BIROn - Birkbeck Institutional Research Online

Open Library of Humanities

C21 LITERATURE: Journal of 21st-century Writings

Article

Peer Review:

Copyright:

Open Access:

Digital Preservation:
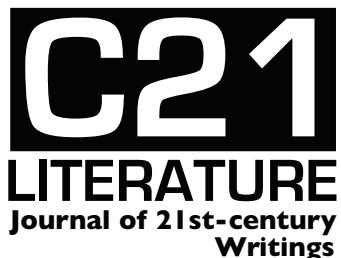
ARTICLE

The Historical Imaginary of Nineteenth-Century Style in David Mitchell's *Cloud Atlas*

Martin Paul Eve

Birkbeck, University of London, GB

martin.eve@bbk.ac.uk

The first section of David Mitchell's genre-bending novel, *Cloud Atlas* (2004), purports to be set in 1850. Narrative clues approximately date the intra-diegetic diary object of this chapter to the period 1851–1910. This article argues for the construction of a stylistic historical imaginary of this period's language that is not based on mimetic etymological accuracy. Specifically, I show that of the 13,246 words in Part I of The Pacific Journal of Adam Ewing, there are at least three terms that have an etymological first-usage date from after 1910: spillage, variously attested from ~1934; latino, from ~1946; and lazy-eye, from ~1960. Instead, I show that racist and colonial terms occur with much greater frequency in *Cloud Atlas* than in a broader contemporary textual corpus (the Oxford English Corpus), indicating that the construction of imagined historical style likely rests more on infrequent word use and thematic terms from outmoded racist discourses than on etymological mimesis.

The Pacific Journal of Adam Ewing

David Mitchell's *Cloud Atlas* (2004) is a novel written in six different generic styles (on sources and influences, see Book World 2004; Mitchell 2010; Begley 2010). These range from detective thrillers to farce to far-future SF (speculative/science fiction) (Hopf 2011; Eve 2014; O'Donnell 2015). Perhaps the chapter that must perform the most work, though, is The Pacific Journal of Adam Ewing. Certainly, one could argue that Sloosha's Crossin' has chronological priority, looking back on the other sections of the text from a historical vantage point. One might also assert that the Letters

from Zedelghem section is central to the text, for it is there that the metonymic Cloud Atlas Sextet is composed. However, Ewing begins and ends the novel, thus placing its language and themes under more intense literary-critical scrutiny (Adamo 2000). The chapter must not only introduce this strange novel but, due to *Cloud Atlas*'s unusual narrative structure, also convey that sense-making function of an ending towards which Frank Kermode gestured (2000). The diary object that is later read by Robert Frobisher certainly has an important role in this text.

The diary itself is written in the supposed style of a seafaring narrative of the mid-nineteenth century; "Ewing puts me in mind of Melville's bumbler Cpt. Delano in 'Benito Cereno'", remarks Frobisher in both the E and P editions of the text (Mitchell 2004a, 463, 2004b, 445; for information on the differences between the editions and the use of E and P notation, see Eve 2016b. Citations in this article are to both print variants of the novel). Indeed, Melville looms large over this novel, not only because Mitchell claims the author as a source in an interview but also since the novel gives us, early on, the line "it's not down on any map", which echoes Melville's famous "it is not down on any map; true places never are" in *Moby-Dick; or, The Whale* (1851) (Mitchell 2004a, 3, 2004b, 3; Melville 2008, n.p.; Mitchell 2010). Yet, despite the novel's own statements being a metatextual gesture that undoubtedly "expos[es] its concerted effort to 'forge' the form of a historical journal" (Hicks 2010), previous work that I have conducted shows that computational authorship attribution techniques correlate the text neither with Melville's *Moby-Dick* nor with his novella "Benito Cereno" (1855) (Eve 2017; for more on these methods, see Burrows 2002; Hoover 2004; Stein and Argamon 2006; Argamon 2007).

The novel also gives its own internal dating, though, for the Ewing narrative. We are told, by Frobisher, that "[m]ention is made of the gold rush, so I suppose we are in 1849 or 1850" (Mitchell 2004a, 64, 2004b, 64). If we take the diary at face value, then Frobisher is almost right. In fact, the year must be precisely 1850, since this is the only year in the 1850s range that has the 7th November (the first dated entry in the diary) falling on a Thursday. Hence, also, by the internal chronology, when Ewing notes that "[t]oday is [his] thirty-fourth birthday" on Sunday the 12th January 1851, Ewing's precise birthday is the 12th January 1817 (Mitchell 2004a, 527, 2004b, 506).

In its tight internal chronology that *does* match the historical record, the text even manages here to parody the (or, my) act of literary interpretation; Frobisher is akin to the paranoid critic who would seek out such information.

However, the first thing to note is that we cannot take the date of the diary at face value. As Frobisher again notes, in a perhaps defensive authorial move for Mitchell, there is "[s]omething shifty about the journal's authenticity—seems too structured for a genuine diary, and its language doesn't ring quite true" (Mitchell 2004a, 64, 2004b, 64). Frobisher clearly suspects the entire thing to be a literary forgery; which, of course, it is. Mitchell is the ultimate forger here (although it is by license of the reader), but in the intra-diegetic setup of the text, Jackson Ewing looks likely to have doctored the diary.

The reader knows, from the final pages of the diary, that Jackson Ewing, the son who has "edited" this published diary, was born before Ewing set sail in 1850. However, we are also told that Jackson Ewing is the same age as the first hazing victim aboard the ship: "Rafael was Jackson's age" (Mitchell 2004a, 518, 2004b, 499). Assuming, then, an approximate earliest birthdate of 1st January 1835 for the late-teenager Jackson Ewing, it seems likely that the furthest date within the text's internal chronology for editing and publication of the diary, taking an optimistic human lifespan average of 60 years for the time, might reasonably be 1895. The diary would also have to have been edited after Ewing's return at a later date. If one wanted to be generous, one could extend this by 15 years to 1910, so as to also chime approximately with the date of the diary's "discovery" by Frobisher in the Letters from Zedelghem section of the novel, a few years later.

The date range that this yields for Mitchell's Ewing chapter is, then, 1851–1910. However, this chapter is not a traditional "historical fiction". Certainly, it possesses some of the tropes that we traditionally ascribe to historical fiction: a sense of "heft and authenticity" and a time-frame beyond the lived experience of present-day human readers (see de Groot 2010). It is also clearly the case that the chapter required intricate research to write, as per the rules of the Walter Scott Prize for historical fiction. Yet, where Mitchell's text differs from other works of contemporary historical fiction – such as Mantel's *Wolf Hall* (2009) – is that the *linguistic style* purports to be

of the time. That is, Mitchell aims to write as though the diary was actually produced in the ~1850–1910 timeframe.

By most accounts, Mitchell's novel is successful at imitating the linguistic style of the period in which it purports to be set. However, the questions that I ask here in this article are: how does Mitchell achieve this? What are the limits of linguistic mimesis in *Cloud Atlas*? And what kind of historical imaginary could function as a model against which we could measure Mitchell's prose? Specifically, this article addresses the extent to which Mitchell's chapter is accurate in the use of language from its claimed period. That is, I wanted to know whether, of the 13,246 words in The Pacific Journal of Adam Ewing, any would have been inaccessible to a writer living between 1851 and 1910. Upon discovering that such anachronistic words do exist within the novel, I then turn to other hypotheses about how Mitchell constructs his verisimilitude of nineteenth-century prose for a late-twentieth-century reader. In particular, I conjecture that Mitchell deploys terms that are simply less frequently written in contemporary English – including racist language – and test this against a contemporary corpus. The overarching argument that I make from this investigation is that, as per Mitchell's own remarks on his style to which I will later turn: to write in a passable faux-nineteenth-century style at the beginning of the twenty-first century, one need not only use words that would be accessible to those of that era, but one should certainly pepper one's text with colonial discourses and racist terms.

Etymological Mimesis

Assuming that Mitchell's diary object attempts an accurate depiction of language from the time of its purported authorship, an obvious first question is: are there words in the diary whose first usage falls later than the date of the Ewing section of *Cloud Atlas*? In order to gauge the "authenticity" of the diary through the appropriateness of its linguistic register, I made two initial preparatory modifications to the E edition's first section of The Pacific Journal of Adam Ewing, the portion to which Frobisher refers. The first pass that I made was to split all words within the text into their own lines and to then eradicate any words that appeared in the Project Gutenberg version of Herman Melville's *Moby-Dick.* Using this text as a filter enabled

me to eradicate words that were clearly in use in 1851, the first publication date for Melville's novel. This greatly reduced the effort involved in sequential etymological date checking of these words. The second step was to produce a piece of software that would "scrape" sets of open-access dictionary sites for claimed "first usages" of words and to run the remaining words through that software (Eve 2016a). The idea behind this was that it should give an indication of any obvious outlier words, which I would then be able to more thoroughly check.

It is worth, at this point, making a brief digression to outline some of the difficulties of trusting etymological source data. There are two dictionary sources that I used for this project: one was Dictionary.com and the other was the experimental Oxford Dictionaries API (that is: the Oxford English Dictionary). In the case of Dictionary.com, the sources upon which this site draws for its etymological data are not entirely clear. That said, a sampled check of their etymologies compared to other dictionaries – such as the OED and Merriam Webster – indicates a close correlation here. Yet, of course, etymological research is a historical process like any other, prone to flaws, revision, bias, and the perils of document destruction. The science of etymology is far from precise. Furthermore, the science of data-mining such sources, as used here, is even less precise. There were many words that I was unable to automatically classify and that were simply marked as having an unknown etymological start date. That said, because I was specifically looking for words that fall *outside* those accessible to Jackson Ewing in the novel, this presents less of a challenge. Indeed, so long as there were some anachronistic results, there would be something happening in the novel's style that would have a knock-on effect on its interpretation. In other words, this type of approach is good for answering a simple, well-defined (but nonetheless limited) query that I would phrase as: "return as many as possible, but not necessarily all, words in a text that have etymological first-usage dates after 1910".

In order to militate against the above challenges of etymological research data, I decided to further reduce the terminologies studied (in addition to de-duplication and the *Moby-Dick* filter above) to words that appear in Ewing Part I that have etymological data in both the OED and in Dictionary.com. At the time of authorship, the OED had just released an experimental API that allows for word lookup. This
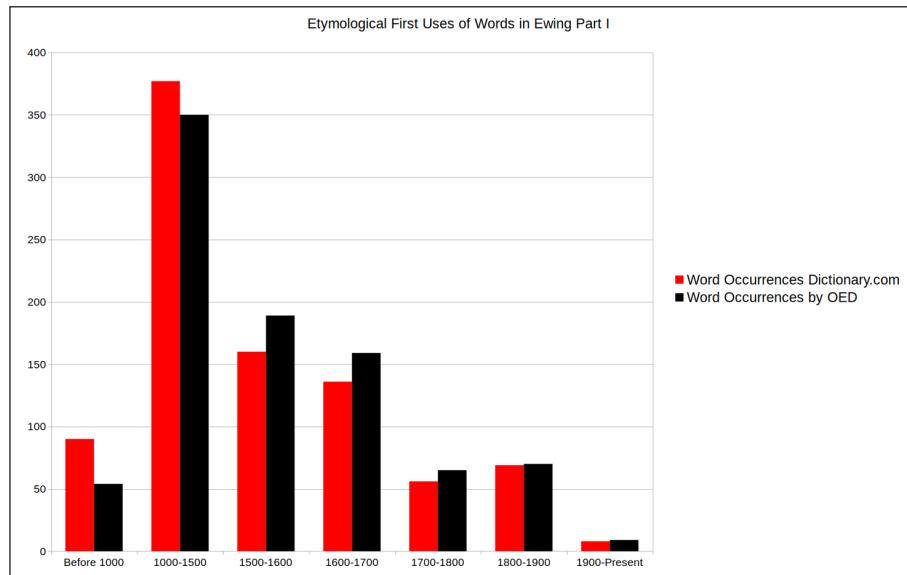
Figure 1: Word distributions by first-usage according to Dictionary.com and the OED.

includes a date-range parameter. However, since there are multiple senses for many lemmas, with different first-use dates, after an initial computational filter, I had to manually check the majority of the remaining terms. Nonetheless, this resulted in a final unique vocabulary of 896 words out of an original 13,246 terms for which I now have two sets of etymological first-use dates.

As can be seen in **Figure 1**, there is generally a close correlation between the two dictionary sites for the distributions. However, the words that fall after the 1900 cut-off point are different between the two sources.

Taking, then, a latest in-text "publication date" for the Pacific Journal's first section as 1910 yields, in my search of Dictionary.com, just six anachronistic words that would definitively not have been available to either Adam or Jackson Ewing and that occur in both editions of the text: "home-town" (P)/"hometown" (E) [1910–1915] (Mitchell 2004a, 5, 2004b, 5), "spillage" [1920–1925] (Mitchell 2004a, 15, recurring in a more appropriate register at p. 89, 2004b, 14, recurring in a more appropriate register at p. 90), "lazy-eye" [1935–1940] (Mitchell 2004a, 9, 2004b, 9), "return-ees" [1940–1945] (Mitchell 2004a, 31, 2004b, 30), "latinos" [1945–1950] (Mitchell 2004a, 29, 2004b, 28), and "A-frame" [1960–1965] (Mitchell 2004a, 6, 2004b, 6).

Yet, the Oxford English Dictionary disagrees. For "hometown", it tells us, was coined in 1851; "returnee" in 1870; and "A-frame" as far back as 1827. The OED also yields a number of terms from the novel as being after our cut-off date that Dictionary.com does not. In the OED, through the automatic approach, we are given: "bizarreness" [1920] (Mitchell 2004a, 27, 2004b, 26), "slumped" [1937] (Mitchell 2004a, 6, 2004b, 6), "pulsed" [1942] (Mitchell 2004a, 21, 2004b, 20), "colour" (P)/"color" (E) [1944] (Mitchell 2004a, 16, 2004b, 16), and "scuttlebutt" [1945] (Mitchell 2004a, 37, 2004b, 36). There are some strange things going on here that are worth briefly unpacking.

In the case of "bizarreness", "slumped", and "pulsed", the OED API simply disagrees with Dictionary.com, claiming that the specific forms of these words, deriving from older ancestors, were not used until these later points. This is probably because my software is pulling out the incorrect part-of-speech definitions for first-usage within the specific contexts. Two words have more interesting stories behind them, though.

"Color"/"colour" seems an unlikely candidate to have been coined, even in its American spelling, in 1944. Indeed, this is the case. What has actually happened here is that the OED API has taken "color" in the sense of "Any of various musical devices or techniques used to enhance the performance of a piece, esp. a repeated melody in late-medieval isorhythmic motets. Cf. talea n.": a very specific definition of "color", with the main entry for perceptions of electromagnetic radiation listed instead under "colour". This later usage of color in the musical sense comes later in *Cloud Atlas* but hardly applies to the initial use here: "a Bonapartist general hiding here under assumed colo[u]rs" (Mitchell 2004a, 53, 16, 2004b, 53, 16).

"Scuttlebutt" also has two different meanings. The older, given by Dictionary.com as first occurring around 1800, means "an open cask of drinking water". The usage in the text, though, is that "Henry shall inform the 'scuttlebutt' that Mr Ewing has a low fever", meaning in this case a person who puts a rumour about. This second definition as a colloquialism, according to the OED, comes from 1945, while Dictionary.com yields 1905. Interestingly, Mitchell puts this term in quotation marks, as though the speaker is using an informal or new word. Although there is disagreement between my two etymological sources, scuttlebutt is definitely an edge case here. It is very unlikely that it would have been used in the informal sense during

the period of purported authorship of the document. On the other hand, Dictionary. com does put such a use at 1905, so it makes sense to exclude this from the final definitive list.

This leaves, then, just three terms that, I feel, can be said with certainty to have been absolutely inaccessible either to Mitchell's historic author or editor: spillage, from ~1934; latino, from ~1946; and lazy-eye, from ~1960. In the case of spillage, the text is here recounting the debate between the Moriori elders as to whether "the spill-age of Maori blood" will "also destroy one's *mana*". Interestingly, the Online Etymology Dictionary disputes this entry, claiming it for the nineteenth century ("Spillage" 2017). Mitchell could have avoided this slip through reverting to the verb form, "spilling". On the other hand, latino is definitely a twentieth-century construction: "'Passionate Latinos,' observed Henry, bidding me a second good-night". While this term did not actually come to prominence until after the Second World War, the use, here, of a racial epithet has an important different effect for the construction of a stylistic imaginary of the nineteenth century, to which I will turn shortly. Finally, Mitchell gives us a "par-lour [...] inhabited by a monstrous hog's head (afflicted with droop-jaw and lazy-eye, killed by the twins on their sixteenth birthday". The sources that I consulted give this slang term for amblyopia as first appearing in the middle of the twentieth century.

The first thing to note is that this is a very good attempt at linguistic mimesis within a work of purported historical fiction. Even with Mitchell's disclaimer through Frobisher, to have used only 33 terms in total after 1850 is a substantial achieve-ment, even accounting for double that number due to inaccuracies in my program-ming. At the same time, the admission that "the language doesn't quite ring true" is either a tacit defence for or an outright admission of linguistic inaccuracies.

The second important aspect that this language use changes, however, is our understanding of the text's metaleptic slippage. In fact, the precise datings of first usage here alter the slippage twofold. Firstly, the use of the words that entered the language after 1850 but before 1910 (as a generous estimate) validates Frobisher's assessment that the diary has been subsequently edited within the narrative. However, our knowledge, also, that three of the linguistic terms in the portion of narrative that Frobisher reads were not coined until *after* the time of that section

introduces a far stranger violation of diegetic layers. For, just as Adam Ewing's narrative is told by Jackson Ewing, and Sloosha's Crossin' is told by Zachry's son, this linguistic dating gives an authorial intrusion by Mitchell at this moment, the type of authorial self-inscription seen in much metafiction, here played out in a more subtle, undoubtedly unwitting, linguistic fashion.

Thirdly, though, there are remarks to be made on the linguistic styling of pastiche, parody, and historical fiction in their attempts to become believable. It is clear that readers are very poor at both identifying terms that are anachronistic and at dating the first use of words. I had no idea that "spillage" came from the 1920s. Indeed, I am unsure that, if asked, readers would be able to point to these words as the markers of the language seeming to not "ring true". How, then, does one create a linguistic styling that *appears* mimetic of 1850 when working under the assumption that readers will not know when words are coined? For if readers do not know the truth of the language and the dating of words then they cannot be capable of spotting when the text veers away from linguistic reality.

Signs of the Times

First and foremost, to achieve this warped linguistic mimesis or historical stylistic imaginary, Mitchell deploys archaic language. Within the first few lines of the text we are given "Indian" to refer to any non-European, a "hamlet" for a settlement, a spelling of "trousers" as "trowzers", a jacket of 18th-century origin (the "Pea-jacket"), an ampersand ("&") repeated for conjunction instead of the more common "and", and the term "eyrie" to refer to a homeland (Mitchell 2004a, 3, 2004b, 3). This "archaic overloading", as we might term it, is not strictly accurate for the time period. Looking at the first passage of *Moby-Dick* as a correlative text, ampersands do not appear instead of "and", and several passages would be totally acceptable in contemporary spoken English (were they not so well known already): "Call me Ishmael", "There is nothing surprising in this". That said, there are also a set of terms in the first paragraph of Melville's text that resonate with Mitchell's opening: Ishmael reports himself to be "grim about the mouth" and notes that "[w]ith a philosophical flourish Cato throws himself upon his sword; I quietly take to the ship" (Melville 2008).
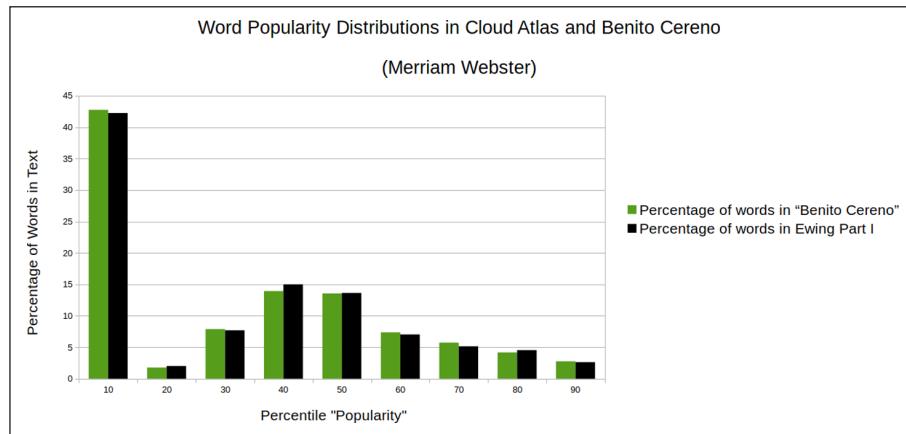
Figure 2: Word "popularities" in The Pacific Journal of Adam Ewing Part I and *Benito Cereno.*

Second, though, I hypothesized that Mitchell might simply be using *uncommon* language to create the perception of a stylistic affinity with Victorian prose for the twenty-first-century reader. For, to achieve the effect of archaic language within an environment where readers do not know when words were actually first used, it might make sense to present readers with a range of words that they are *less likely* to have encountered. This unfamiliarity might be construed, then, as outside of the bounds of conversational tone, which a reader could take to mean that the words are older than those used in day-to-day speech. Or more simply put, the less familiar the language, the more archaic it sounds.

In order to test this, I noted that the Merriam Webster dictionary has a feature that ranks the "popularity" of words and I decided to profile the first portion of the Ewing narrative using this tool. If, I thought, my hypothesis was correct, we could expect to see a distribution of words skewed dramatically towards the "unpopular" end of the spectrum. I also thought that it would be worthwhile and important here to profile a work by Melville of the time, to see whether these works too genuinely chose "unpopular" words (I chose *Benito Cereno* for the same reasons above). The results of the experiment can be seen in **Figure 2**.

Figure 2's distribution of The Pacific Journal of Adam Ewing seems interesting. It shows an approximately normal distribution of the vocabulary's range but

with a strong positive skew applied by the fact that approximately 42% of the words used in the text fall within the *bottom ten percent* popularity of the Merriam Webster account. This seemed to confirm my thinking that unpopularity of terms was a better indicator of how to achieve the prose style of the 1850s than strict mimetic linguistic accuracy.

Perhaps more importantly, though, when we plot the same graph against Melville's *Benito Cereno*, the patterns are *almost identical*. Indeed, the percentage of words that falls in the bottom 10%, according to Merriam Webster, is just 0.5% different to that in *Cloud Atlas*. The remainder of the distribution is also nearly identical to that in Mitchell's sub-novel. At this point in proceedings, I began to wonder whether or not there might be an underlying linguistic pattern at work here that pervades all language. Perhaps the long tail of these distributions is actually a feature that is intrinsic to language more generally? Or might it be an underlying feature of how Merriam Webster measures "popularity"?

There is indeed a problem with this methodology that I have not yet addressed. The underlying question that must first be answered is: what does the Merriam Webster dictionary mean by "popular"? It turns out that the Merriam Webster score for popularity is calculated by *the number of times that each word is looked up by users* online. In other words, "popularity", as defined in the Merriam Webster online dictionary, is not taken from any representative corpus of contemporary use, but is determined by how frequently users visit the definition page in question. This, in turn, raises questions as to what "popularity" might actually mean that turns upon the reasons that people *turn to online dictionaries*. By Merriam Webster's measure, "popularity" is actually constituted by a range of socio-behavioural and technological aspects.

This is to say that, in actuality, it is not really possible to use the Merriam Webster "popularity" measure as an example of frequency of contemporary use, although it may appear as such upon first glance. If we wish to get a true "popularity" measure of word-term usages in contemporary English, we would need a broad reference corpus against which to compare the language in our target texts. One such corpus is called the Oxford English Corpus (OEC), which is used by the makers of the Oxford English
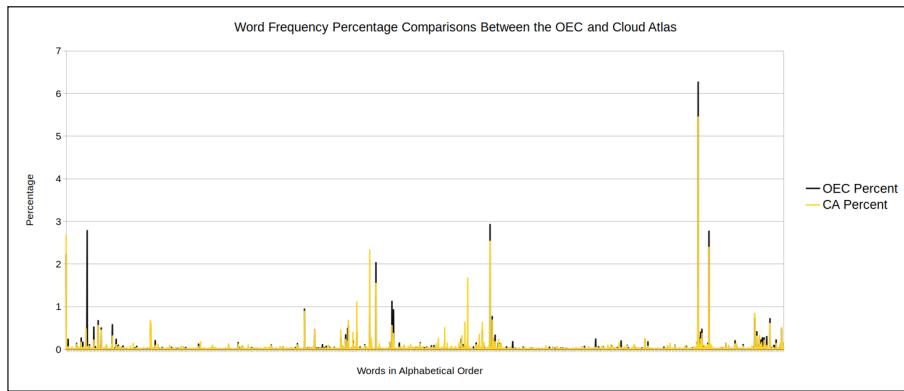
Figure 3: Relative percentage frequency of terms from Ewing Part I (E edition) vs
the Oxford English Corpus. Where a yellow line appears higher, the usage in *Cloud
Atlas* is higher than in the OEC. Where a black line appears higher, the usage is
higher in the OEC than in *Cloud Atlas*.

Dictionary to study evolving language use. It consists of approximately 2.5 billion
words of twenty-first century texts, which gives a far better sample basis for studying
the most frequent words in contemporary usage.

To study the relative frequency of terms within Mitchell's novel, I plotted the
frequencies of the terms in Ewing Part I as percentages of that text and then did the
same for those terms within the OEC. The resultant overlayed graph can be seen in
Figure 3.

This graph serves as a handy locating aide for those instances where the usage dif-
fers between the texts. If we ignore all words that are below 1% of the total usage and
only include those that have any degree of difference from the OEC, there are a num-
ber of interesting points. I also refer to this as a "locating aide", rather than a defini-
tive map of *Cloud Atlas* since there are all kinds of problem with the computational
approach here, the most pressing of which is stemmatization. That is, usually, were
we searching for uses of "abet", we would need to make a conscious decision about
whether to include "abetting", "abets", and a whole raft of other terms. As per the above
exercise where I instead opted to narrow the problem ("find any but not all words after
1910") the same might be said here: find instances of linguistic frequency discrepancy
that we can use as a starting point for a more thorough, manual investigation.

First, the word "and" only occurs three times in Ewing Part I; a mere 0.04% of the text. In the OEC, the term occurs 57,716,722 times, a far higher 2.78% of the corpus. This astonishingly low usage of among the most common terms in the English language can be attributed to Mitchell's frequent deployment of the ampersand, in its stead; the same technique that Pynchon deploys in *Mason & Dixon* (1997) but also seen in China Miéville's *Railsea* (2012) to achieve a strange temporality for the "weird" environment of that novel. Less common in contemporary usage, for sure, the ampersand was at one point in the nineteenth century taught to schoolchildren as the twenty-seventh letter of the English alphabet (Houston 2013, 76–77). While there are multiple convergent histories of the ampersand and its usage, there is no evidence that I have seen of such wholesale replacement of "and" with the ampersand in nineteenth-century prose, such as Melville. That said, the Pacific Journal is supposed to be a hand-written document, so the contraction of "and" to "&" would have saved writing effort in the fictional landscape.

Again, due to the first-person diaristic nature of this segment, we also see a far higher usage of the first-person pronoun "I" in *Cloud Atlas* than in the broader OEC (2.33% vs 0.81%). This is less a stylistic remark than simply a reflection on the specific object type that Mitchell uses: a diary. Likewise, there is a marked difference in usage of the term "is" between *Cloud Atlas*'s diary and the OEC (0.56% vs 1.13%). This is curious. Certainly, the Pacific Journal moves between tenses; some portions of the diary-in-a-novel are written in the present simple and present continuous while others are written in various past tenses. One would assume that the same would be true, though, of the OEC.

Likewise, some of the other differences between Mitchell's frequencies and the OEC's are harder to understand. For instance, the term "in" has almost a half percentage point difference between *Cloud Atlas* and the OEC (2.03% vs 1.55%). It is possible that the micro-tectonic shifts in word-frequency that occur as a result of Mitchell's forced grammatical changes could have caused this difference. It is also possible that the difference has occurred purely by chance. The 0.4% difference between "of" also falls in this category, as do the differences for "the", "that", and "to".

The challenge here, of course, is that the more frequent words (>1%) tend, in both the OEC and the novel, to be function words. On the other hand, those rarer terms in the OEC that occur frequently in *Cloud Atlas*, such as "Moriori" (*Cloud Atlas*: 0.32% vs OEC: 0.000004%) tend to be thematic terms related to the novel's focus on the "Chatham" isles (Rēkohu). This coincides with a thematic focus on empire and tropical medicine (Eve 2018) alongside the use of racist language, prevalent in nineteenth-century English society ("blackamoor" at 0.01% in *Cloud Atlas*'s Ewing compared to 0.000003% in the OEC).

The clearest instance of this prevalent racist language is when Goose – mirrored in the text's later ornithologically named MD, "Dr Egret" (Mitchell 2004a, 457, 2004b, 439) – claims that Ewing has begged him to "keep that d–d nigger away from me" (Mitchell 2004a, 523, 2004b, 502). The alienating shock effect here that constructs a different, historical linguistic imaginary is the same as that exemplified in the earlier corpus percentage analysis: that the word "damn" might need censoring while the truly more offensive racial slur "nigger" remains in plain sight should disconcert and dislocate the reader. This is a type of uncanny effect in which we realize that we are not "at home" in our own time period but instead in a world of warped racial abuse (even as it remains a stain on twenty-first-century society that such language is still used). In this case, though, the reader is asked to question what is here generic and what is specific. At this point, terms of abuse ("damn" or the earlier "bitch"/"bastard" (Mitchell 2004a, 33, 2004b, 27)) are redacted; they are made into generic forms that are cross-substitutable with others in order to construct an era that we imagine to have had a more delicate disposition towards offensive language in print. At the same time, the knowing reader is expected to interpolate a specific term within the blank mark on the page. The juxtaposition of the more offensive term also complicates our ability to read generically. For if we assume that we know the set of terms that might fit beneath generic redaction – those that are offensive – then the twenty-first-century reader should be disconcerted at the proximate inconsistency of encountering the unredacted, specific term, "nigger". This also accounts for how the anachronistic slippage of "Latino" in the work, although mimetically inaccurate, fits with the other modes of linguistic construction of the nineteenth century.

Certainly, though, there are some terms that are used that are just strange – as opposed to offensive – to our contemporary ear and that do not circulate in twenty-first century parlance. "Hugger-mugger", for instance, although occurring but once in *Cloud Atlas* has a significant deviation from the OEC (0.007% vs 0.0000005%). Similarly, the abbreviated "kerchief" occurs at 0.015% in the novel but only at 0.00002% in the OEC. Maladies also has a 0.015% occurrence in *Cloud Atlas* but constitutes a mere 0.00005% of the OEC.

In some ways, this lends credence to my hypothesis: Mitchell does tend to use archaic/unusual terms with a greater frequency than we see in a general corpus of contemporary English. Specifically, though, colonial terms of racist abuse occur in the Ewing section of *Cloud Atlas* at a far-higher frequency than in a broader contemporary corpus. There are, though, a number of additional limitations to this method that must be discussed. For instance, consider that, because the size of the Ewing chapter is relatively small, a low number of usages is often enough to produce a distinctive skew against the OEC. For instance, the above term "blackamoor" is only used twice within the Ewing narrative. However, this is enough to substantially weight its relative percentage against the OEC. This particular method, then, over-weights words with low frequencies in the smaller corpus. In a sense, though, this is helpful; the small number of usages constituting a relatively large percentage within a smaller text here is a distinctive linguistico-thematic intersection to which we *should* pay attention.

Secondly, there is the question of the composition of the OEC. The blurb for the OEC indicates that it seeks to build a representative corpus of twenty-first-century English from across the spectrum of writing types. As the makers note:

> It represents all types of English, from literary novels and specialist journals to everyday newspapers and magazines, and even the language of blogs, emails, and social media. And, as English is a global language, the Oxford English Corpus contains language from all parts of the world – not only from the UK and the United States but also from Ireland, Australia, New Zealand, the Caribbean, Canada, India, Singapore, and South Africa.

The extensive use of web pages has allowed us to build a corpus of unprec-
edented scale and variety – the corpus contains nearly 2.5 billion words
of real 21st-century English, with new text being continuously collected
(Oxford Dictionaries 2017).

Indeed, although the OEC is not composed entirely of fiction, the breadth it offers
in terms of sourcing provides for a comparison environment that is more repre-
sentative of global language usage in the early twenty-first century. Time and time
again, Mitchell uses words that occur less frequently than in the OEC (some further
examples: "Hollander" at 0.007% against 0.00006%, or the extreme "simulacrums"
at 0.007% against 0.0000004%). Certainly, this comparative frequency disjunct
contributes to the stylistic historical imaginary of the nineteenth century in *Cloud
Atlas*.

Themes, Language, and Style

In the construction of literary style, where does the theme or topic end and language
begin? Is it even possible to speak of literary style in such terms of sub-components
that are crudely divorced from one another? Certainly, debates have raged for many
decades over the definition of style and it is not likely that they will be resolved here
(Ellis 1970; Lang 1987).

Mitchell himself has noted that the construction of a historical stylistic imagi-
nary is not about total mimetic accuracy. Instead, he notes of the form that:

Historical fiction isn't easy; it's not just another genre. How are they going
to speak? If you get that too right, it sounds like a pastiche comedy—people
are saying "thou" and "prithee" and "gadzooks," which they did say, but to an
early 21st-century audience, it's laughable, even though it's accurate. So you
have to design a kind of "bygone-ese"—it's modern enough for readers not
to stumble over it, but it's not so modern that the reader kind of thinks this
could be out of House or Friends or something made for TV ("Interview with
David Mitchell" 2010).

In this reading, Mitchell notes that complete accuracy sounds alien and over-performed, though he does not go so far as to write about specific vocabularies and their (un)availability.

What I have shown in this article is that the construction of an imagined stylistic profile of a nineteenth-century text, as performed by David Mitchell in *Cloud Atlas*, has several unexpected characteristics, some of which chime with Mitchell's statements, others of which bring out something fresh. First, it is not about etymological mimesis. Readers are poor at identifying the first-usage dates of words and Mitchell's language – while extremely close to nineteenth-century reality – betrays itself in a small number of edge cases. Even if, as Rose Harris-Birtill has suggested, Mitchell's texts are concerned with reincarnation and repetition, the reincarnation of the nineteenth-century prose style that *Cloud Atlas* presents is a *differentiated* repetition; it is the same but different, a re-imagined representation of the target century from a twenty-first-century perspective, a historical imaginary of nineteenth-century stylistics, a "reincarnation time" for language (Harris-Birtill 2017, 166). It is as though there is a transmigration of language, accurate for the most part, but punctured by time (see Childs and Green 2011 for more on transmigration in Mitchell).

Second, the comparison of the frequency of outlandish terms that Mitchell's Ewing uses against a contemporary English corpus seems to affirm my conclusion. Certainly, such a claim is on shakier ground and there is a chance that some of the terms occur more frequently simply by chance. On the other hand, future studies may wish to test whether this hypothesis/tentative conclusion holds: that attempts to write in the style of the nineteenth century involve higher frequencies of archaic and racially abusive language than contemporary writing.

At the end of the day, though, it may be that the language in the Ewing section is not quite so outlandish compared to our own, or compared to that in the Sloosha's Crossin' section of the novel. Thematic concerns (such as the text's focus on tropical medicine – even though mediated through a Derridean *pharmakon* of medicine that both cures and kills (Dimovitz 2015, 71) – colonial violence, and seafaring narrative) inflect the word choices that Mitchell can make. It is not as though the use of

language is here selected in isolation from the thematic content. To set one's generic mode to the nineteenth century, in the early twenty-first century, requires not only a focus upon *how* one writes but also *what one writes about.*

There is another side to the language usage in this section of Mitchell's novel to which I must finally turn though. Thus far, aside from one mention, I have assumed here that Mitchell aims to straightforwardly achieve a mimesis of a nineteenth-century prose style and have used various computational techniques to appraise this. It could be, though, that there is an element of parody or pastiche in Mitchell's writing that deserves closer attention (even if Mitchell states, in the above-cited interview, that he aims to avoid pastiche). Theories of parody, such as Linda Hutcheon's famous formulation of ironic repetition and distancing (Hutcheon 1985), stress the need for both repetition and deviation. A parody must resonate with but also clash against its target so that readers may at once identify the target work of parody but also feel a distance from it. The fact that the language in Ewing Part I "doesn't ring quite true" and that this is acknowledged in the text itself should give us pause for thought. In the way in which the dissonance of etymological deviation can shine through – if one knows how and where to look – very little is lost in the "inaccuracy" of Mitchell's linguistic *parody.* It may in fact be this slightly off-kilter accuracy that transforms this section into a parody, thereby at least in part critiquing and neutralising the offensive colonial discourses of its characters. There is, therefore, at least one way in which David Mitchell's language choices in The Pacific Journal of Adam Ewing, whether conscious or not, must be read as political choices. For it is through the eyes of pastiche and parody that the discourses of this past space are made to seem ridiculous and outmoded (for more on this, see Dunlop 2011; also, Shoop and Ryan 2015 write of the politics of "big history" within the novel). Yet at least part of this effect comes from a re-imagined but punctured imagined stylistics of the nineteenth century.

What I also hope to have shown in this piece is that a close attention to quantitative aspects of *Cloud Atlas*, using computational methods, can bring us closer to understanding the text's features and style while unearthing fresh evidence that can then be re-incorporated into our existing humanistic methods: a form of symbiosis between the so-called digital humanities and their longer-standing traditional

disciplinary counterparts. This is not an attempt to replace people (or the humanities) with machines, as those hostile to the digital humanities often claim. But it is a type of cybernetic or bionic reading. Alone, the computational methods bring us data about the novel, but little more. What we make of the statistics of word frequencies and dates and how we understand their novelistic and political import remains a matter of hermeneutics. Overall, though, we need to also remember, when reading The Pacific Journal of Adam Ewing, that the language here is an excursion from our present, a mental voyage to an imagined stylistics of the nineteenth century. And, as Frobisher puts it, no matter how much evidence we excavate to the contrary of the section's accuracy, our suspension of disbelief is likely to remain within Mitchell's trip to the past. For, even with its twenty-first-century intrusions, it seems likely that, for most readers, "time cannot permeate this sabbatical" (Mitchell 2004a, 490, 2004b, 471).

Data Accessibility Statement

A full dataset for the work in this article, also using an expanded corpus, will be available in 2019. In the meantime, interested parties may contact the corresponding author, Martin Paul Eve.

Competing Interests

The author is a CEO of the Open Library of Humanities, which publishes *C21*. Double-blind peer review was overseen by an independent editor.

The editor would also like to acknowledge that whilst every attempt was made to create an anonymous version of this article for double-blind peer review, references to the author's previous research meant that the author's identity may have been discernible. However, both peer-reviewers noted that they felt that this did not prevent them from being objective.

References

Adamo, G., 2000. "Twentieth-Century Recent Theories on Beginnings and Endings of Novels." *Annali d'Italianistica* 18: 49–76.

Argamon, S., 2007. "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations." *Literary and Linguistic Computing* 23(2): 131–47. DOI: https://doi.org/10.1093/llc/fqn003

Begley, A., 2010. "David Mitchell, The Art of Fiction No. 204." *Paris Review*, Summer 2010. http://www.theparisreview.org/interviews/6034/the-art-of-fiction-no-204-david-mitchell.

Book World. 2004. "Q&A: Book World Talks With David Mitchell." *The Washington Post* August 22, 2004. http://www.washingtonpost.com/wp-dyn/articles/A17231-2004Aug19.html.

Burrows, J., 2002. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17(3): 267–87. DOI: https://doi.org/10.1093/llc/17.3.267

Childs, P., and Green, J., 2011. "The Novels in Nine Parts." In: *David Mitchell: Critical Essays*, Dillon, S. (ed.), 25–47. Canterbury: Gylphi.

de Groot, J., 2010. "Walter Scott Prize for Historical Fiction: The New Time-Travellers." *The Scotsman*. June 19, 2010. http://www.scotsman.com/lifestyle/books/walter-scott-prize-for-historical-fiction-the-new-time-travellers-1-813580.

Dimovitz, S., 2015. "The Sound of Silence: Eschatology and the Limits of the Word in David Mitchell's *Cloud Atlas*." *SubStance* 44(1): 71–91. DOI: https://doi.org/10.1353/sub.2015.0009

Dunlop, N., 2011. "Speculative Fiction as Postcolonial Critique in *Ghostwritten* and *Cloud Atlas*." In: *David Mitchell: Critical Essays*, Dillon, S. (ed.), 201–23. Canterbury: Gylphi.

Ellis, J.M., 1970. "Linguistics, Literature, and the Concept of Style." *WORD* 26(1): 65–78. DOI: https://doi.org/10.1080/00437956.1970.11435581

Eve, M.P., 2014. "'Some Kind of Thing It Aint Us but yet Its in Us': David Mitchell, Russell Hoban, and Metafiction After the Millennium." *SAGE Open* 4(1). DOI: https://doi.org/10.1177/2158244014521636

Eve, M.P., 2016a. *DateText*. https://github.com/MartinPaulEve/dateText.

Eve, M.P., 2016b. "'You Have to Keep Track of Your Changes': The Version Variants and Publishing History of David Mitchell's *Cloud Atlas*." *Open Library of Humanities* 2(2): 1–34. DOI: https://doi.org/10.16995/olh.82

Eve, M.P., 2017. "Close Reading with Computers: Genre Signals, Parts of Speech, and David Mitchell's *Cloud Atlas*." *SubStance* 46(3): 76–104. DOI: https://doi.org/10.3368/ss.46.3.76

Eve, M.P., 2018. "'What Was Knowledge for, I Would Ask Myself': Science, Technology, and Identity in David Mitchell's *Cloud Atlas*." In: *David Mitchell: Contemporary Critical Perspectives*, Knepper, W., and Hopf, C. (eds.). London: Bloomsbury.

Harris-Birtill, R., 2017. "'Looking down Time's Telescope at Myself': Reincarnation and Global Futures in David Mitchell's Fictional Worlds." *KronoScope: Journal for the Study of Time* 17(2): 163–81. DOI: https://doi.org/10.1163/15685241-12341382

Hicks, H.J., 2010. "'This Time Round': David Mitchell's *Cloud Atlas* and the Apocalyptic Problem of Historicism." *Postmodern Culture* 20(3). DOI: https://doi.org/10.1353/pmc.2010.0012

Hoover, D.L., 2004. "Testing Burrows's Delta." *Literary and Linguistic Computing* 19(4): 453–75. DOI: https://doi.org/10.1093/llc/19.4.453

Hopf, C., 2011. "The Stories We Tell: Discursive Identity Through Narrative Form in *Cloud Atlas*." In: *David Mitchell: Critical Essays*, Dillon, S. (ed.), 105–26. Canterbury: Gylphi.

Houston, K., 2013. *Shady Characters: The Secret Life of Punctuation, Symbols, & Other Typographical Marks*. New York: W.W. Norton & Company.

Hutcheon, L., 1985. *A Theory of Parody: The Teachings of Twentieth-Century Art Forms*. New York: Methuen.

"Interview with David Mitchell." 2010. *Goodreads*. July 5, 2010. https://www.goodreads.com/interviews/show/537.David_Mitchell.

Kermode, F., 2000. *The Sense of an Ending: Studies in the Theory of Fiction*. Oxford: Oxford University Press.

Lang, B., (ed.). 1987. *The Concept of Style*. Ithaca, N.Y: Cornell University Press.

Melville, H., 2008. *Moby-Dick; or, The Whale*. Project Gutenberg. http://www.gutenberg.org/files/2701/2701-h/2701-h.htm.

Miéville, C., 2012. *Railsea*. London: Pan.

Mitchell, D., 2004a. *Cloud Atlas.* London: Sceptre.

Mitchell, D., 2004b. *Cloud Atlas.* New York: Random House.

Mitchell, D., 2010. "Guardian Book Club: *Cloud Atlas* by David Mitchell." *The Guardian.* June 12, 2010. http://www.theguardian.com/books/2010/jun/12/book-club-mitchell-cloud-atlas.

O'Donnell, P., 2015. *A Temporary Future: The Fiction of David Mitchell.* New York: Bloomsbury Academic.

Oxford Dictionaries. 2017. "The Oxford English Corpus." Oxford Dictionaries: English. 2017. https://en.oxforddictionaries.com/explore/oxford-english-corpus.

Pynchon, T., 1997. *Mason & Dixon.* London: Jonathan Cape.

Shoop, C., and Ryan, D., 2015. "'Gravid with the Ancient Future': *Cloud Atlas* and the Politics of Big History." *SubStance* 44(1): 92–106. DOI: https://doi.org/10.1353/sub.2015.0011

"Spillage." 2017. Online Etymology Dictionary. https://www.etymonline.com/word/spillage.

Stein, S., and Argamon, S., 2006. "A Mathematical Explanation of Burrows' Delta." In: *Proceedings of Digital Humanities 2006.* Paris, France.