

**DRAFT DO NOT
CIRCULATE OR QUOTE
OR REPRODUCE**

White Mythologies
Richard Jean So (Chicago)

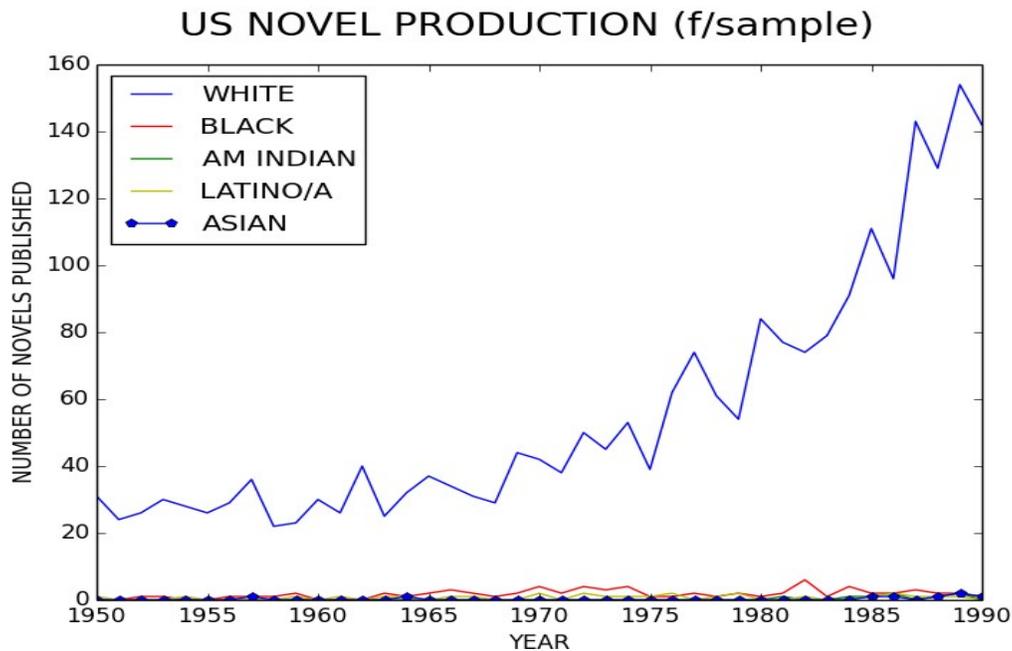
A Quantitative Provocation

The postwar American novel is *white*. By this, I mean that the number of novels written by authors who identify (or are identified) as racially white vastly outweighs the number of writers who do not. The number of novelists who appear on best-seller lists are disproportionately white. The number of novelists who are reviewed in influential magazines and newspapers, and later, appear in anthologies of writing, are also disproportionately white. Along axes of commerce, prestige, and quantity, the postwar American novel is white. Literary scholars are understandably not happy about this. We tend to deal with this somewhat uncomfortable fact in one of two ways. One, *implicit recognition with self awareness but without intervention*. Say, if one wanted to write a new literary history of the American novel in the few decades after the war in relation to religion, the fall of the middle-class, or “the crisis of man,” one might have 4-5 chapters, and the chapters would probably feature white novelists like Updike, DeLillo, Pynchon, and O’Connor, and one non-white novelist – Ralph Ellison. This is an unhappy ratio. To say that in the general unfolding of the history of the American novel after the war, non-white writers contribute a mere 20% of that story is to say something very unsatisfying. However, 20% is better than 0%. The other approach to this quandary is *an implicit critique*. Here, we might think of scholarship that exclusively examines minority fiction. Despite the relative sparseness of novels produced by racial minorities in the early decades of the second half of the century, one could still write a history of postwar US culture by writing it through this corpus of minority fiction. And many indeed have done this. We can think of this scholarship as interventionist insofar as it takes an unacceptable reality of the literary field and disrupts it by imagining a different way to perceive it.

What we have is a problem of *inequality*. This inequality looms in the background of our critical investigations of the postwar novel, but we don’t have a good conceptual way to deal with it, or even talk about it. When I read accounts of the formation of the postwar US literary field, the word rarely comes up (amazingly, it does not appear a single time in either Morris Dickstein’s *Leopards in the Temple* or Mark McGurl’s *The Program Era*), and I think this is part of the problem. “Inequality” is of course a problem that is lived and felt socially, politically, and culturally. But in its first articulation, it is a quantitative problem. $X > Y$. It’s one thing to say that the divide between the rich and the poor in America is vast and increasingly untenable. It’s another thing to say the wealthiest 10% of households in America control 76% of the wealth in the country. The latter gives the scholar a certain concrete traction in approaching the former.

So, say we could do the following: identify the 10,000 most commonly held novels at US libraries published between 1950 and 1990. This list would index some combination of prestige and popularity: the most commonly held novels in libraries represent books that people want to read, as well as books people should read, as determined by “experts” in the field of literature (professors, librarians). Now, once you have that list, say you could identify the racial identity of many of those writers based on official records, biographies,

public information, scholarship, and self-identifications. (And while we're at it, let's also grab information about their gender, where they went to college, where they lived most of their lives, whether they got an MFA and/or a NEA grant, religion, etc.). I think this list of books, combined with the information about their authors, gives a pretty decent view of the field of American literature after the war (admittedly, a "sample," but a very good sample selected on reasonable principles). My collaborators and I did this. What would you guess is the racial breakdown of this list? I asked colleagues and the average guess was 75%. No one said more than 80%. The actual result: *the American novel from 1950 to 1990 is 95% white*. And even more striking and counter-intuitive: *it gets worse as we move closer to the present*. Here is a rather terrifying graph:



That's my quantitative provocation. In writing histories of postwar American literature and culture, I'm not sure we can keep ignoring this inequality. It's been a blindspot that we have longed felt but have not been able to quite name. It's one thing to say that white authors have a kind of hegemony in the US literary field and it shouldn't be that way; it's another thing to say that white writers constitute 95% of the field.

Language that is White and not White

I suspect this graph will startle most of my colleagues. But I also suspect it'll read as coarse, "merely sociological," and even a bit dull. It tell us things about how what kinds of people, and how many of them, wrote novels yet it doesn't tell us about the things we actually care about, such as language, narrative, and aesthetics. These "facts", however precise and disturbing they are, are potentially irrelevant for our critical work. A strong assumption of our discipline is that the force of language and representation within the domain of the aesthetic and cultural precisely exceeds and deforms what are felt to be somewhat crude "on the ground" or "sociological" structures of determination, such as how many white people wrote books versus how many non-white people wrote books. This is

the justification, I think, for writing a literary history of postwar American culture by closely interpreting only four or five novels by four or five novelists.

And this gets to a likely objection to my original critique of what I find to be an *unhappy ratio* in contemporary scholarship regarding race. I implicitly accuse, say, Michael Szalay of believing that black writers merely contribute 20% to the general story of post-war US culture if he only includes one black author out of a total of 5 or 6 in a book ostensibly about race and the novel after the war. In truth, this is a somewhat facile accusation. Szalay precisely argues that black culture and language had a particularly powerful effect in saturating white culture such that black and white bodies became inextricably intertwined. I think this is right. You might say then: *your numbers tell us nothing about the actual dynamics of language and meaning in the wider literary field.* Does it say anything more than ‘lots of white writers’? *No.*

But say we could do the following: we could get our hands on the actual texts of those 10,000 novels and try to find patterns of language and meaning in them. The question is: is there a relation between the vast racial inequality within the field of US cultural production and the way that white versus non-white writers express themselves in language? Does the fact that the American literary field is awfully crowded with white people condition the way that a non-white or white writer writes books?

The real question is: *at scale*, is there even something cogent and identifiable as “white” language versus “non-white” language? At scale, beyond a small cohort of writers such as Pynchon, Updike, and Cheever, do the mass of writers in the 1970s who identify as racially “white” actually possess a vocabulary distinctive from non-white writers that we can then identify as a “language?” This is where computation helps. First, I selected a sample of 1000 texts from, say, the 1970s that included ~900 novels by white authors and 100 non-white authors (that’s low but remember, the overall number of non-white writers in this decade is still miniscule, look at the graph!). I used a sample to enable quicker computation and the sample was randomly chosen. Next, I wrote a computational algorithm that does the following: it’ll look at all the words in the 1000 novel corpus (about 80,000), throwing out words that are too common (like “the”) and too rare (like “cephalopod”), and figure out which words most differentiate the books by white versus non-white authors. It wants to figure out how many words and which words does it need to know in order to perfectly distinguish one type of book from the other. If it sees the word “beauty,” does that word help it to decide whether it is a novel written by a white person? And if it does, how many more words does it need like “beauty” to *always and consistently* distinguish the two types of books? It eventually will come up with a finite list of distinguishing words and it has a way to test its own accuracy.

Here is the list of words for the 1970s.

people, woman, black, end, street, children, quickly, laughed, watch, stand, legs, hit, yeah, hurt, ain't, everybody, sister, skin, reach, dressed, lie, master, sweet, strength, easily, cause, placed, mountain, falling, natural, soul, laughter, realize, studied, ancient,

hills, image, poured, faint, fuck, students, colored, dad, dirt, fucking, quarter, weak, win, freedom, rush, reality, object, terror

53 words. These are the words that separate the races within novelistic language. A writer articulates his or her disposition towards or against whiteness, his or her implicit affiliation with a constructed category of racial identity, based on how he or she uses these words and in what relative combination and frequency. What's most striking is that the machine indicates that with these 53 words alone, it can determine the racial identity of an author with *a 99% accuracy*. This says: the novelistic language that indexes a segregation of different racial categories of writers can be known.

A Methodological Digression

This, I think, is interesting, but will perhaps again strike my colleagues as rather coarse and even obvious. In the history of American race relations, it is precisely language that has indexed patterns of social and cultural division and domination. Our best historians of racial hegemony in the US, like David Roediger and Eric Lott, have typically adopted, following Raymond Williams, a "keyword" approach to reconstructing the forms of authority that white culture has exercised over black subjects. It is language that enables one to best identify the "constructions of identity" that one group enforces upon another, as well as (as later scholars have shown) attempts to build counter-constructions as a measure of resistance. Roediger and Lott track a series of "white power" keywords. Say, if one finds in a newspaper article in the late 19th century the word "coon" or "buck" in a non-ironic or critical context and without quotation marks, one can be fairly sure that the author belongs to a certain racial class and ideological disposition. This approach is very efficient and feels intuitively correct. If one today meets for the first time a person who uses the words "goddamn Asians" and "anchor baby" unironically and non-critically (and without "hand quotes" of course), one could reasonably assume that the person falls into a certain racial class, bears certain ideological leanings, and is a "racist."

However, you'll notice that my list of "white" and "non-white" words does not include obvious keywords of race, such as racial epithets like "nigger" or vernacular words, such as "jessup," which might indicate a text's membership to a certain racialized group. I'm not so much interested in racial behavior, as indexed in novels, as an explicit phenomenon – as one group so obviously exercising force and hatred over another group to maintain racial hegemony and discursive dominance. Rather, I'm interested in racial identity – the perception of racial identity – as a mood or disposition that permeates US society in ways that exceed more obvious flare ups of struggle and conflict. The keyword approach is very effective in identifying and tracking a racial group's most self-aware and organized articulations of control or resistance to another group; it maps the sites of conflict extremely clearly. It doesn't capture, though, the ways that those articulations also can live in more ordinary language, sublimating into language that evades linguistic obviousness but performs just as forcefully. I don't need to tell you that one can still be a racist even if one has never used the word "nigger" or "chink."

Here's a quantitative way to think about this: for the 1970s, we have about 1000 novels, or 100,000 unique words. One approach would be to go looking for all the instances of "gooks" in these texts to flag a group of novels as "white nationalist." But again this approach would lead to a predictable outcome: books by white nationalist authors tend to use a word explicitly marked as white nationalist or racist. It doesn't tell us much about the diffusion of a sensibility or disposition. So, instead, let's try to identify a cluster of words that indicate membership to either a white or non-white group of writers. First, we need to cut out common words that both groups might use without distinction: these are so-called "stopwords," i.e. "the," "a," "what," etc. But we also want to cut out words that are too uncommon. In truth, words like "nigger" or "gook" in American literary culture are on the whole rather rare, probably because *they too strongly* index an ideology. Even a raging racist who doesn't care what you think of him or her will not use the word "nigger" in daily interactions precisely because it is too strong. So we are left with a set of words that are not extremely common or uncommon. These are words that *all* writers will use commonly. But what matters is *the frequency* and *the combination* by which the writers use them that will reveal membership to a category. In a statistical frame, the use of language has a baseline probability distribution; how you, as an individual, use your words puts you somewhere in that distribution between 0 or 1, where 0 is white language and 1 is non-white language. What I describe here is basically what my algorithm does for our sample of 1000 novels published between 1970-1979.

I want to name the cluster of words that arise from this process *mythologies*. I take this concept in part from Roland Barthes. I'm drawn to his idea of "myths" as composed of individual linguistic signifiers (words) that in combination aggregate to a higher level of meaning, or what he calls "ideas-in-form." Myth, he says, essentially "robs" words of their individual signifier meaning to rearrange them in relation to other words or images, and in doing so, they become "transformed" into a greater form. Barthes likes myths, as I do, because they suggest that the really forceful semantic entities in society are these clusters. While we don't really know how they came together, they nonetheless wield tremendous force that far exceeds whatever individual meanings they had as solitary words. All we know is that like a bright and warm light in winter, different categories of writers gravitate toward them, perhaps unknowingly, yet inexorably.

The Inexorable Force of Literary Whiteness

Myths are different from keywords. They are a latent substrate of consciousness. They permeate speech and language in unknowing ways; they permeate society, traveling far and wide, pouring into the edges of places they shouldn't be, yet are. Because they are not explicitly marked words, but rather, constellations of ordinary words that in seeming random combination mark something much more, they precisely take on a "viral" quality that exceeds intention or awareness. I'm interested in mythologies associated with race that emerge and disseminate in the literary field after 1950.

I ran the following experiment. First, I identified the words that separate white and non-white writers in the 1950s based on a sample of about 500 novels. Again, I identify this cluster of words as a "mythology" of racial distinction. Here:

eyes, brother, bar, beneath, American, tears, self, shouted, nature, admit, reminded, angry, cause, image, voices, dying, holy, singing, weak, prison, youth, drink, mama, ain't, leg, dance, folks, bread, understand

Note that we have less words than for the 1970s. This means that we need less words in total to distinguish white and non-white writers in the 1950s, likely because there are simply less non-white writers but also, because white and non-white writing is more distinct and thus requires less words to define. *The myth is simpler in form.*

Now, as with the 1970s, this “myth” can nearly perfectly (99% of the time) distinguish 1950s white and non-white writers, which itself is remarkable. It indicates the strength of this myth in separating the literary languages of the different races. But say we could do the following: take this 1950s myth and see what kind of force it has in the 1960s in terms of identifying differences between white and non-white writers. That is, do the myths of the 1950s persist into the next decade, and if they do, how do they partition the literary field of this decade into white and non-white? Which 1960s writers, white and non-white, gravitate towards the white aspects of this mythology of the 1950s and which gravitate toward its non-white aspects? This might tell us something about the diachronic force of white or non-white language in directing the shape of the American field of literary production, the persistence of certain ideas. I built a statistical model to do this work. It goes through a large sample of the 1960s corpus (here about ~900 novels) and identifies which novels, in terms of their quantitative distribution of language, gravitate more towards the white or non-white aspects of the 1950s myth. It then reports back three categories: books by white writers that it classifies as “non-white”; books by non-white writers that it classifies as “white”; and books by non-white writers that it classifies as “non-white.” Here are the novels that represent the outer extremes of each category, the books that definitively belong to each category:

white as non-white	non-white as white	non-white as non-white
A Wizard of Earthsea (Le Guin 1968)	Hog Butcher (Fair 1966)	This Child's Gonna Live (Wright 1969)
If the South Had Won the Civil War (Kantor 1960)	The Martyred (Kim 1964)	The Free-Lance Pallbearers (Reed 1967)
Trout Fishing in America (Brautigan 1967)	Sissie (Williams 1963)	Seduction of the Minotaur (Nin 1961)
The Arrows of Hercules (de Camp 1967)	Cotton Comes to Harlem (Himes 1965)	
A Bad Man (Elkin 1967)	City of Night (Rechy 1963)	
Where the Boys Are (Swarthout 1960)	Catherine Carmier (Gaines 1964)	
Something Wicked this Way Comes (Bradbury 1962)	Been Down So Long it Looks Up to Me (Farina 1966)	
Imaginary Friends (Lurie 1967)	Of Love and Dust (Gaines 1967)	
The Ticket that Exploded (Burroughs 1962)		
Spencer's Mountain (Hamner 1961)		
Galactic Pot-healer (Dick 1969)		

The Hunter and the Trap (Fast 1967)		
-------------------------------------	--	--

Some obvious things, some not so obvious things. The first category indicates spaces in the literary field where “non-white” language seems to seep into white writer aesthetics, and the results suggest that it in part does through genre fiction. Note the incidences of science fiction (Le Guin, Dick, and Bradbury). But also, note the examples of straight popular fiction: *Trout Fishing in America* and *Where the Boys Are*. This, I think, is where white language gets “blackened” or at least less white. Does this proceed or merely follow the “blackening” of white language we see in Norman Mailer as presumed avant gardes of this effect? I have to run more tests.

The next category I find quite interesting. It interests me because it doesn’t make sense to me. Broadly, it indexes the force of white language through the 1950s to the 1960s on emergent minority authors. In some ways, it uncomfortably confirms what scholars have intuited: that it is Asian-American and Latino/a writers in this period who appear most likely to assimilate into a white literary discourse. We have John Rechy and Richard Kim, the latter who studied at the Iowa Writer’s Workshop under Engle, and deliberately tried to mimic the writing style of Roth and Updike. However we also have writers like Ronald Fair and Chester Himes, who we hardly suspect of such assimilation; we expect the reverse, really. I’m not sure what to do with this. This is where close reading and historical research become essential after this “distant reading.” Overall though, we can say the following: there are a lot of writers, from various racial backgrounds, who are classified as “white” – deeply acted upon by the 1950s white mythology – and this, to me, indexes the diachronic force of such mythologies, how it pulls to its fold even the most seemingly reluctant to be assimilated writers, like Fair. This is not to label any particular writer as assimilationist: it’s to describe at scale how a certain structural inequality in the literary field morphs into discursive effects.

And the last category is perhaps the most interesting. These three texts represent the novels in the 1960s that are *relentlessly* resistant to any incursions of white language. Their language is utterly unlike the white language of the 1950s. In a sense, we might think of these novels as representing a category of “the emergent” (in the phraseology of Raymond Williams) insofar as they indeed represent a tiny minority of this era’s fiction, but produce an incipient voice that is defined by its total otherness to a subsuming white language, a language that even captures the likes of Chester Himes.

Finally, as for the overall field dynamics: 8% of writers in the 1960s were identified as strongly belonging to the category of “white writers identified as non-white”; 6% as “non-white identified as white”; and 1% as “non-white identified as non-white.” Also, for what it’s worth, these numbers stay the same in the next decade.

What Does a White Mythology Look Like?

For all its precision, I suspect that these results too will seem excessively coarse or “merely sociological” for my colleagues. It also aggressively posits some claims which precisely in their stark definitiveness (Chester Himes writes more like a “white” person

than a “non-white” person) are likely to be seen as merely inflammatory. The problem, I suspect, is that the entire process of “discovery” is (in Latour’s terminology) “black boxed” by the machine. What does it even mean for a writer to gravitate towards “white language” or “non-white language” in this decade or any other?

Let’s take a look at a specific example of a novel written by a Black writer that has been classified by the computer as “white”: Chester Himes, *Cotton Comes to Harlem* (1965). Himes wrote the novel after he moved to Paris in the 1960s; it represents an example of his popular “Harlem Detective novels,” which feature a pair of African-American detectives solving crimes based primarily in Harlem, New York City. *Cotton Comes to Harlem* begins, indeed, in Harlem and ends in Africa. Its cast of characters features “Grave Digger” Jones and “Coffin Ed” Johnson, the two Black detectives Himes had become well known for creating, and its villain is Deke O’Malley, a crooked Reverend who attempts to trick legions of Harlem dwellers to “Go Back to Africa” and leave America. Just with the barest description of the plot, few readers – expert or otherwise – would confuse this for a “white novel” or an African-American novel that seeks to be “white” in readerly orientation. How do we possibly explain the computer’s classification?

The mystery thickens when one reads the first few pages: the text is dialogue heavy, and the dialogue consists significantly of Black characters speaking in Black vernacular or dialect: “Here I is been cooking in white folk’s kitchens for more than thirty years.” The setting demands a high density of Black characters, and the high density of Black characters demands a density of language marked as African-American.

We can simply reject the machine’s conclusions; or we might use its findings as an occasion to build a new conceptual framework by which to rethink the very idea of “racial assimilation.” When one reads the entire novel, one is struck by the high ratio of narrative description; although there is a fair amount of dialogue, there is significantly more description, particularly by the end. And what one discovers is that the description is largely shorn of language that might mark it as “non-white,” such as dialect. Instead, it reads like traditional detective fiction, and here, this is what the machine might be picking up on. If detective fiction has historically been practiced by white writers, and there are parts of *Cotton* that require a mastery of that descriptive form, perhaps that style is in fact subtly “white” in linguistic form, and that is what the machine detects.

For example, during one intense scene: “Their legs were tied together like their arms but their feet touched the floor. They were straining with arched bodies and gripping feet to push each other into the wall. The chairs slid on the concrete floor, back and forth, rocking precariously. Arteries in their necks were swelled to bursting, muscles stretched like frayed cables, bodies twisting, breasts heaving, mouths gasping ...”

By my lights, this is a passage that could easily appear in a novel by Raymond Chandler. My first thought is relatively straight forward: Himes is actively assimilating aspects of the detective novel form, and in doing so, he himself is assimilated into the form. My second thought though is more provocative: if this form has largely been dominated by white male authors, such as Chandler, Himes is also unconsciously assimilating into a form of “white” writing; the assimilation of genre has merely disguised the potential racial assimilation. The dynamics in the novel are complex: there are elements that clearly mark it as a work

of “African-American fiction,” and elements that appear to compel it towards a literary whiteness. What matters though – if we take the machine seriously – is that by the lights of the computer, the latter *decisively* overwhelms the former.

This, I think, is “what a white mythology looks like.” It’s something that exists at the edge of more obvious language, something disguised by more explicit dynamics of writing, something that only gains clarity when taken out of a pattern, and assumes force precisely as itself in certain combinations – but remains inexorably “white.”

Conclusion, More Provocations

We obviously know a great deal about the post-45 American novel, but in some ways, we don’t really know enough. We don’t know much about how the entire field of literary and cultural production has evolved, and how various majority and minority groups have emerged in relation to each other and interact. Through historical contextualization and close reading, we’ve long inferred dynamics – quite well I think. But, the lack of a kind of empirical and measurable precision – *95% white* – obstructs deeper insights into the full nature of obvious inequalities and why those inequalities persist.

The above work I present is admittedly quite rough and still speculative. I admit that a serious weakness of this essay is my monolithic idea of “whiteness,” which needs to be refined. I believe though that in creating such reductions, we can begin to see some useful wider dynamics, even as we lose nuance.

Broadly, in writing new histories of American culture and literature after the war, I’m not sure we can keep ignoring such vast structural inequalities in the field. We need to understand better the empirical fact of a culture dominated by a majority racial group, and its relationship to the broader ecology of words. How does this majority group exercise a hegemony of what kinds of things can be expressed in that ecology? How do minority groups nonetheless find ways to disrupt this hegemony, and facilitate (again in Williams’ language) the “emergence” of new forms of expression? Today now more than ever we live in an “ecology of words” populated by different racial groups. And now more than ever, each group vies for the right to be heard. Who gets to talk, who has the right to write, and who gets to be heard is inseparable from the question of one’s racial identity, and increasingly, it is a question of life and death.

My provocation is this: the rise of ethnic and minority writers in American literature after the war laid the foundation for the emergence of discrete cultural publics based on racial identity but our scholarship has partly misapprehended that emergence because we have not been able to see it at scale and at once. The “unhappy ratio” of say 1 out of 5 authors in a major study of postwar culture as not white, what scholars might feel embarrassed by, *actually quantitatively over-represents non-white writers*. And while research in Ethnic Studies has produced invaluable work in recovering these literary traditions, it has potentially *obscured the full extent of the quantitative hegemony of white writers in the literary field*. The 95% figure is real. Its related white mythologies are real too. If today we are startled by the unremitting amount of white racism on social media, and the persistence of the “mainly white room” in US culture, it might be because we have over-estimated the rise of minority cultures in the 1970s and 1980s.