

# Voice

Nika Mavrody, Laura B. McGrath, Nichole Nomura, Alexander Sherman

**Nika Mavrody**, Stanford University  
**Laura B. McGrath**, Temple University  
**Nichole Nomura**, Stanford University  
**Alexander Sherman**, Stanford University  
 Peer-Reviewer: Sarah Wasserman  
 Data Repository: 10.7910/DVN/1NTRIF

## ABSTRACT

The concept of the writer's voice is central to the way that contemporary literature is read, evaluated, circulated, and criticized, appearing everywhere from the creative writing classroom to online reader reviews. Yet, voice remains a slippery and tendentious concept: Is voice something a writer has, or is it something a writer is? Does everyone have a voice? Are some voices voicier, and how? What form does voice take on the page? How is voice different from style or constrained by genre? In this essay, we track voice's many meanings across a large corpus of what we call "vernacular literary criticism." First, we consider the ways that voice is used in different communities and the writers it is used to describe. Second, we develop a conceptual model of voice's many uses, based on our reading of a (limited) version of our composite corpus. Finally, we build a word-embedding model to track voice's use in a larger discourse. Ultimately, we show that voice, style, and genre operate in a unified vernacular critical system, that voice (along with genre) is a subcategory of style, and that voice consists of the parts of style not otherwise captured by genre.

If trade publications like *Publisher's Weekly* and *Writer's Digest* are any indication, there is no more desirable a trait for a would-be writer than voice.<sup>1</sup> Not talent or technique, not craft or style, not a meticulously plotted story or glittering prose: *voice*. Agents want to represent writers who have it: "I tend to be seduced by voice, so voice-driven fiction and nonfiction are high on my wish list."<sup>2</sup> Editors champion it: "If a new voice speaks to you, persist in your crusade on behalf of that writer."<sup>3</sup> Critics are likewise spoken to and seduced by voice. The *New York Times* alone christened "the voice of a generation" 96 times in just as many years. Literary organizations, such as PEN America, celebrate voice when recognizing "emerging writers whose voices are adding to the literary experience," or hosting public programs designed to "amplify . . . emerging and marginalized voices."<sup>4</sup> And voice sells. The hashtag #OwnVoices is used to designate books written by and about members of historically marginalized communities.<sup>5</sup> To sum up: voice is a quality of writing — an aspect of style — but it is related to lived experience and identity. A writer is a voice, but only certain types of writers are voices — Voice of a Generation, on one hand, Emerging Voices, on the other. And, despite (or perhaps because of) the stubborn inconsistency and downright contradictory nature of the

concept, voice is central to the way that contemporary literature is read, evaluated, circulated, and criticized.

This should come as no surprise to scholars of contemporary literature. As Mark McGurl has argued, the imperative to “find your voice” was both “a scholarly-pedagogical and an artistic preoccupation” of the creative writing program of the 1960s and 70s, and it came to dominate “postwar fiction as a whole.”<sup>6</sup> McGurl traces the concept’s genesis to Wayne C. Booth’s rejection of New Critical impersonality, in which voice emanates from a unifying Implied Author. Dorothy Hale summarizes this new definition of voice as “authentic self-expression of identity that is integral to and inevitable in any act of novelistic communication.”<sup>7</sup> Critics attached the term primarily to writers associated with the Program Era’s “high cultural pluralism.” But just as that category was capacious enough to include such differing writers as Philip Roth, Flannery O’Connor, and Toni Morrison, so voice is a concept that adheres diverseley, ubiquitously. Voice, in Hale’s understanding, exists in all writers and obtains to all writing. But whatever was intended by the enjoinder to “find your voice” in the Program, the concept has had an active extracurricular life.

The evolution of this concept *outside* of the creative writing classroom might well attest to the significance of McGurl’s thesis: the Program has been so pervasive that its practices and logics have become commonplace. The focus of this essay is not the Program, but the extracurricular application of its vocabulary. Despite the proliferation of accounts of literary institutions and the institutionalization of literary studies, we know little about the ways that readers — Janice Radway’s “general readers,” Laura Heffernan and Rachel Sagner Buurma’s “common readers,” Merve Emre’s “bad readers” — have responded to these formations outside of the academy.<sup>8</sup> Voice’s transformation shows us that whatever knowledge institutions produce is both received as and recycled into something quite different. Literary criticism has a very active life outside of the seminar and away from the purview of peer review, replete with its own discourses and vocabularies. Such criticism contains vital literary-sociological information that, as Emre notes, has the potential to “generate new theories about reading and the social logics it reproduces.”<sup>9</sup> Attention to readerly networks of production and meaning-making is just as crucial for large-scale computational literary criticism as for literary sociology — and just as rare.<sup>10</sup> For both literary sociology and cultural analytics, the reader’s absence poses a methodological and a critical problem.

In this essay, we examine voice's conceptual development and usage in vernacular literary criticism. "Vernacular criticism" was coined by Bryan Droitcour in 2014 to describe the work of amateur art critics who review museums, installations, and exhibits on Yelp; Lisa Nakamura used the phrase independently, noting the centrality of digital media platforms, such as Goodreads, in the proliferation of this discourse.<sup>11</sup> These online forums, Scott Selisker has argued, have hastened the uptake of concepts such as the Bechdel Test — an analytical framework that is significant to readers but generally not taken seriously by scholars. Though not unrelated to Aarthi Vadde's formulation of "amateur criticism," vernacular criticism does not hinge on the critic's professional status. Vernacular critics may write in the *New York Times* or *Bustle.com* or on *Reddit* or all of the above. (Indeed, well-known authors maintain their own Goodreads accounts, regularly post reviews, and engage with other Goodreaders; and as Selisker has noted, scholars with formal academic training increasingly participate in vernacular critical networks.<sup>12</sup>) Instead, we adapt the concept of vernacular literary criticism from Miriam Hansen's influential concept of "vernacular modernism," used to describe the persistence of modernist style in "mass-produced, mass consumed phenomena," such as cinema, fashion, and architecture.<sup>13</sup> Like Hansen, we mean to emphasize discourse: the language that surrounds these concepts — its use, transformation, connectivity. And following Selisker, we examine vernacular literary criticism on its own terms, as — a discourse concerned with its own modes of evaluation, replete with its own critical idiom.

Voice, a slippery and tendentious concept, makes for a useful case study. In what follows, we consider the uses of voice across vernacular literary criticism, toggling between methods (counting, reading, modeling) and objects of study. Beginning with a working definition of voice based on Hale's, as an "integral," "inevitable," and "authentic self-expression of identity" in a novel, we ultimately find that voice provides readers a unitary framework for grappling with problems both textual and paratextual. We begin by tracking voice's many meanings across a large corpus, considering high-level statistics about its usage in different communities and the writers it is used to describe. Second, we develop a conceptual model of voice's many uses, based on our reading of a (limited) version of our composite corpus. Finally, we build a word-embedding model to track voice's use in a larger discourse. Ultimately, we show that voice, style, and genre operate in a unified vernacular critical system, that voice (along with genre) is a subcategory of style, and that voice consists of the parts of style not otherwise captured by genre.

## A Good Voice is (Not) Hard To Find

Voice: an “authentic self-expression of identity that is integral to and inevitable in any act of novelistic communication.” We adopt this definition from Dorothy Hale — seemingly the most agreeable, and not entirely inconsistent with many popular usages of the term briefly sketched above. But Hale says nothing of form or style, nothing to suggest voice can be read or perceived by Goodreads users or reviewers in the *Times*. Voice is a type of “self-expression” that is present in “act[s] of novelistic communication” — what form does it take? While Hale would likely assert that everyone has a voice, PEN America and #OwnVoices use the term as an imprecise nominalization for “underrepresented writers” or “minoritized writers.” Is voice something a writer has, or is it something a writer is? *Does* everyone have a voice? Are some voices *voicier*, and how?

These are not questions that individual examples, no matter how interesting, can satisfactorily answer. Our century-spanning, composite corpus (see table 1) is comprised of vernacular criticism produced by different communities (reviewers, authors, and common readers), published in different forms (book reviews, literary journalism, author interviews, Goodreads), and in response to different types of literature (canonical, prizewinners, bestsellers).

Corpus	Review Publication Dates	Number of Texts	Notes
American Periodicals Online Review Corpus	1900-2003	1,698	Book reviews from U.S. newspapers, magazines, and literary journals; augmented with newspaper data from the Literary Lab newspaper corpus.
<i>The Paris Review</i> Author Interviews	1953-2019	318	Interviews with 311 authors, translators, or publishers.

Blurbs	2013-2018	26,139	Blurbs used in publishers' seasonal catalogs, 2013-2018. Blurbs for 748 books.
Goodreads Reviews: Canon	2013-2019	77,913	User-generated reviews of the 250 texts compiled in Literary Lab Pamphlet #8, "Between Canon and Corpus" (Algee-Hewitt and McGurl).
Goodreads Reviews: Prizewinners	2013-2019	105,993	User-generated reviews of the prizewinners and nominees for the National Book Award in Fiction, the National Book Critics Circle Award in Fiction, and Pulitzer Prize for Fiction.
Goodreads Reviews: Bestsellers	2013-2019	116,293	Bestsellers from 1953-present, according to <i>Publisher's Weekly</i> .

Table 1. Composite corpus of Vernacular Criticism.

*Professional Book Reviews:* From the American Periodicals Online series (ProQuest), we gathered book reviews from popular magazines (like *Cosmopolitan*) and literary journals (like *The Dial*). To this, we added book reviews from the Literary Lab's historic newspapers corpus, which includes publications such as the *New York Times* and the *Los Angeles Times*. We also gathered, by hand, *New York Times* reviews of prizewinning and bestselling novels that were not available from ProQuest (see below). Together, these corpora amount to 1,164 book reviews, published between 1900 and 2003, by professional reviewers.

*Author Interviews:* *The Paris Review* is a well-known quarterly literary magazine that has been publishing lengthy interviews with authors since its founding in 1953. This corpus contains the full text of every interview published by the magazine — 318 interviews with 311 authors, translators, or publishers.<sup>14</sup>

*Blurbs:* These blurbs were scraped from seasonal catalogs from the Big Five publishers, used to advertise books published between 2013 and 2018. While many of these blurbs have been printed in newspapers that may be in our APO corpus, we believe that these excerpts are differently useful, indexing what publishers believe is meaningful praise. This corpus is more recent than all of our others, with all books under consideration published since 2013. While Goodreads reviews (below) were also published recently, books under review can be much older.

*Reader-Generated Book Reviews:* Our largest collection is comprised of Goodreads reviews. Goodreads is a valuable resource, but unwieldy due to its scale. We developed three subsets of novels (nearly 1,500 books) and scraped the first 300 reviews for each: bestsellers, prizewinners, and canonical novels. Bestseller data was gathered from *Publisher's Weekly*. Prizewinners include books that won and were shortlisted for the National Book Award, the Pulitzer Prize for Fiction, and the National Book Critics Circle Award.<sup>15</sup> Our list of canonical novels comes from the Literary Lab's twentieth-century corpus, which was generated from six lists (the development of this corpus is detailed in Algee-Hewitt and McGurl, "Literary Lab Pamphlet #8: Between Canon and Corpus").<sup>16</sup>

Corpus	Number of Reviews	Number Voicey	Percentage Voicey
American Periodicals Online Review Corpus	1,698	196	11%
<i>The Paris Review</i> Author Interviews	318	285	89.6%
Blurbs	26,139	955	3.6%
Goodreads Reviews: Canon	38,933	2,027	5.2%
Goodreads Reviews: Prizewinners	105,993	5,372	5%
Goodreads Reviews: Bestsellers	116,293	1,499	1.2%

Table 2. Some data about the use of "voice" in each corpus.

Across all of our corpora, one consistent trend emerged: voice is used primarily in discussions of literary fiction, rather than genre fiction — prestigious fiction, rather

than popular fiction.<sup>17</sup> We'll begin with the professionals (table 2). 11% of reviews in the APO corpus mention voice at least once. However, the *New York Times* alone accounts for 70% of those uses. Given the centrality of *Times* reviews in the larger book review ecosystem, it is hardly surprising to see the term popularized here, especially as only a small group of in-house reviewers are employed at the *Times* at once. Even more shocking is the frequency with which voice appears in *The Paris Review*: a staggering 89.6% of interviews mention "voice" at least once, and the concept is invoked by 78% of authors interviewed.<sup>18</sup> This result is perhaps predictable: the writers profiled here are the closest to the MFA seminar, whether graduates or employees.<sup>19</sup>

Though professional reviewers frequently discuss voice, their comments are infrequently excerpted for advertisements. Our blurb corpus contained very few mentions of voice: only 3.6% of blurbs for books published between 2013 and 2018 praise a writer's voice in catalogs or on a cover. Voice is used most frequently for more literary works (as opposed to genre fiction): 45% of the blurbs that mention voice are for works classified by publishers as literary fiction, with the remaining 55% split between genres; no other genre receives as large a share of mentions.<sup>20</sup> This distinction likely carries over into book reviews, as well, in the form of a selection bias for the types of books reviewed by major outlets. In other words, critics at the *NYT* might invoke voice often because they are reviewing more literary fiction than genre fiction. Literary fiction might be an "anti-genre genre," as Matthew Wilkens has described it, in which "it is difficult to find any broadly agreeable set of . . . feature[s] by which literary fiction might be consistently identified."<sup>21</sup> But the vernacular literary-critical discourse surrounding literary fiction suggests that voice may be one of those features distinct to, or at least more pronounced in, literary fiction.

This finding is consistent with our Goodreads corpus as well, though somewhat complicated owing to the amateur status of reviewers and the variety of works reviewed in our corpus. Reviewers on Goodreads do not use the term as often as professional reviewers, nor are they as withholding as publishers. The Goodreads reviews of both canonical novels and prizewinning novels invoke voice at a comparable rate; it is mentioned at least once in approximately 5% of reviews. Yet voice does not appear to be an especially commonly used phrase in the bestseller corpus. Either Goodreads users comment on voice only when discussing a certain type of prestigious novel — prizewinning and canonical novels — or, the type of

Goodreads user who is likely to read and review a canonical or prizewinning novel is also likely to discuss voice. Or this may simply be an instance of feedback: voice is included in blurbs for literary fiction and discussed by reviewers (many of whom are active on Goodreads), and so common readers are primed to read for and discuss voice.

As #OwnVoices makes clear, voice is often shorthand for discussing complex issues of identity and cultural appropriation. We wondered: is voice used more often when discussing people of color? Is it used to extol, essentialize, or both? Table 3 includes the Top 5 (or so) “voiciest” writers reviewed in each corpus. These are the writers with the highest percentage of reviews that mention voice.

<b>Corpus</b>	<b>Top 5 Voiciest Writers, per Corpus</b>
American Periodicals Online Review Corpus	1. Eudora Welty 2. John Updike 3. Jayne Philips, Reynolds Price, and James Thurber (tie) 4. Isaac Bashevis Singer, Don DeLillo, William Gaddis (tie) 5. James Baldwin, Dave Eggers, Louise Erdrich, David Gates, Adam Haslett, Alice McDermott, Joyce Carol Oates, William T Vollmann (tie)
<i>The Paris Review</i> Author Interviews	1. E.L. Doctorow 2. Joy Williams 3. Hilton Als 4. Frank Bidart 5. Sam Shepard
Blurbs	1. Wiley Cash 2. Adam Haslett 3. Lila Bowen 4. Jami Attenberg, Lisa O'Donnell (tie) 5. Elizabeth Crook, Emma Donoghue, Maria Semple (tie)



Goodreads Reviews: Canon	1. William Faulkner 2. William Gaddis 3. Samuel Beckett 4. Edward Morgan Forster 5. Robert Heinlein
Goodreads Reviews: Prizewinners	1. Louise Erdrich 2. Barbara Kingsolver 3. Philip Roth 4. Marilynne Robinson 5. Saul Bellow
Goodreads Reviews: Bestsellers	1. Stephen King 2. Kathryn Stockett 3. James Patterson 4. John Grisham 5. Suzanne Collins

Table 3. The voiciest writers in each corpus.

In no corpus does voice appear to be over-applied to writers of color. It may be that this usage — voice as central to a liberal “diversity” discourse — is somewhat more recent than the average date of reviews in our corpus (1994 and 2003 in the *Times* and *Paris Review*, respectively). Or it may be that an author-centric approach to studying voice, as we have taken above, is less useful than a topic-centric approach. That is to say, voice may be used in relation to topics like discrimination in publishing (like #OwnVoices) or a desire for inclusivity (“Emerging Voices”), but less frequently for discussing any specific writer. Still, voice is a frequent enough concept in vernacular literary criticism to warrant investigation beyond these high-level word counts. How is voice used when it is a topic of concrete discussion? Where do reviewers find voice in a text, on the page? What makes a voice distinctive or appropriative? How do you know when you’ve found it?

## The Voice You Find May Not Be Your Own

Regardless of the outlet, there are clear trends in voice’s use in contemporary vernacular literary criticism. Having located examples of voice across corpora, we winnowed our total number of reviews down to a manageable selection and began to read. Our analysis suggests that vernacular literary critical evaluations of voice fall into two broad categories: the internal and the external. These two categories

may also be conceptualized as the textual and the paratextual, or the formal and the critical. Indeed, the term voice is sometimes used to refer to paratextual or extratextual elements — like a writer’s ethnic affiliation, generational status, or professional reputation — and sometimes used to distinguish a writer’s prose style or her handling of character and dialogue.

Reviewers invoke the concept of voice at its broadest to designate an author’s relationship to a community, presuming a direct affiliation between an author and that community. This is the type of voice that operates in the publishing industry’s desire for “emerging voices.” In this context, critics use voice to imply that a writer can speak *to* or *for* a community, be the voice of a generation, or “give voice” to a historically underrepresented group or an overlooked problem. For instance, as Philip Caputo writes in *The New York Times*, Viet Thanh Nguyen’s *The Sympathizer* “fills a void in the literature, giving voice to the previously voiceless while it compels the rest of us to look at the events of 40 years ago in a new light.” Nguyen is able to rectify the “voicelessness” of the Vietnamese people, Caputo argues coyly, because of his own heritage: “Nguyen, born in Vietnam but raised in the United States, brings a distinct perspective to the war and its aftermath.”<sup>22</sup> The Implied Author is irrelevant; what matters is that Nguyen’s “distinct perspective” is tied to his Vietnamese heritage, which allows him to speak on behalf of the Vietnamese (whether or not he claims to do so himself).

Caputo’s assumptions about Nguyen’s racial identity and those of the *Times*’s audience (“the rest of us”) are hardly new or surprising; indeed, these sorts of statements are made regularly about minoritized writers. Marco Portales’s 1984 review of *Love Medicine* praised Louise Erdrich, who had just “found her Native voice.”<sup>23</sup> By 1993, Josh Getlin was writing in the *Los Angeles Times* that Erdrich had become “A Voice No Longer Ignored” — though Getlin’s lede elides the distinction between Erdrich’s voice and that of Native Americans broadly: “Louise Erdrich’s success helped ensure that Native American fiction became a flourishing genre. But she says publishers have much to learn.”<sup>24</sup> Though these reviewers clearly mean to praise Erdrich, their discussion of her voice is deeply essentializing. Whether or not Erdrich or Nguyen purport to be representatives for or speak on behalf of their respective communities, reviewers presumed that these were the goals of American ethnic writers, or of works that are about marginalized groups. (This is the stated goal of #OwnVoices.)

More narrowly and traditionally, reviewers use voice to refer to the writerly qualities that make an author uniquely identifiable as himself, without any presumption about an author's relationship to an external community. This is the voice purportedly found in the MFA classroom, the authorial position from which authors write what they know. This type of voice is developed over time, a kind of skill that reviewers evaluate with endless adjectival clarifications. Janet Maslin's *Times* review of Michael Chabon's *Amazing Adventures of Kavalier and Clay* works on each of these registers: "In a cameo-studded book . . . that echoes *Ragtime*, just as it sometimes suggests John Irving in fanciful mode, Mr. Chabon tells a bustling, convoluted story in an eloquent, exceptionally precise voice."<sup>25</sup> Maslin emphasizes Chabon's literary-historical *bona fides* when arguing that the novel is indebted to both Doctorow and Irving, but *Kavalier and Clay* is ultimately written in Chabon's "precise voice" — that is, precisely *not* Doctorow's or Irving's. Chabon's voice is eloquent, commanding a "convoluted" story with exceptional skill.

Voice is also used to refer to internal dimensions of a text, as in the voice of the narrator or characters, and direct speech or dialogue. Speaking of his novel *World's Fair* in *The Paris Review*, E.L. Doctorow discusses voice in each of these ways, demonstrating the interconnectivity of voice as formal concept:

To me the more interesting change has to do with the voice of the major narrator, the protagonist, Edgar, who as he recalls more and more of his childhood, as he passes from infancy to youth, takes on the voice of an articulate child. The diction changes, the tone changes, as if Edgar is gradually possessed by his memory. So there's a kind of two-voiced effect, I think, the man recalling, but in the boy's higher pitch.<sup>26</sup>

Here, Doctorow uses the term to refer to "the major narrator," who is also "the protagonist," whose speech ages according to the passage of time, as Edgar is "gradually possessed by his memory." The change in Edgar's extradiegetic narration ("takes on the voice of," "the man recalling") becomes embodied in a way that is normally associated with characterization ("as if Edgar is gradually possessed"), and is marked by changes in the representation of oral speech ("diction," "tone", "the voice of an articulate child", "the boy's higher pitch") — this is diegetic voice, speech represented in writing and rooted in sound, represented textually. Doctorow uses the term flexibly: Voice slips through the different levels of narration, binding

the represented speech of characters to the body of the imagined narrator to the hand of the author.

Voice is often used to describe narrators and narration, particularly in employing free-indirect discourse or first-person narration, as in the case of Paul Berman's NYT review of Philip Roth's *The Plot Against America*: "The dignity, the formality, sometimes even the hint of academic reserve in the narrator's voice produce two vibrating timbres, and these dominate the novel — a timbre of explosive anger, and, when the clapper swings to the other side, a timbre of husky pathos."<sup>27</sup> The voices of narrators and characters are often linked: Roth's narrator, Philip, is also the central character of the novel. As Berman's review makes clear with its invocation of timbre, vibrations, and tone, sound is at the very core of voice — that is, speech, as a form of sound emanating from a body, represented in writing. Reviewers use voice to refer to dialogue, particularly when rendered in dialect. And herein lies voice's central problem. The affinities between Roth, the author, and Philip, his narrator, are quite clear. However, the (perceived) relationship between Character Voice and Author Voice is often incredibly fraught, concerning both skill and cultural appropriation.

Anxieties about an author's ability or right to represent others are often expressed in terms of "ventriloquizing." A 1986 review of Kate Vaiden by Reynolds Price addresses the mismatch between the gender of the author and his central character, calling Price's "successful creation of a female voice [...] a tour de force," in contrast to a "showy ventriloqual act." Price defends himself against the implicit critique, telling interviewer Rosellen Brown, "I felt that I could write in her voice — I was given birth by a woman and I was reared by women. A writer can function outside the narrow mental and physical confines of a particular gender."<sup>28</sup> In this case, and in 1986, Price is commended for his authorial skill, for transcending his gender to create a wholly convincing voice for a female character. But more often, the misalignment between author and character voices is rife with questions of appropriation. Voice is thus a strongly normative term, rarely used in a way that doesn't imply approbation or disapprobation.

Consider, for instance, the mixed reviews for Kathryn Stockett's *The Help*, a 2009 white-savior narrative about a journalist named Skeeter who writes an exposé on the poor treatment of African American domestic workers in Jackson, Mississippi, after witnessing the poor treatment of her own maid, Aibileen. Stockett (a white woman) claims that she wrote *The Help* while homesick, trying "to comfort myself by writing

in the voices of the people I missed,” drawing on her personal relationship with Aibileen’s real-life analogue.<sup>29</sup> Janet Maslin called the novel “ultimately winning,” but expressed reservations: “The trouble on the pages of Skeeter’s book is nothing compared with the trouble Ms. Stockett’s real book risks getting into. Here is a debut novel by a Southern-born white author who renders black maids’ voices in thick, dated dialect.”<sup>30</sup> Goodreads reviewers were not as forgiving as Maslin:

I felt that the author played to very stereotypical themes, and gave the characters (especially the African American ones) very inappropriate and obvious voices and structure in terms constructing their mental character. I understand that the author wrote much of this as a result of her experiences growing up in the south in the 1960's, and that it may seem authentic to her, and that she was even trying to be respectful of the people and the time; but, ultimately, I thought that it was written from a very narrow, idealized, almost childish perspective of race relations without a true appreciation of the humanity and soul of the characters. . . . The author would benefit from exploring authentic African American voices (Richard Wright, James Baldwin, Zora Neale Hurston, Langston Hughes, Toni Morrison, Alice Walker, Maya Angelou) and understanding the scope, range and (most important) the foundation of the emotions genuine African American characters express as a result of their journey as a people in the US (hope, frustration, drive, passion, anger, happiness, sadness, depression, joy).<sup>31</sup>

Here, the dialect employed by Stockett is a stereotypical appropriation, a sort of literary blackface. A good voice is connected to being “authentic” and “genuine,” having a “true appreciation of the humanity and soul of the characters,” and to “understanding” real feelings (“hope, frustration, drive, passion, anger, happiness, sadness, depression, joy”); a bad voice manifests in “stereotypical themes,” being “inappropriate and obvious,” having a “very narrow, idealized, almost childish perspective.” Voice carries this reviewer anywhere they need to go: the representation of human speech, characterization, the author’s biography, depictions of emotions, political and racial consciousness. At the same time, voice is not just a

catch-all: it maintains a link to the bodies of both characters and the author, and to the relation between text and community.

## Every Voice that Rises Must Converge

How, then, are we to understand how voice fits in a larger discourse? To answer this question, we turn to Goodreads and to computational modeling. Goodreads is a prime site of vernacular criticism because of the diversity of perspective and the intensity of engagement it offers. Goodreads users are professionals, amateurs, and in-between. Well-known writers (such as Roxane Gay), book reviewers (like Ron Charles), and what we might call “literary influencers” (like Anne Bogel, blogger and podcast host better known by the moniker *The Modern Mrs. Darcy*) maintain their own accounts. These writers often repost reviews published elsewhere to Goodreads and engage more directly with their audiences than would be possible in a magazine like *The New Yorker* or a newspaper like *The Washington Post*. Goodreads facilitates lively conversation amongst reviewers in comments and replies; these are not formal reviews but engagements in critical conversation about books under review. Unlike published book reviews, Goodreads book reviews are not static. Often, Goodreaders will update their reviews in response to other reviewers or conversations in comments, indexing shifts in their thinking and contributing to the development of a critical vocabulary. Vernacular literary criticism flourishes in this sort of conversational, dialogic space: users develop, critique, and transform a shared critical language that is specific, but not limited, to the platform.

In other words, Goodreads enables us to ask the question: How does voice function within vernacular literary criticism? We’ve seen that voice appears to be related to genre, primarily restricted to conversations surrounding literary fiction; we’ve seen, too, that voice relates to style and technique, something evident in writing itself. We restate these questions computationally: what is the semantic connection between “voice” and other related terms, like “style”? Word Embedding Models (WEMs) are an effective computational method for modeling these relationships, enabling us to extrapolate the ways in which words are *used*. WEMs are built on the premise that word-to-word co-occurrence can help us understand semantic meaning; by this logic, words that appear together very frequently are not only statistically correlated but semantically related. For instance, *ice* and *solid* are likely to occur frequently with the word *water*, and infrequently with the word *tuxedo*.<sup>32</sup> WEMs also allow for

mathematical operations. In addition to querying individual words (the top 40 words associated with voice, for instance), we can also add one word to another to come up with a new set of meanings (what is the result of combining Voice and Genre?), or subtract one word from another (what is Style without Voice?) to understand the differences between concepts. In what follows, we investigate the relationships between voice, style, and genre that circulate on Goodreads. We find that Goodreads reviewers use the word voice in order to identify aspects of style that cannot be attributed to genre.

Figure 1 shows the Top 50 words associated with voice in our Goodreads corpus. This table can be read from left to right, top to bottom, with each word descending in its strength of association with the word “voice.” In Goodreads reviews, voice is used when discussing different people. Conceptually, voice is most nearly related to a Narrator, as well as other person-centric words: Person, Character, and Author, His, and Her. For Goodreaders, voice is primarily internal to the text; though both terms are present, “Narrator” is more closely related to voice than “Author’s.” Likewise, Goodreaders more strongly associate voice with sound than with writing — an implied body from which voice emanates. Voice includes sound and bodies, both a speaker and a perceiver: Speaks/Speaking, Tone, Sound, Dialogue, Hear, Ear. Voice implies a physical presence, sound emerging from a physical, embodied point of origin. Voice is physiological, not only emanating from bodily organs (Mouth) but also perceived by other bodily organs (Ears).

This is not to say that voice is merely a synonym for speaking or narration. Goodreads reviewers also trade in the language of individuation when discussing voice. Voice is personal — Distinctive, Unique, Own — and related to a particular Perspective. Presumably, this reflects moments in which Goodreaders critique the voice of the author. This does not necessarily imply hierarchy, but distinction; if voice is “inevitable in any act of novelistic communication,” as Hale asserts, then it is to be expected that Goodreaders would take notice, and that voice would distinguish one author or book from another. But the presence of “Own” in this vector perhaps suggests that Goodreaders have adopted the language of #OwnVoices — that corporate language surrounding diversity has infiltrated vernacular literary criticism as a shorthand for addressing thornier issues of representation, signification, and appropriation. Regardless, most Goodreaders understand that voice is something a person possesses that is in some way related to identity, whether that person is a narrator or character, or an author.

## “Voice”

*Goodreads*

Voice	Narrative	Dialogue	Way
Narrator	Mouth	Author’s	Inner
Consciousness	Stream	Writing	Her
Personal	Perspective	Omniscient	Ear
Narration	Unique	His	Name
Voices	Tone	Manner	Speaking
Style	Character	Own	Prose
Speaks	Telling	Storytelling	Strong
Distinctive	Sound	Her	Whose
Mind	Approach	Characterization	Makes

Figure 1: The top 40 words associated with the word “voice,” comprising the voice vector, in a GloVe model of the Goodreads Corpora. Table should be read from left to right, top to bottom, with each word descending in its strength of association.

By contrast, Goodreads reviews suggest style is not endemic to a person, but a skill that one develops; style is related less to authenticity than technique (figure 2). When Goodreaders discuss style, they are more likely to discuss the *writing* rather than *communication*. While Goodreads users do not seem to closely associate voice and craft, they clearly associate style with literary technique: Writing, Prose, Narrative, Characterization, Tone, Structure. Voice is embodied and aural, whereas style is technical and written. Here, the terms that Goodreads users most closely associated with voice and style prove instructive. Voice is most closely associated with Narrator, Consciousness, and Person; style is most closely associated with Writing, Prose, and Narrative. Narrator vs. Narrative. This raises the question: how does one perceive the voice of a narrator if not through writing, prose, and narrative?

## “Style”

*Goodreads*

Style	Approach	Quality	Though
Writing	Descriptive	Structure	Very
Prose	Dialogue	However	Literary
Narrative	Novel	Way	Tone
Language	Consciousness	Poetic	Form
Storytelling	Terms	Use	Text
Stream	Fashion	Characterization	Although



---

Author's	Work	Description	Voice
Unique	Experimental	Rather	Written
Works	Narration	Usual	Found

---

Figure 2: The top 40 words associated with the word "style," comprising the style vector, in a GloVe model of the Goodreads Corpora. Table should be read from left to right, top to bottom, with each word descending in its strength of association.

Remember: these results are not mere synonyms — “voice” doesn’t mean “narrator” and “style” doesn’t mean “narration.” Rather, they map a conceptual relationship, showing how these words accrue meaning relative to voice or style. These vectors demonstrate that each term is used in distinct ways and is laden with different associations. Voice is not merely the non-specialist’s way of talking about style. Yet, there is clear conceptual overlap between voice and style; Voice appears in the Style vector, and vice versa. Goodreads reviews that discuss Characterization, Tone, and Author may be discussing either voice or style. The mathematical functions of our model enable us to understand how these concepts are organized in a complex system — how style and voice relate to one another, how one might perceive voice through style, how voices may be stylish or stylized, or how one’s style might be voiced or voicey. By subtracting voice from style and vice versa, we can extrapolate differences between the concepts, inferring areas of overlap and singularity.

Style – Voice = Genre

*Goodreads*

---

Novels	Prose
Style	Works
Experimental	Descriptive
Graphic	Century
Genre	Dated
SF	Enjoy
Sci	Romance
Fi	Thriller
Writing	Descriptions
Dense	Fiction
Victorian	USA
Gatsby	Structure
Urban	

Figure 3: Vector math: the top 25 words that result from the operation  $V(\text{Style}) - V(\text{Voice})$  in the Goodreads corpora.

Figure 3 shows the Top 25 words that result from the operation  $V(\text{Style}) - V(\text{Voice})$ . The answer seems clear: Genre. When Goodreaders discuss style but not voice, they turn to generic conventions — that which is independent of an author or character. A number of the words in this Vector name genres specifically: SF, Sci Fi, Romance, Thriller, Urban. The strong appearance of genre suggests that a writer's adherence to generic conventions supersedes the sort of unique, individual, or personal style that we might call voice. Different genres require different styles and stylistic conventions, easily identifiable by even a casual reader of westerns, fantasy, or detective novels. A distinct voice is not necessary to categorize a novel within a genre and may in fact distract from the typical conventions associated with genre fiction. To frame this perspective another way, Goodreaders who read for style without considerable attention to voice are more concerned with formulas and convention; style without voice tends toward the formulaic.

Science Fiction, romance, and thrillers are traditionally associated with the popular, the middle- to low-brow, and pulps. Graphic novels and “urban” novels — an unfortunate industry term referring to fiction by and about African Americans that has not been deemed worthy of the more general (and higher-status) descriptor, literary fiction — have not historically been highly esteemed. Each of these genres has been castigated as predictable and conventional; in the case of romance novels, Janice Radway has demonstrated that the formulaic nature of the plot is part of what makes the genre so appealing to its die-hard readers.<sup>33</sup> The stylistic conventions demanded by these genres, we argue, take precedence over any individual or unique claim to authorial style. Genre is not where voice is found.

Style – Genre = Voice

*Goodreads*

Descriptive	Elegant
Prose	Sometimes
Spare	Description
Style	Eloquent
Repetitive	Sentences
Lyrical	Choppy
Voice	Writing
Verbose	Occasionally
Poetic	Rhythm

---

Dialogue	Repetition
Sparse	Dialog
Tedious	Ornate
Descriptions	

*Figure 4: Vector math: the top 25 words that result from the operation  $V(\text{Style}) - V(\text{Genre})$  in the Goodreads corpora*

Goodreaders do, however, discuss style with no mention of genre, presumably engaging in some type of formal analysis (figure 4). In these instances, they emphasize the language, and in particular, its aural qualities — lyric, poetic, rhythm. While style without voice left us with an entirely impersonal vector, style without genre emphasizes the personal and individual dimensions to writing: voice ( the fifth most related word in the vector). In other words, in vernacular literary criticism, voice is a type of style.

As a type of style, voice can be evaluated. Note the presence of adjectives: Descriptive, Spare, Repetitive, Lyrical, Verbose, Poetic, Sparse, Tedious, Elegant, Eloquent, Choppy, Ornate. This vector is far more attentive to the quality of language than generic or stylistic conventions. While the results of Style – Voice are preoccupied with identification, Style – Genre is far more interested in evaluation. Although generic systems can be identified based on conventions (a spaceship is unlikely to appear in a Western, say), these adjectives are contradictory, reflecting an act of readerly aesthetic judgment: one reader’s Tediousness is another reader’s Lyricism. Voice is not a standalone term; it requires adjectival qualification.

Given that style without voice is genre ( $\text{style} - \text{voice} = \text{genre}$ ) and style without genre is voice ( $\text{style} - \text{genre} = \text{voice}$ ), we should expect that the combination of voice and genre should equal style ( $\text{voice} + \text{genre} = \text{style}$ ). As we can see in figure 5, the equation holds: the top result is style. These concepts are, in fact, tightly related. In both of these equations, when we subtract voice and genre from style ( $\text{style} - \text{voice}$ ;  $\text{style} - \text{genre}$ ), style remains. By contrast, when we subtract style from either voice ( $\text{voice} - \text{style}$ ) or genre ( $\text{genre} - \text{style}$ ), we are left with nonsense — as figures 6 and 7 demonstrate, the concepts do not meaningfully cohere.<sup>34</sup> That is to say, style, voice, and genre are not conceptually synonyms within vernacular literary criticism, though they are intertwined. In sum:

1. Style consists of voice and genre (i.e., voice is a subset of style)
  - a. Style is bigger than and inclusive of both voice and genre

- b. One can have style without voice (genre), but one cannot have voice without style.

Goodreaders do not discuss voice or genre without also discussing style. Likewise, Goodreaders always discuss style when discussing voice and genre. That is to say, in vernacular literary criticism, voice and genre are both parts of — but not synonymous with — style; voice and genre are subsets of style that, together, make up a complex whole. Goodreaders do not discuss voice as an independent entity; it's not something that can be perceived or evaluated in isolation. Rather, to hear or acknowledge or evaluate a writer's (or narrator's) voice is to discuss a facet of their style — a personal, individual stylistic expression, wholly distinct from inherited generic conventions.

Voice + Genre = Style

*Goodreads*

---

Style	Text
Voice	Works
Fiction	Novel
Genre	Mind
Sort	Kind
Writing	Usual
Consciousness	Storytelling
Type	Author's
Narrative	Fashion
Form	Stream
Fantasy	Unique
Perspective	Science
Literary	

*Figure 5: Vector math: the top 25 words that result from the operation  $V(\text{Voice}) + V(\text{Genre})$  in the Goodreads corpora.*

Voice – Style = ?

*Goodreads*

---

Fens	Seacoast
Cryptographer	Unremembered
Limba	Invalid
Clansmen	Unwed
Shawl	Ganged
Effecting	Leukemia
Javelins	Unborn
Fido	Spotless
Sulky	Latins
Hame	Moms
Undefeatable	Scientists
Prying	
Replying	

*Figure 6: Vector math: the top 24 words that result from the operation  $V(\text{Voice}) - V(\text{Style})$  in the Goodreads corpora.*

Genre – Style = ?

*Goodreads*

---

Pontiff	Transnational
Alcove	Pirated
Foursome	Bodybuilder
Ilium	Spoofs
Aficionados	Craze
Tub	Regent
Reservations	Classicists
Noblewoman	Incontrovertible
Revivifying	Lapdog
Apologists	Synchronize
Spayed	Prelude
Pianos	Endowment
Have	

*Figure 7: Vector math: the top 25 words that result from the operation  $V(\text{Genre}) - V(\text{Style})$  in the Goodreads corpora.*

Voice is internal to a text; it is not reducible to an author's identity but, because voice is perceived to be both personal and embodied, it can be evaluated in relation to the

author. Voice might provide readers a unifying narrative by which they understand and argue about a number of features internal and external to a text, as our qualitative analysis suggests; however, each of those features are ultimately stylistic, whether approached from a formal or a sociological angle.

The most important finding of this model, in our view, is not the content of these conceptual relationships (which largely aligns with our qualitative reading) but that these concepts cohere into a unified system. It is less meaningful, in other words, that voice and style are used in different ways to refer to different things — a rather obvious finding — than that the critical vocabulary of Goodreads is systematic and specific. Vernacular literary criticism does not only concern itself with issues of identification, concepts like the Bechdel Test, or discussions of tropes and “Easter eggs.” Rather, through conversation and interactions, common readers on Goodreads have developed vocabularies and systems of critique and evaluation. Just as romance readers’ choice of books is governed by distinct codes and theories, as Radway demonstrated nearly forty years ago, so too is vernacular literary criticism governed by its own internal logic, so too does it maintain its own systematic critical idiom, transforming inherited concepts into a distinct, though not unrelated, discourse. Voice is just one part of this discourse, and perhaps not even the most significant concept. But it demonstrates the value, we believe, in devoting further, sustained attention to the ways that critique flourishes outside of formal literary institutions-- and the value of exploring this discourse computationally, and at scale.

Whether a flashpoint for political debates or a part of a complex system of evaluation, it is clear that voice can be found (almost) everywhere. But it does not mean everything. Because voice is related to issues of authentic self-expression and identity (per Hale), it allows readers to use it in order to approach a text sociologically, linking a text to its larger social milieu via authorial identity. And because the concept of “voice” is finally figurative — a writer’s or narrator’s voice cannot literally be heard; it can only be read, conveyed in writing — it also provides readers a way of approaching and discussing style, particularly those technical aspects of style that evoke sound: rhythm, rhyme, etc. The alignment between these two modes of perception (sociological and formal) and sites of analysis (paratextual and textual) is the subject of much analysis and controversy: does the author’s voice align with the voices she means to represent? Much (digital) ink has been spilled over this question. Voice has become a shorthand through which readers intelligently negotiate questions of contextualization and paratextual politics, authors relate their

biographies to their writing, and the publishing industry structures the literary marketplace.

---

## References

1 We would like to thank the members of the Stanford Literary Lab, and especially Mark Algee-Hewitt, for enthusiastic support and engagement with this project. Many thanks to Alexander Manshel and Annika Butler-Wall, whose contributions on earlier iterations of this project contributed to central questions. J.D. Porter proved an invaluable interlocutor, and his insights helped the project take this final shape.

2 Kevin Larimer, "[We Mean Business: Twelve Agents Who Want to Read Your Work](#)," *Poets & Writers*, June 14, 2017.

3 Jonathan Karp, "[10 Rules for Book Editors](#)," *Publishers Weekly*, October 20, 2017.

4 "[About Us](#)," *PEN America* (blog), September 20, 2016.

5 Lee & Low gives a New Voices Award annually to a writer of color or a Native/Indigenous writer. "[New Voices Writing Contest for Picture Books | Lee & Low Books](#)," accessed January 30, 2020.

6 Mark McGurl, *The Program Era: Postwar Fiction and the Rise of Creative Writing* (Cambridge: Harvard University Press, 2009), 230.

7 Ibid., 232.

8 Laura Heffernan and Rachel Sagner Buurma, "The Common Reader and the Archival Classroom: Disciplinary History for the Twenty-first Century," *New Literary History* 43, no. 1 (2012): 113-135.

9 Merve Emre, "[Post-disciplinary Reading and Literary Sociology](#)," *Modernism/modernity Print Plus*, February 1, 2019.

10 Critics such as Katherine Bode have called for a more reader-focused model of a "literary system," in which readers are a crucial component of how a work "accrue[s] meaning. Katherine Bode, *A World of Fiction: Digital Collections and the Future of Literary History* (Ann Arbor: University of Michigan Press, 2016).

11 Droitcour writes in *The New Inquiry*, "Vernacular criticism can reject the guidelines set by cultivated artistic tastes, or it can guilelessly speak in ignorance of them, or in its naive fascination with them can inadvertently expose their falseness. Vernacular criticism is an expression of taste that has not been fully calibrated to the tastes cultivated in and by museums. Vernacular criticism inscribes bodies in public spaces that would otherwise erase them." Brian Droitcour, "Vernacular Criticism," *The New Inquiry*, July 25, 2014. <https://thenewinquiry.com/vernacular-criticism/>; Lisa Nakamura, "Words with Friends: Socially Networked Reading on Goodreads," *PMLA* 128, no. 1, (2013): 238-243. Nakamura uses the phrase "vernacular criticism" to describe Goodreads in her PMLA article, yet she does not develop the concept at length.

12 Scott Selisker, "The Bechdel Test and the Social Form of Character Networks," *New Literary History* 46, no. 3 (2015): 505-523.

13 Miriam B. Hansen, "The Mass Production of the Senses: Classical Cinema as Vernacular Modernism," *Modernism/modernity* 6, no. 2 (1999): 59-77.

14 We would like to thank our undergraduate research assistants, Eun Ji Lee and Shana Hadi, for their invaluable work building this corpus.

15 Our thanks to Alexander Manshel for providing us with this data.

16 Some of these lists were generated by lay readers, some by specialists, and some by publishing houses to boost sales as the 20th-century came to a close. Owing to the distinct selection committees involved, this corpus contains everything from the highly canonical (*The Great Gatsby*), to lesser-read experimental novels (*The Ticket That Exploded*), to novels beloved by niche — and highly vocal — reading communities who advocated strongly that their books make the list (Ayn Rand's *Atlas Shrugged* or L. Ron Hubbard's *Battlefield Earth*).

17 We recognize that these are contested categories; we use these terms as broad heuristics, rather than normative or stable categories. As we explain below, we take "literary fiction" as a genre category for publishers' BISAC codes; we take novels shortlisted for prizes and novels appearing on the bestseller list as proxies for "prestigious" and "popular" fiction.

18 The high number of words in each interview might be why this 78% figure is so much higher than that of, say, our results for the MLA International Bibliography: why wouldn't an author eventually use the word Voice if they were prompted to talk about their work for long enough?

19 Foote had originally been a novelist, and he explains how he transitioned to writing history: "Always, anything I write takes place at a certain time and a certain place. I think time, era, and place — geographical location — are very, very important to me and what I'm doing . . . I guess I have trouble with *Waiting for Godot* because it doesn't matter when or where it is." In his description, Foote presents his writing as a mute document of times and places: what matters is the setting, not who's standing in it. In that light, his critique of *Godot*, a play about the incoherence of speech — the plot, characters, and themes strip down to pure voice — makes sense. Even as a historian, Foote's methods are textual, not oral: the book he says he relied on the most for his history of the Civil War was *The War of the Rebellion: A Compilation of the Official Records of the Union and Confederate Armies*.

20 We used a book's BISAC codes, as identified by publishers, to determine genre.

21 Matthew Wilkens, "[Genre, Computation, and the Varieties of Twentieth-Century U.S. Fiction](#)," *Cultural Analytics*, November 1, 2016.

22 Philip Caputo. "'[The Sympathizer](#),' by Viet Thanh Nguyen," *The New York Times*, April 2, 2015.

23 Marco Portales, "People With Holes in Their Lives," *The New York Times*, December 23, 1984.

24 Josh Getlin. "[A Voice No Longer Ignored: Louise Erdrich's Success Helped Ensure That Native American Fiction Became a Flourishing Genre. But She Says Publishers Still Have Much to Learn](#)," *Los Angeles Times*, December 13, 1993.

25 Janet Maslin, "[A Life and Death Story Set in Comic Book Land](#)," *The New York Times*, September 21, 2000.

26 George Plimpton, "[E. L. Doctorow, The Art of Fiction No. 94](#)," *The Paris Review* 101 (1986).

27 Paul Berman, "[The Plot Against America](#)," *The New York Times*, October 3, 2004.



28 Rosellen Brown, "[Travels with a Dangerous Woman](#)," *The New York Times*, June 29, 1986.

29 Motoko Rich, "[A Southern Mirrored Window](#)," *The New York Times*, November 2, 2009.

30 Janet Maslin, "[Racial Insults and Quiet Bravery in 1960s Mississippi](#)," *The New York Times*, February 18, 2009.

31 "[Caroline's Reviews > The Help](#)." *Goodreads: The Help*, September 22, 2009.

32 See Jeffrey Pennington, Richard Socher, and Christopher D. Manning, "[GloVe: Global Vectors for Word Representation](#)," Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (2014).

33 Janice A. Radway. *Reading the Romance: Women, Patriarchy, and Popular Literature* (Chapel Hill: University of North Carolina Press, 2006).

34 Similarly, these equations do not work with any other corpus in our collection. We take this failure to be less a matter of Goodreads' relative coherence than a matter of corpus size. Our Goodreads corpus is larger than any of our other historic corpora, aside from the Historic Literary Criticism corpus.