Abstract

This project puts forth two goals. First, how did Du Bois understand the interrelationship between democracy and the sciences? (The latter for him include the social sciences, the natural sciences, and often history). Second, how can we approach such an interconnection of ideas within Du Bois's numerous writings?

I outline an approach called interpretive concordancing, which allows us to explore a corpus via a concordancer using specific search protocols (i.e., regexes: regular expressions). For this project I created a partial corpus of about 200 texts. Interpretive concordancing involves an analytical phase in which the texts are studied in terms of their component parts: namely, the words and their synonyms that express Du Bois's ideas. During this phase I list n-grams of 2- and 3-word phrases, as well as seek single words or phrases (called node words) within their textual contexts. Also, I conduct proximity searching for word pairs that occur a certain number of characters apart. Lastly, in the synthesizing phase of interpretive concordancing I assemble the ideas of the texts, as evidenced in the search results, into an argument that conveys Du Bois's understanding of the democracy/science relationship.

Interpretive concordancing leads me to this argument. For Du Bois, democracy and science each supports the other and each limits the other. Within a specified domain of action and policy (my phrase) one predominates due to the other's limited scope of knowledge. The sciences, by grasping the natural laws of economic production, could tell us "how" to produce and distribute goods and services, but not the "whats", the "whys", and the "how-muchs". That was the realm of citizens and democratic participation. Each generates only a portion of the overall
I. W.E.B. Du Bois

[1] W.E.B. Du Bois (1868-1963), the civil rights activist and scholar, is increasingly studied in various academic fields. **Political theorists have studied his views on democracy**, especially voting, civil rights, and political equality and freedom (*e.g.*, Balfour 2011; Basevich 2018 & 2019; Gooding-Williams 2009 & 2017). **The social sciences** of sociology and urban studies, as well as the fields of education, history, and public health, have examined his social research (*e.g.*, Alridge 2008; Jones-Eversley & Dean 2018; Monteiro 2008; A. Morris 2015; Outlaw 2000; Rabaka 2010; R.W. Williams 2006 & 2018a; E. Wright II 2016). Du Bois hoped that such research would change White views on racism and could inform governmental policies to alleviate discrimination, exploitation, and poverty by means of equity in education, health care, and job opportunities (MEPF 1944).

### A. Need for Scholarship on Democracy and Science

[2] Given the academic emphases of study, the relationship between Du Bois's views on democracy and the sciences can be a fruitful area of research. However, the two focal points are not fully explored. When scholars approach Du Bois on democracy, the role, including the delimiting role, of the sciences is not typically examined. And when the social sciences are discussed, the natural sciences are not typically covered (although evolutionary theory might be briefly mentioned: *e.g.*, A. Morris 2015).

[3] Why is studying Du Bois's understanding of democracy *vis-à-vis* science important? That particular interrelationship speaks to us in the 21st Century in the form of current and recent public debates—an ideal of liberal democracies—over the effects of second-hand smoking, human-induced climate change, and policies to mitigate and halt the COVID 19 pandemic. In the public spaces scientific information as well as misinformation and distortions abound and abide. Du Bois sets us on the path to examine the respective roles of science and democracy. Admittedly, he did not fully address all of the nuances (which I briefly mention in the conclusion).

### B. Two Questions Animate My Project

[4] First: How did Du Bois understand the interrelationship between democracy and the sciences? (The latter for Du Bois included the social sciences, the natural sciences, and often history). Examining this relationship will lead us to explore the ways in which he expresses his understanding of democracy and science in terms of related (*i.e.* companion) ideas.

[5] Second: How can we approach such an interconnection of ideas through an exploration of Du Bois's vast body of writings? (According to Aptheker (1973), Du Bois's corpus of works, spanning 60 years, exceeds 1900 pieces, including 21 books, hundreds of essays and newspaper articles, and seemingly countless pieces of correspondence, mostly unpublished. Various repositories archive...
unpublished pieces as well as some drafts of published works).

C. Outline of the Project
[6] The project unfolds below as follows:
• Section 2: Interpretive concordancing allows us to explore a corpus using specific search protocols (regular expressions) via a concordancer. Interpretive concordancing involves an analytical phase in which the texts are studied in terms of their component parts: namely words as expressive of Du Bois's ideas.
• Section 3: I will discuss seeking single words or phrases (called node words).
• Section 4: Here I sketch proximity searching for separate words that occur a certain number of characters apart.
• Section 5: I present the synthesizing phase of interpretive concordancing in which the components of the texts—as delineated by the concordancer's search results—are assembled into an argument that sketches Du Bois's understanding of the relationship between democracy and science.

[7] A note on the conventions used in this project:
• The in-text citations to Du Bois's works include an abbreviated title and year of publication. In the bibliography his works are alphabetized by those abbreviated titles.
• Regular expressions, especially those used for proximity searching, are denoted as such: (?i)\bscience.
• Words (or their fragments) that are part of the research process are designated as such: »democra«.

II. Concordancing for the Theoretically Inclined
[8] Over the years I have utilized the conventional techniques of political theory: close reading of documents and their historical contextualization. Both are well established tools for understanding texts and their explicit or implicit political ideas (Charette & Skjönsberg 2020). In this project I wish to outline concordancing as a tool to employ on a corpus of texts, and to sketch the associated techniques of regular expression matching. All of these I have adapted from corpus linguistics and computer science.

A. The Tools and Techniques
[9] A corpus generally includes a massive number of texts or words written or transcribed from spoken sources (Meyer 2004; Sinclair 1991 & 2003). A corpus typically consists of many authors/speakers. Corpora can be composed a writings from a country, genre, time period, intellectual tradition, or any combination. My corpus is focused on only one author: W.E.B. Du Bois. It is a partial corpus of his public domain texts.

Partial corpus of public domain texts by Du Bois: 133 files
• Types: 28,751 • Tokens: 1,203,341
[Note 1—Corpus Creation]
In preparing the texts to be studied with a concordancer, various decisions must be made (J. Sinclair 1991).

- First, locate the author's work: was it published, or is it unpublished? (Do we utilize the latter?)
- Second, is the work in a plain-text, "machine-readable" format? If not, convert it.
- Third, prepare the plain-text file for scholarly study by addressing what to retain of the content (ex., the text), what to add (ex., paragraph numbers; [sic]), and / or what to exclude (ex., photographs, drawings).

We must locate the document of the author's work that we wish to study. Online searches can yield a wealth of sources. Any document so located might already be in the format that we need: plain text without the display codes and formatting of a word-processor. If not we must convert it.

There are several ways to convert an existing document into an e-text that a concordancer can process. If the extant document is in print form, the words can be retyped. Or the pages can be photographed and then scanned via an OCR (Optical Character Recognition) application, which converts the image file into machine-readable format, such as plain text or a word processor document. Whether retyped or OCRRed the resulting text must be double-checked for accuracy with the text of the extant document. Errors are easy to introduce into the concordancer-usable file, regardless of whether there were typos in the starting document.

Once we have a plain text file, we will still need to prepare the e-text for scholarly study. I made several choices about what to include, omit, and modify within the texts. My choice build upon a document that involves the choices of translators, archivists, librarians, and scholars, as well the associated professional norms. Accordingly, I did not start with an object created by natural processes. A text is not a fossil generated by biological and geological forces.

I fashioned the documents for use by concordancers with at least one or more of the following choices:

(a) converting characters that do not render correctly when displayed (ex., accent marks [é è ï ñ etc.], curved ‘single’ and double “quotation” marks, en and em dashes [— —], and ellipsis…) to their counterparts in the less varied character set of ASCII;
(b) adding paragraph numbers and / or retaining pagination—both of which might not have occurred in the original versions; and both of which could be [bracketed] in the processed version; and
(c) designating typos, misspellings, and outdated words by "[sic]" (and also in some cases specifying conventional spellings, so as to enhance the possibility of matching the word in a search);
(d) identifying speakers by name when there may be no name present in the original (which is often encountered as the first word of a passage, and which may or may not have square brackets enclosing the name (ex., CRITO, ANTIGONE), and which I probably retained from the edition that I used);
(e) clarifying places, background details, staging directions, etc., by typically
enclosing such words and sentences in square brackets (which editors and translators also did with some of the texts).

[c6] One consequence of such prepared documents: indicators of errors, clarifying details, and proper names for speakers (if not in the original), as well as paragraph and page numbers all will be displayed in various lists generated by the concordancer. Such additions to the documents also would be calculated by any statistical measures employed.

[c7] Although the following situations did not occur in the included set of texts for the Bennett College courses, we also need to address:

(f) incorporating unpublished documents, ex., drafts (do such unpublished materials tell us anything about the author's thoughts on a topic?);

(g) rendering corrections made by hand or machine (which would require some protocol for standardizing corrections, marginalia, etc.); and

(h) excluding graphics, drawings, photographs, because there is little or no text per se (we can retain captions or add them); as well as excluding tables if their data is not relevant to the research inquiry itself (again, adding or retaining a label describing what is omitted).

[c8] We will recognize that by excluding tables, graphics, photographs, etc. we have lost not only some information that the original document conveyed, but also we are not fully presenting what the author had intended. Ultimately, we must decide whether the losses are an acceptable trade-off in the name of digital analysis and interpretation.

[c9] There are other formats that can and should be considered for texts to be used in digital research. The Text Encoding Initiative format is recognized as a standard. For future versions of a Du Bois corpus such will be important.

[End of Note 1.]

---

**A concordancer** is a software program designed to explore and analyze texts; they excel at working through corpora. Concordancers perform tasks that are integral to corpus linguistics and digital humanities research (Bradley 2004; Hockey 2004; Luz & Sheehan 2020; McEnery & Hardie 2012; Sinclair 1991 & 2003). With a concordancer, we can:

- view a word, or a part of one, within its surrounding passages (KWIC, or Keyword in Context);
- count the frequency of words or characters in a document;
- list adjacent words by groups of 2, 3, 4, or 5 (n-grams);
- discover the words in the vicinity of a search term (collocations).

In addition, AntConc and others incorporate statistical tools to perform various quantitative measures.
Figure 1: AntConc Concordancer (Laurence Anthony)

Numerous concordancers are available in both free and paid versions. Useful websites listing such tools are:
http://martinweisser.org/corpora_site/concordancers.html
https://corpus-analysis.com/tag/concordancer.html

I use AntConc, which was created by Dr. Laurence Anthony
https://www.laurenceanthony.net/software/antconc

#Lancsbox (which the developers call a "corpus toolbox")
http://corpora.lancs.ac.uk/lancsbox

CasualConc (for MacOS)
https://sites.google.com/site/casualconc

[End of Note 2.]

[11] A regular expression (regex) is a set of characters and metacharacters (which are alphanumeric characters performing certain functions) that is designed match a pattern. The pattern sought is a string that itself could be composed of characters forming words or computer code. We write a regex
as a sequence of "tokens", as they are called. The regex tokens include ways to repeat sequences and/or to find alternatives at a position within the text (string) being searched. Applied to a document, we use regexes to find potential match(es), which then are displayed in the concordance window of the concordancer.

[Note 3–Regexes]

Several very useful sources of information on regular expressions include:
- Jan Goyvaerts at [https://regular-expressions.mobi](https://regular-expressions.mobi).
- [https://www.rexegg.com](https://www.rexegg.com)
- [https://regular-expressions.mobi](https://regular-expressions.mobi)
- [http://regextutorials.com](http://regextutorials.com)
- [https://riptutorial.com/regex](https://riptutorial.com/regex)
- Microsoft: .NET Regular Expressions
- Ryan's Tutorials: Regular Expressions

[End of Note 3.]

B. Interpretive Concordancing

[12] Concordancing as performed by corpus linguists seeks to discern patterns in word usage and recurring phrases within the corpora, from which explanations can be hypothesized and tested as regards why such patterns appear (McEnery & Hardie 2012; Stefanowitsch 2020). I advocate here for a research approach that can be called interpretive concordancing. Its goal is to facilitate searching and discovering the ways—indeed, the paths by which—Du Bois conveyed ideas to us by means of and across his texts. It does not utilize the statistical procedures of the social sciences. (Interpretive concordancing, as I label it, is similar to the procedures proposed by Rockwell and Sinclair, who subheaded a chapter in *Hermeneutica* "Concordance as an Interpretive Tool" (2016: p.46). They, however, offer more tools than concordancers for computer-assisted analysis on their website <voyant-tools.org>.

[13] Interpretive concordancing supplements, not supplants, close reading. We could and should read every text individually, for this is still a necessary part of interpretation. Without a concordancer, we still could search individual files with a browser or e-reader. Such remains a useful method to the extent that the digital files are searchable with a text layer. But the difficulty lies in viewing all of the results as a collection of search hits. Here emerges the domain and power of concordancing (Bonelli 2010; Rockwell & Sinclair 2016: Ch.3), interpretive or otherwise.

C. Similarities: Interpretive Concordancing vis-à-vis Quantitative Approaches
Both the (chiefly) quantitative and conventionally interpretive approaches utilize analytical and synthesizing phases in the research process (Luz & Sheehan 2020; Rockwell & Sinclair 2016: Ch.10).

- In the **analytic phase** of research the text is broken into constituent elements, such as words. Comparisons can be pursued with other words or via word frequencies relating to other texts.
- In the **synthetic phase** of research the scholars intervene to understand what she/he/they have uncovered via the algorithms. After all, the search results of the located words and the statistical representations of the texts do not speak for themselves.

Both the quantitative and interpretive approaches seek patterns in the evidence (data for the former, and ideas for the latter). Patterns provide insights into the evidence being explored. If some phenomenon reoccurs in the data, then this might indicate that it is non-random and is not necessarily by chance. Hence, the pattern could be intentionally and/or causally created, whether by human decisions or by social and natural forces at work. In short, there is something meaningful to be studied (Bod 2013).

Despite the commonalities, the interpretive and quantitative approaches are also distinctive.

**D. Differences: Interpretive Concordancing vis-à-vis Quantitative Approaches**

The social sciences that model themselves after the natural sciences will gather and aggregate data relevant to human thoughts and behaviors. As such, the data are treated as discrete bits of information devoid of intrinsic connections with other bits of information, rather than as indicators of human experiences, irreducible and meaningful. (cf. Drucker's 2012b critique of reductionism in digital humanities).

For example, for some quantitative-based techniques the documents studied are simply bags of words which can be statistically compared to other corpora to determine whether the word pairs or combinations (n-grams) are more likely to occur together, especially when compared to a reference corpus.

As a consequence, science-based approaches that study concepts and language typically do not consider as scientifically relevant the author's agency over the creation of the sentences and paragraphs. Authors are working within systems and structures of language which transcend them.

To be fair, the humanities have types of structuralist and post-structuralist theories that proclaim, as Barthes's wrote, the "death of the author" (Barthes 1968).

Interpretive approaches of the humanities, including in many ways political theory, typically seek to foreground the authors and their agency in the creation of the artifacts. Such approaches ask: How do the ideas express the author's understanding of the world—or a textually based plausible range of understandings? Such perspectives of the writers studied by theorists evoke various questions: How does that (political) world or cosmos operate? What political order is being
suggested? Who has agency? What are the duties of the ruled and the rulers? What is human nature? Are the ideas and ideals of freedom, equality, and justice present and, if so, what do they mean normatively and practically.

[19] In another area of difference, the sciences and humanities address the patterns they encounter in the phenomena in alternate ways. For the sciences, patterns in the data can prompt further study to discern possible generalizations that arise from (probabilistic) causal factors. Such causal factors are discovered via statistical tools and are conveyed with statistical measurements. For the humanities, patterns bespeak the themes of an author or authors, or else might point to a theme that unites various authors into a tradition or genre, perhaps as posited by scholars (Hirsch 1967). Moreover, patterns in the textual evidence help us to discover the uniqueness and contributions of an individual writer within a discernible tradition of thought. Statistics are not necessary to convey the interpretation of the phenomena (i.e., the textual evidence).

E. Assumptions of Interpretive Concordancing

[20] First, the digital texts are artifacts derived from the original documents themselves. The digitalization process renders, indeed reduces, the original physical object into the bits and bytes of files, whose zeros and ones convey the words in the documents so rendered (Dobson 2019; Rockwell & Sinclair 2016: Ch.2). The digital version represents—models (McCarty 2005)—the original for the purposes of algorithmic analysis. By so doing, the scholars make choices as to what they wish to retain from the original artifact (e.g., its pagination and hyphenation at line breaks; marginalia and added commentary; manuscript deletions, substitutions, additions, and transpositions; and so forth).

[21] Second, any results discovered via the procedures are mediated by the concordancer and the techniques used with it, such as regular expressions. Computer mediation refers to the ways that computers and digital artifacts facilitate what we are seeking, but paradoxically also influence what we find (Dobson 2019; Ramsay 2011; Rockwell & Sinclair 2016). And such paradoxes could also become lethal in discriminatory situations with racially marginalized communities (R. Benjamin 2019).

[22] Third, concordancer-mediated exploration is only as good as the corpus that was created (Reppen 2010; Sinclair 1991). In my personal experience, corpus creation has been the most time-consuming activity of the overall research process.

[23] Fourth, ideas conveyed via words in the document are not limited to specific words alone, but can be defined by synonymous terms which could be ideas—hence, companion ideas—in their own right (Halliday 2004; Hampsher-Monk et al. 1998; Richter 1995; also see Alfano 2018; Danis & Meunier 2012; Davies 2015; Hunston 2010; Tribble 2010). That is to say, Du Bois can express an idea like democracy in morphologically different but conceptually related ways across multiple texts. Suffrage can be the synecdoche of democracy, but democracy involves more than voting. Democracy also could be defined (ideally) in terms of the freedom and equality of the citizens...
practicing it. In turn, freedom and equality are not limited only to democracy as a form of government and its participatory ideals. Indeed, freedom and equality also could be studied in relation to individual actions outside of institutional political activities, such as the social equality of interpersonal relations between races and ethnicities.

[24] Fifth, the meaningfulness of ideas does not derive from only one text, but includes other texts, as well as the overall intellectual milieux of the writer's time and place. Hence, to understand ideas requires us to grasp their intertextuality (G. Allen 2000; Kristeva 1986a & 1986b). Auto-intertextuality holds that the ideas of a single author in all of their nuances, applications to events, and modifications, are conveyed across multiple texts within the author's corpus (Kreiswirth 1996; R.W. Williams 2016).

F. Caveats of Interpretive Concordancing

[25] Assumptions denoted in the negative I will call the caveats to interpretive concordancing. First: We must not assume that there is a necessary coherence to the person's or school's thought. There may be contradictions and inconsistencies between and within particular works. Second: Common themes in an author's works may be directly expressed by the author her-/himself or else inferred by later scholars, but only insofar as there are textual bases for the inferences. Third: We must not assume that there is some singular and/or "pre-destined" path to the development of the author's or school's thought. Fourth: We must not assume that the author had the same purpose or goals as later readers and scholars. (cf. Borges, "Kafka and His Precursors"). Fifth: We must not assume that the writer's ideas allow for only one interpretation or only one set of meanings and implications.

The preceding caveats are based on, and are my extension of, Adolph Reed's admonitions about interpreting Du Bois (Reed 1997).

[26] In general, the work flow of concordancing research is iterative in the manner of close reading. We move through the texts in a sequential, or perhaps a back-and-forth manner, while we are pursuing the ideas in their similarities and differences. With those iterations we could be examining n-grams, which are groups of 2 or several words, or else "node words", which are the individual words or phrases on which we focus. We also can examine such node words with the KWIC (Keyword in Context) function of the concordancer. Moreover, we can reach beyond single words or phrases by means of proximity searching which seeks to find (at least) two ideas near each other in the texts.

III. Work Flow: N-Grams and Node Words (Analytical Phase)

[27] In this section, I use the concordancer to generate bigrams and trigrams. Such n-grams can point to patterns in word usage, especially in terms of higher frequencies. I will also utilize the Keyword in Context (KWIC) of the concordancer which will allow use to examine the surrounding texts of the search terms that we locate within the corpus (Sinclair 2003). Both node words and
n-grams can provide the basis for further exploration of the corpus.

A. Bigrams and Trigrams: "Democracy", "Science, and Variants"

[28] The following results derive from a partial corpus of Du Bois's public domain texts. AntConc presents the results as lists with frequency counts, but no specific document is associated with the n-grams.

**Figure 2: Bigrams and Trigrams of »democra«**

| democratic despotism                      | democratically controlled state |
| democratic weapons                        | democratization of industry    |
| free democracy                            | free democratic organizations  |
| world democracy                           | free democratic government     |
| democratic control                        | all embracing democracy       |
| industrial democracy                      | experiment in democracy        |
| abolition democracy                       | extend the democratic         |
| a pure democracy                          | fundamental democratic foundation |
| a real democracy                          | ideals of democracy           |
| a vulgar democracy                        | theoretical democracy but     |
| advance of democracy                      | theory of democracy           |
| road of democracy                         | republicans and democrats     |

**Figure 3: Bigrams and Trigrams of »scien«**

| agricultural scientist                     | adequate scientific foundation |
| inorganic sciences                         | scientifically demonstrated to |
| scientific accuracy                        | scientifically made certain   |
| scientific induction                       | scientifically treated to     |
| scientific spirit                          | slightest scientific warrant  |
| supposed scientific                        | splendid scientific work      |
| scientific dogma                           | strong scientific backing     |
| scientific inquiry                         | unbiased scientific study     |
| scientific law                             | unprejudiced scientific interest |
| scientific measurement                     | scientific truth              |
| boundaries of scientific                   | this false scientific         |
| scientific exploitation of                 | unjustifiable scientifically  |
| government and science                     | without adequate scientific   |

[Note 5–N-Gram: Details]

AntConc 3.5.9. (2020) was used to generate the n-grams.

• **All Bigrams** in the Public Domain Partial Corpus of W.E.B. Du Bois

Parameters:
### B. Keyword in Context (KWIC) with a Concordancer

[30] Here we use the concordancer to examine "democracy" and "science" as separate search terms (node words). The n-grams from the earlier lists can also be sought, but space and time does not permit that in this incarnation of my project.

[31] When we concordance »democra« we discover over 400 matches in the public domain, partial corpus. The matches reside in about 45 individual documents (as determined via the "Concordance Plot" feature in AntConc). The figure below displays the AntConc program with 20 concordance lines visible.

---

**[Notes]**

- All Trigrams in the Public Domain Partial Corpus of W.E.B. Du Bois
- Sorted by Freq | Cluster Size: 3 | Max. 3 | Min. Freq. 1 | Min. Range 1

<table>
<thead>
<tr>
<th>Trigrams of »(?i)democra«</th>
<th>Results: 921</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigrams of »(?i)scien«</td>
<td>Results: 1258</td>
</tr>
</tbody>
</table>

---

[29] N-gram results permit us to find patterns in the word groups, which are integral to corpus linguistics and digital humanities. Via the KWIC feature phrases can be sought, as well as the the individual words of the n-grams. Moreover, the n-grams in the figures above disclose various companion ideas to "democracy" and "science". For example, the idea of "foundation" occurs with both concepts. Also, science is often characterized by experimentation, and Du Bois connected democracy to experiment. In addition, is there a conceptual analogy between the bigrams of "vulgar democracy" and "false scientific"?

This great vision of the black man was, of course, at first the vision of the few, as visions always are, but it was always there; it grew continuously and it developed quickly from wish to [p.139] active determination. One cannot think then of democracy in America or in the modern world without reference to the American Negro. The democracy established in America in the eighteenth century was not, and was not designed to be, a democracy of the masses of men and it was thus singularly easy for people to fail to see the incongruity of democracy and slavery. It was the Negro himself who forced the consideration of this incongruity, who made emancipation inevitable and made the modern world at least consider if not wholly accept the idea of a democracy including men of all races and colors. [GBF 1925: pp.138-139]

This passage repeats a common theme across Du Bois's life: African Americans possessed the agency to struggle against the oppressors. Also note his criticism of the founding of this country.
[33] For another example of Du Bois considering "democracy" in an historical and world-wide perspective, we turn to his "African Roots of War" published during World War I.

[¶ 31] [...] We shall not drive war from this world until we treat them [Black men and women] as free and equal citizens in a world-democracy of all races and nations. Impossible? Democracy is a method of doing the impossible. It is the only method yet discovered of making the education and development of all men a matter of all men's desperate desire. It is putting firearms in the hands of a child with the object of compelling the child's neighbors to teach him not only the real and legitimate uses of a dangerous tool but the uses of himself in all things. Are there other and less costly ways of accomplishing this? There may be in some better world. But for a world just emerging from the rough chains of an almost universal poverty, and faced by the temptation of luxury and indulgence through the enslaving of defenseless men, there is but one adequate method of salvation—the giving of democratic weapons of self-defense to the defenseless. [AROW 1915]

Notable in this passage are the provocative ideas of a "world-democracy and "democratic weapons". Questions arise: Where else did he use those terms? Is he talking about a world governmental system? It is quite interesting to contemplate why he evokes the imagery of weapons during a global war.

C. KWICing "Science" and Its Variants

[34] We continue our concordancing with the node word "science" and its variants. The regex `(?i)\bscience` matches "science" or "sciences" in public domain partial corpus (approximately 250 results in about 50 files). The regex `(?i)scientific` as the node word yields about 180 results in approximately 40 files. Searching with a word stem, such as `\bsci\b` for the node word will match more variants on its passage through the corpus. That word stem finds over 400 matches. The figure below display the resulting concordance lines.
[35] Here is a quotation from "The Atlanta Conferences".

The present condition of sociological study is peculiar and in many respects critical. Amid a multitude of interesting facts and conditions we are groping after a science—after reliable methods of observation and measurement, and after some enlightening way of systematizing and arranging the mass of accumulated material. Moreover the very immensity of the task gives us pause. What after all are we trying to do but to make a science of human action? And yet such a task seems so preposterous that there is scarce a sociologist the world over that would acknowledge such a plan. Rather, turning from so startling a task, they have assured the world that the object is to study a certain metaphysical entity called society—and when they have been asked earnestly and rather insistently just what society is, they have replied in language and once curious, mystical and at times contradictory. Has not the time come however when we should face our problem? In reality we seek to know how much of natural law there is in human conduct. Sociology is a science that seeks to measure the limits of chance in human action, or if you will excuse the paradox, it is the science of free will. [...] [ATLC
Notice the ideas in the vicinity of science. Du Bois mentioned his quest after a "science of human action" in many places across his life. Indeed, in the terminology of corpus linguistics it is a collocate within Du Bois's corpus. Also, the ideas of "chance" and "free will" are interesting companion ideas that would bear further study.

[36] We continue our searching for "science", "scientific", even "scientist", with \textit{scient}. Below is an important passage for Du Bois's interpretation of the role of evolutionary theory. Within a section detailing how science has been used to justify the oppression of persons of color, Du Bois wrote:

[14] Why is this? It is because the splendid scientific work of Darwin, Weissman, Galton and others has been widely and popularly interpreted as meaning that there is such essential and inevitable inequality among men and the races of men as no philanthropy can or ought to eliminate; that civilization is a struggle for existence whereby the weaker nations and individuals will gradually succumb and the strong will inherit the earth. With this interpretation has gone the silent assumption that the white European stock represents the strong surviving peoples and that the swarthy, yellow and black peoples are the ones rightly doomed to eventual extinction. [EORP 1909: ¶ 14 (also JB 1909: pp.375-376)]

We will notice Du Bois's praise of Darwin and Galton. The latter is infamous as an advocate for eugenics. The passage above and the next one in his essay emphasize a DuBoisian argument against the so-called survival of the fittest that characterized what is called Social Darwinism.

[Note 6–Du Bois's Criticism of Galton]

In \textit{Black Folk Then and Now} (1939) Du Bois criticized Galton:

\begin{quote}
English science in the hands of Galton and Pearson and their ilk made English aristocracy rulers by divine right. The vast possibility of a pool of human knowledge as wide as the living world never arrested their attention. They could not imagine that the freedom and development of all men would make in time a world Renaissance, beside which the little European Renaissance would seem small and petty; that art and science could look forward to greater and more wonderful conquests, if they contemplated using the ability of all the world and not simply a narrow section. [BFTN 1939: p.373]
\end{quote}

[End of Note 6.]

[37] The passage just quoted does not specifically connect science directly with democracy per se. But by reading a few paragraphs further into the essay, we glimpse Du Bois linking aspects of democracy to evolutionary theory (as per Du Bois's understanding of Darwinism).

[19] What the age of Darwin has done is to add to the eighteenth century idea of individual worth the complementary idea of physical immortality of the human race. And this, far from annulling or contracting the idea of human freedom, rather emphasizes its necessity and eternal possibility—the boundlessness and endl...
possible human achievement. Freedom has come to mean not individual caprice or aberration but social self-realization in an endless chain of selves, and freedom for such development is not the denial but the central assertion of the evolutionary theory. So, too, the doctrine of human equality passes through the fire of scientific inquiry not obliterated but transfigured; not equality of present attainment but equality of opportunity for unbounded future attainment is the rightful demand of mankind. [EORP 1909]

This passage from the "Evolution of the Race Problem" clearly expresses the important role of science in social justice. The passage also contains numerous possible companion ideas to search for in Du Bois's corpus. Ideas such as "equality" and "freedom" are obvious choices. But an idea like "social self-realization" intrigues me, as does Du Bois's idea of time that is conveyed in words like "eternal", "future", "boundlessness", "endlessness", and so forth.

[38] The previous examples illustrate not only how Du Bois defined democracy and science, but also links them to companion ideas. Such ideas can emerge as part of the search results derived from simple concordancing with a word or phrase. Nonetheless, an idea articulated by a phrase like "science of human action" can also be arranged differently, such as via grammatical metaphors (Halliday 2004): for example, "Sociology is a science that seeks to measure the limits of chance in human action" (also in ATLC). With such alternate forms of expressing and expanding the intellectual reach of one idea has emerged the importance of proximity searching.

IV. Work Flow: Proximity Searching (Analytical Phase)

[39] The purpose of proximity searching is to locate ideas in relation to one another within a certain distance. This is accomplished—is mediated (both influenced and facilitated)—by searching a corpus with regular expressions (regexes) by means of a concordancer. Such a process will narrow the number of texts so as to allow us to more closely examine how Du Bois interconnected those ideas.

A. The Technique of Regular Expressions

[40] The central relationship discussed in this project: democracy and its variants vis-à-vis science and its variants can be addressed via a regex such as this:

\{re-1\} (?i)democra(\?::)\{0,400\}?\bscien|\bscien(\?::)\{0,400\}?democra

[41] Running that regex locates an interesting pairing in a chapter of Darkwater, "Of the Ruling of Men".

That the problem of the democratization of industry is tremendous, let no man deny. We must spread that sympathy and intelligence which tolerates the widest individual freedom despite the necessary public control; we must learn to select for public office ability rather than mere affability. We must stand ready to defer to knowledge and science and judge by result rather than by method; and finally we must face the fact that the final distribution of goods—the
question of wages and income is an ethical and not a mere mechanical problem and calls for grave public human judgment and not secrecy and closed doors. All this means time and development. It comes not complete by instant revolution of a day, nor yet by the deferred evolution of a thousand years—it comes daily, bit by bit and step by step, as men and women learn and grow and as children are trained in Truth. [DARK 1920: Ch. VI (=OROM: ¶ 61)]

Highlighted is one of the recurring themes in Du Bois's corpus: industry must be democratized, especially in terms of government oversight and constraint on the oligarchs of business, as he terms them. Moreover, several questions arise from Du Bois's passage. What do the following ideas mean?

- "defer to knowledge and science"
- "judge by result rather than by method" [Is this pragmatism?]
- "the question of wages and income [...] calls for grave public human judgment"
- "development [...] comes not complete by instant revolution of a day, nor yet by the deferred evolution of a thousand years"

What other companion ideas arise from this passage?

[42] Because interpretive concordancing augments other research techniques like closer reading, then to address our question we can read more of this chapter inDarkwater. For example, a few paragraphs earlier Du Bois became very explicit about the relationship of democracy and science (OROM ¶¶ 56-59). Du Bois set up his mountaintop/valley metaphor.

[...]. Here in the heavens and on the mountaintops, the air of Freedom is wide, almost limitless, for here, in the highest stretches, individual freedom harms no man, and, therefore, no man has the right to limit it.

On the other hand, in the valleys of the hard, unyielding laws of matter and the social necessities of time production, and human intercourse, the limits on our freedom are stern and unbending if we would exist and thrive. This does not say that everything here is governed by incontrovertible "natural" law which needs no human decision as to raw materials, machinery, prices, wages, news-dissemination, education of children, etc.; but it does mean that decisions here must be limited by brute facts and based on science and human wants.

Today the scientific and ethical boundaries of our industrial activities are not in the hands of scientists, teachers, and thinkers; nor is the intervening opportunity for decision left in the control of the public whose welfare such decisions guide. On the contrary, the control of industry is largely in the hands of a powerful few, who decide for their own good and regardless of the good of others. [...]. [DARK 1920: Ch. VI (=OROM: ¶¶ 56-58)]

[Note 7–OROM: ¶¶55-59]

In "Of the Ruling of Men" Du Bois wrote:

[55] In this intricate whirl of activities, the theory of government has been hitherto to lay down only very general rules of conduct, marking the limits of extreme anti-social acts, like fraud, theft, and murder.
The theory was that within these bounds was Freedom—the Liberty to think and do and move as one wished. The real realm of freedom was found in experience to be much narrower than this in one direction and much broader in another. In matters of Truth and Faith and Beauty, the Ancient Law was inexcusably strait and modern law unforgivably stupid. It is here that the future and mighty fight for Freedom must and will be made. Here in the heavens and on the mountaintops, the air of Freedom is wide, almost limitless, for here, in the highest stretches, individual freedom harms no man, and, therefore, no man has the right to limit it.

On the other hand, in the valleys of the hard, unyielding laws of matter and the social necessities of time production, and human intercourse, the limits on our freedom are stern and unbending if we would exist and thrive. This does not say that everything here is governed by incontrovertible "natural" law which needs no human decision as to raw materials, machinery, prices, wages, news-dissemination, education of children, etc.; but it does mean that decisions here must be limited by brute facts and based on science and human wants.

Today the scientific and ethical boundaries of our industrial activities are not in the hands of scientists, teachers, and thinkers; nor is the intervening opportunity for decision left in the control of the public whose welfare such decisions guide. On the contrary, the control of industry is largely in the hands of a powerful few, who decide for their own good and regardless of the good of others. The making of the rules of Industry, then, is not in the hands of All, but in the hands of the Few. The Few who govern industry envisage, not the wants of mankind, but their own wants. They work quietly, often secretly, opposing Law, on the one hand, as interfering with the "freedom of industry"; opposing, on the other hand, free discussion and open determination of the rules of work and wealth and wages, on the ground that harsh natural law brooks no interference by Democracy.

These things today, then, are not matters of free discussion and determination. They are strictly controlled. Who controls them? Who makes these inner, but powerful, rules? Few people know. Others assert and believe these rules are "natural"—a part of our inescapable physical environment. Some of them doubtless are; but most of them are just as clearly the dictates of self-interest laid down by the powerful private persons who today control industry. Just here it is that modern men demand that Democracy supplant skilfully [sic: skillfully] concealed, but all too evident, Monarchy.

[End of Note 7.]
The previous passages highlight how we might explore for the science/democracy relationship in Du Bois’s thinking: namely, via synonyms derived from Du Bois's own words. Ideas that could be explored via keyword in context or via proximity regexes include:

From "Of the Ruling of Men" in Darkwater, we might wish to explore:
- "freedom" and "individual freedom"
- "natural" law [with and without the double quotation marks]
- "scientific boundaries"
- "human decision"
- "air of Freedom"

From the "Evolution of the Race Problem" essay (as quoted in previous passages throughout), we could seek:
- "survival of the fittest" and "fittest" ["fittest" is found in some interesting places]
- "Darwin" and "evolutionary theory"
- "social self-realization"
- "equality of opportunity"

[Note 8–Proximity Regexes: Examples]

**Useful Proximity Regexes**

Ideas conveyed via words within Du Bois's corpus can be sought via node words or by regular expressions.

(?i)fit finds "fit", "fittest", and "unfit", but also "profit" and "benefit".

Does it change our interpretation of "to be fit for" a job, for example, if we consider that the author might be using the word in a Darwinian-inflected context (whether the text or the society)? Possible regexes include:

**Note**

Before trusting the results of more complex regexes, I typically try to determine whether some possible variants exist. I will, for example, focus on possible matches of one word in its entirety, not a word stem. Does the corpus contain "democratically" or "colour" or "world democracy" (without a hyphen)? If so, then first seeking their variants with (?i)democra will prove beneficial.

Of course, if I wish to exclude "democratically", but wish to match other word forms,
then I could fashion (?i)democra[^y]\b (to emphasize the end of the word). Yes, that regex will exclude "democratically" but might locate "democratical", if that spelling variant appeared in the texts.

The regexes that I employ herein start the matching process from the beginning of the string to its end. As regards the texts of Du Bois's corpus, the matching proceeds from top to bottom and from left to right. For example, proximity searching for (?i)democra first then (?i)scien, will not locate those places where "science" precedes "democracy". Hence, I need to reverse the word variants in order to find any reverse arrangement.

Technically speaking, regexes can backtrack and go forward within the patterns specified, but I am not using the lookaround procedures in the regexes mentioned herein. Visit the resources listed in the note on "Regexes".

[End of Note 8.]

B. Other Avenues

[44] Further avenues can be pursued with proximity regexes seeking democracy vis-à-vis science in terms of race (and its synonyms)

\{re-10\} (?i)Af[rio]{0,400}?democra|democra(?:){0,400}?Af[rio]

Referring to white workers in European colonial countries Du Bois wrote in "The Negro Mind Reaches Out":

[....] Color hate easily assumes the form of a religion and the laborer becomes the blind executive of the decrees of the masters of the white world; he votes armies and navies for "punitive" expeditions; he sends his sons as soldiers and sailors; he composes the Negro-hating mob, demands Japanese exclusion and lynches untried prisoners. What hope is there that such a mass of dimly thinking and misled men will ever demand universal democracy for all men?

The chief hope lies in the gradual but inevitable spread of the knowledge that the denial of democracy in Asia and Africa hinders its complete realization in Europe. It is this that makes the Color Problem and the Labor Problem to so [p.408] great an extent two sides of the same human tangle. [....] [NMRR 1925: pp.407-408 (=WCFA 1925: pp.442-443)]

The "denial of democracy in Asia and Africa" is conceptually related to "The African Roots of War" and its evocatively critical phrase "democratic despotism" (AROW 1915: ¶¶ 12-15). Although Du Bois did not use the word "despotism" in "The Negro Mind Reaches Out", he expresses the idea of democratically despotic actions in the paragraph preceding the search result.

[Note 9—"The Hands of Ethiopia"]

• Also read Ch. III "The Hands of Ethiopia" in Darkwater: "no industrial democracy can be built on industrial despotism, whether the two systems are in the same country or in different countries, since the world today so nearly approaches a common industrial unity."
• Or we could read "The Black Man and the Wounded World" (January 1924):

This New Imperialism has widely prevailed and its way has been cleared by a new Propaganda. This Propaganda bases itself mainly on Race and Color -- human distinctions long since discarded by Science as of little or no real significance. But this false scientific dogma which the 18th century rejected with avidity making freedom the basis of a new and world wide Humanity has been revamped by 20th century Industrialism as an Eternal Truth, so that most modern men of the masses believe the advancement of civilization necessarily involves slavery, lust and rapine in Africa. [WND1 1924: p.11]

[End of Note 9.]

[45] Another proximity-oriented regex allows us to explore Du Bois's "science" in context with "Africa", "Afro-American" (he did not seem to write "African American"), and related words. For example, in "The Afro-American" (unpublished, ca. 1895) we read an early statement of Du Bois's implicit philosophy of (social) science and its idea of the knowledge embodied in both the researcher and the researched.

Meantime one of the most important elements of the problem is without doubt, the attitude of the Afro-American himself, his opinion of his situation, his aspirations, and ideals. For it is the peculiarity of problems in social science, as distinguished from physical science, that the thing studied as well as the student, is a living breathing soul, all of whose numberless thoughts and actions must be ascertained and allowed for in the final answer.

Many are the companion ideas in that passage, and exploring them through Du Bois's corpus permits us more insight into how he approached the humanity of those being studied and the vital need to incorporate their experiences into the knowledge produced.

[46] The searching could continue and the subsequent perusal of the search results would find useful passages and a few false positives. At some point, of course, we must make sense of the results and an argument must be crafted.

V. The Synthesizing Phase of Interpretation

After gathering much information via concordancing, I am confronted with numerous possible paths by which to craft an argument. This section presents one path through the materials, a path somewhat less labyrinthine than others, by which I reassemble the components into an interpretation of Du Bois's understanding of the relationship between democracy and science.

Du Bois's views on democracy cover a range of dimensions, including the familiar normative
dimensions of equality and freedom, and the practices of disfranchisement and of voting (ATTC 1906; OROM 1920). Notably, Du Bois began to argue for an extension of democracy into the hitherto private areas of industrial decision making (OROM 1920; WND1 1924) and also for a world-wide democracy to challenge the trajectory of country-based democracy which historically had resulted in a "democratic despotism" over peoples of color, especially via colonialism (AROW 1915: ¶ 13).

[49] In a few texts Du Bois also connected democracy, in particular voting, to governance. All governments regardless of their regime type, required the "excluded wisdom" of the disfranchised to provide knowledge that governments could not about (Darkwater: Ch.VI–OROM 1920). Yet, in some of those very same texts, Du Bois also indicated that democracy will face limitations, limitations placed on some policy decisions because the scientific knowledge possessed by experts would preclude in a sense the input by the citizens (OROM 1920; NEUR 1941). Seeking some explanation of this seeming contradiction requires us to study his views on science. What was science for Du Bois?

[50] For Du Bois, science involved rigorous methods by which to explain the laws governing actions in the world and the cosmos in general, as well as rocks, trees, astronomical objects, and also humans. With regards to the studies of rocks and trees—and the economic production of goods—Du Bois's understanding of science focused on the natural laws. (As he wrote once, we cannot abolish gravity, IFRE 1949). The social sciences study human action (APLP 1940; SNP 1898; PSOP 1940). Humans can be understood in terms of primary rhythms and secondary rhythms (SOCH ca.1904). Primary rhythms are regular patterns of behavior like birth and death rates. Secondary rhythms point to the potential for human agency and free-will to alter predictable behaviors. This conceptually permits Du Bois's to avoid reducing humans into non-agents (like rocks and trees), devoid of conscious intentionality. It also allows for the potential for humans to self-develop and to achieve "social self-realization" (EORP 1909; cf. TDOP 1904). Here we glimpse how Du Bois's understanding of Darwinian evolutionary theory as processual eschews static traits; this, in turn, allows us to theorize humans developing their capacities over time (R.W. Williams 2018b).

[51] Du Bois challenged the social and natural sciences whenever they considered African Americans and other persons of color as not (or as less) capable of human agency (e.g., RTCE 1911; WND1 1924: p.113). Today we would call such studies pseudo-science. Moreover, many of Du Bois's writings over time criticized the social and natural sciences for static conceptualizations of humans, especially for not studying the social-environmental factors on the development of persons of color (HAPS 1904).

[52] In "Of the Ruling of Men" we encounter the core of Du Bois's argument that science limits democratic input to the extent that the sciences can discern more-or-less deterministic laws of nature and of the production of goods. Yet because humans can change their minds and are not robots (my word), living as they are in a world that is far from the ceteris being paribus, then the sciences themselves face limits on their scope of knowledge production (Darkwater 1920: Ch.VI–OROM).
Du Bois, ultimately, distinguished between information that the sciences could provide about how to make stuff, on the one hand, from what the sciences could not predict about human wants and needs—based as they are on irreducible human will, joy, suffering, and experience (IASC 1905). Hence, the sciences cannot tell us what to make and what people want because this is the realm of human agency (CDCP 1945: pp.83-84). Citizen participation into governance, at least via voting, allows the "excluded wisdom" of the people to help the government govern (Darkwater 1920: Ch.VI).

[53] The optimistic dimensions to Du Bois's thoughts on the interrelationship of democracy and science requires recognition of the limitations and assumptions that he is making. Several caveats follow.

• Science was assumed to be non-controversial and scientists were assumed to be in agreement.
• There is no indication of how to reconcile potential conflicts between citizens and scientists, or between citizens and government, or among politicians, bureaucrats, and scientists.
• There is no focus on manipulation of science for partisan or bureaucratic purposes.
• Scientists in general are assumed to be pursuing the public good (which for Du Bois would include at the very least racial and social justice, education, and an end to poverty and war).
• There is no detailed examination of how agreement among citizens on public policies might be reached.

VI. In Closing

[54] Political theorists can use concordancing for their research as a means to supplement the common technique of close reading and contextualization (both of which are still necessary). Although corpus linguists typically concordance as a way to seek generalizations and explanatory laws of language use, concordancing itself can be oriented to interpretive-style research, research that seeks meanings conveyed via ideas in the collected texts of a corpus.

[55] Interpretive concordancing is a means to assist close reading because it facilitates

• access to large numbers of documents (creating an assemblage of texts composed by one or more authors);
• locating ideas and themes (as represented by the words in the digitized texts); and
• interconnecting various texts in terms of similar ideas (whether with reference to the same words or to textually derived synonyms).

[56] Several implications follow from interpretively concordancing Du Bois's thinking on democracy vis-à-vis science.

• The emphasis is on the importance of the ideas themselves, regardless of their frequencies, as the means by which Du Bois made—and we make—sense of the world.
• Du Bois's agency and his understanding of social problems and their potential solutions are foregrounded via the intertextuality of his ideas, both the initial ones of democracy and science, as well as the companion ideas. Such intertextuality points us to Du Bois's various definitions and applications and accordingly, emphasizes that our interpretations must take into account the
incompleteness of any one expression of an idea. Hence, finding and interconnecting Du Bois's ideas in various texts does not imply any necessary consistency in them across time and texts. We can note, however, any similarities and differences (R.W. Williams 2016).

- The intertextuality of Du Bois's ideas also points us to his interdisciplinarity, by which is highlighted how he worked within and contributed to the social sciences, and utilized the insights of some of the natural sciences, especially Darwinian evolutionary theory (R.W. Williams 2018).

[57] Emerging here are three intertwined caveats to my project and its interpretive concordancing. The first caveat relates to the texts that comprise the Du Bois corpus that I am studying, while the second focuses on how the corpus is created for the purposes of interpretive concordancing. The third caveat pertains to the use of regular expressions (cf. Atwood 2008). These caveats are specific to my project and extend the more general caveats on interpretive concordancing that are mentioned above.

[58] Regarding the first caveat, the interpretations discussed herein, although textually based, focus on certain strands of Du Bois's thinking because they are the ones locatable in the corpus that is available. The interpretations cannot be comprehensive in the sense of a full understanding of Du Bois because the corpus itself is not comprehensive due to its incompleteness. The partial corpus I created and the standard canon of texts read by Du Bois scholars do not exhaust what is available in published and unpublished forms. In addition, new texts by Du Bois might be, and indeed have been, recovered (e.g., R.W. Williams 2012 & 2014).

[59] The second caveat to my project centers on the digital corpus that is at the heart of the research process. I made choices about what to include in the digitized texts that I rendered for concordancing. Would I include photographs, tables, figures, drawings, and images? I excluded those items but retained their captions or table descriptors. I even added an editorial comment in cases where no caption existed. Would I include marginalia, and corrections of various types, among other considerations? Yes, and I devised an editorial system to so designate those additions by Du Bois in his texts, especially the unpublished ones.

[60] The third caveat to my project focuses on the computer-mediated qualities of the analysis itself. Algorithmic approaches to texts, which includes concordancing, influence what we actually locate in the texts of the corpus. Will the regular expression searching miss a match due to the parameters established when I construct the regex itself? Consequently, testing the regexes becomes very important. For more commentary, visit Note 8 on "Proximity Regexes: Examples".

[61] As a consequence of these specific caveats, my project is a starting point, and attempts to provide some insights as part of an intermediary stage in studying Du Bois's ideas and texts. In no way is my project the final word on any topic discussed herein.

[62] In the spirit of the afterwords that he penned, I end with Du Bois's closing to his 1906 "Address to the Country" (ATTC)—words still resonating 115 years later.
"The morning breaks over blood-stained hills.
We must not falter, we may not shrink.
Above are the everlasting stars."

VII. Bibliography

A. Works Created or Edited by W.E.B. Du Bois

URL: http://credo.library.umass.edu/view/full/mums312-b211-i124.


URL: www.webdubois.org/dbAtlantaConfs.html.


AUPs. 1896-1916. Atlanta University Publications, various years. [Links to the AUPs are available at www.webdubois.org/wdb-AtlUniv.html].


URL: www.webdubois.org/wdb-darkwater.html.


OROM. 1920. "Of the Ruling of Men." Ch. VI in Du Bois, Darkwater (DARK 1920).


SOCH. Ca. 1904-1905. "Sociology Hesitant." Credo Repository. URL: http://credo.library.umass.edu/view/full/mums312-b212-i003 [Metadata indicates "ca. 1905" as the possible date of creation].

A. Works Created or Edited by Others


Borges, Jorge Luis. "Kafka and His Precursors". *Labyrinths*.


London: Routledge.


