

# Three Models of World Literature

## Abstract

In the early 2000s, digital literary studies took a new direction. For much of the twentieth century, computational scholars had focussed on details, compiling exhaustive concordances of classic texts, densely tagged databases of stylistic examples, or comprehensive digital editions of famous writers. But in the 2000s, scholars acquired a new aspiration: to data mine the archive, and model world literature as a whole. In this paper I look back on these 20 years of effort, and take stock. What does it mean to ‘model’ world literature? And how is such modelling affected by digital technology? (§1) I contrast three models of world literature: the *canon* (§2), the *cosmopolis* (§3), and the *archive* (§4). In each case, I show how the particular model implies a certain research method, and then consider how that method may be assisted or transformed by digital technology. To conclude (§5), I consider Kath Bode and Roopika Risam’s powerful materialist critiques of digital literary scholarship, and suggest that scholars should embrace *orthogonality* as the underlying principle of world literary history.

## Bio note

Michael Falk is a digital humanist and scholar of eighteenth-century literature. He is an interdisciplinary scholar, who works with computer scientists, designers, historians and linguists around the world, to determine how AI can enrich the study of literature, and how literature can reshape the science and design of AI systems. He has published on colonial Australian literature, the history of the sonnet, design fiction, artificial stupidity, quantitative book history, character network analysis and the evolution of modern tragic drama. He is currently developing a new project, which uses digital methods to explore how Swahili, Hindi-Urdu and Australian English developed as literary languages from 1700 to 1900.

## §1 Making Worlds, Making Models

South Asian literatures present some of the most important frontiers in world literary research. South Asia is home to an extreme variety of different literary, linguistic, and cultural traditions. These traditions present a severe challenge to conventional understandings of literary modernity, as scholars such as Sheldon Pollock and Nirmala Menon have demonstrated (Pollock 2005, 2006, Menon 2016). They also present a severe practical challenge for digital literary scholarship. While scholars have made considerable progress in digitising Indic- and Dravidian-language texts, and the Digital Humanities community is growing rapidly in India in particular, the sheer multilingualism of South Asian literature makes digital analysis difficult. It is only recently that literary scholars have begun to grapple with the challenges of multilingual digital analysis, and the tools required for South Asian languages—OCR and HTR engines, part-of-speech taggers, reference corpora, stopword lists, practical know-how—are just starting to appear.

In this essay, I sketch out a vision for a planetary digital scholarship that can accommodate South Asian literatures in all their complexity. As we develop new archives and new techniques, what should our guiding principles be? Twenty years ago, Franco Moretti famously called for a new kind of ‘distant reading’, which would make the whole ‘slaughterhouse’ of world literature visible for the first time (Moretti 2000). But he and his followers have largely failed to incorporate non-Western literatures into their models of world literature. In what follows, I consider why this is, and suggest that a reformulated version of Moretti’s approach could provide a way forward. I contrast Moretti’s *archival* approach to modelling world literature with two other approaches: the *canonical* approach of David Damrosch, and the *cosmopolis* approach of Sheldon Pollock. I then consider how these different approaches can be instantiated digitally. To conclude, I consider the powerful materialist critiques of Kath Bode and Roopika Risam, and propose *orthogonality* as a methodological ideal for digital scholars of world literature.

The concept of ‘world literature’ can never be settled, because both the ‘world’ and ‘literature’ are constructed, contingent and contained. We humans don’t simply inhabit the world, or read or write literature. We make a world, and make literature. This is the sphere of activity Hannah Arendt calls ‘work’. We work, argues Arendt, to make ‘things’, which

'have the function of stabilizing human life' (Arendt 1958, p. 137). '[T]he most important task of human artifice', she continues, 'is to offer mortals a dwelling place more permanent and more stable than themselves' (1958, p. 152). Such a meaningful 'dwelling place' is a 'world'.

For scholars of world literature, there is a double making at stake:

The first making: Storytellers, poets, listeners and readers create literature in order to make sense of the world, or in other words, to make worlds.

The second making: Scholars unite these diverse literary worlds into a single world—world literature—which they describe in their research.

A 'world' literature is not necessarily universal. The worlds we inhabit are generally small: the household, the nation, the garden, the coterie, the neighbourhood, the desk, the earth. Writers generally invoke one or more of these small worlds in their works. Has any writer ever succeeded in addressing everyone simultaneously? Meanwhile scholars generally aspire to encompass the widest possible world in their constructions of *world literature*.

There are various proposals for the widest possible world, the most influential of which in recent years has been Gayatri Spivak's 'planet'. Her proposal is persuasive because it accepts that no 'world' we make can ever fully encompass reality: 'When I invoke the planet, I think of the effort required to figure the (im)possibility of this underived intuition.' (Spivak 2003, p. 72) The planet is the limit of our world-making, at least until our fantasies of interstellar travel are realised. Everything we make *is* the planet. Each thing we make is an attempt to 'figure' the planet, but the planet always exceeds our figuration. The ideal, '(im)possible' world literature would be a planetary literature—all actually existing world literatures fall within this limit.

Spivak's proposal meshes nicely with theories of 'modelling' that guide research in Digital Humanities. Spivak argues that when a scholar sets out to describe world literature, they are merely 'figure' the ideal planetary literature. Likewise, Willard McCarty argues that when digital scholars set out to analyse literature on the computer, they merely 'model' the ideal literature that lies beyond the computer's grasp.<sup>1</sup> This does not doom digital scholarship. In

---

<sup>1</sup> In fact, he is making a broader argument about modelling *anything*, but his arguments can be applied to literature in this way.

fact, 'the misfit between model and target not only points a way forward but is the point of the exercise.' (McCarty 2005, p. 179) By trying to model literature computationally, the scholar is forced to formalise their assumptions about literature, and test these assumptions against the available data. When the model inevitably fails to encompass reality, the scholar realises the inadequacy of their assumptions, tries to improve their model, and triggers an 'interactive, heuristic process' of criticism and re-modelling (McCarty 2005, p. 81).

A figure is not precisely the same thing as a model. But Spivak's notion of *figure* is broad enough to include McCarty's notion of *model*. A model is just an especially precise figure for something, which can be manipulated computationally. If we accept this synthesis of Spivak and McCarty's views, then questions arise: How can scholars model planetary literature? Or, how can scholars make planetary models of literature? In this essay, I answer these questions by considering three different kinds of model:

1. *Canons*: Selections of texts that transcend their local context because of their supreme literary qualities, cosmopolitan character or global reach. Writers make world literature by writing translatable works; scholars make world literature by translating them.
2. *Cosmopolises*: Communities of writers and readers who look beyond their local context, and address a cosmopolitan audience in a cosmopolitan language. Writers make world literature by making a cosmopolitan language; scholars make the cosmopolis by describing the institutionalisation of that language.
3. *Archives*: Maximally inclusive or non-selective collections of texts. The ideal archive of world literature is simply the set of all texts ever written. Writers make world literature by writing; scholars make world literature by accounting for what is written.

All these approaches to world literature predate Digital Humanities and the rise of literary 'modelling'. As I hope to show, however, modelling has the potential to fundamentally alter how *canons*, *cosmopolises* and *archives* are made (§§2-4). Moretti and the distant readers have generally striven to make digital *archives* of world literature, though as we will see, digital *canons* and digital *cosmopolises* are also of great value in contemporary literary scholarship. Of all these kinds of model, it is the *archive* that seems most likely to embody Spivak's ideal of planetarity, and indeed in *Death of a Discipline* she cautiously concurs with

Moretti's archival approach (2003, p. 101 but see also 107n1). Although it is indeed easy to think of all the planet's literature as a single giant archive, materialist scholars such as Kath Bode and Roopika Risam have demonstrated how incomplete and overdetermined any actual archive is bound to be. In the final section of the essay (§5), I take stock of Bode and Risam's arguments, and propose a new ideal of digital literary scholarship, *orthogonality*.

## §2 Model 1: *Canons*

Let us begin with world literature as a canon of texts. 'Canon' is a controversial word and is often rejected by scholars who adopt this approach. David Damrosch is a prominent example. In *What is World Literature?* (2003), he argues that world literature is an ever-changing set of literary works which are absorbed from one 'national tradition' to another:

... works become world literature by being received *into* the space of a foreign culture, a space defined in many ways by the host culture's national tradition and the present needs of its own writers. (Damrosch 2003, p. 283)

He denies that such works form 'a set canon of texts':

... different groups within a society, and different individuals within any group, will create distinctive congeries of works, blending canonical and noncanonical works into effective microcanons. (Damrosch 2003, p. 298)

In his view, there are as many world literatures as there are reading communities. In my household, we may particularly prize Shakespeare, Hafez and Cao Xueqin. In your household, perhaps Germaine de Staël, Ngugi Wa Thio'ngo and Mir Taqi Mir are the key writers. At this level, Damrosch's theory seems extremely sceptical. There is really no world literature at all, but rather a whole chaotic series of world literatures, spontaneously rising and falling all over the earth as readers' preferences change.

Despite this apparent scepticism, Damrosch does argue that world literature is a particular kind of literature, which can be understood as a single thing. First, he supposes that translation is necessary for world literature to exist, and that therefore only texts that 'gain in translation' are capable of becoming part of world literature. If texts are not translated, they cannot move easily between national traditions in the way Damrosch describes. There are many things that can impede translation: lack of interest in foreign texts, lack of skilled

translators, lack of time and money. But the ultimate limit to translation, argues Damrosch, is the translatability of literary works themselves.

Literary language is ... language that either gains *or* loses in translation, in contrast to nonliterary language, which typically does neither. ... literature stays within its national or regional tradition when it usually loses in translation, whereas works become world literature when they gain on balance in translation, stylistic losses offset by an expansion in depth as they increase their range ... (Damrosch 2003, p. 289)

Is it true that some texts are innately more translatable than others? The statement seems dubious, but is a question for another article (or book! or library!). For our purposes, we can simply note that Damrosch's concept of world literature as the set of 'those texts that gain in translation' implies a certain research methodology: Damrosch argues that the study of world literature requires more, and better translations, so that a wider circle of readers has access to a larger set of world-literary texts. There is a latent world of translatable texts, waiting for the touch of the translator's wand. There is a single, potential, planetary canon of translatable texts. Each 'microcanon' of translated texts is simply a fragment of this ideal planetary canon.

The digital equivalent of Damrosch's *canon* is the *digital edition*. To make digital editions, scholars select particular authors who are particularly worthy to be edited, and expend enormous scholarly resources representing all the folds and complexities of the writer's *oeuvre* in digital form. Following Damrosch's argument, for a digital edition to form part of the canon of world literature, it should contain the work of a 'translatable' writer, and present the translation to the reader. But in fact, this is not what the most interesting and successful digital editions do. Digital editions can be highly dynamic and unstable, and can present a very different kind of 'translation' to the one Damrosch describes. Damrosch imagines that world literature is made up of texts that slip easily from one language to another. But the most exciting digital editions often showcase writers whose work is highly resistant to translation. One of the best examples is Frances Pritchett's digital edition of the great Urdu *ghazal*-writer, Mirza Asadullah Khan 'Ghalib'.

If any writer is 'untranslatable', it is surely Ghalib. His poetry is highly culturally specific. Much of it takes place in what Frances Pritchett calls the 'ghazal world', a universe of

imagery, tropes and concepts which are highly unique and interdependent (Pritchett 1994, p. 104). Moreover, his poetry is dense and structurally complex. He uses particular features of Urdu grammar to maximise the ambiguity and euphony of his verses, and he delights in puns, which often rely on meanings that are very specific to particular Urdu words. In a traditional book-based translation, where space is limited, and readers and publishers expect the book to be pleasant to read through sequentially, the translator has to make stark choices about what aspects of Ghalib's verse to represent in their translation. But in a digital edition, where space is unlimited, and readers expect a more fragmented reading experience, a very different kind of translation is possible.

Here is the translation of the first *sher* of Ghalib's 32<sup>nd</sup> ghazal, from Pritchett and Owen Cornwall's paperback translation of 2017:

When there was nothing, then God existed; if nothing existed, then God would exist.  
Existence itself drowned me; if I were not I, then what would I be? (Ghalib 2017, p. 57)

There are no footnotes to this version, and the Urdu is reprinted at the back of the book in Urdu script. Here is the version from Pritchett's online edition of Ghalib's Urdu ghazals:

nah thā kuchh to ̣hudā thā kuchh nah hotā to ̣hudā hotā  
ḍuboyā mujh ko hone ne nah hotā maiñ to kyā hotā

1a) when there was nothing, then God was; if nothing was, then God would be

1b) when I was nothing, then God was; if I were nothing, then God would be

1c) when I was nothing, then I was God; if I were nothing, then I would be God

2a) 'being' drowned me; if I were not I, then what would I be?

2b) 'being' drowned me; if I were not, then what would I be?

2c) 'being' drowned me; if I were not I, then what would be?

2d) 'being' drowned me; if I were not, then what would be?

2e) 'being' drowned me; if I were not I, then so what?

2f) 'being' drowned me; if I were not, then so what?(Ghalib and Pritchett 2020, v. 32,1)

The reader can swap between four different orthographies: Urdu script, Devanagari, this Roman script with diacritics, or a more basic Roman script. Below the poem, Pritchett

provides seven commentaries (six by Urdu writers and one of her own), with numerous hyperlinks that take readers to related parts of the site. In this digital edition, it is actually Ghalib's *untranslatability* which makes him exciting to read. In the book translation, the translators had to make painful choices about how to string the different parts of the *sher* together to make a comprehensible whole. In the online version, different possible translations proliferate on the page. It's an exciting reading experience, as the reader experiments with combining the different possible renderings, and checking them back against Ghalib's original text, which can be read in the reader's script of choice. The commentaries provide a range of other perspectives, sparking further experiments in literary interpretation. In a way, the online translation is a kind of anti-translation. It demonstrates how hard Ghalib is to translate, and makes the English reader confront their inability to understand Urdu. This results in its own kind of pleasure.

For Damrosch, world literature is process of mutual understanding, a translators move texts from one language to another. Digital editions allow scholars to make a very different canon of planetary literature, founded on mutual bafflement. Digital editions can baffle the reader more effectively than books because of their hypertextuality. '[T]he hypertext,' argues Jerome J McGann, 'unlike the book, encourages greater decentralization of design.' (McGann 2001, p. 71) The text itself becomes decentred: 'One is encouraged not so much to find as to make order—and then to make it again and again, as established orderings expose their limits.' (McGann 2001, p. 71) A digital edition can thus embody Emily Apter's ideal of the 'Untranslatable'. A plural translation like Pritchett's online rendering of *nah thā kuchh to khudā thā* is not a finished, polished translation that allows Ghalib to slip easily into English poetry. It is rather a 'linguistic form of creative failure with homeopathic uses' (Apter 2013, p. 20). Pritchett's failure to find a single good translation for Ghalib's poem increases our appreciation of the text, and encourages the reader to read more sceptically and creatively. Pritchett's digital edition undermines Damrosch's conception of the canon of world literature: translatability is no longer the criterion that distinguishes world literature from merely local literature. With the help of digital editing, any text may be lifted into the planetary canon and transported about the globe, regardless of how easily or reluctantly it slips from one language into another.



Damrosch makes a second key argument to support his idea that there is a single implicit canon of world literature: he argues that there is a single world. The earth is divided, he argues, into discrete 'cultures', 'nations' or 'national traditions':

The modern nation is, of course, a relatively recent development, but even older works were produced in local or ethnic configurations that have been subsumed into the national traditions within which they are now preserved and transmitted.

(Damrosch 2003, p. 283)

Since the world is divided up in this way, there can logically be only one world literature. World literature is simply the set of literary works that have transcended these national or cultural boundaries. Certain texts may circulate more widely than other texts, but still we have a clear criterion that separates the set of world-literary texts from the set of merely national texts. It is very difficult to credit this argument. It is of course true that most of the world's states can be described as 'modern nations', but Damrosch's description overlooks internal divisions within states and transnational institutions that cut across states. The English language, for example, is the 'national language' of a dozen or so countries, an everyday medium of communication in several dozen more, and a useful *lingua franca* pretty much everywhere. When writers such as R.K. Narayan, Alexis Wright, or Abdulrazak Gurnah choose to write in English, they do not address a merely 'national' audience. Likewise, when a Quechua, Arrernte or Shertukpen writer publishes a work in their own tongue, it cannot be said that their work is simply 'subsumed' by the national traditions of Peru, Australia, or India. It seems that even today, the cultural-linguistic world is lumpy, and no clear distinction can be made between local/national and non-local/world literature. Some languages, like English, Urdu, or Haka, transgress national boundaries. Others, such as Quechua or Arrernte, exist independently of any particular nation, and have affiliations that cut across national groupings. Do Quechua and Arrernte belong to the Hispanophone or Anglophone worlds in the same way Peruvian Spanish and Australian English do? Different languages and traditions seem to bear different relationships to the planet as a whole. Some languages have a regional or global reach. Others are more local. This realisation leads to the second kind of model scholars build: the *cosmopolis*, world literature as literature composed in a cosmopolitan language.

### §3 Model 2: The Cosmopolis

Damrosch sees world literature as a canon of translatable texts that transcend national boundaries. Theorists of the cosmopolis define world literature in a very different way. A world literature or 'cosmopolis' is a particular literature which is recognised as a world literature by its own writers and readers. The most famous proponent of this model is Sheldon Pollock. In *The Language of the Gods in the World of Men*, Pollock argues that a 'Sanskrit cosmopolis' arose in South Asia in the first millennium of the common era. In this period, Sanskrit literature was not a literature that belonged to any particular polity or region. It was a 'transregional culture-power sphere', a 'great tradition over against the local or vernacular as the indigenous, little tradition' (Pollock 2006, pp. 11–12). What is most striking about Pollock's study is the way it allows us to compare and classify different kinds of world literature. The 'globalisation' of the last two centuries is only one way that literature can become cosmopolitan. Unlike English or French, Sanskrit was not imposed upon anyone through invasion or corporate takeover, but spread throughout South and South-East Asia through a relatively peaceful process of 'transculturation' (Pollock 2006, p. 133). Unlike English or French, it was not used as a weapon of religious conversion, but seems indeed to have been an essentially secular language during the period of the Sanskrit cosmopolis (Pollock 2006, p. 98). And unlike English or French, Sanskrit did not have a core and a periphery—it was a truly cosmopolitan language, a language of nowhere and everywhere, rather than a metropolitan language, the dialect of a prestigious city such as London or Paris. Pollock provides a strong antidote to some of the nostrums of world-systems theory. In our late phase of globalisation, perhaps English itself has escaped its metropolitan roots, and has become a universal currency that exceeds any of its dialects.

*Canons* and *cosmopolises* imply different research methodologies. To make a canon of world literature, the scholar must select and translate a set of special texts. To make a cosmopolis (or a model of a cosmopolis), the scholar must examine the particular writing and reading practices of a particular historical community. If a particular community use a particular language (or languages) as universal languages, addressing a world audience, then they form a cosmopolis. In fact a cosmopolis is rarely as universal as it seems to its inhabitants. There are probably few scholars today who would accept Sanskrit literature's claim to universality. Other examples are even more extreme. Jennifer Dubrow argues that

Urdu literature formed a cosmopolis in the nineteenth century, even though it was basically confined to British India. This is because Urdu writers and readers formed ‘a transnational language community that eschewed identities of religion, caste, and even class’ (Dubrow 2018, p. 109). Nineteenth-century Urdu was a language confined to a single state. If a language confined to a single state can nonetheless be ‘transnational’, then what exactly is the difference between a ‘national’ and a ‘transnational’ language?

Pollock provides a powerful answer to this question, rooted in classical Sanskrit theories of knowledge:

In short, when the absolute perspective of science (*pāramārthika sat*) is at odds with the representations produced from within the traditions of language thought (*vyāvahārika sat*), it is to the latter that we must defer if we are to understand the history made by knowledgeable agents. (Pollock 2006, p. 66)

To crudely paraphrase: if the writers and readers of Urdu and Sanskrit *represent* their language as a cosmopolitan or transnational language, then it is one. A cosmopolis is a social institution, like a state or a church, which exists essentially by the consent of its members. It is not an object that is independent of human actors, though of course any individual will encounter the cosmopolis as something larger and less changeable than themselves.

The *cosmopolis* puts the scholar in a less powerful position than the *canon*. Scholars can, in principle, create a planetary canon easily, by progressively selecting texts from around the world and translating or failing to translate them. Such work, however inadequate so far, has begun. But cosmopolises of the past no longer exist, and cosmopolises of the present are shaped by large forces. At best, scholars can probably only ever *describe* the workings of cosmopolises past and present, and hope to raise awareness of them among the public.

Digital Humanities has embraced the cosmopolis model, particularly in the field of quantitative book history. Such approaches qualify Pollock’s dictum that only the *vyāvahārika sat* of a cosmopolis can really be known. The *Oceanic Exchanges* project, for example, has developed tools and resources ‘to examine patterns of information flow across national and linguistic boundaries in nineteenth century newspapers’ (Cordell *et al.* 2017). In the nineteenth century, facts, poems, stories or articles that appeared in Calcutta or Sydney one month could appear in Cape Town or Boston the next. The project is able to

study such information flows in great detail by linking together national newspaper archives and using clever digital methods to locate related articles. In one study, the project team first searched for articles from around the world that mention Nikolay Bobrikov, the Russian Governor-General of Finland who was assassinated in 1904. Then they used string-alignment techniques to locate instances of text reuse—in other words, they used software similar to anti-plagiarism software to find when an article had been quoted, plagiarised or reprinted (Oiva *et al.* 2020, pp. 393–394). In this way, they were able to describe the actual workings of a ‘cosmopolis’ in some detail. Pollock relies on close interpretations of scattered inscriptions and manuscript texts to painstakingly reconstruct the Sanskrit cosmopolis. By interpreting these texts, he is able to reconstruct the *vyāvahārika sat*, the way Sanskrit writers understood the cosmopolis of which they formed a part. By contrast, the *Oceanic Exchanges* team use quantitative methods to reconstruct the *pāramārthika sat* of media cosmopolis of the nineteenth century. They describe how far, how fast, and how much writing travelled around the nineteenth-century world. For example, they are able to show that in nineteenth-century Australia, news went ‘viral’ more quickly than in other parts of the world, with a smaller number of news articles spreading further and faster than elsewhere (Oiva *et al.* 2020, p. 401). Really, the two approaches are complementary, and Pollock himself uses some rudimentary quantitative techniques to provide an objective framework for his interpretations (e.g. observing when and where the first and last inscriptions of a particular kind appeared).

There is an aspect of the *Oceanic Exchanges* project that points beyond the idea of the cosmopolis. The information flows that the authors describe are *interlingual*. Bobrikov was a Russian functionary in Finland, yet the news that was first announced in Finnish newspapers (presumably written in Finnish, Swedish or Russian) quickly made its way into French, German, English and Spanish newspapers. In the modern media landscape, facts quickly become detached from words—and words from facts—zipping away along copper wires to the far reaches of the earth. Pollock’s model of the ‘cosmopolis’, like Damrosch’s model of the ‘national tradition’, more or less assumes that a particular literature must be defined by a particular literary language (though Pollock does also discuss the subordinate roles that Prākṛit and Apabhraṃśa played in the Sanskrit cosmopolis). What if world literature

transcends language itself? What if texts form part of wider systems, which are tied together by something other than translation or the sharing of a cosmopolitan language?

#### §4 Model 3: The Archive

This brings us to the third model of world literature, the *archive*. On this model, world literature is an arbitrary collection of texts, and the ideal planetary literature is simply the collection of all the texts ever written. The scholar's task is to examine this massive archive and try to determine how it fits together. This archive model is essentially Moretti's gambit in those influential articles from the early 2000s. Distant reading may change 'how we look at *all* literary history: canonical and noncanonical: together' (Moretti 2000, p. 208). This '*all*', which Moretti himself italicises, is the key. Moretti proposes a radically inductive form of literary scholarship. First take *all* of literature, and take it 'together', in a single archive that stretches over time and space. Second, find the implicit order in the archive, by searching for '*the repeatable element of literature*: what returns fundamentally unchanged over many cases and many years' (Moretti 2000, p. 225). Rather than imposing their own structure on literary history, scholars should let the archive speak to them, as regularities and patterns emerge from the great mass of data.

This is an arresting model, and it has the potential to solve a common problem that bedevils the study of world literature. The problem is that the concepts we use to explain literary history have generally been developed to explain the literary history of Western literatures, and are often unsuitable to describe literature from elsewhere. This problem often places the postcolonial critic in a tight spot, as Nirmala Menon indicates in *Remapping the Indian Postcolonial Canon* (2016):

The critique of the limitations of 'Western' theory (here, deconstruction) is not to elevate a 'nativist' theory but [leads to] the very practical conclusion that the multiple ways of narrativizing multiple postcolonial experiences have probably not been factored into the Western theoretical structure. (Menon 2016, p. 142)

Menon walks a tight line between imperialism and nativism, as she draws on Sanskrit *Bhava-Rasa* theory to try and redefine the act of translation for an Indian context. With her tactful, careful argument, Menon largely succeeds in this particular instance; but Moretti's approach may offer an easier way out. By flattening all the world's texts into a homogenous

archive, Moretti suggests we can reset world literature, and temporarily set aside all the conceptual baggage that so often gets in the way. We can then generate fresh, new concepts of literary history by observing the patterns that emerge from the data as we apply different algorithms to it.

This is potentially a democratic, planetary vision of world literature, which erases the distinction between ‘world’ and ‘local’ literatures entirely. Both the *canon* and the *cosmopolis* presuppose that only certain texts are world literature: only the canonical or cosmopolitan texts circulate globally. But if world literature is simply an *archive*, it becomes possible to think of world literature as a planetary commons, where every literary tradition, every act of writing, every instance of reading, is equally present.

The obvious problem is that there is no planetary archive, no database or library where a scholar can easily access the entire literary heritage of humanity—or those parts of it that survive. The scholars who have taken up Moretti’s call, such as Matthew Jockers, Ted Underwood and Stanford’s Literary Lab, have not had the resources nor the linguistic expertise required to try and sample world literature as a whole, and have devoted their efforts almost exclusively to English literature (Jockers 2013, Underwood 2019).<sup>2</sup> In recent years, the situation has started to improve. There are digitisation projects underway across Asia, Africa and the Americas, and digital humanists in these regions are developing the teams and the tools required to perform the kind of data-rich analysis that has hitherto only really been possible for prestigious European languages. As the practical barriers recede, however, more fundamental barriers to a planetary literary archive emerge.

The first reason is that the archive is uneven. Roopika Risam makes this point forcefully in *New Digital Worlds*. She argues that the digital cultural record is colonialist in two ways: not only as ‘a function of *what* is there—what gets digitalized and thus represented in the digital cultural record—but also *how* it is there—how those who have created their projects are presenting their subjects.’ (Risam 2019, p. 17) Moretti himself provides a good example of this *how*, the way an archive can itself be structured to present the subject in a certain way. In *Graphs, Maps, Trees: Abstract Models for Literary History*, he presents a graph of the ‘Rise of the Novel, 18<sup>th</sup> to 20<sup>th</sup> Century’ (Moretti 2005, p. 6). It is an arresting graph, placing the

---

<sup>2</sup> The Stanford Literary Lab’s series of pamphlets are available at <https://litlab.stanford.edu/pamphlets/>.

history of British, Italian, and Spanish literature alongside that of Japan and Nigeria. Each country appears as a line in the graph, representing the number of novels published in each country in a certain year according to rigorous scholarly bibliographies. It has two axes: the x-axis indicates year of publication, the y-axis the number of novels published that year. The data underlying the graph relies on four key assumptions:

1. The passage of time in literary history is measured in years.
2. The quantity of literature is to be measured by the number of printed books.
3. Texts are to be grouped into national traditions by place of publication.
4. The novel is a (or possibly *the*) crucial genre of modern literature.

While each assumption may be justifiable, the graph as a whole resolves every assumption according to traditional Western models of literary modernity. Books published per year is a unit that only makes sense for a large country with an extensive printing industry—Luxembourg and Samoa are simply unplotable on the graph. Grouping texts according to place of publication only makes sense if the published literature is national in character—but the texts that Moretti groups under ‘Britain’ include Irish, Scottish, and colonial texts. Most problematic of all is the focus on the novel. Moretti’s vision in ‘Slaughterhouse’ was to search the archive for the ‘repeatable elements’ in literature, and to allow a pattern to emerge from the data. But here the ‘repeatable element’ is assumed *a priori*—any literature that falls outside the definition of ‘novel’ simply cannot be included. This makes Moretti’s argument in the piece somewhat circular. He wants to prove that world literature is convergent in the modern world, and cites the ‘rise of the novel’ as a key piece of evidence. All literatures are becoming similar, because they are all becoming ‘novelised’. But by setting up the graph this way, it could *only ever* provide evidence for such convergence. It’s true that such a graph can display the *absence* of the novel in a particular time and place, but if there is a process at work in any of these places that is *not* the ‘rise of the novel’, the graph simply will not show it.

This is what Risam is targeting when she argues that it matters ‘*how* [the data] is there’ in the digital cultural record. The very structure of the archive can predetermine what kinds of question we are able to ask, and what kinds of answer we are able to give. If your archive is an archive of novels, grouped into nation-states, with the year of publication recorded, then the only real question you can ask is the old question of the ‘rise of the novel’. How can

digital scholars ever discover other patterns in planetary literary history with archives like these?

There is a second, deeper problem with approaches like Moretti's—they are too idealistic, too inattentive to the grubby material reality of things. Moretti hopes that if an archive is large enough, then the selection of texts within it will be essentially random, and therefore free from prior interpretation. This is what allows him to clear the deck and generate new concepts for world literary history. But as Kath Bode shows through her close study of Australia's *Trove* database, one of the largest and most comprehensive newspaper databases in the world, even the largest and most comprehensive archives have been edited and shaped by human hands—first the hands of the editors who edited the newspapers, then the hands of the librarians who preserved and catalogued them, then the hands of the computer scientists who wrote the software that digitised them, then the hands of the readers who correct them, and on and on (Bode 2018). All archives have such a grubby, material history, that if Bode is correct, then the time will never come when we can flatten the world's literature and start afresh in the way Moretti describes. Every archive will always be lumpy, encoding not only the 'raw data' of literature, but the inadvertent data of the people who made the archive. What has been preserved? By whom? Why? There are gaps in the data—what is their shape? Even a postcolonial archive, such as Risam advocates, will have this same lumpiness and incompleteness.

There is actually nothing wrong with Moretti's novel graph in and of itself. It is a masterful piece of literary-historical storytelling. But the constraints of this particular graph have become the blind spots of digital literary scholarship. And this undermines the whole enterprise. For ultimately, *canons* and *cosmopolises* rely on *archives*. All our models of world literature lie strung between the *planet* and the *archive*. The planet is the unattainable target of our aspiration. The archive is the inescapable source of all our evidence. If Risam and Bode are right, and even the best archives are radically incomplete, then what is the future for digital literary history?

## §5 Between Worlds, Orthogonality

I would like to conclude by indicating how I think a version of Moretti's vision may still be possible. Where Moretti goes wrong is in his hope to grasp world literature as a whole—'all



literary history ... together'. As Risam and Bode so convincingly argue, this whole is simply not present in the archive, and may never be present. Indeed, arguably there is no whole, no single literary history into which every act of writing can be integrated. Where Moretti goes right is in his hope for a more inductive literary history—searching for the 'repeatable elements', the waves and fluctuations in the flow of words that spark comparisons and analysis. I propose *orthogonality* as the methodological ideal that can guide this search for 'repeatable elements'. In its most homely meaning, *orthogonal* is roughly synonymous with *perpendicular*. More abstractly, it refers to the separation of different features of a system. In practical terms for digital literary scholars, *orthogonality* means the *search for axes*. In Moretti's graph, the three axes were time in years (x-axis), novels published (y-axis) and country of publication (grouping variable). What other axes of comparison can we find to compare and contrast traditions, texts, tropes, words, or characters in the disparate literatures of the planet? In Moretti's graph, the selection of axes was more-or-less determined by the data itself: he derived the graph from a set of bibliographies that listed novels by year of publication divided by country. But as full-text databases proliferate for more and more languages, digital literary scholars will have more and more 'unstructured' textual data at their disposal. What are the measurable quantities of a mass of text? What measurements can we make that apply with equal validity to texts in languages A, B, and C? Will the same measurements still be valid if applied to texts in languages B, C, and D? If I compare the distribution of personal pronouns in English and Hindi texts, to reveal something about the literary representation of subjectivity, how can I then include Italian or Swahili texts in my analysis, when they use personal pronouns so rarely? We are just now approaching the time when such questions can really be asked.

*Orthogonality* presupposes that there isn't a single world with a single literature. It sets aside the goal of creating a single unified model for all of literary history, and sets a humbler goal in its place. If there is no single map of world literature, then there is no single latitude and longitude by which to plot the literary history of the world. We should aspire for a profusion of maps, each revealing the particular gridlines that demarcate that little pocket of literary reality. We may never know the whole, but we can systematise our bafflement. In this way digital scholars can help to figure the (im)possibility of a truly planetary literature.

## Bibliography

- Apter, E., 2013. *Against World Literature: On the Politics of Untranslatability*. London and New York: Verso.
- Arendt, H., 1958. *The Human Condition*. Chicago: University of Chicago Press.
- Bode, K., 2018. *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor: University of Michigan Press.
- Cordell, R., Beals, M.H., Russell, I.G., Nyhan, J., Priani, E., Priewe, M., Salmi, H., Verheul, J., Alegre, R., and Hauswedell, T., 2017. *Oceanic Exchanges*.
- Damrosch, D., 2003. *What is world literature?* Princeton: Princeton University Press.
- Dubrow, J., 2018. *Cosmopolitan Dreams: The Making of Modern Urdu Literary Culture in Colonial South Asia*. Honolulu: University of Hawaii Press.
- Ghalib, M.A., 2017. *Ghalib: Selected Poems and Letters*. Bilingual edition. New York: Columbia University Press.
- Ghalib and Pritchett, F., 2020. A Desertful of Roses: the Urdu Ghazals of Mirza Asadullah Khan Ghalib [online]. Available from: <http://www.columbia.edu/itc/mealac/pritchett/00ghalib/index.html#index> [Accessed 18 Dec 2020].
- Jockers, M., 2013. *Macroanalysis: Digital Methods and Literary History*. Urbana, Chicago and Springfield: University of Illinois Press.
- McCarty, W., 2005. *Humanities Computing*. Houndmills: Palgrave.
- McGann, J.J., 2001. *Radiant Textuality: Literature After the World Wide Web*. Houndmills: Palgrave.
- Menon, N., 2016. *Remapping the Indian Postcolonial Canon*. London: Palgrave Macmillan UK.
- Moretti, F., 2000. The Slaughterhouse of Literature. *Modern Language Quarterly*, 61 (1), 207–228.
- Moretti, F., 2005. *Graphs, Maps, Trees*. London and New York: Verso.
- Oiva, M., Nivala, A., Salmi, H., Latva, O., Jalava, M., Keck, J., Domínguez, L.M., and Parker, J., 2020. Spreading News in 1904. *Media History*, 26 (4), 391–407.
- Pollock, S.I., 2005. *The ends of man at the end of premodernity*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Pollock, S.I., 2006. *The language of the gods in the world of men: Sanskrit, culture, and power in premodern India*. Berkeley: University of California Press.
- Pritchett, F.W., 1994. *Nets of Awareness : Urdu Poetry and Its Critics*. Berkeley: University of California Press.
- Risam, R., 2019. *New Digital Worlds*. Northwestern University Press.
- Spivak, G., 2003. *Death of a Discipline*. New York: Columbia University Press.
- Underwood, T., 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: Chicago University Press.