

READING, WRITING, RESISTING: LITERARY APPROPRIATION IN THE READERS PROJECT

Daniel C. Howe, Run Run Shaw
Creative Media Center, City University
of Hong Kong
email: daniel@rednoise.org

John Cayley, Department of Literary
Arts, Brown University, Providence, RI,
02912-1923, U.S.A.
email: John_Cayley@brown.edu

Abstract

The *Readers Project* is an aesthetically-oriented system of software entities designed to explore the culture of human reading. These entities, or 'readers', enact specific reading strategies and function as autonomous text generators, networked writing machines visible beyond the texts they 'read'. As the structures on which they operate are culturally implicated, the project's readers shed light on a range of institutional practices surrounding the digital literary and the aggregation of the linguistic commons by corporate interests. In this paper, we present the practical and theoretical considerations guiding the project's development, and consider various strategies to resist the commodification and enclosure of literary culture within the corporate 'cloud'.

Keywords: linguistic commons, networked language systems, critical algorithmics, copyright and intellectual property, network services, aesthetic computation, natural language processing, conceptual writing/literature, digital language art, writing digital media, literary visualization, computational writing, (human) reading.

Introduction

The *Readers Project* deploys algorithmic readers onto inscribed surfaces that model and explore the multidimensional culture of human reading. For a description of the design and behaviors of selected 'readers', please see our earlier article '*The Readers Project*: procedural agents and literary vectors'. (*Leonardo* 44.4)

We concluded this essay by positioning the project as both a celebration and a critique of the institutions of reading and writing. In today's world, so much of our symbolic, language-driven engagements are now transacted with algorithmic agents that what we think of as reading and writing, those practices heretofore considered as proper to human 'authors', have been fundamentally altered. The 'cloud'-based services provided by networked aggregators such as Google and Facebook give us instant and convenient access to all the linked references that we need, habitually and professionally, as we write or read. By default, these sources have superseded previous reference works and fact-

checking tools. Few human authors, however, acknowledge the fact that they have become prosthetically 'enhanced' by algorithm. By way of example, the text we are now composing 'lives' on 'Google Drive', allowing its authors to collaborate more or less effectively. As we write hundreds, if not thousands, of algorithmic agents work for us in various capacities: saving, transferring, attributing and time-stamping our edits; bringing up windows of 'research'; framing our writing with indicative paratext. (Linked advertisements, for the moment, are just off screen) [1].

Some portion of this computational compositional work is displayed on screen 'in plain sight', but much more takes place 'in the cloud', as a function of the proprietary indexing and analysis of Google. It feeds the 'big data' that we hand over with each keystroke to massive corporations behind proprietary firewalls, with whom we have—tacitly, all but by default—agreed to terms: terms of use, terms of service, statements of rights and responsibilities, etc. The primary function of these terms is to grant such companies the right to pass the 'data' we are generating on to the algorithms they own, for the derivation of hugely profitable, tangentially related sales [2]. When we make a thing—this sentence, its phrases—in 'private', do we expect it to be instantly and automatically appropriated for purposes that are distinct from our own—and of which, perhaps, we disapprove? Too late. We have already agreed that this should be the case. However, this agreement is quite clearly not reciprocated. Perhaps strangely, the appropriation of what we create does not seem to occur in other domains of our lived experience, and might well be found to be illegal or immoral if it did. It is precisely here that the dimensionalities of language, of reading and of writing have already changed, radically and by default, without our comprehensive agreement or understanding; and this vectorization, this appropriation of symbolic practice by 'big data' and 'big software', threatens to coerce, by statistical force majeure, the totality of culture, including material culture and exchange, into those channels of activity and attention that it considers 'best', 'not evil', or most profitable; but for whom, exactly?

Which brings us to the question of how the *Readers Project* relates to, and engages with, these massive and ubiquitous networked agendas? To start, the project recognizes these circumstances

as a profound reconfiguration of the predominant modes of reading and writing. The algorithmic, compositional, and configurative agents of big software's network services have changed the very nature of these activities far more deeply than even the development of hypertext in the 1990s [3]. The project consciously strives to articulate the relationships between these software agents and the new dimensionalities of reading; configuring and setting out, for example, 'motivations'—specific vectors—for its readers. This perspective requires that any critical understanding of the project takes adequate account of its readers' motivations, and provides a model for the understanding and critique of other reading agents on the net, as they are silently or pointedly deployed, both by big software and by the even less benevolently motivated engineers of the Dark Net [4].

Reading

The process of designing variously motivated software 'readers' for the *Readers Project* has transformed our understanding of reading, both as an institution and as a set of cultural practices. Reading only ever takes place, only ever brings language into existence as readable, in a culturally predetermined and located context, predominantly what we have called the 'typographic dimension'. This fact necessitates the precise specification of a textual field or 'neighborhood' which defines the possible foci of attention for each reader. Further, the neighborhood is determined by typography per se, and not, in the first instance, as might be expected, by linguistic or stylistic considerations. For graphically inscribed language, it is the visual form of the text that determines its interpretative context, rather than only its linguistic features or formal structure.

This realization highlights the fact that typical reading practices in all of human culture are as unnatural as they are highly developed [5]. When we read we do something that is unusual, even for animals that have evolved as we have. The requisite reading-attuned visual facility and acuity is remarkable. The way our eyes move as we read is extraordinary, not to say bizarre. In particular, for western readers, there is the hyperspeed movement from right margin to left that, literally, collapses the typographic dimension of prose and makes relatively distant words 'neighbors', both in the terms of our project and in the culture of reading itself. In other words, it is just

such strange, embodied, human, cultivated practices that ‘make reading possible’; that constitute ‘what reading is’.

Thinking further, it becomes clear that, from the perspective of a phenomenologically engaged philosophy of language, the practical art and science of reading that our project delineates is, substantially, a privileged metaphor or allegory, in so far as it is constrained to the particular support media associated with ‘graphically embodied’ language. The symbolic practices of language may be inscribed in, literally, anything [6]. Their (typo)graphical instantiations are, ultimately, pure convention, however elaborate, powerful, or ancient. While the ways in which we read are not necessarily, or even predominantly, determined by the ‘typographic’, our exploration of this dimension in the project shows that reading *has* dimensionalities that may be overlooked or undervalued, and that these are, and will always be, constituted by institutional histories and conventions.

Writing

Since Christopher Strachey’s coding of a love letter generator in 1952, the institutions of writing and reading have been challenged by novel programmable media [7]. Until recently, the efforts of those devising and proposing new aesthetic or expository practices of reading and writing have been transfixed by the novelty of what before seemed impossible: that the graphic word might move, for example, or that text might otherwise change over time [8]. Clearly it may, and it does. However, work focusing on innovation may do so without a sufficient regard for those dimensionalities of reading which, precisely, make reading possible. To explore and critique exactly those dimensions of practice that *maintain* readability during the processes of transformation, requires a degree of artistic and material-cultural care that can be missing in much digital literary work.

We make both stronger and more specific claims. We argue that text generation—as demonstrated by the project’s Perigram, Lookahead, or Mesostic Readers [9]—is more ‘readable’, both literally and in terms of significance and affect, when it takes systematic account of the dimensionalities of reading practice, typography as one example, that continue to determine what is readable for contemporary humans. We hypothesize, for instance, that a Markov model without orientation in the typographic dimension

is less readable, and less likely to generate significance or affect, than one that is projected into some appropriately designed and configured space-time for reading [10]. The consequences are not only literary. As ephemeral screen fodder, a Markov model may continuously spew its syntax without disturbing any institution other than idleness; however, when such a model writes websites and link farms, it populates the ‘Dark Net’ with post-human language, working a textual terraformation of the very world in which so many of us now symbolically exist.

Moreover, the project’s readers themselves reach out and ‘use’ network services. Explicitly non- or post-human, but programmed with aesthetic or cultural-critical motivation on behalf of human readers, the project’s agents query and transgress the access points provided by big software, attempting to negotiate and interact with the software architectures and algorithmic agencies they discover [11]. The service we have used most consistently and with significant results is, simply, search. For exact-match phrases, the counts of search results provide, effectively, real-time probability models across a huge corpus. Constructing searches that allow us, for example, to filter out those results associated with a specific author have enabled us to develop works—distributable and performable outcomes of the project—that engage with conceptual literature, rigorously problematizing contemporary conceptions of authorship, of copyright, of moral rights. These works explore just those characteristics of language that make it common. Language, inscribed or freely performed, is an inalienable commons which is nonetheless under threat precisely due to network services’ intrusions and enclosures. The project’s readers read, but as they do so, they test and transgress the algorithmic agencies of reading’s new dimensionalities. They resist. They resist the enclosure of language, its vectorialization, the automation and alorization of symbolic practice. They seek out strategies of resistance in aesthetic vehicles.

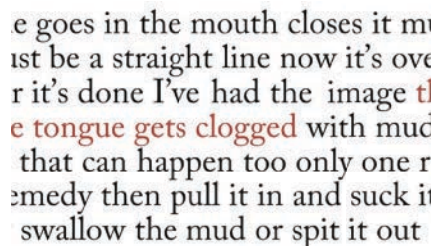
Resisting

The *Readers Project* is inherently long-term, and in continual development. Rather than a work in itself, it is more a framework [12] for the development of aestheticized outcomes. In a note from our previous brief introduction to the Project, we wrote that:

the existence of “services” (or pretended cultural vectors) such as those provided by Google, combined with a burgeoning, aesthetically motivated “use” of these services, has profound implications for contemporary artistic practice. Such use also allows artists to engage critically and productively with important socio-economic and political developments in an unprecedented manner. [13]

To address these implications, and elucidate those aspects of the project that critically engage with and resist the aggregation and enclosure of culture by big software, we turn to recent installations and publications revolving around the rubric *Common Tongues* [14].

In one recent installation, visitors enter the space opposite a projection displaying a ‘perceptual reading interface’ [15]. This interface maintains a focus of proposed reading attention at the center of the screen, while continually reassembling a typographic neighborhood around it [16].



e goes in the mouth closes it must be a straight line now it's over it's done I've had the image the tongue gets clogged with mud that can happen too only one remedy then pull it in and suck it swallow the mud or spit it out

Figure 1

Figure 1 shows a Phrase-Finding Reader navigating a short text by Samuel Beckett, entitled ‘The Image’, later incorporated into his last longer prose work *How It Is* [17]. Unlike Readers described previously, the Phrase-Finding Reader follows a strict left-to-right linear progression through a text, as would a typical western reader. Rather than moving word-by-word, however, it reads the text in short phrases according to what we have called the ‘Longest Common Phrase’ algorithm. The procedure for the selection of phrases follows.

A ‘Longest-Common Phrase’ (or LCP) is the longest sequence of words, beginning from a specific point in a text, that can be found on the web, *not* written by the author or about the text in question. The *LCP algorithm* is the procedure, generally employed by Phrase-Finding Readers, to locate such phrases, via queries to public search engines like Google and Bing. The algorithm begins

```

LCPDriver(text, 10)    # start the driver

LCPDriver(text, max) :: Phrase[]

    var cursor = 0, phrases = Array
    while:
        var p = LCP(theText, cursor, max)
        if (p):
            phrases.add(p)
            cursor += p.words.length
        else
            exit('no phrase')
        end if
    end while
    return phrases

end

LCP(text, cursor, max) :: Phrase

    var found = false, offset = max
    var words = text.substring(cursor, cursor+offset)


    while (!found):
        var matches = doSearch(words + "-beckett")
        if (matches > 0):
            return Phrase(cursor, words, matches)
        end if
        words = text.substring(cursor, cursor + --offset)
    end while

end

type Phrase:
    var textindex      # the phrase's position in the text
    var words           # list of words in phrase
    var matches         # list of matching URLs
end

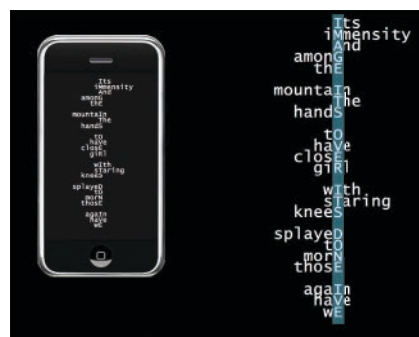
```

In a *Common Tongues* installation, the Phrase-Finding Reader may employ the LCP algorithm to read through the Beckett text via the perceptual interface; at the same time, on a screen recessed into a lectern, we may present the text in an animated codex-like opening, onto which the Project's readers—Perigram, Mesostic, Lookahead—have also been released. Here, viewers can experience the readers moving through the text as originally set in type, rather than via the perceptual reading interface. Additionally, on a set of smaller, iPad-like screens, visitors may view the generated output of each reader as described above. Networked clients, through which visitors



what else
the little by little
among their white spots
lamb
little by little by little
among white spots
lamb
little by little
among their dams
what else
the bluey bulk
closing the mountain

For Perigram and Mesostic Readers, such clients manifest primarily as text generators (see figures 3 and 4).



For the Phrase-Finding Reader, however, we have developed a series of textual visualizations that reveal when a phrase is or is not ‘common’, when it is unique to an ‘authored’ text, and when it has been commonly or uncommonly inscribed by one (or many) others.



An additional ‘client’ deserves special attention here: specifically the printed, full-length book, *How It Is in Common Tongues*. We refer to this book as the ‘client’ of its processes of manufacture; while it is not a client in the usual sense, animated in real or networked time, it is nonetheless a client in that it bears the programmatic traces of its generation. It is also, arguably, the artifact of the *Readers Project* that most persistently

past⁸ moments old dreams⁹ back again or fresh like those⁸ that
pass⁹ or things things always and⁹ memories I say them as I¹⁰
hear them murmur¹¹ them in the mud

in¹² me that were without¹³ when the panting stops¹⁴ scraps of
an ancient voice in¹⁵ me not mine

my¹⁶ life last state last version¹⁷ ill-said ill-heard ill-recaptured¹⁸
ill-murmured in the¹⁹ mud brief²⁰ movements of the lower face
losses²¹ everywhere

recorded^{1c} none the less it's^{1d} preferable somehow^{1e} somewhere

[illegible]

As a 'book', *How It Is in Common Tongues* resists interpretive assimilation in terms of copyright or the related moral rights of association and integrity; as a client of the *Readers Project* it is a critical aesthetic artifact.

How It Is in Common Tongues also resists conventional reading. As a process that rediscovers common language in an authored text, the book represents a conceptual interruption in the reading of a modernist classic; one that directs our attention toward the words that ‘belong’ both to all of us and to none of us. Less obviously, perhaps, the book and its making also resist the control of the search interfaces and indexes that were employed to enact and prove its principles of generation. These processes reclaim information about our language from services that have been built from the appropriation of our collective linguistic practice. Our literary aesthetic agents ignore and transgress network services’ unilaterally-asserted ‘terms of use’, and build from this resistance a conceptual literary artifact intended as both commentary upon, and critique of,

the vectoralization of search; especially of search understood as linguistic practice and as practice-based research [20].

In Conclusion

We are far from having exhausted descriptions, analyses, or elaborations of the implications that might be derived from the Project's readers and their readings. When we say that the *Readers Project* is aesthetically oriented, we are aware that we may be constraining its practices and outcomes within certain specific institutions through which critical gestures are unlikely to be translated into action or event as such. Nonetheless, it is an important aspect of digital, networked practices that they act in, and upon, exactly those networks of which they are nodal parts. When the readers execute, they act on the networked world of big software and big data — the same world that is reconfiguring the ways in which humans read and write; and while the project's readers are denied and rejected by the unilateral terms set out by proprietary, global, transgovernmental, historically and contingently predominant, vectoralist 'services', the project's resistant appropriative algorithms are coded by, and for, those humans who are invested in the exploration of the contemporary literary, and in the cultures of reading and writing. These quiet readers, by demonstrating how they read, and by generating new readings, perform aesthetically and also 'act'. In doing so, perhaps, they can help us to do the same.

References and Notes

1. John Cayley writes from within Google Apps for Education as commissioned by Brown University, relatively protected from commercially motivated software agencies. Daniel Howe shares this document through his 'private' Gmail account, the terms of which may be more open to such agents' intrusions, though browser plugins like Ghostery and Adblock Plus mitigate their effects. As Howe works, his network monitoring software tells him that 3 servers (internet addressable computers) for Google Drive and 13 servers for Google Chrome are managing the transactions concerned with this 'writing'. For several of these servers, imagine hundreds of agents handling specific tasks and queries, many of these quite unrelated to the task of composition that Cayley (the human) pretends to undertake.
2. See Cayley, John. 'Terms of Reference & Vectoralist Transgressions: Situating Certain Literary Transactions over Networked Services', *Amodern* 2 (2013): <<http://amodern.net>>.
3. Once the hype had settled—as was often pointed out in the literature—it became clear to both scholars and practitioners that most of the formal characteristics of hypertext (linking through cross-reference or indexing, intertextuality, non-linearity) already existed in print culture, both materially and conceptually.

4. The Dark Net generally refers to those portions of the web not indexed by commercial search engines, and thus invisible to most users. In this essay we refer as much to the textual content of the net that is visible but not authored by humans, as to the invisible and or sequestered regions.
5. Dehaene, Stanislaus. *Reading in the Brain: The Science and Evolution of a Human Invention*, New York: Viking, 2009 provides a good, popular overview of cognitive science approaches to reading with, in chapter one, some indication of the unusual perceptual and cognitive facilities that reading requires.
6. Writing as language-in-the-graphic clearly dominates this and many other related discussions, but humans are all-but-universally familiar with the most common support media for language, namely orality. Moreover, there is the real-world culture of signing: a kind of 'visual orality'. In the project and in this essay, we practise and write after Jacques Derrida's *Of Grammatology*, and intend what we say about reading (and writing) to be applicable to linguistic practice in general.
7. Wardrip-Fruin, Noah. 'Digital Media Archaeology: Interpreting Computational Processes', *Media Archaeology: Approaches, Applications, and Implications*, eds. Huhtamo, Erkki and Parikka, Jussi. Berkeley: University of California Press, 2011, 302-22 —describes and analyzes Christopher Strachey's love letter generator, developed in 1952.
8. Two books by Christopher Funkhouser provide numerous examples: Funkhouser, Christopher T. *Prehistoric Digital Poetry: An Archaeology of Forms, 1959-1995*, Tuscaloosa: University of Alabama Press, 2007; and *New Directions in Digital Poetry*. New York: Continuum, 2012. A well-argued celebration of the openness of potential digital literary form can be found in: Eskelinen, Markku. *Cybertext Poetics: The Critical Landscape of New Media Literary Theory*. London and New York: Continuum, 2012.
9. Descriptions of these readers can be found on the project website at <http://thereadersproject.org> and also in Howe, Daniel C., and Cayley, John. 'The Readers Project: Procedural Agents and Literary Vectors', *Leonardo* 44.4 (August 2011): 317-24.
10. Markov models or chains, based on the relative frequencies of collocations for arbitrary-length sequences of symbolic or linguistic tokens, are commonly used engines for text generation.
11. Deliberately avoiding official APIs and pointedly transgressing terms of service. See Daniel C. Howe's Processing and Javascript libraries for natural language processing and generative literature, *RiTa*, at <http://www.rednoise.org/rita>; and Howe, Daniel C. and Nissenbaum, Helen. 'Trackmenot', 2008, <http://trackmenot.org>.
12. The *Readers Project* framework is, in turn, built upon the RiTa toolkit, described in Howe, Daniel C. 'RiTa: creativity support for computational literature', in *Proceedings of the seventh ACM conference on Creativity and cognition (C&C '09)*. ACM, New York, NY, USA, 205-10.
13. Howe, Daniel C., and Cayley, John. 'The Readers Project: Procedural Agents and Literary Vectors', *Leonardo* 44.4 (August 2011): 317-24, note 18.
14. First installed for the exhibition associated with the ELMCIP organized conference 'Remediating the Social', at Inspace Gallery, University of Edinburgh, Nov 1-17, 2012. See <http://thereadersproject.org/index.php?p=installationn/rt2012/commontongues.html>. *Common Tongues* is also documented in the ELMCIP Knowledge Base with permanent URL <http://elmcip.net/node/4677>. *How It Is in Common Tongues* is at <http://elmcip.net/node/5194>.
15. We have sometimes referred to this as a 'Perceptual Reader', but this is to use 'reader' in a different sense from that where we write about, for

example, a Perigram Reader. A perceptual reading interface is capable of tracking the reading strategy of any of the Project's readers, always maintaining its neighborhood as the 'focal point' of the interface.

16. Our perceptual reading interface is derived from research on the cognitive science of reading—as centered on the visible field around the point of foveal vision/focus while reading—but it is also inspired by Bob Brown's 'readies', see: Brown, Bob. *The Readies* (originally *Bad Ems: Roving Eye Press*, 1930), with an afterword by Saper, Craig (ed.), Houston: Rice University Press, 2009.
17. Both 'The Image' and *How It Is* were originally composed by Beckett in French. See Cayley, John. 'Writing to Be Found and Writing Readers', *Digital Humanities Quarterly* 5.3 (2011): <<http://digitalhumanities.org/dhq/>>, for further bibliographic details and closely related discussion.
18. The initial K-value for a text is selected via a trial-and-error process that attempts to balance efficiency (the number of real-time searches required) and accuracy (the possibility that too low a value will miss longer common phrases). One optimization of the algorithm involves the selection of a high K-value and a binary search strategy where K is halved (rather than decremented) after each miss.
19. In the 'worst' case, for a totally unique text, the number of phrases will equal the number of words; each word only being found separately. Thus far in the project, no single word has returned a zero count.
20. Some time after completion of this work we learned, via our colleague Clement Valla, of a similar, if visually distinct web-based project, by Stéphanie Vilayphiou (2013), engaged with Michel Hoellebecq's 'La carte et le territoire' <http://bcc.stdin.fr/LaCarteOuLeTerritoire/>