Sequencing Literary Reception: Trajectories of Russian Novels over the Afterlife Course

Alexander Kim

School of Philological Studies, Higher School of Economics, Russia

Abstract

The study explores literary recognition as a continuously generating new results process, as opposed to static models. Introducing sequence analysis to cultural analytics, it examines how literary trajectories of Russian pre-revolutionary novels unfold over time (1919-2022). After analyzing variations in publishing trajectories among school and non-school clusters, the paper explores the relationship between school and reprint domains. The results support the idea that there is a positive spillover effect from being included in the school curriculum, as it is an advantage for a long-term publishing career. Conversely, they do not support the idea of a negative spillover effect, as being excluded from the curriculum does not seem to be associated with a deterioration in publishing trajectories. Except for school ones, there is a group of novels characterized as the "second canon," which have survived on the book market for more than a century.

Keywords

literary reception, canon/archive, cultural success, sequence analysis

1. Introduction

Why do some cultural objects receive recognition while others do not? Since Bourdieu [8], the most common approach to this problem is to measure the impact of different institutions and social properties. In the case of literature, these include prizes, publishing houses, artistic networks, socioeconomic background, gender and racial/ethnic identities of writers [33, 13, 11, 19]. Recent computational literary studies have introduced new approaches that allow researchers to identify textual patterns that contribute to the literary distinction [6, 36, 9, 7].

However, assessing both social and linguistic ingredients of literary success is challenging for the analysis of century-spanning reception. Most of the studies outlined above focus primarily on short-term recognition, rather than dynamics of long-lasting reception. Furthermore, if we are to sharpen the argument, regression-based research deals with the "timeless present," failing to acknowledge the temporality of social processes [1]. As Abbott pointed out, the question of success often implies a "final outcome" paradigm, which treats the research object or its state as "the state of the trajectory at its end" [1]. Literary recognition is nothing but an "outcome-at-a-point." Rather than asking what predictors of success are, it is worth thinking about the reception of cultural objects as a "process outcome," which continuously generates new results.
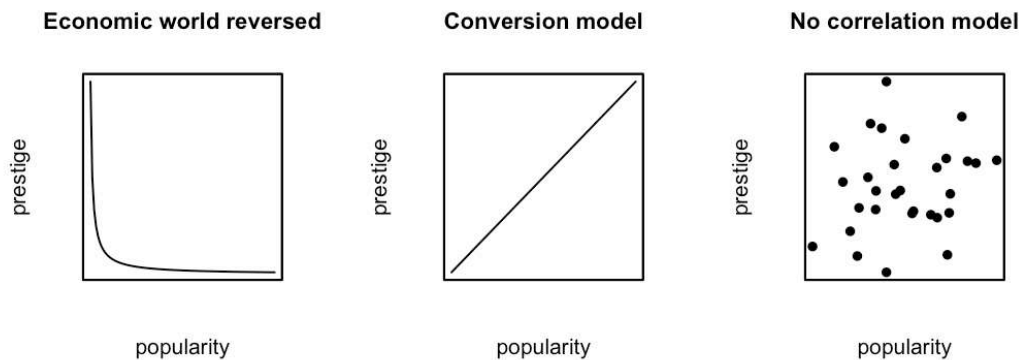
Figure 1: Ideal-typical models of the interplay between popularity and prestige.

One possible strategy for doing this can be found in the two pamphlets from Stanford Literary Lab [6, 28]. In order to test Bourdieu's idea of the literary field as an "economic world reversed," scholars charted anglophone fiction on the axes of "prestige" and "popularity," serving as proxies for economic and symbolic capitals. Algee-Hewitt and his co-authors used the nineteenth-century data on reprints and translations of novels (popularity) and the number of mentions of these works in the MLA Bibliography and the length of the Dictionary of National Biography entries (prestige) to model the "British novelistic field" of 1770-1830. Porter took a similar approach to anglophone writers of different genres, but he measured their popularity based on Goodreads ratings. What did they expect to find and what did they actually find? Consider the ideal-typical models of reception in Figure 1.

The first model, "economic world reversed," is what Stanford Lab scholars had expected, drawing on Bourdieu's works. The relationship between popularity and prestige can be described as a reciprocal function. This model denies the conversion of capitals: bestsellers fall into academic obscurity, while highbrow literature does not provide commercial opportunities. The second model, the conversion one, has two variations: prestige (artistic consecration) leads to popularity (market success) or vice versa. I have represented this using an identity function, but it can also be interpreted as a different monotonically increasing function. In the third model, the two variables do not correlate at all. This is approximately what the Stanford team found in their charts.

Since scholars have relied on belated data (what is the time lag between the writing of Daniel Defoe and MLA articles?), their research is anachronistic and certainly misleading when it comes to testing Bourdieu's theory, which emphasizes synchronic rather than diachronic relations within a field (Baudelaire was opposed to his contemporaries, not to authors from earlier or later periods). However, the Stanford analytical strategy could be easily improved by providing non-anachronistic data and charting a series of "snapshots" of the field of reception throughout time (see an example in [35]). This approach allows us to examine the structural dynamics of the field. Charting snapshots is a useful but not subtle strategy for grasping the

processual nature of literary recognition. Modeling à la Bourdieu, whether through popularity/prestige graphs or multiple correspondence analysis, focuses on the group-level oppositions and distribution of capitals but fails to capture individual trajectories of capital accumulation and conversion over time [32].

In their research on the reception of French sociologists in the United States, Ollion and Abbott considered one level of their analysis to be the study of authors' "afterlives" [27]. Taking this metaphor seriously, one can examine the literary reception in the same way as life-course scholars do using sequence analysis (SA). The standard approach in this research field is to construct sequences of successive lived states across different domains, such as employment, education, and marital status, in order to build a typology of life-course trajectories and quantitatively characterize them [5, 18]. For example, in their study of work-life trajectories in the US and Germany, Aisenbrey and Fasang found that in both countries, individuals with stable, high-prestige careers had two or more children, while those with stable, medium-prestige careers remained childless until their forties [5].

Within a bourdieusian framework, Rossier applied SA to examine how the academic paths to a professor position were structured in the Swiss economics field [32]. Similarly, SA can be a useful tool for studying artistic careers. In their research on hierarchies within the field of contemporary French poetry, Dubois and François found that poets' publishing trajectories were highly irregular, and their successive decisions to collaborate with high- or low-status publishers had little impact on their dominant or dominated positions within the field [13]. Using data from art history textbooks, Accominotti explored how the number of illustrations dedicated to each five-year period of painters' professional trajectories was distributed over their life course, constructing a typology of "creativity careers." [4]. However, it is not just writers and painters who have distinct career paths; the works themselves follow "career" trajectories that can outlast their creators. Thus, it is surprising that SA remains largely unknown in the sociology of cultural reception and, more broadly, in cultural analytics, especially considering the suitability of longitudinal historical data on cultural reception for sequence-based approaches.[1]

Instead of examining the timeless causal effects of different factors on literary recognition, considering literary reception as a career or a life-course sequence provides an opportunity to uncover how the pathways of literary works jointly unfold over time. The main problem here is data availability. If data on reprints is relatively easy to gather, data on "prestige" across centuries is a needle in the haystack. Although academic citations and critics' reviews contribute to the artistic consecration, school is the major social institution that preserves wide and long-lasting literary recognition [17]. This is especially true for countries with a unified education system. This study employs data on school curricula in the Soviet Union and post-Soviet Russia to quantify the prestige of pre-revolutionary Russian novels given by the state. As I will demonstrate, even though official circulation of books in the Soviet command economy can be considered as another indicator of the government approval, publishing trajectories can not be reduced to the state apparatus even in the Soviet Union, let alone the post-Soviet book market.

Taking into account the school and reprint domains, I ask: how do the trajectories of school

---

[1]In a somewhat ironic twist, Abbott introduced SA from biology into sociology as a way to strengthen historical research [2, 3], however, SA is now fully integrated into more or less presentist life-course scholarship.

and non-school novels differ in the long term? It is clear that the presence of a novel in school is a strong incentive for publishers to print it. However, what happens after a text is excluded from the curriculum? Is the school advantage necessary for long-term publishing survival?

As in popularity/prestige studies, the analysis presented here focuses on the interplay of cultural and economic capitals, but in a diachronic perspective. The models of reception outlined in Figure 1 are consistent with those developed by scholars while studying the relationships between life-course domains. In addition to work-family independence, Fasang and Aisenbrey distinguished three types of work-family interdependence: mutual support, competing alternatives, and instability spillovers [14]. While mutual support relationships correspond to the conversion model ("activities in one life domain can generate resources that open up opportunities in the other life domain"), competing alternatives align with the "economic world reversed" model, suggesting a zero-sum situation. Instability spillovers between domains take place when deteriorating trajectories mutually constrain each other. For example, job loss may lead to separation, which in turn limits economic resources.

In the case of school-reprints interrelations, I expect a one-way dependence of reprints on the presence in school. Being integrated into the curriculum is an advantage for a successful publishing career, but not vice versa. The same can be true in the case of negative spillover: exclusion from the curriculum can be associated with deteriorating publishing trajectories. I will address the question of school-reprint (inter)dependence by clustering novelistic sequences across domains and measuring their integrative capability.

2. Data and methods

2.1. Data

In this study, I depart from a bibliographic database of 2026 Russian novels published before the Russian Revolution of 1917, which was collected by the HSE project "Russian Novels." The dataset was manually compiled on the basis of six volumes of the dictionary "Russian Writers 1800-1917," "Dictionary of Eighteenth Century Writers" and other bibliographic sources. This is the most comprehensive list of novels written during the Russian Empire period.

To enrich the database with "popularity" measurements, I scraped data on reprints of these novels from the Digital Catalogue of the Russian State Library (RSL). There were three main criteria for excluding reprints: 1) the inability to automatically identify the novel from the book title (e.g. "The Selected Works of Leo Tolstoy"), 2) the lack of information on the publication year, and 3) the publication place is not the Russian Empire, the Soviet Union or Russia, depending on the historical period (RSL has highly incomplete data on translations).

Moving on to the prestige measurements, the two datasets on Russian school curricula from the last 100 years, available in the Repository of Open Data on Russian Literature and Folklore [21, 20], are exceptional sources for measuring the extent to which different novels were consecrated by the state throughout history. As both datasets include information about the curricula of secondary and high schools, they can be compared. However, there is a problem of incomplete data. In the case of post-Soviet Russia, reading lists are only available from 1998 onwards, which is the year when the first official curriculum was published after the collapse of the Soviet Union. The database for the Soviet school covers the period from 1919 to 1991,
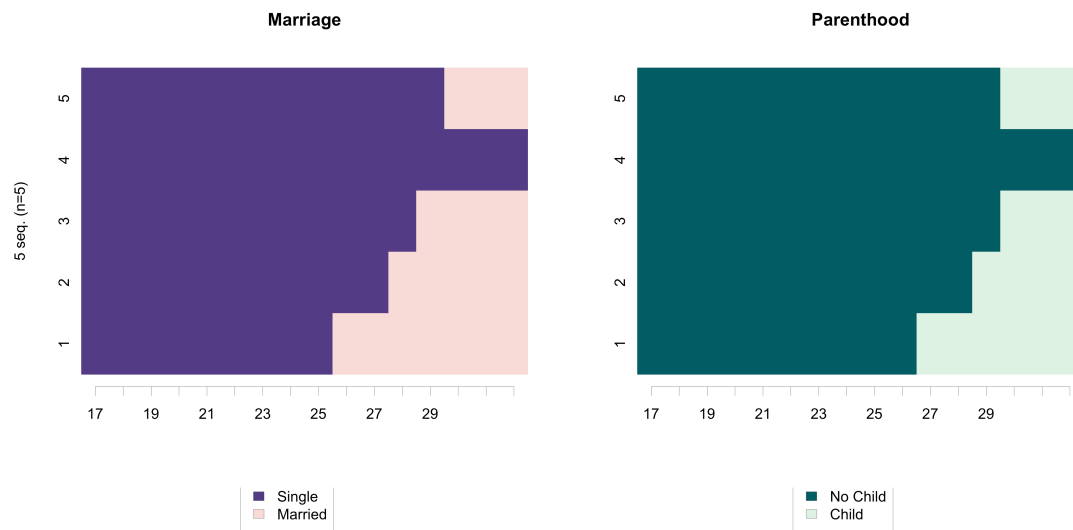
Figure 2: Toy example of an MD sequence index plot.

with an absence of curricula for 23 years, particularly during the 1920s and 1930s, which were decades of great political turbulence. For details on how curricula changed over time, see Appendix A.

2.2. Sequence analysis

The first step in performing a simple SA is to define state tokens in a domain, for example, "single" and "married." Rather than creating an extended alphabet of combined state tokens when working with multiple domains (e.g. "single, no child," "single, 1 child"), recent studies have tended to construct sequences for each domain separately and then apply different strategies to build a joint typology [31]. This approach has various labels, such as multidomain, multichannel, multidimensional, and joint SA. Out of necessity to choose one, I will call it multidomain (MD). A toy example of an MD sequence plot can be seen in Figure 2. Here, each horizontal line represents an individual's trajectory within a given domain (marriage or parenthood). For example, individual 1 married at the age of 25 and became a parent at the age of 26.

In this study, the "afterlife" course of novels has two distinct domains: reprints and school curriculum. Each comprises three state tokens: "no reprints," "1 reprint," "1+ reprints" (more than one) in the first case, and "curricula," "no curricula," and "no data" in the second case. Each position in a sequence is defined by a one-year time interval from 1919 to 2022. The "no data" token is used in the absence of data on the curriculum in a given year (see the previous section).

The common way to obtain MD clusters is to compute pairwise MD distances. As Ritschard and colleagues have shown, different methods of deriving distances have their own implicit assumptions about the association between domains [30]. The nature of the data I use requires

a method that is not prone to any independence presuppositions. Although researchers suggest using distances independent from domain costs and distances (IDCD), it is also possible to construct a MD typology through an independent analysis of each domain. The advantage of this strategy is that it "ensures coherence between MD and domain types by construction" [30]. It also allows me to control for groups of school and non-school novels, which is crucial to my research question.

Thus, I follow two steps to obtain MD typology. The first one is clustering within the school domain using the partition around medoids algorithm (PAM) [34]. Sequences are compared to each other based on the optimal matching distance with INDELSLOG costs. The optimal number of groups is determined using various quality measures (see Appendix B). The second step is the following: after getting the four-cluster solution, I take the non-school group and cluster it in the same way within the reprint domain. The whole procedure results in a six-cluster solution of three school and three non-school groups.

2.3. Integrative potential

To quantitatively assess the behavior of individual sequences, SA scholars have developed various metrics that measure diversity, complexity, as well as desirability and undesirability of the sequence [30]. Considering the question of processual literary success, it is of interest to measure the long-term publishing survival of the novels. If a text X remained in obscurity for, say, a century but was later resurrected and reprinted frequently over a short period, this does not represent the long-term publishing survival. I define the latter as a text being reprinted in a more evenly distributed manner over time. To measure this, I use the integrative potential proposed by Brzinsky-Fay [10] as a proxy.

Unlike many other metrics, such as turbulence, complexity or entropy, this indicator takes into account the nature of the states by dividing them into positive and negative ones and calculates a tendency to reach and end in a positive state ("1 reprint" and "1+ reprints" for publishing sequences, "curricula" for school ones). The formula is as follows [30]:

$$Integr = \frac{\sum_{i=1}^{l} is.pos(x_i)i^{\omega}}{\sum_{i=1}^{l} i^{\omega}},$$

where $is.pos(x_i)$ is a logical function that equals 1, if ith token in the sequence is positive, and ω is a power parameter that controls the importance of the last tokens. I apply the default TraMineR value of 1.

In this study, the integrative potential measures the tendency of a novel to either be reprinted or included in school curricula, and then remain in these "positive" states. The metric is calculated separately for each domain, allowing us to assess the relationship between long-term school and publishing survival.

All SA calculations and visualizations are made using the TraMineR and WeightedCluster R packages [16, 34]. Data and code are available at https://github.com/kimchs/sequencing-reception.
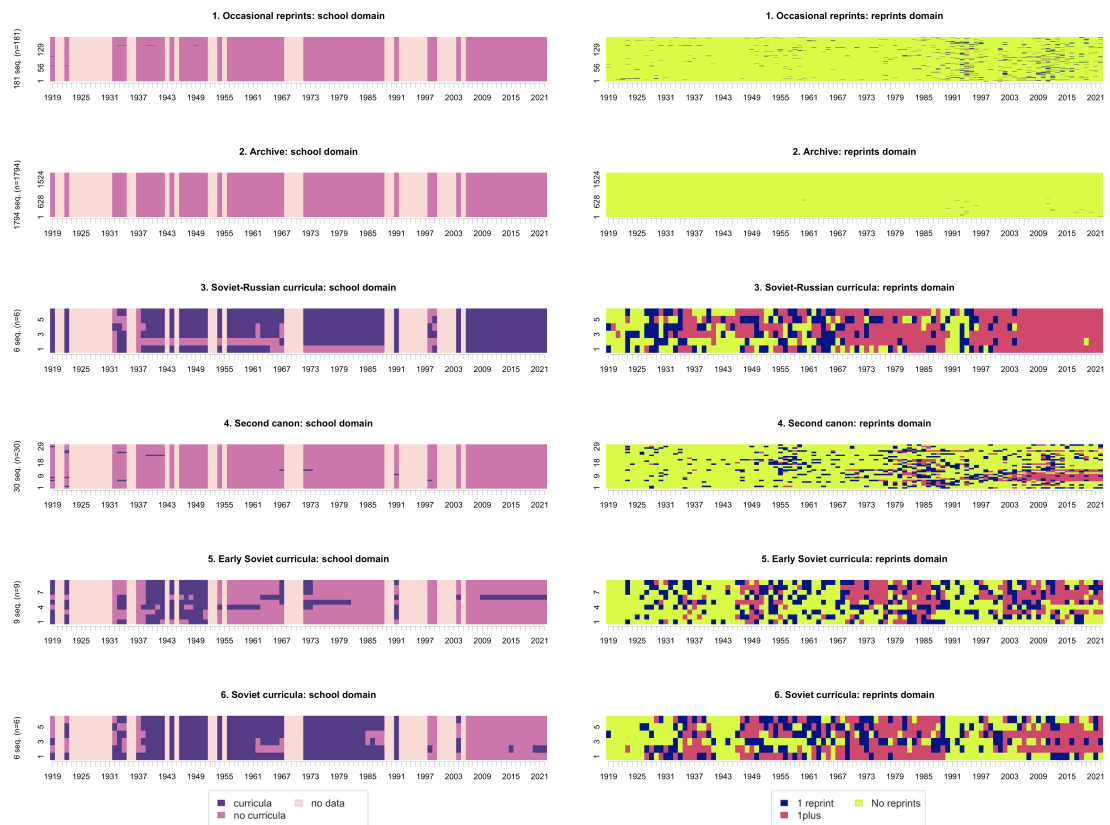
Figure 3: Sequence index plots of five school-reprints clusters.

3. Results

3.1. Typology of novelistic trajectories

Figure 3 presents a MD typology of novelistic trajectories. For both domains, horizontal lines at the same level plot trajectories of the same novel. As it was designed, there are clusters of school novels (clusters 3, 5, and 6) and non-school novels (clusters 1, 2, and 4). The way in which the former are divided shows the process of shrinking of the novelistic school canon. There were two main events after which school novels were cut off: the death of Stalin and the fall of the Soviet Union. I will distinguish these groups as early Soviet (cluster 5), Soviet (cluster 6), and Soviet-Russian novels (cluster 3).

Cluster 6 comprises novels that were at the core of Soviet curricula, but only Soviet. It includes Maxim Gorky's revolutionary novel Mother (1907), Nikolay Chernyshevsky's political fiction What Is to Be Done? (1863), which was banned in the Russian Empire for four decades, and Alexander Radishchev's eighteenth-century political travelog A Journey from St. Petersburg to Moscow (1790), for which the author was exiled to Siberia. The other three were Leo Tolstoy's Anna Karenina and Resurrection and Pushkin's Dubrovsky. In post-Soviet Russia, the

988

curriculum was clearly derevolutionized. While Radishchev and Chernyshevsky disappeared altogether, Gorky, Pushkin, and Tolstoy remained at the core of new reading lists, but in the form of other texts.

The early-Soviet novels in cluster 5 were written by almost the same authors as those in the Soviet-Russian and Soviet clusters. Goncharov, Turgenev, Gorky, Saltykov-Shchedrin, and Herzen were deeply embedded in the Soviet curricula, so the exclusion of some of their novels can be explained simply by the limited size of the reading lists. Interestingly, there are no pre-revolutionary novels that have been "resurrected" in the post-Soviet Russian school, except for Saltykov-Shchedrin's The History of a Town from cluster 5. The list of the most successful novels (cluster 3) results from a shrinkage or "optimization" of the Soviet legacy. Each novel in this group represents only one author. If, for example, school students in the Soviet Union were required to read both Tolstoy's Anna Karenina and War and Peace, then in the post-Soviet Russian school curriculum, only War and Peace remains.

All school novels were more or less widely reprinted through Soviet history and survived on the Russian book market, even if they disappeared from curricula in the second half of the twentieth century. In contrast, non-school novels did not receive a significant number of reprints. This cluster represents what is often referred to as the literary "archive" — a collection of published texts that have been preserved but have not garnered significant attention over time [6]. Cluster 1 is characterized by occasional reprints.

Cluster 4 shows trajectories that differ from all those described above. It represents non-school novels that nevertheless had a publishing career in the Soviet Union and subsequently in post-Soviet Russia. In the 1980s, there was a growing interest in these texts, which has persisted to the present day. With the exception of school novels, which pre-revolutionary, "imperial" novels could enjoy relative popularity within the Soviet command economy? It is plausible to suggest that these would primarily be books written by canonical school authors. However, only six out of thirty novels formed such an extended curriculum. The others were written by Dostoevsky (8), then Mamin-Sibiryak (3), Danilevsky (3), Zagoskin (2), Lazhechnikov (2), Melnikov-Pechersky (2), Bely, Sologub, Stepnyak-Kravchinsky, and A. K. Tolstoy.

Except for Mamin-Sibiryak, who was a well-known children's writer [25], none of these authors fully received the Soviet state's ideological approval. As a Christian and blatantly conservative author, Dostoevsky was considered as a "reactionary." The Soviet school tried from time to time to appropriate his legacy and, finally, "rehabilitated" Crime and Punishment in the late 1960s. Despite this, his novels, including even anti-revolutionary Demons, were reprinted throughout Soviet history. Along with authors of popular fiction and modernist novels, Dostoevsky constitutes, so to speak, the "second canon" in the Soviet Union (popularity without governmental ideological approval). As Dobrenko demonstrated, although the early Soviet authorities struggled against the "bad tastes" and attempted to "clean" libraries from the mass literature à la Zagoskin and Lazhechnikov, readers continued to demand removed authors, and their demands "had to be satisfied somehow" [12]. This was not only the case for popular fiction, but also for modernist and other "bourgeois" literature. In this context, the "second canon" can be seen as a market sector of the Soviet publishing economy. The authorities reprinted the most read non-school novels, which could also be justified by economic considerations [2].

---

[2]It is worth noting that the wider circulation of the second canon coincided with the advent of Perestroika (Restruc-

3.2. Behavior of publishing trajectories

While trajectories within the school domain are easy to interpret, it is of interest to provide quantitative metrics to understand how publishing trajectories differ across clusters and within them. Looking at Figure 3, it can be posited that regardless of the curricula spell duration, the presence in school is a key factor in ensuring a prolonged publication lifespan. The analysis presented in this section is not sufficient to prove the causal statement, however, it provides support for the hypothesis.

Figure 4 displays the integrative potential (Integr) computed for each school and reprint sequences. This index measures a novel's tendency to reach a positive state (being included in school curricula or reprinted) and remain in that state. I use it as a proxy for long-term literary survival.

Although school integrity is not a prestige metric in the sense that it was operationalized in Stanford Literary Lab's prior studies, the graph can still be compared with those in Figure 1. Figure 4 illustrates a one-sided conversion model or a positive spillover effect. The greater the degree of integration of a given novel within the school curriculum throughout Soviet and Russian history, the more it was reprinted in the long term. In the absence of instances where "market" popularity drove its inclusion in the school curriculum, the relationship between domains is asymmetrical.

In order to assess how the behavior of publishing trajectories change depending on the school factor, I computed the integrative potential of reprint sequences for each cluster both before and after the fall of the Soviet Union. Given that clusters 3 and 6 both contain Soviet novelistic canon, yet their respective school trajectories diverge after 1991, it is especially of interest to compare their reprint characteristics (see Figure 5).

In relation to the Soviet period, Integr values in reprints domain correspond to the degree novels were integrated in the school curricula. Clusters 3 and 6 show almost the same values, followed by the group of early Soviet curricula novels and then non-school novels. How has this changed in contemporary Russia? The Soviet-Russian novelistic canon has become deeply integrated into the book market, with the Integr variance being extremely low. While the mean values of the early Soviet and Soviet curricula clusters did not change much, indicating that in the absence of school support, their publishing trajectories were shaped by conservative (or inertial) forces, there was a notable increase in variance. Non-school novels demonstrate remarkable alterations in their levels of integrative capability.

4. Discussion

The interplay of school and publishing trajectories can be described through a positive spillover effect. A stable school "career" is associated with a long-lasting publication life. However, the analysis presented in this paper does not provide evidence to support the negative spillover effect. When a novel is excluded from the school curriculum, it is likely to remain visible on the book market, with the range of possible publishing trajectories expanding. Furthermore,

---

turing). Drawing on the description of the literary landscape of the 1970-80s as a period of the rise of conservative and nationalistic forces [23, 29], it could be argued that one consequence of this process was the increasing popularity of Dostoevsky and historical fiction writers (A. K. Tolstoy, Danilevsky, Lazhechnikov, Zagoskin).
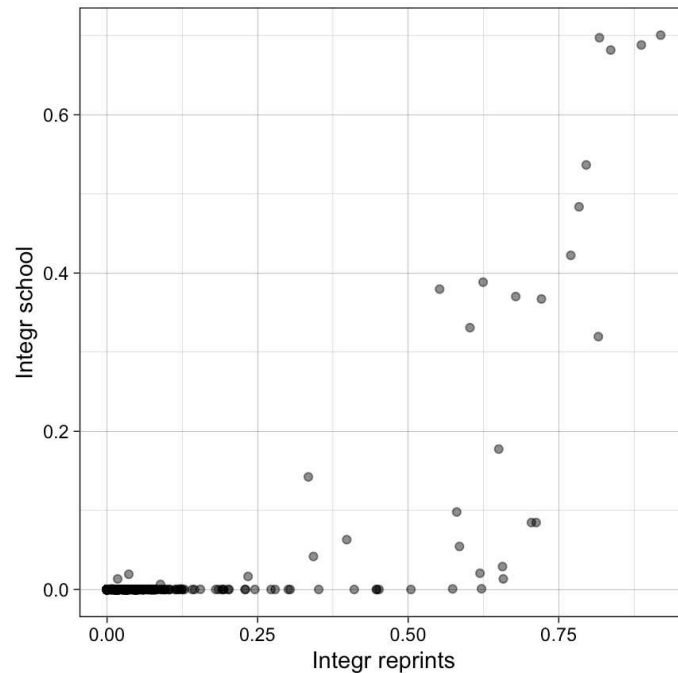
Figure 4: Integrative potential of novels in school and reprints domains (1919-2022).

the school advantage is not a prerequisite for long-term publishing survival, as evidenced by the existence of a group of non-school novels that were reprinted throughout Soviet and post-Soviet Russian history. I proposed the "second canon" label for these novels.

The persistence of the "second canon" and ex-school novels in the book market could be explained by publishing inertia. Once a text surpasses a certain threshold, say, 10 years of continuous reprints, a self-reinforcing process takes hold. The inertia leads to more frequent and longer-lasting reprints compared to texts that have not crossed a threshold. This hypothesis requires thorough testing and further exploration of alternative models.

As for school trajectories alone, with a few exceptions, both Soviet and post-Soviet Russian curricula do not give pre-revolutionary novels any chance of "resurrection." The novelistic canon is not being updated, it is simply shrinking. However, we take into account all texts ever included in either the Soviet or post-Soviet Russian curricula and cluster them, we would get a group of texts (N=179), mostly poems, that only appeared on school reading lists after the fall of the Soviet Union (see Appendix C). Apart from extensions of nineteenth-century classics, these are texts by modernist poets (Akhmatova, Bunin, Mandelstam, Pasternak, Tsvetaeva) who were considered ideologically inappropriate. Along with Dostoevsky, they form a core of post-Soviet "resurrections."

Although the shrinkage of the novelistic canon is certainly connected to context-based biases, such as organizational and ideological decisions, it can also be driven by the short-list effect that is "the tendency for selection based on content to be stronger when an agent has
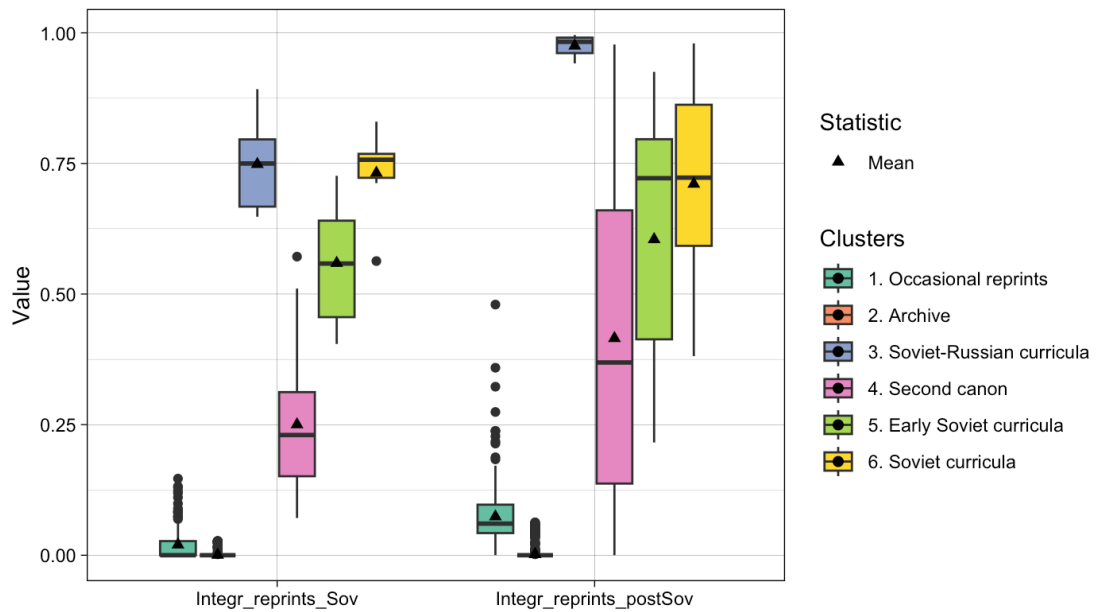
Figure 5: Integrative potential of reprint sequences in each cluster over two periods: 1919-1991 and 1992-2022.

fewer resources to devote to the acquisition of cultural items" [26]. While the number of all texts ever written continues to grow exponentially, the "nestedness" of the school curriculum may increase as well. Future research should decouple these content- and context-based selection mechanisms. However, this is not possible without a reliable metric of appeal, which is difficult to establish in the case of literature [15].

Despite its maturity, or at least "young-adulthood" [22], SA is still almost neglected in studies of cultural reception and diffusion. I have provided an example of exploratory research on literary trajectories within this methodological framework. There are several ways in which it can be deepened, ranging from data enrichment to combining it with causal and generative inference. In any case, the study of "afterlives" using SA has the potential to advance both theoretical and empirical arguments in cultural analytics.

Acknowledgments

References

[1] A. Abbott. *Processual Sociology.* Chicago, London: University of Chicago Press, 2016.

[2] A. Abbott. "Sequences of Social Events: Concepts and Methods for the Analysis of Order in Social Processes". In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 16.4 (1983), pp. 129–147. DOI: 10.1080/01615440.1983.10594107.

[3] A. Abbott and J. Forrest. "Optimal Matching Methods for Historical Sequences". In: *The Journal of Interdisciplinary History* 16.3 (1986), pp. 471–494. DOI: 10.2307/204500.

[4] F. Accominotti. "Creativity from interaction: Artistic movements and the creativity careers of modern painters". In: *Poetics* 37.3 (2009), pp. 267–294. DOI: 10.1016/j.poetic.2009.03.005.

[5] S. Aisenbrey and A. Fasang. "The interplay of work and family trajectories over the life course: Germany and the United States in comparison". In: *American Journal of Sociology* 122.5 (2017), pp. 1448–1484. DOI: 10.1086/691128.

[6] M. Algee-Hewitt et al. *Canon/Archive: Large-scale Dynamics in the Literary Field.* Stanford California, 2016.

[7] J. Barré et al. "Operationalizing Canonicity: A Quantitative Study of French 19th and 20th Century Literature". In: *Journal of Cultural Analytics* 8.3 (2023).

[8] P. Bourdieu. *The Rules of Art: Genesis and Structure of the Literary Field.* Trans. by S. Emanuel. Stanford, California: Stanford University Press, 1996.

[9] J. Brottrager et al. "Modeling and Predicting Literary Reception. A Data-Rich Approach to Literary Historical Reception". In: *Journal of Computational Literary Studies* 1.1 (2022).

[10] C. Brzinsky-Fay. "Lost in Transition? Labour Market Entry Sequences of School Leavers in Europe". In: *European Sociological Review* 23.4 (2007).

[11] C. Childress, C. M. Rawlings, and B. Moeran. "Publishers, authors, and texts: The process of cultural consecration in prize evaluation". In: *Poetics* 60 (2017).

[12] E. Dobrenko. *The Making of the State Reader: Social and Aesthetic Contexts of the Reception of Soviet Literature.* Trans. by J. M. Savage. Stanford University Press, 1997.

[13] S. Dubois and P. François. "Career paths and hierarchies in the pure pole of the literary field: The case of contemporary poetry". In: *Poetics* 41.5 (2013).

[14] A. E. Fasang and S. Aisenbrey. "Uncovering Social Stratification: Intersectional Inequalities in Work and Family Life Courses by Gender and Race". In: *Social Forces* 101.2 (2022), pp. 575–605. DOI: 10.1093/sf/soab151.

[15] P. Feldkamp, Y. Bizzoni, I. M. S. Lassen, M. R. Thomsen, and K. L. Nielbo. *Measuring Literary Quality. Proxies and Perspectives.* en. Tech. rep. 1. Darmstadt, May 2024, 27 Seiten. DOI: https://doi.org/10.26083/tuprints-00027391. URL: http://tuprints.ulb.tu-darmstadt.de/27391/.

[16] A. Gabadinho, G. Ritschard, N. Müller, and M. Studer. "Analyzing and visualizing state sequences in R with TraMineR". In: *Journal of Statistical Software* 40.4 (2011). DOI: 10.18637/jss.v040.i04.

[17]   J. Guillory. *Cultural Capital: The problem of literary canon formation.* University of Chicago Press, 1993.

[18]   M. Jalovaara and A. E. Fasang. "Family Life Courses, Gender, and Mid-Life Earnings". In: *European Sociological Review* 36.2 (2020). DOI: 10.1093/esr/jcz057.

[19]   K. W. Kim and P. Chong. "Writing by women or for women? Either way, You're less likely to be reviewed". In: *Poetics* 96 (2023).

[20]   A. Kokorin. *Literaturnyye proizvedeniya v gosudarstvennykh standartakh i programmakh dlya sredney shkoly 1998–2022 gg.* Repository of Open Data on Russian Literature and Folklore. 2023. DOI: 10.31860/openlit-2022.12-B009.

[21]   M. Kondra and E. Kazakova. *Programmy po literature dlya sredney shkoly s 1919 po 1991 gg.* Repository of Open Data on Russian Literature and Folklore. 2022. DOI: 10.31860/openlit-2022.9-B006.

[22]   T. F. Liao, D. Bolano, C. Brzinsky-Fay, B. Cornwell, A. E. Fasang, S. Helske, R. Piccarreta, M. Raab, G. Ritschard, E. Struffolino, and M. Studer. "Sequence Analysis: Its Past, Present, and Future". In: *Social Science Research* 107 (2022), p. 102772. DOI: 10.1016/j.ssresearch.2022.102772.

[23]   M. Lipovetsky and M. Berg. "Literary Criticism of the Long 1970s and the Fate of Soviet Liberalism". In: *A History of Russian Literary Theory and Criticism: The Soviet Age and Beyond.* Ed. by E. Dobrenko and G. Tihanov. University of Pittsburgh Press, 2011, pp. 207–229. DOI: 10.2307/j.ctt5hjn1z.15.

[24]   O. Malinovskaya. "Reading Russian Classics in Soviet Schools under Stalin: Analyzing normative material, 1922-1941". In: *Reading Russia. A History of Reading in Modern Russia, vol. 3.* 2020.

[25]   S. Maslinskaya. "Pochemu D. N. Mamin-Sibiryak stal glavnym detskim pisatelem (k probleme stanovleniya kanona russkoj detskoj literatury)". In: *Detskie chteniia* 15.1 (2019), pp. 9–25. DOI: 10.31860/2304-5817-2019-1-15-9-25.

[26]   O. Morin and O. Sobchuk. "The Shortlist Effect: Nestedness Contributions as a Tool to Explain Cultural Success". In: *Evolutionary Human Sciences* 3 (2021), e51. DOI: 10.1017/ehs.2021.48.

[27]   E. Ollion and A. Abbott. "French Connections: The Reception of French Sociologists in the USA (1970-2012)". In: *European Journal of Sociology / Archives Europ'eennes de Sociologie* 57.2 (2016). DOI: 10.1017/S0003975616000126.

[28]   J. D. Porter. *Popularity/Prestige.* Pamphlet 17. Stanford California, 2018.

[29]   A. Razuvalova. *Pisateli-derevenshchiki: literatura i konservativnaia ideologiia 1970-kh godov.* Moskva: Novoe Literaturnoe Obozrenie, 2015.

[30]   G. Ritschard. "Measuring the Nature of Individual Sequences". In: *Sociological Methods and Research* 52.4 (2023), pp. 2016–2049. DOI: 10.1177/00491241211036156.

[31]   G. Ritschard, T. F. Liao, and E. Struffolino. "Strategies for Multidomain Sequence Analysis in Social Research". In: *Sociological Methodology* 53.2 (2023), pp. 288–322. DOI: 10.1177/00811750231163833.
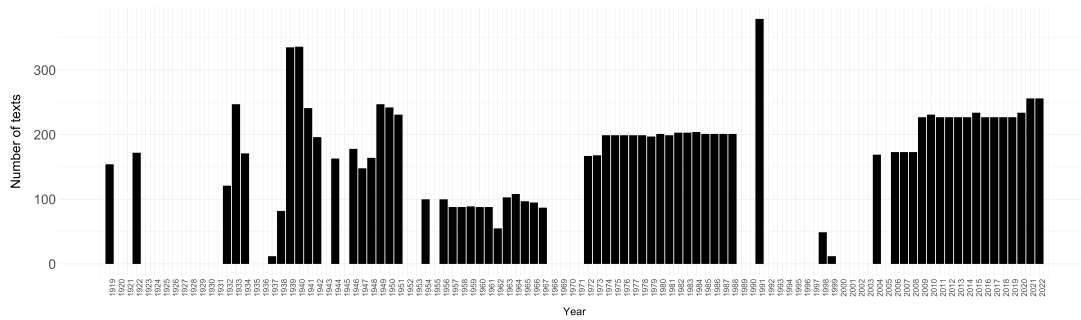
Figure 6: The number of school texts from 1919 to 2022.

[32] T. Rossier. "Prosopography, Networks, Life Course Sequences, and so on: Quantifying with or beyond Bourdieu?" In: *Bulletin of Sociological Methodology* 144.1 (2019). DOI: 10 .1177/0759106319880148.

[33] G. Sapiro. "The Structure of the French Literary Field During the German Occupation (1940-1944): A Multiple Correspondence Analysis". In: *Poetics* 30.5 (2002).

[34] M. Studer. *WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R.* LIVES Working Papers 24. Switzerland, 2013. DOI: 10.12682/lives.2296-1658.2013.24.

[35] T. Underwood. "Distant Horizons: Digital Evidence and Literary Change". In: University of Chicago Press, 2019. Chap. The Long Arc of Prestige.

[36] T. Underwood and J. Sellers. "The Longue Durée of Literary Prestige". In: *Modern Language Quarterly* 77.3 (2016).

Appendices

A. Dynamics of school curricula

The data on Soviet and post-Soviet Russian curricula are available in the Repository of Open Data on Russian Literature and Folklore [21, 20]. Figure 6 shows how the size of literary curricula (i.e., the number of included texts) has changed over time (values of 0 indicate years for which data is missing).

The number of texts in curricula varied significantly across different periods, reflecting broader changes in the educational system. To more precisely measure the similarity between lists of texts, I apply the overlap coefficient. Unlike the Jaccard index, the overlap coefficient is not prone to differences in set sizes. For example, if curriculum A is half the size of curriculum B, but all texts from A are included in B, the Jaccard index would be 0.5. The overlap coefficient, by contrast, identifies the common core of curricula, regardless of their size differences. It is calculated as follows:

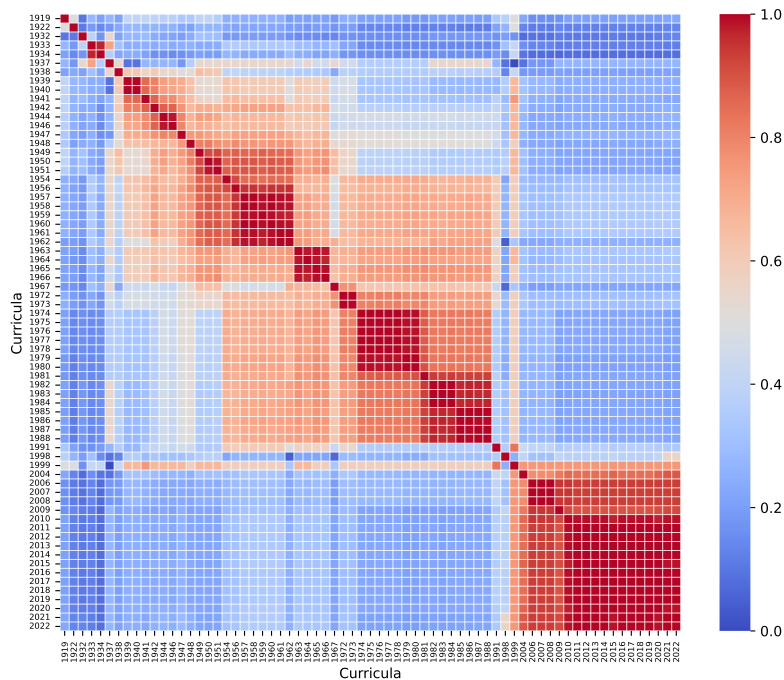$$\text{overlap}(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}.$$

995

Figure 7: The overlap coefficient between curricula (comparison by texts).

In the hypothetical example above, where curriculum A is a subset of curriculum B, the overlap coefficient is 1. I apply this similarity metric to all possible pairs of curricula and plot the results in Figure 7.

The heatmap can be read as a history of canon formation and maintenance. Prior to 1939, Soviet literary curricula had little in common with each other and with the subsequent ones. As historians have shown, the pre-World War II period saw significant turbulence in school organization [12, 24]. It was a time of "methodological" wars: from flexible to aesthetical approaches, from class-based to "individual self" ones. Stalin's era was characterized by a gradual canonization and subsequent solidification of the so-called "literary generals." The process of canon building started at least in the early 1930s, with Maxim Gorky's *The Mother* being included in curricula every year and being widely published.

From 1939 to 1966, there was a notable increase in the degree of overlap between curricula, reflecting the ossification of the canon. The reading lists from 1956 to 1961 were identical, which contrasts with the common perception of the Khrushchev Thaw as a period of de-Stalinization. It appears that the cultural and political shifts of the time had little impact on the content of school curricula. The next period of similarity lasted from 1972 to 1988. The "period of developed socialism," also known as the "Era of Stagnation," coincided with two decades of curricula stability. While the curricula from the first half of the 20th century differ significantly from
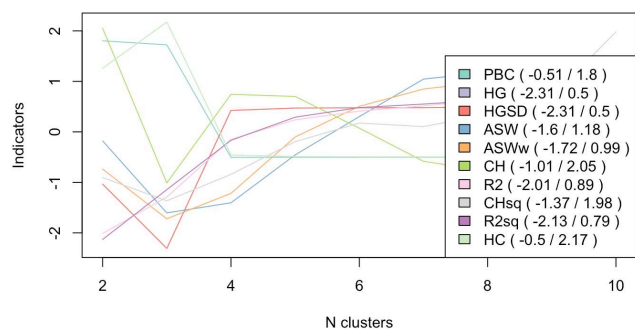
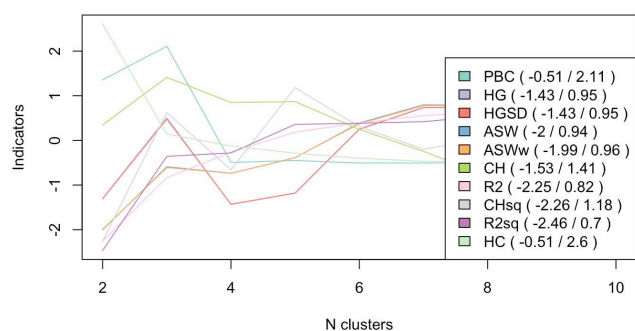**Figure 8:** Measures of the quality of partition for school sequences.

Figure 9: Measures of the quality of partition for non-school sequences.

those of the final two decades of Soviet history, the Thaw-era curricula serve as a bridge between them.

Since 2006, school curricula in contemporary Russia have been either identical or subsets of each other, reflecting the standardization and unification of education under Putin's authoritarian regime.

B. The number of clusters

Figures 8 and 9 illustrate the partition quality metrics for the school and reprint domains, respectively. For the school domain, the optimal number of clusters is 4, while for the reprint domain, it is 3.
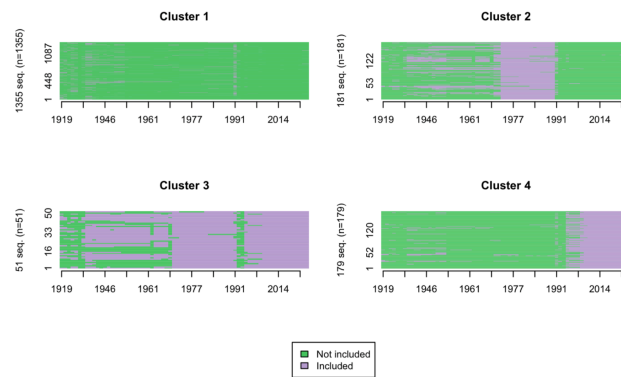
Figure 10: Sequence index plots of four school clusters computed for all texts that have ever been included in the Soviet or Russian curriculum.

C. Sequencing all school trajectories

Figure 10 portrays curricula trajectories of all school texts. The majority of them had a limited and short-lived presence in the literary curricula (cluster 1). The other groups can be characterized as late Soviet (cluster 2), Soviet-Russian (cluster 3) and post-Soviet Russian texts.