 2018). "Toponyms as Entry Points into a Digital Edition: Mapping Die Fackel". In: *Open Information Science* 2, 23–33. 10.1515/opis-2018-0002. 297  
298  
299
- Bartholomew, J. G. (John George) (1914). *A literary & historical atlas of Europe*. In collab. with University of California Libraries. London ; Toronto : Dent ; New York : Dutton. 272 pp. <http://archive.org/details/literaryhistatlas00bartrich> (visited on 07/15/2024). 300  
301  
302  
303
- Bodenhamer, David J., John Corrigan, and Trevor M. Harris, eds. (June 28, 2010). *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Edition Unstated. Bloomington, Ind.: Indiana University Press. 203 pp. ISBN: 978-0-253-22217-6. 304  
305  
306
- Bowers, Katherine (2022). *Writing Fear: Russian Realism and the Gothic*. University of Toronto Press. ISBN: 978-1-4875-2692-4. <https://www.jstor.org/stable/10.3138/j.ctv2hvfjgr> (visited on 02/09/2025). 307  
308  
309
- Brunson, Molly (Sept. 10, 2016). *Russian Realisms: Literature and Painting, 1840–1890*. Cornell University Press. 283 pp. ISBN: 978-1-5017-5753-2. 310  
311
- Döring, Jörg (2013). "How Useful Is Thematic Cartography of Literature?" In: <https://www.semanticscholar.org/paper/How-Useful-Is-Thematic-Cartography-of-Literature-D%C3%B6ring/27e15ff9349db2aeeddfb73e59e853068633c88d> (visited on 07/12/2024). 312  
313  
314  
315



- Douglas Clayton, J. (2016). "Russian Modernism (1890–1934)". In: *Routledge Encyclopedia of Modernism*. 1st ed. London: Routledge. ISBN: 978-1-135-00035-6. 10.4324/0123456789-REM1879-1. <https://www.rem.routledge.com/articles/russian-modernism-1890-1934> (visited on 02/07/2025).
- Du, Keli, Julia Dudar, Cora Rok, and Christof Schöch (2025). *Project – Zeta and Company*. <https://zeta-project.eu/de/project/> (visited on 02/08/2025).
- Fanger, Donald (1998). *Dostoevsky and Romantic Realism: A Study of Dostoevsky in Relation to Balzac, Dickens, and Gogol*. Northwestern University Press. 332 pp. ISBN: 978-0-8101-1593-4.
- Fridlender, G. (1971). *Pojetika russkogo realizma: Ocherki o russkoj literature 19 veka*. "Nauka", Leningr. otd-nie.
- Ioffe, Dennis (2009). "The Poetics of Personal Behaviour: The Interaction of Life and Art in Russian Modernism (1890-1920)." PhD thesis. Universiteit van Amsterdam [Host].
- Kazakova, Elena (Jan. 18, 2024). *Zabytye romany russkih pisatelej iz fondov Pushkinskogo Doma (1857–1917)*. Version 2. 10.31860/openlit-2023.12-C007. <https://dataverse.pushdom.ru/dataset.xhtml?persistentId=doi:10.31860/openlit-2023.12-C007> (visited on 02/08/2025).
- Kuzmenko, E. and Boris V. Orekhov (2016). "Geography Of Russian Poetry: Countries And Cities Inside The Poetic World". In: Digital Humanities Conference. <https://www.semanticscholar.org/paper/Geography-Of-Russian-Poetry%3A-Countries-And-Cities-Kuzmenko-Orekhov/4e506e7c533f219c5ade38e6193162f6d35d2cf5> (visited on 06/19/2024).
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (Sept. 7, 2013). *Efficient Estimation of Word Representations in Vector Space*. 10.48550/arXiv.1301.3781. arXiv: 1301.3781[cs]. <http://arxiv.org/abs/1301.3781> (visited on 02/08/2025).
- Moretti, Franco (Sept. 17, 1999). *Atlas of the European Novel: 1800-1900*. Google-Books-ID: ja2MUXS\_YQUC. Verso. 232 pp. ISBN: 978-1-85984-224-9.
- natasha/natasha (July 14, 2024). original-date: 2016-08-03T09:49:51Z. <https://github.com/natasha/natasha> (visited on 07/16/2024).
- Sobchuk, Oleg and Evgenija Lekarevich (Jan. 17, 2025). *Korpus narrativnoj prozy 19 veka*. Version 4. 10.31860/openlit-2020.10-C004. <https://dataverse.pushdom.ru/dataset.xhtml?persistentId=doi:10.31860/openlit-2020.10-C004> (visited on 02/08/2025).
- Vdovin, Alexey, Margarita Vaysman, Ilya Kliger, and Kirill Ospovat (2020). "«Realizm» i russkaja literatura XIX veka". In: *Russkii realizm XIX veka: Obshchestvo, Znanie, Povestvovanie*, edited by M. Vaisman, A. Vdovin, I. Kliger, and K. Ospovat. Moscow: Novoe literaturnoe obozrenie, 431–451.
- Wilkens, Matthew, Elizabeth F. Evans, Sandeep Soni, David Bamman, and Andrew Piper (Sept. 26, 2024). "Small Worlds: Measuring the Mobility of Characters in English-Language Fiction". In: *Journal of Computational Literary Studies* 3.1. Number: 1 Publisher: Universitäts- und Landesbibliothek Darmstadt. ISSN: 2940-1348. 10.48694/jcls.3917. <https://jcls.io/article/id/3917/> (visited on 02/08/2025).