sciendo
Hungarian Journal of English and American Studies
30.2 (2024): 331-345
doi: 10.2478/hjeas/2024/30/2/5

Searching for Latent River Cultures in English-Language Literature Using Word Embeddings

Dez Miller

ABSTRACT

Throughout the nineteenth and twentieth centuries in the United States and the United Kingdom, rivers and streams were piped, dammed, reversed, straightened, and dried-up, all in service of a growing demand for clean, reliable water in every household. This paper uses an interpretive distant reading methodology for asking how this dramatic change was reflected in English-language literature. As an imaginative space of reflection on culture and material life, how does literature accommodate and make sense of changes in environmental realities? Looking at the diachronic word embeddings surrounding the word "river" in the Novel TM corpus housed within HathiTrust Digital Library, this study identifies a number of trends over time in the shifting semantic fields surrounding "river." It argues that these results indicate a possibly less intimate conceptualization of rivers over time, one more defined by rivers' geographic attributes than by their ecologies and specific natures. (DM)

KEYWORDS: rivers, word embeddings, cultural analytics, computational text analysis, gensim, nlp

Over the course of the nineteenth and twentieth centuries, a process was underway that would result in a drastically different relationship between human beings and the substance most intimately connected to life: water. In many industrialized nations around the world, rivers, creeks, springs, and streams would be systematically piped, dammed, reversed, straightened, and dried up in service of a growing demand for reliable, clean water on demand in every household. After many millennia of daily interaction with naturally occurring water flows, most human beings in urban environments of the US and the UK would come to interact with water primarily mediated through a tap, faucet, or showerhead. Increasingly, with the growth of cities and then subsequent chemical pollution, natural water flows that remained would become highly polluted.

In this paper, I explore an interpretive distant reading methodology for asking how this change in human water cultures was reflected in English-language literature. As an imaginative space of reflection on culture and material life, how does literature accommodate and make sense of changes in environmental realities? By analyzing diachronic word embeddings in the Novel TM corpus housed within Hathi Trust Digital Library, I offer a generalized, randomly sampled portrait of how English-language literature created different semantic fields around rivers over time. This study focuses on the word "river" for several reasons: 1) It appears in the vast majority of novels in the corpus, between 442 and 2749 times per decade-corpus (see Figure 1). 2) It lacks the multivalence of a word like "stream," which can be a verb. 3) Rivers are and have historically been repositories for cultural values and power dynamics. 4) Rivers have served as water sources for a huge number of cities, and they have been the object of extreme measures of control, pollution, and extraction as city populations have grown exponentially over these last two centuries.

Background

In order to know what to look for in literary landscapes, it is useful to understand the infrastructural realities that underpinned cultural meanings of rivers. Starting in the 1800s, most American rivers were just beginning to be employed for water infrastructure, with early waterworks just starting out in the most advanced northeastern cities (Melosi 19). In cities like London, on the other hand, waterwheels were established as early as 1581 (Broich 2). By the 1800s, however, both cities in the UK and the USA were facing similar problems: with the exponential growth of cities, and an increase in waterborne diseases, a demand for more sanitary cities was growing (Melosi 12). So began the more large-scale infrastructure projects, which would fastidiously bury most streams and creeks in urban environments into sewers underground.

Major cities in the US and the UK took slightly different tactics for controlling water flow depending on their geography, social realities, and eras in which they undertook these projects. Early waterworks in northern US cities usually relied on steam engines and wooden pipes to pump water from the closest rivers available. These projects were similar to the ones that would be untaken eventually by southern US cities almost a century later (Pierce). Initial waterworks were largely unfiltered until basic sand or gravel filters began to gain popularity. As cities grew, rivers and lakes quickly became polluted as street runoff and sewage ran directly into the same

sources where water was drawn from. Until the twentieth century, it was little understood what exactly made a water source safe to drink. Martin V. Melosi notes a telling piece of guidance from the president of the New York Board of Health in 1873: [A]lthough rivers are the great natural sewers, and receive the drainage of towns and cities, the natural process of purification, in most cases, destroys the offensive bodies derived from sewage, and renders them harmless (57). As an understanding of water-borne illness slowly dawned, sewers and more advanced freshwater pumping systems were built to a greater and more ambitious extent. In the USA, the number of waterworks went from 244 in 1870 to 9,850 in 1924 (Melosi 82). Along with this came more ambitious waterworks—canals, dams, and aqueducts—to help source water from more abundant flows, farther away from where sewage was sent into the river. Throughout the Great Depression and beyond, huge public works projects would reshape rivers and redefine their ecologies. One such example was the Columbia River, the extensive damming of which would nearly decimate the salmon run, as famously elucidated in Richard White's *The Organic Machine*. These public works and modern water systems, as is apparent in the 1941 Woody Guthrie's *Columbia River Songs*, were often culturally coded as symbols for Western man's dominance over nature (Bonneville Power Administration).

With underground sewage systems came the beginning of the end for natural streams and creeks in the heart of downtowns. As cities were paved, these smaller water flows were controlled to an even greater extent, and stormwater became an increasing issue. Perhaps nowhere was stormwater a greater issue than in Los Angeles, with the fickle Los Angeles River, which ran sometimes in a rush and other times dry, and which altered its route with every storm. The river, which allowed for the founding of the city, would become a menace to the expensive new real estate, flooding the city in 1914, 1934, and 1938 before miles of concrete canal were laid, to straighten it out and make it more predictable (Gumprecht 3). In other cities, the control of water meant dredging marshes. In *Cities and Wetlands*, Rob Giblett profiles a number of cities built on wetlands, which were then drained. He argues that these wetlands remain sites of repression for city identities themselves, as well as sources for metaphors that continue on in a city's life-long past the "end" of the wetland.

Outside of water infrastructure, what role did rivers play in cultural narratives? And why should it matter? As Tracy Scott McMillin explains in his book *The Meaning of Rivers: Flow and Reflection in American Literature*

> What rivers have meant can help us think about what rivers do mean and perhaps what rivers might mean. . . . Many scientists, including Luna Leopold, believe that the people of the United States "have acquiesced to the destruction and degradation of our rivers, in part because we have insufficient knowledge of the characteristics of rivers and the effects of our actions that alter their form and process." (xviii)

Part of that insufficient knowledge, I argue, comes from a marked decrease in intimacy with rivers, creeks, and streams, due to the control of water for infrastructure. Another aspect of this lack of intimacy has to do with the kinds of stories we tell about rivers and river flows. McMillin divides river stories into being defined by their distance to the "river's energy." They position themselves as either 1) "overlooking the river"; 2) being "by the river"; 3) going against the flow, or "up the river"; 4) going with the flow, or "down the river"; 5) crossing the river; or 6) going "up and down the river." Many of McMillin's categories speak to rivers' roles in travel and transportation. Upriver trips were facilitated significantly by the invention of steamboats throughout the nineteenth century before ultimately being supplanted by train travel (Burton et al.). While travel still occurs on rivers in the US, it is frequently more in the form of shipping barges than human transportation. Given that change, it is worth wondering whether rivers are still associated with travel in the cultural imagination, and if the kinds of travel rivers are associated with mirror historical realities.

This essay deviates from other research like McMillin's in that it is interested in narratives that are not explicitly "river stories," but cultural products that contain rivers as setting, background, or marginal elements. In the aggregate, how are rivers coded semantically? That is, what are they associated with, and how, like the wetlands in *Cities and Wetlands*, might they show up as marginal, polluted, repressed, or with an emphasis on their unimportance? This research also responds to calls within more materially minded ecocriticism to "think with water" as a way to reveal formerly ignored locations of power, modes of response, or methods of relation (Chen et al.).

Finally, ever since the activism leading up to the 1972 Clean Water Act in the United States, there has been a growing movement to clean up rivers, as profiled in Paul Stanton Kibel's *Rivertown: Rethinking Urban Waters* and other books. Therefore, another research question was how this return of attention to the ecology of rivers might show up in the stories we tell about them over the last few decades?

Corpus selection and research questions

Because of copyright restrictions, legally employing digital humanities text mining methods on most twentieth and twenty-first-century novels is mainly possible through the HathiTrust Digital Library. This Library contains over seventeen million volumes, digitized by partnering library collections (Underwood et al.). While 3.2 million of these volumes are public domain, HathiTrust also makes available its full-text volumes for "non-consumptive research" (HathiTrust) including text mining through their Data Capsule, a secure, virtual computing environment. For this project, I used Ted Underwood et al.'s curated dataset, which is one of HathiTrust's "Recommended Worksets," entitled "NovelTM Datasets for English-Language Fiction: Manually Checked Subset." This collection is a 2,730-volume randomized subset of a larger 210,776-volume list, which was identified as fiction "by trawling, predictive models" (12). The NovelTM subset was then manually checked by Underwood et al. to make sure that each title was indeed fiction and had accurate metadata attached.

Volumes between 1800 and 2009 were selected in order to capture the period in which water infrastructure developed most rapidly. I then divided the list by decade and built models for each decade within a HathiTrust Data Capsule. Each of these decade lists were around 130 volumes. It is worth mentioning that this is not meant to be a "representative" corpus of the past. Representativeness is something that computational literary scholars have hotly debated, and a random sample of a digital library that does not contain all titles cannot represent the totality of English-language fiction. There are other subsets of this NovelTM corpus that might have been chosen for this inquiry instead, including a subset of frequently reprinted titles, which some have argued better represent the past. The authors of the NovelTM corpus note that their compiling of the corpus goes against the recommendations of other digital humanists like Katherine Bode, who in *A World of Fiction,* recommends corpora in which the context of circulation can be well-understood.

For this reason, the following distant reading can be seen more as a roadmap for future inquiries in trying to understand how historical realities in river control have affected the roles of rivers in narratives. My overall research question, therefore, is whether or not it is possible to identify trends in the semantic fields of rivers even in a general corpus like the NovelTM subset I used. One of the purposes of this study, therefore, is to test out the usefulness of generalized corpora like the NovelTM Corpus for

identifying trends over time. I wanted to know whether it was possible to determine certain ways of talking about rivers that were possible in the 1800s, which, with the changes in historical realities, became impossible in the 2000s, or vice versa.

Methodology

Word2Vec is a Gensim word embedding algorithm that uses shallow, two-layer neural networks to place each word in a corpus within a vector space model. For digital humanists who use word embedding models, a word's particular connotative meaning can be represented by the words that the seed word is close to in vector space. These spatial relationships are determined by the context of a word across a particular corpus, as well as by the context of related words. This means, effectively, that even if "river" very rarely appears within the direct context of a particular word, for example "forest," it may still be considered by the Word2Vec algorithm to be highly semantically close to the word "forest" as long as another closely related word, for example "stream," appears frequently in the context of "forest."

The idea that a word is represented well by its context has been explored by language theorists of the past several centuries. Examples include the maxim attributed to John Rupert Firth, "you shall know a word by the company it keeps!," (Firth 11) and Jacques Derrida's refutation in *Of Grammatology* of the structuralist idea of the signified, arguing that behind every signifier is a chain of signifiers which constitutes the meaning of the word.

A common operator with Word2Vec is getting cosine similarity results between two words. These results are, theoretically, both physically closer in vector space and *semantically* closer in meaning. For this inquiry, I asked for the top twenty cosine similarity results for the word "river" in each decade-specific corpus. I then removed and identified the proper names from this list (Table 2) and kept the top ten similarity results (Table 1).

Parameters

Word2vec has a number of parameters, and these help to determine the kind of semantic results that are possible to glean. There is no one "correct" way to do Word2Vec, but different parameters offer different kinds of results. Below I detail the reasoning behind my choices for each of these parameters.

Window. The default window value for Word2Vec is 5, meaning that the context for a word is five words before and five words after. Generally, the guidance is that smaller context windows give similarity results wherein the similar words are *interchangeable*, whereas larger windows (15 or more) give results that are more highly *related* (Konstantinovskiy). This can be a somewhat perplexing spectrum, given that those two words—*interchangeable* and *related*—are not opposites. While multiple windows were tested, I chose a window of 25, given that that would represent approximately the previous and following sentence and a half in relation to a word. It allowed the model to get at some attributive qualities and highly related words that were particular for each corpus rather than producing interchangeable words that might be true for many corpora.

Skip-gram vs CBOW. Word2vec has two types of modeling: skip-gram and CBOW (Continuous Bag of Words). While CBOW trains by predicting a word from context words, skip-gram does the opposite: it predicts context words from a single input word. Skip-gram is known to perform better with a smaller dataset, which would describe the approximately 130-volume corpora I was working with.

Min-count. I chose the default minimum count of five, meaning that words that appear four or fewer times will be discarded from the training data before training occurs. The logic behind this default minimum count is that words that appear four or fewer times will not have very accurate or meaningful word vectors, since their context may be overly limited.

When using quantitative tools for literary analysis, it is important to recall that the tools were not necessarily designed for literary methods. In theory, setting a minimum count to 1 instead might be useful for literary analysis, because even if a word's context is overly skewed by a novel or passage in a novel, it might be a useful lens through which to do literary interpretation on that novel. For example, if I am studying a novel or group of novels that only uses the word "river" four times in total, similar words may not make a lot of sense as far as being interchangeable to "river." Let's say the word "lunch" is a similar word to "river" in that corpus. That would indicate a not very "accurate" model, given that "lunch" is not a semantically similar word to "river." However, the dissonance between those two words might invite a new research question: are people in this corpus eating lunch frequently by rivers? And if so, how are rivers being framed as a site of recreation? For this particular inquiry, however, it was more useful to glean general results than results determined by a novel or set of novels. Therefore, the minimum count chosen was 1.

Results

1800	1810	1820	1830	1840
banks	navigable	banks	banks	banks
rocky	coasted	slope	coasted	stream
declivity	serpentine	creek	stream	rivulet
moon-light	banks	meandered	lake	southeast
craggy	southerly	stream	dammed	rapids
tinkling	declivity	fordable	narrows	creek
woody	skirted	rivulet	widens	southerly
repassing	rivulet	streamlet	empties	water-gap
conies	creek	sloped	south-east	reconnoitering
moss-grown	bluff	footpath	rafts	valley
1850	1860	1870	1880	1890
stream	stream	banks	stream	banks
creek	*wooded*	stream	northerly	stream
widens	fordable	pebbly	*estuary*	foothills
ferried	banks	affluents	confluence	barges
banks	shallowed	hilly	tributary	rapids
rapids	foot-bridge	rivulet	barges	cascades
lake	precipitously	widens	boat-house	*wooded*
fording	gorge	wharves	cascades	rafts
islet	slopes	water	northward	inlet
bayou	creek	mountain	gravelly	*freshet*
1900	1910	1920	1930	1940
fording	stream	banks	tributary	northernmost
stream	sloped	stream	wooded	inlet
rapids	cut-off	confluence	down-stream	piers
headwaters	upland	tributaries	flows	weir
sweetgrass	tributaries	zigzagging	waterfall	navigable
banks	*cornfields*	fording	stream	ferries
coulées	narrows	creeks	sampan	bluffs
south-east	*grassy*	delta	cliffs	*marshy*
cañon (canyon)	bayou	gravelly	gorges	southward
tributary	promontory	shelving	uplands	gorges
1950	1960	1970	1980	1990
stream	*swampy*	forded	tributary	upstream
barges	downstream	down-stream	rapids	riffles
sampans	creek	lowlands	*wooded*	expressway
willows	upstream	traversing	creek	gorge
narrows	spanned	stream	banks	rapids
embankment	northeast	upstream	fishing	downstream
ferry-boat	banks	headland	stream	marshlands

2000
Tributary
Turbid
Banks
Upstream
Creeks
Waters
Stream

"By the River"
River as Boundary
Transportation
Navigation
Ecological Resonance
Directional/
Mapping

rivulet	lough (lake)	lakes	upstream	**bridge**	Downstream
creeks	westwards	banks	half-submerged	banks	Creeks
juts	stream	shallows	lagoon	***mangrove***	Silty

Table 1.
Top ten similarity results for "river," proper names excluded

1800	1810	1820	1830	1840
Wye (U.K.)	Baltic (northern, sea)	Aar (Switzerland)	Potomac (U.S.)	Saluda (U.S.)
Rhone (U.K.)	Dwina (Russia)	Mernene (fictional)	Landinsburgh (U.S., reservoir)	Champlain (Quebec)
Euphrates (Middle East)	Ob (Russia)	Mersey (U.K.)	Hadley (U.S., waterfall)	Gavarnie (France, village)
Loire (France)	Tagus (Spain)	Augustine-Bay (U.S., bay	Tahquamenon (North America, bay)	Saliko (U.S.)
Rhine (Germany)	Indus (Central Asia)	Neckar (Germany)	Trent (Ontario)	Portage (U.S.)
Tigris (Middle East		Kishon (Israel)	Saco (U.S.)	
		Slaney (Ireland)	Chaudière (Quebec)	
		Kei (South Africa)	Flesk (Ireland)	
		Manderra (fictional)	Fishkill (U.S., creek)	
		Tamar (U.K.)		
1850	**1860**	**1870**	**1880**	**1890**
Susquehanna (U.S.)	Rough (Ireland, possibly fictional)	Adur (U.K.)	Harlem (U.S.)	Lucerne (Switzerland, town)
Dee (U.K.)	Sarapiqui (Costa Rica)	Medway (U.K.)	Pend d'Oreilles (U.S.)	
Tiber (Italy)	Nightach (Ireland, possibly fictional)		Elbe (Germany)	
Seine (France)	Caftan (U.K., pool)			
Moskow (Norway, island)	Earmonth (U.K.)			
Neckar (Germany)	Grimsel (Switzerland, mountain pass)			
1900	**1910**	**1920**	**1930**	**1940**
Teton (U.S.)	Mississippi (U.S.)	Dordogne (France)	Braes (Jamaica)	Allegheny (U.S.)
Catawba (U.S.)	Missouri (U.S.)	Sauty (U.S.)	Bogongs (Australia, region)	Avon (U.K.)
Crois (U.S.)	Platte (U.S.)	Terek (Caucacus)	Taronga (Australia, park)	Cowford (U.S., fictional)
Rockies (U.S., mountains)		Vézère (France)		Rannals (U.S., fictional)
Pend (U.S.)		Tinto (Spain)		Paddock (U.K., town)
Meeker (U.S., town)		Avon (U.K.)		Canaan (U.S.)
Musselshell (U.S.)				Matanzas (U.S., inlet)
				Severn (U.S.)
				Haly (fictional)
				Tamplin (mountains, fictional)
1950	**1960**	**1970**	**1980**	**1990**
Erie (U.S., canal)	Abati (Tanzania)	Styx (mythical)	Thames (U.K.)	Varada (India)
Chenab (Central Asia)	Illawarra (Australia, region)	Suong (Laos)	Nile (Northeast Africa)	Portage (U.S.)
Guyas (ecuador)	Eisak (Italy)	Rhone (France)	Ganges (India)	Wandsboro (U.S., town)
Diz (fictional)	Arkansas (U.S.)	Louthe (fictional)	Ota (Japan)	Kalang (Wales)
Paducah (U.S.)	Struilitsa (Bulgaria)	Bori Khan (Thailand, city)	Ouse (U.K.)	Mekong (East Asia)
Tamiami (U.S., canal)	Chesapeake (U.S., bay)	Evoron (Russia, lake)	Mocobila (Honduras, fictional)	Meuk (Laos)
	Nid (U.K.)	Muang (Laos, city)		
2000				
Amaria (unclear, possibly fictional)				
Exe (U.K.)				
Koel (India)				
Vistula (Eastern Europe)				
Taff (Wales)				
Ganges (India)				
Thames (U.K.)				

Table 2.
Place names within top 20 similarity results by decade
Note: these are all river names unless otherwise noted

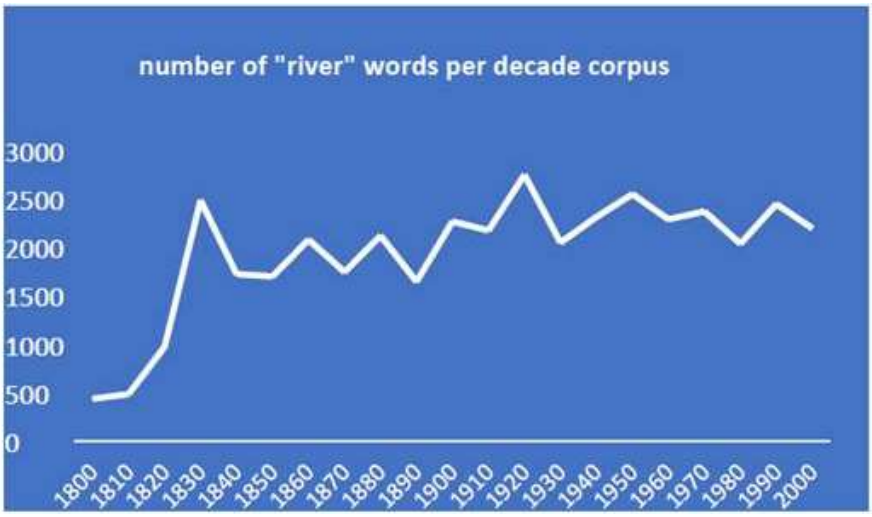

Figure 1.
Number of "river" words per decade corpus

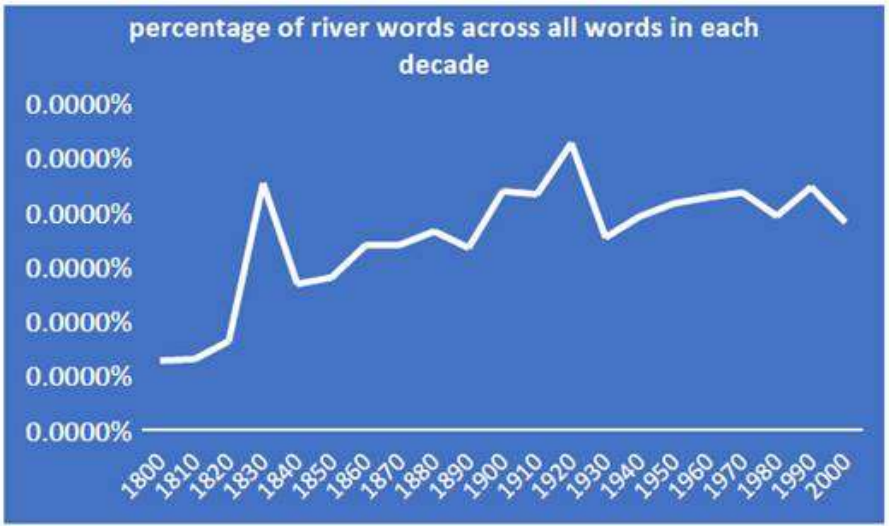

Figure 2.
Percentage of "river" words across all words in each decade corpus

Discussion and limitations

Several categories of words emerged from the results. Inspired by Tracy Scott McMillin's descriptions of river stories, "By the River," words, that is, words that could only come from direct observation of the river and

its surroundings, emerged as one category. (Note: I excluded very common results like "banks," which appeared in nearly every decade-corpus within the top 30 results). Perhaps the most significant finding was that noticeably more of these words appeared in the first half of the 1800s than in later decades, with seventeen words in the first half of the nineteenth century, and only six words in the latter half of the twentieth century. This might be indicative of a mimetic response to the reality discussed in the "Background" section of this paper, in which authors became less and less intimate over time with naturally occurring water flows due to the increasingly comprehensive water infrastructures, and so represented them directly less often in literature.

Words within this category like "moon-light," "tinkling," and "moss-grown" (1800s) each speak to an experiential closeness to the river, as well as, perhaps, a romantic relationship to the riverine environment. Other words within the "By the River" category include words that seem to describe the nature or shape of the river flow, like "serpentine" (1810), "meandered" (1820s), "narrows" (1830s), and "widens" (1870). A few words within this category described the specific nature of the river bottom, its banks, or surroundings: "pebbly" (1870s), "gravelly" (1880s and 1920s), and "silty" (2000s). "Silty" required further investigation, as silt pollution is a significant issue in North American rivers. However, upon looking at the novels in which this word appeared, it became clear that these were exclusively novels set on the Asian continent, which tends to have naturally siltier rivers even without pollution (Gordon).

This speaks to a shift in the geographical content of this dataset over time, which is also apparent in Table 2, showing place names that appeared in the Top 20 cosine similarity results. For the first hundred years of the dataset, almost exclusively American and Western European place names appear, with some Middle Eastern or fictional place names as well. By the mid-twentieth century, place names from Asia and Africa began to appear much more frequently, reflecting, perhaps, a general shift in what is catalogued as English-language fiction in different eras. It is worth underlining that these are not the only locations, or even rivers, that appear in the corpora. These are only the locations that the algorithm deemed to be significantly tied to the seed "river." Therefore, while this pattern may indicate an increase in English-language fiction that is not set in North America, the UK, or Europe, this is not sufficient evidence, and other methods, like Named Entity Recognition (NER), would be stronger tools

for investigating the question of how different geographical areas are represented in this dataset.

Again continuing with McMillin's categorizations of river meaning, a "River as Boundary" category emerged, which contained words that appeared to contextualize rivers as boundaries to be crossed. Many of these were related to fording rivers, as in the 1820s, 1860s, 1900s, 1920s, and 1970s decades. The "ferry"-related words, though coded as "transportation," could also be included in this category. The fact that by the last decades rivers no longer show up as to be contended with speaks to the level to which infrastructures all but eliminated in the human psyche this natural quality of rivers to contain or resist human movement.

There was a similar pattern to "Transportation/Navigation" words, which did not appear after the 1950s, and which seemed to mostly mirror transportation from their historical eras. The one transportation-related word that did appear after the 1950s was "expressway" (1990s), which speaks more to the experience, perhaps, of seeing rivers while traveling on expressways rather than relating to river-transportation itself. Words like "navigable" (1810s, 1940s) and "reconnoitering" (1840s) speak slightly more to a small or non-mechanized boat experience, whereas "ferried" (1850s), "ferry-boat" (1950s), and "barges" (1950s) connote larger-scale boats with engines. The word "sampan" appeared in the 1930s and "sampans" in the 1950s. A small Chinese or Malaysian boat, this might speak to the inclusion of colonial narratives, or perhaps the increase in narratives written in English and set on the Asian continent. I was surprised to find that sampan did appear quite a bit in these decades: twenty times across five different novels in the 1930s and sixty-eight times across six novels in the 1950s.

Words of the "Ecological Resonance" category also decreased slightly over time, but to a lesser extent than with the other categories. This category was defined as words that indicate some connection between rivers and other parts of the ecosystem. Therefore, words that related to trees, as appeared in the 1800s, 1860s, 1890s, 1930s, 1950s, 1980s, and 1990s, were noted. Additionally, words that related to sensitive ecological spaces were included in this category, like "estuary" (1880s) and "swampy" (1960s). There were also words that indicated a relationship between rivers to plant life, such as "moss-grown" (1800s), "sweetgrass" (1900s), and "cornfields" (1910s), as well as one animal word, "conies" (hyrax, 1800s).

The one category that showed a remarkably different pattern from the others was "Directional/Mapping." There were markedly more of these words in the latter half-century corpora than in earlier decade corpora.

These were words that seemed to refer to rivers as landmarks as a way to locate other places. These included cardinal directions, as well as words like "upstream" or "downstream." The frequency of "upstream," appearing five times in the latter five decades, and "downstream," appearing five times in the latter six decades, was surprising, and would be interesting to explore through future close reading. A research question for that close reading would be whether, in addition to being a directional indicator, it might reveal an anxiety about what water pollutants are up or downstream of a given location.

It is possible to read an unfortunate reduction in meaning of rivers over time in these results. While in the 1800s rivers' semantic fields were rich with references to particular ways of flowing, to plant and animal species, and to particular geologies, by the 1990s and 2000s, they are represented somewhat more generically. Beyond the categorizations I have laid out here, it is possible to read a general decrease in specificity over time. Have totalizing water infrastructures limited the meaning of rivers in the contemporary era to places on a map?

Altogether, these are preliminary results and might be best conceived as a guideline for future close readings. Digital work often works best in tandem with close reading, as has been elucidated by digital humanists like Andrew Piper and Richard Jean So, among others. In future work, I intend to perform close readings of several novels within this corpus in the post-45 period in order to gain a more nuanced context for what these patterns that I have identified might mean for a cultural analytic understanding of what rivers mean. What can be gleaned from this exploration is that Word2Vec is a useful heuristic tool for conceptualizing general resonances of rivers in different eras. In future digital ecological readings, it would be interesting to explore more geographically curated corpora to see whether material changes in human-river relations specific to particular locations can be tracked onto literary semantic imagination. It is also clear, however, that rivers have occupied, and continue to occupy, a significant space in fiction. Rivers not only showed up quite frequently in these novels, but both the number of mentions and percentage of "river" words across all words increased over time (see Figures 1 and 2). While this study may point toward a less intimate relationship with rivers being reflected in the cultural imagination over time, it does not point to them disappearing entirely from the zeitgeist. Additionally, while in some senses literature is mimetic of historical realities of the human-river relationship, resonances from earlier eras persist. This speaks to the particular role of

literature in culture, which may be either mimetic or a space in which reality is constructed.

Emery University, Atlanta, USA

Dez Miller is PhD Candidate in the Department of Comparative Literature at Emory University, Atlanta, USA.

Works Cited

Bode, Katherine. *A World of Fiction. Digital Collections and the Future of Literary History*. Michigan: U of Michigan P, 2018. Print.

Bonneville Power Administration. "Woody Guthrie's Columbia River Songs in the Columbia" (1949 Film). Web. 14 Oct. 2023.

Broich, John. *London: Water and the Making of the Modern City*. Pittsburg: U of Pittsburgh P, 2013. Print.

Burton, O., et al. "The Golden Age of the Steamboat, 1851–1900." *Northern Illinois University Digital Library*. Web.

Chen, Cecilia, et al. *Thinking with Water*. Montreal: McGill-Queen's UP, 2013. Print.

Firth, R. John. *Papers in Linguistics*. London: Oxford UP, 1957. Print.

Gordon, Stewart. "Major Asian Rivers of the Plateau of Tibet: The Basics." *Education about Asia* 15.3 (2010). Web. 3 Dec. 2023.

Gumprecht, Blake. *The Los Angeles River: Its Life, Death, and Possible Rebirth*. Baltimore: John Hopkins UP, 1999. Print.

HathiTrust. "Non-Consumptive Use Policy." Web. 20 Nov. 2023.

Kibel, Paul Stanton. *Rivertown: Rethinking Urban Rivers*. Cambridge, MA: MIT P, 2007. Print.

Konstantinovskiy, Lev. "Text Similarity with the Next Generation of Word Embeddings in Gensim." *PyData Berlin 2017*. Web. 14 Oct. 2023.

McMillin, Tracy Scott. *The Meaning of Rivers: Flow and Reflection in American Literature*. Iowa: U of Iowa P, 2011. Print.

Melosi, V. Martin. *The Sanitary City: Environmental Services in Urban America from Colonial Times to the Present*. Pittsburgh: U of Pittsburgh P, 2000. Print.

Pierce, Morris. "Atlanta, Georgia." *Documentary History of Atlanta Waterworks*. Web. 14 Oct. 2023.

Piper, Andrew. *Enumerations. Data and Literary Studies*. Chicago: U of Chicago P, 2018. Print.

So, Richard Jean. *Redlining Culture: A Data History of Racial Ineqaulity and Postwar Fiction*. New York: Columbia UP, 2021. Print.
Underwood, Ted, et. al. "NovelTM Datasets for English-Language Fiction, 1700–2009." *Journal of Cultural Analytics* 5.2 (2020): 1–30. Print.