CA Journal of Cultural Analytics

February 17, 2021

Cultural Capitals: Modeling 'Minor' European Literature

Matt Erlin, Andrew Piper, Douglas Knox, Stephen Pentecost, Michaela Drouillard, Brian Powell, and Cienna Townson

Matt Erlin, Douglas Knox, Stephen Pentecost, Cienna Townson, Washington University in St. Louis
Andrew Piper, Michaela Drouillard, **Brian Powell**, McGill University
Peer-Reviewers: Rebecca Braun, Thomas Weitin
Dataverse DOI: 10.7910/DVN/SDMGRC

ABSTRACT

Conceived against the backdrop of ongoing debates regarding the status of national literary traditions in world literature, this essay offers a computational analysis of how national attention is distributed in contemporary fiction across multiple national contexts. Building on the work of Pascale Casanova, we ask how different national literatures engage with national themes and whether this engagement can be linked to one's position within a global cultural hierarchy. Our data consists of digital editions of 200 works of prize-winning fiction, divided into four subcorpora of equal size: U.S.-American, French, German, and a collection of novels drawn from 19 different "minor" European languages. We ultimately find no evidence to support Casanova's theory that minor literatures are more nationalistic than literature produced within major cultural capitals. Indeed, the evidence points to the exact opposite effect: all three of the models we employ suggest that novels written in more minor languages tend to be significantly *less* nationalistically focused than those written in European centres like France or Germany. Nevertheless our data do confirm Casanova's larger hypothesis of the existence of visible stylistic effects associated with a book's location within a global cultural hierarchy of languages.

"For I, being a Pole, can attain humanity only in my nation."
- Witold Gombrowicz, *Diary*

Among the key insights of Pascale Casanova's *World Republic of Letters* is a recognition that world literature can be construed as a complex adaptive system.[1] Though most conspicuously indebted to the work of Ferdinand Braudel and Pierre Bourdieu, Casanova's study highlights one of the foundational hypotheses of social network analysis, namely, that "meaning emerges from relations among cultural elements rather than inhering in the elements themselves."[2] According to Casanova, specific features of these cultural elements – in this case works of literature – can be explained on the basis of their position in a "world literary space" characterized by a competition for status. A case in point are representatives of the so-called "small"

or "minor" literatures. As Casanova explains, writers from countries on the literary periphery, regardless of the particular language in which they write, occupy structurally analogous positions of dependency. These positions constrain them in such a way that they are left with a limited number of strategies that can be employed to achieve recognition.

Whether or not one agrees with her specific conclusions, Casanova's relational model has proven highly suggestive, pointing toward new possibilities for comparative investigations into the tensions between the national and the global and between "core" and "periphery" in world literature. Rather than think of world literature as a distinct object, whether as a traveling canon constructed through time and space[3] or an evolving construct of the increasingly globalized publishing industry,[4] Casanova's model places the emphasis on relational aesthetic and stylistic norms and their co-determination by one's location within a broader cultural hierarchy.[5] World literature for Casanova is national literature produced under the pressures of global cultural capital.

According to Casanova's research, these pressures produce distinctive aesthetic effects. Writers from "deprived spaces" (191), that is, writers from peripheral countries without significant literary capital, have a powerful yet conflicted attachment to the idea of a national literature. As she writes, "Within deprived spaces, writers are condemned, in effect, to develop a national and popular theme: they must defend and illustrate national history and controversies, if only by criticizing them" (191). In contrast, "the autonomy enjoyed by the most literary countries is marked chiefly by the depoliticization of literature: the almost complete disappearance of popular or national themes" (199). A preoccupation with "pure" writing, in other words, with purely literary questions, or with one's relationship to literary history in universal terms, is the privilege of writers from countries with the highest levels of cultural capital.

Casanova's work contains an impressive range of examples, including references to canonical writers like Kafka, Beckett, and Michaux, as well as less well-known authors like the Lithuanian Saulis Kondrotas (b. 1953), the Croatian Miroslav Krleža (1893-1981), and the Polish writer who lived most of his life in exile, Witold Gombrowicz (1904-1969). As this list should indicate, while Casanova's claims are

often phrased in transhistorical and transspatial terms, ("the" world republic of letters), her evidence is largely concerned with writers from a particular time (the early- to mid-twentieth century) and place (Europe). More significantly, her selection of writers is only ever drawn from those whose work fits her thesis. Throughout Casanova's extensive monograph no negative examples are provided of writers at the heart of Europe's literary culture who chose to write national tales or those at the margins who wrote on more universal themes.

Despite these shortcomings, or perhaps in light of them, Casanova's project offers, in our view, a great deal of potential for developing a more generalizable framework for the study of international literary relations, one that is delimited by the opposing poles of national self-absorption on the one hand and universal extrospection on the other. Against the backdrop of ongoing debates regarding the status of national traditions in world literature, her model can be productively extended to an analysis of how national attention is distributed in contemporary fiction across multiple national contexts.[6] How do different national literatures relate to the question of national "themes," in Casanova's words, and how might this differ depending on one's position within a global cultural hierarchy? To the extent that it holds, Casanova's model postulates a greater pre-occupation with national themes in literature produced by countries that occupy more minor languages and traditions when compared to those that have traditionally dominated the global literary marketplace (i.e. works in English, French, and German).

In this essay, we provide a discussion of our attempt to computationally model and test Casanova's hypothesis as a first step towards a greater understanding of a relational model of literary production at global scale. Our hope in doing so is to address the current lack of comparative computational literary studies, with most research still overwhelmingly focused on single national literary frames.[7] While prior work has focused on the aesthetic effects of cultural capital within single linguistic contexts,[8] no work has to do date considered national contexts as themselves subject to forms of cultural capital and aesthetic distinction. By taking a computational approach to these questions, our aim is to inject not simply more evidence into debates about large-scale categories like "world literature," but also crucially more *independent* evidence. Unlike in Casanova, our data was chosen prior to, not as a function of, our analysis. Nevertheless, independent does not mean

without bias or limitations. Like Casanova's, our evidence is also limited by time and space, as well as generic conventions: it comes from two hundred prize-winning novels from twenty-two different countries translated into English and published since 2000. These selection criteria were chosen with specific aims in mind and in full awareness of the biases introduced by choosing works in translation that have been selected for literary prizes. Scholars such as James English, Graham Huggan, and Chantal Wright, among others, have written persuasively on the feedback loops through which prize competitions can shape literary production, reinforcing perceptions of what kind of fiction from different countries is "of value."[9] Gisèle Sapiro has shown how shifting power relations in global publishing have skewed the world market for translation toward the tastes of English-speaking readers.[10] And Andrew Piper and Eva Portelance have shown how for at least anglophone novels from different national backgrounds there exist detectable stylistic criteria that underpin prize-winning fiction that are consistent across different national contexts.[11]

Inspired by Casanova's interest in "consecration," however, our aim is to evaluate precisely the kind of literature that international arbiters of literary prestige consider to be of high quality and that has entered into world literary circulation. In choosing prize-winning work in translation, we are trying to control for works of similar cultural capital and thus potential readerships, which research has suggested has important stylistic effects. Moreover, with regard to translation in particular, it is important to note that, in most cases, national consecration occurs prior to global circulation and that such circulation is not limited to English translation alone but indicative of more widespread circulation. Most (though not all) of the books in our corpus were national or European prizewinners (*Deutscher Buchpreis*, *Prix Goncourt*, *European Union Prize for Literature*) prior to translation, with the prize serving as a key impetus for translation, including but by no means limited to translation into English.[12] In other words, there is little reason to think that our corpus is capturing fiction that is primarily of interest to Anglo-American readers, as opposed to offering a cross-section of works with generally high global prestige. Future research will want to test our findings and methods against further samples, either by drawing from a similar cultural subsection of prestigious works, focusing on translations into languages other than English (or comparing with English), expanding the analysis to different types of writing beyond prizewinners (such as

fan-fiction or genre fiction), or observing non-European regional frameworks. As trained German and European comparative literature scholars, we are starting with what we know best.

As our analysis indicates, we ultimately find no evidence to support Casanova's theory as it relates to European literature written after 2000. Indeed, the evidence suggests the exact opposite effect: novels written in more minor languages tend to be significantly *less* nationalistically focused than those written in European centres like France or Germany according to the measures used here. The pressure of achieving aesthetic recognition beyond one's borders appears to be reversed: those on the periphery must strive to be more, not less universal than those at the centre. Our results suggest two important insights: first, that modeling literary relations as one of centre/periphery or major/minor may indeed be a valid way of understanding international literary relations and, second, that the dimension of national attention (whether as theme or discourse) is a significant differentiator of aesthetic behaviour, though in the opposite direction than Casanova theorized.

Corpus

Our data consists of digital editions of 200 works of fiction, divided into four subcorpora of equal size: U.S.-American, French, German, and a collection of novels drawn from 19 different "minor" European languages. We use the term "minor" here in a deliberately self-conscious way, both to reference less commonly spoken languages as well as call upon the influential theoretical work by Deleuze and Guattari that would situate these languages within the provisional hierarchical framework envisioned by Casanova.[13] Casanova for her part departs from the use of "minor" in favour of the term "small literatures," both to differentiate herself from Deleuze and Guattari's interpretation of Kafka and to occupy a less pejorative framework ("under-resourced" is the term favoured by those in the NLP community). Nevertheless, given that our operating framework is to test hierarchical literary relations at the continental level, we consider the use of "minor" to be a clear indicator of our guiding assumptions. For our purposes here, we define minor languages as those European languages that are not among the ten world languages (excluding English) with the most titles included in the Virtual International Authority File (VIAF) database, a resource that aggregates the contents of library

catalogues worldwide.[14] Table 1 provides the list of languages in our dataset and the number of novels from each language.

Table 1. Languages and number of works in our dataset.

Language	Novels	Language	Novels
American English	50	Icelandic	3
French	50	Latvian	3
German	50	Lithuanian	3
Bulgarian	2	Macedonian	2
Catalan	3	Greek	2
Croatian	3	Norwegian	3
Czech	3	Portuguese	3
Danish	3	Romanian	3
Estonian	3	Serbian	1
Finnish	3	Slovenian	3
Hungarian	3	Ukrainian	1

Because we expect different kinds of works aimed at different readerships to behave differently with respect to Casanova's theory of national themes, we only include works that have been nominated for or won a general literary fiction prize (these include the Prix Goncourt, the Deutscher Buchpreis, the PEN/Faulkner Award for Fiction, and the European Union Prize for Literature, among several others). We thus condition our cross-national analysis on a single local dimension of cultural capital. Following Casanova, we are focused primarily on the European context, but we chose to include a US-American corpus to reflect recent shifts in the global distribution of literary capital as well as to engage with some recent scholarship on US-American literary exceptionalism.[15]

For the purposes of our analysis, the non-English language portion of our corpus is represented by English translations of the original works. We take this step for two

reasons: first, it ensures the consistency and comparability of results for the natural language processing we describe in the next section, and second, it allows us to condition on novels that have not only received national literary recognition through the prize selection process but have gained international literary recognition through circulation within other languages. We use translation as a mechanism of assessing a work's acknowledgment within a hypothetical "world" or at least transnational literary framework.[16] Future work will want to further study the effects that translation has on stylistic norms compared with untranslated works of similar cultural prestige. To ensure the diversity of our national corpora, we try to limit the number of multiple times an author appears in any corpus. No authors appear with more than one work in the US and minor European collections, while eight German and three French authors are represented by two novels. We should also note that in referring to novels as representative of a particular national literature, we mean that a work was published in the relevant country and written in the national language of that country. Our corpus includes no novels in French published in Switzerland or Belgium, for example, and no novels in German published in Switzerland or Austria, but it does include a small number of authors who could be considered bi-national. Finally, all of the novels were published in 2000 or later, with the majority appearing after 2005. Our choice of periodization is aimed at addressing distance to the major geo-political changes that transpired in Europe after the end of the Cold War in the early 1990s as well as perceived changes to the publishing industry that transpired through major consolidations that also occurred during this same period. While our time frame is different from Casanova's, it has the advantage of representing the most recent past in European literary fiction. Future work will want to explore whether the aesthetic pressures that we are seeing are different for different time-periods in European literary history.

To prepare our data for analysis, the spelling of all novels was normalized to American English to account for differences between British and American translations. Novels were then run through the LitBank tagger developed by David Bamman et al, a specially developed annotation tool for literary texts to assist with tasks such as named entity recognition.[17] As shown in Bamman et al, the accuracy of annotations using the LitBank tagger is considerably higher for literary texts than when using standard tools such as the Stanford NER. The output of the tagging process is a file for each work that includes additional labels for any word in the text

that represents one of several categories, including, most significantly for our purposes, that of "location."

Modeling Nationalism

To return to Casanova's thesis, our goal in this paper is to test whether writers in less central European literary cultures demonstrate a greater pre-occupation with "national and popular themes," which Casanova equates with a heightened degree of "politicization" of the novel. Are there contextual pressures *of a geographic nature* that are exerted on writers, influencing the stylistic qualities of their narratives? While we assume that there are a variety of ways that context shapes writers' choices (from genre to social prestige to biographical details), here we want to test the hypothesis that one's geographic location has an identifiable impact on one particular narrative feature -- that of national thematics. We refer to the phenomenon under investigation as literary "nationalism," but we want to be clear that we are using the term as a neutral designation for attention to the nation as an object of literary interest. It should not be understood as implying a positive or negative representation of the nation or a particular political orientation (i.e., "patriotic" or "critical"). In order to quantify these various concepts, we propose the following three models.

Our first two models are premised on the assumption that national content is reflected in the "geographic imaginary" of a literary work. Following in the footsteps of the work of Matthew Wilkens, Michael Gavin and others, we take geographic references as linguistic indicators of larger national or international concerns.[18] This can take the form of explicit references to the national context in which a novel is written. For example, the novel *Shadow Country* (2008), Peter Mathiessen's tale of the nineteenth-century Florida sugar cane planter and outlaw Edgar "Bloody" Watson, includes 40 instances of "U.S." or "U.S.A." and 77 instances of "America" or "American," putting it in the top 10 percent of all novels in our corpus. Alexis Jenni's *The French Art of War* (2011), which tells the tale of the fictional French veteran Victorien Salagnon, refers to "France" or things "French" over four-hundred times and begins this way: "[Salagnon] talked, and I wrote, and through him I witnessed the rivers of blood that cut channels through France, I saw the deaths that

were as numberless as they were senseless and I began finally to understand the French art of war."

We thus take the rate of explicit national self-reference as a scalar proxy for the nationalism of a novel's content. As these rates rise, we assume a novel to be more explicitly about one's national context, and as they fall to be less so. Jose Saramago's highly allegorical *Seeing* (2004), for example, the sequel to his acclaimed novel, *Blindness* (1995), mentions "Portugal" and "Portuguese" exactly once each and in a way to discredit the national specificity of the novel (though not without a touch of irony, a point to which we will return in our discussion section):

> Indeed, given the delicate nature of his message, it would be little short of insulting to say My dear compatriots, or Esteemed fellow citizens, or even, were it the moment for playing, with just the right amount of vibrato, the bass string of patriotism, that simplest and noblest mode of address, Men and women of Portugal, that last word, we hasten to add, only appears due to the entirely gratuitous supposition, with no foundation in objective fact, that the scene of the dire events it has fallen to us to describe in such meticulous detail, could be, or perhaps could have been, the land of the aforesaid Portuguese men and women. It was merely an illustrative example, nothing more, for which, despite all our good intentions, we apologize in advance, especially given that they are a people with a reputation around the world for having always exercised their electoral duties with praiseworthy civic discipline and religious devotion.

In addition to referencing national frames, novels may also manifest engagement with national themes through the use of geographical locations *within* the country of origin. For example, while Andreas Maier's *The Room* (2010) has a below-average rate of explicitly national mentions of either *Germany* or *German*, it registers the highest frequency of in-country mentions of any novel in our corpus, as it narrates the life of the intellectually disabled "Uncle J," who lives in a small German town during the 1960s and 70s. As the first in a series of works that span outwards (*The Room*, *The House*, *The Street*, etc.), it is in many ways a fictionalized version of Karl Ove Knausgård's autobiographical project, giving expression to the minutiae of

everyday experience within a highly concrete geographic framework. One can compare *The Room* to the Portuguese novel, *The Implacable Order of Things* (2000) by José Luís Peixoto, which also recounts a tale of small-town life, this time in a fictional village in Portugal. Peixoto's novel, however, has only a handful of references to concrete geographic places: the Rua da Palha, one mention of Lisbon, and a number of references to the Mount of Olives. While it is entirely reasonable to interpret Peixoto's novel as "about" Portugal and the sufferings of rural villagers, it does so in a narratively drastically different way than Maier. Besides the lack of concretization (Maier's Germany in the 60s and 70s), there is also in Peixoto a high degree of allegorical allusion -- while the Mount of Olives can certainly be imagined to be in Portugal (mountains with olives), its more famous location is in Israel and thus carries Biblical overtones -- as well as magical realism. There is a "giant" who plays an important early role in the novel (raping the main character's wife and beating up the husband repeatedly), the devil makes an appearance, and a Siamese twin who lives to 80 impregnates a seventy-year-old cook, after which the three of them happily raise their child. The paucity of concrete geographic attention aligns with the paucity of overt national or political themes. While not quite as allegorical as say, Kafka, Peixoto's novel is far more similar in nature to the Kafkan tradition than that of someone like Maier.

Our first model is thus the most straightforward and measures the rate of either adjectival or noun entities for the novel's country of origin (France/French, German/Germany, Bulgaria/Bulgarian, etc.). Model 2 uses a more elaborate approach that relies on a two-step process of named entity recognition and a large gazetteer of place names (ca. 12 million locations) to resolve geographic mentions of various kinds (cities, landmarks, topographical features) to their national frameworks. We describe the workflow for this model in more detail in the Appendix (located in the supplementary material). The output for the model is a table of geographic mentions for each novel that have been resolved to a particular country based on a set of heuristic rules. These results allow us to calculate the frequency of place name mentions overall; the rate of within-country (national) mentions; the rate of out-of-country (international) mentions; as well as the entropy of national and international mentions to capture the heterogeneity of spatial construction in the novels.

Our third and final model is designed to capture textual features that are indicative of “national and popular” themes but that are at least partially independent of the geographic attention measured by the first two models. One can certainly imagine scenarios where a preoccupation with the nation is mediated through narrative elements other than references to specific locations. The content that we focus on in this case reflects a different conception of literary nationalism, one based on what we would describe as an actor/event model of national history. If a novel engages with historical content that is significant for the nation, one way it presumably does so is at the level of key historical personae or events. For example, Patrick Deville's *Plague and Cholera* (2012) is a biographical novel about the French-Swiss scientist and explorer, Alexandre Yerson, who discovered the bacterium of the bubonic plague and lived most of his life in South-East Asia. The novel is framed around the onset of the Second World War, which is marked as an ending of a particular global French national consciousness. A variety of historical actors make appearances, from Louis Pasteur to Adolf Hitler to Ho Chi Minh, along with references to well-known historical events in France's past, such as the "Treaty of Versailles," as well as temporal markers like the "belle époque" or the "Second Empire." Deville's novel is as much, if not more so, about "France" as it is a tale of scientific discovery.

To capture these kinds of features, we turn to a general representation of each country's history that is ideally as consistent as possible across different national contexts. With this goal in mind, we use the "history of" Wikipedia page for each novel's country of origin (History of Germany, History of Bulgaria, etc.) and extract a list of the people, places, and events that contain hyperlinks within those pages, limiting ourselves to bi- and tri-grams to avoid ambiguity. These pages have the value of being highly standardized across each country as well as reliably encompassing major events and personae. The linked entities provide reliable consensus around high value information relevant to each country, as can be seen in the examples listed in Table 2 for Germany and Bulgaria. It is important to point out that links do not exclusively point to national references but can also include extra-national references that are important to that country's history. For example, "Roman Empire" appears in the list of important German bigrams alongside more explicitly national references like "Nazi party." Our third model thus calculates the rate at which novels explicitly refer to these lists of nationally relevant historical actors, events, and regions.

Table 2. Selected list of links in Wikipedia "history of" pages for two sample countries.

Bulgaria	Germany
Asen dynasty	Holy Roman Empire
Ottoman Empire	Harz Mountains
April uprising	Teutonic Knights
Todor Zhivkov	Thirty Years' War
Khan Kubrat	Peace of Westphalia
Ivan Vladislav	Treaty of Versailles
Constantinople Conference	Werner Heisenberg
Alexander of Battenberg	Nazi Party
Treaty of Craiova	Final Solution

Given the variety and complexity of the discursive markers that could be said to indicate national content in a novel, it is certainly the case that any single approach will only capture particular facets of the topic, and ours is no exception. No model can be universal. By the same token, however, this qualification applies equally to the numerous, more traditional investigations of the concept of "nation and narration" within existing scholarship, including Casanova's own study. “A sense of ‘nationness’” might be unearthed in one interpretation in idyllic representations of home and hearth and in the depiction of industrial prowess in another.[19] Casanova herself insists in one passage on the “genuine hegemony of realism” in the most politicized (i.e. deprived) literary spaces and then transitions a few pages later to a discussion of Kafka – certainly not a realist in any conventional sense – as the preeminent representative of “the necessarily political position of writers in emerging nations.” One of the advantages of a computational approach is that it demands a high degree of specificity with regard to the definition of the phenomenon under investigation as well as the features selected to identify it. While some may see this as a limiting factor, we see it as an important step for the process of generalizing about real-world phenomena, especially those as large in scale as "world literature" or international literary relations. To reliably make such

generalizations, we need to condition on particular and consistent literary features and estimate their prevalence in the world.

Although quantitative methods may appear to encourage more categorical judgments about topics such as literary nationalism, a further advantage of their use is that they can allow for more nuance and differentiation when it comes to understanding the aesthetic practice under investigation.[20] An interpretation of, say, the colonial unconscious in Austen's *Mansfield* park, however sophisticated, encourages us to view colonialism as either central to the novel (if we agree) or not (if we don't). Observing semantic variation across a corpus, on the other hand, for all of its association with precise measurement, can help us to think in terms of differing levels of intensity. Literary nationalism should be viewed as both a relative phenomenon (subject to gradations in magnitude) and a relational one (co-determined by the structures of world literary space). Our models aim to foster this kind of relative and relational thinking.

Results

We begin our analysis by observing corpus-level effects for each of our models and then move to the document level. For Model 1, we find that the average frequency of national references across our whole corpus is 0.000459 or roughly 46 per 100,000 words (where 100K words represents roughly the length of an average novel). We also find that our subcorpora use national references at significantly different rates ($\chi^2(3,200)=430.83$, $p<2.2e-16$) (Table 1). Interestingly, it is the French corpus that references France/French considerably more often than all of the other collections' invocations of their own national frames, with an estimated 70% more mentions than the Minor Literature collection.

Table 1. Ratio of national mentions to all tokens by corpus.

	France	Germany	US	Minor
national mentions	2498	2109	2125	1818
all tokens	3,798,096	4,718,696	5,385,678	4,708,922
ratio per 100K	65.7	44.7	39.4	38.6

For Model 3, our actor/event model based on Wikipedia data, we find that the overall rate of occurrence is 0.000102 or about 10 mentions per 100,000 words. Once again, the different corpora use these mentions at significantly different rates ($\chi^2(3,200) = 184.98$, $p<2.2e-16$), where this time it is the US collection that indicates significantly higher rates than the other groups, with an estimated 250% more mentions than the Minor Literature collection and 25% more than the next closest collection, which is the French corpus (Table 2). The Minor Literature collection is once again the lowest among the collection and statistically significantly lower than the next highest corpus (Germany).

Table 2. Ratio of Wikipedia mentions to all tokens by corpus.

	France	Germany	US	Minor
Wiki count	431	454	756	260
all tokens	3,798,096	4,718,696	5,385,678	4,708,922
ratio per 100K	11.3	9.6	14.0	5.5

For Model 2, instead of using our observed counts, we adjust the overall rates of occurrence based on estimated levels of detection error. Because NER is not error-free in its prediction of place names, and because we assume that it may perform differently on our different corpora, we undertake a process of manual validation to assess the degree of error for each corpus. In other words, if places mentioned in the small Europe corpus are more obscure, it might be the case that NER has a lower degree of accuracy in identifying those locations than the locations in the French or German or US corpus. Thus the aim of this step is to address the confounding

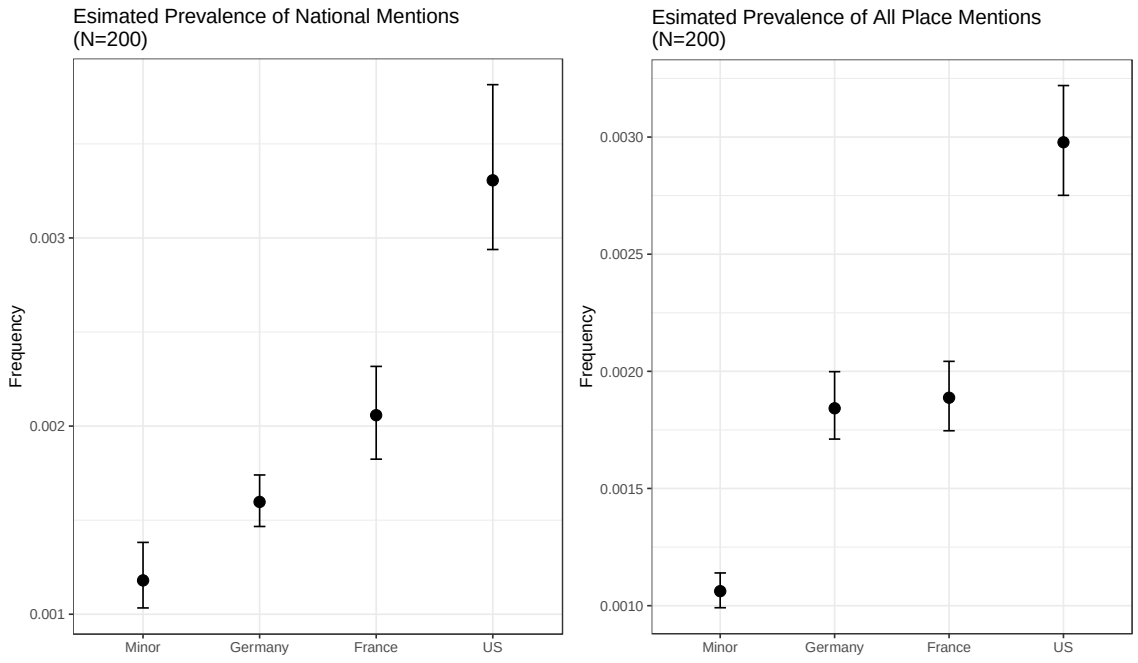
problem that any observed differences in the levels of national place names may be due to differential levels of detection of place names rather than being due to their actual prevalence. To account for this possibility, we hand-validate 1,200 passages drawn equally from each of our four subcorpora, which provides us with estimates of the degree of error by corpus, which we present in Table 3. We focus here exclusively on recall as precision and specificity approach 1 in all cases. Missed place names (false negatives) are a much greater problem in our case than the over-prediction of place names (false positives). Doing so, we observe an overall rate of place name occurrence in our corpus of 0.00591 or roughly 590 per 100,000 words.

We then use a Bayesian process of estimating the amount of error for each corpus recommended by Messam et al, which not only adjusts the estimated prevalence of place names based on the degree of error (i.e., recall), but also provides more conservative (i.e., wider) confidence intervals surrounding the prevalence estimation.[21] Because we can only validate a small number of samples and because place-names are rare events, we model this "ground truth" as itself prone to uncertainty. The net effect of these efforts is not only an adjustment of the observed rates of place names upwards in accordance with the differential recall rates of our four subcorpora, but it also generates more conservative estimates of uncertainty surrounding the process. We provide a full description of this process in the Appendix in the supplementary material.

Table 3. Recall (True Positive Rate) by corpus and place name type.

	All	International	National
France	0.593	0.710	0.543
Germany	0.653	0.641	0.738
Minor	0.757	0.819	0.683
US	0.596	0.673	0.541
All	**0.642**	**0.714**	**0.600**

After adjusting for differential error rates across the four corpora, we find that our Minor Literature collection still appears to use the lowest levels of national mentions when compared to the other corpora, and that the US collection appears to use significantly more (Fig. 1). To put this in context, we estimate that the Minor Literature collection includes national mentions at a rate that is roughly 26% lower than the next highest corpus (Germany). This would amount to an average of about 40 extra place names per 100,000 words, a not insignificant amount. When compared to the French and US corpora, we see how those other collections use 75% and 180% more national mentions than the Minor Literature collection. We also see how these rates are consistent when we consider all place mentions, with both the German and French corpora using roughly 75% more place name mentions overall than the Minor Literature collection and the US once again using almost two-times as many (Fig. 2).



Figs. 1 & 2. Estimated prevalence of place mentions by category and corpus. Error bars represent 95% confidence intervals.

When we move to the document level, we see similar results, though somewhat less strongly differentiated. Here we combine all three models using the adjusted values from Model 2 to arrive at what we call a "nationalism quotient" for each work. Figure

3 indicates that our Minor Literature collection once again demonstrates the lowest levels of this quotient, with a degree of difference that indicates slight statistical significance when compared to the next lowest group of Germany (according to a Wilcoxon rank-sum test with continuity correction, W=1564, p=0.03068). A Kruskal-Wallis rank sum test suggests that the variances among these four groups are independent of one another (χ^2(3,200)=29.087, p=2.147e-06). Both France and the US are estimated to use significantly higher levels of these combined national mentions, with the US using the most. To put these numbers in perspective, the median level of our nationalism quotient in the Minor Literature collection is about 63% lower than that of the German collection and 8 of the 10 novels with the lowest levels of our nationalism quotient belong to the Minor Literature collection.

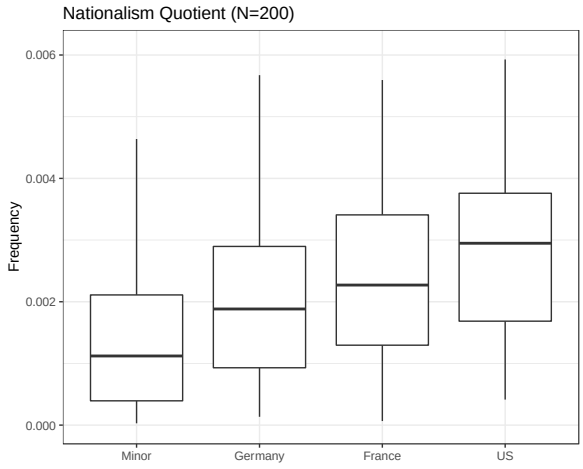


Fig. 3 Nationalism quotient at the document level by corpus with outliers removed.

Finally, we present a graph of the ratio of national to international mentions by corpus (Fig. 4). Here we see how the US corpus is the only corpus where there exists a sizable portion of novels that have more than twice as many national mentions as international, echoing findings by Matthew Wilkens on a considerably larger corpus of U.S. novels.[22] Otherwise, in general across our European corpora there is a bias towards favouring international place mentions over national ones.

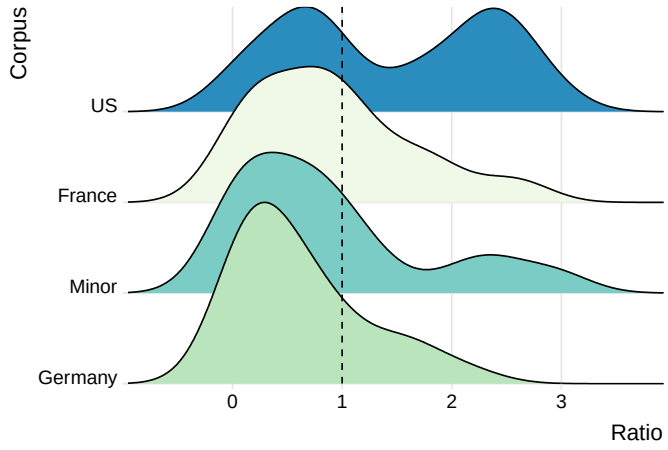


Fig. 4 Ratio of national to international place mentions by corpus with outliers removed.

Discussion

According to our models, then, we see consistent evidence to suggest that so-called minor literatures exhibit significantly different behaviour when it comes to pre-occupation with national self-reference. Whether our proxy is national mentions (e.g., France/French), mentions of in-country locations, historical actors and events, or our broader overall nationalism quotient, each measure indicates a lower degree of preoccupation with national self-reference by authors writing in less common European languages. All of our measures point in the same direction and exhibit statistical significance. Indeed, our findings suggest a clear ordering principle at work, one that aligns with perceived levels of cultural capital in a global context: the United States marks an extreme when it comes to explicit national self-expression, followed by France, Germany, and then our collection of less common languages.[23] Germany's position within this hierarchy is itself revealing of the overall trend. Because of the particularly dynamic and conflicted nature of German nationalism since World War II, it makes sense that we see this corpus playing a middle role between the low-levels of national self-reference in the minor literatures and the more unabashedly self-absorbed literature of France and the U.S.[24] Were we to remove the German collection from our experiment, the differences between the minor literatures and the other two cultural capitals would be extreme.

If we take our models at face value, our findings suggest that we ought to reject Casanova's primary claim. We find no evidence that minor literatures are more likely to, in her words, “defend and illustrate national history and controversies.” In fact, our measures indicate that the opposite is likely to be true. It is worth underscoring the novelty and significance of this finding: within a corpus of literature drawn from 19 different languages, we see an overall consistent and significant behavioural difference when it comes to a particular kind of national self-expression compared to corpora drawn from more widely-spoken languages.

Our findings nevertheless raise two larger, more philosophical questions. The first concerns the extent to which our models adequately capture the idea of "national history and controversies" or "nationalism" more generally, which we might label as a question of "construct validity." How valid are our models for the question we are posing? The second, even more general question is what we might call the "theoretical validity" of our models: how valid is it to model international literary relations in terms of Bourdieuian distinction? As we will indicate below, our findings suggest that on this front Casanova's framework appears highly appropriate, just not in the way she indicates in her book.

One of the questions we have had throughout our project is whether more national self-references, whether to place names or key events and historical actors, are actually an indication of "nationalism" in the novel and whether the inverse is equally true (i.e., fewer references indicate lower levels of nationalism). Said more colloquially, what, precisely, are our measures capturing? If we look at the novels that indicate some of the highest levels of our nationalism quotient, we find that they are indeed overtly "nationalistic." Deville's *Plague and Cholera*, the highest scoring novel in our collection, is an extended reflection on French colonial identity reflected through the life of a single scientist. Ófeigur Sigurðsson's *Oraefi*, sixth overall and highest for the Minor Literature collection, is littered with Icelandic place names and concerns the travels of an Austrian toponymist, whose primary goal is to piece together place names and historical lore in a famed region of Iceland. This narrative of national references is quite literally concerned with the link between nation and narration. Andreas Maier's *The Room*, fourth highest overall, chronicles everyday life in a small German town. While seemingly local in attention, as it chronicles

Uncle J's route to work at the Post Office with obsessive detail, such particularity is self-consciously set against the nationalizing pressures of post-war German commerce and infrastructure, most notably through the grandiose concept of the *Ortsumgehungstraße* [ring road or orbital highway] that encircles and bypasses the local in favour of the national. As the narrator writes, foreshadowing the novel's trajectory: "Germany in the year 1969, a land that still stands before the first economic collapse. A land without *Ortsumgehungsstraßen*." To take another example, Pauline Jiles's *News of the World,* which ranks third overall, tells the story of Captain Jefferson Kydd and his 10-year old charge Johanna, an orphan who has been held captive by the Kiowa for four years and is now being returned to her relatives by the former soldier. Their precisely documented journey through Texas in 1870 -- "the Captain and Johanna would travel through Cross Timbers to Spanish Fort and then on south to Dallas and eventually four-hundred miles farther south" -- provides the backdrop for discussions of American political and social upheaval in the aftermath of the Civil War. The novel tellingly ends with a perfect trinity of property and possession, as the Captain takes parental ownership of Johanna, his daughters take ownership of his wife's former Mexican estate, and Johanna is married off to a highway-robber-cum-successful cattle rancher at the novel's close. A series of frontiers and wild-spaces, from the ambiguous boundaries with Mexican culture to the psychological profiles of Native-Americans, are closed off at the end of this place-bound journey in favour of a more contained and self-possessed America.

Each of these novels engages the idea of the nation in a different way: for Deville it is within an international network of knowledge discovery; for Sigurðsson it is as an engagement with historical folklore; for Maier as the foregrounding of specific local coordinates as symbols of the hollowness of national structures; and for Jiles as an act of conciliatory enclosure. And yet they are at the same time united through a particular linguistic dimension that foregrounds a kind of national self-consciousness. The explicit mention of national frames or national places or national actors and events represents what we would call a form of "performative nationalism," where narration is overtly and repetitively situated in a national context.

What this suggests is that for novels that exhibit lower levels of such national self-reference, what we expect to observe is not necessarily the absence of nationalism *per se*, but the absence of a particular kind of nationalism. Our models, for example, are not able to capture what we might call a form of "allegorical nationalism," where a national context is alluded to but never (or almost never) explicitly mentioned. To take one example, György Dragomán's *The White King* (2007), which has the lowest nationalism quotient of all the novels in our corpus, is a novel narrated by a teenager who comes of age in a nameless and utterly oppressive totalitarian country, one that bears a good deal of "resemblance" to the Romania where the author grew up. Such resemblances surface through character surnames like Gyurka (a former national soccer player), fictional street names like "Street of the Martyrs of the Revolution," mentions of the Young Pioneers, and a handful of mentions to the "Danube Canal." While there are a number of Danube canals in different European countries there is at least one in Romania where the river flows into the Black Sea; alongside Gyurka, there are also a number of Hungarian nicknames in the novel like the narrator's friend Szabi, and while to our knowledge there is not actually a "Street of the Martyrs of the Revolution" in Romania, there is a "Martyr's Way (Calea Martirilor)" in Timisoara and a "Revolutionary Plaza" in Bucharest with the inscription, "Glory to our martyrs (Glorie martirilor nostri)." As Éva Bányai has explained in her detailed reading of the novel, all of the names in *The White King* are designed to produce both local and transnational echoes, destabilizing the clarity of the novel's nationness, while simultaneously giving a sense of *connoted* place (most acutely through auditory allusions that are harder to capture in English translation).[25]

Where *The White King* expressly withholds specific place names in its narrative, other novels in our Minor Literature collection allegorize nationhood through a focus on other cultures. László Krasznahorkai's *Seibo There Below* largely takes place in Japan, though in a mostly fantastic and philosophical mood. Referring to the city of Kyoto, the author writes, "the Allusion floats across the entire city...so that it may represent the ungraspable, the inconceivable, in other words: unbearable beauty." Kyoto, in a classic case of European Orientalism, comes to stand not for a concrete sense of national Japanese identity, but for a more universal literary abstraction known as beauty. Such moments of transnational allegorization in the Minor Literature collection almost always have a directional aspect to them: according to

our data, minor literatures almost never set themselves in foreign places which, from a European perspective, tend to be viewed as having lower cultural status, while major literatures do. German novels can be set in Syria (Rafik Schami's *The Dark Side of Love*), U.S. novels in Nigeria (Julie Iromuanya's *Mr. and Mrs. Doctor*), French novels in Vietnam (*Plague and Cholera*), but when minor literatures are set abroad, they most often look to the center as it has been traditionally understood. The Macedonian novel, *Freud's Sister*, is set in Germany, while the Hungarian novel *Trieste* is set in Italy; Milen Ruskov's Bulgarian *Thrown into Nature*, a replica of the Spanish picaresque novel, is set in Spain, while notably Daniel Kehlmann's most recent German novel, *Tyll*, which is also a picaresque novel set in the same time period, takes a national work as its reference point (named after Till Eulenspiegel, a German Don Quixote).

These examples highlight for us the value of specificity when it comes to talking about what we mean by "nationalism" and the novel. We are not capturing a universal concept, but a particular manifestation expressed in a particular way. While novels like *The White King* (or Saramago's *Seeing* or Ruskov's *Thrown into Nature*) can certainly be read as "national" in some sense, they are national in ways that are significantly different from novels like Matthiessen's *Shadow Country* or Michel Houellebecq's *The Map and the Territory* (2010), where the litany of references to Paris and its various neighborhoods are complemented by a host of invocations of the material details of everyday modern life (Mercedes cars, Western Digital hard drives, iPods, Fujinon lenses, and Michelin maps). As Bányai explains, the allegorical nationalism of works like *The White King* allows a novel to transcend its local framework and potentially aspire to some other kind of frame of reference. There is no doubt, to return to a more famous example, that Kafka's novels are in some sense about the bureaucracies of the Austro-Hungarian Empire and the Workers Insurance Institute in Prague. But the almost complete absence of concrete geo-historical references in his work allows his fiction to be read as part of a very different literary tradition. Works like the Estonian author Rein Raud's *The Brother*, about a nameless brother who shows up in a nameless town to right his sister's disinheritance by a cabal of nameless men, is a direct inheritor of this tradition. The question of "estate" or "statehood" is central to this novel without names, but it is addressed in ways completely differently from the bulk of novels produced in France

or the U.S. Indeed, the author has explicitly called it a "spaghetti Western," one of the great transnational genres in existence.

This dualism or double-voiced quality provides a useful perspective from which to nuance Casanova's claims about the nationalism of minor literatures and our models' rejection of these claims. Rather than indicating that minor literatures are exclusively less nationalistic overall than their major kin, our measures suggest that so-called peripheral writers are potentially more likely to adopt an anti-realist aesthetic than writers from the center, and that, *pace* Casanova, their resistance to realism, not their embrace of it, constitutes a key political or national moment in their works.[26] It is this distinction, between novels that are explicitly and *performatively* nationalistic and those that are more allegorically so, that our models are able to foreground. Interestingly, this result aligns with the controversial hypothesis about peripheral world literary production floated by Fredric Jameson back in the mid-1980s, where he argued that "all third world literature" was "necessarily allegorical."[27] Jameson's understanding of allegory was rather less straightforward than the one our models are registering, and we would certainly not endorse his universalist claim here. Nonetheless, our work supports the idea that literature from at least the European periphery does indeed appear to be *more likely* to be allegorical, if not always so. Even more significantly, it suggests that allegory may be a fruitful category for conceptualizing the impact of an asymmetrical distribution of cultural capital in the world literary system.

This brings us to our final discussion point, which we see as a provocation for how future research into transnational literary relations is framed. One residual question we are left with is whether our modeling of national literary production as part of a transnational system of distinction continues to be a valid one. Much recent theorizing on translation and the novel has suggested that the past few decades have witnessed a dramatic transformation of literary culture towards a model of global brands and a single, homogenous "style." One could argue that our inability to confirm Casanova's hypothesis regarding the nationalism of minor literature is due to a transformation in world literary space (i.e., our data simply looks at a time period after her book). In the current global literary system, it may be the case that peripheral novels are subject to a *denationalizing* pressure in order to achieve recognition by larger audiences. Shifts in the nature of publishing over the past few

decades have potentially eliminated the traditional and centralized instances of consecration that underpin her view of world literary space as stratified on the basis of cultural capital. Instead, the (global) market has become the sole arbiter of quality. If true, this transformation would challenge the central premise of Casanova's model, namely, that "the key to understanding how this literary world operates lies in recognizing that its boundaries, its capitals, its highways, and its forms of communication do not completely coincide with those of the political and economic world" (11). Perhaps the literary and economic worlds are now entirely isomorphic.[28] This reading would align with some recent commentaries on contemporary literature, including Tim Parks lamentations regarding the "dull new global novel," which eliminates "culture-specific clutter" and simplifies language to ease international acceptance.[29]

The emergence of a truly global market for literature, characterized, among other things, by the consolidation of the publishing industry, the reduced influence of mediating arbiters of prestige, and the proliferation of international prizes, has no doubt changed the calculation for authors from the European periphery. And yet our models indicate that those at the so-called centre are not similarly affected by these pressures. While a purely market-based explanation would suggest a convergence of novels toward a standard form along the lines of Parks' "global novel," our models indicate that substantial differences among national corpora continue to exist and those differences cleave along differences in the distribution of cultural, rather than just economic capital. As the fictional Icelandic poet in Sigurðsson's *Oraefi* laments, "If I were not an Icelander, I would have earned the Nobel Prize… I'm stranded on a deserted rock, banished from the raging sea of languages! ... I got the worst curse of Babel, born an Icelandic poet, no one understands me, no one hears me at all." As our findings indicate, the conditions of "being heard" on the world literary stage continue to be unequally distributed.

While our models contradict Casanova's primary hypothesis about the explicit national focus of minor literatures, we would argue that her larger claim continues to remain generative as an interpretive framework for international literary production. It appears to be the case that "the practices and traditions, the forms and aesthetics that have currency in a given national literary space can be properly understood only if they are related to the precise position of this space in the world

system” (39). We look forward to future efforts to build out the map of this “world republic of letters,” efforts based on more nuanced models of the stylistic pressures impinging on writers from the periphery as well as on an expanded corpus of texts from beyond Europe and North America.

Appendix

Geo-spatial Model Description

The following is a description of the geo-spatial model used in predicting place names and their country of location.

1. **Litbank Tagging**. All novels are run through the Litbank tagger. Tagged locations with multiple words are outputted as a single line. Thus "Grand Canyon" would resolve to:

start_token	end_token	label	text
2161	2162	LOC	Grand Canyon

2. **Subset by place names**. Litbank tables are then subsetted by GPE and LOC tags and stored in a separate table, including the novel's title, country of origin, corpus, and the token number of the location. This allows users to study the narrative timing of locations. This table is located in the supplementary data under "min_geoTags_all.csv."

This gives us a total of **85,485 predicted locations** in our data of 200 novels, or about 427 per novel.

3. **Creating a gazetteer**. The next step is subsetting this table by names that match our gazetteer. This is created using the following steps:

- Our geo-names derive from the dataset allCountries.txt from geonames.org, an open-source resource of place names. This dataset contains 12,013,739 place names and is freely downloadable.

- We create tiered place names based on generality, using the following schema:

    - Continents. We use the following list of supra-national regions:

"Europe", "South America", "North America", "Africa", "Antarctica", "Asia", "Middle East", "Far East", "Near East", "Central America"

- Countries. We use a customized version of the country-designation in geonames.org located in the table "min_countries_allNames.csv." Country names are normalized to colloquial uses such that "People's Republic of China" becomes "China" etc.

- States. Here we use the geonames admin taxonomy "ADM1" with populations greater than 10,000. We normalize Canadian spellings and names (such as removing French partner names (New Brunswick/~~Nouveau-Brunswick~~).

- Cities. We keep all city names according to the geoname taxonomy of "P" with populations > 10,000.

- Regions. We also keep regional names such as bodies of water, mountains, and landmarks (like bridges). We use the following variables from geonames: "LK", "LKS", "SEA", "STM", "PRK", "GNL", "BDG", "DSRT", "MT", "MTS". We remove any names included in this set if they are already listed in the higher order sets.

- Custom. We provide two custom lists that were derived through rounds of data cleaning. The first, "min_geo_custom1.csv," is a list of place names that did NOT match the above gazetteer. We examined all non-matching place names that occurred more than 5 times in our data and resolved these locations to their country-level locations. There are 634 unique place names that result in an additional 10,477 matched place names. The second custom list, "min_geo_custom2.csv," is based on the manual examination of all matching place names that occurred more than 2 times in our data. Any errors detected at this level were manually cleaned and this custom list is used to override the resolution of place names to their country-level assignments. Here there are 257 unique place names resulting in 3,327 adjusted assignments.

- Using all of these steps, we reduce our data from 85,485 potential place mentions to **58,578 matched place mentions** or roughly 292 per novel.

4. **Country Prediction**. The next step is assigning a country-code to every place name in our dataset. We use the following hierarchical process:

- First, place names are looked up in our custom lists. These serve as the first level assignment. If there is a match, this is the place name's country assignment. If no match, then:

- We move to the next levels in descending order of generality: continents, countries, states, cities, and then regions. Once a match is made all other levels are ignored. If there are multiple matches made at a single level (i.e. a place name is listed numerous times in our geo-data), then we use the following heuristics to resolve to the proper country:

  - For multiple city matches (i.e. Paris, Texas v. Paris, France), we take the country with the city whose population is the highest. Thus "Paris" always resolves to "France" and "Warsaw" always resolves to "Poland."

  - For regions, if there are multiple regions from multiple countries matched, if the region is in the novel's country of origin, we resolve the region to the novel's country of origin. Thus the "Danube" will resolve to "Romania" if the novel is originally written in Romanian and "Germany" if the novel is originally written in German because the Danube is a body of water that belongs to multiple countries. Sometimes, as with mountains, places will have the same name but be in different countries. Thus we assume that a natural toponym is "national" if any of its locations falls in the novel's country of origin. If there is no match to the novel's country of origin, and there are multiple national regions, then we simply resolve to the first country because this will be registered as "foreign" in our subsequent workflow (where we measure national v. international mentions). Thus, for at least some of our data, we do not identify the *actual* country of a place, but

merely the in- v. out-country nature of the place, which is what we validate on below.

- A final note in our annotated data: we resolve internal bodies of water or mountain ranges to *all countries in our dataset* (rather than all countries) that include those natural regions. The "Alps" thus resolve to "France" and "Germany" because we don't have novels from Italy, Switzerland, or Austria. Similarly, historical entities such as "Yugoslavia" or "the Soviet Union" resolve to their respective countries in our dataset. Thus our annotations are not fully transferable to other data sets.

- The resulting output is a two-letter iso code for each place name. This final table is: "min_geoTags_Annotated_All.csv."

Validation and Error Detection Description

We manually validate our method on a random sample of 1,200 passages drawn equally from our each of our subcorpora (i.e. 6 random passages were selected per novel). Student coders read each passage and annotated the actual country of location of each place name, the results of which were then reviewed by the two Principal Investigators.

From this we can measure the rates of false positives (identified locations that are not place names, such as "Charlotte") and false negatives (locations that are not identified by the tagger but are place names) for each corpus to estimate differential accuracy across the 4 corpora.

"min_Validation_All.csv" contains all passages that were annotated and "min_Validation_ErrorTable.csv" subsets this table by all place mentions, either those tagged by our process or observed by our student RAs. We use this table to estimate our detection error.

In order to estimate the amount of error in the detection process for each subcorpus, we use the procedure recommended by Messam et al. discussed in our paper and

outlined in our accompanying code. Rather than use the observed error, we assume that there is a degree of uncertainty with respect to our ground truth. Because place mentions are rare events and because we cannot manually validate all passages in our data, it could be the case that there are biases encoded in the sample passages used to validate our detection accuracy (some books have significantly higher amounts of place mentions and thus the observed error could be skewed because of those books, etc.). For example, on 140,000 observed tokens, we find only 358 instances of national mentions overall, with only 122 false negatives across four subcorpora. Thus any estimate of uncertainty is based on very sparse data.

The Bayesian procedure outlined by Messam et al. allows researchers to specify the parameters in advance that will be used to estimate the uncertainty in their prevalence estimations. As with all Bayesian modeling, there is no a priori correct selection criteria for these parameters, which are intended to reflect researcher beliefs. Thus, to account for this uncertainty and our understanding of it, we undertake the following steps:

1. We first remove the most extreme books in our error table, i.e. for each subcorpus we remove the book with the most number of place mentions. Because books can have very different levels of place names, a single book can account for a high percentage of observed errors. In the case of the Minor Literature collection in particular, a single book from a single language could disproportionately skew our understanding of that corpus's error. This step results in the removal of 48 instances of observed place names in our data.

2. Using this error table, we then use a process of bootstrap sampling to estimate the upper and lower bounds of our detection error (i.e. recall). We take 1,000 samples of our error table with replacement and calculate the recall for each corpus for every sample. We sample each corpus with replacement separately so that their overall rates of place names are preserved. We then use the 10th and 90th percentile as the bounds for the Bayesian estimates as they reflect for us a reasonable degree of uncertainty regarding expected levels of recall. Table 4 presents the values used while Figure 5 represents the overall distribution of recall by corpus. We did not feel that the upper and lower quartiles were sufficiently broad, while we did feel that the 1.5 IQR was too broad. For example, we felt it was incredibly unlikely that the true

recall for either the German or Minor Literature collection would approach 85%, while the US or French rates as low as 40% struck us as similarly unrealistic. Table 4 presents the bounds that were implemented in our model which we consider conservative estimates of just how much uncertainty there is with respect to the true recall of our detection process.

Table 4. Upper and lower bounds used in our Bayesian estimation of the error surrounding our true prevalence estimates of national place references by subcorpus.

	Lower Bound (0.10)	**Upper Bound (0.90)**
France	0.481	0.604
Germany	0.672	0.803
Minor Lit	0.61	0.781
US	0.475	0.602

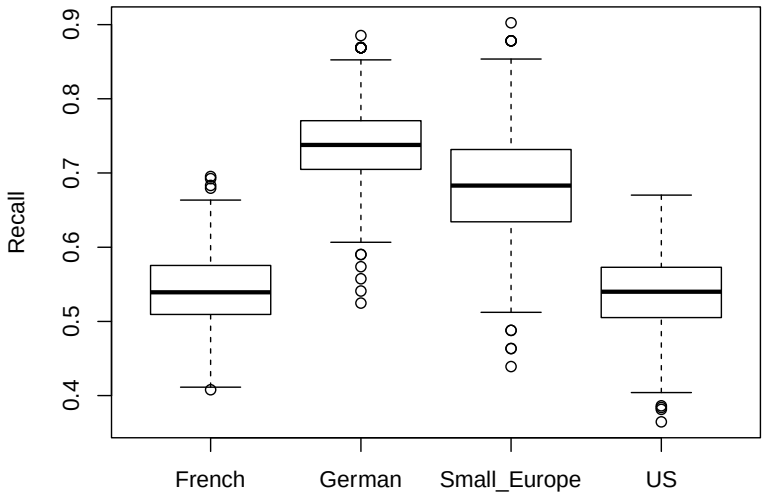


Fig. 5 Distribution of observed recall using bootstrap sampling per subcorpus.

References

We would like to thank Jack Pruett, Alireza Taheri Araghi, and Lily Wu, all fellows with the Humanities Digital Workshop at Washington University, for assistance and consultation during the early phase of this project.

[1] Pascale Casanova, *The World Republic of Letters*, trans. M.B. DeBevoise (Cambridge: Harvard UP, 2004).

[2] Paul DiMaggio, "Cultural Networks," *The Sage Handbook of Social Network Analysis*, ed. John Scott and Peter J. Carrington (Los Angeles: Sage, 2011), 294.

[3] David Damrosch, *What Is World Literature?* (Princeton, N.J.: Princeton University Press, 2003).

[4] Rebecca L. Walkowitz, *Born Translated: The Contemporary Novel in an Age of World Literature* (New York: Columbia University Press, 2015). Tim Parks, "The Dull New Global Novel," *New York Review of Books*, February 9, 2010. https://www.nybooks.com/daily/2010/02/09/the-dull-new-global-novel/

[5] For an elaboration on this theory, see Warwick Research Collective, *Combined and Uneven Development: Towards a New Theory of World-Literature* (Liverpool: Liverpool University Press, 2015).

[6] Writing in *The Routledge Companion to World Literature* in 2011, Jin Tsu asserted that "there is as yet no consensus as to where world literature stands in relation to national literatures." Little has changed since then. Jin Tsu, "World literature and national literature(s)," *The Routledge Companion to World Literature, eds.* Theo D'haen, David Damrosch and Djelal Kadir (Abingdon: Routledge, 2011).

[7] For exceptions: Hoyt Long and Richard Jean So, "Turbulent Flow: A Computational Model of World Literature," *Modern Language Quarterly* 77:3 (2016): 345-367. doi: 10.1215/00267929-3570656; Matt Erlin, "Topic Modeling, Epistemology, and the English and German Novel," *Journal of Cultural Analytics* (May 1, 2017). doi:10.22148/16.014; Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy* (Evanston, Illinois: Northwestern University Press, 2019).

[8] James F. English, "Now, not now: Counting time in contemporary fiction studies," *Modern Language Quarterly* 77.3 (2016): 395-418. doi: 10.1215/00267929-3570667; Alan Riddell and Karina van Dalen-Oskam, "Readers and their roles: Evidence from readers of contemporary fiction in the Netherlands," *PloS one* 13.7 (2016): e0201157; Mark Algee-Hewitt, Sarah Allison, Marissa Gemma, Ryan Heuser, Franco Moretti, and Hannah Walser, "Canon/Archive. Large-scale Dynamics in the Literary Field," *Stanford Literary Lab Pamphlet* 11 (2016).

[9] James English, *The Economy of Prestige: Prizes, Awards, and the Circulation of Cultural Value* (Cambridge, MA and London: Harvard University Press, 2005); Graham Huggan, "Prizing Otherness: A Short History of the Booker," *Studies in the Novel*, 29, no. 3 (1997): 412-433; Chantal Wright, "Before Chamisso: The Role of the Munich *DAF* Writing Competitions and Anthologies in the Promotion of a 'Deutsche Literatur von Außen, 1979-1987," *Oxford German Studies*, 43, no. 1 (2014): 20-36. doi: 10.1179/0078719113Z.00000000046.

[10] Gisèle Sapiro, "Globalization and cultural diversity in the book market: The case of literary translations in the US and in France," *Poetics* 38, no. 4 (2010): 419-439. See also Sapiro, "Translation and Symbolic Capital in the Era of Globalization: French Literature in the United States," *Cultural Sociology* 9, no. 3 (2015): 320-346.

[11] Andrew Piper and Eva Portelance, "How Cultural Capital Works: Prizewinning Novels, Bestsellers, and the Time of Reading," *Post-45* (2016).

[12] A few examples may be informative here. Per Petterson's *Out Stealing Horses* won the Norwegian Booksellers Prize in 2003. It was translated into Danish in 2004, Swedish in 2005, English in 2005, French in 2006, Dutch in 2006, and German in 2006. It won the Independent Foreign Fiction Prize in 2006. Jáchym Topol published *Gargling with Tar* in Czech in 2005 and it was selected as the Lidové Noviny Book of the Year. It was translated into German

in 2007, Polish in 2008, French in 2009, and English in 2010. It was longlisted for the Independent Foreign Fiction Prize in 2011.

[13] Gilles Deleuze and Félix Guattari, *Kafka: Toward a Minor Literature*, trans. Dana Polan (Minneapolis, MN: University of Minnesota Press, 1986).

[14] This list was also filtered to include only works with at least five expressions, which can be either editions or translations. The top ten languages based on these criteria (valid as of December 2018) are: French, German, Italian, Spanish, Swedish, Latin, Russian, Dutch, Polish, and Japanese.

[15] See Matthew Wilkens, "'Too isolated, too insular.' American Literature and the World" (under review).

[16] Casanova also emphasizes the importance of translation, calling it "the foremost example of a particular type of consecration in the literary world" (133).

[17] David Bamman, Sejal Popat and Sheng Shen, "An Annotated Dataset of Literary Entities," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies, Volume 1* (2019): 2138-2144. doi: 10.18653/v1/N19-1220.

[18] Matthew Wilkens, "The Geographic Imagination of Civil War-Era American Fiction." *American Literary History* 25.4 (2013): 803-40; Michael Gavin and Eric Gidal, "Scotland's Poetics of Space: An Experiment in Geospatial Semantics," *Journal of Cultural Analytics* (Nov. 16, 2017). doi:10.22148/16.017.

[19] The phrase "A sense of 'nationess'" is from Homi Bhaba, "Introduction," *Nation and Narration* (London: Routledge, 1990) 2.

[20] Ted Underwood, "Distant Reading and the Blurry Edges of Genre," *The Stone and the Shell* October 22, 2014. https://tedunderwood.com/2014/10/22/distant-reading-and-the-blurry-edges-of-genre/

[21] Locksley L. Messam, Adam J. Branscum, Michael T. Collins, Ian A. Gardner, "Frequentist and Bayesian approaches to prevalence estimation using examples from Johne's disease," *Animal Health Research Reviews* 9.1 (2008): 1-23.

[22] See Matthew Wilkens, "'Too isolated, too insular.' American Literature and the World" (under review).

[23] For an empirical study of language centrality that confirms our ranking, see Shahar Ronen, Bruno Gonçalves, Kevin Z. Hu, Alessandro Vespignani, Steven Pinker, César A. Hidalgo, "Links that speak: The global language network and its association with global fame," *Proceedings of the National Academy of Sciences* 111.52 (Dec 2014): E5616-E5622. doi: 10.1073/pnas.1410931111.

[24] Post-1945 German literature has certainly featured intensive, if sometimes opaque, reflections on Germany. While an early generation of post-war authors was explicitly preoccupied with the legacy of the war and with the division of the country (such as the *Gruppe 47*), younger authors writing after 1989 have often focused on the aftermath of re-unification. Authors from both groups, such as Christoph Hein (b. 1945) and Ingo Schulze (b. 1962), are represented in our corpus, as well as representatives from an even younger generation less burdened by the past and more attuned to global concerns, which may account for this middle-position that we are seeing with respect to the German case.

[25] Éva Bányai, "An East-Central European Success Story. György Dragomán's *The White King*," Acta Universitatis Sapientiae, *Philologica* 1 (2009): 78-89.

[26] Casanova, as mentioned previously, aligns the national/political orientation of small literatures with realism and associates works from the center with an aesthetics of experimentation: "the political dependence of emerging literary spaces is signaled by recourse to a functionalist aesthetic" (199).

[27] Fredric Jameson, "Third-World Literature in an Era of Multinational Capitalism," *Social Text* 15 (1986): 65-88.

[28] Casanova herself suggests a similar telos when she argues that: “the new organization of production and distribution, together with the emphasis at all levels upon immediate profitability, favours the transnational circulation of books conceived for the mass market” (171).

[29] Tim Parks, "The Dull New Global Novel." Franco Moretti made a parallel argument with regard to the international success of action films back in 2001. Franco Moretti, *Planet Hollywood*, *New Left Review* 9 (2001): 90-101. The idea that the market has a purely leveling effect has also recently been challenged by Warwick Research Collective, *Combined and Uneven Development: Towards a New Theory of World-Literature* (Liverpool: Liverpool University Press, 2015).