

## Writing to be Found and Writing Readers

John Cayley <John\_Cayley\_at\_brown\_dot\_edu>, Brown University

### Abstract

Poetic writing for programmable and network media seems to have been captivated by the affordances of new media and questions of whether or not and if so, *how* certain novel, media-constituted properties and methods of literary objects require us to reassess and reconfigure the literary itself. What if we shift our attention decidedly to practices, processes, procedures — towards *ways of writing* and *ways of reading* rather than dwelling on either textual artifacts themselves (even time-based literary objects) or the concepts underpinning objects-as-artifact? What else can we do, given that we must now write on, for, and with the net which is itself no object but a seething mass of manifold processes?

Part one of the essay presents a brief analysis of recent experiments in “writing to be found” with Google, making some claim that such writing may be exemplary, that its aesthetic and conceptual engagements are distinct, and that there is something at stake here for “the literary” or rather for certain practices of literary art. After very brief discussion in part two of some broader implications of writing with the Google corpus and its tools, part three addresses more examples of writing to be found, and introduces a collaboration with Daniel Howe, *The Readers Project*, many processes of which engage with “writing to be found” “in” Google and making use of its tools.

### One

In order to begin to write this essay, I set out to make some appropriate use of what I have come to think of as “writing to be found”. Originally I had thought that this would be by way of simply beginning to write, embarking on my usual process of writing while checking, periodically, to see whether the sequences of words that I was in the midst of composing were still “found” in the corpus and then at what point they became “not yet found”.<sup>[1]</sup> How many words would I have to add, composing my syntagmatic sequences, before they were not found in the corpus of language to which the Google search engine gives me access, before they were, perhaps, original sequences? How difficult would I find it to produce unfound sequences? Would I be able to continue to write as I usually write once I was aware that, at some perhaps unanticipated moment, the words I write are suddenly penetrating and constituting the domain of sequences that are not yet found in our largest, most accessible corpus of written English?

1

There have proven to be many questions raised by any and all of my attempts to engage with these processes and their contexts. Moreover, I remain convinced that many of these processes may be productive of significance and affect, to an extent that will allow aesthetic, not only critical, practices some purchase.

2

This way of working with language is enabled by unprecedented, convenient and articulable access to the network, a world of language, a media-constituted diegesis, that is still “powered” – as the contemporary technologically inflected usage would have it – by text, by encoded representations of inscription, in what we usually call writing. The net is still largely composed from all the privileged instantiations of our languages’ singular materialities that we, as irrepressible language-makers, have so far written to be found.

3

By which I mean to make it clearer, that when I write with these processes, I’m both writing, and writing with Google.<sup>[2]</sup> Is Google my collaborator? Does Google become the space within which I write? I want to make it clear that I don’t

4

consider myself necessarily to be writing in the space of the network nor collaborating (directly) with other artists. At this point, I also want to make it clear that I do not consider myself to be using Google, not, at least, in the usual way that Google is used for gathering instances of language by search. I'm not refashioning myself as a Flarfist.<sup>[3]</sup> I'm not casting a faux-*puerile*, post-everything, absurdist net over the net using the net, gathering glittering detritus, spectacular disjuncture, in endless anti-syntactic listlings. I'm not composing searches in order to find the language for what I'm making. I've got my language already, one way or another. I just want to know whether it's found or it isn't. The Flarf-poetic approach is – although this is only a small part of Flarf – a *détournement* of the affordances that Google offers us as a portal to text on the network. My “writing to be found”, on the other hand, is in itself a way of writing that is shaped by the way that Google is shaped, by the way in which Google curves the space of the network. And Google does also, in a sense, write with me: constraining, directing, guiding, and, especially, punctuating my writing.

It occurs to me, broadening the scope of these experiments' relevance, that Poetic writing for programmable and network media seems to have been captivated by the affordances of new media and questions of whether or not and if so, how certain novel, advanced, media-constituted properties and methods of literary objects require us to reassess and reconfigure the literary itself. What if we shift our attention decidedly to practices, processes, procedures – towards ways of writing and ways of reading rather than dwelling on either textual artifacts themselves (even when considered as time-based literary objects) or the concepts underpinning such objects as artifacts? What else can we do, given that we must now write on, for, and with the net which is itself no object but a seething mass of manifold processes? Google itself signals the significance of process since Google both is and is not the net. Google is not the inscription that forms the matter of the net. Google is merely (almost) everyprocess (not *everything*) that makes it possible for us to find and touch and consume what was always already there in front of us.

When you collaborate you are more or less obliged to get to know your collaborator. Getting to know Google better, in a practical sense, as a collaborator, is one of the most interesting results to emerge from even the relatively simple and preliminary processes that have been set in train.

This is probably the moment to introduce some details of the procedures with which I am writing. First, a classical epithet via Montaigne in John Florio's translation, “The Philosopher Chrisippus was wont to foist-in amongst his bookes, not only whole sentences and other long-long discourses, but whole bookes of other Authors, as in one, he brought Euripides his Medea. And Apollodorus was wont to say of him, that if one should draw from out his bookes what he had stoln from others, his paper would remaine blanke. Whereas Epicurus cleane contrarie to him in three hundred volumes he left behind him, had not made use of one allegation.” [Montaigne 1910]

Process: Write into the Google search field with text delimited by quote marks until the sequence of words is not found. Record this sequence. Delete words from the beginning of the sequence until the sequence is found. Then add more words to the end of the sequence until it is not found. Repeat. Each line of the resultant text (although not necessarily the last line) will comprise a sequence of words that is “not yet found”. At the time of composition these lined sequences of words had not yet been indexed by Google and were thus, in a certain (formal) sense, original:

“If I write, quoting,”

“I write, quoting, “And”

“write, quoting, “And the”

“quoting, ‘And the earth”

““And the earth was without form and void; and darkness was upon the face of the deep,’ these words”

“upon the face of the deep,’ these words will”

“deep,’ these words will be found”

“these words will be found. Perhaps”

“will be found. Perhaps they will now”

“Perhaps they will now always be found”

“will now always be found. I”

5

6

7

8

“always be found. I write”

“be found. I write, in part”

“I write, in part, in the hope that what”

“in the hope that what I write will be found.”

with Google, Sat Oct 3, 2009, completed 2:04am EST.

I was induced to explore this way of writing by the remarks of a philosopher and cognitive scientist, Ron Chrisley, at a workshop on Neuroesthetics.<sup>[4]</sup> In discussing robotic perception, he was making some use of the concept of the “edge of chaos”. I understood this phrase loosely as referring to a threshold of information processing, the point at which an artificial cognizer can no longer assimilate – typically by compression or by rule formulation – the information that comprises its inputs. Somehow, to me, this suggested or rhymed with that moment in our now common encounters with search engines when what we are looking for is not yet found, when it could still be anything, because, as yet, it is nothing to the corpus. It isn’t there. It isn’t in any way predictable. It’s still maximal, raw information in Shannon-Weaver’s sense – the edge of chaos that we are about to make, literally, readable.

9

Since I have some practical experience with Markov models for text generation, I also pretend to recognize this as a closely related phenomenon.<sup>[5]</sup> If we think of Google as giving us access to a vast Markov model, I believe I am right in saying that as I build up my sequences of words delimited by quotes and test them after adding each word, I am testing the model’s ability to be able to find me an n-gram where n is equal to the number of words in my sequence. Non-zero results mean that there are probabilities to play with. Not only is it the case that other people before me have produced instances of this sequence of words, but an n-gram model, constructed from the Google corpus, would also have some chance of generating my search phrase. However, once I’ve reached an unfound sequence, the model breaks down. I’m at the edge, and I may also, perhaps, be about to extend, by some minuscule amount, the readable, the unchaotic territory of the textual, perhaps even that of the literary. I’m about to write, and to add my own writing to the corpus.

10

And then suddenly it gets interesting. I was just writing, and now I’m writing with Google and beginning to wonder what that means. Google is where we search for language and for forms of all kind that are made from language, including aesthetic forms. It’s become our default portal to the default corpus. It is not yet all writing but we feel that we are close to the historical moment when the extraordinary possibility – Ted Nelson’s Docuverse – has become an actuality for, at least, a major portion of the existing textual corpus of writing in English. Already, I wager, we type our searches into Google expecting that it will find anything and everything that we might expect to be found in the world of letters, of conventionally inscribed textuality. What do I mean by that? I mean at least all of those sequences of words that have been written by authors who are known to us. All of the writing that is known, all of the writing that will have been found. And much besides.

11

“The purpose of this writing is to address”

“an edge of chaos.”

“Specifically, the point or points”

“in sequences of words that”

“delimit phrases”

“found to be unique in our”

“most accessible corpus.”

with Google, Sat Oct 3, 2009, completed 10:27am EST.

The two singularly lineated sentences above are made with a slightly different process, a retreat from the not yet found sequence – at the time this was, for example, “The purpose of this writing is to address an” – to the longest sequence that was still found in the accessible Google corpus. Although the sentences are original to me they are expressed in phrases that can be shown to be plagiarized from the corpus. They have all already been written.

12

For we do seem to be addressing something like the palpable, objective edge of authorial originality. “The purpose of

13

this writing is to address” was always unoriginal before I set out. When I wrote, “The purpose of this writing is to address an”, the indefinite article made me an author.

Those of us who are educators will be aware of the way that Google and other search engines are used as simple detectors of student plagiarism. Type the suspected sentence into Google and it is very likely to find the source from which it may have been copied. Writing to be found with Google reveals, however, the singular, perhaps unprecedented nature of its, Google’s, co-authorial authority. By definition Google changes shape. As we’ve said before, it’s a process. By providing access Google seems to *be* the corpus of reference while remaining a protean manifold of processes that continually reconfigure themselves while crawling over *our* networked body of language (the actual corpus), even unto the edge of chaos, finding new readable things and indexing them relentlessly and swiftly, remarkably swiftly. Less than three hours after I’d posted my not-yet-found texts to the netpoetics blog, they were suddenly found.<sup>[6]</sup> Thus, taking the same text and putting *it* through the same procedure produced an entirely different text and a new measure (or textual visualization) of my originality.

14

Returning to my first process, with the supply text just quoted, for example:

15

“The purpose of this writing is to address an”

“is to address an edge of”

“address an edge of chaos.”

completed with Google at 9:17 EST on Oct 1, 2009, became:

“The purpose of this writing is to address an edge”

“is to address an edge of chaos.”

a little over two hours later at 11:30 on the same day. (By the way, although the second iteration of the process reduces the number of unfound sequences in this initial extract; for the entire supply text, the second iteration actually increased the total number of unfound sequences from 17 to 21.)

This potential for iteration was not only expected but it was something with which I desired to experiment, using it to produce a series of texts, evolving over time in relation to the findableness of their constituent sequences of words.

16

But imagine my surprise when I tried the procedure again and found it regenerating the earlier version. My new, original writing was no longer found. I could see it there in the corpus (at netpoetics) but as far as Google, the “index of reference”, was concerned it was, apparently, no longer there. I could not yet have produced it. Uncanny. But easily explained by my arbitrary access, at the first instance of checking, to Google servers that had already published the indexing of their busy spiders. Later, I had been less lucky: my client must have connected to other servers (I have no obvious control over this) onto which the new indexes had not yet propagated. Google had temporally denied my originality, my authority. It had changed the shape of my authorial persona. I wasn’t writing with it. It was writing with me, against me, withholding what I thought I had inscribed.

17

## Two

Why hadn’t I considered this before? Why don’t we think of it now, and then more often? As a culture, we are in the seemingly ineluctable process of handing over the digitization and indexing of our entire surviving published textual legacy to Google, in order for them to include that part of it which they have not already indexed. I, we, have no idea how they are going to index our literature or how their indexing of it might change over time. On the other hand there is considerable evidence of uncertainty and inconsistency.<sup>[7]</sup>

18

I should of course mention in passing that there are already and will likely remain some checks and balances to Google. So far, the other internet search engines have access to most of the same corpus, and they do not index this corpus in the same way.<sup>[8]</sup> Without huge investment we could all write and set up our very own search engines. Nonetheless it is remarkable the degree to which Google has become, as I say, initially the search engine of reference and now in some

19

sense the reference of reference. This is so obvious to us that it has become banal to point out that whatever Google is, it may be the most remarkable and significant agency for cultural change on the planet.

Of course, the scholars amongst us (and within us) will defer. We cannot rely on anything that the folksonomic internet provides, although relying, admittedly “by default”, is exactly what all of us having access actually do. Neither can we defer from Google in the same way that we defer from Wikipedia, on the basis of what it “contains”. Google is not Wikipedia and, in a sense, it does not contain anything. Practically and in other critical senses, it stands between us and Wikipedia while also providing – in so far as it indexes all the writing that can be found – much of the material from which Wikipedia is built. Wikipedia is something that arose contemporaneously with the Googlization of everything but is more a symptom than a cause. Whatever Google is, is a problem that remains to be addressed, and written with.

20

Here is one brief working statement of what Google is becoming or what it may already be: *Google is the preferred or default agency to which our existing institutions of cultural production and critique delegate the symbolic processing of our inscribed material culture in exchange for unprecedented access to the results of that symbolic processing.* I am, of course, bracketing all the important questions concerning what exactly is handed over to Google for processing, how is this done, who owns it, and where it is – all of which are irreversibly complicated by the fact that any answers will be radically different “before” and “after” these processes that were already in train long “before” any actual exchanges – such as agreements to digitize libraries – were made explicit, let alone regulated in any publicly agreed and articulated manner.

21

Let’s say it again in more polemical terms. We hand over our culture to Google in exchange for unprecedented and free access to that culture. We do this all but unconscious of the fact that it will be Google that defines what “unprecedented” and “free” ultimately imply.<sup>[9]</sup> As yet, we hardly seem to acknowledge the fact that this agreement means that it is Google that reflects our culture back to us. They design the mirror, the device, the dispositive, as the French would put it. They offer a promise of “free” access in many senses of that word including zero cost to the end-using inquirer and close to zero cost to the institutions that supply the inscribed material culture that Google swallows and digests. But Google does not (some might here add “any longer”) conceal the fact that this free access does come at a cost, another type of cost, one that is also a culture-(in)forming cost: Google will process all (or nearly all) this data in order to sell a “highly-cultivated” positioning of advertisements. The deal can’t go ahead without this underlying engine of commerce and commercialization. In a sense, Google is the predominant global corporation a major proportion of whose capital is literally cultural capital. Now, what was already a huge backing investment is being freely augmented by the traditional investors in this market of culture, the universities in particular. Bizarrely, these institutional investors are not asking for shares in the business, or rights to vote on the board. All they seem to want is to have what they *already* had, but processed, indexed, reformed and reflected back to them, to us, in, as I say, a manner that allows many of us unprecedented access.

22

This is not, primarily, an essay about Google, and the situation was and is far more complicated than this polemical outline suggests. Google did, after all, emerge from the *popular* culture that was born on the internet itself, long (in net history terms) before institutions began to contribute to this culture to any significant extent. Thus the initial cultural capital that Google amassed may be seen as fairly won, and the access that Google provided to a suddenly vast, ever-accumulating resource was truly unprecedented, rendering the culture of the net useable, manageable, findable, beyond all expectation.<sup>[10]</sup> We learned quickly that “unprecedented access” meant that Google was better than any other agency at managing the “more than ever before” of everything that is digitally inscribed, the exponential increase in information. But now this simple, if overwhelming, quantitative fact is *all* that we and our institutions know with any surety. We know that Google will deal with the scale of it all, and manage it all better, and give more of it back to us,<sup>[11]</sup> but we may never know, unless we ask or demand, exactly *how* they do this or how they *will or will not* do this in some speculative future when they have already disposed of the problems of processing it all, displacing it all, continually rendering it back to us through manifold devices with post-human artificial intelligences.<sup>[12]</sup>

23

## Three

So now all my writing to be found has been recast in the light of this shared, would-be universal engagement or struggle

with Google to retrieve or reform culture. And immediately, as in the work of writing digital media that underlies these remarks, I return to specifics with a heightened awareness of their potential significance, especially as critique of these relations.

24

For example, in the course of investigating writing to be found, it occurred to me that any material that is quoted in a text from a well-known, and therefore much indexed, source will emerge very differently in the procedures outlined above. It seems that in what may be standard original composition, you can expect sequences of words that you are writing to be found to be unique after about five words, depending on diction. However, arbitrarily long sequences of words recalled or quoted from many texts, like the English Bible in one of the standard translations, will already and will always be found ... by Google. The conceptualist in you might want to test this to some absurd aesthetic extreme, typing all of Genesis into the Google search box delimited by quotes and discovering thousands of hits. I didn't get this far although I made attempts with lengthy sequences until I noticed, in light grey type, the legend:<sup>[13]</sup>

25

“what” (and subsequent words) was ignored because we limit queries to 32 words.

I hadn't noticed or been aware of this limitation before. And I am still unsure about when and how it was instituted. How long had this been a Google limitation? Who decided it was needed and why? Why 32 words? It's clearly not surprising that this limitation exists. The point here is that it gets in the way of using or, in my case, writing *with* Google in the way I believed would be interesting and might lead to further aesthetic or critical cultural production. What if I wanted to continue with what I had hoped and planned to do? Google's got indexes to my language, my culture. Even if they might not reasonably be expected to give me all the tools I might need or want to explore this material, why should they constrain or reform the tools that they do appear to give me in ways that seem to me to be arbitrary or, at least, unrelated to my own concerns? These questions are already important but not as important as they will become. When Google indexes all books, which institutions will keep track of when and why they change their search algorithms, let alone endeavor to influence Google's decisions in such matters?<sup>[14]</sup>

Never mind, for my immediate purposes at least. Conceptually, I can imagine what the search results would have been for absurdly long sequences from famous texts and how, using writing to be found procedures for lineation, texts that quoted or plagiarized such material (let's say, writing to be found punctuating certain texts of Kathy Acker or Pierre Menard's *Quixote* or Kent Johnson's *Day*),<sup>[15]</sup> would be chopped up where they are “original” and then bulge out where they incorporated what is already found, as the “If I write, quoting ...” example above demonstrates. (Menard's *Quixote* would be all “bulge”.)

26

I say “never mind”, but remain disturbed. A productive engagement had been interrupted by a (ro)bot from Porlock and now this seems as if it will be characteristic of writing and working *with* Google, re-energizing the Anglo-Saxon origins of that preposition. In fact, of course, it is a function of encoded properties and methods that are designed to reassert, where and whenever necessary, the underlying purposes of the Google engine which is, as we recall, to dispose of culture and propose advertisements based on this disposal. Google asserts: “You don't need more than 32 words in your queries in order to determine what you want and what interests you. Making something that requires longer searches will simply skew our data and make it harder for us to know what you want.”

27

Despite Google's assertion, I keep searching. Now my collaborator, Daniel C. Howe, and I keep searching. We've already, like many others, come up against another important limit. If you search too much or too fast (even manually I found), then Google's engine thinks you might be a process (as *it* is) and that you might be making automated queries. This produces the same threat to Google's underlying purpose, the threat of skewed analytical data. However, to us it seems as if we are simply retrieving access to our own linguistic culture. Usually, we are simply mining the corpus that Google makes accessible – in an unprecedented manner – for “natural language data”. In writing to be found, I seek out the chaotic edge of what is being written and is soon to be found by myself and others, the edge of what literary culture acknowledges to be attributable authorship.<sup>[16]</sup> Isn't this a legitimate engagement with what Google promises us? Shouldn't these admittedly or purportedly poetic queries be accepted as a part of the culture with which they also engage?

28

As a matter of fact we continue to write programs that generate automated queries and it is strange that Google – itself a vast conglomeration of processes – rejects them as such. Shouldn't Google be prepared to pass judgment as to whether a process is an innocent cultural address to its services rather than assume that any automated inquiry is an attempt to undermine or deflect it from its prime, commercial objective?<sup>[17]</sup> Returning to a concrete example that engages related concerns with poetics and the author function, I realized that using the Google search query's *not* prefix (a minus sign) I might search for sequences of words from well-known texts (delimited by quote marks) that would be found in the corpus but in places where they were not associated with their well-known "authors". I used this negatively qualified version of the procedure described above, testing successively longer sequences and aiming to find the longest sequences that also satisfied the essential condition of *not* being attributed to the famous author. This produces a text that, paradoxically, is collaged from phrases that are quoted from arbitrary internet unknowns but which, when linked together, will compose a famous text. Before supplying an actual example, I want simply to point out that the program I write to undertake this entirely legitimate essay in conceptual poetics generates a large number of test searches even for a brief text and it will find itself frequently blocked by Google's suspicion of and ultimate denial of my own process's high cultural intentions.<sup>[18]</sup>

" blue and white of sky " " a moment still " " April morning in the " " mud it's over " " it's done I've had the " " image the scene is " " empty a few " " animals still then " " goes out no more " " blue I stay " " there way off on " " the right in the mud " " the hand opens and closes " " that helps me it's " " going let it go I " " realize I'm still smiling " " there's no sense in that now " " been none for a long time now " " my tongue comes out " " again lolls " " in the mud i stay " " there no more " " thirst the tongue " " goes in the mouth " " closes it must be a " " straight line now it's " " over it's done I've had " " the image "

This is Beckett, three fragments from *How It Is* which also correspond to the final part of a short prose work he originally published in French as *L'Image*. But it is also possible to assert that is *not* Beckett but rather something that I have written together with Google, where we have conspired to calculate a maximal syntagmatic association with Beckett's texts while ensuring that these sequences are attributable to others, often many others, and we do this in a manner that can be established by a contemporary form of citation. It is a relatively nice problem to consider whether this text infringes copyright. I might claim, for example, that it is not copied, that it's not even the same text, especially given that I have transcribed it with quotation marks around the phrases. A copyright expert might assert that it was created by a mechanical process, that it is the product of a procedural but regular form of transcription and is, therefore, a copy, to which I would have to reply that a great deal of personal thought and significant indeterminate and unmediated human labor also went into its making. The piece certainly challenges the Beckett estate's moral rights in respect of the text's integrity and its association with the author's name. In US law these rights are not established. In any case, I may both justly claim fair use, and also perversely propose that my first-cited example was actually derived from the following entirely original collage composed from fragments found to have been written on the internet:<sup>[19]</sup>

"a moment *still* " " *animals* still then " " *April morning in the* " " *blue and white of sky* " "been none for a long time now" " *blue* I stay " "empty a few " " *there no more* " "mud *it's over* " "my tongue comes out" " *thirst* the tongue" "goes out no more" " *goes in the mouth* " " *again* lolls" " *closes* it must be a " "straight line now *it's* " " *the hand opens and closes* " "that helps me *it's* " " *going let it go* I" "realize I'm still smiling" " *in the mud* i stay" " *it's done I've had the* " " *image* the scene is" "there way off on " " *the right* in the mud" " *over* it's done I've had" "the image" " *there's no sense in that now* "

Clearly a lot more could and will be done with the procedures of writing to be found including with this latter variation in which one rediscovers how much of what has been written has already been written. Google makes all of this possible and Google also stands in the way of these unanticipated essays. One very significant reason to continue to work in this way is precisely to reveal how Google and other similar agencies will reform what they pretend to enable, and how our existing institutions that support writing as a cultural practice will relate to the profound reformations that must ensue.

## Four

The "writing readers" within a major collaborative project in digitally mediated literary art are underpinned by the critical,

contemporary, quietly hacktivist natural language processing and research initiated in “writing to be found”. *The Readers Project* incorporates “writing with Google”, and it also proposes performative reading as, perhaps, exemplary of how we may write in this, our future. The collaboration, with Daniel C. Howe, produces literary objects that have an extensive computational dimension and will, typically, be realized as screen-based or projected works, for both private viewing and reading, and more public exposure in installations with distributed multi-media and/or mobile displays. As such, they are, in the relatively small world of writing digital media, examples of a variety of work whose real-world instantiations take some place either in the screen real estate of net-based or personal computer-based art, or in the mediated gallery space of digital art. Even the computational aspects of this work have become amenable to critical attention in these days of codework, expressive processing and/or critical code studies.

However – and this may not be the best news for an already over-extended critical community examining aesthetic objects that have still to prove themselves in any wider cultural forum – crucial reading strategies that are already encapsulated in our projects, in our quasi-autonomous readers, are derived from precisely the kind of “writing with Google” that I have outlined above. In other words, one of the more interesting dimensions of these readers is that they are, in significant measure, the result of natural language research and processing undertaken in, arguably, a socio-politically implicated dialogue with our predominant new devices of cultural reflection and disposition. Of course, the readers also have other inclinations and ambitions (apart from any jostling entry into the world of digital art). They may simply wish to offer themselves to open-minded literary critical readings such as are often applied to the literary avant-garde. You can read them as poetry or as a poetics. What I am suggesting, however, is that they may also be read for the way that both they and their making reads and writes with newly mediated culture, with Google in this instance.

This is a final point, a vector for both literary poesis in digital media and for its critical reception, but I must conclude the point with its illustration. Here are three readers from the project, moving through and “reading”, in some sense, an underlying text, a prose poem of my own, “Misspelt Landings”.<sup>[20]</sup> There is a mesostic reader that finds and highlights words containing letters (which it capitalizes as it finds them) in a phrase beginning “READING THROUGH ...”, and there are two other readers: one that tends rightwards and downwards in the conventional vectors of human reading while deviating occasionally, and one that seems to wander while surrounding itself with a halo of erased or faded text. What is far from obvious is that these readers, all of them, chose their next word to read (and hence their deviations) on the basis of simple but quite effective research on the usage of these words in the corpus to which Google gives us access, however reluctantly. An important aspect of the way this and other pieces from *The Readers Project* are deployed is that, for each such manifold display, the readings of all the live readers are separately broadcast to a server, a feed to which you may subscribe by accessing a URL with a browser and with other clients under development. Subscribed to a particular reader, you may read along with it and see clearly the textual path it has chosen, according to its particular reading strategies.

In simple terms these readers check the proximate neighboring words of the word they have just read and they “know” – from the results of their writers’ struggle with Google – whether or not any or all of those proximate words will represent likely natural language phrases.<sup>[21]</sup> Daniel C. Howe and I are the writers of these readers and we, along with other coded processes, struggled with Google, sending queries to its “books” domain to see how many instances of thousands of three-word phrases had already been inscribed as writing to be found and how frequently they had been inscribed in the net’s textual corpus, if at all.

Many of you reading this will understand that this is far from being an entirely novel approach. However, although our readers may seem to be following a simple Markov chain, the actual processes and models deployed in *The Readers Project* conceal some significant differences to a standard Markov model.<sup>[22]</sup> More importantly and finally, these readers were written with processes that hacked near-live statistical data out of the Google-indexed internet corpus of all the inscribed cultural material that can be found. Writers of readers like these could not have made anything approaching their capabilities until very recently, or not without huge, institutionally-maintained resources. We were and are able to make these readers remarkably up-to-the-minute in their model-driven analyses of the texts that they were written to read. They know what they need to know about the latest writing to be found on the net in their domain. This knowledge was mined iteratively from the language that we all gave over and continue to give over to Google and, in so far as

34

35

36

37



Google was uninterested in or threatened by the queries we needed to make in order to gather our readers' simple knowledge, that knowledge is the result of a fascinating struggle that – for this reader at least – is a model in micro-procedure of the struggles that we must all undertake as our institutions of culture pass over their care and disposition to all those strange engines of inquiry that may suddenly reject our search for writing. They reject our queries for reasons that we may not entirely comprehend. Not yet and perhaps, not ever.

## Acknowledgements

Based on and extended from a presentation entitled “Edges of Chaos: Writing to be Found” for a workshop at the University of Bergen, Norway, November 8-10, 2009, this essay resulted from a keynote paper at the *Futures of Digital Studies* conference, University of Florida, February 25-27, 2010.

38

## Notes

[1] Throughout this essay I refer to the Google “corpus”, implicitly treating the inscribed text that is addressed by the Google indexing engines as if it were a body of material similar to or commensurate with other textual corpora such as might be compiled into a particular author’s corpus or the corpora put together and studied by corpus linguists such as the Brown Corpus, the Corpus of Contemporary American English, the British National Corpus, the American National Corpus, etc.

[2] Whenever I use the word “with” in this context, my intention is to highlight the underlying, now chiefly archaic, sense of “against” that was once more active in the Anglo-Saxon preposition, although we do still both work and fight *with* others. This negative apotropaic inclination of “with” is preserved by contemporary English in words like “withhold”, “withdraw”, “withstand”.

[3] *Flarf*, the coinage attributed to Gary Sullivan, is a name for a practice of poetic writing. There exists a “Flarf(ist) Collective” of writers, mostly poets, who have exchanged and published work under its aegis. (See the Flarf feature in the excellent online *Jacket Magazine*, Jacket 30, July 2006) Wikipedia describes its aesthetic as “dedicated to the exploration of ‘the inappropriate’ ” (as of: Feb 16, 2011) and this seems right to me. It’s a significant poetic movement of the late 20th, early 21st century for which, personally and critically, I have a high regard. However, Flarf is now also closely associated with methods of composition that make extensive use of internet searches engines since they are, clearly, well-adapted for gathering large amounts of “inappropriate” linguistic material. The association is unfortunate since there are many, many other ways to explore the inappropriate and gather relevant exempla. The identification of Flarf with Google-mining is, itself, inappropriate Flarf. At this point in my argument, my aim is simply to contrast the Flarfist use of Google-as-grab-bag versus a sustained aesthetic engagement with the cultural vectors that Google both offers and denies. Engagement at the level of computation may be a key to making and maintaining this distinction.

[4] European Science Foundation (ESF) workshop: *Neuroesthetics: When Art and the Brain Collide*, Sept 24-25, 2009, IULM, Milan, Italy. <http://www.esf.org/activities/exploratory-workshops/humanities-sch/workshops-detail.html>

[5] Markov models, processes, chains – named for the Russian mathematician, Andrey Markov (1856-1922) – provide formal descriptions for systems with a finite number of elements in successive states. Using such a model, we only have to know the relative frequency of the elements in a system in order to be able to generate further sequences of these elements, probabilistically, that will be, as it were, characteristic of the system. These models can be applied to language, taking any distinct linguistic element – letter, phoneme, syllable, word, phrase, etc. – as the units being considered. A sequence of n elements considered as a unit is known as an n-gram or n-gram. A three-word phrase may be treated as an n-gram and if we search for such a phrase, double-quoted, in Google, we get a “count” that can be used as a relative frequency for that phrase within the domain of the Google-indexed internet “corpus” of linguistic tokens. Refinements of such purely statistical language models are now proven to be remarkably powerful, and underlie, for example, much automated translation. The existence of the Internet-as-corpus and its Google search boxes puts such linguistic modeling in the hands of everyone. Google and its rival service providers are aware of non-venal uses for this data. Recently there was a short, rather dismissive piece on the Google labs: Books Ngram Viewer in the *London Review of Books* [Diski 2011]. A *Science* article is referred to that describes work underlying the Ngram viewer in more detail [Michel 2011]. Another contextually relevant discussion of Markov chains can be found in Noah Wardrip-Fruin, *Expressive Processing: Digital Fictions, Computer Games, and Software Studies* (Cambridge: MIT Press, 2009) 203-05 [Wardrip-Fruin 2009, 203–5].

[6] See <http://netpoetic.com/2009/10/an-edge-of-chaos>.

[7] Apart from specifics discussed here, I will cite two sensational socio-political examples. Firstly, there is Google’s dubiously or *unprincipled* accommodation of Chinese state censorship as a Chinese language news provider in February 2004, as an investor in the Chinese search site Baidu, by voluntarily blocking politically sensitive searches in January 2006, and its subsequent *purportedly* principled retreat from the Chinese

search “market” in 2010. See [Battelle 2006]; [Watts 2010]; [Bosker 2010b].

[8] For one simple example, Microsoft’s *Bing* treats line endings differently. Line endings (carriage returns, etc.) don’t break sequences as they do for Google. For neither engine however, is this a recognition of differences or distinctions that might be significant for poetics. The fact that we can be fairly certain that differential treatment of line endings is technical *in the service of commerce* rather than poetic or, for example, rhetorical, speaks volumes concerning Google as an engine of mis- or undirected culture formation. Its undoubtedly “powerful” forces are self-trammelled by concerns to which Google is strategically blind and to which we, as producers of culture with other motivations, seem already to have become blind. If we fail to start noticing these motivated distinctions now, it will soon be too late since they will cease to exist. In the ontology of software if a object is not implemented, it cannot have instances.

A further note on line endings: It is interesting to remark that although line endings break word (or token) sequences in Google’s indexing of web pages – chiefly html or html-derived content – token sequences are *not* broken by corresponding punctuation or tagging when Google indexes the predominantly pdf-derived content of Google Books. This is simply one example of many conditions demonstrating that when you search these two domains, you search them differently with no explicit signal of this fact. The underlying software is taking away any care that you might have had for the way in which you are searching. If your relationship to the corpus is transactional and you understand the nature of the underlying contract, this is fine. My point is that now, when you search Google, you increasingly treat it as if you are searching all of inscribed culture. Once again, this is fine, if you realize what you are doing – research that is abbreviated, shorthand, provisional, or pragmatic for example – and yet after having qualified your understanding of the scope of the Google corpus, do you also take responsibility for your failure to know any details of the procedures by which it undertakes the search on your behalf, *how* that search addresses the corpus, the manner in which the results are delivered, and so on?

[9] And ultimately or more accurately: whoever or whatever *owns* Google.

[10] See note 16 below. Daniel C. Howe adds, “Of course Google automatically/procedurally indexes our pages/content, yet makes it illegal or at least, they would claim, a violation of their terms of service for us to do the same to them.”

[11] Daniel C. Howe adds, “in tiny droplets”, that are regulated by: Google.

[12] The fact that we accept – pragmatically, gratefully – Google’s indexing of the corpus represented by inscribed textuality on the internet is the sign, I believe, of an order-of-magnitude shift in the scale of the cultural archive and our engagement with it as humans. I provide brief remarks on these issues here, acutely aware that they deserve extensive and detailed consideration.

In a sense the world and the “knowledge” or “culture” that is in it – call it “content” – has not and will not change. Human life is what it is. Nonetheless we tend to agree that our ability to archive this content in order to make it recordable and manipulable has radically changed during the modern period. Scholars of the age of Francis Bacon began to lose hold of any sense that they might read and thus know “everything”. In the maturity of print culture, we have long ago lost sight of being able to read or “know” everything in a particular discipline, let alone “everything” *per se*. However, we were wont to believe that all inscribed textuality might be collected in libraries or traditional archives and that, at the very least, a “union catalogue”, the product of human labor, would be able to give us access to any necessary article of knowledge, with universities curating and signaling the originality of purported contributions to this sum of content. However, just as the efflorescence of print made it literally impossible to read everything, the explosion of content-creation that is enabled by programmable and networked media now makes it literally impossible for humans to *index* everything in their archives. Humans are already, now, not able to create a catalogue of the articles of culture that they have, collectively, created.

Instead, humans write software, processes that will index these archives. These processes will reflect human culture back to its maker-consumers and consumer-makers. This is already what Google does for us. At first it seemed that the company did this almost gratuitously, more or less as a function of Silicon Valley utopianism and naivety. Now this intensely, importantly cultural service is fundamentally skewed and twisted by commerce, by a requirement to generate advertising revenues that are dependent on the most advanced forms of capitalism. These circumstances may have been all but inevitable, but the time for decisions has come. What computational processes do we want to create and have running for us, in order to index or otherwise represent for us the contents of the cultures that we are making?

Abby Smith Rumsey, Director of the Scholarly Communication Institute, University of Virginia Library spoke cogently to these issues, especially in questions following her presentation, *Digital Archives: the Missing Context*, for the *Animating Archives: Making New Media Matter* conference held at Brown University, Dec. 3-5, 2009.

[13] Saturday Oct. 7, 2009.

[14] Clearly, my underlying argument resonates with traditionalist Humanities anxieties about scholarship and the effects on scholarship of the tools and resources which Google has suddenly provided [Nunberg 2009]. However, I am not so much concerned with the preservation of cultural standards. I am entirely content that institutions should change. I just don't think that such change should be at the whim of unacknowledged, ill-considered, and venal forces. The cultural vectors opened up by Google will only ever be able change our institutions coherently and generatively if they remain susceptible to the values and standards of *all* our institutions, not only our mercantile and marketing institutions.

[15] Daniel C. Howe suggests additional reference to Jonathan Lethem, 'The Ecstasy of Influence: A Plagiarism,' Harper's Magazine February 2007 [Lethem 2007]. More recently there is also the novel-as-manifesto-of-appropriation: David Shields, *Reality Hunger: A Manifesto* (New York: Alfred A. Knopf, 2010) [Shields 2010]. The work of the late American novelist Kathy Acker was known for its techniques of appropriation not to say plagiarism. In the story "Pierre Menard, Author of Don Quixote" Jorge Luis Borges imagines a French writer, Menard, who is so able to immerse himself in the earlier work that he "re-creates" it word for word. Recent gestures in the realm of Conceptual Poetics are relevant here. Kenneth Goldsmith's *Day* consists of a straightforward transcription of the Sep 1, 2000 issue of the *New York Times* within the format and design of a standard 836-page book [Goldsmith 2003]. In a further conceptual gesture, Kent Johnson appropriated this work as his "own" with the connivance of Buffalo-based small press Blazevox by simply pasting over all references to Goldsmith, replacing them with a Johnson overlay. I possess a copy of the altered book, signed by the (latter) publisher.

[16] I could, of course, do this in other domains using the resources of other institutions but the thought of what this would mean is overwhelming – a life-changing shift into research on natural language, with single-minded devotion to finding or building the databases one would need. Google promises me an accessible corpus and even tells me that it is always already mine and everyone else's – in good net-utopian terms – but then denies me service at crucial moments when I am beginning to build a poetic.

[17] Extracts from Google's Terms of Service, supplied by Daniel C. Howe:

"2.1 In order to use the Services, you must first agree to the Terms. You may not use the Services if you do not accept the Terms. ..."

"2.2 You can accept the Terms by: (A) clicking to accept or agree to the Terms, where this option is made available to you by Google in the user interface for any Service; or (B) by actually using the Services. In this case, you understand and agree that Google will treat your use of the Services as acceptance of the Terms from that point onwards. ..."

"4.5 You acknowledge and agree that while Google may not currently have set a fixed upper limit on the number of transmissions you may send or receive through the Services or on the amount of storage space used for the provision of any Service, such fixed upper limits may be set by Google at any time, at Google's discretion."

"5.3 You agree not to access (or attempt to access) any of the Services by any means other than through the interface that is provided by Google, unless you have been specifically allowed to do so in a separate agreement with Google. You specifically agree not to access (or attempt to access) any of the Services through any automated means (including use of scripts or web crawlers) and shall ensure that you comply with the instructions set out in any robots.txt file present on the Services."

[18] The following text is based on a short piece by Samuel Beckett that eventually became, as a final text in English, three fragments from *How It Is* [Beckett 1964, 31]. I searched Google for successively longer sequences of double-quote-delimited words from these fragments with the qualifiers: -Beckett -Beckett's -Beckett's (with prime and apostrophe) looking for pages on which the sequences occurred but are not associated with Beckett. Links to selected pages that I found current on Saturday Feb. 15, 2010 have been added to the relevant word sequences in the constructed passage above. They are not, that is, quoted from Beckett. Most of these links still work, but some are now broken and others may, of course, break over time.

[19] In actual fact, I made this text by first alphabetically sorting the gathered sequences and only then rearranging them as little as possible in order to provide some kind of relatively coherent diegesis.

[20] This preliminary piece from *The Readers Project* may be accessed from <http://thereadersproject.org>.

[21] There is a great deal that could be written about *The Reader Project*: about how it operates and engages literary aesthetics from a critical or theoretical perspective, most of which would not be entirely relevant to the present discussion. However, it may be worth noting and commenting briefly on this sense of "proximate". A proximate or neighboring word may be one that is contiguous with a reference word. In linguistics, such a word, for example, collocates with the reference word if it follows it in the line of the syntagm, in the metonymic dimension as

Roman Jakobson called it. Another notion of proximity – in the complementary metaphoric dimension, that of replacement – would see words such as synonyms or antonyms as (virtually) proximate to particular words of reference in the text. However, proximity or neighborhood may also be defined, in *The Readers Project*, in terms of the *typographic*, and this neglected dimension of textuality reveals itself, in our aesthetic analyses, as vital to if not constitutive of reading. (Typography is not, perhaps, neglected as a *graphic* art but it is, arguably, neglected as an art of reading, as a literary art, *sensu stricto*.) Specifically, readers in the project currently have access to databases of information about all the actual word pairs in a text that they are “reading” combined with any (and all) third words existing in the text. Clearly the vast majority of these three-word combinations will not occur anywhere in the text itself as contiguous syntagms. These sequences of three words we call *perigrams* to distinguish them both from *bigrams* and *trigrams* in standard Markov analyses. Once we have derived a text’s perigrams, we then use Google to find counts for their frequencies in the internet corpus, and (for the moment) discard any perigrams with zero counts. This allows the readers to follow the standard syntagmatic line but to check arbitrary typographically neighboring words to see whether they would form a perigram that occurs in the natural language of the Google-accessible corpus. If they do, a particular reader may be allowed to follow the alternate syntagmatic line of reading that it has discovered in its typographic neighborhood.

Clearly “proximity” may be redefined in accordance with other features of linguistic items, including, for example, orthographic features. Thus the “mesostic” reader mentioned above, looks for words containing particular letters and considers them “proximate” if they contain a letter that it requires to read-while-spelling. In point of fact, the current mesostic reader takes further cognizance of physical typographic proximity and also what one might call the relative “perigrammatic proximity” (just described) of two words that it might be about to read, for example, and that both contain the letter it needs to spell. It will prefer to read a word that is more proximate in the maximum number of dimensions.

For a more extensive methodological and computational introduction to *The Readers Project*, see [Howe and Cayley 2011]

[22] See above. A normal Markov model applied to language is only concerned with the syntagmatic dimension of language and takes no account of any typographic structure that it may have. The above definition of *perigrams* in *The Readers Project* takes some account of typography and thus complicates the standard Markov model.

## Works Cited

- Battelle 2006** Battelle, John. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Portfolio, 2006.
- Beckett 1964** Beckett, Samuel. *How It Is*. Translated by Samuel Beckett. New York: Grove Press, 1964.
- Beckett 1995** Beckett, Samuel. *The Complete Short Prose, 1929-1989*. Edited by S.E. Gontarski. New York: Grove Press, 1995.
- Bosker 2010a** Bosker, Bianca. *Google Instant Blocks Sexy Searches*. *Huffington Post*. September 10 2010. [http://www.huffingtonpost.com/2010/09/09/google-instant-search-blo\\_n\\_710199.html](http://www.huffingtonpost.com/2010/09/09/google-instant-search-blo_n_710199.html).
- Bosker 2010b** Bosker, Bianca. *Google Shuts Down China Search, Redirects Users to Hong Kong*. *Huffington Post*. March 23 2010. [http://www.huffingtonpost.com/2010/03/22/google-leaves-china-googl\\_n\\_508639.html](http://www.huffingtonpost.com/2010/03/22/google-leaves-china-googl_n_508639.html).
- Diski 2011** Diski, Jenny. “Short Cuts”. *London Review of Books* 33: 2 (2011), pp. 20-20. <http://www.lrb.co.uk/v33/n02/jenny-diski/short-cuts>.
- Goldsmith 2003** Goldsmith, Kenneth. *Day*. Great Barrington: The Figures, 2003.
- Howe and Cayley 2011** Howe, Daniel C., and John Cayley. “The Readers Project: Procedural Agents and Literary Vectors”. *Leonardo* 44: 4 (2011), pp. 317-324. [http://www.mitpressjournals.org/doi/abs/10.1162/LEON\\_a\\_00208](http://www.mitpressjournals.org/doi/abs/10.1162/LEON_a_00208).
- Lethem 2007** Lethem, Jonathan. “The Ecstasy of Influence: A Plagiarism”. *Harper’s Magazine* (February 2007), pp. 59-71. <http://www.harpers.com/archive/2007/0081387>.
- Michel 2011** Michel, Jean-Baptiste, et al. “Quantitative Analysis of Culture Using Millions of Digitized Books”. *Science* 331: 6014 (2011), pp. 176-182.
- Montaigne 1910** Montaigne, Michel Eyquem de. “Of the Institution and Education of Children”. In John Florio, trans., *Literary and Philosophical Essays: French, German and Italian*. (*The Harvard Classics*). New York: P.F. Collier & Company, 1910. pp. 29-73.
- Nunberg 2009** Nunberg, Geoffrey. “Google’s Book Search: A Disaster for Scholars”. *The Chronicle of Higher Education*

(August 31 2009). <http://chronicle.com/article/Googles-Book-Search-A/48245/>.

**Shields 2010** Shields, David. *Reality Hunger: A Manifesto*. New York: Alfred A. Knopf, 2010.

**Wardrip-Fruin 2009** Wardrip-Fruin, Noah. *Expressive Processing: Digital Fictions, Computer Games, and Software Studies*. Cambridge: MIT Press, 2009.

**Watts 2010** Watts, Jonathan. "China's Internet Crackdown Forced Google Retreat". *The Guardian* (January 13 2010). <http://gu.com/p/2dnqk>.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.