Royal Holloway, University of London

Information Security Group
&
Department of Geography

Centre for Doctoral Training in Cybersecurity

Language in the Age of Algorithmic Reproduction:

A Critique of Linguistic Capitalism

Author:
Pip Thornton

Supervisors:
Prof. Peter Adey
Prof. Keith M. Martin

March 10, 2019

Abstract

Adopting the framework of Benjamin's 1936 essay 'The Work of Art in the Age of Mechanical Reproduction' (1999, 2008), and thinking through central themes of value, context and circulation, this thesis examines how language moves through digital space. Concentrating on Google's search and advertising platforms, the thesis explores the concept of 'linguistic capitalism' (Kaplan 2014), arguing that the ongoing effects on digitally mediated language are both linguistic and political. Indeed, the politics and power which lurk behind these technologies are often obscured, or normalised, by their ubiquity and aesthetics, a process which, following Benjamin, can perhaps only be exposed by the repoliticisation of language through art. Using poetry and classic texts such as Orwell's 'Nineteen Eighty-Four', as well as existing and experimental writing, the thesis harnesses the power of language and literature to critique and resist the technologies that exploit it in today's digital economy. Adopting what I am calling a post-digital (auto)ethnographic approach, the thesis offers an innovative new method in order to make visible the workings and effects of linguistic capitalism, using data gathered from Google AdWords for both quantitative and qualitative analysis, as well as introducing and documenting the development and reception of my own piece of 'political' art in the form of a critical creative intervention called {poem}.py.

DECLARATION OF AUTHORSHIP

I, Pip Thornton, hereby declare that the work presented in this thesis is the result of original research I conducted whilst enrolled in the Information Security Group and Department of Geography as a candidate for the degree of Doctor of Philosophy. This work has not been submitted for any other degree or award in any other university or educational establishment. Where I have consulted the work of others or worked in collaboration, this is clearly stated.

Pip Thornton 17th December 2018

Acknowledgements

This thesis would never had happened were it not for the faith my supervisors Pete Adey and Keith Martin put in me from the outset and throughout. I really can't thank you enough for the support. I know I was slow, but look - I did it! Thanks and acknowledgment must also go to the EPSRC and the Royal Holloway Centre for Doctoral Training (CDT) in Cyber Security for the funding, and the opportunities it allowed, and also for their open minded and ongoing support of a project that evolved into something unrecognisable from what they actually agreed to fund. {poem}.py would never have got off the ground were it not for the generosity and patience of several members of the CDT who helped me with the more technical aspects of the project. Special thanks to Ben Curtis and Feargus Pendlebury for lending their coding expertise, and to Kenny Paterson, Sam Scott, Giovanni Cherubin, Thyla Van der Merwe, and Steve Hersee, who all lent supporting hands and/or shoulders in one way or another, as did Kui McKay, Harriet Hawkins, Alasdair Pinkerton, Rachael Squire, Anna Jackman and Simon Cook from my other home in the Geography department. From my many adventures outside Royal Holloway, I am incredibly grateful for the time given to me and my project by Sam Kinsley, James Ash and John Morris, and for the various opportunities, help and intellectual stimuli provided me by John Danaher, Jessica Foley, David Kelly, Rachel Parry, Bernadette O'Brien, Agathe Henssien, Sophie Skach, Pete Haughie, Chris Bewick, Julie Freeman and Bettina Nissen. Special thanks to my examiners Louise Amoore and Oli Mould, to the participants of the *Living with Algorithms* workshop, where {poem}.py was first unveiled, to John Morris and Phil Garnett for their original piece, and to Matt Spillman at the Frame Workshop in Egham, who made masterpieces from till rolls. To everyone who has ever supplied them on request, thank you, and I apologise for ruining your favourite poems. Extra special thanks and gratitude must go to Andrew Dwyer, Nick Robinson, Andreas Haggman, Adam Badger, Joe Shaw, Brendan McElhill, Thalia Laing, Mike Duggan and Michael Flexer, for their untiring physical and mental support over the last few years. Lastly, I am indebted to the wonderful NHS, and in particular the staff of the Orient Practice, Whipps Cross and Bart's Hospital who swiftly and compassionately diagnosed me with, and successfully treated me for the cancer I discovered I had shortly after embarking on this PhD. I would have struggled to get through that time without the kindness and support of a host of wonderful people including Helen Clark, Anna Thomas, Tamsin Dakoutros, Ian Floyd, Brendan Quilter, David Saffery, Ben Hardey, Helen Williams, Andy Haigh, Sam Barnes, Emma and Peter Martin, and of course my parents, Viv and Steve Thornton.

Publications

Some of the material in this thesis has been previously published, or is currently in the review and revision system. Chapter 4, 'Geographies of (con)text: Language and Structure in a Digital Age' (Thornton 2017), is adapted from an article published in a special section of *Computational Culture: A Journal of Software Studies* on the 'Geographies of Software' (Lally & Burns 2017). Chapter 5 is based on an article currently undergoing revision. Some of the material in Chapter 6, specifically the section on Hayek, was developed in collaboration with John Danaher during my time working with him at NUI Galway, and a co-authored version is currently under review (Thornton & Danaher forthcoming). My {poem}.py provocation and intervention has previously been published as an article for *GeoHumanities* (Thornton 2018), and the speculative fiction piece in Chapter 8 on 'Subprime Language' is a chapter in a forthcoming collection 'Running a City like a Company, & Other Fables' (Kitchin, Graham, Mattern & Shaw 2019). Some of the ideas on subprime language were developed in collaboration with John Morris and presented at the 2017 RGS-IBG conference as part of the sessions on 'Power 2.0: New Digital Geographies' (Thornton & Morris 2017).

This thesis is dedicated (with a huge amount of love and thanks) to my parents, who thought it would never end, and to the memory of my grandmother, Betty, who until the end told everyone my PhD was in criminology.

Contents

I PART ONE : FRAMING / THEORY 16

1 INTRODUCTION 18

1.1 Benjamin's 'Work of Art' essay 20

1.2 Linguistic Capitalism 25

1.3 Aims and research questions 26

1.4 Thesis outline 28

2 POSITIONING LANGUAGE IN THE AGE OF ALGORITHMIC REPRODUCTION 32

2.1 Introduction 32

2.2 Language 32

2.2.1 Language and capital 34

2.3 Critiques of Google 37

2.3.1 Stop searching, start questioning 38

2.4 The politics and geographies of big data & algorithms 39

2.4.1 Data 39

2.4.2 Algorithms 40

2.4.3 Digital geographies 41

2.5 New media and digital humanities 42

2.5.1 (Critical) digital humanities 42

2.5.2 (New) media 43

2.6 Conclusion 45

3 APPROACHING A CRITIQUE OF LINGUISTIC CAPITALISM 48

3.1 Introduction 48

3.2 Researching digital technologies 49

3.3 What's in the (black) box?: researching proprietary technologies 50

3.3.1 Performativity and variation 51
3.4 Thinking critically about digital technology 53
3.4.1 Reverse engineering 54
3.4.2 The socio-technical assemblage 56
3.5 Developing methods 57
3.5.1 The reclamation of serendipity 57
3.5.2 Reading and writing linguistic data 59
3.5.3 Literature & fiction 61
3.6 Towards a post-digital (auto)ethnography 63
3.6.1 The post-digital 64
3.7 Conclusion 65

4 GEOGRAPHIES OF (CON)TEXT: LANGUAGE AND STRUCTURE IN A DIGITAL AGE 68
4.1 Introduction 68
4.2 What happens when words become data? 69
4.3 What do I mean by (con)text? 70
4.4 Language spaces 72
4.4.1 Searching 74
4.4.2 Translating 77
4.4.3 Defining 79
4.5 Language games 81
4.6 Language markets 83
4.7 Conclusion: The prison house of digitised language 84

II PART TWO: PRACTICE / ENGAGEMENT 88

5 LANGUAGE IN THE AGE OF ALGORITHMIC REPRODUCTION 90
5.1 Introduction 91
5.2 Why Benjamin? 93
5.3 What is language in the age of algorithmic reproduction? 95
5.3.1 Search Engine Optimisation 97
5.4 The death of the reader 99
5.5 The commodification of language 102
5.6 The aestheticisation of politics 106

5.7 Conclusions 110

6 THE POLITICAL ECONOMY OF ADWORDS: LINGUISTIC MARKETS AND ALGORITHMIC GOVERNANCE 114

6.1 Introduction 114
6.2 The anatomy of a large-scale hypertextual Web search engine 116
6.2.1 Algorithmic markets 118
6.2.2 Hayek's argument 120
6.2.3 Algorithmic governance 122
6.3 Google's AdWords market 124
6.3.1 The (un)wisdom of algorithmic markets 128
6.3.2 Google AdGrants 130
6.4 Political advertising: AdWords and the dementia tax 134
6.4.1 Winners and losers in linguistic capitalism 137
6.5 Can the Google AdWords market be Hayekian? 141
6.6 Conclusion : What knowledge is being produced? 143

7 CRITIQUING LINGUISTIC CAPITALISM: PROVOCATION & INTERVENTION 148

7.1 Introduction 148
7.2 Provocation 150
7.3 Interventions 155
7.4 {poem.py} : a critique of linguistic capitalism 165
7.4.1 Coding Poetry 170
7.4.2 Context 176
7.4.3 Linguistic Geographies 178
7.5 Conclusion 181

8 RESISTING LINGUISTIC CAPITALISM: THEORY & PRACTICE 182

8.1 Introduction 182
8.1.1 Co-option / Recuperation 184
8.2 Modes of resistance 186
8.2.1 Framing and Fading 187
8.2.2 Printing and Performing 194
8.2.3 Destruction and Encryption 196
8.2.4 Dodging the Market 201

8.2.5 Reclaiming poetry from algorithmic market . . . 203
8.3 Resistance and Persistence . . . 209
8.3.1 De-bugging {poem}.py . . . 210
8.4 Speculative resistance . . . 212
8.4.1 Subprime language . . . 213
8.4.2 Derivative value and indifference . . . 217
8.4.3 The Circulation of Economic Value . . . 220
8.5 Conclusion . . . 222

9 CONCLUSION 226
9.1 Contributions . . . 226
9.1.1 Language in the age of linguistic capitalism . . . 227
9.1.2 Raging against the machine . . . 230
9.1.3 Exposition / intervention . . . 232
9.1.4 Context / Value / Circulation . . . 234
9.2 Future work . . . 236
9.3 Last words . . . 238
POSTSCRIPT . . . 242

Bibliography 248

List of Figures

1.1 Google search for 'wives and girlfriends sexist', February 2014. Screenshot: author's own. 19

4.1 Safiya Noble's 'Algorithms of Oppression' (2018). Photo: Amazon.com . . . 75
4.2 UN Women campaign (2013) 78
4.3 Oxford Dictionary example of *rabid*, October 17 2016. Screenshot: author's own. 80

5.1 Example page from Cabell & Huff's *American Psycho* (2013.) 104

6.1 Adverts returned for a Google Search for 'join isis', 15 December 2016. Screenshot: author's own. 131
6.2 Front Page of the Redirect Method handbook (Jigsaw 2015). 133
6.3 Google search for 'dementia tax', 22 May 2017. Screenshot: author's own. . 135

7.1 Christophe Bruno's 'AdWords Happening' (2002) 157
7.2 Example of Google Poetics method (Nuotio 2012), June 2014. Screenshot: author's own. 159
7.3 Example of AI generated Harry Potter text (Botnik.org 2018). 161
7.4 William Stafford's 'At the Bomb Testing Site' ({poem}.py 2016). 167
7.5 Daffodils (2016): A critique of linguistic capitalism - poster for RHUL Information Security Group Open Day. 170
7.6 Daffodils (2016): A critique of linguistic capitalism - poster for RHUL Humanities & Arts Research Institute poster competition 171
7.7 Snippet of the original {poem}.py code, courtesy of Ben Curtis (2016). . . . 173
7.8 {poem}.py receipt printer, AKA the Deconstruction Machine / Black Box. Photo: Amy Freeborn. 175

7.9 Suggested bid price (pounds) for complete text of Alan Ginsberg's 'America' from 1 July 2016 to 31 May 2017 in US, UK and ALL Google AdWords markets. 180

8.1 {poem}.py exhibit at Inter/Sections, London 2017. 187
8.2 Paul Muldoon's 'Something Else' ({poem}.py 2016). Photo: Rachel O'Dwyer. 191
8.3 Public Enemy's 'He Got Game' ({poem}.py 2017). Photo: author's own. . . 193
8.4 {poem}.py printer. Photo: author's own. 194
8.5 {poem}.py receipt spike. Photo: author's own. 197
8.6 'Agrippa: Book of the Dead', William Gibson et al. (1992). 200
8.7 Jemima Foxtrot's 'Bog Eye Man' ({poem}.py 2016). Photo: author's own. . 202
8.8 Engineering Fictions workshop, most expensive love poem, 'Bondage' ({poem}.py 2017). Photo: Jessica Foley. 204
8.9 Engineering Fictions workshop, cheapest love poem, 'Ode-Love' ({poem}.py 2017). Photo: Jessica Foley. 208
8.10 The New York Times, 18 January 2017. 219
8.11 Amazon 'Keep Calm' image (Bridle 2017). 221
8.12 Subprime Language and the Crash (Thornton 2019) 224

9.1 Email received from Google, 30 November 2018. Screenshot: author's own. 232
9.2 Orwell's 1984 (beginning) ({poem}.py 2017). 235
9.3 Larkin's 'This be the Verse' ({poem}.py 2016). 237
9.4 Orwell's 1984 (end) ({poem}.py 2017). 239

We live in capitalism. Its power seems inescapable. So did the divine right of kings. Any human power can be resisted and changed by human beings. Resistance and change often begin in art, and very often in our art, the art of words (The Guardian 2014).

Part I

PART ONE : FRAMING / THEORY

Chapter 1

INTRODUCTION

This was not my intended thesis. This thesis began life as a Google search relating to my old thesis, which was about military geographies. It happened like this:

Five years ago I was sitting in the library of military history at the Royal United Services Institute (RUSI) in London, listening to a presentation about how the spread of social media use amongst military personnel and their friends and family might pose a threat to operational and national security. It was an academic presentation, but was attended by a mixture of serving and retired military personnel, and various RUSI defence and security experts. The audience was predominantly male; the significance of which became strikingly apparent to me when several participants began substituting the phrase 'friends and family' for 'wives and girlfriends'; a semantic irrelevance perhaps, but one that seemed so unrepresentative, culturally loaded, and even a little patronising, that it bothered me enough to research it afterwards. So later that day I sat down at my desk and typed the phrase 'wives and girlfriends sexist' into Google, expecting to find some cultural criticism about the portrayal of footballer's wives in the press and media, or even some more nuanced critique of the sexual semantics of the phrase - something along the lines of Cynthia Enloe's 'womenandchildren' (1990), which cleverly conveys the frustrating removal of agency and personhood embodied by such a conflation (Sjoberg 2006, 2007). What I was not expecting was for Google's search algorithm to 'correct' the word *sexist* to *sexiest* (see Figure 1.1) and present me with a page of the top ranking sexiest and hottest WAGs in its index.

It seemed at first that the search engine had 'decided' that it was more likely that the typical Google user had meant to search for sexy, as opposed to sexist women, and had

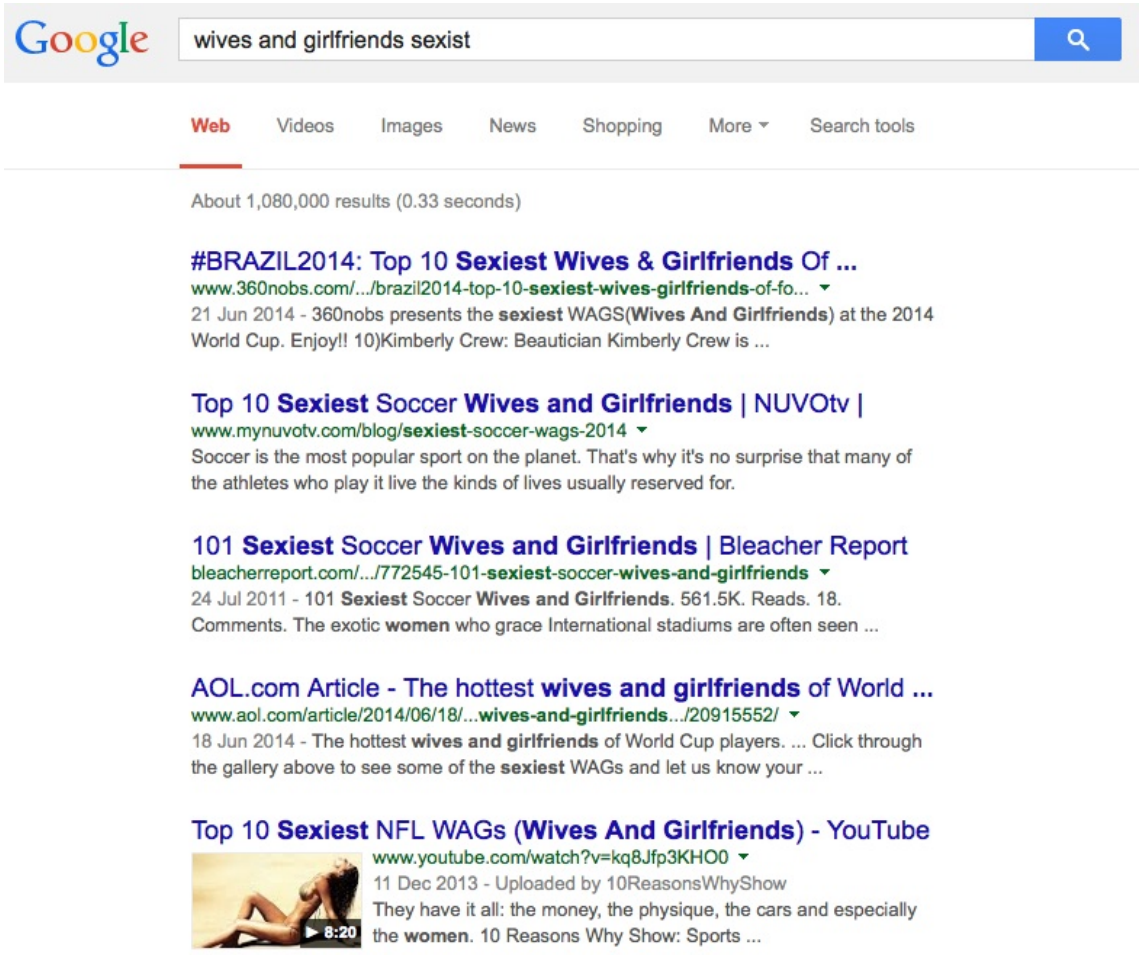


Figure 1.1: Google search for 'wives and girlfriends sexist', February 2014. Screenshot: author's own.

simply made a typographic error. While this may (arguably) seem like a fair assumption, the more I researched how Google and other search engines operate, it became apparent that algorithmic decisions such as this are not based on any cultural or semantic knowledge as such, but purely on the mathematical logics of search technologies, on vectors, on marked up, decontextualised language, and on the analysis of big data by 'keyness and co-location', a method of corpus linguistics which in the field of Critical Data Analysis has been criticised because it 'omits essential qualities of actual language use' (Fairclough 2013: 20). In the pool of data in which the algorithm works, the word *sexiest* is simply statistically more likely to be linked to the phrase *wives and girlfriends* than the word *sexist*; the associations perhaps compounded by the disproportionate volume of sexualised content or sensationalised news copy in the database, by algorithmic reproduction of erroneous typos or synonyms, and by the machinations of the linguistic marketplace and digital advertising. Words in this sense have become reduced to data; their proximity to one another based not on what we might call a cognitive narrative (Hayles 2012: 179), but a logic of mathematics that codifies and magnifies incomplete or biased data. 'At this point', as Fuller and Goffey suggest, 'semantics is largely irrelevant' (2009: 156).

But if digital technology has made semantics irrelevant, then what is the fate of language when our means of communication and the dissemination of information has become increasingly digitised and algorithmically processed, and therefore always already affected by factors and logics out of our control? What does it mean that a search engine gets to decide who or what can be sexy, as opposed to sexist, and why does this matter? It matters because it is not just sexist/sexiest results that get conflated in this way. Every word we enter into the portal of the search bar goes on a quite a journey before it comes back out again. And that is how I conceptualise this thesis. It is about how words (as datafied commodities) move through the structures and processes of the web, and what happens to them on the way. It is therefore about *context* (in a material and literal sense), *value* (in an economic and artistic sense), and *circulation* (in a material and economic sense). These are themes which run through the thesis, and are also the themes I draw from the Walter Benjamin text which inspired my title, and upon whose framework the thesis hangs.

1.1 Benjamin's 'Work of Art' essay

As I will go on explain further in Chapter 5, the title of the thesis is a play on Walter Benjamin's 1936 *The Work of Art in the Age of Mechanical Reproduction*, and as such

uses the essay as a theoretical frame on which to hang my arguments. Benjamin's essay is a staple text in the disciplines of art history, culture and media studies, and more recently has also been used to shine light on how new digital technologies affect those areas of study. The 'Work of Art' essay (1999, 2008) can perhaps be simply described as a Marxist analysis of the means of artistic production, and indeed that is how Benjamin frames his text from the start, arguing that technologically enhanced mass reproduction of art and artefacts is a process of alienation and distantiation of the work of art from its original 'presence in time and space', a process in which the work of art loses what Benjamin calls its 'aura'. Using examples such as coin stamping, woodcutting and engraving to illustrate early reproductions of art, and then the advent of the printing press and photography, Benjamin acknowledges that the work of art has always been reproducible, but points to the significant political and cultural effects that result from advances in technology, from the mechanically facilitated speeding up of this means of reproduction, and from its greater means of dissemination and consumption. Indeed, the essay identifies that the reproducibility of art is 'connected with the contemporary mass movements'. The 'withering' of the aura of the work of art in the age of mechanical reproduction is therefore not necessarily mourned by Benjamin. He sees mass produced art as having been 'emancipated' from its ritualistic origins, having entered the market place, and therefore the realm of the political, with all its emancipatory potential.

The other important point Benjamin makes is that although ancient artworks such as religious effigies or texts are reproducible, they are only reproducible by 'other men', rather than by means of mechanical reproduction, which is one of the main reasons I see the text as a useful starting point for the analysis of words being processed and reproduced by algorithms rather than humans today. Benjamin's description of the aura is not without its critics, and when it relates to language and literature, can easily be challenged with deconstructionist theories of language as being in a constant state of deferral of meaning depending on its context (Derrida 1976, also Chapter 4 in this thesis). In this way language has never had an essence, an authority or an aura which might be ruined by mechanical, or indeed algorithmic reproduction. While I acknowledge, and indeed embrace parts of this critique, I still think there is value in using it in relation to algorithmically reproduced language, as the words which flow through digital spaces are - like Benjamins 'Work of Art' automatically enrolled in a complicated and pervasive political and cultural economy enabled by the ubiquity of digital communication in modern society. My use of Benjamin's

text in this way is based more on the idea of the movement of words through the search bar, imaging them as commodities, as pieces of raw data, moved further away from the context in which they were inputted into the search engine by algorithmic reproduction of the search engine and its complicated market influences. Thus, any authenticity of language relates more to the intention of the search engine user, and how that intent becomes secondary to economic and algorithmic logics as words are processed and reproduced, rather than to any claimed aura of language itself. This idea of the movement of words through digital spaces as commodities has particular resonance in the age of digital technology, when advertising and information can be said to have taken over from heavy industry as a major accumulator of wealth and influence. Thus today, proprietary algorithms have taken over as the facilitators of the movement of raw materials and commodities (in this case raw data or text instead of coal or steel) for capital (re)production. Algorithms are therefore in effect 'do[ing] for information systems what canals did for mercantilism' (Poon 2013), and making their owners extremely rich and powerful in the process.

Although Benjamin's essay is perhaps best known for its musings on the advent of mass produced photography and film, it contains many fascinating insights into language in the digital age, indeed it is a passage about the distinction between the painter and the photographer that highlights another key concept in my thesis. According to Benjamin, there is a distinct distance between a painting and 'reality'. The painter sees the picture in its totality, as a whole, and at a distance, whereas the picture created by the cameraman in a photograph 'consists of multiple fragments which are assembled under a new law' (1999: 229). The photograph is therefore a 'web' of different components that can be analysed and dissected to pull apart the various fragments, laws and influences that make up its production. The idea of the re-assembly of fragments in this way is a great description of the process of deconstruction and reconstruction of text as it flows through digital platforms and portals as data. Like Benjamin's photograph, the words re-assembled through the search engine cannot be viewed from a distance without appreciating the fragmentary process of their reproduction. Referring in particular to the film and its capacity for close ups, slow motion and still frames, Benjamin calls this new mechanically enhanced artistic medium 'the mutual penetration of art and science', and likens the process to Freud's relatively recent method of psychoanalysis which he says 'isolated and made analyzable things which had heretofore floated along unnoticed in the broad stream of perception' (1999: 229).

As I discuss more in chapter 5, Benjamin saw the avant-garde and anti-capitalist Dadaist movement as an extension of this revolutionary means of dissection and perception, for example in the way they chopped up words and phrases into 'word salad' to create poetry, and made art from old buttons and tickets (1999: 231). By deconstructing, reconstructing and repurposing text and objects, the Dadaists were using the very means of production as art in itself, which meant that the works of art themselves were always already reproductions rather than originals with any self-contained authenticity or aura. It might then make sense to suggest that the decontextualisation and deconstruction of language in the process of its algorithmic reproduction is therefore some kind of radical statement. Indeed, many poets and authors do use snippets and fragments from Google search results or autocompletes to create 'digital art'. However, as I argue in the thesis, these types of work run the risk of under-awareness and under-analysis of their means of production and indeed their dissemination, which makes any intended radical creativity or subversion at risk of co-option back into the systems of digital capitalism and algorithmic governance that increasingly control every aspect of modern-day life, which I talk about in chapter 6.

And this is perhaps the starkest, but also perhaps the most controversial parallel I draw between Benjamin's essay and the systems of digital linguistic capitalism that mediate online communication and the flow of information today. Benjamin recognised that technological advancements were giving more and more people access to art and culture, and that this increasing feeling of participation and agency had become a means of control over the masses. He was writing in the 1930s when Fascism was on the rise in Europe and particularly in his native Germany, and saw in the martial spectacle of Nazism a certain aesthetic which had the effect of appeasing the masses by means of including them in this cultural phenomena, while at the same time obfuscating the politics behind it. Benjamin sees this as politics rendered aesthetic (1999: 234), concluding that 'Fascism sees its salvation in giving these masses not their right, but instead a chance to express themselves' (1999: 234). According to Benjamin, this aestheticisation of politics can only lead to a violent outcome, i.e. war. When I began writing this thesis, my comparison of the political backdrop of when Benjamin was writing and the contemporary sphere was questioned by reviewers of an early paper. Benjamin was after all writing in the politically polarised and increasingly violent era of the 1930s, when fascism was growing more popular despite communist uprisings elsewhere. But as global politics grows ever more unstable

and swings further towards the right, it is becoming increasingly easy (and more common) to extend the comparison to the power and influence of modern-day tech companies such as Google and Facebook, and the politics they practise and facilitate. Social media platforms, the fake news phenomena and micro-targeted advertising are all part of a network of communication and participation which, while giving users the sense of expression and agency, at the same time obscures the politics behind its means of production and capital accumulation.

But the politics and neoliberal logics which lurk beneath Google's empire are often hidden, or normalised by the ubiquity and aesthetics of the near mandatory networked technologies such as Google search and its connected platforms. These are technologies that have become so embedded in our everyday lives that it becomes not only impossible, but also undesirable to opt out or overthrow them. My thesis argues that this is a digitally enabled contemporary instantiation of a Benjaminian aestheticisation of politics, and that the algorithmic reproduction of language through these technological systems not only enacts authoritarian power over its masses of participants, but is also facilitating the spread of fascistic politics we can see proliferating in modern society today. Just like art in the age of mechanical reproduction, language in the age of algorithmic reproduction is therefore part of a process of distancing, decontextualisation and monetisation that has profound political and social consequences.

But as well as its technological causes, in the mass reproduction of art Benjamin also identified a potential cure. While he warned that mass reproduction leads to 'uncontrolled (and at present almost uncontrollable)... processing of data in the Fascist sense' (1999: 212), he also saw a redemptive use for such fascistic reproduction of art, which he identified as being 'useful for the formulation of revolutionary demands in the politics of art' (1999: 212). Quoting the Italian Futurist Marinetti's impassioned plea to poets and artists of the future to be aware of the dangers of fascist aesthetics, Benjamin's solution to the aestheticisation of politics is to 'politicise art' (1999: 235), a provocation which I interpret as a call to expose and resist the forces of digital linguistic capitalism through the repoliticisation of language through my own creative intervention and acts of resistance (as detailed in chapters 8 and 9), thus reclaiming language from the algorithmic marketplace, and returning it to art.

1.2 Linguistic Capitalism

This thesis is born of a love of language; of Stafford's poetry, Orwell's prose, and of the lyrics that define lives. But language is so much more than literature. Language is communication, it is community, it is wealth, politics, and it is power over both people and places (Bourdieu 1991; Foucault 1972, 1990, 2012; Derrida 1976, 1988; Fairclough 2001, 2013). It builds societies and it subjugates them (Anderson 2006; Ji 2004). While this critique of linguistic power may not be new, what is new today is that in an age of digital capitalism and big tech monopolies, words have become commodities which gain in economic value the more and the faster they circulate through digital spaces. Whether scraped from the text of emails (Cabell and Huff 2013), algorithmically policed for plagiarism and other criminal activities (Introna 2016, Gillespie 2014), corrupted as spam (Brunton 2013, Fuller and Goffey 2009), commodified as keywords for digital advertising (Kaplan 2014) or optimised so that search algorithms can read them (Gillespie 2014), the words that flow through the platforms and portals of the Web are all in someway caught up in a system of what Frederic Kaplan (2014) has called 'linguistic capitalism'. Words have become data. And like data, they have become a valuable and powerful commodity. This is perhaps most apparent in the way Google monetises the language that flows through the search engine via its AdWords platform, which is the main source of the company's wealth. As Kaplan explains in his 2014 article 'Linguistic Capitalism and Algorithmic Regulation', each Google search triggers an auction for the words contained in the search query, with advertisers bidding for particular keywords in order to obtain the most prominent positions on the results page. Indeed, with the near ubiquity of Google's platforms and advertising empire, which in effect strip narrative context from the words we use while at the same time loading them with dissonant capital, it has become almost impossible to critique the system without adding to its economic value. The text of Emails, blogs, news, search queries and literature has all in some way become data, generating capital for one private, opaque and ultimately unchallengeable company. This is a neoliberalisation of discourse at a fundamental and systemic level. But language; the way we, as humans, communicate, and the way information is disseminated, is fundamental to how human society functions, and to how culture, politics and society is constructed and mediated. And more than this, language is art; it displays, conveys, and indeed provokes human emotion. It is so much more (or so much less) than data.

1.3 Aims and research questions

Yet what is to be done about this from within the system? As scholars and researchers have shown (Kitchin 2017, Pasquale 2015), attempting to critique or explain any system or platform that functions with the use of proprietary algorithms is difficult and problematic. The workings of capital generating algorithmic systems such as Facebook and Google are closely guarded 'black-boxed' trade secrets. In addition to this, tweaks, a/b testing, and other personalisations such as geolocation (Zook and Graham 2007), can also make it hard to research objectively. It is today surely almost impossible to conduct critical research into, or with the help of digital technology without somehow utilising - consciously or unconsciously - the very structures we seek to critique.

This thesis acknowledges these methodical hurdles, and instead takes a different approach in order to make visible the workings and politics of the algorithmic systems of linguistic capitalism through provocation and artistic intervention. As Louise Amoore points out, we should 'be sceptical of claims about 'opening the black box' of the algorithm in order to have some kind of accountability... we must begin instead from notions of opacity and partiality' (2018). Citing Donna Haraway's call to 'stay with the trouble', Amoore suggests that the opacity of algorithmic forms of governing might best be punctured with 'partial accounts' (Haraway 2016, Amoore 2017). Haraway's concept of 'situated knowledge' (1988) is a theme I explore further in relation to the political economy of Google Adwords and alongside Freidrich Hayek's 'knowledge problem' in Chapter 6.

Acknowledging the value of 'situated' and 'partial' accounts, I am therefore purposefully making my critical stance from within the master's house (Lorde 2012). As Audre Lorde's famous feminist provocation goes, 'the master's tools will never dismantle the master's house', a metaphor that brilliantly conveys the frustrations of resistance from within a racist and patriarchal society, and which I think illuminates some of the problems of resistance and agency within today's systems of digital capitalism. If the only tools available for critique and resistance (and tools can mean anything from actions, to theories and words), belong to whatever system of control or governance you are trying to resist, then they are in effect useless, or dis-armed; ineffective at breaking down the systems and structures to which they belong. And further to this, anything you do construct with those tools will necessarily become part of those same structures of power; vulnerable to co-option and re-enrollment into the very systems you were trying to dismantle. My ap-

proach to this paradox of resistance is to overtly work from within the system, exploiting Google's platforms as much as possible (in unofficial harvesting of data from Google's AdWords platform), acknowledging my own part in the process, and ultimately trying to turn the system and its algorithms back on itself, or as Mahnke and Uprichard suggest, 'algorithming the algorithm' (2014). As explained in Chapter 3, this is a method I have called a 'post-digital auto-ethnography'. My project thus seeks to explore the power and influence that Google has over the tools we are increasingly coerced and drawn into using in order to communicate and function, and to expose the potentially serious consequences this linguistic power has within society and the wider discourse. In doing so, I plan to chip away at the master's house from within, hoping that the more people who see and understand the structural problems it has, the more people will choose not to visit it, or at least will do so with a heightened awareness of the political and cultural power embedded in its foundations.

As already mentioned, there are broad conceptual themes that thread through this project, those of *context*, *value*, and *circulation*. Before I move on to my research questions, I want to briefly set out what I mean by these terms, and what work I see them doing in working towards a theoretical and practical conclusion to the thesis. 'Context' in this sense is deployed as both a linguistic device which might alter or construct broader narratives poetically or politically, but also in a more physical sense, as in the way data (and in particular the data that represents language, which I refer to as linguistic data) is sorted, moved and constructed in digital spaces. 'Circulation' thus refers to the actual movement of this linguistic data around these spaces, be it a word going into a search bar, or an email being sent and received. As the thesis argues, the circulation of linguistic data is both a political, and an economic issue. The movement of words around the web (be it search results, spam, or news stories) is the main driver of success in an age of digital capitalism, so circulation in this way also refers to the movement of linguistic data as commodities embedded within new political economies. Circulation and context are then intrinsically linked, and in the context of this project combine to determine the 'value' of language within the new political economy of digital - or more specifically - linguistic capitalism. The thesis therefore rests on the tension between the value of language as a data-ised and movable commodity, and the value of language as a means of artistic expression, emotion, or human narrative and communicative agency. It is not my intention in this thesis to valorise artistic, or poetic language, or indeed any particular form of language above an-

other, but instead to make visible how in an age of algorithmic reproduction, the *economic* value of language has become the overriding mediator of the flow of digital information, which has potentially detrimental effects on other types of linguistic value. As I argue in the thesis, when these different types of value clash, there are significant social, political and economic consequences.

Having explained the central themes and frameworks, I now want to set out the 3 main research questions they will help to focus on answering in the thesis:

1. How is language affected by digitisation/datafication, monetisation and algorithmic processing (broadly speaking the system of linguistic capitalism), and what social and political effects/consequences does that have?

2. What are the difficulties of studying/critiquing language mediated by digital technologies?

3. What can be done to expose, mitigate, or resist the effects of linguistic capitalism? Is it possible to resist/intervene?

1.4 Thesis outline

With the motives and aims of this thesis outlined in Chapter 1, in Chapter 2 I conduct a wide-reaching review of the literature relating to language in the age of algorithmic reproduction. Because of the interdisciplinary nature of the subject, this review ranges from geography, political economy and the social sciences, to media studies and the digital humanities. Following on from that, Chapter 3 will detail the methods I have used, but as the intervention part of my thesis has been somewhat of an adventure and has relied upon both ethnographic and auto-ethnographic approaches and interactivity with audiences and colleagues, the chapter also draws insight from the methods I chose not to use, or indeed, accidentally stumbled upon. The ethnographic method and the subjectivities of researching digital technology is further complicated by its entanglement with binaries such as the digital and the analogue, human and machine. It is with these critical and existential subjectivities in mind, that I set out the overall rationale for my methodology, which I am calling a 'post-digital (auto)ethnography' of the lived experiences and agencies of language in the age of algorithmic reproduction.

Chapter 4 puts forward the concept of 'Geographies of (con)text' to explain and critique the effects of the physical constructs and constraints of language on the web. Imagining a landscape of words as opposed to a landscape of code (Thrift & French, 2002), language-as-data becomes material in ways very different from both print and spoken word; its physicality represented in bits, bytes and circuitry, and its limits and variations mediated and governed by the processes which order, sort, move and index it. I argue that because of their reproducibility and enhanced means of dissemination, digitised words can have paratextual - and often political - agencies and excesses beyond their linguistic function. The chapter uses examples of online search, dictionaries and digital translation to show the wider impact and consequences of how language is (de)constructed on the web, also drawing on (post)structuralist theories of language and information.

In Chapter 5 I will be exploring what exactly 'Language in the Age of Algorithmic Reproduction' *is*, asking questions such as who/what reads and writes the words we find online, how and why is content created, and in particular, what is the role Google and the Search Engine industry has in mediating language. The chapter draws on research conducted on an SEO training course, and uses Walter Benjamin's 'Work of Art' essay as a theoretical lens through which to examine how language has changed in the face of digital technology, arguing that there is a hidden politics and power behind algorithmically reproduced language which has been aestheticised (following Benjamin) by Google's ubiquity and reach within the digital economy.

Chapter 6 examines the 'Political Economy of Google AdWords', critiquing the economic assumptions and credentials of the system of linguistic capitalism as described by Kaplan (2014), which are often adopted by Google itself as justifications for its workings. Questioning the capitalist ideologies and credentials of the term, and the algorithmic and market-based logics that purport to drive it, the chapter enrols Friedrich Hayek's theories of market governance to unravel the complicated political economy of Google's AdWords market, arguing that algorithmic price mechanisms cannot produce the information needed for the efficient market governance, but rather that they have initiated a new form of algorithmic governance. The chapter also draws on empirical data I have gathered from Google AdWords and examines how 'linguistic capitalism' (Kaplan 2014) impacts on the political sphere.

Putting forward political and literary provocations and potential methods of intervention and resistance, Chapters 7 (provocation & intervention) and 8 (resistance) draw together all the strands presented thus far about the value gained and lost by language as it circulates through digital spaces and between changing contexts. Drawing on literary sources which have provided fictional warnings about the authoritarian restriction of language, in particular George Orwell's *Nineteen Eighty-Four*, this pair of chapters foreground my approach to a critical digital humanities, in that it reverses the power dynamic inherent in much digital critique, harnessing the power of language and literature to analyse digital technology, rather than the other way around.

Chapter 7 introduces and documents the development of my own artistic intervention, called {poem}.py, which uses code, poetry and data gathered from Google AdWords to calculate the price of poems fed through the Google keyword planner. The monetised poems are then printed out as analogue receipts and displayed as art, thus destabilising the market logics of digitised language, reclaiming it from the algorithmic marketplace, repoliticising it (following Benjamin), and returning it to art. The chapter thus aims to expose and make visible the workings of linguistic capitalism, while also asking wider questions around language, art and critique in a digital age.

Chapter 8 continues the discussion of my {poem}.py intervention, but specifically documents the difficulties, hurdles and resistant technologies I encountered along the way. The chapter thus explores how linguistic capitalism resists examination, but also how it can itself be resisted by artistic and theoretical critique and intervention. Inspired by the creative power of literature, the chapter also puts forward the concept of 'Subprime Language' as a way to theorise language in the age of algorithmic reproduction, and includes an original piece of speculative fiction which imagines a world in which the unfettered circulation of monetised language has catastrophic consequences for its value and context within human society.

In chapter 9, I will conclude by drawing together the threads from each chapter, addressing my research questions, and detailing some ideas for future work.

Chapter 2

POSITIONING LANGUAGE IN THE AGE OF ALGORITHMIC REPRODUCTION

2.1 Introduction

This chapter reviews the existing literature relating to an examination of language in the age of algorithmic reproduction. The review necessarily spans, and links together, a expansive array of disciplinary fields, from software studies, critical data and algorithm studies, political, human and digital geographies, linguistics and economics, to computer and data science and new media, E-Literature, literary theory and digital humanities. Much of the literature I review (especially the work on digital technologies), is cross disciplinary in nature, so I will organise the next section into four groupings: Language & Capital, Critiques of Google, The Politics and Geographies of Big Data & Algorithms, and New Media & Digital Humanities.

2.2 Language

Weaponised by centuries of successive invasions and colonisations, and manipulated by repressive regimes or systems of governance (Ji 2004; Golumbia 2009), language has always been a tool of power (Joseph 2006; Fairclough 2001, 2013). Language in this way is always capable of potentially devastating political and societal physical and discursive effects, which gives an extraordinary amount of power to those who have control over it. While this critique of linguistic power is not new, especially in as relates to post-structuralist

and post-modernist theory and feminist literary criticism (Lyotard 1984; Jameson 1974; Cixous 1976; Derrida 1976; Foucault 1972, 2012; Haraway 1991; Bourdieu 1991), what *is* new in today's digital age is the manner in which language is processed by algorithms, as monetisable data, and exploited by the big tech monopolies that control the Web and act as the gatekeepers of the world's information.

Changing technologies have of course always had an affect on language in terms of the manner of dissemination of the written word, which, as Walter Benjamin's 'Work of Art' essay demonstrates, can take a piece of art (be it an inscription, painting or sculpture), and facilitate its movement into both the economic and the political sphere. As Benjamin notes, the advent of the printing press, and the subsequent logistical advancements that facilitated the spread of the written word around the world (or at least the Empire), had significant effects on society in terms of the public's ability to consume and to contribute to popular discourse. The printing press really was an 'agent of change' in political and scientific terms since the fifteenth century (Eisenstein 1980), and indeed for Benedict Anderson, print capitalism was a major factor in the creation of 'imagined communities' and the rise of nationalism (Anderson 2006).

But moving on from the printing press, digital technology has delivered another seismic shift in how language is used and the effects it can have, as addressed by work in linguistic anthropology (Kockelman 2010, 2014) and the philosophy of technology (Stiegler 2015, Coeckelbergh 2017). The algorithmic reproduction of language by reduction and reconstruction through binary code has consequences for the integrity and evolution of language and discourse which reach far beyond the relative stability of a printing press cliché. Print capitalism may, as Anderson wrote, have given 'a new fixity to language' (2006: 46), but digitised language is far more lucrative, and far less stable, and it is in this flux of money and words that new forms of power and influence flourish (see also Lyotard 1984). Mercedes Bunz draws on Habermas to call this new environment a digital public sphere (2013).

The viral spreading of Fake News in the run up to the US Presidential election in 2016 was largely funded by digital advertising systems (Graham 2017), and is a clear example of the impact the language disseminated online can have, as are the 'nudging' effects of Facebook newsfeeds on voting behaviours (Tufekci 2015). Likewise, stereotypes and prejudices are compounded by confirmation bias in auto-completions and auto-predictions, and as the

minds of criminals are moulded by what they see on a Google search results page (Noble 2018; Baker & Potts 2013), or on social media (Tufekci 2015). The actionable effects of the words disseminated online through platforms such as Google and Facebook are now taken so seriously that both companies have been under considerable governmental pressure to combat fake news and hate speech, which I elaborate on in more detail in Chapter 6.

2.2.1 Language and capital

In terms of the history of the phrase 'linguistic capitalism' itself, Frederic Kaplan did not coin it with his critique AdWords' effect on language. Indeed, it was used as far back as the 1970s to push back against the corruption of language through advertising. In his book 'Words in Time: a Social History of the English Vocabulary' (1989), Geoffrey Hughes in particular refers to the advertising industry's effect on language as 'verbicidal massacre' (1989: 177), noting how certain words could be appropriated as brand names for advertising purposes, and could be exploited by either their removal from their original context, or their forcible insertion in to another (see also Ding 2017; Jhally 2014) . Pierre Bourdieu's concept of *Linguistic Capital* (1991), while it does not refer directly to advertising, recognised that linguistic privilege through, for example, knowledge of foreign languages, is directly linked to political economy through systems of education (1991: 57).

In the 1990s, the Italian post-Fordist school of philosophers of technology and economics highlighted the links between language and capital in the post-industrial workplace. Scholars such as Christian Marazzi and Franco 'Bifo' Berardi began to recognise the new role of communication on the production line and in the market place, and the direct link between language and labour (Marazzi 2008, 2011; Berardi 2012). Meanwhile, Matteo Pasquinelli developed ideas on language and capital in relation to emerging digital technologies such as Google in what he called 'cognitive capitalism' (2009; see also Moulier Boutang 2011). Pasquinelli saw that in an era of digital advertising, we, as users of these new technologies, our data, our attention and our labour had all in some way become commodities (Pasquinelli 2009, see also Mager 2012; , Nixon 2016). In a direct link to the 2008 Global Financial Crisis, Appadurai's 'Banking on Words' (2015) argues that the 2008 crash was a 'failure of language' facilitated by the new role of language in the marketplace. By this, Appadurai was talking about language of finance, and the words used to perform finance, as having derivative value (see also Martin 2013; Amoore 2011; Wark 2017). In such a way words become promissory notes, or contracts, which he says systematically and conta-

giously failed to deliver. The Global Financial Crisis was also the catalyst for Franco 'Bifo' Berardi's provocations in his book 'The Uprising: On Poetry and Finance' (2012). Berardi uses Kaplan's ideas on linguistic capitalism and Google AdWords to claim that 'the economy is the new universal grammar', and suggests that in an era of *financial* insolvency, an insistence on *linguistic* insolvency is the only way to counter these forces of capitalised technologies; an idea which resonates with the artistic intervention part of my own thesis. The concept of linguistic capitalism is also explored by Bernard Steigler in relation to wider questions around algorithmic governmentality and the 'automatic society' (2015), and by Warren Sack in relation to the spatial make up of digital text (2017), which are areas I engage with in more detail in Chapters 6 and 4 respectively. Other academic literature that directly addresses linguistic capitalism and Google AdWords, includes Richard Graham's (2017) recent article in which he stresses the epistemic importance of Google's role in mediating online discourse, and particularly in light of the fake news debate, but is theoretically derivative of the Post-Fordist approach to digital capitalism. Some of the most incisive critiques of the social and political power of Google AdWords come from the artist and academic Christophe Bruno, whose work has been a great influence to me. Bruno's interventions into what he called 'generalized semantic capitalism' were a very early - and accurate - warning of the power inherent in Google's monetisation of language. On the website of his 2002 'AdWords Happening', in which he made up poems as adverts to try to subvert the advertising system, Bruno warned readers to 'imagine the day when a search engine will rule the whole textual content of the web, in which the memory of mankind will be stored. Think of the power in their hands.'

In a chapter on 'Google: Words Beyond Grammar', art critic and philosopher Boris Groys describes the Google search engine as a 'philosophical machine' which has become the primary mediator of 'our dialogue with the world' (2016: 147). While I would agree with Google's dialogical role, Groys sees the way the search engine processes and decontextualises language as a quasi-utopian liberation from the constraints of grammar: 'Google dissolves all discourses by turning them into the word clouds which function as collections of words beyond grammar' (2016: 149). What Groys' analysis of Google fails to recognise is that, far from enjoying 'extragrammatical freedom', language as processed by Google is constrained and quantified by other grammars, i.e. what Berardi calls the 'grammar of the digital economy' (see above). Groys also sees in the linguistic melting pot of the search engine a kind of philosophical short cut to knowledge: 'Accordingly, true knowledge as

such is understood as a sum of all the occurrences of all the words of all the languages through which humankind currently operates' (2016: 148). Not only does this approach not take into account the hidden market forces and other ranking and censorship factors that affect search engine results, but it also credits Google with some of the knowledge producing ability I discuss in Chapter 6. It is this optimistic view of Google as a 'philosophical machine' that Rebecca Goldstein so brilliantly critiques in her pastiche 'Plato at the Googleplex', where she imagines Plato coming back from the past for an author's talk at the Googleplex in Mountain View. Still in his toga, and oblivious to any of the new technologies around him, Plato is at first in awe of the search engine as the sum of all knowledge, before being horrified that this knowledge engine, which he thinks holds the keys for the answers of so many philosophical dilemmas is in fact used 'for the insignificant aim of making money' (2014: 72), leading him to conclude that it is not 'knowledge' at all that can be gleaned from the search engine, but merely information. These are questions I address in detail in Chapter 6, with reference to Freidrich Hayek's knowledge problem and market governance (1945), and also Donna Haraway's concept of situated knowledge (1988).

The political economy of search engines has also been a direct subject of research in work from Van Couvering (2004), Fuchs (2011), and Feuz e.al (2011). Rather than 'linguistic capitalism', however, Feuz et al.'s paper on 'Semantic Capitalism' (2011) is an experiment with personalisation and Google search, which, although not directly concerned with AdWords, explores the epistemic effects of the capitalisation of the search engine.

More generally, several scholars have engaged with the wider effects of emerging digital economies not limited to Google or indeed language. McKenzie Wark's 'Hacker Manifesto' (2004) is an early warning cry about the power held by what he calls the 'vectoralists'; the technologically privileged classes who control and commodify the flow of information on the Web. Nick Srnicek's more recent book on 'Platform Capitalism' (2016) (see also Langley 2017) provides a solid overview of a range of technologies and their political economies, as do Christian Fuchs & Vincent Mosco in 'Marx in the age of digital capitalism' (2015), and Ezrachi, & Stucke in 'Virtual Competition' (2016) (see also Bratton 2016 and Pasquale 2015 for further critiques of the governing structural and economic architectures of the Web). Mark Poster's early work in internet studies also exposed the political implications of language and culture subjected to economically driven digital technologies. In his 2001

book 'What's the Matter with the Internet', Poster suggested that capitalism had taken a 'linguistic turn', with language becoming private property, and information no longer free: 'What was previously celebrated as spirit, the highest human aspiration, is now for sale as ïntellectual property"' (2001: 39-40).

2.3 Critiques of Google

Since Google first launched in 1998 it has spawned an enormous amount of academic literature in fields such as marketing and business studies (Jansen 2008; Turnbull & Bright 2008) ,legal studies (Grimmelmann 2008; Tan 2009) and computer science (Mehta et al. 2007; Langville & Meyer 2011; Ferragina & Venturini 2013; Bolukbasi et al. 2016; Caliskan et al. 2017) , as well as the raft of popular literature either warning the public about (Battelle 2011; Pariser 2011; Vaidhyanathan 2012; Rushkoff 2016), or promoting the virtues of the magnitude of Google's achievements and enterprise (Levy 2011; Schmidt & Rosenberg 2014) . In addition to this are the thousands of popular books and articles that prop up the Search Engine Optimisation (SEO) industry, and are indeed critical in informing the public discourse and perceived opinion about the workings of Google, a phenomena I explore in greater depth in Chapters 5 and 6.

In terms of critical academic research specifically into the power and effects of Google's search engine, one of the first responses was Introna and Nissenbaum's prescient paper on the politics of search engines (2000) in which they tried to make sense of the then rapidly emerging phenomena of the search engine, stressing the crucial importance of maintaining some kind of control over its already visible power and influence. They anticipated the commodification of information which online search effected, and also its situatedness in and power over the market, and called for public oversight of search engine operation (see also Hinman 2008) and transparency and disclosure of the algorithms. Elizabeth Van Couvering's PhD research (2004, 2007, 2010) on search engine bias was also an important early examination of the field. But what these early researchers of search technology did not anticipate was that one search engine would quickly gain a monopoly, not only of search, but of the market forces and political weight inscribed within it. In 2018, Google had a search engine market share of 92.51% in the UK, 88.6% in the US, and 93.65% in Europe (Statcounter 2019), and with its ever growing portfolio of interconnected platforms such as Google+, GMail, AdWords and AdSense, it is now one of the most dominant play-

ers in the global digital economy.

Since Introna and Nissenbaum's paper, the verb 'to Google' has entered the vernacular, Google apps, schemes and projects have expanded to mapping the world, digitising its literature, keeping its secrets, anticipating its questions, and even driving its cars. Google has become indispensable, and all of this with a lack of transparency so subtle that most people don't care or think to question its motives, accuracy or objectivity, let alone its political implications or the potential effects it is having on wider society (König & Rasch 2014; Fuchs 2015; Halavais 2017). If in the past there was a certain degree of apathy towards the purely commercial nature of potential search engine bias in industry and academia (Van Couvering 2010), the prominence of the search engine in social, cultural, ontological and epistemological ways, as well as a producer of opinion, has certainly changed that and has galvanised the academy into critical exploration.

2.3.1 Stop searching, start questioning

The bias within Google Search is well documented (Introna & Nissenbaum 2000; Granka 2010; Stalder & Mayer 2009; Van Couvering 2010; Gillespie 2014; Noble 2013, 2018), as is the power the company holds (Diaz 2008; Fuchs 2011; Shaw & Graham 2017; Mager 2012; Rogers 2009). Many of the early critiques of web search (some of which have already featured in this review individually) are to be found in Spink & Zimmer's Web Search: Interdisciplinary Perspectives (2008), in Becker & Stalder's 2009 edited collection 'Deep search: The Politics of Search Engines beyond Google' and in Hillis, Jarrett, & Petit's 'Google and the Culture of Search' (2013). All of these collections were important critical interventions into the emerging power and politics of web search in general, and in particular, Google's growing search empire. In a similar vein, but drawing out more of the underlying economics behind Google, the Institute of Network Cultures' edited reader 'The Society of the Query' (2014), is based around papers given at an eponymous conference in 2009 which was given the tagline 'Stop searching, start questioning'. The book draws theoretically on Guy Debord's 1967 Situationist critique of the capitalist spectacle (2016), and pulls together critical and creative work on Google, as well as potential solutions, interventions and alternatives. Although she does not focus on the economies of advertising that fuel and fund the search engine, or indeed the more technical side, Safiya Umoji Noble's more recent book 'Algorithms of Oppression' (2018) is an important study of the gendered and racialised stereotypes embedded within Google search results (see also Noble

2013; Baker & Potts 2013; Olofsson 2015; Jobin & Glassey 2014).

2.4 The politics and geographies of big data & algorithms

Since the early studies of Google and its political side effects and biases, there has been a huge amount of research into the digital technologies on which companies like Google and Facebook operate, in particular the concepts of (big) data, and the algorithm.

2.4.1 Data

My thesis asks what are the consequences of turning words into data, and more specifically, what are the consequences of turning words into data when data is, as many have suggested, a highly lucrative new raw material of the digital age, which generates captial as it moves through digital spaces. Data is (arguably) the new oil (The Economist 2017; BBC News 2017); ready to be mined, extracted, processed and monetised. And just like those who control the extraction and circulation of raw materials like coal and oil are in possession of huge wealth and power, the companies that control data today, such as Google, Amazon, and Apple, are also the wealthiest and most powerful companies in the world, providing services on which we become more dependant every day. Not only is this a worrying geographic distribution of power which mirrors historic colonial western dominance of trade and information (Thatcher et al. 2016), but the power which comes with the ownership and exploitation of data becomes critically important when it comes to precisely what that data represents. Much of it is highly valuable personal data, harvested from search queries, browsing habits, online purchases, social media interactions, email communications and geolocations, for example, often without the users knowledge, or at the very least wrapped up as the negligible and normalised consequences of free software packages, tools, apps or platforms. When data represents such personal details, the control over it becomes a political and social issue.

The merits and perils of big data, perhaps first articulated on a popular level by Chris Anderson in his 2008 Wired piece 'The End of Theory: The Data Deluge makes Scientific Method Obsolete', have been areas of intense academic interest in recent years, across the disciplines. Kate Crawford's work on 'The Hidden Biases in Big Data' (2013), is particularly important in highlighting the social inequalities and lacunae within harvested datasets (see also boyd & Crawford 2012 and Thatcher & Dalton's call for Critical Data Studies 2014). Providing a critical exploration of the politics and ethics of data and data

infrastructures, Rob Kitchin's book 'The Data Revolution' (2014), conceptualises a critical framework for understanding and researching the rapidly increasing amount of data available in today's society, while others seek to address the ethics of information and data from broader philosophical (Floridi 1999, 2001), and scientific (Floridi & Taddeo 2016, Mittelstadt & Floridi 2016) angles. As well as posing significant questions in relation to security (de Goede 2018; Aradau & Blanke 2015, 2017; Amoore 2006, 2014, 2017; Amoore & de Goede 2012; Crampton 2015; Leszczynski & Crampton 2016) and privacy (Powles 2017; de Goede 2014; Graham & Wood 2003, Wood & Graham 2006; Zuboff 2016), the accumulation and analysis of big data can also affect the way we perceive the world (Halpern 2015, Amoore 2016), reflecting hidden colonial legacies (Thatcher et al. 2016), and technological (Kwan 2016; O'Neill 2017), social (Tufekci 2014) and geographical inequalities (Wilmott 2016; Wilson 2011).

2.4.2 Algorithms

In terms of the literature surrounding algorithms, as Thatcher and Dalton's call for 'Critical Algorithm Studies' (2014) states, algorithms need to be studied as best and as critically as we can, not just as technical, but as political, cultural and social issues. The power embedded in the algorithmic processing of data has been a widely explored topic across the disciplines. Drawing on Scott Lash's work on 'Power after Hegemony' (2007), sociologist David Beer was amongst the first to critically question the inherent power of the algorithm, and the apparent democratising, empowering, and inclusive rhetoric of Web 2.0 technologies and social media (Beer 2009; see also Goffey 2008; Neyland 2014). A 2013 conference in New York on 'Governing Algorithms' spawned a large amount of work on the emerging issue of the algorithm from a wide range of disciplinary perspectives. Many of the papers and the articles which resulted from that conference feature in this review, but particularly pertinent reflections on algorithms and their industrial logistical heritage (Poon 2013), and the affect of algorithms on language (Gitelman 2013), were contained in responses given to those papers, and were an important and formative part of my own early thinking on this thesis. It was Poon's concept of the algorithm as the modern transporter of raw material/data, combined with Gitelman's reflections on how linguistic data is constructed/written with algorithmic, rather than human logics in mind, that first brought to my mind Walter Benjamin's 'Work of Art' essay, with its questions about the mobility of art and literature in new economic markets and political spheres.

The authoritarian and controlling potential of algorithmic systems has now been given its own terminology, referred to across the disciplines as 'algorithmic governance' (Just & Latzer 2017; Yeung 2018; Halavais 2017), 'algorithmic governmentality' (Rouvroy 2013), or as 'algocracy' (Danaher 2016, see also Aneesh 2006, 2009), while the wider power in quantification through computational and statistical methods has been thoroughly researched by Golumbia (2009); Beer (2016), Totaro & Ninno (2014) and Cathy O'Neill (2017). Paul Dourish (2016), Ted Striphas (2015) and Nick Seaver (2017) have highlighted the impact of algorithms on (and *as*) culture , while Ananny (2016), Mittelstadt et al. (2016) have turned their attention towards the ethics of algorithms. It is to these newly identified critiques of the governing power and potential of algorithmically processed data that I will be turning in my own critique, but with the explicit focus on linguistic data and on the algorithms that facilitate Google's search and advertising technologies. Although, as I have shown, much work has been done in relation 'the Google algorithm', but my thesis goes deeper in its exploration of the granular structuring and capitalisation of linguistic data by these algorithmic systems of governance.

2.4.3 Digital geographies

In the discipline of Geography, the work on all things 'digital', broadly defined under the auspices of 'digital geographies' (Ash et al. (2016); Wilson 2018), does not only cover the impact that technologies have had on physical space, place and society (Zook & Graham 2007, Leszczynski & Crampton 2016), but is also amongst some of the earliest academic literature in the social sciences to explore concepts such as 'virtual' and 'cyber' spaces (Crampton 2002, 2003; Zook & Graham 2007), and the materialities of digital technologies (Kinsley 2013). The geographical approach has also facilitated a wide range of research into the spaces of computation themselves (Amoore 2016; also Straube 2016 on topological space), and their co-constitution with physical spaces (Kitchin & Dodge 2011; Graham et al. 2015; Thrift & French 2002; Shaw & Graham 2017), which links in with other archival approaches to digital technologies such as Media Archaeology (Parikka 2013) and digital architectures (Carpo 2011). Warren Sack's article 'Out of Bounds: Language limits, language planning, and the definition of distance in the new spaces of linguistic capitalism' (2017) provides a semi-technical exploration of the 'language spaces' of digitised text. Although Sack's article is an exploration of language in terms of the differences and distances between form and meaning, or as Sack writes, 'the concerns of syntax from

those of semantics' and also draws on Kaplan's concept of Google's linguistic capitalism, it is more concerned with the intricacies of Chomskyan linguistics and pedagogy than with what the political and social consequences of linguistic capitalism might be. The crossover between software studies and geography is a set of literature that has been particularly useful in developing the concepts of context and reproduction in my thesis, but which I address specifically in terms of the physical and discursive constructs and consequences of language in a digital age.

2.5 New media and digital humanities

While the economics and politics of algorithms, data and digital technology might generally be thought of in (social) scientific terms, with more technical work in the fields of Computational Social Science, Software Studies and Science and Technology Studies (STS) sitting alongside Digital Sociology, Digital Geographies and Digital Anthropology (as detailed in previous sections above), there are several strands of scholarship within the humanities that also provide a range of critical and creative responses. Drawing together concepts of language and data in a digital age, matters of linguistic structure and context, and the economics and politics of technology, are the relatively new fields of E-Literature, New Media, and the Digital Humanities.

2.5.1 (Critical) digital humanities

There are of course other fields of research within data science that cross over into humanities based areas, such as Natural Language Processing (NLP) and Computational Linguistics. Studies emanating from these fields can be useful in terms of understanding the mechanics of digital technologies, for example two recent studies on word embeddings in vector space, which reveal how stereotypes and biases appear in large data sets such as search engines (Caliskan et al. 2017; Bolukbasi et al. 2016). However, they can often be frustratingly uncritical of the potential consequences of the manipulation of data, and also in their treatment of its provenance, as shown by a recent analysis of the Urban Dictionary by data scientists at the Alan Turing Institute who claimed to 'make sense' of the project, but failed to acknowledge its problematic racial history (Nguyen et al. 2018).

The discipline of the Digital Humanities has to a certain extent grown out of the logics of

data science and computation, in that many Digital Humanities departments at universities grew out of need to provide support across the disciplines, and particularly in the social sciences and humanities, for researchers using new digital methods to enhance the scope of their work. As such, what we might call the 'traditional' Digital Humanities has drawn criticism from scholars wary of the uncritical use of computational methods when it comes to trying to make sense of language, for example Lisa Gitelman and Paul Edwards' edited collection 'Raw Data is an Oxymoron' (2013; see also Bowker & Star 2000 on the power of classification). Likewise, Stephen Marche's passionate essay 'Literature is not Data: Against Digital Humanities' (2012) insists that 'insight' remains the preserve of humans, not data (see also Kirschenbaum 2014). Some scholars have highlighted the need for a 'critical digital humanities' (Berry 2017, see also Liu 2012), recognising the radical potential of digital humanities to question power (Posner 2015) and digital capitalism (Grimshaw 2018). As I will explain in the next chapter, my methodology for this thesis is very much aligned with the idea of 'critical' digital humanities. Instead of using digital tools and technologies to analyse literature, art, or other creative texts, I turn the emotive and communicative power of language back around, and use literature and art as a way of analysing, making visible, and pushing back against the political power of digital tools and technologies.

2.5.2 (New) media

Although studies of media, information and communication reach back to classical times, they necessarily update with changes and advances in technology (Kittler 1999; Stiegler 1998; Manovich 2001; Gitelman 2006; Berry & Dieter 2015). As the title of my thesis suggests, 'algorithmic reproduction' of language, can be seen as the next step on from the 'mechanical reproduction' of language and art that Benjamin, Adorno and other members of the Frankfurt School were examining in their critiques of culture and politics in the 1930s (Adorno 2005). The title of the thesis is a deliberate play on Walter Benjamin's *The Work of Art in the Age of Mechanical Reproduction*, and as such uses the essay as a theoretical frame on which to hang my arguments about the digital and algorithmic processing of language, and about potential ways to mitigate the linguistic and political consequences of this new mediation by harnessing the power of creativity and art.

Examination, and indeed production of 'new' media today, is fairly variable in the extent

to which it engages with the politics and economics of digital technology. Just as with Digital Humanities projects, aspects of digital art and E-Literature can sometimes be seen as uncritical of the tools and methods they use, as many have recognised (see O'Gorman 2006 on E-Crit; also Hoy 2006; Liu 2012; Berry 2014; Fuller & Goffey 2009; Berry & Fagerjord 2017; Bartscherer & Coover 2011; Cayley 2017). More specifically in relation to language and literature, N. Katherine Hayles provides a helpful balance between the creative and the critical in Electronic Literature (2008), tempering possibilities of digitally enabled literature with critical literary theory and software politics. Hayles highlights the residue of old media structures in the mediation of new languages, both literal and computational (2010), as well as the foregrounding the relationship between database and narrative in digital texts, and the importance of the interface in the production of meaning (2012; see also Manovich 2001; Galloway 2012; Anderson & Pold 2011, 2018 on the agency of the interface). Flowing through all of Hayles' work are questions about the place and agency of the (post)human and the machine in digital media, texts, and information (2008, 2012; see also Braidotti on 'Posthuman Humanities' 2013). Fuller & Goffey's explorations into 'Evil Media' (2009, 2012), Finn Brunton's work on 'Spam' (2013), and Brunton & Nissenbaum's call for 'Obfuscation' of digital data (2015) also provide interesting studies into human and machine mediations of digitised language, as do Parikka and Sampson (eds.) in their exploration into 'Viruses, Porn, and Other Anomalies from the Dark Side of Digital Culture' (2009).

In terms of what actually constitutes art and literature in the digital era, Berry & Dieter's edited collection on 'Postdigital Aesthetics: Art, Computation and Design' (2015) contains several contributions that question the nature and enmeshings of pre/post digital media and art (Cramer 2015), in particular the chapters by Anderson & Pold, and Paul & Levy (2015), which theorise and question the artist James Bridle's concept of the 'New Aesthetic' (2011). Bridle's work suggests that new ways of seeing/reading and being seen/read in digital and physical spaces as mediated by digital technologies has produced a new kind of aesthetics (see also Contreras & Mirocha 2016). Trevor Paglan's artistic representations into what machines or algorithms see and read also speak to these same theories of post-digital aesthetics, bringing to the fore the asymmetries of vision, spatial control and the politics of digital technologies (2016; see also Zach Blas 2016). Artist and critic Hito Steyerl takes this critique of digital art a step further, presenting the provocation that with the political, societal and economic inequalities and iniquities of

the modern age, all art is now compromised by virtue of its immersion in proprietary technologies. According to Steyerl, art can never be entirely 'duty free' (2017), which is one of the problems my own artistic intervention into digital technology also grapples with (see also Vanderbeeken et al. 2011 and Mould 2018).

2.6 Conclusion

In the above sections I have attempted to review the literature relating to my thesis across a range of disciplines from arts and humanities, to the social sciences, digital geography, economics, linguistics, sociology and politics, and data sciences. I want now to bring back into focus why I started this project, which was the realisation that although language has always been a site of asymmetric political, geographic and cultural power, in an age of algorithmic reproduction, that power, and the inequalities and biases embedded within it, has been magnified and compounded by innovations in digital technology and connectivity and by the all-consuming influence of the digital economy. Exploring 'Language in the Age of Algorithmic Reproduction' means engaging with many current debates on (big) data and algorithms, but my thesis takes a distinctly different, and perhaps radical approach. My thesis posits language, or the representation of language as data (Fuller & Goffey 2009, 2012; Thornton 2017), and the algorithmic means by which it is processed, not only as issues pertinent to aesthetic concepts of literature and culture, but also as highly political issues, both by virtue of their embeddedness in the contemporary digital economy and because language is historically such a powerful tool of control of people and places. I argue that when language becomes algorithmically processed data, it becomes just as valuable, lucrative and powerful as the other digital data that mediate and increasingly govern, our everyday lives (Pasquale 2015; Lupton 2016; Moore 2017).

The thesis therefore places Google's digitisation and monetisation of language (primarily through AdWords and associated platforms) as a critical vulnerability to the stability and security of society and discourse, also providing a much needed link between more technical (cyber)security studies, and social science and humanities approaches. Furthermore, although critiques of data, algorithms, and specifically Google, have already focused on the economics and politics of digital society, by placing the emphasis on the human readable words (represented by data) that facilitate the flow of information through and around digital spaces, my thesis therefore questions the integrity of the very tools and techniques with which we are increasingly obliged, urged, or forced to communicate, debate, share

and publish our critiques in an age of ubiquitous digital technology.

Having foregrounded the importance of language in the political economy of digital society, my thesis also brings something new in bringing the critical theory framework of Walter Benjamin's 'Work of Art' essay into direct conversation with contemporary mediations of culture, specifically digital advertising and the search engine. I do not claim to be the only person to use Benjamin's theories on culture in a more contemporary context, and am by no means the first to see the potential in using Benjamin as a means of exploring and critiquing modern digital technology, whether in terms of critical theory (Berry 2014), geography (Gilge 2015; 2016a; Kingsbury & Jones, 2009), machine translation (Nabugodi 2014; Littau 1997), big data (Halpern 2015), or in comparison with other theorists of technology such as Donna Haraway (Franklin 2002). The thesis is also informed by a growing number of scholars who continue to make insights and inroads into the intricacies of language and knowledge when mediated through machines, including N. Katherine Hayles, Lisa Gitelman and Bernard Steigler. My own research adds to this burgeoning debate with a hopefully unique mixture of Benjamin's critical theory and work I have been doing around Google's advertising platforms and the commercial Search Engine Optimisation (SEO) industry. This slightly experimental splicing of Walter Benjamin and Google SEO tips facilitates an innovative and engaging discussion of the socio-technical, economic and political nature of search technology, and the effect it is having on language (whether creative, communicative, or a mixture of both) and the wider discourse, and is particularly timely in our current age of fake news (Graham 2017) and algorithmic bias (Noble 2018).

My thesis also makes unique contributions in that I put forward the theoretical concepts of 'Geographies of (con)text' and 'Subprime Language', which I detail in Chapters 4 and 8. In addition, the thesis also presents a radical artistic intervention in my {poem}.py project (see chapter 7), and an innovative new method for researching and visualising language in the age of algorithmic reproduction, which I will explain in the next section.

Chapter 3

APPROACHING A CRITIQUE OF LINGUISTIC CAPITALISM

3.1 Introduction

As I mentioned in the introduction, this thesis grew out of a Google search. When I conducted that search for 'wives and girlfriends sexist' in February 2014 I had no idea of the journey it would take me on. Not only did it become the subject of a PhD thesis exploring and exposing the (geo)politics and economics of the search engine, but in so doing it also allowed me to indulge in one of my first loves, both academically and personally, and that is language and literature. That it also led to a creative intervention and art project which I have presented and exhibited nationally and internationally seems completely unimaginable in hindsight. My {poem}.py project calculates the 'price' of poetry as the words in the poem flow through the Google search engine. Using data taken from Google AdWords and some code, the poems are in effect 'deconstructed' by the algorithmic process, before being reconstructed back into a readable format and printed out as receipts as a critique of linguistic capitalism. I also gather and analyse AdWords price data for longer term geo-temporal study of the changing value of language according to Google's algorithms. In this chapter I will attempt to explain how this all happened, exploring the methods, processes and pitfalls I encountered along the way. Part quantitative, part autoethnographic, and part practise-led, my methodologies are further complicated by their encounter with 'the digital', a prefix always already unbalanced by the weighty traces of human/machine and analogue/digital binaries, as well as being notoriously difficult to research. With all these complications, contradictions and subjectivities in mind, the overarching method I have adopted for the thesis is what am calling a *post-digital (auto)ethnography*, which has been

an effective way to develop my research theoretically and creatively.

3.2 Researching digital technologies

In recent years, computational concepts such algorithms, data, code and software, have become hot topics across the disciplines and are generating increasing amounts of literature not only in academia, but in the wider media and popular culture. As these topics touch on every aspect of modern life, it is hardly surprising that their critical study has been taken up across a range of academic disciplines, which although it has made a thorough review of relevant literature a difficult task, has also opened up many possibilities in terms of how we approach and research digital technology.

Attempting to research a complicated phenomenon such as Google is, as the literature review suggests, a rich, but complicated matter, not least because of the difficulties of separating the components of the algorithmic assemblage. As Gillespie notes, it is important to differentiate between the database and the algorithm (2014, see also Manovich 2001), a point also made clear by van Couvering (2010) in her research into search engine bias. Van Couvering separates the operation of a search engine into three categories, all of which can be biased: the index (by categorisation bias, seed lists and index cleaning), the algorithm (by its creators, ranking systems and links, relevance etc.) and display results (largely as result of advertising, bigger results, colour, font etc.). These distinctions (which are indeed not always all that distinct), are extremely important when it comes to finding out where the power and agency is situated in technologies such as search engines, and indeed, how that power can be made visible, regulated or challenged. As Nick Seaver observes, algorithms are 'embedded within complex social-technical assemblages made up of individuals, datasets, objects, apparatus, elements, protocols, standards, laws etc. that frame their development' (Seaver 2013, 10).

Rob Kitchin has called for critical study of algorithms from a broad range of complimentary disciplinary angles which bring into consideration their socio-technical assemblages and situations. According to Kitchin, technical approaches should be

> complemented by perspectives that consider: the discursive logic driving the propensity to translate practices and systems into computation; how the practices of coding algorithms are thoroughly social, cultural, political and eco-

> nomic in nature; and how algorithms perform diverse tasks, much of which raises political, economic and ethical concerns (Kitchin 2014: 7).

Bearing in mind the complicated assemblages of code, data, algorithms, and respecting their ontological (if not practical) independence, Kitchin's 'Thinking critically about and researching algorithms' (2017) is an extremely helpful base on which to think about all aspects of researching digital technology, and in particular the proprietary technology used by Google in its search and advertising technology, which is of course where my interests lie. The next section will build on Kitchin's framework in relation to my own research methods and practices.

3.3 What's in the (black) box?: researching proprietary technologies

As scholars and researchers have shown (Kitchin 2014, 2017; Pasquale 2015; Thatcher & Dalton 2016; boyd & Crawford 2012), attempting to critique or explain any system or platform that functions with the use of proprietary algorithms is difficult and problematic. Rob Kitchin identifies four significant methodological challenges to studying algorithms. Firstly, issues of *access*. As mentioned above, many algorithms, especially commercial ones like Google, are 'black-boxed', closely guarded trade secrets. This is both to protect the revenue stream from imitators or competitors and to limit the 'gaming' of the system, for example by link farms, Googlebombers[1], or aggressive SEO or keyword stuffing. Google in particular are famously secretive about their search algorithm(s). Issues of access could, of course, be at least partially overcome with the cooperation of the owner of the algorithm, although as Morozov notes, Google's algorithmic systems may be fundamentally unstudiable, as they have become technologies over which maybe even the Google coders and engineers no longer have control (Morozov 2012). Incidentally, this possibility of algorithms being *out of control* is another challenge to the critique and study of algorithms Kitchin identifies.

Access to Google data is also a methodological problem, making the researcher reliant on Application Planning Interfaces (APIs), commercial Search Engine Optimisation (SEO) material, and the data Google wish to release, although in my own research I did 'work around' this problem by gathering data manually, but it was time consuming and restric-

[1] http://www.searchenginepeople.com/blog/incredible-google-bombs.html. Accessed 19 October, 2016

tive. I tried several times in the course of my PhD to gain access to Google for information about its algorithms, but was ultimately unsuccessful. As well as attempting to contact Google's head of research Peter Norvig on several occasions and by several means, I also had no response from Google's European press liaison Peter Baron. Aware that several scholars had managed to gain access to Google, I also tried approaching them, but with no luck. Ultimately, I devised my own ways to gather data from (and about) Google, and to make conclusions about how their algorithms work without gaining any privileged (or indeed paid) access. In hindsight, I am glad it happened in this way, as an important part of the rationale behind my critique of Google and linguistic capitalism has been to make sure I make my stance from 'outside' the system, as I will elaborate on later.

3.3.1 Performativity and variation

The second issue identified by Kitchin is that algorithms are *heterogeneous and embedded*, meaning that even if access is gained to algorithms, they are too 'messy' to untangle having been merged, tweaked, corrected so many times over long periods of time and by different human and non-human actors and factors. Algorithms are also *ontogenetic and performative*. They are 'never fixed but constantly unfolding', and responsive and reactive to input. The workings of capital generating algorithmic systems such as Facebook and Google are constantly altered with tweaks, a/b testing, and dynamic advertising systems, as well as other personalisations such as geolocation (Zook & Graham 2007), which can also make them hard to research objectively. Not only can they, and what they produce, change from minute to minute, place to place, person to person in terms of various levels of personalisation such as device use, histories and geolocations, their workings can be altered or tweaked by Google for many reasons and purposes, for quality control, censorship or commercial gain, at any time, with no notification or justification. Richard Rogers notes how it used to be possible to gather a limited amount of data on Google search results through a dedicated API, but that several researchers who relied on that method had projects ruined or cut short when Google decided to withdraw the feature without warning. As I will explain in more detail in my final chapter describing my intervention, I came up against similar problems in gathering AdWords data, when Google changed details such as search volumes, how the keyword planner uses synonyms and the formatting of data outputs. These were significant obstacles in my research, and, much like a coder spends a significant amount of time 'de-bugging' a piece of code, all had to be overcome

by first recognising, diagnosing, and 'working around' the problems.

Following on from issues around their performativity, the results of algorithms cannot be anticipated; they can be unexpected, and affected by factors of which we are unaware. It is therefore almost impossible to conduct reproducible empirical research on algorithms and their data as we cannot see when the results are comparable. There is no base line, or constant, as Feuz, Fuller & Stalder showed with when they tried to engineer the imagined personalised search engine results of famous philosophers (2011), an intervention I discuss in greater depth in Chapter 7. Google's algorithmic systems may indeed be fundamentally unstudiable, as they have become technologies over which maybe even the Google coders and engineers no longer have control, as Morozov notes (2012). This possibility of algorithms being *out of control* is the fourth challenge to the critique and study of algorithms that Kitchin identifies. My own work recognises challenges such as these, but instead of letting it become a problem, I instead worked with the instabilities in the system, making the variabilities and performativities a central part of the critique. That my results are un-reproducible is therefore perhaps a triumph rather than a hindrance.

Despite these obstacles, however, as my project developed, it became clear that it was the very performativity of Google's algorithms that provided the greatest insights into their hidden logics, especially when thinking about the geographical variations in the output they produce. When I began collecting data from Google AdWords, in particular the suggested bid prices of words through the Keyword Planner, I was keen to find out how much these prices changed in different countries, and areas within countries. I thought this would provide clues as to how embedded within, or reflected they were of popular discourse, local and world events on geographical and temporal scales, a process which I have called 'Linguistic Geographies'. This is of course a method used to great effect by Mark Graham and others in work on the localised and often highly political bias within Wikipedia and Google Maps, for example (2010; 2014).

Some of my own options here might have been to access the AdWords data from different geographical areas, using either VPNs or outsourcing the data gathering via personal networks in different countries or through crowd sourced marketplaces such as Amazon Mechanical Turk. To have taken any of those options, however, would have added yet another layer of variables to the equation in terms of keeping the data set constant and free

from issues, such as individual personalisation via device or search history, not to mention the increased labour for myself or colleagues, and the complications of out-sourced labour (Irani 2015), had I decided to set up several VPNs or simply rely on the goodwill of international friends. What I ultimately decided to do was to conduct every Keyword Planner query through the same personal laptop, conducting the same three searches each day for a year, and, no matter where I was in the world, to always use the Royal Holloway VPN while gathering the data. The AdWords Keyword Planner allows users to target geographic markets by selecting different regions on the interface, thus giving potential advertisers an indication of how expensive their ad campaigns might be in a global market, or if they are advertising a regional service or product. While this method may not have been perfect (and indeed it would be almost impossible to check for skewing or to reproduce), it meant that at least I had some control over the standardisation of such volatile and dynamic data. From June 2016 to June 2017 I gathered the price data of 1200 words in 3 different geographic areas, the UK and US markets, and finally the whole of the AdWords market. In keeping with the literary and ludic themes of my project (explained in more detail below), I wanted to choose the words as culturally produced corpora, rather than using a dictionary or other more prescriptive and structured method, so I picked 5 poems, a song lyric and a speech which I thought threw a loose kind of net around a particular concept or event. Gathering data in this way allowed me not only to monitor the fluctuating prices of culturally pertinent words and phrases, for example 'prosperity' and 'austerity' in Billy Bragg's *Between the Wars*, or 'borders' and 'maps' in Bernadine Evaristo's *Heart of Exile* in the context of immigration and the refugee crisis, but also to calculate the 'price' of whole poems and texts over particular periods and regions, such as Alan Ginsberg's poem *America* in the build up and aftermath of the US Presidential election. I also wanted to monitor some words which did not occur 'naturally' in poetic form, such as 'Trump', 'Clinton', and 'algorithm', so I also produced a list of words I thought might be interesting to study from a geopolitical, as well as cultural angle. A full analysis of the data gathered in this way will hopefully be part of a further work, but some of the initial results of this quantitative study, will be elaborated on in Chapter 6.

3.4 Thinking critically about digital technology

Having identified these four methodological challenges, Rob Kitchin suggests six possible approaches to study and 'think critically' about algorithms. Some of Kitchin's suggestions

are highly relevant to my own methodological considerations, but others are not so. The first two approaches, *Examining pseudo-code/source code* and *Reflexively producing code* both involve a degree of technical knowledge far beyond my capabilities and the time constraints needed to learn them. One of the difficulties in researching digital technology is the vast disciplinary gap between 'technical' and 'non-technical' study of digital media, processes and software. A detailed knowledge of, and interest in, both the maths/computer science side and the social sciences/humanities is comparatively rare, although historically such a mix has produced the likes of Donna Haraway, N. Katherine Hayles, and Andrew Goffey, for example, and more recently a handful of geographers such as Rob Kitchin and Sam Kinsley, Till Straube, Andrew Dwyer, Louise Amoore have (or are developing) a more mixed background which has proved productive in recent scholarship. Although throughout the course of my PhD I have gained a significant (for me) amount of technical knowledge, the intervention part of my thesis in particular would not have been possible without collaboration and support from colleagues in the Information Security Group at Royal Holloway who wrote the code for the {poem}.py project. Indeed, the different perspectives and priorities of staff and student colleagues often gave valuable insights into wider issues of privacy, integrity and pedagogy, as I explain in chapter 7. In this way my lack of technical knowledge, brought into disciplinary and methodological tension by collaborations with more technically minded friends and colleagues, has actually been a productive method in itself.

3.4.1 Reverse engineering

The third of Kitchin's suggestions, *Reverse engineering* is more relevant, despite on the face of it also being a fairly technical method of research. The main problems with reverse engineering anything reflect the same old problems of access, and the black-boxed nature of proprietary algorithms. Reverse engineering for my project would involve examining the input and output of what makes up the algorithmic process to try to determine (or more likely guess) exactly what happens inside the hidden/secret/opaque part. Even if I could conduct satisfactory quantitative research using this method, it is still open to huge problems such as the instability of the data and how Google works. For example, as I have already mentioned, results will differ depending on the time of day, location, the device used, personalisation and potential A/B testing which can be happening unbeknownst to me at any time, and makes any conclusions fairly inconclusive. Kitchin does suggest, how-

ever that there is methodological potential in qualitative work around reverse engineering in trying to show how the input data is made 'algorithm ready', or in other words how the data is prepared for processing in terms of being easily structured, read, tagged or categorised (see also Gillespie 2014 here). Kitchin suggests interviewing marketers, media strategists, PR firms and following debates on forums. In line with this suggestion, part of my methodology has been attending a commercial Search Engine Optimisation (SEO) course to learn motives and techniques of advertisers and their role in the construction of online language both as adverts and as the digitised text and copy on which search algorithms feed (see Chapter 5). This method was hugely revealing in many ways, giving me a glimpse into the mindset and motives of an SEO expert, and the somewhat evangelical regard in which Google and its products appear to be held. In order to better understand the motives of AdWords users and the political economy of AdWords, I also engaged in email correspondence with different actors involved in bidding on the term 'dementia tax' in the run up to the 2017 UK General Election (see Chapter 6). Despite being limited by a lack of technical knowledge, and without the access and capabilities to be able to examine exactly what goes on within the black box of Google's algorithmic systems, I can still see what goes in a search engine, and what come out. I also have access to the same data as Google in terms of the corpus of searchable data on the Web from which the algorithm extracts results, and I can see how the embedded contexts of the words and their proximities to other words has an effect on what appears in search results and autopredictions. This is what I explore in terms of 'Geographies of (con)text' in Chapter 4.

Also, as I show in Chapter 8, part of my intervention and resistance into Google search and advertising has been to 'reverse engineer' the algorithms that try to boost keyword performance in AdWords. One of the functions of the keyword planner in AdWords is to suggest alternative keywords and phrases that might assist an advertiser in creating a successful advert and also then securing a successful bid. I re-use (or hack, maybe), this function to try to make visible the market logics and contexts that are applied to words as they flow through the search engine and therefore to imagine how language is 'read' by algorithms (see also Chapter 5).

3.4.2 The socio-technical assemblage

Kitchin's fourth suggestion, *Interviewing designers or conducting an ethnography of a coding team* is something I considered, but due to issues of access and technical restrictions, decided against. Another approach relevant to my project would have been an ethnographic or autoethnographic study with an academic setting such as a computational linguistics environment, or within the commercial SEO industry itself, perhaps similar to Jamie Woodcock's research in call centres (2017). Ethnographic methods of research in digital environments have been extremely insightful in recent scholarship, for example Taina Bucher's study on Facebook and the 'algorithmic imaginary', which used an ethnographic approach to try to gauge how people 'experience algorithms' in order to 'understand their social power' (2017: 30; see also Duggan 2017; Miles 2017; Bonner-Thompson 2017). Andrew Dwyer's authoethnographic account of his time training as a malware analyst (Dwyer 2017) is also a rich and innovative engagement with the complex ecologies within which software, algorithms, code, and of course humans interact. All these methods might perhaps respond to Rob Kitchin's fifth suggestion, *Unpacking the full socio-technical assemblage of algorithms*. Unpacking the relationship between the social and the technical, as well as Kitchin's final suggested method, *Examining how algorithms do work in the world*, is perhaps the most pertinent to how my project developed, and form a base for thinking about ideas of the post-digital and alternative ways of representing the complex assemblages that make up all digital technologies today.

The nature of my research has led to the development of methods and techniques that build and expand on Kitchin's suggestions, and are particularly relevant when thinking specifically about the proprietary nature of the software I am studying. There have been several approaches towards what we might call 'digital methods' in the last 20 years, many of which are concerned with the use of digital tools to learn more about a particular subject or discipline, be that (digital) humanities, (digital) geography, or geotagging and GIS techniques (Leszczynski 2018). Types of study in these areas tend not to be critically interested in the tools they use, but instead in what the tools can do in the furthering of knowledge and existing scholarship. As I mentioned in my literature review, there are many recent critical takes on the study of 'the digital' across the disciplines that position the integrity and agency of the tool of study, for example the computational means of (big) data analysis) at the foreground of analysis (see Thatcher & Dalton 2014), and that is very much where I position my own work.

3.5 Developing methods

As mentioned above, the methodology I ultimately adopted for this thesis revolved around extracting data from the Google AdWords keyword planner, using the data both as empirical analysis into the relationship between the value of words and the wider political and cultural sphere, and also as the ingredients for my artistic intervention. As I detail extensively throughout the thesis, artistic and academic critiques of Google have relied on interaction with various Google interfaces such as GMail (Cabell and Huff 2013), personalised search (Feuz et al. 2011), Google maps (Zook & Graham 2007), and also specifically AdWords (Bruno 2002, 2012; see also Berends et al. 2016; Ter Heerdt (in progress)), but as far as I am aware, AdWords data has not been the subject of enquiry into politics and society on a more empirical level (although touched upon by Kaplan 2014 and Jobin & Glassey 2014), a gap I address at various points throughout the thesis, but particularly in Chapter 6, *The Political Economy of Google AdWords*.

3.5.1 The reclamation of serendipity

Another key angle to my research has been the ongoing insights generated by the interaction with audiences and colleagues. Since I started the poem.py project in early 2016 I have used poems chosen by fellow panelists and chairs at workshops and conferences to feed through the keyword planner and turn into receipts. This might perhaps be classed as a practise-led method of research, as I very much followed the insights that were revealed in this way, analysing the process of words as they presented themselves to me. According to Scrivener and Chapman (2004):

> Practice-led research is characterised by a focus on issues, concerns and interests that are explored and manifested through the production of creative artefacts. This implies that, as an object of experience, the creative product is as important as any knowledge embodied in it (2004: 2-3).

I preferred this method over a systematic analysis of all the poems in an anthology (for example), as it sparked a very intense relationship with participants, and also avoided the risk of 'bias by canonisation' were I to rely on existing edited collections. I was asking for people's favourite poems, and the reasons for them giving them to me was sometimes very personal, and very emotional. To be able to use words which actually mattered to people on this level to demonstrate the violence of their algorithmic capture by the logics

of linguistic capitalism was a powerful tool. As I will explain more in chapter 8, one participant reacted so strongly to the monetisation of their favourite poem that they ripped it up in front of me. An interactive workshop I led at Trinity College Dublin in 2017 also proved a fruitful and fun method of learning from the AdWords data. I asked participants at the workshop to write their own poetry, trying to predict what the cheapest and most expensive love poems might be. The results were beautiful poetic artefacts that also gave fascinating insights into the value of language in literary and monetary ways (see chapter 8). Another rationale behind my choice of an experimental, random approach to the gathering of data in this way, was to subvert the restrictive logics of the anthology, the poetry website, and the filter bubbles and echo chambers of the internet, by relying purely on chance and serendipity. The poems I used came from the heart, from emotions, and half remembered schooldays (although as a source, the poetry on school curriculae also carries the risk of over-canonisation). They were, up until I scraped their facsimiles from the web, seemingly untainted by algorithmic and market logics (although it is their hidden para-textual functions this thesis strives to examine). This reclamation of 'digital serendipity' (Maccatrozzo 2012) in the face of automation and computation is something several scholars have called for. One of Fuller and Goffeys 'Evil Media' strategems is to 'make the accidental the essential', while Geert Lovink suggests that 'if we can no longer stumble into islands of reason through our enquiries, we may as well build them ourselves' (2012), which reflects my argument about the 'prisonhouse of digital language' that I put forward in Chapter 4.

Much criticism of personalised search results focuses on the negative effects of losing the element of random luck which might send you in an unexpected but fortuitous direction. This has been referred to as 'the death of serendipity' (Barnet 2009), and is reflected in the hierarchical tracts of knowledge which undermine the randomness of the rhizomatic structure of the Web (Hess 2008: 41), in echo chambers or filter bubbles (Pariser 2011), or in the rise of personalised, aggregated content (Barnet 2009). Many techniques and interventions have been developed to reclaim serendipity through 'ludic subversion', for example (Mahnke & Uprichard 2014), but as I will explain in the final chapter it is extremely difficult to avoid methods which aim to critique the iniquities and inequalities of digital technology without being co-opted back into the logics of digital capitalism. My project recognises this problem of almost unavoidable re-capitalisation, and instead makes that issue centre stage, demonstrating how it is almost impossible to keep our words free

from these influences by playing along with the system. John Cayley suggests that

> [a] major challenge in the programming era will be to develop linguistic aesthetic practices that intervene significantly and affectively in socio-ideological spaces thoroughly saturated with synthetic language that are largely controlled by commercial interests (2017).

It is indeed today almost impossible to conduct critical research into, or with the help of digital technology without somehow utilising, consciously or unconsciously, the very structures we seek to critique. My thesis acknowledges these methodical hurdles, and as Richard Rogers suggests in his book *Digital Methods*, seeks its insights by 'thinking along' rather than against digital technologies; in order to 'strive to follow the evolving methods of the medium' (2013: 1). By this method, thinking along with the algorithms, and being guided by a degree of serendipity, I have been able to assemble several 'collections' of poem-receipts, which have been rich sites for analysis.

3.5.2 Reading and writing linguistic data

The advent of big data prompted some in the humanities (Moretti 2013), and in the media (Anderson 2008) to believe that superior insights could be gained by computational 'distance reading' of large datasets, although this method of literary analysis is contested, and even met with some degree of scorn in some circles, for example in Stephen Marche's provocation against the emerging uses of digital humanities:

> Through these vast accumulations of ciphers the robots now endlessly scour for significance much the way cockroaches scour for nutrition in the enormous bat dung piles hiding in Bornean caves.... The algorithmic analysis of novels and of newspaper articles is necessarily at the limit of reductivism. The process of turning literature into data removes distinction itself. It removes taste. It removes all the refinement from criticism. It removes the history of the reception of works (Marche 2012: para.1).

Some scholars in the fields of New Media and E-Lit have also recognised the continued importance of 'close reading' of texts. Whether the words are on printed paper (which is of course not without its own politics of making), or are a mediated and dynamic amalgam of interfaces, software and commercial forces, close reading is still a critical means

of understanding a text (see Pressman 2014; Rogers 2013; Golumbia 2014; Liu 2012). Close reading of the poems re-rendered through {poem}.py has been an innovative, yet incisive, way to prise apart the politics of linguistic capitalism, and the reciprocal way language is shaped in a digital age, as well as the impact it has in shaping the wider discourse. The importance some scholars of E-Literature give to techniques learned in the study of pre-digital literature such as close reading and deconstruction (for example Hayles 2008, 2012; Pressman 2014; Landow 2006), is also an important frame for my own approach. Chapters 4 and 5 of my thesis use key figures of literary post-structuralism such as Roland Barthes, Jacques Derrida, Jean-Francois Lyotard as lenses through which to analyse digitised language. Digital Humanities scholars are also increasingly turning to pre-digital critics of politics and culture such as Walter Benjamin in order to contextualise modern media ecologies (Berry 2013, 2014; Halpern 2015; Parikka 2013), so my literary and cultural approach to the study of language mediated and structured through emerging technologies is not without a theoretical home.

As I mentioned earlier, as well as the artistic intervention, part of my method has also been to collect Google's suggested bid prices for several poems and texts collected daily and in different markets over a 12 month period. The poems for this part of my project were picked because I thought they created a kind of serendipitous 'word cloud' around particular topical events (this was 2016) which might provide insights into popular culture and politics as they fluctuated in the AdWords market. The data I gathered in this way has been analysed and visualised in both a traditional quantitative way (via Tableau software), as well as more creatively, but in any case, the way I have gathered and analysed these large datasets does not negate or preclude a 'closer' reading of the data therein, in fact, as Jessica Pressman suggests, in an age of 'digital modernism' the practice of close reading must be 'renovated' if we are to make sense of the modern world and the texts it produces (2014: 23).

My {poem}.py project falls very much into Richard Rogers' idea of what digital methods can and should achieve. For Rogers

> digital methods repurpose or build on top of the dominant devices of the medium, and in doing so make derivative works from the results, figuratively and literally. That is, the initial outputs may be the same or similar to those from online devices, but they are seen or rendered in new light, turning what

> was once familiar - a page of engine results, a list of tweets... a collection of comments, or a set of interests form a social networking profile - into indicators and findings (2017: 3).

Thus the poems given to me, so familiar to participants on a deep emotional level, are re-rendered in the new light of digital linguistic capitalism. When I was at school I was taught that the function of poetry was the 'making strange' of language so as to force tension, ambiguity, imagination, and interpretation. Once fed through the keyword planner, poems are rendered unfamiliar by their forcible restructuring into the form of a receipt. Reading downwards, like a list, words are aggregated so the poem struggles to maintain its form; their exchange values are inescapably revealed, visibly questioning their standing and position within the poem. In this way, my poem-receipts become an innovative means of data visualisation; not only challenging the decontextualising process of linguistic capitalism, but also subverting the quantitative restrictions of more traditional methods of visualisation such as graphs and charts. This can also perhaps be seen as a form of translation, but much like the computational way Google Translate works (see chapter 4), the rendering of the new poem bears the scars of its algorithmic reproduction. Knowing, or guessing, what is inside the 'black box' is not what is ultimately important here. Instead it is making visible the *effects* of those opaque systems. If using poetry as a kind of sacrificial offering to the algorithmic gods feels wrong (and it sometimes has to me), then it only goes to show the power that language and literature can have. Harnessing the innate power of language by using poetry in this manner is in fact a way of restoring its critical potential in the face of digitisation and monetisation, as well as a means to reclaim its narrative integrity from the the algorithmic logics of the linguistic marketplace.

3.5.3 Literature & fiction

As well as tool of epistemic, geographic and colonial power, all of which are of course magnified in a digitally mediated and connected age, language also has considerable power as culture in the form of fiction, poetry, and news media. My thesis harnesses this cultural power of the written word, turning it against the political power of language by using literature itself as a central tenet of my critique. As I mentioned in the introduction, language has always been a tool of control over people and places in a myriad of ways. Today more than ever, when analogies of Orwell's *Nineteen Eighty-Four* and Huxley's *Brave New World* fuel the discourse around authoritarian politics and fake news, literature proves it-

self to be a similarly powerful force. Just as I have used the power of poetry to explicate the effect of Google's advertising systems on language, my thesis also uses literary fiction as a critical lens.

Drawing on my own background in English Literature, which also explains the rationale for my application of literary criticism and post-structuralist theory to my work, Chapter 7 begins by exploring the political consequences of language control through the idea of 'Newspeak' presented by Orwell in *Nineteen Eighty-Four*, which is a striking example of the power of literature in critiquing and parodying power. Chapter 8 therefore also includes my own short piece of dystopic speculative fiction, which imagines a world in which the monetary value of words through Google's linguistic capitalism has made language so unstable and worthless that it becomes 'subprime'. While this short piece of creative writing is not a practice-based component as such, it does form an important part of my method, conveying speculative and conceptual ideas far more eloquently than a descriptive chapter or section would. In terms of practice 'led' research, the story becomes an artefact that 'translate[s] messages between concrete objects and abstract requirements' (Mkel 2007: 158). As Maarit Mkel writes, 'the knowledge and the skills of a practising artist or designer form a central part of the research process, and this has produced a new way of doing research, and a new 'way of knowing' (2007: 157).

My project uses a mixture of poetry and literature that existed in a pre-digital age, but also includes what might be called 'born digital' poetry written with Google's AdWords prices in mind, as well as a new piece of fiction through which I can explore and explain more conceptual ideas. There are of course many examples of 'born digital' creative writing that aim to explore the politics or philosophies of digital technology, however, by virtue of their medium, they run the risk of being co-opted into the economic logics of the system. By using pre-existing literature as a means of provocation and intervention, my thesis shows that incisive critique into digital language, and human and machine cognition does not need to be solely based on born-digital literature, and that experimenting with analogue (as in the receipts I use to show how AdWords works), rather than purely electronic representations is perhaps a more effective method. Approaches such as E-Literature and New Media Poetics typically restrict their outputs to electronic only formats in order to assume 'a synergy between human beings and intelligent machines' (Morris & Swiss 2006: 10). Indeed, as Johnston (2016) claims, digital poetry and other born digital texts can

usefully be found in, and created from, technologies such as algorithmic mediation and digital advertising, and the 'living' ecosystems, and dynamic representations of such media can be innovative expressions of critique, as long, of course, that they remain self-reflexive in their precarious digital ontology.

But I would argue that more informative and innovative synergies between man and machine can be created by questioning and bringing those assumptions into tension by mixing methods and media. Critical analysis of man/machine and analogue/digital binaries might be more insightfully executed by not dismissing pre-digital work in favour of born digital texts. Using existing literature as a tool of analysis, as Alan Turing did in his original 1950 paper 'Computing Machinery and Intelligence' by using a Shakespeare sonnet in his prototype Turing Test, is still an important intervention. Indeed, although I have collected and analysed AdWords data in all its temporal and geographic fluctuations, the culmination of my project, the framed poem as 'work of art', is deliberately no longer a 'living' artefact, in that my intervention is to violently freeze a particular version of the poem in time; thus chaining it to a context and a price. My own method of research and practice therefore troubles the assumption that 'living' interventions made possible by, and dependant on, digital technology are the only, or the most effective way of progressing the study of language in a digital age.

3.6 Towards a post-digital (auto)ethnography

The problematic materialities of so called 'virtual' technologies have been pointed out by geographers (Kinsley 2013; Ash et al. 2016) as well as other scholars in the field of digital technology. Christophe Bruno notes that

> After shifting from materiality towards immateriality at the end of the twentieth century, we have recently been experiencing opposite shifts. Whatever we may call these trends, 're-materialization', 'vintage media', 'neo-analog', 'post internet', 'post-digital', etc., they all deal with the inverse paths starting from the immaterial or conceptual and heading towards material or physical space...(2014: para.2).

Christophe Bruno has been an inspiration to my own work, even before it developed into an artistic project, so his observations about the 'material turn' in digital methods are

key to me here, and are also an excellent way to begin thinking about what this might mean in terms of what has been called the 'post digital', which in itself is a disputed and ambiguous term.

3.6.1 The post-digital

As Florian Kramer explains, the 'post' in 'post-digital' does not necessarily indicate a period 'after' the digital, or indeed a return to analogue methods and approaches (2015; see also Parikka et al. 2017; Berry & Dieter 2015), but rather refer to a period where 'the digital' is so embedded within society and culture that it loses its meaning as a standalone concept. As I mentioned in the literature review, the post-digital is often read alongside James Bridle's idea of the 'New Aesthetic', a term he coined to try to describe new ways of seeing and representation in a digital age. Bridle's work often aims to reveal the tensions between analogue and digital worlds by virtue of aestheticising the anomalies or glitches which present themselves in different types of technology, for example the 'rainbow plane' images captured by satellites which he says 'provide a glimpse into the ways machines see the world' (2014). Nathan Jones and other glitch artists, poets and scholars (see Menkman 2011 and many others) also suggest the exposure of the anomaly as the 'random' (Goriunova 2008) or not so 'random' (Parikka & Sampson 2009) moment which reveals the hidden machinations behind technology. 'The glitch-incision', writes Jones, 'does not claim to reveal a beyond-media here, but rather is a figure for a lapse in a system that reveals another system, a gap that opens transversal perspectives through one skin, into another' (2017: 237).

Although some examples of 'glitch' poetry are based directly on Google search and autocompletions (for example Riviere 2015; Nuotio 2012), and are not, as I will suggest in the last chapter, entirely unproblematic in that method, I do not situate my own poetic intervention in the same vein as either the New Aesthetic or Glitch Art/Poetry. By demonstrating the logics behind the monetisation of language by Google, it is not the oddities, glitches or limitations I am exposing, but rather *the grain* of the digital economy; the new standards, norms and structures of digitised language, against which, in an age when Google in some way mediates so much of the data that circulates online, it is becoming increasingly hard (or unappealing) to go, as I will argue further in Chapter 6. Linguistic capitalism has become a normalised, virtually unnoticed part of everyday life, and this is ultimately reflected in and represented by the output of a receipt in my

intervention. A symbol of capitalism we carry round with us in our pockets, screwed up in the bottom of our bags, and increasingly in our email trash folders, the receipt is so familiar, yet so expendable, and as I mentioned above, it is the 'making strange' of the familiar that can have such a profound and incisive effect as creative intervention. The receipt is an analogue residue of the materiality of exchange and the mundanity of existence, and indeed is a medium that has been used to the same effect by artists before, including Ceal Floyer's exhibition of a supermarket till receipt at the Tate Gallery (1998; see also Perry 2016). With all of this in mind, I find that my {poem}.py project falls best into Christophe Bruno's definition of the post-digital, which for Bruno is an intervention by which 'somebody creates an artwork that uses a digital tool; once the work is set up, remove the digital tool and observe what remains afterwards; if the work still holds, one may say it is post-digital' (2014: para.3).

3.7 Conclusion

While the way I have self-reflexively interacted with and learned from audiences and collaborators has been a journey of discovery that might be best described as autoethnographic, it does not seem to fit into some established examples of 'digital' or 'computational' autoethnographic approaches in which researchers harvest and analyse their own digitised data (see Brown 2019). In fact I have gone to some lengths to avoid the personalisation of my data and any kind of self-tracking or self-quantifying, as I describe in Chapters 7 and 8. Despite this, I am still inclined to think that my approach is autoethnographic in a more general way, at least as it is defined as 'research, writing, story, and method that connect the autobiographical and personal to the cultural, social, and political' (Ellis 2004: xix). As Ellis et al. conclude, 'when researchers do autoethnography, they retrospectively and selectively write about epiphanies that stem from, or are made possible by, being part of a culture and/or by possessing a particular cultural identity' (2011), and that is certainly how I see my own engagement with my research. My cultural situatedness as a woman with a background in the military was after all the catalyst for the whole project, and to varying extents continues to inform my research.

Alternatively, Nick Seaver's definition of an algorithmic ethnography also seems relevant to my approach, particularly his ideas about the creative performativity of algorithms. Seaver suggests that

> critical researchers might seek to enact algorithms ethnographically, seeing them as heterogeneous and diffuse sociotechnical systems, rather than rigidly constrained and procedural formulas. To do so, I suggest thinking of algorithms not *in* culture but *as* culture: part of broad patterns of meaning and practice that can be engaged with empirically (2017: 1).

Perhaps the difficulty in categorising my method might lie not only with my personal interaction with, or situatedness (or not) within the digital structures I am researching, but rather in its complicated relationship with whatever 'the digital' might be.

In his article 'Questioning Digital Ethnography in an era of ubiquitous computing' (2017), Duggan unpacks the differing approaches which currently operate under the umbrella of 'the digital', arguing that digital ethnography (as defined by Pink et al. 2016), along with other iterations he lists such as Robinson and Shultz's 'Cyberethnography' (2009), Dominguez et al.'s 'virtual ethnographies' (2007), and Kozinets 'net-nographies' (2017) are in fact not 'new' methods as such, but are just technogenetic developments in ethnographic study, and that the term digital (or its variations), is therefore not necessarily helpful. He also suggests that these 'digital' ethnographies run the risk of 'pull[ing] apart notions of technogenesis by privileging the digital in ethnographic approaches' (Duggan 2017: 4). Similarly to Seaver, Duggan instead advocates a 'nonmedia-centric' approach to 'digital' ethnographies which 'focuses on digital culture and practice rather than necessarily life online' (2017: 6), which is perhaps the closest description of my own method; the ubiquity of digital technology making the nominative distinctions between digital and non-digital negligible in much the same way as practitioners of the post-digital do. Bearing in mind Duggan's concerns about digital ethnographies, along with Christophe Bruno's definition of post-digital aesthetics, and the added intricacies of subjectivity and immersion in digital technology, I have weighed up the merits and drawbacks of this mix of approaches in relation to my own theory and practice, and have come to the conclusion that my own method might therefore best be described as a 'post-digital (auto)ethnography'.

Chapter 4

GEOGRAPHIES OF (CON)TEXT: LANGUAGE AND STRUCTURE IN A DIGITAL AGE

Il n'y a pas de hors-texte (Jacques Derrida 1976)

4.1 Introduction

As I mentioned in the first chapter, there are three themes running through this thesis (*context*, *value*, and *circulation*), which pull together my overall argument from several different angles. This chapter concentrates on the idea of context, or rather (con)*text*, as I want to emphasis that the way words are organised, structured and processed on the web can have consequences on other scales (local and global) and in other contexts, be they cultural, economic or geopolitical. As well as applying a post-structuralist critique to the geo-linguistic makeup of digitised language through the works of Lyotard (1984), Jameson (1974) and Derrida (1976, 1988), the chapter also uses Michel de Certeau's concept of tactics and strategies (1988) to imagine how and why different actors produce, mediate and traverse digital spaces not, as the 'landscapes of code' imagined by Thrift and French (2002), but as a 'landscape of words', or a *geography of (con)text* (Thornton 2017). This conceptualisation of the structural constraints (or otherwise) of language is a key part of the thesis as a whole, linking in with ideas of circulation (of data and capital), and the creation/representation of value which culminate in the creative interventions explored in

Chapters 7 and 8.

4.2 What happens when words become data?

Now that almost all language is in some way digitised, 'data-ised' and submitted to computation, the ownership, control and organisation of linguistic and other data can have a profound, yet not always intentional or indeed predictable, impact on social, cultural and political discourse, and on personal freedoms and securities. As well as significant power and profits for those who harvest and process it, the 'data-isation' of language brings the risk of serious collateral damage. By virtue of their reproducibility and enhanced means of movement and dissemination, words-as-data can have paratextual agencies and excesses beyond their linguistic function; the granular configurations and distortions of data can be instantly and exponentially magnified to a global geopolitical and discursively significant scale. But in contrast to this apparent post-modern untethering of language from its locational and referential functions, language that appears non-normative, or somehow unexpected, such as poetry or other creative variations of text, resists computation, and as a result has its movements through digital space restricted. It becomes suspect; penalised for its originality, or its unmarketability, as it fights its way through search engines, firewalls and new libraries of spam.

Once reduced to data, language loses the linear, or narrative order it might have had on the printed page or in speech. It is deconstructed in the process of digitisation, becoming part of the fluid pool of decontextualised data which flows through the spaces of the web, commoditised by the companies that increasingly control and mediate those spaces through advertising platforms and social networks. Indeed, language-as-data produces new contexts, in both a linguistic and physical sense, and creates new narratives. Processed on huge scales, digitised data has a wide geographical reach, so when it is reconstructed back into what we might call ordinary language (for example via search results, online translations or dictionaries), the manner in which that data has been stored, ordered and moved becomes extremely important.

Imagining context in this way facilitates the study of language in digital spaces from both material and theoretical angles. Firstly, it becomes necessary to think about how context as a kind of space might be produced, and the structures, biases and lacunae of the data

within it. Secondly, we can start to think about the movement of data though those spaces. How is language-as-data made retrievable, searchable, (in)visible, and exploitable by a range of different actors, from tech giants such as Google and Apple, to individual web users, advertisers, activists and politicians, and how their competing motives and agencies change the contextual makeup of that space. Not only is the ownership of language online a lucrative business (as we see from Google advertising platforms AdWords and AdSense), but as Lyotard (1984: 46-7) suggests, 'context control' in the production of knowledge is a powerful means for the legitimisation and therefore dominance of information. The physical and epistemic context of language reduced to data is perhaps at once liberating and restrictive. When reduced to data, words carry an entirely new set of values, motives, and meanings. They may become trapped within the logic of the digital economy, but can also become magnified to new scales and agencies by virtue of algorithmic systems of processing and interpreting. In conceptualising the de/re/constructions of linguistic data in this way, this chapter therefore considers both the physical and theoretical implications of the geographies of (con)text.

4.3 What do I mean by (con)text?

Much of the work in the nascent field of 'digital geographies' (Ash et al. 2016) has to date been focused on how digital technologies affect and produce geographies (for example Graham 2005; Zook & Graham 2007), or how space has become intrinsically linked with code (Kitchin & Dodge 2011). Yet more work has been done on algorithms and their work in the world (Kwan 2016; Dourish 2016; Neyland 2015; Gillespie 2014). The ideas in this chapter, and the concept of 'geographies of context' offer something novel by focusing on the data geographies and structures within and on which these algorithms work in the hope of opening up thought to the material and theoretical 'spaces of calculation' (Amoore 2016: 11) within which the data and algorithms interact, and in particular when the data is representing language. Even if such space is considered to be primarily topological, as is the case with vector space, the spatial relations within it are still important. As Till Straube contends 'the various articulations of relational systems making up the layers found within digital infrastructures should be taken seriously as spaces proper' (Straube 2016: 6). Although the context of words on pages have always been linked to their economic value (as a book, a telegraph, a newspaper article, for example), the manner in which data-ised words sit in relation to other words on web pages has taken on epistemic, economic, and even political possibilities unimaginable within the relative constraints of

print capitalism. As I will discuss further in subsequent chapters, new kinds of capitalism mediate the contextual spaces of data. Forms of digital, platform, semantic, or linguistic capitalism (Srnicek 2016; Fuchs & Mosco 2015; Feuz, Fuller & Stalder 2011; Kaplan 2014) have created new workers, owners, users, products and commodities, and their currencies are based predominantly on data. But even if we think of (con)text as an abstract space, the generation of value by and within that space lends it a 'very real social existence' (Merrifield 2015) wherein those who dominate have corresponding agency in 'real' space. As Shaw and Graham (2017) have shown, there is great power (and great money) in the control of informational flows by large technology companies such as Google. Of course, the 'text' of (con)text does not refer only to words, in fact its Latin root comes from *texere* (to weave). If we think how the weaving together of different threads determines different patterns, meanings and values in textiles, then the use of (con)text as a metaphor for the abstract space of linguistic data becomes even more relevant and politically significant. Often regarded as a forerunner of digital computer programming, and also the catalyst for new sets of labour relations, technologies such as the Jacquard loom not only revolutionised the weaving of silk and other materials, but also had wide-scale societal and political impacts similar to those of the digital revolution. Just as big data technologies and the patterns and narratives they produce - have become contentious issues today, the automation of the textile industry in nineteenth century Europe saw widespread resistance from workers facing unemployment and exploitation. In an era of AI and machine learning, data, and its (con)text, has become a similarly controversial raw material.

This chapter thus sets out to develop a critique of 'the digital' that concentrates on the physical constructs of language on the web. Rather than the 'landscapes of code' imagined by Thrift & French (2002: 309), instead I want to imagine a 'landscape of words', or a 'geography of (con)text'. A geography of (con)text can thus be imagined as the relational makeup of an ever-changing body of text, through and amongst which paths and trails of significance are intentionally or unintentionally traced, woven, diverted or created, producing dynamic contextual spaces, in and with which different actors, with differing motives, continuously engage. This chapter will begin by thinking about bodies of linguistic data as relational, yet often problematic and incomplete spaces. It will then discuss how competing actors contribute to the (in)visibility and (im)mobility of this data on the web; how it moves and is moved around, exchanged and manipulated. I find it useful here to imagine these movements as maneuvers, played out like de Certeau's every-

day tactics in attempts to (re)gain power over or subvert certain narratives, controlling conduits of meaning, or perhaps 'poaching' in lexical spaces (de Certeau 1988: 165-176). As de Certeau writes, 'on the blank page, an itinerant, progressive, and regulated practice a "walk"- composes the artifact of another "world" that is not received but rather made' (1988, 134-5), but 'walking' through digital space is not necessarily such a progressive or liberating linguistic experience. The mathematical and often binary logics that make and mediate the language within this space set paths from which it is hard or inadvisable to deviate; systemic processes which are only complicated and exacerbated by the commodification of linguistic data by technology/media companies such as Google, Facebook or Apple. It is therefore to the dominant structures of (con)text we turn first, before considering the other actors within these spaces, their tactics, motives and intentions, and then finally thinking about the unintentional or collateral consequences of maneuvers within this space which are made perhaps inevitable by the technologically magnified scale and reach of language-as-data.

4.4 Language spaces

It is estimated that only 5% of the data which exists on the web is indexed by commercial search engines. The machinations of the deep web and the dark web are beyond the scope of this chapter, but suffice it to say, there are still over 100,000,000 gigabytes of data for search algorithms to work on (Google n.d). This is what I refer to as the searchable database,[1] although given its dynamism and inherent hierarchies, it might better be described conceptually as a 'data environment' (Parisi 2017). The data that makes up the searchable database is, however, not necessarily democratically spread or representative of society, and neither is its visibility (or findability) upheld by any explicit or implicit norms of 'fairness'. Factors such as categorisation, language, tagging and indexing make some data easy to find, and some not, and the algorithmic systems managing the data are also always influenced by the motives, skills and potential prejudices of the flesh and blood programmer. The linguistic data which makes up the searchable database is also not an accurate reflection of all the analogue data produced in the world. As many scholars have noted (Crawford 2013; O'Neill 2017, Noble 2018; Graham 2014), there are significant weightings and gaps in the technological, linguistic, geographic or social agency of some

[1] Although not a database in the strict sense of the word, this phrase attempts to describe the quantity of indexed data on which search algorithms are able to act.

groups or areas which means that the 'searchable database' is a distinctly hierarchical and undemocratic dataset before commercial ranking algorithms have got anywhere near it.

Mark Graham has highlighted this content bias in the case of Wikipedia (2014), a platform which by nature of its user interaction and editing, is often held up as a success of the democratic Web. But an analysis of geotagging shows that the 'facts' presented by Wikipedia are constructed by and of the world's dominant languages and places. 'Knowledge created in the developed world appears to be growing at the expense of viewpoints coming from developing countries', writes Graham, with 84% of geotagged articles being about Europe and North America, and more articles about the Middle East in English than in Arabic languages. It is easy to see how this might also effect Google Search, especially as Google now use Wikipedia as a source for its 'intelligent' Knowledge Graph feature which pulls facts and images together into a handy box displayed at the to of your search results. Added to this, by virtue of its link structure and reputation, Wikipedia has been found to be on Page One of search results in 99% of searches (Search Engine Watch 2012). Therefore as Graham suggests, 'Wikipedia might not just be reflecting the world, but also reproducing new, uneven, geographies of information' (2014). The tendency for search results to favour collected information or analysis sites is, according to Reilly, also at the expense of original (potentially controversial or partisan) source material, which can lead to minority non-state groups or actors being made less visible, (Reilly 2008) and results being dominated by bland links and under the influence of groups with more commercial or political 'organizational clout' and 'off-web prominence' (Gerhart 2004). Thus, because of the ranking algorithms, the actual objects of the search become invisible, replaced by representations of themselves, whether an opinion, a encyclopedia type entry or an advert (Mager 2012), all the time the search engine is further abstracting the user from the object of their search. The corpus of linguistic data that exists in this digital space is therefore far from inclusive. What can be contributed to the searchable database is highly contingent and can exclude groups of people on multiple social-economic levels. Despite the oft-touted democratic and analytic qualities of big data, as Kate Crawford writes, although 'data are assumed to accurately reflect the social world there are significant gaps, with little or no signal coming from particular communities' (2013: para. 3). As I will detail more in relation to Walter Benjamin in the next chapter, those with better technology, better skills, better social standing, better connectivity and with access to better platforms, are the main authors of the searchable database, and it is this

privileged data that is used by the wealthiest and most powerful technology companies to search for answers to the world's questions. As organic search results, auto-completions and auto-suggestions are based on a mixture of previous queries and the data already existing on the web, the volume and structure (including any bias) of this data becomes crucially important in how these results are algorithmically extracted and reconstructed.

4.4.1 Searching

The surprising, confusing, and even distressing results sometimes produced by search engines are well documented, and can have a significant impact on the wider cultural and political discourse, reinforcing stereotypes (Noble 2018; UNWomen campaign 2013), marginalising the less powerful or wealthy, or spreading 'fake news' and even potentially influencing election results.[2] As I mentioned in the introduction, a personal example of this stemmed from a search for the phrase *wives and girlfriends sexist*, which I had typed in Google after hearing 'wives and girlfriends' being used problematically at a military security briefing. Rather than finding some cultural criticism about the portrayal of footballers' wives in the press and media, the word 'sexist' was substituted/mistaken by the algorithm for the word 'sexiest', and my search results consisted entirely of references to the top ranking *sexiest* and *hottest* WAGs within its index.

Apparent anomalies and glitches such as these have tended to be held up popularly either as proof of some kind of inherent sexism/racism within the system, or as sad but inevitable reflections of society. However, the algorithmic decisions that generate them are not based on any cultural or semantic knowledge as such, but on the mathematical logics of search technologies, on vectors, on marked up, decontextualised language, and on the analysis of big data by 'keyness and co-location', a method of corpus linguistics which in the field of Critical Data Analysis has been criticized because it 'omits essential qualities of actual language use' (Fairclough 2013: 20). In the pool of data in which the algorithm works, which I have called the 'searchable database', the word *sexiest* is statistically more likely to be linked to the phrase *wives and girlfriends* than the word *sexist* the associations perhaps compounded by the disproportionate volume of sexualised content or sensationalised news copy in the database, by algorithmic reproduction of erroneous typos or synonyms, and, as I will expand upon at length later, by the machinations of the linguistic marketplace.

[2] In the aftermath of the 2016 US presidential election, both Google and Facebook have been forced to acknowledge the potential impact of politically biased fake news stories or advertising click-bait hosted on their platforms (The Financial Times 2016).

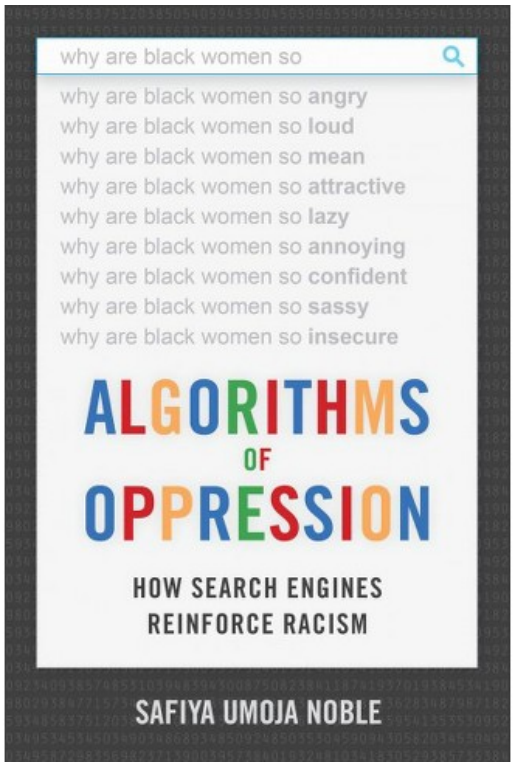


Figure 4.1: Safiya Noble's 'Algorithms of Oppression' (2018). Photo: Amazon.com

It is due to the lacunae and/or overloading of the searchable database with certain types or strings of information that the autocomplete/autocorrect function of Google search has so often been criticised for the controversial and stereotypical results it can produce. Far from simply reflecting the questions people have typed into the search bar, Google's autocomplete function, just like search, reacts not only to the search query, but is an amalgamated reflection of the entire corpus of data available to the algorithm. Words and phrases that appear next to each other more often within this digital corpus, will therefore be more likely to be linked in autocompletions, which is the reason why racist and sexist autocompletions are still fairly common, despite the number of times they have been called out in recent years (Baker & Potts 2013; Noble 2018) (see Figure 4.1 and Figure 4.2).

As I mentioned in the previous chapter, due to the black-boxed nature of the proprietary technology which operates commercial search engines such as Google, it is almost impossible to reverse engineer the search algorithm, as many scholars have pointed out (Pasquale 2015; Kitchin 2017). However, it is possible to understand why such apparent anomalies occur by studying the development of the information retrieval systems on which search technology is based, although some of the more scientific literature in this area can itself be fairly uncritical. In order to combat natural language processing (NLP) problems such

as synonymy and polysemy (Langville & Meyer 2011: 6), most search engines now use semantic searching based on Vector Space Analysis, an algebraic method of measuring the spaces between words in a large corpora data set. While Vector Space Models differ, their use in language corpora is based on the assumption that 'words that are closer together in the vector space are semantically closer in some sense' (Caliskan-islam et al. 2016). In this way, one recent study claims to have revealed the 'human' biases in semantics found in language corpora using a process of word embeddings. Caliskan-islam et al.'s 2016 report used various already acknowledged stereotypical word associations and compared them to the results of vector space analysis on data collected from the web, finding a correlation between human bias and prejudice and that found in the machine learnt results. The analysis revealed, for example, that 'flowers are significantly more pleasant than insects, and insects more unpleasant than flowers', musical instruments are more pleasant than weapons, and that European American names are more pleasant that African American names (2016: 3). Similar results were found in another recent report that used a corpus of text from Google News as its training set (Bolukbasi et al 2016). Bolukbasi et al. found that 'the closest word to the query BLACK MALE returns ASSAULTED while the response to WHITE MALE is ENTITLED TO' (2016: 41). Although these reports stress the importance of the contextual ordering of words in the production of meaning - even in non-absolute topological space - and show how a result like the 'wives and girlfriends' example can have occurred due to a reflection of loaded or incomplete linguistic corpora, they are not without their problems. There is an important distinction to be made between the way algorithmic models work and the data they work upon, although the distinction is often ignored. Lev Manovich writes that algorithms and data are 'two halves of the ontology of the world according to a computer' (Manovich 2013, 67), and as I have explained above, how the database is produced and constructed, and the relationships between the data within it is crucially important to the outcome of search results. Caliskan-islam et al.'s report notes that the statistical machine-learning model they used 'knows' the properties of flowers and insects 'with no direct experience of the world, and no representation of semantics other than the implicit metrics of words' co-occurrence statistics that it is trained on (Caliskan-islam et al. 2016: 3). As discussed above, given the incomplete, loaded and skewed nature of the data search engines can work on, co-occurrence alone is no basis for absolute knowledge. Their model does not 'know' the properties of flowers any more than the Google algorithm 'knows' women are sexy and not sexist; what it 'knows' is how the properties of flowers are represented in particular body of data, which is not

the same thing.

4.4.2 Translating

The systemic problems faced by search data and algorithmic processing can be seen also in online translation, where similar vector space analysis is deployed to produce the most accurate predictions of corresponding words in different languages. Although its accuracy has improved over time (and there are of course other online translation tools), Google Translate has always been infamous for its amusing and intriguing inaccuracies and anomalies, a good example of which is when in early 2016 the word 'Russia' began to be synonymised with 'Mordor' when translated from Ukrainian to Russian. Reports at the time suggested either that the translations had somehow been manipulated by Ukrainian hackers to ridicule Russia (How Stuff Works 2016), or were caused by a 'bug' or an 'automated error' in the algorithm (The Guardian 2016a). Google's own response to the incident simply reiterated what has long been known; that Google Translate does not actually translate from language to language in a semantic way, but relies on the corpus of data available on the web in a particular language to provide most likely predictive matches based on the frequency and proximity of words and phrases in vector space. Google put out statements explaining how its translation tool 'looks for patterns in hundreds of millions of documents to help decide the best translation for you... automatic translation is very difficult, as the meaning of words depends on the context in which they're used' (The Guardian 2016a). As the context for the translate algorithms is the same indexed dataset used by search algorithms, then there is always the possibility that 'translations' will pick up on cultural or regional anomalies, or stories most recently disseminated. As several sources pointed out, there are a significant number of references to Russia and Mordor in Ukrainian language popular literature, news and satire (Molchanov 2015; BBC News 2015; Bershidsky 2014). Bershidsky also notes that a translator of Tolkien's text suggested that Sauron was based on Josef Stalin - although this is a suggestion refuted by Tolkien himself - and mentions a rumour that in 2012 'a design company was planning to light up an enormous Eye of Sauron over a Moscow skyscraper'. Such is the strength of association between the two terms, that if the words 'Russia' and 'Mordor' become interchangeable in the limited quantity of Ukrainian text which Google Translate has to work on, then 'glitches' like this will happen. It isn't a 'bug' or an 'error', in fact the algorithms are doing exactly what they are programmed to do. Of course there is the possibility that the anomalies could have been deliberate, and similar ones could easily be replicated with the

Figure 4.2: UN Women campaign (2013)

appropriate amount of effort; it is after all relatively easy to 'play around' with Google, as some Google Map hackers (The Guardian 2018) and Googlebombers (Search Engine People 2010) have shown, although presumably it is becoming harder to do so anonymously as more and more Google services require logging in. Whether or not it was a concerted tactical attempt to manipulate and subvert the system, or whether the data available to the translate algorithms was already loaded with cultural references which might cause mistranslations, the (de)contextual ordering of words on a digital level in this case was quickly magnified to a geo-political and potentially dangerous diplomatic level.

A similar thing happened in 2013 when researchers contacted US intelligence officials after noticing that Google Translate had begun 'translating' the cod-Latin placeholder text Lorem Ipsum into 'apparently geopolitical and startlingly modern' English words and phrases such as China, NATO and Internet (Krebs 2014). Rather than uncovering some kind of spy communication network, however, this was simply an insight into how translation technology projects vector representations and 'word neighbors' from a source 'language space' (in this case Lorem Ipsum), to target 'language space' (English), thereby predicting the most likely translation (Mikolov et al. 2013). The incident therefore reveals far more about the types of website which use Lorem Ipsum than it does about covert operations. The web only holds a finite quantity of real Latin text and corresponding translations, just as it only holds a finite amount of Ukrainian text and translations, so when the Lorem Ipsum text is used as a placeholder which corresponds with the content of a website which can be viewed in different languages, in effect diluting the word-stock of that language, then the algorithm knows no better than to match the 'Latin' word to the English one. As Lorem Ipsum tends to be used as placeholding text on the multilingual websites of official organisations, governments, diplomatic pages or multinational businesses, so the 'translations' will reflect the language used on those sites.

4.4.3 Defining

Online dictionaries also suffer from algorithmic predictions of meaning based on large corpora data sets. Although Samuel Johnson was criticised for stamping his own personality on some of the definitions in his 1755 'Dictionary of the English Language'[3], the objectiv-

[3] A British Library webpage states how Johnson's definition of the word oats is 'very rude to the Scots. He defines the word as "A Grain, which in England is generally given to horses, but in Scotland supports the people"' (The British Library n.d.).

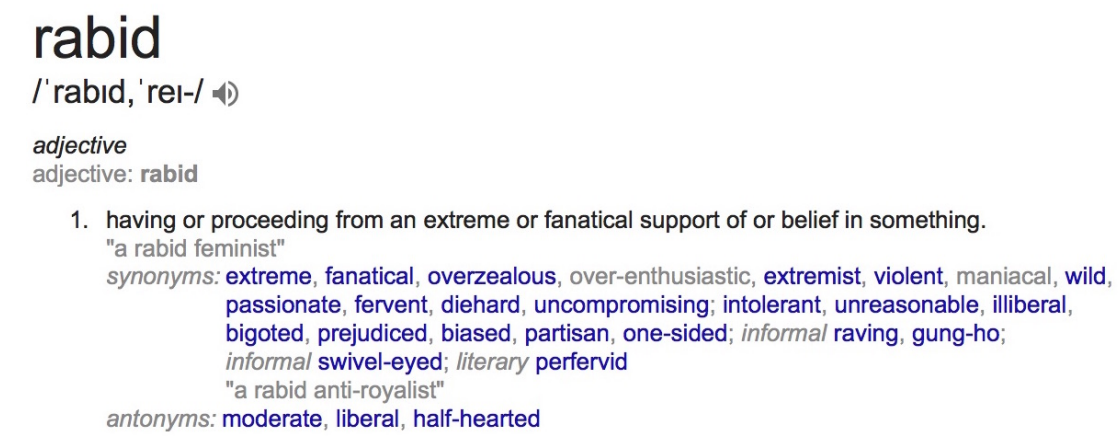


Figure 4.3: Oxford Dictionary example of *rabid*, October 17 2016. Screenshot: author's own.

ity of an equally popular but more modern, online dictionary was recently also called into question when it started to use the phrase *rabid feminist* as a contextual usage example for the word *rabid* (see Figure 4.3). Oxford Dictionaries Online, the digital version of the Oxford Dictionary of English which (amongst other things) supplies Apple products with their built-in British-English dictionary[4], responded to criticism of this apparently sexist word pairing by explaining that 'our example sentences come from real-world use and aren't definitions'. Their website explained how the phrases are drawn from 'a vast bank of more than 1.9 million example sentences (around 38 million words) of real English, extracted from the world's newspapers and magazines, academic journals, fiction, and blogs' (Oxford Living Dictionaries 2016).

Just like Google's search and translation tools, the Oxford Dictionary's examples are seemingly just a reflection of the linguistic data that exists online, which as I explained above, is not necessarily representative, democratic, or untainted by the technologies and commercial interests which mediate it. Unlike in Johnson's day, definitions based on digital data are calculated mathematically, so their potential inaccuracies or controversies are based on the accumulated mass of available data, rather than on the qualitative bias of one person or small group. The phrases generated as likely usage examples will therefore mirror the most likely pairings or orderings of already existing words based on the semantically irrelevant factors of frequency and proximity. It would therefore seem (rightly or wrongly) that the word 'feminist' is topologically closer to the word 'rabid' than any other word in the corpus of available linguistic data, including, we must assume, the word 'dog', which is an interesting linguistic development in itself.

[4]see BuzzFeed 2016 for a comprehensive 'storyfied' report of the controversy

What can be drawn from the examples of online search, translation and definitions, is that when language becomes data - due to the nature of data processing, language technologies and digital networking - it becomes volatile and invasive; its effects spreading more widely and more quickly, than the printed word. It can perpetuate stereotypes and inequalities, confirm biases, create diplomatic incidents, and with the increasing ubiquity and indispensability of the technologies that employ it has a perhaps unprecedented impact on web users. The Oxford Dictionary has what is presumably a very lucrative contract with Apple for the supply of words for its database of definitions and examples, all of which are made visible at the touch of a screen though the millions of Apple devices worldwide. As Attig (2016) responded to the incident on Twitter, with examples such as 'rabid feminist', the Oxford Dictionary is in effect 'beaming a sexist lexicography straight to students' iPads'. While there may be no tactical agenda for Apple beyond a convenient and cost-effective business agreement, the physical reach and volume of their products, coupled with their reliance on algorithmically mediated samples effectively puts an extraordinary and unintended epistemic power in the hands of another private company. As useful and ubiquitous as the tools provided by companies such as Apple and Google are, what we can see from this is that the way big tech companies use mathematical methods to extract search results, translations and definitions from large corpora can have serious side effects. But there are other actors in this space; activists, politicians, academics, advertisers, and even the hand of the market are all part of these contextual maneuvers.

4.5 Language games

Recent debates around privacy and encryption have highlighted the specific nature of the interests of the state, civil society and commercial companies in the visibility or invisibility of linguistic (and other) online data. Text can be hidden or redacted for private or political purposes - through the use of passwords, paywalls or even through site architecture such as sitemaps and robot.txt. Such was the case, for example, in 2003 when it was reported that the White House had hidden webpages which referenced Iraq from being indexed by the webcrawlers (Elmer 2009). Privacy activists have suggested ways to deliberately obscure personal data. Brunton and Nissenbaum's users' guide to 'Obfuscation' calls for the 'deliberate use of ambiguous, confusing, or misleading information to interfere with surveillance and data collection' (2015: 1). They see obfuscation as a tactic, or a 'weapon of the weak' (2015: 55) to be deployed from within the surveillance society using 'every-

day' language as a subversive tactic against the powerful (de Certeau 1988, 30). Criminal activities such as phishing attacks often vary a web address or keyword by one letter in order to exploit the inadvertent user who may have misspelled a search query, or clicked on the wrong link, thus directing them to a fake site. 'In the lexicon of the World Wide Web' write Matt Fuller and Andrew Goffey, 'such typos are the homonyms and synonyms, the words that allow a user to pass over into another dimension of reference' (2009: 153). Companies or individuals can guard the portals and spaces of sites by buying up domain names or keywords. Spam emails often rely on algorithmically generated jumbles of words that need to make just enough sense, not to fool humans (at least in the first instance), but rather, to fool firewalls. Language in online spaces can also be disguised in order to deceive commercial copyright or plagiarism software. Evidence of the resulting linguistic mutations this embeds in the database is clearly visible. Tarleton Gillespie uses the example of the misspelling of 'Britny Speers' so that people searching for illegal downloads could find recordings, 'but the record industry software could not' (Gillespie 2014: 184). On the flipside, however, it is sometimes impossible to search for a specific word without being autocorrected, which may seem to give the platform doing the 'correcting' a more 'strategic' method of 'producing, tabulating and imposing'(de Certeau 1988: 30) power within the contextual space of the search, yet, as I will suggest, power balances within geographies of (con)text are constantly being challenged by actors on all levels, and it is hard to imagine an unconnected strategic vantage point within that context.

Data-ised language in this sense becomes either a tool, or means of access or restriction; its linguistic function as a means of human communication becoming secondary to its use and exchange value for other purposes. But as well as being used as a tactical tool, language can in itself have its movements restricted as it passes through digital spaces. Non-normative, creative language can be actively 'criminalised', not only by Google, but also by firewalls or anti-virus software which blocks text that does not conform to whatever 'natural language' databases or templates it has been taught. The condensed, carefully worded nature of poetry can sometimes be construed by algorithms as key-word stuffing; a tactical manipulation of the search rankings which is frowned upon by Google, and which can lead poetry websites to be 'buried' in the search rankings (Hoy 2006). Artist Sophie Mayer (2013) noticed how the poetry magazine *Poems in the Waiting Room* had to issue advice to potential contributors, asking them to send their submissions in the body of an email rather than in an attachment, as the attached poetry might be mistaken for a virus

by the spam filtering software. All this starts to beg the question just how communicable is poetic or creative language in a digital age? Is it less communicable than a virus? How restricted or restrictive - is language when it becomes data? And further to this, what is the function (and fate) of language if it is written for, or increasingly by in (the case of algorithmic content generating) machines or algorithms for the sole consumption of other machines or algorithms? As the method for this type of exchange is all acted out mathematically, then the onward effects on non-mathematic language becomes little more than collateral damage; the leftovers of a linguistic power-grab perhaps, or a post-modern language game.

4.6 Language markets

Poetry is of course also the lens through which I conduct my intervention into another major player in the landscape and mobility of online language the advertising industry. As I will explain in subsequent chapters, in a system Frederic Kaplan has called 'linguistic capitalism' (2014), the language which flows through Google's search and advertising platforms becomes infused with the logics and values of what might be termed a linguistic marketplace; its grammars, patterns and frequencies formed not through generations of linguistic happenstance, but as advertisers compete for lucrative keywords, and (as we have seen earlier) as machines learn and magnify algorithmic glitches. Words have gained a currency detached from, for example, their narrative function. Not only do their prices fluctuate according to apparently neoliberal free-market competition, but there are also many distortions to this market due to centralised interference and control from Google in the form of regulations, censorship and other linked applications, as I explore in chapter 6. The data-isation and monetisation of language in this process has a subsequent effect on the density and frequency of certain words within the searchable database, privileging both paid-for words and phrases, and also organically optimised language in the search results. To paraphrase N. Katherine Hayles, language does not emerge unscathed from its encounter with code (Hayles 2005: 39). Indeed, data-ised language tells other, paratextual stories as it moves and is moved through digital space. It gains economic value as it draws eyes to adverts and negotiates spaces controlled by 'cognitive rent' (Pasquinelli 2009: 157) or 'lexical squatting' (Fuller & Goffey 2009: 153), and as we have seen, it becomes weighted with the residue of the wider pool of decontextualised, yet hierarchised data, or 'tracts of knowledge' (Hess 2008: 35). This residue turns tools such as translate and search into 'authoring device(s)' (Rogers 2009: 176), and turns (machine) translations into new

linguistic art forms (Nabugodi 2016). But the stories told and the paths forged in this manner are always products of the structures from which they originate and through which they flow. Even the vectors which quantify data-ised language do not just point to, or connect things, but are carriers of meaning; they add to the story. We too have become vectors in the digital assemblage, or as Mackenzie Wark puts it:

> Both the flâneur and the facebooker are voluntary wanderers through the signage of commodified life, taking news of the latest marvels to their friends and acquaintances (2015: para.40).

But Wark's flâneur is very much the Benjaminian flâneur of the arcades, not the modern day dérive-drifter whose goal is to subvert, not admire, the spectacle. It is surely impossible to drift through digital space today without picking up the 'signage of commodified life' in the form of 'likes', 'shares' and click-bait, whether voluntary or not. As de Certeau identified, with the ubiquity of the mass media 'instead of an increasing nomadism, we thus find a "reduction" and a "confinement": consumption, organised by this expansionist grid takes on the appearance of something done by sheep progressively immobilized and "handled" as a result of the growing mobility of the media as they conquer space' (de Certeau 1988: 165). There is of course much more to be said about the changing values of data-ised language and the circulation of linguistic capital, which I will expand upon in Chapters 5-7.

4.7 Conclusion: The prison house of digitised language

To conclude this chapter, it is apparent that language is an integral element of digital space, and one way to conceptualise this is through thinking about the composition of data-ised language as '(con)text'. Just as in other spaces, there are social and political factors at play, and power is exerted and subverted by competing actors. Following de Certeau, the manipulation and movement of language within this produced space can therefore be read as kinds of tactical maneuvers which 'make use of the cracks that particular conjunctions open in the surveillance of the proprietary powers' (1988: 37). We can see in the decontextualised database, or, indeed, in such a simple action such as cutting and pasting what Derrida called the 'citational graft' of a sign; the capacity to put quotation marks around any word or phrase and move it elsewhere, thus 'break(ing) with every given context, engendering an infinity of new contexts in a manner which is absolutely illimitable' (1988, 12). As Ming Lim suggests, in the mechanics of the search engine industry we can

see that there really is nothing outside of the (con)text:

> In this space, signs truly refer only to other signs and all pretense at 'presence' or 'essence' or 'authenticity' is no longer necessary. Marketers who use SEM [Search Engine Management] are now subject to the peculiar laws of an architecture which have barely begun to be theorised (2008: 10).

But Derrida's illimitable contexts were imagined before the structures and restrictions of modern digitised language were fully realised. In this context, words are not 'free' to be eternally deferred, but become exponentially re-infused with the residue of the dominant structures, market forces, biases and stereotypes that make up the corpus of the searchable database. Inextricably linked with code (Hayles 2005, 16), their meaning has been 'arrested' (de Certeau 1988: 165)[5]; chained to the past by algorithmic association and given a record it is impossible to expunge. Followed around by 'semantic escorts' (Metahaven 2009: 189) which have the power to decide who or what can and cannot be sexist, (un)pleasant, rabid or racist, they have in effect been 'reconstructed'; squeezed through binary systems which force an either/or logic on words even as they squirm to get away. And neither is digitised language free from an organising metanarrative. Words-as-data, already (re)constructed by virtue of the way they are stored and transmitted are also monetised within the system of 'linguistic capitalism' (Kaplan 2014). As Franco Berardi states, 'today the economy is the universal grammar traversing the different levels of human activity' (Berardi 2012: 158).

But what of de Certeau's tactics? Who benefits from the structural and economic restrictions of language-as-data? If, as Gunnar Olsson believes 'it is in the interest of social cohesion to impoverish language' (1978: 110), then the power wielded by the big technology companies such as Google, Apple and Facebook becomes even more terrifying, as I will detail in the provocations of Chapters 7 and 8. In terms of '(con)text' that is the geo-linguistic space of the web - it might perhaps be easy to think of these big technology companies as strategic actors in de Certeau's sense of an institution which has established itself in 'a *place* that can be delimited as its own and serve as the base from which relations with an *exteriority* composed of targets or threats (customers or competitors) can be managed' (de Certeau 1988: 36). But it is more complicated than that. As the content and the currency of the web is made up of interactive, interconnected, data-ised

[5] 'To arrest the meaning of words once and for all, that is what the Terror wants', quoted in de Certeau 1988: 165

language, none of the actors (including us as users) involved in the kinds of (con)textual maneuvers I have described can ever act from a completely strategic (air gapped) vantage point. While there are different levels of access to other resources that shape digital infrastructures, such as source code, what we might call 'natural' language - which has been deconstructed and decontextualised into data - has a liquid quality that permeates through the spaces of the web, in some way making users of us all. It may seem that with their control over the mediation of language, the big technology companies have the upper hand, but they have, to a certain extent, created something over which they no longer have complete strategic control. What does seem to be the case, however, is that like the spaces of consumerism described by De Certeau, language-as-data always seems to have an ulterior, paratextual motive and, perhaps because of the mathematised logic of code, can never just be language. So just as we are trapped within the structures of a digitalised discourse, so are we too involved to be able to look in from the outside from a strategic vantage point. 'There is' de Certeau says 'no longer an elsewhere' (1988: 40). We are truly coopted; practicing out everyday lives within what Jameson might have called the 'prison house of digital language' (1974). Yet despite its structural constraints, the manner in which linguistic data is disseminated online can, as I have shown, have far reaching effects untethered to and unpredicted by its topological environment or physical location. In this way, it could be concluded that digitised language falls somewhere in the middle of a structuralist/post-structuralist critique; being at the same time both free from and constrained by the geographies of (con)text.

This chapter has set out my ideas on how the granular digital and algorithmic structuring of linguistic data can have significant consequences on a magnified global, societal, and political scale. I conclude that the way language is stored, processed and manipulated online reconstructs stereotypes and inequalities into the wiser discourse. I introduced the concept of 'geographies of (con)text', as a way to explore the different actors who operate within this geo-linguistic space, and to analyse their motives. The chapter has also introduced some of the work I present in subsequent chapters about the power and politics of language in the age of algorithmic reproduction (Chapters 5-6), the value, circulation and liquidity of digitised language (Chapters 6-9), and my own tactical intervention which, like Jameson in his structural analysis of Russian formalism (1974), uses poetry as means to critique the structural 'prison-house' of language in an age of linguistic capitalism (Chapter 7). Before that, however, the next chapter sets out the framework of the thesis

as a whole, 'The Work of Art in the Age of Mechanical Reproduction', an essay written by Walter Benjamin in the 1930s about the impact of technology on culture and politics, but which I argue provides an insightful lens through which to study *language in the age of algorithmic reproduction* today.

Part II

PART TWO: PRACTICE / ENGAGEMENT

Chapter 5

LANGUAGE IN THE AGE OF ALGORITHMIC REPRODUCTION

Using Walter Benjamin's 1936 essay 'The Work of Art in the Age of Mechanical Reproduction' (1999) as a theoretical springboard, this chapter continues my investigation into what happens to language when reproduced algorithmically through search engines such as Google, and sets the scene for the unfolding of the Benjaminian frame in the following chapters. Reflecting both the political and economic frame through which Benjamin examined the work of art, I argue that the processing of language through the search engine is similarly based on the distancing and decontextualisation of language from its 'source', or from the intent of the user; an exploitation of media which is politically no less important today in an age of fake news (Graham 2017) and algorithmic bias (Noble 2018). Using the Benjamin text as a focus, the chapter also discusses techniques learned from participation in a search engine optimisation (SEO) training course, which reveal how much of the language we see online has been structured for the consumption of algorithms. Setting up ideas for a wider critique about the power and political economy of language (Chapter 6), and my own intervention (Chapter 7), I then suggest that the power and politics lurking behind these algorithmic processes is facilitated by the aesthetics of the Google empire, and that, following Benjamin, the most effective way to counter the aestheticisation of politics is to make art political. But in re-politicising digital artistic critique we must take care not to re-aestheticise the politics of technology, paying close attention to the means of its production.

5.1 Introduction

The title of this chapter, and indeed the whole thesis, is a fairly obvious and unapologetic appropriation of Walter Benjamin's 1936 essay 'The Work of Art in the Age of Mechanical Reproduction' (1999), however I have used it less of a template than a springboard from which to explore ideas about reproduction of a distinctly more modern kind, that of words and language by algorithms, and more specifically through search engines such as Google and its associated applications.

When researching language in a digital age, there are questions frequently asked which I think must be tackled straight away, one of which is: how is the reproduction of language - or indeed art - different in an age of digital as opposed to mechanical, analogue technology? And following on from that, is language as an evolving part of human ontology not inseparable from the progress of technology (Stiegler 1998); 'textural structures being... always already technological' (Barnet 2003), and is there not therefore some kind of conceit in decrying the corruption of 'natural' language and meaning at the 'hands' of digital technology? Surely post-structuralism tells us that language is never fixed anyway, and indeed, why is language important to humans at all?

The proliferation of text and information afforded by the printing press and new methods of dissemination was perhaps just as much of an information revolution as the start of mass access to the Web. Indeed, Robert Darnton (2009: 33) argues that 'every age was an age of information, each in its own way, and that information has always been unstable'. And if computerisation of language is seen as a threat to the integrity of the printed text, perhaps through inadequate machine reading such as Optical Character Recognition (OCR), or machine translation, then likewise copies of all kinds of text have always deteriorated, varied, or been open to corruption through translation, plagiarism and bootleggers selling pirate copies (Bridle 2015). To a certain extent these are valid points, but James Bridle's defence of digitised language highlights important differences between different forms of reproduction. Yes, language generated by printing presses had a monetary value in that there has always been a market for cheap copies, and likewise much of the text created for new readerships of newspapers was controlled by monetary concerns (in terms of distribution and physical space on the printed page), as were new communications methods such as the telegram. But the way language is valued online is not the same as that of the column inch, or the 'price-per-word' method of the telegram, where words are valued more

by the physical space or effort they take up, rather than any value inherent in the words as narrative, in a literary or aesthetic sense, or purely as a function of human communication.

On the Web and in the telegram, language is indeed moulded by either the proliferation or brevity needed to make or save money for both consumers and producers, but what I argue throughout this thesis is that the algorithmic reproduction of language is infused with a different kind of market value, whereby words themselves can be elevated beyond or outside both their material manifestation and their semantic function in order to create capital. This is language being 'othered' by what Katherine Hayles calls the ontology of the database; its narrative order, and its creative and communicative functions (to humans) compromised by classification and enumeration (Hayles 2012: 179). Further to this, I want to suggest that the algorithmic reproduction of language by reduction and reconstruction through the enumerative and economic structures of search engine technology has consequences for the integrity and evolution of language and discourse which reach far beyond the relative stability of a printing press cliché, which we might perhaps see as the forerunner of the algorithmically reproduced 'glitch'. Print capitalism may have given 'a new fixity to language' (Anderson 2006: 46) but 'linguistic capitalism' (Kaplan 2014) and 'semantic capitalism' (Feuz et al. 2011) is far less stable, and it is in this flux of money and words that new forms of power and influence flourish.

Apart from the appropriation of Benjamin, part of the methodology used in this chapter has included researching the commercial Search Engine Optimisation (SEO) industry by attending a training course, which has given me an important insight into how search rankings (specifically Google in this case) work, but more importantly, made clear not only the role of the market in those rankings, but of an industry aimed at adapting or manipulating language in order to succeed in that market. In this chapter I will therefore be using the basic structure of an SEO training guide to organise the chapter through three main themes, the first of which deals with the 'organic' optimisation of web content through links, keywords and tagging which leads to a discussion of language and linguistics, raising questions of authorship, readership, content and context. The second section will focus on how language is commodified as it circulates through digital spaces such as commercial search engines and email, and the inherent linking of text to market forces, while the third section explores issues of power, monopoly and influence which the control of the means of processing, producing and reproducing language entails. These factors

the linguistic, the economic, and the inevitable political debate in which they synthesise were central to the atmosphere of cultural/ technological debate which informs Benjamin's 1936 essay; their historical combination, according to Benedict Anderson, having enough discursive power to affect (and effect) national consciousness. In exploring the 'explosive interaction between a system of production and productive relations (capitalism), a technology of communications (print), and the fatality of human linguistic diversity', Anderson was of course writing about the development of print-capitalism since the 14th Century, but I argue that the collision of language, capital and technology are equally explosive and discursively powerful today. Bringing Benjamin's essay to its natural conclusion, the chapter concludes by suggesting that the power and politics lurking behind these algorithmic processes (Introna & Nissenbaum 2000), is facilitated by the aesthetics of the Google empire, and that, following Benjamin, the only way to counter the aestheticisation of politics is to turn this tool of power back on itself, to make art political; reclaiming language (be it literary, artistic, political or prosaic) from the algorithmic marketplace in order to re-politicise artistic critique, which is the focus of my intervention detailed in Chapter 7.

5.2 Why Benjamin?

> *That which withers in the age of mechanical reproduction is the aura of the work of art (Benjamin 1999: 215).*

I cannot claim to be the only person to use Benjamin's theories on culture in a more contemporary context, and am by no means the first to see the potential in using Benjamin as a means of exploring and critiquing modern digital technology, whether in terms of critical theory (Berry 2014), geography (Gilge 2015, 2016a; Kingsbury & Jones, 2009), machine translation (Nabugodi 2014; Littau 1997), big data (Halpern 2015), or in comparison with other theorists of technology such as Donna Haraway (Franklin 2002). The argument I put forward in this chapter is intended to add to this burgeoning debate with a hopefully unique mixture of Benjamin's critical theory and work I have been doing around Google and the highly lucrative and competitive commercial Search Engine Optimisation industry (SEO). I hope that this slightly experimental splicing of Walter Benjamin and Google SEO tips facilitates an engaging discussion of the socio-technical, economic and political nature of search technology, and the effect it is having on language (whether creative, communicative, or a mixture of both) and the wider discourse, particularly in our current age of fake news (Graham 2017) and algorithmic bias (Noble 2018).

Benjamin's essay recommends itself particularly to my examination of digital technologies in that his critique of mechanically reproduced art necessarily, and implicitly, includes mechanically produced language. It is perhaps the idea of creative, or literary language - in a distinctly humanistic way - being somehow corrupted or exploited by algorithmic reproduction, that drives my own critique. However, when reduced to data for transmission or storage, all digitised language (whether artistic or otherwise), is reduced to what we might think of as a 'binary soup'. The consequences for artistic language in a digital age are thus necessarily mediated by the processes that govern and mediate all other types of digitised language, and indeed any other piece of data. In an age when data has become such a lucrative and politicised commodity, the collateral effects on the language it represents and reproduces can have serious repercussions, as I cover in Chapters 6 and 7.

In recent years there has been something of a revival in Benjamin studies, with 2015 marking the 75th anniversary of his death. Mackenzie Wark's (2015) Benjamedia essay, written as part of the anniversary events, identifies distinct parallels between the political and historical situations of today and when Benjamin originally published, asking not only if Benjamin still speaks to us today, but 'what he can do for us' in the future. Benjamin was fascinated, and perhaps conflicted by the tensions between past, present and future (Kingsbury & Jones 2009). The sadness displayed for the 'withering' of the aura in the Work of Art essay, and a love of artefacts, history and high art which was perhaps at odds with his supposed revolutionary credentials (Werckmeister 1996), might suggest a less than progressive philosophy, but Benjamin's apparent nostalgia does not exclude his recognition of technology as a potential force for good. As Tim Beasley-Murray (2012: 781) observes, even in the face of the mechanical violence of the war machine, he 'does not resort to nostalgic technophobia', instead seeing a perhaps unrealised emancipatory potential in the reproducibility of media such as film and photography.

Following Benjamin's lead, my critical stance on search engines, and Google in particular, does not preclude any notion of the benefits and creative possibilities of digital technology. Perhaps answering David Beer's (2009) call for a critical counteraction to the 'Web 2.0 bandwagon', my work is instead a call for continued caution towards the tools we increasingly rely on to communicate and express ourselves in everyday life, making the power and effect they have visible through critiques and interventions such as I describe in this

thesis. As Benjamin himself wrote, 'technological revolution... are the fracture points in artistic development where political positions, exposed bit by bit, come to the surface' (2008: 329), and these fracture points are perhaps just as evident today. In the age of digital technology, where advertising and information can be said to have taken over from heavy industry as a major accumulator of wealth and influence, proprietary algorithms have become the facilitators of the movement of raw materials and commodities (in this case raw data or text instead of coal or steel) for capital (re)production. They are in effect 'do[ing] for information systems what canals did for mercantilism' (Poon 2013), and making their owners extremely rich and powerful in the process.

The ownership of these new means of production has therefore never been more important, and nor has the ownership of the commodities themselves. Algorithmically reproduced words have become the vital textual currency of the Web and the lifeblood of 'linguistic capitalism' (Kaplan 2014). But as I showed in the previous chapter, as words are digitised, they become little more than constellations of data, re-ordered in storage and transit as they are moved further and further away from their original context. In Benjaminian terms their 'place in time and space' (1999: 214) is compromised as they circulate through the Web, and especially through text-based information and marketing platforms such as Google's search and advertising technologies. Here they perhaps risk losing their 'aura'; stripped of their contextual meaning, their value becomes subject to the fluctuating laws of the market (Bruno 2002; 2012), rather than any residual narrative function, or indeed even the intent of the human searcher (Jarrett 2014). As I explained in Chapter 4, and will again in more detail in Chapter 8 (with the concept of Subprime Language), as words flow through the portal of the search bar, the subjective context from which they originated is negated in favour of their potential profit as optimised keywords, and each time they make that journey, circulating around the web, their economic value increases.

5.3 What is language in the age of algorithmic reproduction?

Language isnt what it used to be (Hayles 2006).

According to Google itself, there are over 200 different 'signals' which can affect the rankings of search results, including location data, personalisation the device you use, browsing history, and the time of day, not to mention the intricacies of the sorting and ranking al-

gorithms themselves (Search Engine Land 2010). These issues have been discussed at length by other scholars, but what I want to do in this section is to take some of issues I discussed in the previous chapter about data structures and the searchable database, and relate them to the cultural and political context in which Benjamin wrote his 'Work of Art' essay. To briefly recap, in Chapter 4 I put forward the concept of 'Geographies of (con)text' in order to explore how the language on the web is constructed and distributed, and the wider effects of its algorithmic processing through platforms such as Google search. In particular, I discussed what is perhaps the most fundamental factor in determining what comes out of a search engine, and that is what goes into it, i.e. the data available on which the algorithms can work. Without such data structures, algorithms are, as Fuller and Goffey (2009) point out, 'useless'. What I call the 'searchable database', however, which produces the context for all web searches, auto-predictions and auto-corrections is far from uncomplicated, and holds within it the same hierarchies of links, language, categorisation and social and political agency as any socially produced space (Massey 2013; Lefebvre 1991).

What is uploaded to this corpus of data is also contingent on who has the technology and the skills to upload to it, which can be heavily socially and geographically skewed, depending on different levels or types of internet access in different parts of the world. In this respect, one of the things which first struck me about the 'Work of Art' essay was Benjamin's fairly damning description of how advances in the printing and distribution industries had enabled those he considered unworthy and unqualified to contribute to the production of written texts. Noting that 'the greatly increased mass of participants has produced a change in the mode of participation', Benjamin complained about how an

> increasing number of people became writers at first occasional ones. It began with the daily press opening to its reader's space for 'letters to the editor'. And today there is hardly a gainfully employed European who could not in principle find an opportunity to publish somewhere or other comments on his work, grievances, documentary reports, or that sort of thing (1999: 225).

What this shows, however, is that literary agency a century ago was also dependent on economic, geographical, social, and class factors, i.e. those 'gainfully employed', those in Europe, those with access to the necessary technologies, and, in this context, those who happened to be male. The advent of Web 2.0, to which 'anyone' has access and the potential to contribute might indeed have brought about a 'change in the mode of participation',

but far from being a democratising leveller of society, or a measure of equal participation, what can be contributed to the 'searchable database' is highly contingent and can exclude groups of people on multiple social-economic levels. As Kate Crawford (2013: para. 3) writes, 'data are assumed to accurately reflect the social world, but there are significant gaps, with little or no signal coming from particular communities'. As I mentioned in the previous chapter, there are also both technical and economic factors that also skew the overall 'level' of the content available online.

But even amongst the privileged who had the means, skills and opportunity to 'become writers', Benjamin noted the changing quality of what was being contributed to the literary archive. When those who have traditionally only been readers or consumers of text suddenly gain access to authorship, Benjamin (1999: 225) worried that 'literary licence is now founded on polytechnic rather than specialised training'. This seems perhaps a familiar argument with regards to the amount of unedited material that circulates on the web today. There is, therefore, no guarantee of the reliability or quality of the dominant content, or as Geert Lovink (2008: 45) puts it 'with the rise of Internet search engines it is no longer possible to distinguish between patrician insights and plebeian gossip'. This is of course further complicated with the increasing popularity of computer generated language in the creation of news stories or weather forecasts, for example.

5.3.1 Search Engine Optimisation

The concept of the reader-turned-author is perhaps most relevant when it comes to my research into SEO. Search Engine Optimisation has become big business. Whether self-taught, through one of thousands of dedicated SEO companies, or via Google itself (there are many guides and instruction videos online), millions of individuals, businesses, journalists, bloggers, academics and other authors are using what Benjamin might have considered distinctly 'polytechnic' skills not only to upload content and add to the corpus of online data, but crucially to adapt that content for the specific consumption of algorithms. 'Google is', as Lisa Gitelman notes, 'certainly making us all into wordsmiths of a very particular, very narrow sort' (2013b). Perhaps the most basic of the methods of SEO is the strategic manipulation of what are, somewhat ironically, called the organic search results, which are the non-paid for results. This method necessitates close attention to the textual content of a website, to keywords and synonyms and to the linguistic construction

of links, tags, Universal Resource Locators (URLs) and Hyper Text Markup Language (HTML), but although language here is key, it is a language so infused with the practicalities of algorithmic enumeration and discovery, that it has perhaps developed its own agency, telling stories beyond its narrative function (Hayles 2012), or creating what Ferragina and Venturini call 'parallel language', or a 'folksonomy' which runs along side and through already existing language (2013). I will go into more detail about these alternative narratives constructed in the process of SEO and linguistic capitalism in later chapters.

According to Benjamin, when a work of art is mechanically reproduced, and that reproduction in itself has no artistic authenticity, it becomes subject to other social and economic forces and so necessarily enters the sphere of the political. Using the example of photography, he explains how the reproduction is no longer a work of art with an aura but has instead been 'designed for reproducibility' (1999: 218)[1]. Much like the keywords carefully implanted into a webpage, the reproduced work of art becomes an instrument of production in itself. Language on the internet has, it seems, taken on a different purpose. In order to optimise the efficiency of your website you must optimise it for reproducibility. You must plan your language with logical, as well as semantic, precision, picking keywords which will maximise your chance of exposure in much the same way (yet somewhat ironically) as we all have to supply searchable keywords which help optimise the visibility of our academic articles. What is interesting here in terms of SEO, however, is the language which is used *about* language. The 'Google verified' SEO instructor who led my intensive training course[2] suggested we 'make ourselves a list of synonyms' with which to bolster the keyword associated data on a website without having to repeat the keyword itself too many times ('keyword stuffing' is frowned upon by Google as an unfair practice). 'We dont talk about dictionaries, we talk about lexicons', my instructor continued, before showing an example of a list of 'synonyms', which were in fact words associated with each other in the loosest sense but certainly not synonymous. What SEO fails (or does not care) to recognise, is that words in a lexicon, unlike those in a dictionary, do not necessarily mean the same thing. The list which the instructor gave as an example was more like an entry in a thesaurus; the words all had something to do with each other (even if by virtue of meaning the opposite), however they were not synonyms. But by creating these connections with keyword lists and sitemaps, and then re-affirming them with clicks and

[1] This emphasis on the 'reproducibility' as well as the 'reproduction' of art is central to Benjamin's essay. A recent translation of selected works reflects this in its title: 'The work of art in the age of its technological reproducibility, and other writings on media' (2008).

[2] I attended a one-day intensive SEO course.

site visits, we are all - as search users of varying degrees - in effect teaching the algorithms that they are synonyms. It is this kind of erroneous word association that will return results for vegetarian when you search for vegan, for example, or Muslim when you search for Asian (Baker & Potts 2013).Benjamin2008

5.4 The death of the reader

Ah, humanity! (Bartleby the Scrivener, Melville 1998)

Perhaps an even greater destabiliser of meaning, however, is the process of decontextualisation through which language goes when processed online. When encoded for machine processing, 'all difference of media dissolves into a pulsing stream of bits and bytes' (Lunenfeld 2000: 7). In fact, in their digital form, there are no 'edges' between different media at all. 'Digital documents... are materially, bibliographically the same as the windows that they appear in and the programs that manipulate them' (Gitelman 2014: 17). So not only is the potential difference between what is art and what is literature, or a novel and a blog eroded, but the distinction between a body of text (as literature) and the type of text which is retrieved by algorithms through search engines is too.

As detailed in Chapter 4, even so-called 'semantic' searches, which aim to gather meaning from groups of data rather than spatial relationships between single words, are just matching data patterns, and when you add the auto-correct and auto-predict functions into the mix, the potential for the proximity and frequency of certain data to be compounded, re-confirmed and therefore perpetuated ad infinitum leads to the kind of anomaly which I encountered, where the word sexiest was presumably statistically more likely to be linked to the term wives and girlfriends than the word sexist. Safiya Noble's (2018) recent book *Algorithms of Oppression* provides numerous examples of the racial implications of this kind of process. As Mark Fuller and Andrew Goffey put it in their Evil Media manifesto,

> harvesting data from websites is a matter of using and then stripping off the mark-up language by which web pages are rendered in order to retrieve the data of interest and returning this to a database, ready for mining. At this point, semantics is largely irrelevant (2009: 149).

But to return to SEO, although Google's PageRank algorithm started out as a hierarchy of citations (Brin & Page 1998), as I explore further in Chapter 6, it was a system which

was easily gamed by link-farming or other black-hat techniques (see also chapter 4). This type of manipulation of the rankings is deemed unfair by Google, mostly because they are less in control of the demand, supply and income generated by web traffic. However, according to my SEO instructor (and confirmed by Matt Cutts at Google 2014), incoming links to a website are becoming less important than the actual content and construction of a site. As well as carefully selected keywords, 'Google likes fresh content', we were told, but by 'fresh content', my instructor meant any recycled content which has been altered slightly, commented on, or added to just enough to keep the web crawlers interested, and recommended in-site blogging as an effective tactic. This recycling of words in this manner, however, is curiously at odds with the recycling of other commodities in that the explicit goal is the production of more, rather than less, waste, and that this excess of waste, rather than its reduction, is the means of capital gain. But what Benjamin called 'locust swarms of print' have become part of what it means to exist online, and as long as web presence and search rankings hold such an existential power over society and the market, these swarms will 'grow thicker with each succeeding year' (2007: 78).

It is, however, not just commercial websites that have to adopt optimisation strategies. It is newspaper articles, blogs, and headlines, which also have to tailor their text to court the algorithms. And this text and those keywords will necessarily be reflective of the already popular; they cannot be new, or creative, or challenging otherwise they just would not serve their purpose. The result is the emergence of 'a new conformity in the language... of news online' (Dick 2011: 475), and various other types of content. This kind of practice, which is in effect mandatory if you want your site or your words or profile to 'exist' online, can only result in the clogging up of the searchable database with repetitive, unimaginative copy, or 'empty phrases'; a term used by Benjamin to describe the flood of cheap journalism which mechanical reproduction facilitated. 'The empty phrase', he wrote,

> is an abortion of technology... the expression of the changed function of language in the world of high capitalism... the label that makes a thought marketable, the way flowery language, as ornament, gives it value for the connoisseur (2007: 242).

Flowery language, however, means nothing in Google search, where an army of algorithmic 'Bartlebys' (Melville 1998) are not reading, but merely copying, and therefore reproducing empty words and empty phrases from a decontextualised pool of linguistic data. Similarly, the Optical Character Recognition (OCR) software which digitises text 'chronically

"misreads", not because of any hardware malfunction or programming error but precisely because scanning is not reading' (Gitelman 2006). This is not so much the Death of the *Author* (in a Barthean sense) (Barthes 1977), but what we might call the Death of the *Reader*, and evidence of a slow strangulation of creativity and authorial autonomy. A democratisation of literary agency perhaps, but also the creation of a proliferation of authors forced to write text for algorithms to read in order to remain visible. As Murray Dick concludes in relation to Google's effects on journalistic standards,

> SEO is applying pressure to these standards not in the perceived interest of the reader or because of publication constraints (as once may have been the case), but in the interests of a third-party commercial arbiter in online distribution: Google (2011: 475).

I explore George Orwell's vision of politicised language in Chapter 7, but his views on the 'decline' of the English language in a pre-digital age are also of interest here. To Orwell, the popular recycling of lazy metaphors, phrases and images rather than inventing new ones turns speakers and writers into machines, a state of consciousness which he sees as 'favourable to political conformity'. As we see from *Nineteen Eighty-Four*, this machinic and constricted mode of writing is indeed inherently political in that it also governs and constricts thought. Like Benjamin's 'empty phrases', to Orwell, the authoritarian power of language also comes from unimaginative repetition of already existing language, or what he calls 'gumming together long strips of words which have already been set in order by someone else, and making the results presentable by sheer humbug', which could perhaps be used as a presciently sharp critique of the commercial SEO industry.

But despite his apparent disdain for 'the empty phrase', Benjamin did also see the emancipatory or even revolutionary - potential in new, more democratic and technologically enhanced forms of media and art, especially in the medium of film. In the 'Work of Art' essay, he describes the visceral, shocking effect that rebellious new forms of writing such as Dadaism might have on the masses. Eschewing the traditional economic and linguistic structures and values of the old bourgeois order of language and literature, Dadaist work smashed together apparently disparate montages of words and phrases; what Benjamin (1999: 231) refers to as 'word salad'. In another essay on 'Experience and Poverty', Benjamin (2006: 733) elaborates on the new kind of language made possible by the experience of modernity, writing that 'what is crucial about this language is its arbitrary constructed nature, in contrast to organic language'. It would at this point be easy to suggest that

the 'arbitrary' radicality of surrealist language, poetry and art in Benjamin's era has parallels to some of the digital art made today from 'arbitrary' algorithmic language such as search results and autocompletions, for example Google Poetics, or what Dan Hoy (2006) (critically) calls 'Google-sculpted' flarf poetry, which are issues I go into in some detail in chapter 7. But as the next sections will describe, algorithmic reproducibility is different to mechanical reproducibility to two key ways. Firstly, due to the responsive, learning and controlling potential of the proprietary algorithm, and secondly, because far from being arbitrary (and also ironically far from being 'organic'), algorithmically reproduced language is in fact mediated by the rules and regulations of companies such as Google, and also by the neoliberal laws of the digital market (see chapter 6). So 'rather than stimulating a participatory revolution, new technologies have doused us even deeper in new, ever-more intrusive forms of capitalism'.[3]

5.5 The commodification of language

> *Google has created the first global, real-time, and multilingual linguistic market (Kaplan 2014: 59).*

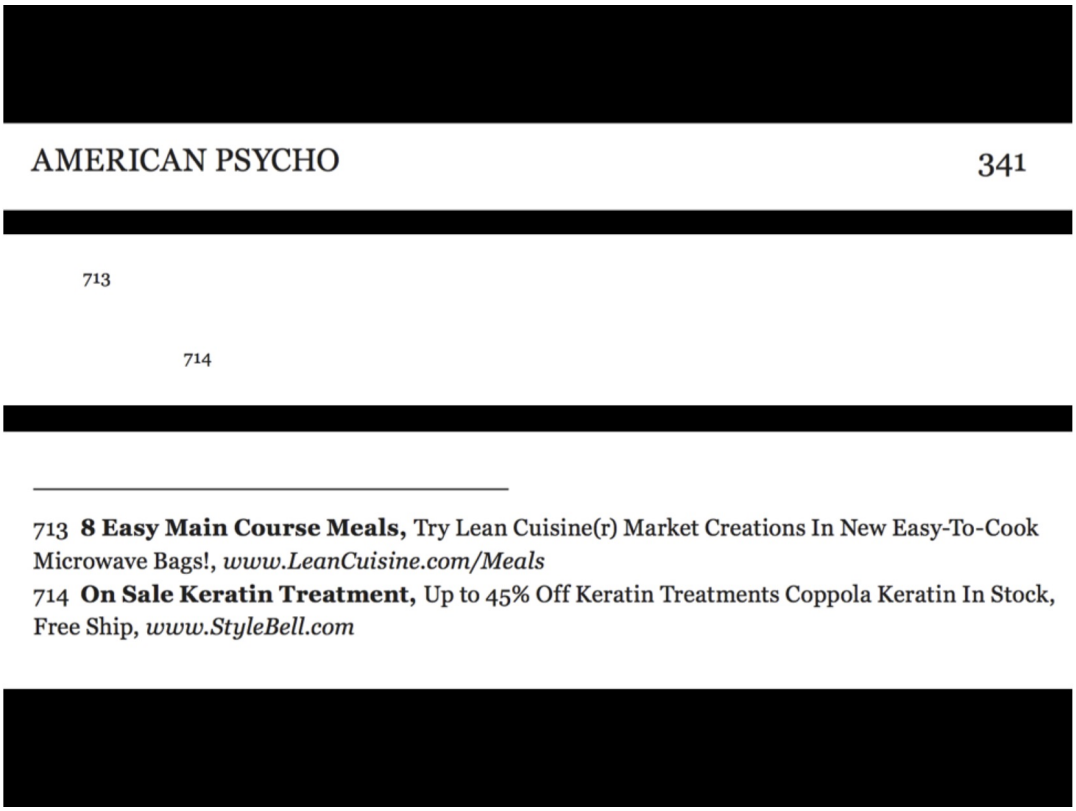
Optimisation of the 'organic' results might appear to be a 'natural' way to manipulate the search rankings, but as I have shown, are far from free of economic incentive. The machinations of the market are, however, far more clearly visible in other ways, most noticeably in issues around plagiarism, the online auctioning of keywords (AdWords), and targeted advertising through web browsers and email (see Chapters 6-7)).

To one extent or the other, language has always been commodified ever since things were written down, became reproducible, mobile, and therefore entered the market. What Benjamin (1999: 212), calls the ritualistic, cult or 'auratic value' of pre-mechanical art or language is lost when it is reproduced 'by third parties in pursuit of gain' therefore becoming a commodity. And 'once a monetary value has been established', writes Mira Seo (2009: 582), 'plagiarism can enter the literary discourse'; the original 'aura' of the work of art becomes less important than its exchange value and the preservation and ownership of its rights for commercial gain. Mechanical, and now algorithmic, reproduction can only accelerate this process. If we think of this in terms of text, this immediately translates into who has the 'right to copy' text, and how this copying is regulated.

[3]Feedback received on an early version of this chapter, submitted to a journal for publication in 2015.

As Lucas Introna (2015) writes on plagiarism detection algorithms, now language is so widely available online through search engines this adds to the problem of who can claim 'ownership' of it intellectually, physically, or financially. But importantly, plagiarism algorithms do not actually identify someone else's 'work' (as in the essence, or aura of their work), or their 'ownership' of it. Instead, they merely match the character strings to which language is reduced when processed through computers. But taking a step back, can language really be 'owned'? Obviously, a book can be owned, and there are laws around the protection of intellectual property, but even the relative stability of the written word is complicated by the paratextual phenomena which arise through new technologies of reading such as Amazon's Kindle. The commercial e-text is in effect 'leased' to a reader and can be monitored or indeed revoked at any time. As Ted Striphas writes, 'economic imperatives and legal considerations have become more deeply implicated than ever in routine cultural practices, such as book reading' (2010: 310).

As I will explore in more detail in the following chapter, another way in which to see the infusion of language with capital is analysing the way that AdWords works, which is by applying economic values to the words which flow through the Google search bar. Although Google may have started off just as a search engine; an online version of a public library or records office, 'over time, it transferred itself into an advertising company, producing not search results, but audiences as its primary commodity' (Feuz et al. 2011). But in order to gain these users or audiences and their lucrative data, Google first has to commodify the words which attract that audience, a process which Kaplan has called 'linguistic capitalism' (2014). In short, if you want to guarantee your website its place at the top of the search engine rankings (without all the time-consuming bother of optimising your site), you have to bid for the word(s) you feel will attract the most traffic. This creates a strange world in which words have value in a way curiously abstracted and sometimes seemingly disproportionate to their usual 'ranking' of importance or relevance in everyday language and life. They have quite literally been auctioned off for their exchange value; their integrity compromised by the distance between the 'original' and the reproduction in much the same way as Benjamin describes. As in the case of the organic search rankings, here too creative writing is superfluous to the needs of the market. Google actively police the AdWords system to make sure it is functioning correctly as a generator of Ad capital. AdWords which fail to generate enough click-through revenue are removed, as the artist Christophe Bruno (2002) found out when he attempted to buy AdWords which



AMERICAN PSYCHO 341

713

714

713 **8 Easy Main Course Meals,** Try Lean Cuisine(r) Market Creations In New Easy-To-Cook Microwave Bags!, *www.LeanCuisine.com/Meals*
714 **On Sale Keratin Treatment,** Up to 45% Off Keratin Treatments Coppola Keratin In Stock, Free Ship, *www.StyleBell.com*

Figure 5.1: Example page from Cabell & Huff's *American Psycho* (2013.)

sought to provoke people's imaginations rather than their wallets. Like the Dadaist poets and artists described by Benjamin, he had 'sacrificed market values in favour of higher ambitions' (1999: 231). As it turned out, however, the exchange value of Bruno's poetic keywords was not enough for Google, and his adverts were swiftly removed from the listings.

Linguistic capitalism is also evident, although by its absence or erasure (Marczewska, 2015), in the way language moves through other digital spaces such as electronic mail. In 2010, design graduates Cabell and Huff (2013) set up an artistic experiment by sending each page of the novel 'American Psycho' (Ellis, 2010) between each other via GMail accounts. Aware that GMail algorithms read email traffic for the purpose of targeted advertising, they then recorded the adverts that were generated and, having removed the original text, reconstructed the book by footnoting the adverts on the otherwise empty pages. This was not only a brilliant nod to the kind of destructive consumerism Ellis was critiquing, but it reveals the shortcomings of these types of semantically trained algorithms when presented with creative or artistic language. Patrick Bateman's skinning and mutilation of women generated adverts for skin tightening and teeth whitening procedures; his

violation of their bodies with rats and chainsaws prompts adverts for rodent control and topiary. In a particularly chilling example, an advert for healthy soup options is served around the scene where Bateman is cooking the head of one of his victims in the microwave (see Figure 5.1). I found this a fascinating way of showing what is happening to words when they move through digital space in this case via electronic mail, which has invisibly stripped the words of their linguistic or creative value in favour of their exchange value. The words of the author, turned involuntarily into monetised data have been 'stripped of everything but that data's end product [which] constitutes perhaps both the purest document of digital capitalism and the sharpest critique of that capitalism' (Benzon and Sweeney, 2015), giving a glimpse into the meaning and value that Google's algorithms extract from narrative fiction when they 'read' the text.

The mediation and manipulation of language in an age of algorithmic reproduction is also visible through the advertising platform AdSense, where Google pays websites to host adverts. Harvesting payments for each ad exposure, a website becomes more profitable the more times it is looked at. AdSense is closely linked to the AdWords platform, and some of its adverts are targeted to keywords in the websites they populate, but perhaps the more problematic linguistic side effect of AdSense is that it has become a vehicle for the spread of fake news, which I will discuss in greater detail in Chapter 8. In the run up to the US Presidential election in 2016, it was reported that 'fake news factories', some based in small towns in Eastern Europe, earned thousands of dollars from AdSense by making up and circulating politically controversial and viral stories. The more controversial the story, the more lucrative its advertising revenue potential. As with AdWords, although in a different way, the language constructed to earn advertising dollars via fake news stories, is therefore similarly compromised by its entanglement with market logics and the political sphere.

Once again, it is worth remembering that texts were being adapted and manipulated to accommodate advertising from the emergence of print-capitalism and the proliferation of the printing presses. As Benjamin (2007: 77) eloquently notes in his 'One-Way Street' essay, 'script - having found, in the book, a refuge in which it can lead an autonomous existence - is pitilessly dragged out into the street by advertisements and subjected to the brutal heteronomies of economic chaos'. But in the digital, or algorithmic production and reproduction of language, advertising and propriety operate under the surface rather than in the street, relying on what Dan McQuillan (2015: 565) calls an 'implicit social contract'

(see also Beer 2009: 67) which uses 'free' services to conduct data-gathering projects just as expansive as those of the intelligence agencies, but they are largely based only on trust, not - ostensibly at least - enshrined in law. Of course, the interrogation of Mark Zuckerberg by US Congress over the data gathering and sharing practices of Facebook in 2018 has dragged the role and implications of digital advertising and politics right to the fore of public knowledge, and it may not be long before Google is forced to face similar public scrutiny.

5.6 The aestheticisation of politics

Dont be evil (Google motto)

Google dropped their famous 'Dont be evil' motto in October 2015, when new parent company Alphabet took over, but the phrase has always been a tacit clue as to how aware its founders were of the potential power of their invention. Following on from the previous two sections which looked at the linguistic and economic aspects of reproducing language through Google's organic and paid search and advertising platforms, the third method suggested by my SEO training instructor to get your content visible and readable at the top of the search rankings is, in contrast, incredibly simple; you can promote your site through a Google+ account. While I am not suggesting that encouraging users to sign up to Google services is necessarily 'evil', it is hard not to contemplate the inherent power that one company now has over the ability of people to 'exist', and to make their voices heard on the web. In fact, each of three SEO methods suggested in my research requires a certain degree of interaction with Google; whether it is transacting with money or with data, by contributing to what can be searched for and what can be found, or by helping to develop and hone the algorithms with click confirmations or 'implicit feedback' (Granka 2010: 367). We have all, to one extent or another, become part of the system; assisting the algorithmic reproduction of language away from its source and, like Benjamin's Work of Art, making it a political process, not just a linguistic one.

Central to Benjamin's essay is the observation that the extreme and fascistic politics of that era had been 'aestheticised' by its co-option of popular culture. Facilitated by the technologically enhanced reach and influence of mass media such as photography, news, advertising and art, Benjamin believed that the violence of an authoritarian regime had

been disguised as an awe-inspiring show of cultural and militaristic strength, and that the population had been appeased by their apparently voluntary immersion into new, and beguiling levels of cultural awareness and participation, or as David Berry puts it, 'fascism introduces aesthetics into political life as a way of giving the masses a chance to express themselves' (2014: 33). As I said earlier, Benjamin did not see mechanical reproduction as necessarily a *bad* thing, but what the essay does tell us is that he recognised that potentially bad mechanical advances like those made in warfare, which used technology as a force of destruction rather than construction, could be neutralised by the co-option of the masses into the aesthetically subjective cultural power of the political and martial spectacle. In 1930s Germany, according to Koepnick, '[transforming] the political itself into an item of mass consumption, a commodity concealing its status as a commodity (and allegory), a symbolic spectacle meant to produce lonely crowds and unite the masses as separate' (1999: 198). It is not hard to extend this to modern day social media and tech companies such as Google and Facebook, whose business models are based on the financial exploitation and commodification of user data masked behind the quasi-philanthropic rhetoric of wanting to connect people, or to get them access to the world's information.

Although we now live in a very different political era, it is also not difficult to see the more explicit political and military parallels between Benjamin's era and today. Not only does Google (Alphabet) executive and former chairman Eric Schmidt advise the Pentagon as part of the Defense Innovation Board, but there has been considerable recent controversy over Google's involvement in a Pentagon program which 'uses artificial intelligence to interpret video imagery and could be used to improve the targeting of drone strikes' (The New York Times 2018), and indeed Google are currently bidding to provide cloud services to the Pentagon. 'We believe that Google should not be in the business of war', is the opening line of a letter from Google employees protesting at this involvement, but when the identification of military hardware becomes part of the everyday free labour Google users unknowingly undertake by completing RECAPTCHA puzzles depicting military trucks and helicopters, their (and our) total (and voluntary) co-option into the system (and into the spectacle) is surely almost complete. In recent years we have also witnessed how the exploitation of user data and digital advertising have facilitated the spread of fake news and misinformation on a scale that may have had an effect on Donald Trump's election as US President and continues to cause dangerously heightened diplomatic tensions between Russia and the west.

As well as the potentially dangerous power unleashed by the spread of language in the form of fake news, the power of algorithmic systems has also been the subject of much recent scrutiny. Scott Lash (2007) sees a 'sinister' and 'military' control in what he calls the post-hegemonic power of cybernetic systems, and Taylor Owen (2015) and others have expressed concern about the potential or inherent 'violence' of mathematically mediated decision-making processes (see also Weizman 2012), while Cabell & Huff's American Psycho project shows what Benzon and Sweeney (2015) call the 'invisible violence and excess of big data'. David Beer (2009) extends Lash's argument specifically to the perhaps 'unconscious' power of algorithms in this process, and I think it is possible to see in a company like Google not only a frighteningly wide-reaching power over people in terms of the data it holds, which, at any time can perhaps be subpoenaed, hacked, leaked, or corrupted, but in Google it is so easy to be lulled into a false sense of security by the bright colours, the silly name, the almost childish format and interface and the 'free' and easy applications. This 'front-end Googlization' (Rogers 2009: 173); the interactive Google Doodles and puzzles which co-opt and 'rebrand the past' (Hoad 2015), the charming simplicity of the search box and the apparent ease with which anyone can interact with and contribute to public discourse, in effect disguises the politics of search with a layer of aesthetically pleasing innocence. However, as Benjamin reminds us, this aesthetic is no accident. Afterall, 'fascism sees its salvation in giving these masses not their right, but instead a chance to express themselves' (1999: 234).

And it is precisely this sense of agency and expression, afforded by the ability to comment on news stories, connect with friends and create content online, that is so attractive, and perhaps addictive, today, despite the growing number of users who are completely aware of their own exploitation, the power and value of their data, as well as the potential safety and privacy risks of total immersion into the world of digital commerce and capitalism. Following Deleuze and Guattari, Cheryl Gilge (2015, 2016b) refers to this seemingly counter-intuitive voluntary co-option of the masses into Web 2.0 as 'microfascism', with Google leading the way with an addictive aesthetics of utility and inclusion which continues to enable its powers of monopoly and manipulation.

There has been a substantial amount of both popular and academic critical attention given to discussing, or challenging, Google as a monopoly, as a form of 'algorithmic governance'

(Rouvroy 2013; Danaher 2016; Just & Latzer 2017, see also next chapter), as perpetuating class divide, as the (un)authorised holder of knowledge (as in the 'right to be forgotten'), as a government source, as a stealer of data, censor of the world, exploiter of users or as a peddler of racist, sexist or politically incorrect narrative (see Noble 2018). But despite these concerns, Google's disarmingly user-friendly interface remains our most popular portal into the Web, despite the existence of other search engines and the increasingly obvious post-Snowden data security implications.

The obvious defence to any such claims is, of course, to fall back on what Tarleton Gillespie (2014) calls the 'illusion of algorithmic objectivity'. Algorithms, data, or any other concept with a scientific basis cannot be 'evil', just as much as they cannot be 'good', 'happy', or 'jealous', for the very obvious reason that they are non-sentient mathematical formulae or constructs, although as Lovink argues, it is possible that they can be out of control (Lovink 2009; see also Kitchin 2017). Dan McQuillan (2018) has, however, recently suggested that algorithms can be 'thoughtless', in the sense that their very 'banality' (he uses Hannah Arendt explicitly) can escape regulation, ethical concerns and accountability, and can therefore be considered so reliable and objective that they can be used in pre-emptive actions of law enforcement, governance and oppression. Gillespie suggests that simply by calling search rankings 'results', they are automatically given a sense of unquestionable mathematical legitimacy, which is, as I have shown, not always the case. Specifically in terms of language, Matthew Fuller and Goffey (2012) have explored the (deliberately anachronistic) potential of 'Evil Media'. But naturalising, or anthropomorphising digitised data is misleading, and perhaps disguises more important issues and performances. As Orit Halpern suggests in her book 'Beautiful Data',

> despite the seeming naturalness of data and its virtues, therefore, there is nothing automatic, obvious, or predetermined about our embrace of data as wealth. There is, in fact, an aesthetic crafting to this knowledge, a performance necessary to produce value (2014: 5).

Going back to the introduction to this chapter and the thesis, I think this is a distinctly industrial performance, with all the embedded social and economic structures which that method of production entails, but these inequalities are hidden either in the sheer banality of process, or by the 'aesthetic crafting' of the owner of both the raw material (in this case Halpern's 'Beautiful Data') and of the means of production, and in both cases that owner is increasingly likely to be Google.

5.7 Conclusions

This chapter has considered the linguistic, economic and political factors which influence, and are influenced by, the algorithmic reproduction of language through commercial platforms such as Google Search, and to do this has taken Walter Benjamin's thoughts on mechanically reproduced art and language and contrasted them with experiences of how SEO and linguistic capitalism work today. I have suggested that algorithmic (re)production is the new mechanical, i.e. a logistical facilitator, moving and circulating information and data around to produce capital, and that the critical scrutiny of the ownership of this means of production and the potential power that brings is therefore a priority, after all, language is power, and at the moment Google has hold of a significant proportion of the narrative. But in addition to that we also have to take into account how algorithmic reproduction differs from mechanical. Not only does it replicate (or perpetuate) the process of industrial production, but it also adds to that process in a way in which mechanical replication did not (Carpo 2011); each click or query 'changes the tool incrementally' (Gillespie 2014).

As Martha Poon (2013) states, 'algorithms execute action with a degree of responsiveness... [their] internal mathematical structure allows [them] to adjust depending on the changing input conditions'. So when algorithms work on language, and when text is encoded and then decoded through computational processes like online search, the tensions between numbers and words, or logic and creativity (Fuller & Goffey 2009), machine and ideology (Galloway 2012), or database and narrative (Manovich 2001: 225), have a performative and practical effect (Totaro & Ninno 2014), and can therefore result in a kind of anomalous and ontogenetic excess of language. As Hayles suggests, the 'processes of signification change when speech and writing are coded into binary digits... [and] they do not emerge unchanged by the encounter with code' (Hayles 2005: 39). These generative powers become potentially problematic, however, because they 'are also pathways through which capitalist power works' (Lash 2007: 71). Language has always to some extent been linked to capital ever since the first instance of material reproduction, but when digital technology is underpinned not only by different forms of capital, but also different, perhaps linguistically incompatible mathematical processes, the discursive power it has is not only greater in terms of reach, but also as a result of the unpredictable, evasive and easily

manipulated nature of algorithmically reproduced language. Although Google claims its search engine *indexes* and *sorts* the world's information in some kind of definitive, reliable and perhaps passive way, in reality, changing and contingent human input, new kinds of market forces, and the responsiveness and potential performative capacities of the algorithm sometimes make it seem more like the Dow Jones than a library index.

But although the 'hand of the market' might be said to have potential epistemic power in this context, there are other actors in this assemblage too. As Lucas Introna writes, there are risks in 'locating, or placing too much agency "in" the algorithmic actor, rather than in the temporal flow of action in the assemblage as a whole' (2016: 22), which is why I have also emphasized the part that we as human actors play not only in writing the code and programming the algorithms, but in producing the content and context in which they can work. Although Artificial Intelligence (AI) and Natural Language Processing (NLP) technology has perhaps begun to destabilise and blur boundaries between human and machine in terms of authorship and readership, there are still considerable geographical and cultural gaps in the data environment which also necessarily affect the outcome of language reproduced through search. Even when not directly affected by overt advertising strategies, the words which pass through the search engine are re-infused with the residue of the dominant structures, market forces, biases and stereotypes which make up the corpus of available searchable data. Their meaning as such may not be fixed, but their use - and therefore their agency - will always be linked to their economic value, which I show in detail in my intervention in the final chapter.

If we think how important language can be as a tool of power over people, places and opinions in a logocentric, epistemological or physical way, then just as the mechanical reproduction of art was for Benjamin, the algorithmic reproduction of language is an inherently political process and a fundamental means of the distribution of power. And just like Benjamin was frustrated by the lack of revolutionary potential in a population appeased by the mass accessibility of art and the aestheticisation of politics, and therefore co-opted into the system, so it is now nearly impossible to avoid being a part of the Google language machine, whether we intend to be or not.

Indeed, this is critical point when it comes to cultural, academic, and artistic critiques of digital technologies, and in particular Google. If they are unaware of, or ignore their

place in this language machine, then their standing as critique is necessarily compromised. As we see at the end of the 'Work of Art' essay, Benjamin's solution to 'the situation of politics which Fascism is rendering aesthetic', is that 'Communism [should] respond[s] by politicising art' (2009: 235); a call, perhaps, to recuperate art from the clutches of the political and economic systems which use it as a means of power and control. But if this interpretation of Benjamin is to be of any use, it is imperative that the precarious cultural and political situation in which we find ourselves today by virtue of the 'microfascisms' of algorithmic technology is taken seriously (Gilge 2016b), and that the potentially redemptive and emancipatory artistic interventions spawned by these technologies do not in themselves become part of the spectacle. Political art as Benjamin saw it must, as David Berry writes, play a central role in 'demystifying the production, the distribution, the form and the content of art, in an attempt to make art serve the cause of the masses and not vice versa' (2014: 33). If we are to further Benjamin's vision, we must, therefore, guard against 'the uncritical use of corporate algorithms as a generator of poetic chance' (Hoy 2006), and be wary of the dangers of 'aestheticising Google critique' itself (Rogers 2018) by omitting to foreground the politics of production that mediate all representations of proprietary digital technologies. We must therefore make sure that language in the age of algorithmic reproduction, threatened and at the same time made threatening, by digital commercialisation, can still be mobilised against the aestheticised politics of technology, be it mechanical, algorithmic, or whatever comes next. My own attempt to reverse the aetheticisation of the politics behind Google's search and advertising technologies (by way of a political artistic intervention) is explained in detail in Chapter 7. But before that, the next chapter aims to shine a light on the sometimes opaque relationship between the marketplace of words which Google has created through its AdWords platform, and the wider political sphere, exploring the tensions between market governance, and the systems of algorithmic governance that have emerged as a result of these new digital economies.

Chapter 6

THE POLITICAL ECONOMY OF ADWORDS: LINGUISTIC MARKETS AND ALGORITHMIC GOVERNANCE

6.1 Introduction

This chapter stems from two events; unconnected in time and space, but which in different ways prompted me to start thinking more about the political economy of Google's search and advertising platforms, in both theoretical and practical terms. My research is based firmly around Frederic Kaplan's concept of 'linguistic capitalism', using the idea of a linguistic marketplace to critique and make visible the power of monetised language in today's digital economy and society. However, I feel my thesis would not be complete without an interrogation of what exactly linguistic capitalism means in a wider political economy setting; questioning the capitalist credentials of the term, and the algorithmic and market-based logics that purport to drive it. In this sense, although throughout the thesis I have been critiquing Google's monetisation of linguistic data as a type of linguistic capitalism, in this chapter I turn my critique to the term itself, asking questions such as what kind of market am I talking about and what type of 'capital' does it circulate and produce; on a political scale, just how 'capitalist' is this linguistic market, and to what extent it is a 'free' market; what the role of Google and algorithmic systems are in this market, and subsequently, what kind of governing and/or knowledge producing properties it may or may not have.

The first of the two events that guided this chapter occurred at a workshop I presented at during my time as research assistant at NUI Galway in 2016. Organised by John Danaher, the workshop, 'From Algorithmic States to Algorithmic Brains', was one of the first times I had presented my {poem}.py project as a full paper, and I remember being very nervous about the other participants who were internationally known experts in fields such as AI and robot ethics, transhumanism, and the morals of neuro and bio enhancement. It was, however, from a completely different field that the most challenging response to my paper came. One of the delegates turned to me and suggested that far from being a potentially problematic mediator of language - Google AdWords was in fact a triumph of Hayekian free market economics; a linguistic market that effectively, and brilliantly, regulates the distribution of goods around the globe. There were a number of responses to this challenge that I wish I had had the wherewithal to articulate effectively on the day, but despite being an unexpected comment, I was in hindsight very grateful for the provocation, because it presented me with the language, and with the theoretical framework, with which to delve deeper into the sometimes ambiguous and opaque political economy of Google and its AdWords platform.

The second event that inspired this chapter is more empirical in nature. It occurred in the run up to the 2017 UK Parliamentary General Election, when several different actors entered an AdWords bidding war over the keyword 'dementia tax'. Over the course of one 24-hour period, the three main political parties, charities, and also independent actors began to compete against each other to secure the most prominent positions for their own particular narratives on the search results page. Analysis of this incident provides a fascinating snapshot into how the system works, and is worked, by actors with different motives (and budgets), and also into the wider effect that the AdWords platform has in the political sphere. So in this chapter I will first give an overview of the origins of AdWords, before considering how Google's search engine came to be perceived as what Kaplan has called a 'linguistic marketplace', which as I will show is a sometimes ambiguous conceptualisation of a system which is perhaps based less on the market than on what we might call a centrally planned, opaque and algorithmically mediated system of governance, which in effect hides behind free market rhetoric for justification and defence against criticism. The chapter goes on to examine the political economy of Google AdWords through a Hayekian lens, using his ideas about the knowledge producing capacities

of free markets to test whether Google's platform really is capitalist in nature, and, helped by some original empirical work of my own, to gain insight into what knowledge and social capital might be produced by such a system, and to what effect. But first, back to where it all began

6.2 The anatomy of a large-scale hypertextual Web search engine

When Sergey Brin and Larry Page invented Google as students at Stanford in 1998, they knew from the beginning how advertising could interfere with the efficiency and integrity of their proposed search engine. In an appendix to their original paper, on 'The anatomy of a large-scale hypertextual Web search engine', they noted that 'the goals of the advertising business model do not always correspond to providing quality search to users', and that 'advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers' (2012: 3832). They were right to be concerned that their goal to 'organise the world's information' might be compromised by advertising, but they had to fund their project somehow, and so adopted paid advertising by developing the AdWords platform. Although its advertising empire has expanded significantly with platforms such as AdSense and Gmail, AdWords remains Google's main stream of revenue, and is certainly the reason the company exists in its current successful and dominant form today. But despite their initial caution, Brin and Page could surely not have anticipated just how much the search engine would be compromised by advertising and the Search Engine Optimisation (SEO) industry. Indeed, a typical Search Engine Results Page (SERP) today is probably unrecognisable from what they first imagined. Neither would they have imagined that not only have search results become biased towards the needs of the advertiser, but, as I will argue, they have taken on a political economy of their own.

As explored in Chapter 4 (Geographies of (con)text), digital technologies such as Google search are used by many different actors, with many different motives. While the search engine 'users/consumers' as imagined by Brin and Page in 1998 were the people using Google to look for the 'world's information', the 'users' of the engine today include advertisers of commercial products, the SEO professionals and amateurs who are constantly attempting to manipulate the information displayed on results pages in order to gain eco-

nomic capital, but also a whole new set of 'users' intent on the creation and retention not only of economic, but also cultural, social, and even political capital.

These new users of the search engine range from celebrities who use SEO techniques and specialist companies to manufacture and maintain their reputations, to those wanting to spread fake news stories (either for financial or political gain - or both), those wanting to spread extremist material for the purposes of recruitment (and those trying to counter these narratives), or indeed mainstream political party broadcasts trying to win office. Of course, the user typing in search queries and looking for answers is still ostensibly the main user of Google, and indeed without the searcher-as-user, the rest of Google's digital ecology would cease to function. The search bar, and all that flows through it, is after all the liminal pivot on which Google rests. However, because of the complex geographies of (con)text I described in Chapter 4, the core 'use' function of Google search is now not only hampered by systematic structural bias, but is also irrevocably linked to the often invisible political and economic undercurrents which lurk behind the search bar and its associated assemblages (see Chapter 5 on the aestheticisation of politics). This means that although the way Google works is sometimes still seen, and, as I will explain, indeed promoted, as a simple mechanism for the organising and extraction of the 'world's information', it is in practice a long way from a free-flowing marketplace of either ideas, language, or commodities. This chapter will therefore suggest that, while Google AdWords presents as free marketplace in which 'linguistic capitalism' operates for the optimal production of knowledge and distribution of goods, in effect the idea of a knowledge producing and self-regulating marketplace is a construct perpetuated by Google (and supported by the SEO industry) in order to mask an inherently 'centrally planned' and judgement laden system which hides its responsibilities and accountability behind the facade of the algorithm and the market (Pasquale 2016).

Rising to the challenge presented to me in Galway, this chapter therefore goes back to Hayek's original arguments about the 'use of knowledge in society' and argues that the AdWords market is far from 'free' in Hayekian terms, but is highly regulated and corrupted by various political, ethical and financial distortions, which makes it virtually impossible for any tacit knowledge to be produced from it (Thornton & Danaher forthcoming).[1] Algorithmically-mediated price mechanisms do not, and probably cannot, accumulate and

[1] Or at least calls into question any apparent 'wisdom' that is produced by this market, as I will go on to show.

signal that same kinds of knowledge that Hayek felt were essential to effective market governance. Instead, the linguistic market can be seen not as a form of Hayekian market governance, but as an opaque, yet powerful, system of algorithmic governance (see Chapter 5).

6.2.1 Algorithmic markets

In September 1945, Friedrich Hayek published 'The Use of Knowledge in Society', in which he made a bold claim about the information processing powers of the free market. Hayek argued that the price mechanism contained within the market functions to automatically collate and communicate information from diverse and partial sources, thereby enabling people to better coordinate their actions to mutually beneficial ends. By this logic, the price mechanism was a governance tool par excellence, preferable to centrally controlled governance tools in almost every way (Hayek 1945).

We live today in an era characterised by the use of another information processing tool in governance: the computer-coded algorithm (Danaher 2016; Pasquale 2015; Rouvroy 2013). In a networked and digitised age, an age in which nearly every move we make is recorded, logged and 'quantified' (Moore 2017), we have come to expect, and possibly even demand, the perceived support and assistance made possible by such algorithmic systems. We look to the information processing powers of big data algorithms to help us make sense of the complexity we have created. These algorithms feed us information, often packaged into simple metrics like scores and ratings, that we use to coordinate and cooperate with one another, and to facilitate the smooth running of our everyday needs and consumption. As Karen Yeung notes, these algorithmic decision-systems:

> manage risk or alter behaviour through continual computational generation of knowledge from data emitted and directly collected (typically in real time on a continuous basis) from numerous dynamic components pertaining to the regulated environment in order to identify and, if necessary, automatically refine (or prompt refinement of) the system's operations to attain a pre-specified goal (2018).

There are of course many examples of algorithmically mediated markets, from shopping sites to high frequency trading, but the type which lend themselves to a Hayekian analysis

are what we might call 'algorithmically constructed market places'. These arise when digital platform providers (such as Google, Uber, Facebook and Amazon) match consumers to suppliers (where the suppliers are sometimes the platform providers themselves) via a shared platform and use automated algorithms to set the prices that suppliers can charge and consumers can pay on that shared platform. These constructed marketplaces often include other non-price based scoring systems (e.g. rating or ranking systems) that match consumers to suppliers. As such, the entire experience of being a consumer or supplier on this marketplace (and it is important here to think back to the 'users' imagined by Brin and Page), is mediated through an automated algorithmic governance structure. This structure tries to automate the knowledge accumulation and signaling needed for the marketplace, rather than relying on traditional feedback loops. What is particularly interesting about some of these new forms of algorithmic market governance is that they attempt to recreate, in automated form, the price mechanism beloved by Hayek, or more importantly, they attempt to recreate a *semblance* of the Hayekian model. While markets are not, and never have been, perfectly 'free' in a Hayekian sense, algorithmically mediated auctions and bidding processes, as well as algorithmically-controlled price-setting tools, are being used by some of the leading digital service providers, and their benefits are often touted and defended in Hayekian terms, suggesting that there is no tension between this automated algorithmic governance and free market governance.

One of the main questions I try to examine in this chapter, is therefore a direct response to the challenge presented to me in Galway. Does Google AdWords, as an automated, algorithmically constructed marketplace, perform the same kind of knowledge accumulation and communication functions as the Hayekian price mechanism? And further to that, and to delve into the political economy of Google and its advertising platforms, is there any tension between algorithmically-mediated market governance and the classical, Hayekian form of market governance?

This chapter argues that there is considerable tension between these modes of governance. Algorithmic price-setting mechanisms do not (and probably cannot) share the benefits that Hayek claimed for the price mechanism. On the contrary, algorithmically constructed price mechanisms suffer from many of the same limitations and distortions that Hayek lamented in centrally planned forms of governance. Importantly, this argument therefore critiques a particular rhetorical defence of algorithmically constructed markets and not the reality. I

am not presuming that markets are or can be perfectly Hayekian, or that the algorithmic model is akin to the Hayekian model. Instead I am questioning the quasi-Hayekian rhetoric that has grown up around these systems as justification or as a method of abdication of responsibility or accountability, and am doing so in order to critique the power that such perceptions might engender when touted, encouraged or facilitated by companies such as Google. This is what I mean by algorithmic governance in this context; it is the power created by the artificial construction (or even co-option) of the idea of free market governance as a justification for the apparent 'wisdom' of algorithmic price mechanisms. In this way, flipping the rhetorical strategy on its head and critiquing these mechanisms in Hayekian terms, I argue that Hayek's knowledge argument (as presented to me as a challenge in Galway) can be turned against those who defend these systems in a Hayekian way.

6.2.2 Hayek's argument

Hayek's knowledge argument starts with the observation that 'planning' is essential to any economy. Decisions (plans) have to be made about what goods and services to produce, how to produce them and who to produce them for. The question is who should do the planning (Hayek 1945: 520). When Hayek was writing, the answers to this question divided into two camps. There were those who thought governments should do the planning (Lange 1936; 1937) and those who thought that individual actors, free from government interference, should do the planning (Von Mises 1920). Hayek sided with the latter, feeling that the planning problem could be distilled down into a knowledge problem (Hayek 1945: 519-520). There was lots of data (information) available to putative economic planners, but not all of that data was interpretable or meaningful (i.e. counted as 'knowledge'). The critical question, therefore, was not who was best placed to make use of existing data on resources, productive processes and consumer preferences, but rather who was best placed to discover and translate data into knowledge, and communicate this to others.

The problem, according to Hayek, was that the knowledge required to plan effectively was dispersed throughout a network of actors and could only be discovered and rendered meaningful through market-based transactions:

> The economic problem of society is a problem of how to secure the best use of resources known to any of the members of society, for ends whose relative importance only these individuals know. Or, to put it briefly, it is a problem of

> the utilisation of knowledge which is not given to anyone in its totality (Hayek 1945: 519-521).

To be more precise, Hayek argued that the knowledge required to solve the planning problem had three distinct properties (see Bronk 2013; Hayek 1945):

> Discreteness: It did not come in a single package; it was dispersed among many different actors, who had partial access to the totality.

> Tacitness: It could not always be easily articulated or codified; it rested on practical know-how and subconscious, tacit understandings of how productive processes worked and how consumers behaved (see also Polanyi 1961).

> Subjectivity: The values and preferences of consumers and producers were subjective - known ultimately only to themselves.

Once these properties are appreciated, Hayek's knowledge argument in favour of free markets is easy to make. Hayek claimed that central planners were ill-placed to acquire and communicate the forms of knowledge needed to solve the planning problem. They could not hope to amass the discrete perspectives of many individuals into a single coherent plan for the market; they could not codify and articulate the tacit knowledge that underlay many market processes; and they could not hope to know the minds of the market actors. He thought that free markets could do a better job. Free markets had a wonderful knowledge accumulation and communication device at their heart: the price mechanism. The discrete, tacit and subjective knowledge of market actors could be translated into prices. These prices would tell people which goods and services were worth producing and supplying, and which were worth buying. This would enable market actors to coordinate towards mutually beneficial ends. Any interference from a central planner with the price mechanism would necessarily disrupt and distort the knowledge accumulation and communication functions that the price mechanism performs, sending people off in the wrong direction, leading to the over - or under-production of vital goods and services.

The question I explore in this chapter is therefore whether an automated algorithmically-constructed price mechanisms used by digital platform providers (such as Google's AdWords) can perform the same coordinating and accumulating functions as imagined by Hayek. Although Hayek's argument was focused on the economy as a whole, not on particular markets within the economy, it is still a useful model for understanding and critiquing

the claims made on behalf of these specific algorithmically-constructed marketplaces. As I will go on to argue, despite being only a section of the economy, Google commands a near monopoly of the web search market, and is increasingly recognised by users and regulators as being the only search engine of significance. This perception of Google as the only relevant search engine (and therefore the only market) has been perpetuated by the SEO industry and by Google itself from the start, and indeed from the beginning Google has defended itself and its algorithmically constructed markets in quasi-Hayekian terms, making a Hayekian analysis an incisive and informative method of critique with which to pick apart the wider political economy of Google's search and advertising platforms.

6.2.3 Algorithmic governance

I have mentioned algorithmic governance in the previous chapter in relation to the political power wielded by Google by its dominance of the internet, but here I want think a bit more about how that power is contextualised within a wider history of political economy which is now necessarily complicated by the advent of algorithmically constructed markets. Of course, technically, all governance related decision-making is algorithmic; i.e. decisions are made on the basis of the input and output of information, but what I am talking about here are forms of algorithmic governance that are made possible by modern information communications technology, specifically Big Data systems which work via a combination of mass surveillance and collation of data from networked technologies (computers, phones, smart devices etc.) and data-mining (descriptive and predictive analytics). The algorithms that power such systems rely on statistical analysis and statistical learning rules, as I explored in terms of online search and translation tools in Chapter 4. These systems are increasingly familiar, with virtually every internet company or service provider taking advantage of them in marketing and selling to consumers. Governments are also taking advantage of them in making risk-related decisions, such as who to audit for tax purposes, who poses a terrorist threat, or who should be released from jail (O'Neil 2017).

In his study of the rise of these contemporary algorithmic governance systems, which he terms 'algocratic' systems, A. Aneesh (2006; 2009) contrasts them with pre-existing governance structures. In particular, he contrasts them with market-based governance structures and legal-bureaucratic governance structures. The former are characterised by their use of the price mechanism to 'govern' human behaviour (i.e. to nudge it in particular directions, to structure and constrain its scope); the latter by their use of rules

and laws. He argues that algocratic governance constitutes something new: the use of computer-coded architectures to govern human behaviour. These new systems then work alongside the old governance structures, sometimes being grafted on top of them, and sometimes complementing them (Danaher 2016), creating multiple interlocking and overlapping layers of governance in human life.

The critical question is: are these newer modes of algorithmic governance compatible with market-based governance? Can you simply recreate the virtues of the market through an automated, algorithmically constructed price mechanism? Or is this simply impossible? The answer might seem obvious if we consult the historical record. The early use of automated systems of algorithmic governance seemed to be very closely-aligned with centrally-planned, bureaucratic modes of governance. In many ways, the quantitative, statistical models used in algorithmic governance structures are the bread and butter of the central planner, as Hayek himself pointed out (1945).

Furthermore, if we look into the history of socialist governance we see some obvious attempts to use information communications technology to solve the knowledge problem that Hayek identified and to enable socialism to flourish. The clearest example of this is in the Cybersyn project run by the Allende government in Chile in the early 1970s (Medina 2011; Morozov 2014), which was designed to be 'a real-time control system capable of collecting economic data throughout the nation, transmitting it to the government, and combining it in ways that could assist government decision making' (Medina 2011: 3).

All of this suggests that the history of algorithmic governance is positively un-Hayekian, missing the benefits of the free market's price mechanism that he endorsed. But the historical tide has now started to shift. In recent years, several algorithmically constructed price mechanisms have come into operation, and while market actors have been using algorithmic systems of calculation for years, what is important is that these newer systems are trying to replicate the benefits of the free market within new structures of algorithmic governance. This practice is defended on the grounds that it finds the most efficient 'market-clearing' price for a good or service. Thus, for example, Uber mediates the price of ride-sharing between the driver and passenger, or Google sets up algorithmically mediated auctions to sell words to potential advertisers. These auctions include a price-setting mechanism that tries to replicate what happens on a free market. Indeed, and despite

Brin and Page's original intentions to keep their search engine untainted by market forces (Brin & Page 2012: 3832), Google has defended its use in explicitly free market terms, claiming that a competitive auction leads to efficient pricing (Kordestani 2008).

The rise of such algorithmically-constructed pricing mechanisms suggests that Hayek's original suspicions about quantitative, statistical models, and the historical association between cybernetic control systems and socialist forms of governance, are misplaced. If we are to follow the logic of these more recent efforts at algorithmic market governance, it would seem that the use of automated price-setting mechanisms, powered by big data algorithms, can actually help to perfect the knowledge accumulation and communication functions identified by Hayek, at least within discrete markets. In which case, the challenge given to me in Galway may have been valid. Algorithmically constructed markets such as Google AdWords, stitched together through a Big Data infrastructure, might well be an ideal form of the Hayekian market.

But is this really the case? Can we really perfect the information processing powers of the market through our modern technological infrastructure? I argue that this is unlikely (see Thornton & Danaher forthcoming). Algorithmic price mechanisms cannot accumulate and signal the kinds of knowledge that Hayek felt were essential to the success of free markets. Indeed, a proper understanding of Hayek's argument provides the tools for dismantling the claims made on behalf of such systems. The next section explores the genesis of AdWords and the market-based rhetoric which still surrounds it today, before digging a bit deeper into the often opaque algorithmic systems and policies which involve distortions to the price mechanism that are much closer to what we would expect from centrally planned economies than decentralised free markets, and which necessarily involve compromises that block the knowledge-generating virtues touted by Hayek. The final section will examine what knowledge is produced by the AdWords market, and what insights that gives us into the political economy of Google and its role within contemporary society.

6.3 Google's AdWords market

In his 2014 paper 'Linguistic Capitalism and Algorithmic Mediation', Frederic Kaplan suggested that with their search-based advertising platform AdWords, 'Google has created the first global, real-time, and multilingual linguistic market' (Kaplan 2004). First

launched in 2000, Google AdWords adopted its current price-per-click auction format in 2002, controversially borrowing its business model from early rival search company Overture (formerly GoTo.com), whose founder Bill Gross was a firm believer that the market provided all the information needed to set the price of advertising keywords. Interviewed in 2005, Gross explained he had always been confident that 'bid prices would increase to their true value over time... as they have for mesothelioma, a rare cancer that- in a gruesome twist of capitalist fate- affords a high chance of recovering damages in a lawsuit' (Battelle 2011: 113).

As explained in previous chapters, AdWords operates on an auction model, whereby advertisers bid on the keywords most likely to attract customers to their adverts. Each time someone searches for a keyword on Google, a mini-auction takes place, and the advertiser who wins the auction has their advert displayed in one of a number of ranked spots at the top or bottom of the search page, made visibly separate from the non-paid organic results by a small Ad box next to the paid result. The winning advertiser pays Google one cent more than the second highest bidder every time someone clicks on the advert. In this way, the AdWords system supposes that 'every word of every language has a price, that fluctuates according to market laws' (Bruno 2012: 144).

Despite Brin and Page's initial reservations about advertising corrupting the purity of their search engine, Google soon adopted a quasi-Hayekian rhetoric to defend this model, claiming that the AdWords auction provides a platform for free market competition which is 'by far the most efficient way to price search advertising' (Kordestani 2008). Distancing themselves from potential accusations of anti-trust and price-fixing over a potential ad-sharing deal with Yahoo in 2008, senior business officer Omid Kordestani confirmed that 'Google does not set the prices manually for ads; rather, advertisers themselves determine prices through an ongoing competitive auction' (Kordestani 2008). AdWords can therefore appear to be set up as the most efficient producer of 'commercial information' (Kordestani quoted in Battelle 2005), gathering signals from a distributed marketplace in order, as Google's then CEO Eric Schmidt put it 'to provide a platform that mediates supply and demand for pretty much the entire world economy' (in Battelle 2005: 248). The AdWords auction can therefore be seen as a clear attempt to instantiate Hayek's vision of the marketplace in an automated, algorithmically-constructed form (Mirowski 2009, 11).

The Hayekian narrative also persists in some areas of academic literature. In the US in particular, where several court cases have arisen over potential trademark infringements when rival companies purchase keywords relating to their competitors - or even company/brand names - in order to divert custom towards their own sites, some of the more libertarian-tinged literature in law and legal studies [2] advocates the benefits of AdWords as an unregulated leveller of the advertising playing field (2012: 157). In a paper entitled 'It's Google's world and we're just clicking in it', Aferiat argues against 'excessive intervention' in the AdWords market (157), arguing that courts deciding trademark laws should 'refrain from stifling the sponsored links and allow these innovative, comparative advertisements to expand consumer choice' (187). Some go even further in their defence of the 'free' advertising market. Ashley Tan's 2010 article on AdWords and trademark infringement contends that 'AdWords transcends territorial boundaries in a manner that makes it an extremely effective vehicle for liberalizing international trade' (500). The Hayekian rhetoric in article is remarkable. Indeed, it is literature in this area that seems to mirror most closely the challenge I received in Galway. Tan continues:

> AdWords has become a private sector experiment in applying free trade policies to a truly global marketplace... it gives consumers access to the widest available range of providers of services and goods while drastically reducing the barriers to new market entry for foreign firms (2010: 504).

Even in what might be viewed as more critical literature, the organic feedback and knowledge producing capacities of Google and its platforms are sometimes assumed without much critical depth. Although his analysis is not quite that simple, and he does acknowledge the effect of ad-ranking on the linguistic market, Kaplan's original description of AdWords is of a system that 'relies heavily on the blind mechanisms of so-called collective intelligence' (2014: 58). In his book 'The Stack', Benjamin Bratton refers to 'geolocative advertising schemes' such as AdWords as 'real-time Cloud-based user-response driven systems' (2016: 255) without really thinking about the different 'users' involved in this

[2] The Chicago School philosophy still has a firm foothold in the field of internet economics and regulation, for example Bork and Sidak's 2012 article 'What does the Chicago School Teach about Internet Search and the Antitrust Treatment of Google?', which defends Google against anti-trust accusations, complaints about monopoly and unfair practices on the grounds that its competitors simply arent trying hard enough. The article, which was commissioned by Google, claims to 'demonstrate competitors' efforts to compete not by investing in efficiency, quality, or innovation, but by using antitrust law to punish the successful competitor (663). Bork and Sidak claim that the Chicago School (backed up by the Supreme Court) teaches that antitrust laws are to protect consumers, not competitors, therefore they are not relevant (as they claim the critics are all competitors, not consumers). 'Penalizing Google's practices as anticompetitive would violate that principle, reduce dynamic competition in search, and harm the consumers that the antitrust laws are intended to protect'(663).

apparatus, or their differing motives and levels of agency.

Although he does go on to suggest a tension between the free market and centrally planned economic models of digital platforms (2016: 330), Bratton also seems to accept a little too readily that any knowledge (and therefore value) produced by these systems cannot in itself be part of a system of control:

> Google's cloud economics draws on capitalised cognition, networked value production, and incremental value accumulation and reappropriation. In this sense, it not foremost an apparatus of surveillance and control per se but a medium for the capture and transformation of living, thinking, and knowing into platform value (2016: 138).

As this chapter will go on to show, the distinction Bratton draws between the knowledge produced by these new marketplaces (and AdWords in particular), and their capacity to control and surveil are not mutually exclusive. Indeed, the way the market is manipulated actually manifests as a new kind of algorithmic governance, mediating and manipulating flows of communication and information in opaque, yet significant ways.

Of course, for Google AdWords to be seen as having any kind of 'governing' capacity, its authority and dominance would in effect have to be total, and to truly satisfy the Hayekian model - and indeed to validate the critique - the AdWords market would have to be the only market. As previously discussed, there are alternative search engines for advertisers and consumers to use, and other means of advertising. But despite these other options, and indeed - as discussed in Chapter 5 (Language in the Age of Algorithmic Reproduction) - despite knowledge of the compromises and drawbacks, Google is often - and increasingly - treated as 'the internet', the gatekeeper of information, and in one way or another, an indispensable part of most people's daily lives. More specifically, Google AdWords is increasingly seen as the only player in the field, and indeed in the wider global economy. Tan's article on trademark infringement takes the governing capacities of Google AdWords even further, hailing the platform as a 'truly global marketplace with the potential to advance trade liberalization far beyond any governmental or even intergovernmental action could achieve' (2010: 508).

A recent US court case illustrates how this vision of Google AdWords as global regulator has also become consolidated in policy circles too. In this respect, Google's algorithmic

price-setting systems have become ubiquitous and normalised, and their complete control over the market is increasingly held by legislators and governments as providing conditions positive for fair competition. In the case between Google and optician company '1-800-Contacts', official documents show how a perceived 'army' of Federal Trade Commission lawyers argued that Google was being unfairly harmed by the contact lens company 1-800-Contacts 'fixing' AdWords keyword search results by collaborating with other companies. As the FTC's pre-trial brief states:

> in addition to not being able to serve up a large volume of potentially relevant advertising, these artificially-imposed restraints hamper the search engines' ability to learn by analyzing what users are choosing to click on (or not to click on) (FTC 2017, 22).

So not only did the FTC conclude that this was financially damaging to Google, but crucially, that by 1-800-Contacts meddling in the market, Google was being denied access to the information needed to provide consumers and advertisers with the necessary decision-making information. The impression given here is that Google has some kind of assumed right, or an existential authority over an un-hampered market, at the expense of any other actor. Google AdWords thus becomes the manifestation of Hayek's theory, but in a digital age.

6.3.1 The (un)wisdom of algorithmic markets

But despite the examples above, the reality is that AdWords is not a decentralised producer of unhindered knowledge facilitating a level and fair market, and is in fact full of the 'artificially-imposed restraints' deemed so undesirable by the US trade authorities . It is in fact a distorted marketplace where Google's software engineers interpose their own knowledge and ideology between the advertisers and the consumers. The bid prices and auction-winning prices are not reflections of discrete, tacit or subjective knowledge about the value of certain words. AdWords advertising dominates the ad-market today, and is Google's main source of revenue. The success of the platform lies not purely in a 'highest bidder wins' formula. The potential for adverts to win auctions (and therefore the top ad slots), or indeed to appear on the results page at all, will depend on their perceived quality and effectiveness. As well as the bid price, adverts are given algorithmically-generated quality scores to determine what is known as their Ad Rank. A poorly performing advert will not necessarily be shown, even if its keyword bid won the auction. If it fails to attract

enough clicks, it is not cost effective to host the advert, and more importantly, in free market rhetoric, it has not provided a good enough service; it has failed to read the signals from consumers. This sensitivity to distributed consumer feedback might seem like a self-regulating mechanism reacting to the knowledge produced by the market, but before the decision on whether to click on an advert or not reaches the consumer, the process has already been heavily mediated by the quality ranking algorithms, and by Google's own internal policies. The price is not a pure reflection of distributed knowledge.

Far from being a self-regulating decentralised system, Google is constantly changing the goalposts in its constructed marketplace. Trying to keep on top of Google's evolving advertising systems has become an industry in itself. Search Engine Optimisation (SEO) experts are in a constant struggle between penalisation for attempting to 'game' the system, and remaining visible amongst the search results. The rules of the SEO game are often modified or tweaked with no warning, and at great expense to advertisers, whose revenue streams can be drastically reduced when their adverts effectively disappear overnight. SEO experts recount how in the earlier days of AdWords, if there was no competition for the word you wanted to bid for, then you could win your advertising place for pennies. But this policy changed, and as well as negotiating the often opaque algorithmic Ad Rank score, which can be seen as a way to artificially inflate prices while hiding behind a hidden formula and the rhetoric of the market, there are artificial base rates applied to certain top spots on the results page which means that some low bid ads 'simply do not show up - even if you are bidding against nobody' (Wall 2008).

Due to the opacity of the AdWords algorithms, it is of course almost impossible to empirically prove anecdotes such as these, but they serve to provide an important insight into how the narrative of free market competition pervades the SEO industry. It is also clearly a narrative that Google itself is happy to cultivate. Many SEO forums contain discussions expressing outrage at any suspected artificial distortions of this perceived market, including accusations of cost per click (CPC) inflation caused by Google raising lowest bid prices in the face of increasing competition for finite linguistic resources.

Apart from potential internal manipulation of bid prices, certain keywords are removed from the marketplace entirely through Google's ethical policies on censorship. As Google co-founder Sergey Brin was quoted as saying, '[w]e dont try to put our sense of ethics into

the search results, but we do when it comes to advertising' (Sheff 2004, quoted in Diaz 2008, 23). Google's current AdWords guidelines state that:

> we value diversity and respect for others, and we strive to avoid offending users with ads, websites or apps that are inappropriate for our ad network. For this reason, we don't allow the promotion of any of the following:
>
> - hatred; violence; harassment; racism; sexual, religious, or political intolerance or organisations with such views
>
> - content that's likely to shock or disgust
>
> - content that's exploitative or appears to *unfairly capitalise* at the expense of others [my emphasis] (Google 2016).

Of all the above opaquely and artificially-imposed restraints, perhaps the most pertinent is the banning of content which 'appears to unfairly capitalise at the expense of others' - the precise meaning of which is unclear, but certainly does not indicate that this algorithmic price mechanism would be able to collate the information necessary for the creation of true distributed knowledge.

So, far from a 'free for all' marketplace, the mechanisms of Google's advertising are closely controlled and monitored. Google's 2015 year-end US Securities and Exchange Commission report details how the company has been 'removing hundreds of millions of bad ads from our systems every year [and] closely monitoring the sites and apps that show our ads and blacklisting them when necessary to ensure that our ads do not fund bad content' (US Securities and Exchange Commission 2015). The list of prohibited keywords is confidential. Far from levelling the market playing field with such knowledge, according to an AdWords engineer, 'having it public would make it too easy for people to work around our policies' (Google 2012).

6.3.2 Google AdGrants

Another apparent distortion of the linguistic market can be seen in Google Ad Grants, a scheme that allows non-profit and charitable organisations to use up to 10k dollars a

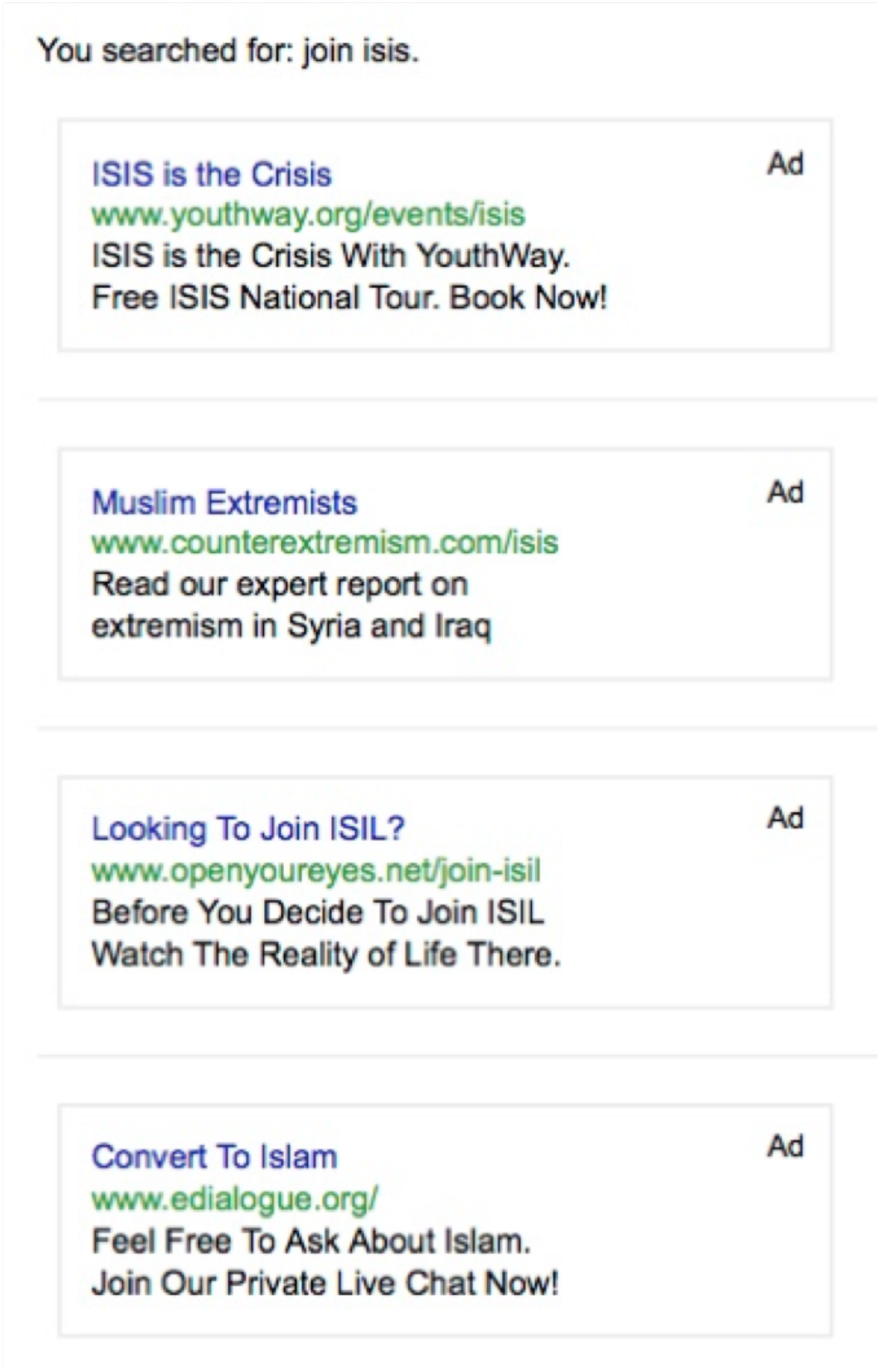


Figure 6.1: Adverts returned for a Google Search for 'join isis', 15 December 2016. Screenshot: author's own.

month worth of AdWords. In effect, the scheme enables small groups and charities to enter into the paid ads market by 'fixing' the prices. The words they can buy are capped at 2 dollars per word, meaning that they often lose out to big businesses competing for the same word. If the AdWords used against the grant are successful, the click through rate is capped at 2 dollars, but if a commercial ad enters the market and bids over the 2 dollars then they will win the auction and pay Google the click through fee as normal. Although Google frames its Ad Grants project as a philanthropic enterprise, its distortion of the AdWords market is potentially huge. In 2013 Google's UK Head of Public Policy Sarah Hunter (HAC 2013) told UK parliament that '[b]y the end of 2012 we had donated over 33 million dollars to over 11,000 UK charities through giving them this free advertising'. The wording is misleading here. There is no actual upfront donation of cash from Google. The 'donations' are in kind, and as such are artificially constructed market influences.

Despite Google's restrictions on political ad content, Google Ad Grants are also actively encouraged for use by non-profits to fund an anti-extremism agenda. As Google's head in Europe, Dr. Anthony House (HAC 2016) told the UK Commons Home Affairs Select Committee in February 2016, '[w]e offer Google AdWords Grants to NGOs so that meaningful counter-speech ads can be surfaced in response to search queries like *join Isis*'. Figure 6.1 shows some of the questionable AdGrants funded advertisements that were returned when I searched for the phrase 'join isis' on 15th December 2016.

The scheme, which works the same way as commercial AdWords, with the 2 dollar cap, has been criticised as 'a precedent-setting maneuver that permits the company to tailor its results based on perceptions of users' ideologies' (Knibbs 2016). As well as adverts served in search results through AdWords, Google's innovation centre Jigsaw has also recently been behind a scheme it calls 'The Redirect Method' (Jigsaw 2015), for which they selected [3] 1700 keywords which might indicate intent to radicalise when searched for (see Figure 6.2). During the eight-week trial in 2015, amongst the returned results were adverts which redirected users to (Google owned) YouTube videos, carefully selected from existing YouTube content to provide a 'counter-narrative' to the perceived intent of the searcher. As well as the distinct tinge of Orwellian thought control (or even a Clockwork Orange, see final chapter for an Orwellian provocation), the Redirect Method also defends criticism of their manipulation of the market and potential discrimination against certain

[3] And presumably bid for- although it is not clear if Jigsaw's involvement in this scheme allows the Redirect Method to bypass the auction process, or receive 'free' words.

Figure 6.2: Front Page of the Redirect Method handbook (Jigsaw 2015).

opinions with a curious neoliberal logic. Their method, they say on their FAQ page:

> is the same tactic that businesses use to advertise to consumers. [W]hile this method fuels a trillion-dollar ecommerce business, it's hardly been used as a tool to provide alternative messages to people who are looking for extremist content online (Jigsaw 2015).

In Google's algorithmically constructed marketplace, therefore, it seems it is acceptable to harness the power of the market to influence ideologies as well as consumer choice. These are, in effect 'centrally planned' schemes that actively bypass market mechanisms and prices. Some of Google's interventions might be laudable from a policy perspective, but by presuming they know better than the market, we end up with a mode of governance that is actually far from the Hayekian model. What we have now is quite literally a 'marketplace of ideas', which, far from the liberatory roots of the phrase, becomes quickly oppressive, fascistic (see Chapter 5), and even dystopian (something I will explore in more depth with an analysis of Orwell's Nineteen Eighty-Four in Chapter 7).

6.4 Political advertising: AdWords and the dementia tax

I want now to turn to the second of the incidents that guide this chapter, which I think provides a fascinating insight into the political economy of Google AdWords, and how the system works, and is worked, by a host of actors.

On the morning of Monday 22nd May 2017, just over three weeks before the UK Parliamentary General Election, Google search users began to notice and report that Conservative Party adverts were appearing at the top of the search results pages (see Figure 6.3). The adverts were being served against the keyword 'dementia tax', and appeared to be straightforward AdWords purchases. Dementia tax was the (derogatory) label used by the Labour Party and other critics to describe controversial social care reforms proposed by Theresa May's Conservative party in their 2017 election manifesto. The proposed policy had been to apply similar care charge rules to people in residential care as well as those being cared for at home. The existing policy means that people in residential care have to pay for their care if they have assets of over £23,250. The application of this rule to people being cared for at home, often for long periods of time with ailments such as dementia, meant that the value in their homes could be used to pay for their care,

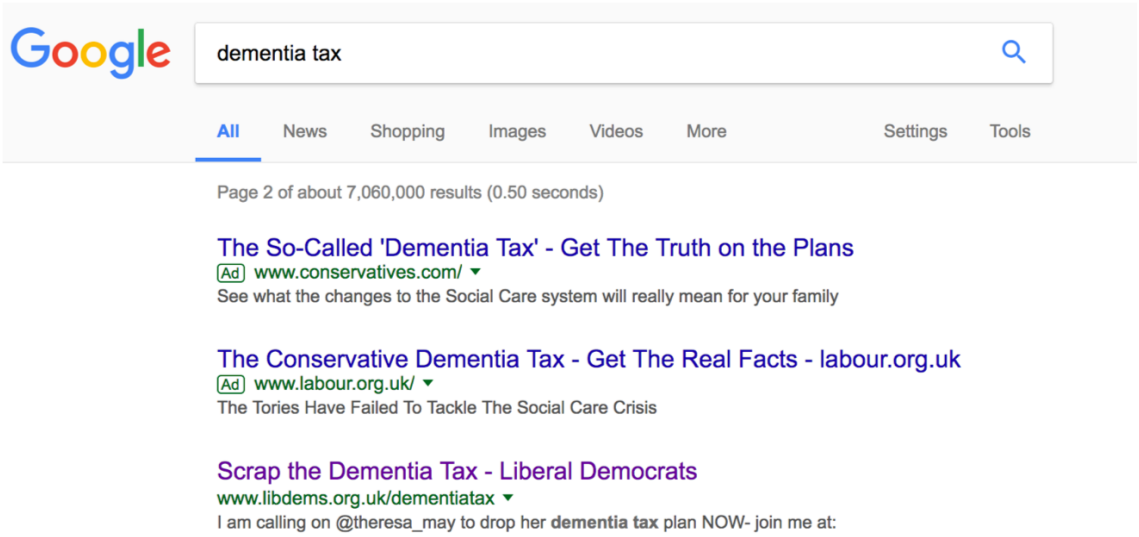


Figure 6.3: Google search for 'dementia tax', 22 May 2017. Screenshot: author's own.

meaning the amount of money left to pass on to further generations would potentially be reduced (although the Conservatives did raise the protected asset bar to £100,000. The perceived targeting of long-term conditions such as dementia, sufferers of which are often cared for in their own homes, is what led to the term 'dementia tax'. It was such a contentious and divisive issue in the run up to the election that it was eventually scrapped.

In the backlash that followed the announcement of the proposed policy, the term dementia tax became a huge PR problem for the Conservatives, and was utilised by the opposition and in the left-wing press to evoke the apparent cruelty of the proposed reforms. What is interesting to point out here is that the term 'dementia tax' was not phrase that the Conservatives used about their own policy, so it was somewhat ironic that it was that precise phrase they decided to bid for via the Google AdWords auction in order to secure the dominant position on the search engine results page. Although the phrase is qualified in the advert and presented as the 'so-called' dementia tax, it was a keyword the Conservative Party apparently had no problem embracing, harnessing and indeed purchasing in order to exploit its commercial capabilities on Google. A Tory party spokesman defended the tactic, releasing a statement declaring that 'It is quite right we take steps to tackle the misinformation and fear being spread by Jeremy Corbyn and the Labour Party' (Huffington Post 2017). Google AdWords in this instance became an overt means of taking control of the political narrative, and what was an ostensibly commercial marketplace became the vehicle of a direct attempt to influence voters. Political advertising in spaces that normally carry commercial adverts is of course not a new thing. Perhaps the most

famous example in UK politics was the Conservatives 'Labour Isnt Working' campaign by Saatchi and Saatchi which adorned normally commercial advertising billboards, and was credited with bringing Margaret Thatcher to power in 1979. Likewise today, political parties pay to take out adverts in newspapers and TV slots, either as adverts, or as party political broadcasts. But political advertising in a digital context has proved a different matter. As well as the unprecedented scale and reach of digital advertising, social media platforms such as Facebook have facilitated the micro-targeting of voters based on online profiling and the collection of personal data. These are processes so opaque, covert, and potentially illegal, that since the current scandal around political advertising in the Trump and Brexit campaigns was revealed, they have prompted international investigations into voter fraud, unprecedented fines for those involved, and the demise of companies such as Cambridge Analytica. Like Facebook's unique eco-system of clicks, likes, comments and shares, all of which generate billions of advertising revenue, Google AdWords, linked as it is intrinsically with the Search Engine, has also become a complicated and opaque part of not only the digital economy, but also a ubiquitous part of the day to day existence of digitally connected society.

Because these platforms rely on their 'users' to generate revenue through interaction via Facebook or the Google search engine, these everyday social interactions, facilitated by the revenue chasing tech companies by means of their platforms and apps, necessarily become infused with the residue of the vehicle on which this capital circulates, be that an advert, a fake news article, or a political campaign. Digital social relations therefore cannot exist in isolation from the political economy of the platforms that created them. This is part of what Jonathan Beller has called computational capitalism. As he puts it, the 'dominant economy', in whatever form of digital capitalism, 'is now parasitical on its own metabolism. The social revolution in planetary communicativity is being farmed and harvested by computational capitalism' (Beller 2018). Indeed, in an age of digital capitalism, the facilitation of communication and information exchange (via social media, search, email etc) is inherently and symbiotically (co)dependent on its own compromise by virtue of capital extraction. Google AdWords, its economy dependent on the dissemination of the very commodities it needs to function (i.e. words), is therefore an important example of this perhaps paradoxical (or at least precarious) form of computational capitalism, something I expand upon in Chapter 8 when I put forward the concept of 'subprime language'.

6.4.1 Winners and losers in linguistic capitalism

It might at first seem quite surprising that the keyword 'dementia tax' would be such a valuable commodity, but the Conservatives were neither the first not the last actor to bid for the phrase in the AdWords auction. Indeed, for a couple of weeks 'dementia tax' became a fascinating example of the co-dependent relationship between Google's (re)production of language via the search engine, the 'parasitical' (Beller 2018) harvesting of the economic capital produced by that language, and the ensuing competition for the political capital made possible by the AdWords market.

It was not just the Conservatives who entered the market for 'dementia tax' that day. There was a substantial media reaction to their original adverts, and it was not long before the Labour Party and the Liberal Democrats also started bidding for Google advertising space. In fact the screenshot at Figure 6.3 provides a fascinating snapshot into the politics, economics, and strategies at play in the run up to the election. Taken on 23rd May 2017, the screenshot shows a set of results for the keyword 'dementia tax'[4], showing the Conservative advert on top (and therefore probably the highest bidder), the Labour party in the second of the paid advertising spots, and the Liberal Democrats in third place, as an organic result. But although this screenshot potentially reflects the budgets of the three main parties at that particular political moment, what is perhaps more interesting, is that in order to capture the three parties like this, I had to venture to the second page of the search results. There is only a certain amount of space on the result page available for purchase via Google AdWords, as otherwise the search results would be exclusively adverts (as long as enough distinct actors were in the market for the same keyword, of course). Therefore when companies (or political parties) enter the market with new bids and fresh budgets, anyone who had previously been at the top of the search results will in effect be priced not only out of the market, but also out of the most prestigious digital real estate. The 'organic' voice of the Liberal Democrats was at that time effectively buried as it was pushed out of the front page by the mounting level of paid adverts.

Such was the media coverage of the Conservative's use of AdWords, that as well as the main political parties entering the market for 'dementia tax', privately funded adverts began to appear against a search for 'dementia tax' too. One such 'independent' advert did not reveal the identity of the person or group doing the bidding, but linked to a site

[4] As searched by me, not logged in to Google, from the UK, but using the RHUL VPN.

created that day (22nd May 2017). The politics behind the advert became even clearer on the site (www.dementia.tax), which stated it had been created 'by a voter that really doesnt want mayhem in power', along with an unflattering photograph of Theresa May. I was not able to find out any details of the owner of www.dementia.tax, or who was paying for the AdWords, but I did have more luck with another privately funded intervention. As reported in the Huffington Post (2017), Jason Scales, a 22 year old 'internet entrepreneur' from the Isle of Man also seized the moment to push back against the Conservatives attempt to control the narrative. When he saw what the Conservative Party were doing with AdWords, Scales quickly created a website with a landing page which claimed to reveal 'the truth about the Conservatives' plan to screw you', and what he called Theresa Mays 'Strong and unstable deception' in terms of her social care policy. His website then linked to the Labour Party website, thus balancing out what Scales believed to be an unfair distortion of the available information. 'If you dont have a counter to the ads to state the facts as they are, then its sending a wrong message', he told the Huffington Post. Scales' intervention was entirely politically motivated, his website describing him as 'a concerned individual on behalf of the British Electorate'. He responded to my request to speak to him via email, explaining that although he did already have a knowledge of AdWords, he had ' never used it to this affect before.'

> Bidding was difficult, initially I set an unusually high maximum bid and then I set for Google's own algo to dictate the appropriate max bid prices. One thing that is has shown to me is that the parties are spending a lot of money on it. My small funding of the campaign lasted but hours and received thousands of clicks, the main parties are still pushing it (email correspondence, May 2017).

Scales had set an upper limit of £750 to spend on his counter-information campaign, which he said ran out very quickly due to the high number of clicks on his advert. He also stated that his average cost per click (CPC) was £0.26, which would equate to 2884 successful diversions of eyeballs away from the Conservative's advert, a figure he seemed pleased with. His intervention was also widely circulated via other channels such as Twitter, as indeed was the Conservative's original campaign, with several people calling for as many clicks on their advert as possible, in a bid to break the Tory Party budget.

One immediate reaction from some opponents of the Conservatives and their proposed social care policy was to call for people to click multiple times on the advert. The AdWords platform is the original and fundamental source of the wealth and power Google

enjoys today, but advertising this way can be expensive - and ultimately unsustainable - if your advert attracts clicks that you are unable to convert into sales. As described earlier, an advertiser only pays Google the price per click PPC each time the advert is actually clicked on[5], so multiple clicking on the adverts would potentially mean a greater strain on the Conservatives budget, and a subsequent shortening of the campaign. But as explained earlier in the chapter, the AdWords system is not quite as simple as that.[6]

Although AdWords is based on an auction model, quality scoring and other algorithmic ranking factors also help to determine which bids 'win' the top spots on the search page, and it is not necessarily the highest bid that comes out top (as already explained). Advertisers can also buy bundles of ad placements for a fixed price (called PPM - Pay per Impression). This would normally be used for customers who just want their brand or message to have more exposure, and do not necessarily need people to click-through. In addition to this, Google has systems in place to detect and counter apparent click fraud, whether automated or part of a physical campaign, and there are numerous independent anti-click fraud companies too. Google AdWords is a very complicated and confusing economy, which is why an enormous multi-million dollar Search Engine Optimisation (SEO) industry has grown up to sustain and perpetuate it.

Google AdWords is a strange and opaque marketplace, but whatever its distortions, and the seemingly low cost of a click (£0.26 for Jason Scales), the bidding wars that went on for 'dementia tax' that week significantly drove up the price of the phrase. This is potentially great news for Google, who get paid the winning bid price on the pay-per-click system (which is in effect 1 penny more than the second highest bid), but, less good for the budgets of actors such as Jason Scales and the Conservatives and other parties. Published figures for the 2017 election expenditure on Google AdWords show that between 18th May - 30th June 2017, Google earned £562,152.59 from the Conservatives and £254,659.72 for a similar period (9th May - 30th June 2017) from the Labour Party's AdWords budget. Even with these incomplete figures, what is illustrated here is that the two main political parties paid Google over £800,000 for AdWords in a 6 week period in the run up to the General Election (The Electoral Commission 2016). There is no breakdown available to

[5]There are other packages that involve payment for each exposure and also payment per interaction such as a purchase.

[6]The process is in fact much more complicated than that, however. For example, the more successful click-throughs an advert generates, the higher its 'quality score' becomes, thus driving down the cost of each click.

indicate how much of this was on keywords relating to dementia tax, but it is still a huge exchange of economic capital for political capital.

There are also other actors in this market for whom the political interventions that day caused significant effects. As mentioned earlier in the chapter, Google runs an AdGrants system which gives free AdWords to charities and not-for-profit organisations. As it happened, up until the political parties and actors began bidding on 'dementia tax', the Alzheimer's Society had occupied the top spot on a Google search for the phrase, and had done so free of charge under the AdGrants scheme.

When I began to research what was happening on Google search and AdWords on 22nd May 2017, however, the first Alzheimer's Society advert with keyword 'dementia tax' appears alongside other dementia charities at the bottom of the second page of the search results, which in eyeball terms is pretty much oblivion. While a Huffington Post article speculated that 'the best place to hide a dead body is page two of Google' (2014), there have also been other more rigorous studies into browsing habits that show that people rarely scroll beyond the first page (Agarwal et al. 2011). The adverts for the 3 main political parties, however (and the Lib Dems had taken out a paid ad out at the time of this screenshot) have pride of place at the top of the first page. This economic and political hierarchy of results not only takes up valuable real estate space from the organic results, but reveals the real-time effects this politicisation of AdWords is having. The Alzheimer's Society has in effect been priced out of the market; its own narrative on the 'dementia tax' silenced by those with bigger budgets, which is somewhat ironic given that, as a representative from the charity told me, it was the Alzheimer's Society that first coined the phrase in a 2011 report, although 'we only started using it as a PPC keyword once the public debate around the term started and we decided to use PPC as a way to present searchers with informed, unbiased information about social care costs alongside the political messaging'. Just as with the other actors, the Alzheimer's Society used AdWords as a way to make their voice heard:

> As far as I know we have not bid against political parties on a keyword term like this before - I think it would be quite unlikely to happen regularly, but as it was such a key election issue we wanted to make sure we were included in the discussion and that people were getting accurate information on the 'dementia tax' (email correspondence, June 2017).

But in order to continue making their voice heard after the political parties began bidding for the phrase, the Alzheimer's Society was forced to switch from free AdGrants advertising, to using their paid account.

> We knew we'd have to use paid ads to compete against the ads from the political parties. We use our paid account for popular keywords generally for things like charity events (i.e. London Marathon places) or other fundraising search terms (email correspondence, June 2017).

A saving grace for the Alzheimer's charity budget was that although the click prices rose significantly when the political parties entered the market, their domain is ranked very highly on Google's AdRank system (as explained above) due to the longevity and quality of content of their website, which gives their site a 'high landing page score', all of which are factors in bringing down the PPC. Sites not ranked as long-standing authorities on dementia tax (such as Scales' brand new site, or the political party sites) would have had a lower AdRank score and a higher PPC.

Just as official political advertising and also revenue generating fake news articles turned Facebook into a political weapon in the run up to the 2016 US Presidential election, so the AdWords keyword system has knock on effects that did not exist in a pre-digital era. Commodifying keywords via a competitive auction pushes other advertisers out of the marketplace, and therefore out of the public discourse, in the case of Alzheimer's charities, quite literally out of sight on the third page of the search results. As mentioned before, political parties have of course always paid for advertising, but this new way of harnessing linguistic capitalism through Google AdWords speaks volumes not only about the state of digital democracy; a new fusion of politics and proprietary technology with strong and far-reaching collateral effects, but also about the nuances of individual campaigns and the political economy (and indeed the anatomy) of the search engine itself.

6.5 Can the Google AdWords market be Hayekian?

In this chapter I have shown how, despite the rhetoric of justification that surrounds Google's AdWords system, the practicalities of specific algorithmically constructed price mechanisms such as AdWords are far from Hayekian. Indeed, they will never have the same knowledge accumulation and signalling powers that the Hayekian price mechanism for three main reasons.

The first reason for this is that algorithmic price mechanisms always intermediate between suppliers and consumers, and thus have distorting effects on prices. Hayek imagined that prices were emergent functions of discrete actors working in response to local variations in discrete, tacit and subjective knowledge. Algorithmic price mechanisms might rely on mass surveillance technologies that aggregate from such discrete sources of information, but the algorithm's engineers (or the algorithm itself in the case of a machine-learning system) will necessarily suppose themselves to have some greater insight or knowledge than these local actors. As mentioned in Chapter 3 (and developed in the final chapter), Google's algorithms are tweaked and changed hundreds of times a year, often without warning or explanation, and the effect this has on the AdWords market is, as this chapter has shown, significant.

Contrary to the spirit and purity of the Hayekian model of the price mechanism, the algorithms that govern AdWords also embed assumptions about the preferences of the market actors and make predictions about their future behaviour in an effort to better tailor the prices to the market.

The second reason why the AdWords price mechanism is not Hayekian has to do with the distinction between information and knowledge, and the specific types of knowledge that Hayek's argument focused on: the tacit and subjective forms of knowledge. Big data algorithms such as Google's collect and organise objective quantifiable data. Consequently, they miss important sources of knowledge that Hayek held to be crucial to the price mechanism. They only collect and act upon information whose relevance was foreseen by the system's designers, and that can be seen and read by digital technologies. This has led to systems that omit important information and display systematic biases against certain populations (Crawford 2014; ONeil 2016; Noble 2018). Indeed, even with the power of technology, it is surely impossible to collect all sources of information needed to truly reflected the distributed knowledge of the masses. Although his ideas are distinctly unfashionable today (for many reasons), Hayek's argument had a surprisingly humanistic and anti-objective ethos to it. As mentioned earlier, Hayek saw a distinction between information and knowledge. Knowledge was a qualitative interpretation or organisation of information, and he claimed that there were specific types of knowledge that were available only to humans in certain 'circumstances of time and place' (Hayek 1945, 524). These are

forms of knowledge not easily articulated or codified and that are ultimately subjective in quality. Given that algorithmic structures feed upon objective and codifiable forms of information, it is difficult to see how an algorithmic price mechanism such as AdWords could ever accumulate and signal the types of knowledge that Hayek felt were essential to price-oriented governance.

The third reason why AdWords is not an example of a Hayekian market relates to the what kind of goods or commodities can be traded within that market. While what I have described above about the AdWords keyword 'dementia tax' in the run up to the 2017 General Election could perhaps be seen as a perfect instantiation of a competitive market in a Hayekian sense, in reality the opacity of Google's search and AdWords systems render invisible - or enclose - the information needed for the full functioning of a knowledge producing market. The other important thing to remember here is of what exactly does the 'capital' in what Kaplan calls 'linguistic capitalism' consist, or what does is represent? As we have seen above, the AdWords market is made up of a far broader range of 'products' than a Hayekian marketplace.

The commodities traded on AdWords are not necessarily goods or services, or even the keywords that led to those goods and services. What can be traded, and what therefore becomes subject to economic competition, are not products, but ideas. These are not commodities/resources etc which might (arguably) be effectively distributed via Hayekian-style market (algorithmic or otherwise). These are ideologies. With the allowance of political advertising through its AdWords auction platform, Google had created a marketplace of ideas, but a marketplace of ideas so far removed from the inclusive democratic vision of the web as imagined by the early web designers such as Tim Berners Lee, that it borders on the dystopic, a thought that will be developed in more detail in the next chapter.

6.6 Conclusion : What knowledge is being produced?

This chapter has argued that an algorithmically constructed marketplace such as AdWords can never be an instantiation of a Hayekian market-based system of governance, but rather is more of system of algorithmic governance. Indeed, we should be wary and suspicious of any claims made by proponents of algorithmic price mechanisms that tout their objective, knowledge accumulation and communication credentials. Google AdWords

represents the pricing algorithms they employ as objective, motiveless agents. They claim that the algorithm reflects the wisdom of the masses not the policy preferences of the company. But this is clearly not true: in effect the algorithms are used to implement policy preferences and ideologies. As Frank Pasquale (2016) notes, it is easy, and increasingly common, for corporations such as Google, to use both the market and the algorithm as 'excuses' for their behaviour, allowing the algorithm to masquerade as an objective regulator of individual behaviour. This creates a dangerous and oppressive precedent. As Ezrachi and Stucke remind us in their recent book Virtual Competition, the power behind algorithmic regulation is not the invisible hand of the market, but something perhaps far more insidious:

> what might at first glance be seen as competition is, in fact, the creation of a new force - the 'digitalised hand'. That hand, controlled by algorithms, determines the market price in any given market through complex calculations. It is controlled by those who seek to maximise their profits (Ezrachi and Stucke 2016, 209).

The question might then be not only whether algorithmic governance structures can ever be compatible with a Hayekian version of market-based governance, but that if we are to accept a new hybrid 'digitalised hand', what knowledge does this type of market produce? And further to this, what epistemological power is vested in whoever controls the systems producing such 'knowledge'? As Mirowski and Nik-Shah warn us, the 'cadre that gets to construct the markets gets the final say on the nature of truth'. If we are to see Google as the constructor of the linguistic marketplace, then this has serious effects on the narratives being produced by that market. Afterall, 'the visible hand that fashions the auction believes it can govern the world' (2017, 8).

In answering these questions, we should also be wary of attempts made by digital platform providers to 'perfect' or 'complete' the databases of knowledge upon which they construct their algorithmic markets. It is unlikely that such attempts will fill the knowledge gaps needed for Hayekian efficiency. Indeed, there is also an argument to say that more data, particularly data of the wrong type, produces poorer knowledge. The idea that more data is always better, despite its quality, would perhaps even have been an anathema to Hayek, who rejected the idea that 'scientific' aggregate statistics could equal 'the sum of all knowledge' (1945, 521), instead believing in the importance of localised, subjective tacit knowledge. Indeed, if we look to Hayek's suspicions of statistics and scientific ways

of trying to 'know' and 'see' everything of the central planning model, we can see not only Googles 'World Brain', or Uber's 'God View', but also echoes of Donna Haraway's feminist critique of the God Trick, or the problematic claim to objectivity. To Haraway

> objectivity turns out to be about particular and specific embodiment and definitely not about the false vision promising transcendence of all limits and responsibility. The moral is simple: only partial perspective promises objective vision... Feminist objectivity is about limited location and situated knowledge (Haraway 1988, 582-3).

Thus in both Hayek and Haraway we can see a distrust of the obsession to see everything, or know everything, which we might extend to critique the algorithmic 'god-views' of companies such as Google, Facebook or Uber, and the increasing modern obsession with big data. To Haraway, what she calls 'unlocatable' knowledge equals 'irresponsible' knowledge, and '[i]rresponsible means unable to be called into account'. But this is perhaps the main problem with algorithmic governance and its reliance on quantitative big data (and as much of it as it can get) how can it be called into account especially when it operates under the guise of market impunity?

This chapter has been a response to a challenge I received that suggested that Google AdWords should be seen as an instantiation of Hayekian free market economics, with all the wisdom and knowledge producing capabilities such a system might entail. In responding to the challenge in this way, I have used Hayek's arguments to show that far from a model of efficient market governance, Google AdWords is in fact a system of algorithmic governance (see Thornton & Danaher forthcoming).

While I have suggested that AdWords is definitely not a free market economy, ideologically speaking, it might be too simple to conclude that it is therefore centrally-planned. In this respect it is probably more helpful to move beyond such binary views and to see algorithmic governance as something genuinely new and different. As Lucas Introna and Helen Nissenbaum identified, in the early days of the internet, information was pulled in two very different ideological directions. On the one hand, a post-modern narrative saw the opportunity for the dissemination rather than centralisation of knowledge, but on the other hand, the private ownership of technology had a limiting effect (2000: 170). Even before that, there was a confused ideology underlying the development of networked technology, as demonstrated with the early cybernetic projects such as Cybersyn, and

perhaps the somewhat surprising apparent harmony between Hayek and Haraway's views on situated knowledge that I detailed above. Likewise, the advent of Web 2.0 brought possibilities of participation and inclusion, yet the privatisation of information by companies such as Google led to a 'dispossession' of the means of communication (Jakobsson and Stiernstedt 2010).

According to Philip Mirowski, the reason for ideological confusion is that computational systems and neoclassical views of markets are fundamentally incompatible. This is one of the reasons why 'there is not as yet a credible economics of knowledge' in the computational age (2009: 99).

> If the marketplace of ideas is thought to operate like a computer, and then one insists upon neoclassical economic theory as the correct and appropriate model of the market, then economists are dealing in delusion, since they regularly endow the market with capacities that no existing computer can or ever has possessed. Although it is not a popular opinion in the contemporary profession, it seems hard to escape the implication that neoclassical economics and computers just are incompatible. One may wish (as Hayek did) to portray the entire market institution as resembling a computer, but to do so, one must relinquish any commitment to the neoclassical orthodoxy (Mirowski 2009: 143-4).

So what is there to be said of a political economy of algorithmic knowledge, and specifically of Google AdWords? Is there new wisdom in algorithmic markets? Perhaps the same problems exist today as they did when Hayek originally formulated the knowledge problem, for Cybersyn style central planners and for algorithmically governed markets. Not all information can (and perhaps needs to be) codified. Just because we have the means of codifying things in digital, algorithmic systems still doesn't mean that we can get at the discrete, tacit and subjective knowledge needed to create the perfect market (even if that was our objective). More (or bigger) data does not necessarily solve this problem, and such data will in any case always be affected by the biases and motives of the companies and individuals who create and automate the processing algorithms. What Google has created with tools such as AdWords is not a system of algorithmic governance that levels the epistemological playing field and provides efficient distribution of services through the price mechanism, despite the quasi-Hayekian rhetoric on which they rely. Instead, what I have argued is that any knowledge produced through such systems will necessarily carry

the bias not only of its algorithmic structuring, but is also unreliable knowledge based on centrally planned constructed marketplaces and price setting interventions.

Despite my conclusion that Google AdWords does *not* produce the necessary knowledge for the efficient governance of the market, this does not mean that AdWords does not produce *any* knowledge, or that the knowledge it does produce might not have significant effects in the world. Indeed, the final two chapters, which detail my intervention into Google's linguistic capitalism, are based around gaining insights from the aggregated 'wisdom' of the AdWords market. The linguistic market does indeed produce knowledge based on the aggregation and algorithmic processing of information. It just remains crucially important to remember that algorithmically produced 'knowledge' will always be tainted. Indeed, far from a Hayekian ideal of the wisdom of the market, the purported 'knowledge' produced through the market mediated search engines[7], can have dangerous political and social effects, as I described in Chapter 5, and will expand upon in the following chapter.

[7]Google specifically calls search results 'knowledge', as in the 'knowledge box' that appears at the top of the results

Chapter 7

CRITIQUING LINGUISTIC CAPITALISM: PROVOCATION & INTERVENTION

Stop searching, start questioning (Geert Lovink 2008).

7.1 Introduction

In the preceding chapters I have used Walter Benjamin's 'The Work of Art in the Age of Mechanical Reproduction' essay (1936) as a frame on which to hang my arguments and observations about Language in the Age of Algorithmic Reproduction. Using Benjamin has allowed me to think about how the values of words change as they circulate through digital spaces, their ever-changing positions and agency within what I have called a geography of (con)text, and the political, economic and cultural effects that modernising technologies such as Google AdWords and digital advertising might be having today on language and the wider discourse.

As discussed in Chapter 5, central to Benjamin's essay is the observation that the extreme and fascistic politics of that era had been 'aestheticised' by its co-option of popular culture. Facilitated by the technologically enhanced reach and influence of mass media such as photography, news, advertising and art, Benjamin believed that the violence of an authoritarian regime had been disguised as an awe-inspiring show of cultural and militaristic strength, and that the population had been appeased by their apparently voluntary immersion into new, and beguiling levels of cultural awareness and participation. This

phenomena of control through technologically advancing culture is what, in an age of digital capitalism, I have also argued is a kind of new algorithmic governance.

But how - in this digital age- are these political effects of technology to be countered? Benjamin's solution to 'the situation of politics which Fascism is rendering aesthetic', was that 'Communism [should] respond[s] by politicising art' (1999: 235); a call, perhaps, to recuperate art from the clutches of the political and economic systems which use it as a means of power and control. As mentioned before, the 'violence' and authoritarian nature of the technologies that mediate, confuse and influence popular culture and the political landscape today have become more and more obvious in the time I have been researching this thesis (e.g. fake news spreading via advertising, Russian bots on Twitter etc.). As a result, and in response to Benjamin's call, this final section of my thesis is dedicated to exposing, making visible, and resisting the power and politics behind language in the age of algorithmic reproduction, by means of artistic intervention.

As discussed in Chapter 3 [methods] earlier, it is almost impossible to untangle the workings of black-boxed algorithmic systems such as Google's search and advertising platforms, but it is still possible to show their effects both quantitatively and qualitatively. I have already demonstrated how analysis of Google data can help to expose the politics behind our everyday flows of information (Chapter 6), but in this chapter, I will explain and demonstrate my critical intervention called {poem}.py; a project that exposes the workings of linguistic capitalism by creating -and ultimately framing and exhibiting - paper receipts for poetry fed through Google's AdWords platform.

The chapter will therefore explain the genesis of the {poem}.py project as a critique of linguistic capitalism, and also more broadly as a new method for the research of opaque technologies. It will follow its development from poster to commissioned exhibition; documenting some of the methodological hurdles, work-arounds and inspirations I came across along the way. It is in this respect, my answer to calls for artists 'to invent new ways to interact with information, new ways to represent it, and new ways to make sense of it' (Lovink 2008, 9). Or as Martin Feuz et al. state, 'new methods are urgently needed, otherwise the knowledge and power differentials between those on the inside of search engines and those who are mere users of a powerful but opaque machine are bound to grow.' (Feuz et al. 2011). While exposing the power and politics behind Google's search and

advertising technology, and providing a focus for some of the wider problems of researching proprietary and opaque technology detailed in Chapter 3, the project also acts as a standalone piece of politicised/political art and as an act of resistance. Poetry, forced to circulate through systems that strip it of its aesthetic value and context, is thus reclaimed from the algorithmic marketplace, repoliticised, and returned to art.

But before I turn to the intervention, I want to present a provocation. As well as tool of epistemic, geographic and colonial power, all of which are of course magnified in a digitally mediated and connected age, language also has considerable power as culture in the form of fiction, poetry, and news media. My thesis thus harnesses this cultural power of the written word, turning it against the political power of language by using literature itself as a central tenet of my critique. As I mentioned in the introduction, language has always been a tool of control over people and places in a myriad of ways. Today more than ever, when analogies of Orwell's *Nineteen Eighty-Four* and Huxley's *Brave New World* fuel the discourse around authoritarian politics and fake news, literature proves itself to be a similarly powerful force. Just as my intervention uses the power of poetry to explicate the effect of Google's advertising systems on language, in this next section I use Orwell's literary fiction as a critical lens.

7.2 Provocation

> *[I]t is clear that the decline of a language must ultimately have political and economic causes (Orwell 1968).*

Weaponised by centuries of successive invasions and colonisations, and manipulated by repressive regimes or systems of governance, language has always been a tool of power. Language in this way is always capable of potentially devastating political and societal physical and discursive effects, which gives an extraordinary amount of power to those who have control over it. This section explores the political power embedded in Google's commodification of language using George Orwell's concept of Newspeak as a provocative method of critique.

George Orwell's 'Nineteen Eighty-Four', first published in 1949, has been an increasingly popular metaphor for debates around the privacy and surveillance issues of technologies

such as Google and Facebook, as well as for the post-truth era of fake news, and alternative facts. Orwell's idea of Newspeak, the language of thought control and state propaganda employed to further the ideology and control of English Socialism (Ingsoc), is a compelling analogy for some contemporary issues, but I think rather than a straight forward comparison to the misinformation and accusations seemingly employed during (and after) the 2016 US Presidential campaign, there are deeper problems within today's informational infrastructure that a more thorough reading of Orwell's text draws out.

In an appendix to the main text entitled 'The Principles of Newspeak', Orwell imagines that:

> The purpose of Newspeak was not only to provide a medium of expression for the world-view and mental habits proper to the devotees of Ingsoc, but to make all other modes of thought impossible. It was intended that when Newspeak had been adopted once and for all and Oldspeak forgotten, a heretical thought - that is, a thought diverging from the principles of Ingsoc - should be literally unthinkable, at least so far as thought is dependent on words (2000: 343).

The idea that 'thought is dependent on words', and indeed that a government can control the thoughts of the population by controlling language and discourse is central to Orwell's critique of the totalitarian regimes of both Nazi Germany and Soviet Russia, but has also been used in relation to the 'linguistic engineering' that took place in Maoist China (Ji 2004). It is, however, not an entirely unproblematic relationship. There are philosophical objections and indeed multi-lingual practicalities that question the causal effects of words on thought. As Fengyuan Ji points out, 'banning people from using heretical language will not automatically lead to the slow extinction of heretical ideas' (2004:12). This may indeed be so, but as more and more headlines appear about the 'nudging' effects of Facebook newsfeeds on voting behaviours (Tufekci 2015); as stereotypes and prejudices are compounded by confirmation bias in auto-completions and auto-predictions, and as the minds of criminals are moulded by what they see on a Google search results page,[1] there does indeed seem to be a somewhat problematic yet entirely relevant causal linkage between the words people see online, what they think, and what they ultimately do. As Safiya Umoja Noble has shown, search engines have the power to 'reinforce racism' and

[1] I am thinking here of Dylann Roof, who murdered nine people in Charleston, South Carolina in 2015. In Roof's own words, his sympathy with the murderer of black teenager Trayvon Martin in 2012 'prompted me to type in the words "black on white crime" into Google, and I have never been the same since that day' (The Washington Post 2015; Noble 2018).

other oppressive behaviours (2013; 2018). The actionable effects of the words disseminated online through platforms such as Google and Facebook are now taken so seriously that both companies have been under considerable governmental pressure to combat fake news and hate speech. As previously explored, there are now even Google facilitated schemes to 'redirect' searches for Jihadist and extremist material online.

The other important point about the analysis of linguistic engineering through the lens of totalitarian states, is the assumption that only governmental regimes have the power to bring about unquestioned, unchallenged and mandatory changes to language. Ji writes:

> Linguistic engineering in nontotalitarian societies is not effectively controlled by the state, and even when it has political backing, people are free to criticise it and usually to ignore it. Linguistic change is brought about almost entirely by persuasion and social pressure, not by coercion, and it is often accompanied by heated debate and the persistence of rival usages (2004: 4).

While this may perhaps have been true in a pre-digital age, the near monopoly market dominance that Google enjoys today, together with the opaque, proprietary nature of its operations, and the irredressible and almost unchallengeable power that it holds, does indeed bear many of the hallmarks of a state totalitarian regime. Any potential linguistic changes that result from its advertising strategies are not open to competition or debate, nor are they even clearly visible, but are the result of a far more insidious instrument of power. As Golumbia writes, 'we should never expect schemes to regularize, normalize, segment, or even culturally control language to be wholly effective without implementing similarly effective controls on all other social behavior' (2009: 110).

The second point I want to make about the Orwell text concerns the limitations and restrictions of language that is so important to the idea of Newspeak, which was purposefully designed with

> very few words to choose from. Relative to our own, the Newspeak vocabulary was tiny, and new ways of reducing it were constantly being devised. Newspeak, indeed, differed from almost all other languages in that its vocabulary grew smaller instead of larger every year (2000: 352).

We can see echoes of Newspeak in the shrinking of the creative vocabulary of digital language in favour of the most popular keywords, which might be cheaper, easier to find, or

more easily understood to both algorithms or human readers. As discussed in Chapter 5, Search Engine Optimisation techniques encourage the re-hashing of existing content, rather than anything new, and although it could be said that SEO is in itself a creative industry, unless it is also economically lucrative, there is little value to Google in original, or creative language, as my {poem}.py intervention will articulate in the next section. The exception to this rule could be niche words such as mesothelioma, the type of cancer caused by asbestos poisoning, which is searched for relatively infrequently, yet has an extremely high suggested bid price which corresponds with the lucrative litigation industry surrounding it and the increased likelihood of a return on investment (Battelle 2011). People searching for mesothelioma are presumably not Googling for fun.

But for the millions of other words which circulate the web more frequently, and might be used in Newspeak as part of 'the business of everyday life such things as eating, drinking, working, putting on one's clothes, going up and down stairs, riding in vehicles, gardening, cooking, and the like' (2000: 344), it is their proliferation and multiplication that makes them valuable. Although it is one of the factors that might downgrade a site, SEO tactics still advocate moderate amounts of 'keyword stuffing' - the practice of filling web copy with varieties of popular keywords that might relate to the product or site being promoted. These keywords might be duplicates, synonyms, or even antonyms, as one SEO instructor once advised me. As long as the words are popular enough to direct a search to a site, it really doesn't matter.

Important here too is the control over what words can mean. Newspeak words could only be used for one purpose:

> their meanings were far more rigidly defined. All ambiguities and shades of meaning had been purged out of them. It would have been quite impossible to use the vocabulary for literary purposes or for political or philosophical discussion (2000: 345).

Likewise, when you search for a word in Google, it is likely you are being returned the most commercially viable version of that word, which is not necessarily the one you intended. For example, if you enter the word *cloud* into Google, the paid and organic results will relate to *cloud computing* rather than to less monetisable versions of the word, such as the meteorological cloud, or indeed the *lonely cloud* that Wordsworth was imagining in his Daffodils poem. As I go on to show, the meaning and value of the words passing through

Google are therefore 'rigidly defined' by their economic purposes. The poetic - or aesthetic - value of the words you may have imagined is stripped away as it passes through the portal of the search bar, only to be replaced by its exchange value in the linguistic marketplace.

Now if you take those two ideas; i.e. that words have a real effect on how we think (and potentially on what we do), yet the way Google commodifies language encourages the shrinking and restriction of our online vocabulary, then it becomes even more clear how much power there is in being in control of the language that circulates online. Perhaps most disturbing about this situation is the emergence and dominance of a new linguistic economy in which the overriding motive for the regulation of language is not state political control (as in Maoist China or Orwell's dystopia), but private capital gain. Not only do we now have a system which controls language (and therefore the wider discourse) by restricting its possibilities, and its creative potential, therefore presumably doing the same to how we think, but the system rests on a kind of neoliberal logic which is perhaps just as frightening in its scope, power and reach, than an overtly political one.

As Franco 'Bifo' Berardi writes in response to Kaplan's early work on linguistic capitalism (2011), the monetisation of language through Google's AdWords platform means that 'the economy is the universal grammar' (2012). The argument I make in this chapter is that in a digital age, it is this monetised grammar that increasingly both constructs and deciphers - or interprets - language. The way the web works at the moment, with Google mediating and exploiting the circulation of monetised words, the potential for political influence comes often as a side effect of the economic incentive; through Macedonian teenagers exploiting AdSense with lucrative 'fake news farms', or through manipulation of the SEO industry. We might say that, while concentrating on exploiting language for money, Google have in effect let money control the narrative.

As mentioned earlier, the Orwell critique of the state of digital capitalism today is not unique, but is perhaps often used a little too loosely. By paying close attention to the text, and specifically to the idea of Newspeak, I have attempted to highlight the very real dystopian potentialities of the data-isation and monetisation of language in the system of linguistic capitalism. Although the Orwell text was written in a pre-digital age, it is an incisive critique of the power of language when deployed for political or economic purposes. In the next section I continue to use literature as a method of critique, using

poetry as lens through which to make visible the dystopian potential within Google's control of monetised linguistic data. As discussed in Chapter 3, it is almost impossible to untangle the workings of black-boxed algorithmic systems such as Google AdWords, but it is still possible to show their effects both quantitatively and qualitatively. The next section will explain and demonstrate a project called {poem}.py, an artistic intervention which fuses poetry, code and data in order to make visible the political power that Google and linguistic capitalism have over people and places today, as well as exposing the potential linguistic effects of the datafication and monetisation of language.

7.3 Interventions

> *Imagine the day when a search engine will rule the whole textual content of the web, in which the memory of mankind will be stored. Think of the power in their hands (Bruno 2002).*

Whilst I first began to think about using the frame of Benjamin's 'Work of Art' essay as a result of Martha Poon's observations about the movement and replication of information by algorithms as modern day industrial transport systems such as canals (see Chapter 4 - Geographies of (con)text), from the beginning of my project I was also very aware of and interested in, the use of creative intervention to critique the workings of Google's algorithmic systems. Combining what you might call an 'economic' theoretical approach with a 'creative' approach has enabled me to investigate, and to some degree expose, the value generating movement of words (or more accurately, linguistic data) around the digital landscape. Previous chapters have covered the more logistical and economic elements of the algorithmic processing of linguistic data, but this chapter will begin by contextualising my project amongst the growing number of artistic statements and critical interventions into search technology, language, and Google's role in the informational landscape.

As mentioned in Chapter 3, part of my method in this thesis has been to harness the power of language through literature in order to analyse and critique digital technologies and tools. This is perhaps contrary to traditional digital humanities or data science methods which tend to use literature and digitised literary datasets as passive objects on which to be experimented *with* digital methods and tools. In this respect, this chapter harnesses the creative imagination of George Orwell (amongst others), in order to power my critique of the exploitation of language by Google, as well as using a range of writers

from Shakespeare, Wordsworth and Wilfred Owen, to Margaret Atwood, Billy Bragg and William Gibson in order to progress that critique. I want therefore to begin with examples of some of the similarly literary and creative interventions that use the lens of Google and its associated platforms and portals to shine a light on wider issues of (big) data, algorithmic systems, and processes of digital technologies and economies, and which have inspired my own work. It is worth mentioning that while all of these examples have been instrumental in formulating my own method, some of them have done so through their incisive, intelligent and creative critique (which I have sought to emulate), while others have helped to make clear some of the challenges (and also gaps) in using creative methods to critique digital technologies (which I have sought to overcome and build on).

As Google's dominance over the informational landscape has grown, there have been a number of academic and artistic interventions into the potential problems and implications of search technology that have helped to shape my research and to inspire my own creative project. All of them have used playful - even humourous - methods to deliver incisive critiques of the technologies that mediate our everyday lives. Feuz, Fuller and Stalder's innovative experiment into search personalisation and what they call 'semantic capitalism' (2011), consisted of feeding the indexes of the greatest works of Michel Foucault, Emmanuel Kant and Friedrich Nietzsche through the search function of separate logged in Google accounts. By creating search histories for the philosophers, the authors' idea was to see how these imagined 'personalised' profiles affected subsequent search results. The results were quite surprising, suggesting that, far from increased personalisation, search results seemed to do the inverse; reflecting the statistical group profile into which Google's algorithms have sorted the user. As the paper concludes,

> A strong interest in philosophical terms - which can be gleaned from the semantic history - could, for example, be associated with certain age and income groups, which, in turn, become associated with certain preferences in, say, holiday destinations. In such a way, Google infers Immanuel Kant's taste in hotels, or Friedrich Nietzsche's bias for or against open source software (2011).

The experiment thus very cleverly revealed that personalised search results tend to serve the needs of advertisers rather than benefitting the individual user.

I was also fascinated by a 'happening' curated by the artist Christophe Bruno into the then fairly new Google AdWords platform in 2002 (see also 2012). Bruno opened an Ad-

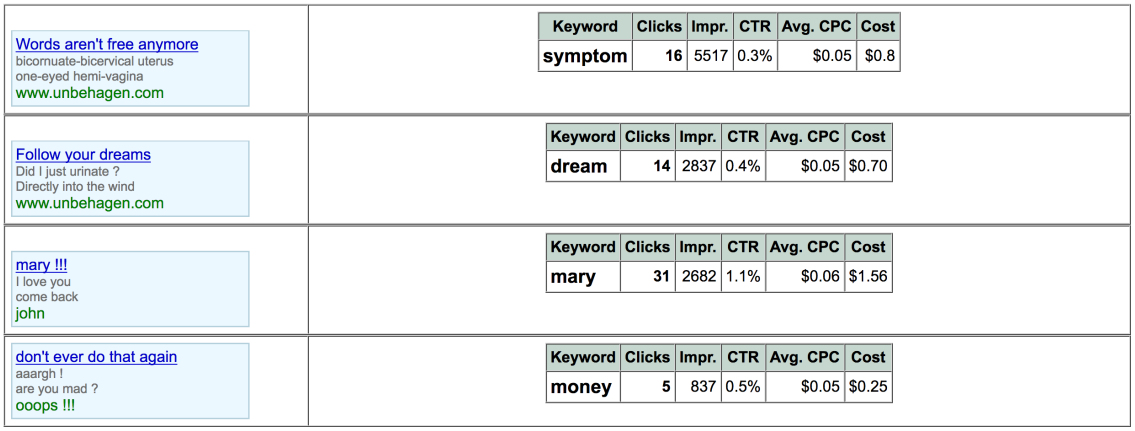


Figure 7.1: Christophe Bruno's 'AdWords Happening' (2002)

Words campaign and bid for a selection of cheap keywords. He then composed a series of adverts which, rather than contextualising the keywords and clearly trying to promote a product or service, were small snippets of poetic nonsense in which the keyword was entirely irrelevant (see Figure 7.1). What Bruno was trying to test was whether or not the people who were served his adverts after searching for one of his keywords would be curious enough to click through to his website on a whim, even though the advert clearly was not what they had been looking for; thus subverting the logic of the market in favour of poetic intrigue. Bruno has conducted similarly interventions which have aimed at 'diverting global symbolic structures like Google search engine' (Bruno Interview 2006). The 'Human Browser' is a kind of mobile installation during which an actor is fed and articulates a text-to-voice stream of search results which are generated by the environment, situation or space through which the actor moves. Bruno's point is to show how speech and sensical language has become subsumed by the ubiquity and invasiveness of the Web, and as lost its individual power and agency. In fact only now in the hijacking and subversion of the global structures of power which make up the Web, is language effectively 'effective' (Bruno interview 2006).

Cabell and Huff's intervention into Google, in which they fed the text of Brett Easton Ellis's 'American Psycho' through GMail and recorded the adverts it generated (which I discussed in Chapter 4) was also a huge influence on the development of my own critique and methods. However, while Bruno's and Cabell & Huff's interventions are perhaps what inspired my own, there are also many other creative interventions that I find interesting, but also potentially problematic. In 2015, poet Sam Riviere published a series of poems based on the online persona of Kim Kardashian, critiquing the blurring lines of

public and private in an age of blogging, social media, and online celebrity. Kim Kardashian's Marriage (2015) consists of 72 'found' poems (one for each day of the titular marriage), which he constructed using existing publicly available material (including Kardashian's own make-up tips), and then exposed to the vagaries of Google's search and auto-complete functions. The result is a series of poems 'that have been produced by harvesting and manipulating the results of search engines to create a poetry of part-collage, part-improvisation', and which proved commercially very successful for a book of poetry. Riviere's method is reminiscent of other forms of 'found' poetry, such as Sampsa Nuotio's Google Poetics project (2012). With its tag line 'Google writes poetry on subjects that people are truly interested in', Nuotios website (www.googlepoetics.com) collects drop-down lists of auto-completed Google searches that 'make sense' as surreal snippets of verse (see Figure 7.2 for my own example of this method). But while we might say they are amusing, and sometimes poignant, distorted reflections of humanity in a digital age, I think the question needs asking: what are these things doing, further than being a bit of a novelty and adding to the ever growing genre of electronic literature or creative digital humanities projects? Is this just Art for Art's sake? or - as discussed in chapter 5 - is it unwittingly contributing to the aestheticising power of Google's particular form of governance?

Yet there is so much more potential in a project like Google Poetics. For example, the site does not credit the contributors of these Google Poems, stating that they get so many similar ones that people may feel aggrieved if someone else got the credit, and because 'we don't wish to differentiate which poems are made by the curators and which by the readers'. These questions of authorship throw up such interesting questions about algorithmic agency (see also Chapter 5, where I discuss the 'Death of the Reader' and Benjamin's views on print capitalism), in terms of questions of new forms and claims of authorship in a digital age. In this way, Google Poetry belongs to everybody who has ever used Google.

As I have explained in an earlier chapter (Chapter 4 - Geographies of (con)text), although auto-completions are based on previous searches AND on the content of the web, there are many other hidden manipulations behind what comes out of the search engine, that are far more important to pursue than capturing the funniest 'poem'. By this I mean the lacunae of data on which autocomplete works, the bias in the data, Google's policing of controversial words and phrases, and of course the complication of paid advertising, all

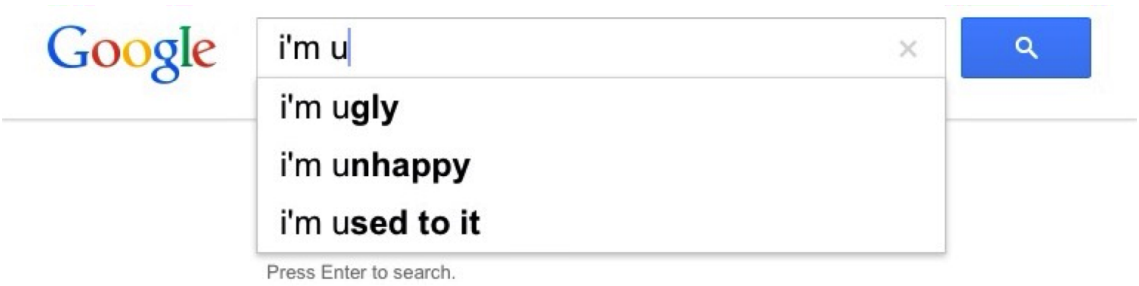


Figure 7.2: Example of Google Poetics method (Nuotio 2012), June 2014. Screenshot: author's own.

of which combine to reveal some of politics behind the search engine which I have been trying to expose in this thesis. Therefore, while 'algorithms can instigate and facilitate imagination, creativity, and frivolity, while saying something that is simultaneously old and new, always almost repeating what was before but never quite returning' (Mahnke and Uprichard 2014), and while there is undeniably something of the absurd and the poignant about the poems, there are much deeper and I believe far more urgent critiques and interventions to be made here. The type of insight Christophe Bruno exposed with his AdWords Happening poems, for example, or Cabell and Huff did with American Psycho, is - for me - a far more penetrating critique of the clash of humanity and digital technology than making poetry from search results, or trying to separate machine written poetry from that of a human.

It is, however, these types of study that seem to gain most attention in the media. In a recent study, data scientists from IBM and the University of Melbourne describe their latest attempts to generate poetry by machine (Lau et al. 2018), which they 'tested' via paid crowd-sourcing and a literature professor. Perhaps unsurprisingly to anyone with any knowledge or appreciation of poetry, the study revealed that form and structure are not the only, or the most important ingredients, in fact, the study concedes that 'our focus on form actually hurts the readability of the resulting poems.' Recognising the importance of its content, rather than the strictness of its meter and rhyme, the machine-generated poetry was then rated on a scale of 1-5 as to 'how much emotion a poem evokes', concluding that 'machine-generated poems... underperform in terms of readability and emotion', and suggesting further work in that area. To me, trying to quantify such a subjective quality as emotion, not only seems futile and distinctly unscientific, but calls firmly into question the value in applying data science techniques to humanities studies; something I have already called into question in terms of some applications of digital humanities schol-

arship[2]. Likewise, projects such as 'Bot or Not' (Botpoet.com n.d.), where website users 'have to guess whether the poem you're reading is written by a human or by a computer' and keeps a running leaderboard of 'most human/computer-like human poems' and 'most computer/human-like computer poems', a system that seems to bestow undeclared - and underexplored - binary value on language. Or 'Botnik' (2018), a platform which 'scours various types of human-created, word-crowded content - from season-three Seinfeld scripts to Yelp reviews - in order to build predictive, idiom-specific keyboards'. Users can then use these corpus specific keyboards to create 'new, inevitably askew versions of well-known works', as this 'alternative' Harry Potter cover shows (see Figure 7.3).

Apart from the fact that it is 'part of the Techstars Alexa Accelerator', and therefore funded by Amazon, projects such as these seem doubly uncritical as they seem to create algorithmically generated content just because they can. They say they are 'a community of writers, artists and developers building and using machine tools to remix and transform language', yet they give no reason why, apart from wanting to 'create strange things'. In effect, they are just mimicking predictive text based on pre-uploaded corpora of data such as 'romantic poetry' or 'harry potter'. The results are funny, but they have no bite. They are just copying commercial predictive text algorithms, and in that way are just reproducing the existing mode of production, rather than saying - or creating - anything new. 'We would like, selfishly, not to replace humanity with algorithms', their site declares, 'instead, we want to find natural ways for people and machines to interact to create what neither would have created alone' (Botnik 2018).

The more I have read about 'digital art', interventions and playful critiques which use new technologies, language and algorithmic platforms, the more I have become aware of the difference between some creative projects in terms of those that used technology to create new forms of (perhaps human-machine hybrid) art, and those that questioned what these technologies were doing to both art and humanity. The latter category (to which I hope my own project contributes) fascinates me because it raises so many questions about digital aesthetics, human agency, machine art, but also because it makes the whole subject so overtly political. It allows technologies to be pulled apart for their motives and methods, and for artists and academics to ask wider questions about what technology is doing not

[2] Another example of this disciplinary and methodological mismatch is a study which claims to be able to 'quantify the beauty of outdoor spaces' using 'online data combined with neural networks [to] provide a deeper understanding of what environments we might find beautiful' (Seresinhe et al. 2017)

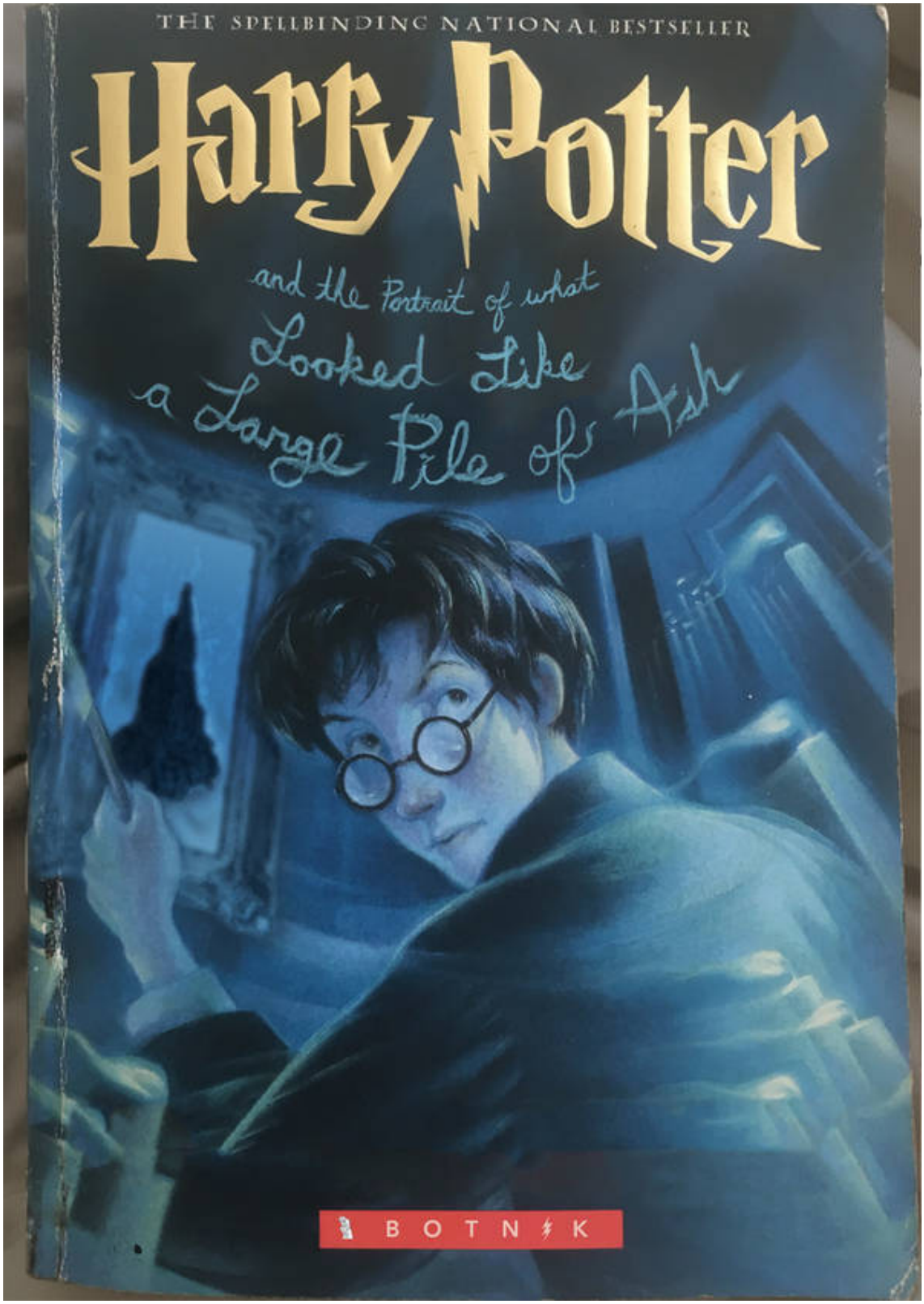


Figure 7.3: Example of AI generated Harry Potter text (Botnik.org 2018).

only to art, but to wider society and culture.

But like many current critiques of AI technologies and data science, exercises, or experiments like these run the risk of being done just because the technology is there to do it, and don't add, or challenge anything, or indeed acknowledge or engage with the politics of the technologies they use. These are at best missed opportunities, but at worst, they serve to consolidate, normalise, and even aestheticise (see Chapter 5) the problematic systems on which they are constructed. They become conveyors of advertising capital themselves as they are shared and spread on social media - giving an air of agency and control to the people who interact with and 'generate' them (again, thinking back to Benjamin). In this respect, the creative, or radical, potential of some of this work is tempered by its passive enrollment in the systems of the digital economy. As Oli Mould suggests in his book Against Creativity,

> Capitalism of the twenty-first century, turbocharged by neoliberalism, has *re-defined* creativity to feed its own growth. Being creative in today's society has only one meaning: to carry on producing the status quo (Mould 2018: 3).

Yes, there is some kind of 'value' in what these things are doing. 'Bot or Not' is, of course, a simplified Turing test, and 'Botnik' could perhaps be read as an (unintentional) critique of the type of big data analytics that makes such a mess of digital search and translation systems (see Chapter 4). Both could usefully be used as foci for more critical and in-depth analysis. Likewise, they are both fun and educational tools, but, like digital 'found' poetry (as discussed above), to my mind they are missing the point of their self-situatedness in (and therefore their reproduction of) the AI / algorithm discourse. Instead of creating reams and digital reams of algorithmically generated texts just because technology now facilitates that type of project, I think that a return to the analysis of already existing literature (despite the technology on hand), might be a better way of grappling with the questions of human/machine interaction that pervade across the disciplines today. For example, I find it fascinating that in his seminal paper on 'Computing Machinery and Intelligence' (1950), Alan Turing turns to poetry in his attempt to explore the question of whether machines can think. In a variation of the imitation game, the intelligence of a witness is tested viva voce style, using questions about Shakespeare's Sonnet 18, which ironically is the same poem quoted in the problematic 'Deep-speare' machine-generated poetry study discussed above (Lau et al. 2018). From Turing's paper:

Interrogator: In the first line of your sonnet which reads 'Shall I compare thee to a summer's day', would not a 'spring day' do as well or better?

Witness: It wouldn't scan.

Interrogator: How about 'a winter's day'. That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.

Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.

Witness: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

Turing here has used existing human poetry in order to test a non-human system, which is precisely the opposite of what the studies I discuss above do, although they purport to belong to the same tradition of data science. What is critical to Turing's experiment is a deep knowledge and understanding of poetry and other cultural and emotional knowledge, not of digital techniques and methods. Indeed to my mind this makes the 2018 Lau et al. paper redundant.[3] Lau et al.'s 2018 conclusion that concentrating on form alone is not enough to create 'human' poetry, was effectively reached by Turing in 1950. As the example shows, the first objection to the changing of 'summer's day' to 'spring day' is one of form. 'Spring', being only one syllable long, does not fit into the iambic pentameter of the rest of the sonnet, in which each line is made up of five metrical 'feet', each of one stressed and one unstressed syllable. Of course, as Lau et al. point out, deviation from strict scansion is also a method of poetic expression, but it is the next part of Turing's

[3]Which in itself bring up further questions about the proliferation of data science funding in the rapidly growing number of research institutes that support this kind of work, but these are questions for another day/forum.

test that shows the importance of subjective emotion over (soley) form in human poetry. In an example of the poetic method of *pathetic fallacy*, in which human emotions or responses are reflected or personified in nature or inanimate objects, the substitution of a 'summer's' to a 'winter's' day (which satisfies the form in terms of scansion), is objected to because 'nobody wants to be compared to a winter's day'. The 'winter's day' alternative is then further complicated (and therefore humanised)[4], by an emotional response to a Charles Dickens story (Mr Pickwick's Christmas), and the suggestion that some people might like to be compared to certain winter's days, especially Christmas Day. I think the importance that Turing gave to a deep understanding of cultural and artistic nuance is critically important, yet has to some extent been lost in the commercialised and perhaps politicised agendas, methods and databases available to modern machinic computation, analysis and critique.

As I have mentioned above, work in this area tends to try to answer the question of whether a machine can write poetry, i.e. computer generated poetry which is 'tested' by seeing if a human can tell it has been written by an algorithm (e.g. Bot or Not). However, such arguments tend to fall into humanistic debates about the nature of poetry and art rather than asking what I think is the more pertinent question of 'can a machine *read* poetry', which is what Turing was getting at, and what are the consequences if it cannot, or if other forces embedded in digitised systems prevent it from doing so or corrupt its interpretation of language? Digital big data technology enables us to 'read from a distance', a technique taken up widely in Digital Humanities (see Moretti 2013 and discussion in Chapter 3), and indeed in large corpora analysis in data science too. However, I think that to abandon 'close reading', as in the traditional 'close reading' method of deep and multi-layered and subjective textual analysis employed in literary criticism, is to miss an incisive way of examining language in the age of algorithmic reproduction. The economic - and subsequent political and cultural - capital generated by words monetised in Google's system of linguistic capitalism need to be incorporated into readings of digitised language. Going back to Benjamin's 'Work of Art' essay, and the new opportunities and insights gained by new artistic technologies such as photography, these paratextual values can be seen as the 'multiple fragments which are assembled under a new law'; in this case, the complicated, yet critically important laws of an algorithmically mediated marketplace (see

[4] According to N. Katherine Hayles, the important intervention Turing makes is to embody the condition of the posthuman; his test revealing a 'cybernetic circuit that splices your will, desire, and perception into a distributed cognitive system' (2008: xiv).

Chapter 6). As mentioned earlier, Benjamin saw this new technological textual layering as 'the mutual penetration of art and science', likening the process to Freud's relatively recent method of psychoanalysis which he says 'isolated and made analyzable things which had heretofore floated along unnoticed in the broad stream of perception' (1999: 229). It is exactly the 'unnoticed' flows of capital that mediate language and affect our 'streams of perception' today that I attempt to 'make analyzable' and visible with my own intervention. In the next section I present and document my intervention into linguistic capitalism, called poem.py, which develops a radical new approach to the 'close reading' of poetry, exposing the competing values and contexts that construct and decipher language in the age of algorithmic reproduction.[5]

7.4 {poem.py} : a critique of linguistic capitalism

My intervention into linguistic capitalism also began with poetry, when I decided to experiment with working out how much poetry is 'worth' to Google (in monetary terms) by feeding poems through the AdWords Keyword planner and printing the results out as receipts. I document the project's development in detail here, because the process, with all its set-backs and successes is critically important to my overall argument and is therefore worthy of further scrutiny, as it draws out many methodological and disciplinary problems of research in this area (see Chapter 3), as well as developing and crystallising arguments and discussions around the themes of value, context, and circulation that draw this thesis together.

The project started off when I was required to provide a poster for the Information Security Group (ISG) Open Day at Royal Holloway in May 2016. Unlike most of my PhD contemporaries in the Cyber Security Centre for Doctoral Training (CDT), I was unable to create a poster filled with the mathematical formulas, graphs and data that made up the required template, yet I had to think of a way to represent and explain the validity and importance of my project to an audience of cryptographers, mathematicians, computer scientists, and industry leaders, and perhaps more inportantly, to satisfy my main funders, the EPSRC (Engineering and Physical Sciences Research Council). It was under the pressure of having to produce a poster capable of translating my work to this audience that I started to experiment with 'valuing' language as it passed through Google's algorithmic

[5] I also see my intervention as being an intervention against digital art, in that it insists on an analogue outcome.

spaces, as I thought it might be a helpful way of visualising the process, suggesting that linguistic capitalism as practiced by Google (and potentially other tech companies) might compromise the integrity and security of communication in digital spaces just as much (if not more) than cryptographic key exchanges and other technical approaches to Information/Cyber security. I therefore decided to see how much my favourite poem 'cost' if I put all the words through the Google AdWords Keyword Planner and output the results on a mock-up of a receipt, which I thought might look nice on a poster. I began this process by copying the text I wanted from a poetry site on the Web, and then manually 'cleaned up' the poem, taking away punctuation, pictures or annotations to make it possible to bulk search the text through the keyword planner. I then created a comma-separated list of the words in the poem (which is the format you can enter multiple words in the KWP), and fed it through the main input bar of the Keyword Planner. The KWP outputs price data in several formats, so I picked Excel, purely because it was what I was familiar with. As well as giving the suggested bid price for each word in the poem, the spreadsheet also gave an indication of search volume, and a competition score for each word. All this data is specific to the date, time and market in which conducted the query. The KWP allows advertisers to target specific geographical markets, ranging from ALL markets, to the US, UK, London, or even smaller areas such as military bases or universities, as well as targeting specific devices - i.e the mobile, tablet, or desktop markets.

The first challenge I faced with this data, is that the words on the spreadsheet are not returned in the same order they are input, so the narrative order of the poem is immediately broken in favour of an opaque algorithmic way of ordering; the power of metrics (Beer 2016) subsuming the power of the poetry, perhaps. For this first experiment, I therefore had to put the poem back into its narrative order manually, and because I wanted the result to look like a receipt, with a price for each word as if it were on a shopping list, I had to calculate and aggregate the value of words used more than once, so that although the poem read downwards in the correct order, as the poem progressed, the conjunctions and more common words are missing. I thought this was important, as it recreates the poem, but also makes the point that monetisation still has a hold over the linguistic narrative function of the words. When I had manually calculated the poem/receipt in this way, I discovered that, at 4:39PM on 7th May 2016, my favourite poem *At the Bomb Testing Site* by William Stafford cost the princely sum of £45.88 (before tax) (see Figure 7.4, pp.153-4).

At the Bomb Testing Site
by William E Stafford

SALE

7th May 2016 16 39 PM
BATCH # 085F2
APPR # ew7378
TRACE # 9

3	at	£1 95
1	noon	£0 31
2	in	£0 86
6	the	£1 92
2	desert	£2 92
5	a	£2 25
1	panting	£0 02
1	lizard	£0 64
2	waited	£0 00
3	for	£0 93
1	history	£1 61
1	its	£0 27
2	elbows	£3 02
1	tense	£0 05
1	watching	£3 96
1	curve	£0 44
2	of	£0 70
1	particular	£0 03
1	road	£0 99
1	as	£0 53
1	if	£0 05
2	something	£0 30
1	might	£0 14
1	happen	£0 18
2	it	£3 92
2	was	£1 16
1	looking	£1 18
1	farther	£0 00
1	off	£1 97

1	than	£0 02
1	people	£1 35
1	could	£0 03
1	see	£0 15
1	an	£0 33
1	important	£0 30
1	scene	£0 15
1	acted	£0 00
1	stone	£0 12
1	little	£0 02
1	selves	£0 00
1	flute	£0 29
1	end	£0 08
1	consequences	£2 09
1	there	£0 65
1	just	£0 24
1	continent	£0 41
1	without	£0 53
1	much	£0 44
2	on	£1 26
1	under	£0 12
1	sky	£0 25
1	that	£0 07
1	never	£0 19
1	cared	£1 18
1	less	£1 75
1	ready	£0 56
1	change	£0 32
1	hands	£0 63
1	gripped	£0 00
1	hard	£0 05

SUBTOTAL	£45 88
TAX	£9 17
TOTAL	£55 05

APPROVED

THANK YOU FOR SHOPPING AT GOOGLE

CUSTOMER COPY

I had, in effect, calculated the theoretical 'value' of the poem as it exists in digital space (in this case flowing through the search engine), as opposed to the 'value' of the same poem if it were to exist only on the printed page. My task then was to make the poem look like a real receipt, so I played around with various free templates online, but none of them would process a poem of that length without payment, so I gave up and created my own template in Excel, which I saved as a PDF, and copied onto a Powerpoint slide in order to make my poster. I took this initial idea - and the mocked up receipt - to my (cryptographer) co-supervisor Prof. Keith Martin, and declared that my poster would consist purely of the Stafford poem enlarged on a portrait-orientated poster. Prof. Martin immediately raised several potential problems with my poster, although to his credit he never said it was a bad idea, or that I should not do something so radically different from the other posters. The first problem was orientation. The Stafford poem was long enough to require a portrait-style poster, yet the required ISG template was landscape. A portrait poster would 'look out of place'. The second issue Prof. Martin raised (and in doing so solved the first issue), was that nobody would be familiar with *At the Bomb Testing Site*, so it would not be immediately obvious it was a poem, which might make the poster even more inaccessible to a technical (or indeed any) audience. I needed something 'even cryptographers' would recognise, something like 'I wandered lonely as a cloud' So that is what I did. I fed the first stanza of William Wordsworth's *Daffodils* through the Google AdWords keyword planner, and used those first few lines as a much shorter receipt which fitted onto a landscape poster template.

Another issue raised by Prof. Martin was also perhaps disciplinary in nature. I had initially wanted only the poem-receipt on the poster, so it stood up as a piece of conceptual art (or political art, following Benjamin), to be 'deciphered' by an audience I hoped might be intrigued by the receipt format, and I wanted to title it *At the Bomb Testing Site (2016)*, perhaps more of a caption than a title, in order to give it the air of an artwork in a gallery.[6] Quite rightly, Prof. Martin pointed out that the receipt needed some explanation, and that the 'conceptual art' format was possibly a step too far for an Information Security audience, and indeed any audience where the piece would have been so out of place. So I wrote a brief explanation of my critique, now called *Daffodils (2016): A Critique of Linguistic Capitalism* and added it to the (landscape) poster, reluctantly conformed to the background branding colours and logos required by the ISG

[6] I'll be the first to admit that I was, at this point, being deliberately contrary, and railing against what I saw as the structural constraints imposed by the ISG.

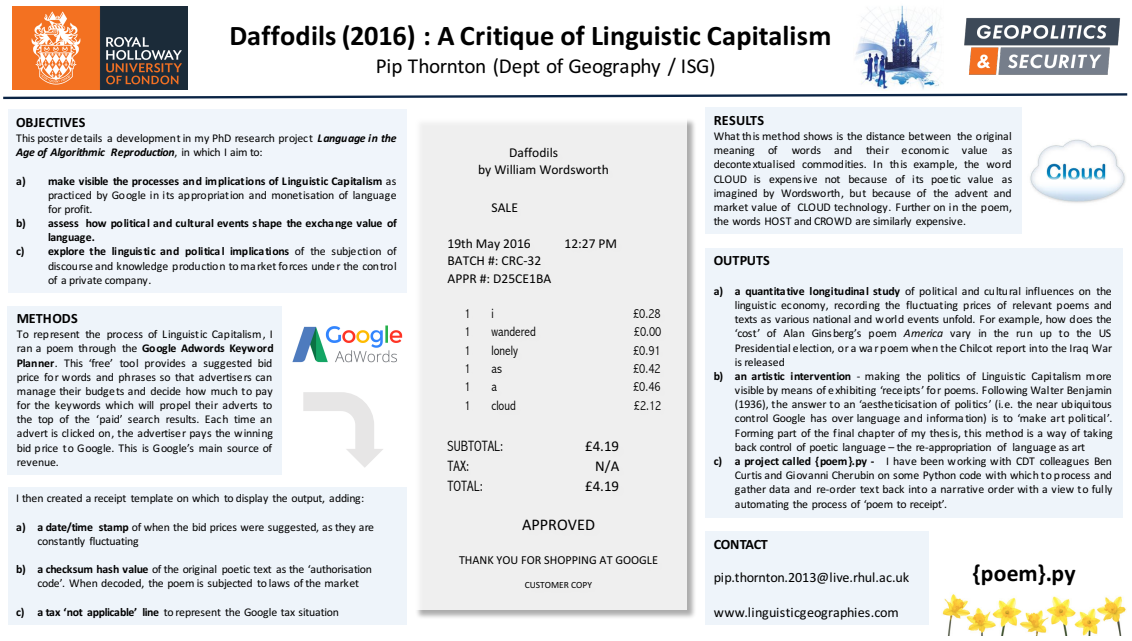


Figure 7.5: Daffodils (2016): A critique of linguistic capitalism - poster for RHUL Information Security Group Open Day.

and Royal Holloway, and submitted my poster. To my surprise and delight - and certainly challenging any assumptions I had initially had about a 'technical' audience understanding it - I ended up winning the prize for best poster at the Open Day. It is also worth mentioning that while I was working on the poster, I also created an aesthetically very different 'non-ISG' version, in portrait and with my own chosen colour scheme, and with a longer extract of *Daffodils*, for submission to the Royal Holloway Humanities and Arts Research Council (HARC) competition. That poster also won a joint first prize (both versions of the poster shown here, see Figures 7.5 & 7.6). Although I was delighted to win these poster prizes, and was lucky to have the means and resources to be able to produce two versions, the difference in aesthetic and corporate presentation needed to do so in two distinct disciplinary competitions, is an interesting example of the difficulties of cross-disciplinary research, dissemination, and visualisation.

7.4.1 Coding Poetry

As I have described above, with my test poem, I had to order the words back into the shape of the poem manually, which was time-consuming and fiddly, and severely limited the scope and potential of the project. I decided to see if I could automate this process using code, and as I was not (and am still not), a competent coder, I asked for help from CDT colleagues. Ben Curtis, a mathematician who wanted to develop his basic coding

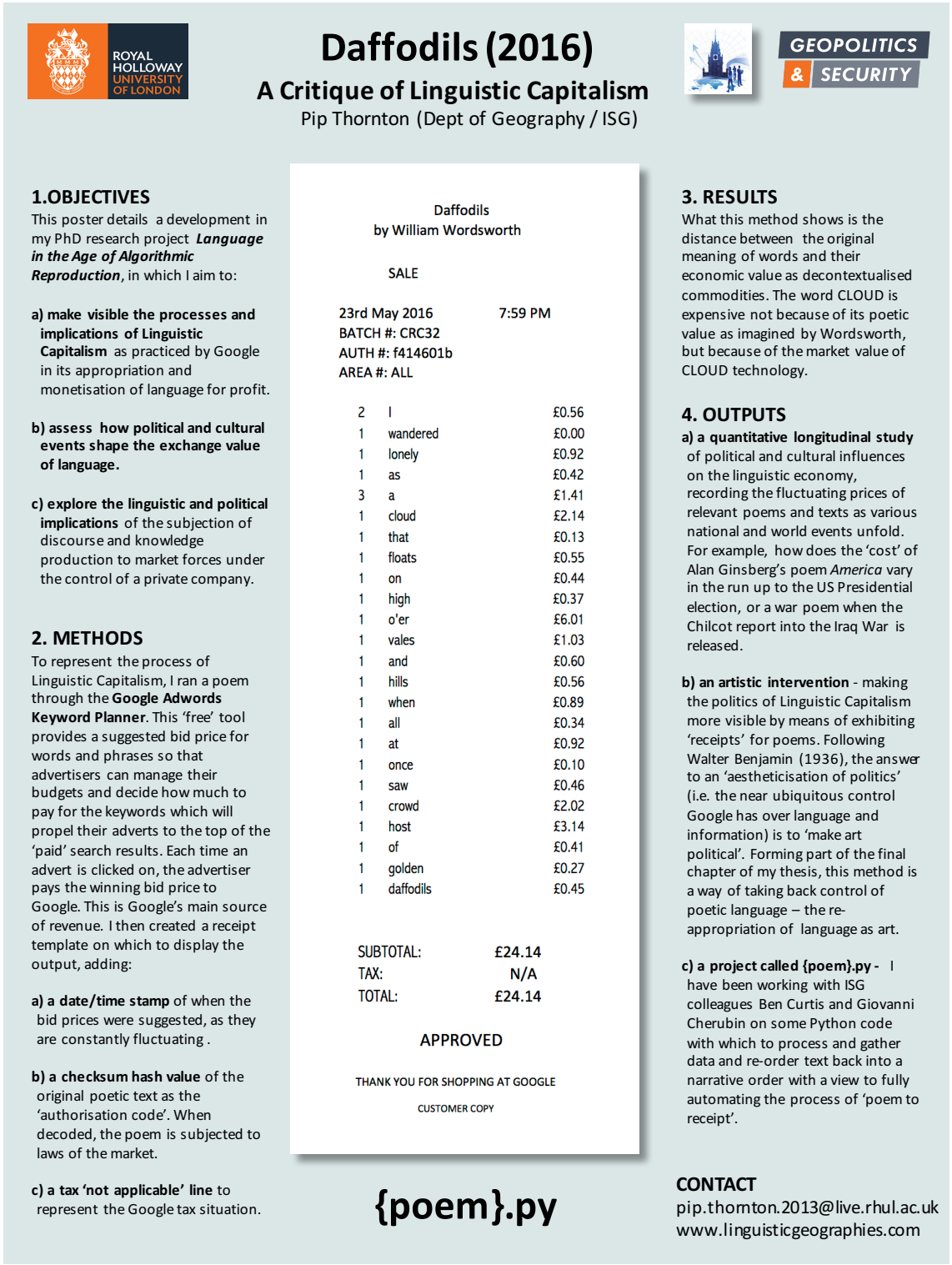


Figure 7.6: Daffodils (2016): A critique of linguistic capitalism - poster for RHUL Humanities & Arts Research Institute poster competition

skills, answered my call for assistance, and so began the co-production of the Python code which gives {poem}.py its name. This union of poetry and code is where the project title {poem}.py comes from .py being the file extension for Python (see Figure 7.7).

I wanted the code to serve two purposes; both to strip the poetry I copied and pasted from the Web of punctuation and repetition, thus preparing it for processing through the keyword planner, and then to reorder it back into readable narrative (downwards) order once it has been monetised. This process involved a lengthy period of debugging and discussion, made both more complicated and challenging, but also more interesting and productive, by the cross disciplinary collaboration. Forced to read the entirety of an Alan Ginsberg poem out loud so that we could work out which word the KWP had not processed (and was therefore skewing the word prices), Ben read the poem like a shopping list while I checked off each word as he spoke. It was a strangely prosaic way of reading a poem, indeed, perhaps comparable to a machinic reading. We eventually realised that the anomaly in the results was because (unless you check an opt-in button) certain words are screened out by Google as 'adult' content. Only by reading the poem and the spreadsheet in unison, could we work out that the word 'fuck' had been invisibly removed from the KWP results. While the opt-in screening of obscenities might seem fairly obvious, and therefore avoidable, what was harder to avoid were the problems thrown up by homonyms. When I processed Simon Armitage's poem 'A Vision', through {poem}.py, a similar corruption of the data occurred, but this time it was because the word *strips* had been screened out. The context of the word in the poem had been 'strip' as in 'portion of a road' (i.e. Sunset Strip), not 'stripping/ stripper', but that was an early indication both of the importance of (de)context in these valuations, and of the hand that Google actually has in this linguistic marketplace; silently screening out and censoring certain words from the results.

With Ben's code, the project progressed fairly rapidly, as I was able to process more and more poems and texts. As part of this process, and to make the receipt format look more visually authentic, I added a CRC32 checksum hash value to the receipt as an 'authorisation code'. A checksum is a mathematical blueprint of a piece of text which is generated to ensure that a transmitted text has not been altered, or more accurately, to ensure that if it *has* been altered, then the recipient will know about it. The sender sends the checksum with the text and the recipient generates the same checksum to make sure it

```
# configure filenames

filename = 'larkin.txt'

output = 'larkinreceipt.xlsx'

input = 'larkingoogle.xlsx'

type = 'all'

if type == 'all':
    zeta = 0
else:
    zeta = 1
```

Figure 7.7: Snippet of the original {poem}.py code, courtesy of Ben Curtis (2016).

has stayed the same in transit. Using this as an authorisation code on the poem receipt is therefore suggesting that when protected by code or encrypted, the poem retains its narrative integrity, but when it is decoded, it is then subject to the laws of the market, as shown on the receipt itself. At this point it might be important to reiterate arguments from previous chapters. Language has of course always been bought and sold, whether as poetry, novels, text books or newspapers, ever since words became written down, movable, and therefore marketable. The difference I want to highlight between language that exists on the printed page and the web page, is that digitised language (or language-as-data) has become infused with a different kind of market value than paper based print, whereby words themselves can be elevated beyond or outside both their material manifestation and their narrative function in order to create capital. Thus one way in which the re-ordering of words as they become subject to the varying metrics valued by Google's algorithmic systems can be resisted, might be by encryption, although that itself is not unproblematic. The idea of encrypting poetry as resistance will be discussed further in the next section.

The final part of the construction of the {poem}.py project was buying a second-hand receipt printer from eBay, and formatting the template to produce realistic receipts (see Figure 7.8). I also added N/A to the tax field as a comment on Google's tax situation in the UK. The production of the physical receipt is an important part of my intended critique of the way language is monetised. Outputting the critique as old-fashioned paper receipts is my way of reclaiming poetry from the algorithmic market, and restoring its literary narrative value as a piece of art. As they pass through the poem.py process, words become quantifiable commodities, taken out of their poetic contexts, and reordered according to their popularity on spreadsheets populated according to quantitative, rather than poetic logics. The coded reconstruction of the poem back into it narrative order is, however, still not a total reversal of the market logic that disrupted its lyrical flow. The poem reads downwards - an unfamiliar format to human eyes - and I thought it was important to keep the format of the receipt in the way it aggregates repeated words, which has the effect that the poem tends to lose conjunctions and popular words as it progresses.

What I have done with {poem}.py is to work deliberately from inside the technology (but hopefully without leaving a trace), exposing the workings and logics of linguistic capitalism, without becoming a fully paid up contributor in that marketplace. I work only with the suggested bid prices, not the actual bid prices, to avoid having to become a part of

Figure 7.8: {poem}.py receipt printer, AKA the Deconstruction Machine / Black Box. Photo: Amy Freeborn.

the market and also to avoid paying any money to Google as an actual advertiser. This method is different to many other so-called 'digital art' projects, in that its outcome is deliberately analogue. In an age when more and more shops issue receipts via email, the old printed receipt is slowly becoming an artefact, which suits my project well. My second-hand receipt printer in effect becomes my own physical, yet conceptual, black-box. As I have said before (in Chapter 2), it is impossible to expose the workings of the black box (at least in an understandable format), but what I *can* do is to show what is input (a poem), and what is output (monetised data).

7.4.2 Context

As the project has progressed, and the more poems I processed through the {poem}.py system, it became clear just how rich a method this would prove to be. As well as an artistic statement (more of this later), {poem}.py also enables me to shine a light on the opaque pricing system which facilitates Google AdWords, and consequently plays a part in structuring so much digitised text. As I ran more and more poems through the system, I began to realise that words relating to health, technology, litigation and finance were particularly - and sometimes inexplicably - expensive in the context of the poem, but once thought of in the context of an advert, their valuations became more obvious. The words of the poem, as they are fed through Google's algorithmic systems, have their poetic context removed. Their value is instead imagined by the KWP in terms of their tradable worth as commodities, a system reminiscent of the novel writing machines in the Fiction Department in *Nineteen Eighty-Four*, where the literary value of language was irrelevant, and 'books were just a commodity that had to be produced, like jam or bootlaces' (2000: 149-150). The suggested bid price for the word 'will', for example, which appears frequently in many poems, including for example William Blake's lyrics to the hymn *Jerusalem*, is relatively high in comparison with the other words around it. On 12th July 2016, when I processed the poem, it was priced at £2.06 per click (in the UK market); its value based not on the future tense of a verb, but on the legal document. Likewise, other poems I processed revealed that the words 'break' and 'receding' are expensive because of holiday adverts and hair products rather than the more poetic heartbreak and tides. William Wordsworth's *Daffodils* is a particularly effective poem to use to illustrate these contextual mismatches, as it is such a well-known poem to most people, and because it includes the words 'cloud', 'crowd' and 'host' in the first stanza, all

of which seem disproportionately valuable at first glance; £2.14, £2.02 and £3.14 per click respectively on 23rd May 2016 in all markets. However, the suggested bid prices of these words do not relate to how Wordsworth felt as he imagined the sky and the landscape on a spring day in the Lake District, but to the far more prosaic variants of 'cloud computing', 'crowd-sourcing/funding' and 'web hosting'. Perhaps unsurprisingly, it seems the cost of a word to Google relates to the size and wealth of the industry it plays a part in advertising.

I was also initially puzzled as to why conjunctions such as 'he' and 'it' hold commercial value, but it soon became apparent that their worth is as acronyms for 'higher education' and 'information technology'. Even in Latin, or indeed any language, conjunctions and acronyms can be valuable linguistic commodities in different geographical or historical contexts. For example, the Latin word 'uti'; in the line 'Quando fiam uti chelidon' (when shall I be like the swallow) in T.S. Eliot's *The Wasteland* has a high estimated bid price because of the market around treatments for 'UTIs' (urinary tract infections) and the advert and click-bait heavy sites that provide (sometimes questionable) online medical advice. Not so poetic now, it seems. But what is so interesting in terms of the discursive and political effects of linguistic capitalism, is that even if you wanted to search for the Latin word 'uti', it is highly unlikely to be returned in that context in a Google search engine results page, purely because irritating and potentially embarrassing medical conditions earn more advertising revenue than Latin verse. Likewise, with the word 'cloud' in relation to Wordsworth's cloud, or indeed the meteorological or psychological renderings of the word, if you type the word into the Google search bar, your intent is irrelevant. In what is effectively a form of algorithmic governance (Rouvroy 2013, Danaher 2016), or a technologically facilitated version of the classic literary 'intentional fallacy' (Jarrett 2014), the version of the word appearing at the top of both the paid for and the organic results will always be the most commercially viable version of the word, which is not necessarily the version you intended. This is what Jobin & Glassey call Google's power of 'semantic determinism', by which 'algorithms may prevent a user's potential search queries from escaping the lexicon with which they are familiar by suggesting words whose meanings make sense for Google' (2014: 158). Thinking back to 'Nineteen Eighty-Four', Orwell describes a similar removal of linguistic intent in Newspeak:

> The word free still existed in Newspeak, but it could only be used in such statements as 'This dog is free from lice', or 'This field is free from weeds'. It could not be used in its old sense of 'politically free' or 'intellectually free'

(2000: 344).

Just as in Newspeak, the 'world-view and mental habits' of the searcher are thus subtly, yet forcibly, shaped by linguistic mediation (2000: 343). A particularly stark example of this digital intentional fallacy is the word 'guttering' in Wilfred Owen's *Dulce et Decorum Est*, which had a suggested bid price of £1.46 per click when I searched it in all markets on 2nd June 2016. When processed through {poem}.py, the word Owen coined to describe soldiers choking on gas in the trenches of World War One - a beautifully constructed, wonderfully alliterative and terribly evocative mixture of 'mud', 'guts' and 'spluttering' - is reduced to the price of an advert for plastic drainpipes. It is difficult here not to be reminded of Orwell's Newspeak, a language 'impossible to use for literary purposes' (2000: 345).

7.4.3 Linguistic Geographies

As mentioned earlier, another factor in the algorithmic reckoning of the linguistic market place is location. The Keyword planner allows the potential advertiser to pinpoint specific geographic markets to hone their campaigns, the price of keywords differs not only temporally but also by location. To illustrate this point, I began creating receipts for the same poems, but in different geographic areas, which led to some interesting observations. Although I have argued in Chapter 6 that the knowledge producing capacity of an algorithmic linguistic market is corrupted by Google's constant interference in the system, there *are* some interesting cultural and political reflections in the data gleaned from the Keyword planner. At a conference in Ireland I had asked the chair to provide me with a favourite poem to print out and use as an illustrative example of the geographic fluctuations of price in AdWords. I processed the chosen poem, Sylvia Plath's *The Arrival of the Bee-Box*, though {poem}.py in three geographically decreasing market areas: firstly, all markets, then the whole of Ireland, and then Galway City, which is where I was giving the talk. Whereas as a general rule, words are worth less in smaller geographical areas, I noticed that the word 'god' in the Plath poem was worth almost three times more in Galway (£1.74 PPC) than in the whole of the AdWords market (£0.66 PPC). This is despite the fact that the poem as a whole was worth less in Galway than in the wider market. It seems 'god' is still a lucrative marketing tool in Catholic Ireland. Another interesting example in the {poem}.py data is the word 'lonely', which seems always to be worth more in urban environments than in rural ones. This perhaps reflects historical cultural fears of the loneliness of the modern metropolis (Simmel 2012), and current research

that suggests people are lonelier in crowded places such as cities than in the countryside, and that dating sites targeting 'lonely' singletons might therefore have more competition in urban areas, thus driving up the price of the keyword[7].

Aside from the process of producing the poem/receipts, I have also been gathering the suggested bid prices for over a thousand words every day since June 2016. Each of these words have been valued by the KWP in the UK, US and in all geographical markets. Some of these words belong to poems or texts which I thought threw a kind of linguistic net around a particular event or subject, such as the lyrics to Billy Bragg's *Between the Wars*, which includes politically relevant words such as 'voting', 'government', 'austerity' and 'prosperity', or Margaret Atwood's *It is Dangerous to Read Newspapers* which describes the immersive terror of reading about war, bombings and atrocities in the media. With this data, I can plot the price of whole poems on a geo-temporal scale, for example showing the price of Alan Ginsberg's poem *America* (in the US and other markets) throughout the buildup and aftermath of the US presidential election in November 8th 2016. As show in Figure 7.9, in the six months either side of the actual election, the overall price of the poem was far more volatile in the US that in the UK market. In the UK and ALL markets, the poem had a fluctuation of circa £50 (from around £285 to £240 in the UK), while in the US the highest and lowest price of 'America' dropped by half in the months after the election, ranging between circa £600 and £350 in a steady decline from July 2016 to end of May 2017, with election day coming in at around £450, although there was a slight rally in price just before my data ended.

But as well as showing the imagined 'price of America' over a politically turbulent time as an artistic statement/intervention, I can also track the trajectory of the individual words within the poems. I have also been gathering the suggested bid prices of another group of words as part of a 'word cloud' made up of words likely to appear in online texts and the media, such as 'trump', 'blair', 'clinton', 'snowden', or 'brexit', so am able to map their suggested prices against world events on a geographical timeline. Although I will have to leave full analysis of this data for a future project (see Chapter 9 on future work), initial findings show how the word 'chilcot' achieved monetary value shortly after the Iraq Inquiry was released on 6th July 2016, going from £0.00 to £1.86 in a period of 7 days. Likewise, the words 'wags' (as in the phrase *wives and girlfriends* which originally

[7] https://www.citymetric.com/horizons/why-living-big-city-so-isolating-lonely-isolation-loneliness-4210

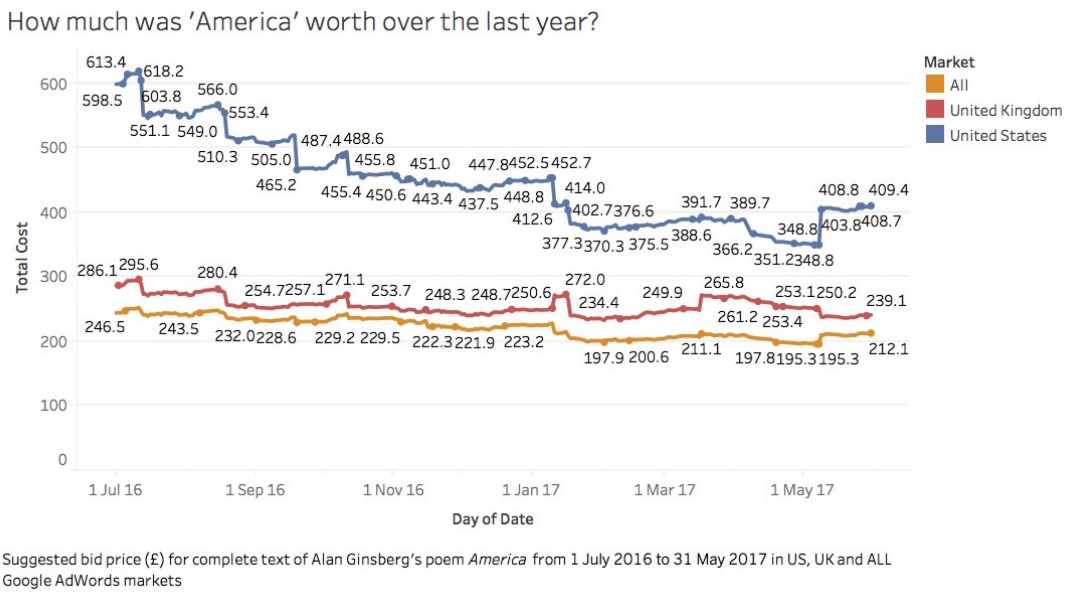


Figure 7.9: Suggested bid price (pounds) for complete text of Alan Ginsberg's 'America' from 1 July 2016 to 31 May 2017 in US, UK and ALL Google AdWords markets.

launched the idea for this thesis), went from zero to £5.44 per click in the UK market 6 days after the start of the 2016/17 Premier League football season. If it is surprising that such ostensibly unmarketable words have money generating potential in the digital adverting market, then we might remember the run up to the 2017 UK general election, when political parties, think tanks and charities found themselves competing on AdWords for keywords such as 'brexit' and 'dementia tax', as I discussed in Chapter 6.

What is important to remember in all this is that lucrative and effective keywords, no matter what their other contexts, or indeed the intent of those who search for them, are more likely to be used in the language which exists on and flows through online spaces, whether this is in AdWords advertising, in organic optimised text, or in the sites that generate income through clicks and views on adverts. Their repeated use by advertisers and confirmatory clicks by users, along with the encouragement of the Google community in the form of SEO help pages and forums (see Chapters 5 and 6), serves to increase their frequency within the wider database, and this is a process which I argue has an important effect on the subsequent information generated by Google functions such as Search, Autocomplete, and even Google Translate (see Chapter 4 and Thornton 2017). As keywords and optimised text become the rules by which online language is organised, it follows that in a world increasingly mediated by Google platforms and infused with the neoliberal logic of the marketplace (see earlier in this Chapter), language which is non-normative,

original, or creative becomes less prominent in the corpora of words available to Google's algorithms. As I discussed in Chapter 5, an important SEO tactic is to recycle content, to reblog or rehash already popular content. Optimised text is rarely original, or new, or it simply would not perform its function of being easily found. It is important here too to think back to Cabell & Huff's 'American Psycho' project, and to think of the words that were stripped of their 'narrative' value and loaded with economic value as they circulated via GMail. My {poem}.py project has shown that the words which are entered into the portal of the search bar suffer a similar fate; de/re-contextualised and de/re-valued, as they circulate through digital space.

7.5 Conclusion

This section has therefore looped back to the Benjaminian framework of imagining the movement of language from its 'original' time/place on a page or in a poem, and making visible the values that are lost and gained as it moves into a digital linguistic marketplace, and into a new political sphere, facilitated by both economic and algorithmic markets and by the embedded politics of proprietary technology. The chapter has also begun to extend Benjamin's framework, to imagine ways that the aestheticisation of politics by Google's search and advertising platforms and also specifically the commodification of language, can be challenged by 'making art political'. I have explained the genesis and development of my {poem}.py project, which has picked up on the themes in previous chapters of context, value and circulation. The next chapter will continue this productive synthesis of theory and practice, and will detail some of the (accidental and deliberate) modes of resistance to linguistic capitalism that have emerged/occurred in the development of the project. With the {poem}.py project I have attempted to reverse the performative logic (Lyotard 1984) of language in the age of algorithmic reproduction by reclaiming poetry from the algorithmic marketplace, repoliticising it, and turning it back into art (Benjamin 2008). However, as I will explain in the next sections, in an age of linguistic capitalism and algorithmic reproduction, resisting the logics and forces of the digital economy has not been an easy task in either a methodological or conceptual way.

Chapter 8

RESISTING LINGUISTIC CAPITALISM: THEORY & PRACTICE

Words forged by revolutionary critique are like partisans' weapons: abandoned on the battlefield, they fall into the hands of the counterrevolution and like prisoners of war are subjected to forced labor (Khayati, Situationist International (1966) 1981).

8.1 Introduction

I have argued throughout this thesis that the curation of language through Google's platforms is an inherently and dangerously political issue. I am hoping that my methods and insights have been innovative and constructive, but I am not the first to suggest or indeed resist the ideas of linguistic, semantic and digital capitalism, and neither am I the first to encounter difficulties in studying and critiquing such an embedded and ubiquitous system of control and governance. Drawing on some of the difficulties of researching algorithms (as in earlier Chapter 3), this next section will explore the specific technical, disciplinary, and also the ethical problems and hurdles which became apparent in the course of my research and, more specifically, as I tried to resist the powers and logics of linguistic capitalism, as well as drawing on existing acts of resistance and documenting some of the ways I have overcome or engaged with such difficulties to produce my own forms of resistance. As well as my {poem}.py intervention, and following the example of existing creative fiction writers such as George Orwell, my methods here also draw on speculative theory and fiction in

imagining the concept of Subprime Language, which I will expand upon later in the section.

As well as Christophe Bruno and Cabell & Huff, another early inspiration for my intervention, and an example of political resistance to search technology, was an edited collection published by the Institute of Network Cultures called *Society of the Query*. The Reader brought together a range of contemporary research and creative critique with an aim to making the invisible workings and influences of search algorithms visible by 'ludic subversion', just as Debord and the Situationists had tried to expose and subvert the 'Society of the Spectacle', their main aim being 'to design visibility campaigns to make their influence apparent' (Konig and Rasche 2014: 10). As Mahnke and Uprichard conclude in their contribution to the book, 'only thinking of algorithms in terms of control, surveillance and hierarchy is to submit to that control' (2014: 260), and is therefore to be actively resisted, something echoed by Peter Olsthoom, who also advocates the resistance of the power of technology companies through playful subversion and disobedience. He writes:

> Through data collection, Facebook and Google are the most omniscient market research bureaus in the world at present. Of course, this will only remain so if we remain faithful to it, searching and commenting in the relevant boxes, like good boys and girls (2013).

But just as with quasi- Situationist activism and acts of détournement, to critique the 'Society of the Query' is also to tread a fine line between subversion and cooption. For example, the editors of 'Society of the Query', Mahnke and Upritchard, imagine what would happen if you could get billions of users to enter specific search terms which 'correct' misconceptions or stereotypes in the database which appear through Autopredict. They suggest that

> perhaps the capacity to act collectively towards a common good is by actively interacting with Google's autocomplete in ways that are similar to activist movements. If Google can shape our views, perhaps we can use Google to change our views?' (Mahnke and Uprichard 2014: 266)

Although it is entirely possible to 'interact with' Google's algorithms (as many Google Bombers and indeed any SEO expert would know), without becoming a paying participant in Google's economy (although this is arguable, as many people/companies are part of a wider 'Google economy' through paying for professional SEO, for example), it seems unavoidable that any attempt to 'change normative attitudes' in this way will itself be too

situated and nuanced a method to avoid criticism, and indeed, 'changing people's views' by manipulating search results is exactly the kind of Orwellian action I have been arguing against in previous chapters. Search results may be biased and controversial, but what right does an individual or group have to impose their own interpretations on other users? It is for reasons such as this that I have never actually used AdWords as a paying customer, nor made any attempt to influence organic or paid search results, as I will elaborate on later.

8.1.1 Co-option / Recuperation

One thing I should maybe stress in my use of Benjamin's call to 'make art political', is that of course there has always been political art. But one of the problems with political art - and in particular digital art - is keeping it politically critical and not letting it be coopted by the capitalist technologies from which it is made and which facilitate its dissemination and continued existence in the world. There is so much 'digital art', but a lot is very uncritical of the technologies which spawn it, and therefore, no matter what its artistic or philosophical influences, it often runs the risk of losing any radical agency by being recuperated into the spectacle of the digital economy. As Conor McGarrigle notes,

> central to the SI theory of the society of the spectacle was the idea that the spectacle had the power to co-opt or recuperate almost anything and that this power could neutralise even the most radical ideas (2010: 59).

As I mentioned before, Google Poetics might be an example of this perhaps; poetry constructed from snippets of auto-completions that do not interrogate the economic mediation of the form. In this way, digital poets have constructed works from Google results which have been compared to surrealist or Dada-esque pieces. But for me such examples are devoid of the really important political and economic structures which, ironically, construct their 'surrealist' form. They take away from the radicality of surrealism, as any critique is re-consumed again by the capitalism machinery of modern digital technology. Google Poems are perhaps reminiscent of Tristan Tzara's 1920 'How to write a Dadaist poem', and such deliberate subversion of traditional linguistic structures and norms was indeed radical in its day. Tzara's instruction for the creation of such 'anti-art' (as Dada was known) is as follows:

> To make a Dadaist poem:

> Take a newspaper.
>
> Take a pair of scissors.
>
> Choose an article as long as you are planning to make your poem.
>
> Cut out the article.
>
> Then cut out each of the words that make up this article and put them in a bag.
>
> Shake it gently.
>
> Then take out the scraps one after the other in the order in which they left the bag.
>
> Copy conscientiously.
>
> The poem will be like you.
>
> And here are you a writer, infinitely original and endowed with a sensibility that is charming though beyond the understanding of the vulgar.
>
> (Tzara 2018: 39)

But as Google Poetry, and similar work shows (see Chapter 8), although they may have the look, and indeed the method, of Dadaist poems, these forms run the risk of immediate recuperation.[1] Constructed of language that becomes data, which is monetisable as it moves through digital platforms and spaces, détourned search results are thus almost immediately coopted. As Sadie Plant writes on the problem of such cooption, 'turned into commodities, works of radical art and political criticism support the system of relations they despised' (2002: 79), and when the very words which construct radical art and political criticism are *themselves* commodities in a linguistic market, it is surely impossible to avoid recuperation into the 'society of the spectacle', or indeed, the 'society of the query'. It may well be 'essential' to capture the 'accidental' (Fuller and Goffey 2009) in order to expose the failings of autocomplete, for example, but the moment of capture is brief; perhaps only caught in the moment of the 'glitch' (Goriunova 2008, Jones 2017). As I will go on to explain, these are all issues I have had when developing my own intervention and resistance to Google search and advertising technologies.

[1] The similarity in method here is particularly obvious if we remember that early search engine technologies quite literally relied on 'bags of words' as methods of data retrieval.

8.2 Modes of resistance

I want now to return to the unfolding development of the {poem}.py project, and to the various forms of resistance it put up as a challenging and new emerging project, but also the modes of resistance it presented as a method and intervention. With the help I received from Ben Curtis with coding, I now had a system by which I could convert a poem/text scraped from the web into a printed receipt in 2-3 minutes (I will elaborate on why it takes this long in the next section). The first time I presented my new project was at a workshop I co-organised with Mike Duggan in May 2016 which we called 'Living With Algorithms'. By then I knew that the poem-receipts were proving a very clear and incisive way of explaining exactly what I was attempting to critique in my wider thesis, and I had found that by processing poems which actually meant something to people, the message became more personal and therefore even stronger. To this end, I emailed the participants in advance of the workshop and asked them - without telling them why - to provide me with a favourite poem. Some of the participants replied immediately that they did not have a favourite poem, so I suggested a favourite lyric or quote would work too. All but one of the participants sent me a poem/lyric, although interestingly several felt they needed to add the caveat that poetry wasn't really 'their thing'. I processed all of the poems in advance of the workshop and set the printer up in the workshop room before anyone else arrived. In this way, I was able to explain the theory behind {poem}.py as a critique of linguistic capitalism, and also show the small daffodils poem (from the poster) on the screen and explain the various parts of the receipt, before pressing print and watching everybody's surprise when the receipt printer started whirring away and churning out their own favourite poems as receipts. In this respect, from that moment, the project became not only a novel method of analysis and critique, but had now also become a performance. I realised that what had begun as a means of creative data visualisation, had also now become artistic practice. As mentioned in the previous chapter and in Chapter 3, the exploratory method I have have used in asking for poem requests, rather than basing my choices around existing anthologies or canons, was a deliberate attempt to avoid perpetuating such literary structures, and so I decided to use the poems from the workshop as the basis for a collection.

Figure 8.1: {poem}.py exhibit at Inter/Sections, London 2017.

8.2.1 Framing and Fading

I had always thought that I would use the 'collection' of poem-receipts from the 'Living with Algorithms' workshop as focus of my analysis, but as I became more interested in them as pieces of art, I started to wonder about having them framed. I had the opportunity to pursue this further when I was accepted as an exhibitor/presenter at the Inter/Sections event run by the Media and Art Technology CDT at Queen Mary University of London (QMUL), and held at the Mile End Arts Pavilion in London in August/September 2017 (see Figure 8.1). Two new (and critically interesting) developments arose from this. Firstly, the framing of the receipts, and secondly, the use of a Raspberry Pi to power the printer, thereby turning it into a centre piece of the exhibition, a black box mysteriously churning out receipts at random intervals, giving an agency of its own to what up until then had just been prop.

The idea of framing the receipts from the workshop was initially problematic on many levels. On a practical level, I knew it would cost a considerable amount of money to get them done, especially as they were such an unusual shape, and also therefore, quite

delicate. I was also dubious about what kind of value I was adding by getting them framed. I was making a bold statement by reconstructing the receipts as framed artworks, and I was worried that, although I was trying to reclaim the poems from the algorithmic marketplace and give them back their aesthetic value, in the act of framing I would essentially be actively re-entering the poems into a different marketplace, with all the value judgments that might entail. I was also very conscious that 'fixing' the poems behind a frame could also be seen as just as violent an act of capture; tying the words for eternity to Google's suggested 'derivative' prices and values. In this way framing halts the iteration of language - it is a violent capture of the words at a particular time, date and place, that echoes the post-structuralist arguments of Lyotard and Derrida that I wrote about in Chapter 4 (Geographies of (con)text). In framing the poems with one economic value I have in effect 'arrested' their 'meaning', which is exactly what Google does constantly when auctioning language. As I will explore later in the chapter, a post-structuralist view of language is that its 'value' lies in its capacity for constant deferral of meaning depending on its context, so in economic terms, the value of language is in its liquidity; its capacity to change shape and meaning. However, with the words that flow through the search engine, their primary 'value' is always an economic one, which makes them potentially illiquid, or to extend the metaphor, they risk becoming toxic. The wider 'violence' of algorithmic, calculative processes have been written about by Beer (2009; 2016), ONeill (2017), Owen (2015), McQuillan (2015) and others, but I think there is also a specific violence of capture in this non-consensual tying of an algorithmically generated economic value to language, or indeed linguistic data. As Kitchin points out, there is also an inherent nominative anomaly in how we even speak about the data that we harvest, scrape and crunch from any source (2014: 2). Derived from the Latin *dare* meaning *to give*, the implication is that data is freely given, rather than selected, ordered and taken as desired by the data analyser. Quoting H.E. Jensen, Kitchin argues that 'data are actually capta':

> [I]t is an unfortunate accident of history that the term datum rather than captum should have come to symbolize the unit-phenomenon in science. For science deals, not with 'that which has been given' by nature to the scientist, but with 'that which has been taken' or selected from nature by the scientist in accordance with his purpose (1950: ix, cited in Becker 1952: 278, cited in Kitchin 2014: 2).

So as well as the algorithmic 'capture of data', by {poem}.py, I was also adding another layer to this process by in effect freezing the 'value' of each word in time and space, and

then physically capturing them behind a frame, which was something I felt might compromise the conceptual integrity of my own project. I was in effect capturing the 'value' of these words and subjecting them the gaze of yet another opaque, yet powerful market. Indeed, resisting the further capitalisation of the poem-receipts has not been easy. People have asked to buy the receipts, framed and unframed. I have had requests to process receipts for people's birthdays, and I have had to provide insurance costs to exhibitions. Various people have suggested the poems would make lovely gifts; novelty toilet rolls or scarves, but I felt that the framing question in particular was especially problematic. How could I remain critical of Google and its system of linguistic capitalism, when I had not only replicated the violence of the algorithmic capture of 'value', but also exploited that data/capta in the name of artistic intervention, and in doing so run the risk of recuperation into the spectacle of the digital art market.

An answer to these potential problems came in part during the actual framing process. I wanted to keep the process as ethical and local as possible, so had approached the local framing shop in Egham. It took quite some time to explain why I wanted to frame eleven receipts ranging in length from 10cms to 180cms, and there were several practical hurdles to overcome too. The longer poems needed extra large sheets of glass and backing, which would have to be ordered in especially, but it was a frantic phone-call from the framer one morning that really helped the project. I had left the eleven receipts at the shop for a couple of weeks until he had time to attempt to frame them, and I had told him that I could always print off more if one got torn or damaged, for example. He had clearly forgotten both this and the fact that he was working with thermal receipt paper, and so when he put the first of the poems into the hot-press to bind it to the backing board, the heat reacted with the paper, and the receipt turned almost black. Horrified that he had ruined my 'artwork', he rang me immediately to apologise and to ask what to do. At first I said I would print another one off and drop it round to him (to his relief), but then I thought what a lovely example of resistance this was. Not of human resistance or intervention, but of an unwitting mechanical intervention into the process of re-aestheticisation. I asked the framer not to throw the darkened poem away, and when I saw it I decided to keep it as the framed version, in all its prematurely aged glory, so Wilfred Owen's *Dulce et Decorum Est* is unique in that collection as being a different colour to the other receipts, perhaps giving it another value, when in reality it is the product of a technological error. This incident with the framer made me think further about the integrity and longevity of the

materials I was working with. The paper I use is normal commercial thermal till roll, so in hindsight it was no surprise that the heat process had that effect. But I then started to think about the other poem-receipts, and how they also would fade with time, just like any other receipt, and it became obvious this was also a serendipitous means by which I could resist the violent 'capture' of the words at a particular algorithmic moment.

Yes, and as discussed above, I had frozen the 'value' of the poems in a specific time and place and 'enclosed forever behind the glass pane'[2], but I could now be safe in the knowledge that one day they would escape; liberated from their linguistic prison house (see Chapter 4), and free to be liquid once more, as I will expand upon in the next section. Several people who have {poem}.py receipts have been concerned - almost apologetic - that they have faded while being stuck to a fridge door, or on a sunny windowsill (see Figure 8.2), but I think that the process of fading must be part of the critique. The Open Data Institute, where the 'Living with Algorithms' collection on display (now in its second year there), originally asked me if I would replace the poems if they faded. I politely said no.

But as with other attempts at subversion, there is also the possibility that a framed poem becomes more 'valuable' in an aesthetic sense, the more yellowed it becomes. The recent prank by Banksy at Sotherby's, where he remotely shredded a piece of his own art the second it was sold, shows just how ridiculous the art market can be in terms of the valuation of individual pieces. Banksy's 'Girl with Balloon' was bought for just over £1 million, but since the shredding has now been estimated to be worth double that figure, some of this extra value perhaps deriving from the revelation that the shredder malfunctioned on the day, unintentionally creating a new work of art. I am not for a minute putting {poem}.py on a scale with Banksy, but his shredding stunt definitely contextualises the difficulties I have had with keeping my own intervention out of the marketplace.

As {poem}.py developed, another example of material resistance occurred when the first receipt to be chewed up by the printer was, appropriately enough, a Public Enemy song that a visitor to the Inter/Sections exhibition had requested (see Figure 8.3). The pre-

[2] A deliberate paraphrasing of a line from the anti-war folk song known variously as 'No-Man's Land', 'Flowers of the Forest' or 'Willie McBride'. The capture of the moment behind the glass has to me always seemed a poignant juxta-position to the imagined violence of the young soldier's death: *And did you leave a wife or a sweetheart behind / In some faithful heart is your memory enshrined / Although you died back in nineteen sixteen / In that faithful heart are you forever nineteen / Or are you a stranger without even a name / Enclosed there forever behind the glass pane / In an old photograph, torn and battered and stained / And faded to yellow in a brown leather frame.*

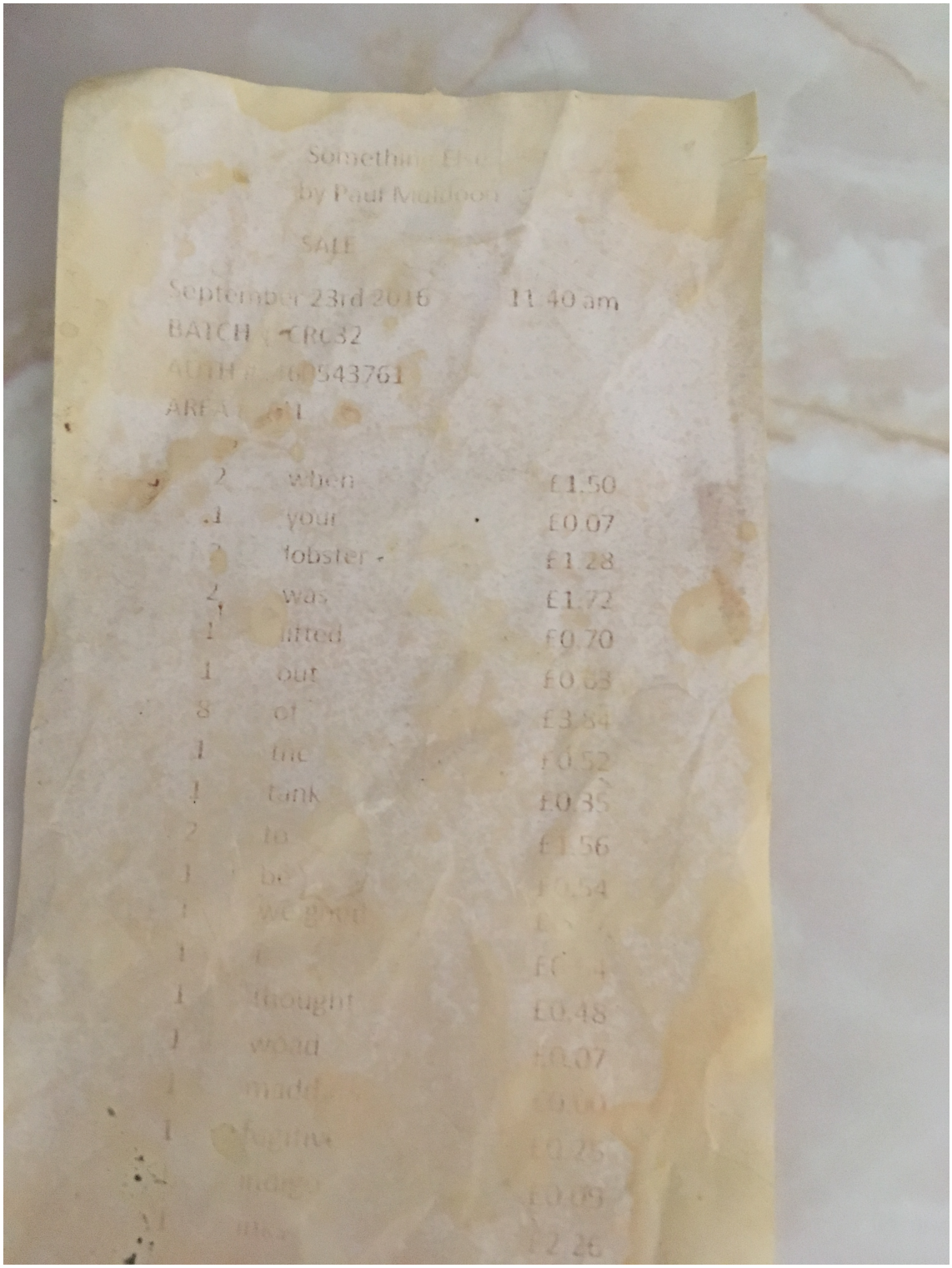


Figure 8.2: Paul Muldoon's 'Something Else' ({poem}.py 2016). Photo: Rachel O'Dwyer.

carious materiality of assembled parts of the {poem}.py process - from the algorithmic pricing, the Google updates, the occasional problems in the code, to the printer itself, the framing process and the paper inside it - can therefore perhaps all be seen as examples of the 'vitality' and agency of 'things' that Jane Bennett describes in 'Vibrant Matter':

> By 'vitality', I mean the capacity of things - edibles, commodities, storms, metals - not only to impede or block the will and designs of humans, but also to act as quasi-agents or forces with trajectories, propensities, or tendencies of their own. My aspiration is to articulate a vibrant materiality that runs alongside and inside humans to see how analyses of political events might change if we gave the force of things more due (2009: viii).

If there has been one moment that embodies Bennett's aspiration, it was when I took my receipt printer from London to Boston for the 2017 American Association of Geographers conference. At Heathrow the security scanners picked out the printer for further examination. *Why was I taking a printer to Boston? They have printers in Boston, you know. Why not use a dot matrix?*, were some of the questions, but what stood out even more was the way the security staff examined the black box of the printer itself. They swabbed it for explosives inside and out, opening it up and pulling out the blank till roll, and swabbing that too. As far as they were concerned it was a suspicious, potentially 'vibrant' object, capable of its own trajectory, but their equipment could only test for one type of political agency. As an object, the printer was benign; just a black box that eventually made it through airport security. But what could not be detected was its capacity for political agency once the chemicals within the thermal receipt paper were activated. The printer uses no ink, as the text is activated by heat, so in this way, the till roll they were swabbing at security already held the potential to say anything and everything. It really does act as a black box which transforms input to output, which is why, as a 'vibrant object', it has been such an effective way of demonstrating the project and making visible the processes of linguistic capitalism and their political potential. In this way the printer, while embodying the sensory regimes and materialities of the airport security process, manages to subvert these security logics and also to resist what Lisa Parks calls the 'venerating capitalist modes of extraction and circulation' of the airport screening site (Parks 2018).

But in subverting its everyday use, and viewing it as an object of 'vibrant' resistance, the printer has itself also taken on its own agency as an almost fetishised black box. It becomes a thing of mystery at conferences when I use it as part of a presentation. It has a

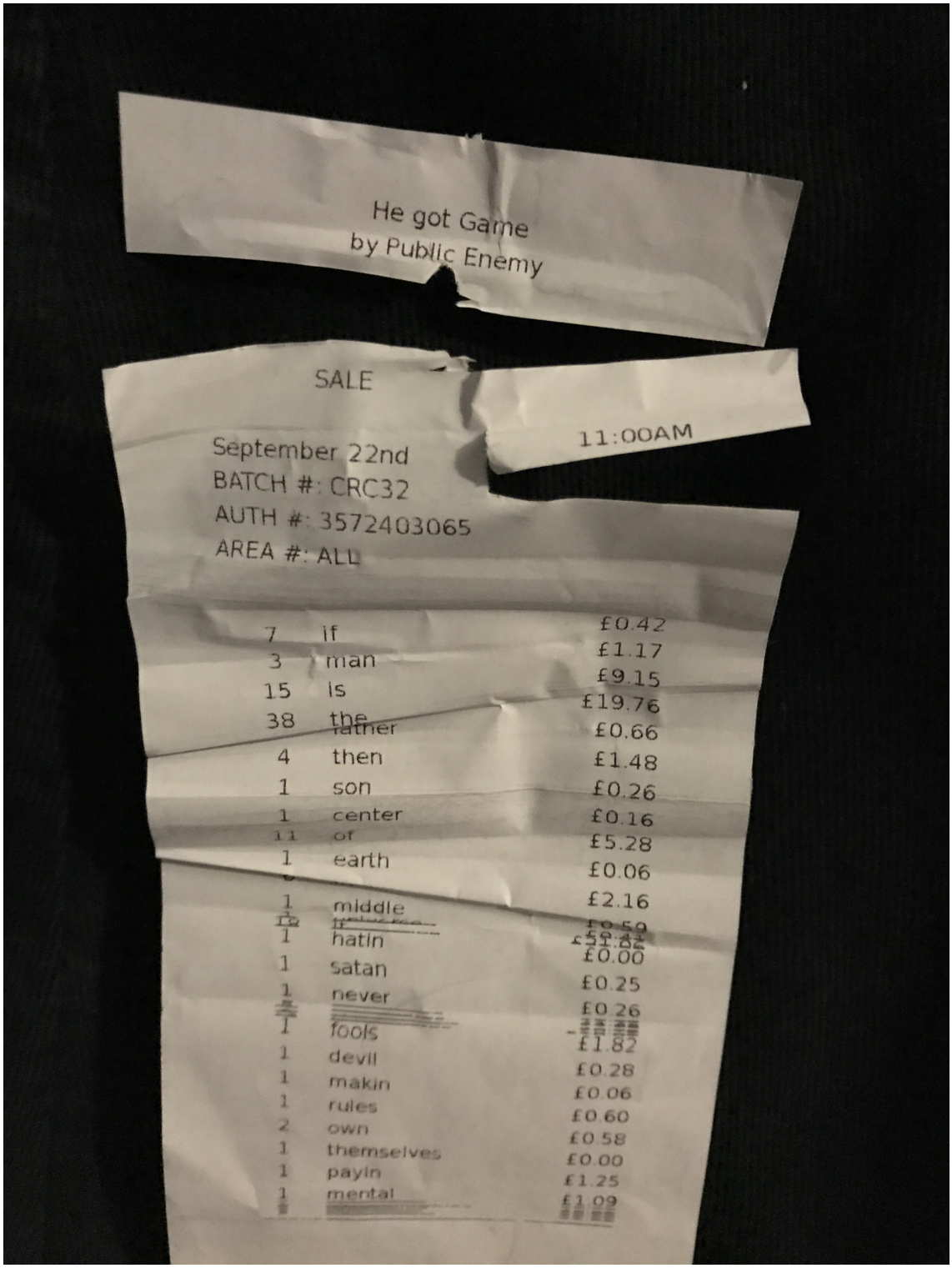


Figure 8.3: Public Enemy's 'He Got Game' ({poem}.py 2017). Photo: author's own.

Figure 8.4: {poem}.py printer. Photo: author's own.

large yellow and black sticker on its side saying CYBER (courtesy of Phil Garnett), which led to a fellow panelist in Boston asking if it was a 'special Google machine', rather than just a shop-used retail machine so dated that it came with an old-school embossed strip of tape with the word EPSON on it stuck to the plug (see Figure 8.4). As I have mentioned before, my methodology for this project has always been exploratory, and I realised early on that opening the algorithmic black box was not only impossible, but also perhaps not desirable. Instead, I have found that the most effective way to reveal the politics behind the system was to let its moving parts speak for themselves, and by moving parts I mean the data, the markets, the printer and paper etc. This is what I mean by working *with* the masters' tools in order to reveal the imperfect structures they build.

8.2.2 Printing and Performing

Another important development at that time was the idea of the 'printer as exhibit'. Previously, the receipt printer, bought second hand from EBay, had been a prop; just a novel means to create the receipts as poems. It is an old model receipt printer, it is heavy and bulky, and has been an absolute pain to carry around, both for security reasons

(see above) and physical ones. But at every opportunity I have carried my printer to talks and conferences, because it has proved such an engaging way to offer an analogue (and therefore understandable) critique of a digital (and therefore necessarily opaque) technology. The printer has therefore become a way of translating between technical and non-technical audiences. For the Inter/Sections exhibition I displayed the collection of framed poem-receipts from the 'Living With Algorithms' workshop, but I wanted to also have the printer on display as a kind of artifact/exhibit to compliment the frames. I was discussing the exhibition with friends when one of them (Joe Shaw) suggested that I could make the exhibition much more fun if I hooked the printer up to a hidden Raspberry Pi computer which would make the printer print out receipts at random intervals, 'as if by magic'. I had never used a Raspberry Pi before, and had no idea how to program one or connect it to the printer, so for the second time I fell back on the expertise and patience of one of my CDT cohort, this time Feargus Pendlebury, who set the printer and Pi up to work together and wrote some code to make it run. At this point we also had to adapt the Excel templates I had been using previously, as Microsoft software is not compatible with Raspberry Pi operating system, so we had to adapt and create a new style of template. Learning to use the Pi, and to run Feargus's code was a steep learning curve, and involved another extended period of collaborative debugging, but we finally got there, and I was able to take the printer and Pi to Mile End for the exhibition. I had asked in advance for two plinths, in addition to the hanging system for the eleven frames. On one plinth I put the printer, with the Pi and all the cabling hidden in the base, so it looked unconnected to any power supply. I had printed out an exhibit label stating 'THANK YOU FOR SHOPPING AT GOOGLE, Please Take Your Receipt', which I stuck to the front of the plinth. On the other plinth I displayed the entire text of George Orwell's *Nineteen Eighty-Four*, which I had managed to run through {poem}.py (not without difficulty, as explained on the next section). On 26th June 2017, when I processed it, the 1984 receipt came to £58,318 and took almost a whole till roll (and 7 minutes) to print. I had wanted to use 1984 because I use Orwell's idea of Newspeak as a critique of the political power of language (see earlier), so I thought it would be a good way of bringing a overtly political message into the exhibition as a whole. I made another exhibit label : 'Nineteen Eighty-Four (2017)', with a quote from the book, and placed it on the plinth in front of the rolled up receipt. I was quite lucky in that the gallery space I had been allocated had a large round ventilation hole in the wall, which doubled up conceptually as one of the 'memory holes' into which Winston Smith and the other employees at the Min-

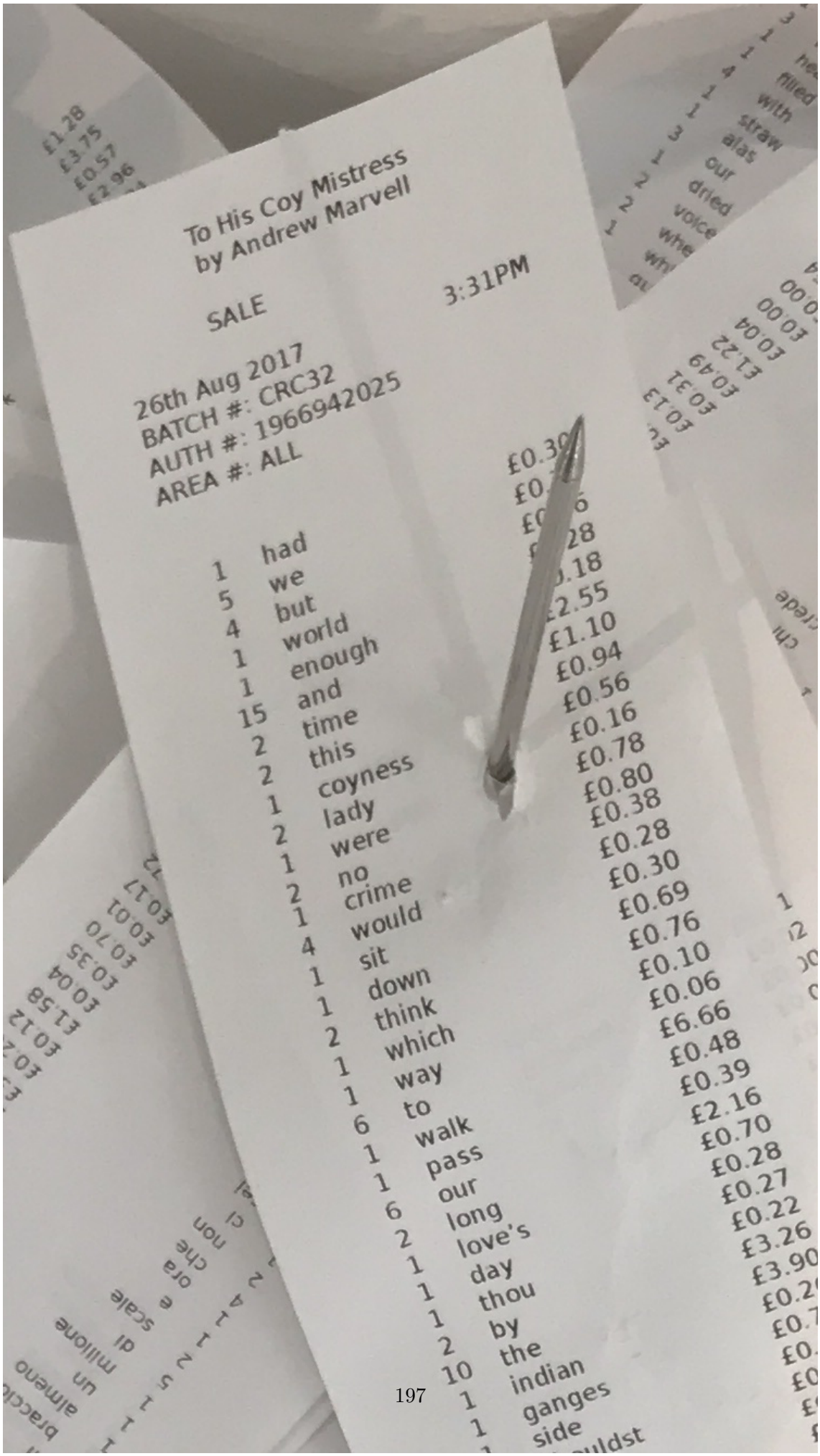
istry of Truth had to throw all their paperwork after altering the factual record in the book.

As the exhibition week went on, with the Raspberry Pi and printer churning out more and more poem-receipts[3] at intervals of between 10 and 30 minutes, I had to start thinking about what to do with them. They were spooling all over the floor around the pedestal which the printer was on, which looked really good, but quickly became messy. Despite the label inviting visitors to 'Please take your receipt', the exhibition curators had told me people were unsure as to whether they were actually meant to be interacting with the display. At this point, another friend (Adam Badger) suggested using a receipt spike - like the ones used in shops and restaurants - to collate the receipts in keeping with the commercial retail theme of the exhibit. I thought this was a brilliant idea. I have always wanted to avoid treating the receipts with kid gloves, as art not to be touched, and have always given them away freely, as (as I mentioned above) I want as far as possible to avoid re-capitalising them. I want them to be expendable, crumpled, lost, torn and forgotten about in the bottom of bags and pockets. Because that is the point. The words on the receipts are expendable, crumpled, lost, torn and forgotten too. I want fading receipts to be pinned to fridges and notice boards, or blue-tacked to office doors until they disappear completely, so skewering them on a spike if people were too slow to collect them off the printer was a perfect way to highlight the expendability of the words to Google. Incidentally, 'To his Coy Mistress' (see Figure 8.5) had been requested by a friend of mine, Michael Flexer. He was about to get married at the time so I decided to give the poem-receipt to him as a wedding present, presenting it to him on his stag night. It was only months later that he admitted he had lost it that same night, probably somewhere on Brighton beach where he had ended up at dawn. He was very apologetic, but I like that one of the receipts was lost in that way. It was rolled up in a plastic tube, so might one day be found. Once again, that is what receipts do.

8.2.3 Destruction and Encryption

The Inter/Sections exhibition also proved critical in tying together the idea of hashing - or encrypting- the poems I fed through {poem}.py with a broader critique not only data processing in an age of digital capitalism, but also of art and its different types of value. One of the visitors to the exhibition had asked me to process a poem called *Agrippa* by William Gibson (author of Neuromancer and coiner of the word 'cyberspace'). I had had a

[3]The receipts coming out were from a database of all the poems I had processed up to that point, so formed a kind of historiography of the project.


To His Coy Mistress
by Andrew Marvell
SALE
3:31PM
26th Aug 2017
BATCH #: CRC32
AUTH #: 1966942025
AREA #: ALL
1 had
5 we
4 but
1 world
1 enough
15 and
2 time
2 this
1 coyness
2 lady
1 were
2 no
1 crime
4 would
1 sit
1 down
2 think
1 which
1 way
6 to
1 walk
1 pass
6 our
2 long
1 love's
1 day
1 thou
2 by
10 the
1 indian
1 ganges
1 side
£0.30
£0.18
£2.55
£1.10
£0.94
£0.56
£0.16
£0.78
£0.80
£0.38
£0.28
£0.30
£0.69
£0.76
£0.10
£0.06
£6.66
£0.48
£0.39
£2.16
£0.70
£0.28
£0.27
£0.22
£3.26

lot of requests during the week and I had processed them all, printing out a version of the receipts from my laptop to give to the person who had requested them, but also saving a Pi-friendly version to the database so that each poem became part of the evolution of the project. But when I was asked to do *Agrippa*, and when I found out about the provenance of the poem, for the first time I refused to do it.

'Agrippa (Book of the Dead)' (Gibson et al. 1992), to give it its full title, is a poem written by Gibson in 1992. It it based around his reflections on finding an old photograph album which had belonged to his father. Somewhat prosaically, *Agrippa* was the brand name of the photo album, but the poem's subject matter of memory and loss was not incidental. It was part of an artistic project and was only ever meant to be an electronic poem, stored on a floppy disk and designed to encrypt itself after its first execution, which happened at a launch event in New York in December 1992, when the poem was shown on a screen and read out by magician/entertainer Penn Jillette (later of Penn and Teller fame). The disk itself had been hidden inside a book, the text of which had been designed to fade when subjected to light, and had been designed as 'a supposedly self-devouring floppy-disk intended to display the text only once, then 'eat itself' (Gibson 2007) (see Figure 8.6). Since the night of the poem's first (and only) official performance there has been much controversy over the project, not helped by Gibson and his collaborators Dennis Ashbaugh and Kevin Begos's hazy and often conflicting recollections of the event itself. In 2012 a US university department launched an international call for cryptographers to decrypt the poem, which was ultimately successful, but despite the supposed destruction of the poem, its text was already in circulation, supposedly the result of a video taken of the poem on screen at the original launch event (for full details see Liu et al. 2005).

Yet despite the original idea of the ephemerality of electronic text its destruction by encryption, Gibson himself did not seem to mind that the conceptual element of the project had not gone to plan, in fact he was pleased that the poem had 'escaped to cyberspace and a life of its own', while also recognising that, in cyberspace, the poem might be subject to 'bit-rot', so he ended up publishing the poem on his website 'with the correct line-breaks, etc.' (2007).

Gibson's *Agrippa* therefore posed a dilemma to me. It was a poem designed specifically to be unreproducible, yet since its performance and execution Gibson seemed to have had

a change of heart. The encryption which 'silenced' the poem became restrictive rather than redemptive, something to be broken free from, rather than cherished, thus echoing the rich and conflicting history of cryptography itself; is it freedom or is it control, and in whose hands? As I mentioned earlier, my {poem}.py project uses cryptography as a conceptual tool with which to 'protect' poetry from the laws of the algorithmic marketplace, thus maintaining its narrative integrity. It is very much a tool of resistance, rather than of destruction. As one paper on the Agrippa case states, 'encryption is not destruction because enciphered text is necessarily subject to cryptanalysis (cracking)' (DuPont 2013). In the crypto wars of the 1990s, cryptography was considered a weapon, which is maybe why the narrative of 'destruction' that DuPont aims to correct in the case of Agrippa captured the public imagination. Indeed, cryptography is still regulated as a potential threat to international security under the Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods (Shehadeh 1999). But true to the original conception of 'Agrippa', and indeed to my own project, encryption does not necessarily mean destruction, but a means of preservation, and resistance. The words in {poem}.py are not escaping from encryption, but from the market. My dilemma, therefore, was whether to honour the original concept of 'Agrippa', and let it rest in peace, so to speak, or to add another layer of intrigue/abstraction onto the poem by putting it through {poem}.py. At the time, and before I had researched the matter, I decided against processing 'Agrippa', but now knowing more about Gibson's reasoning since the poem was first performed/executed (and executed is such a good word here - they did try to kill it!), it might be something I do in the future.

But in thinking about whether encryption is protector or destroyer, it would be wrong to view the poems processed and hashed by {poem}.py as un-damaged by their capture, whether that be by the Google algorithm, or by my framing. One of the things that has most surprised me about the reception of my project, is how keen people are to have their favourite poem monetised, or in effect butchered by the {poem}.py process. Only one person has ever objected to the monetisation of their poem. At one event I had asked participants (as I always do) to send me their favourite poems so I could use them in my presentation and print them out. I do not explain why I ask for these poems. Before the workshop one of the participants had responded to my request to send me a favourite poem. Initially they had emailed me to ask if the poem had to be in English. I said it did not matter what language the poem was in, as processing foreign language poems through

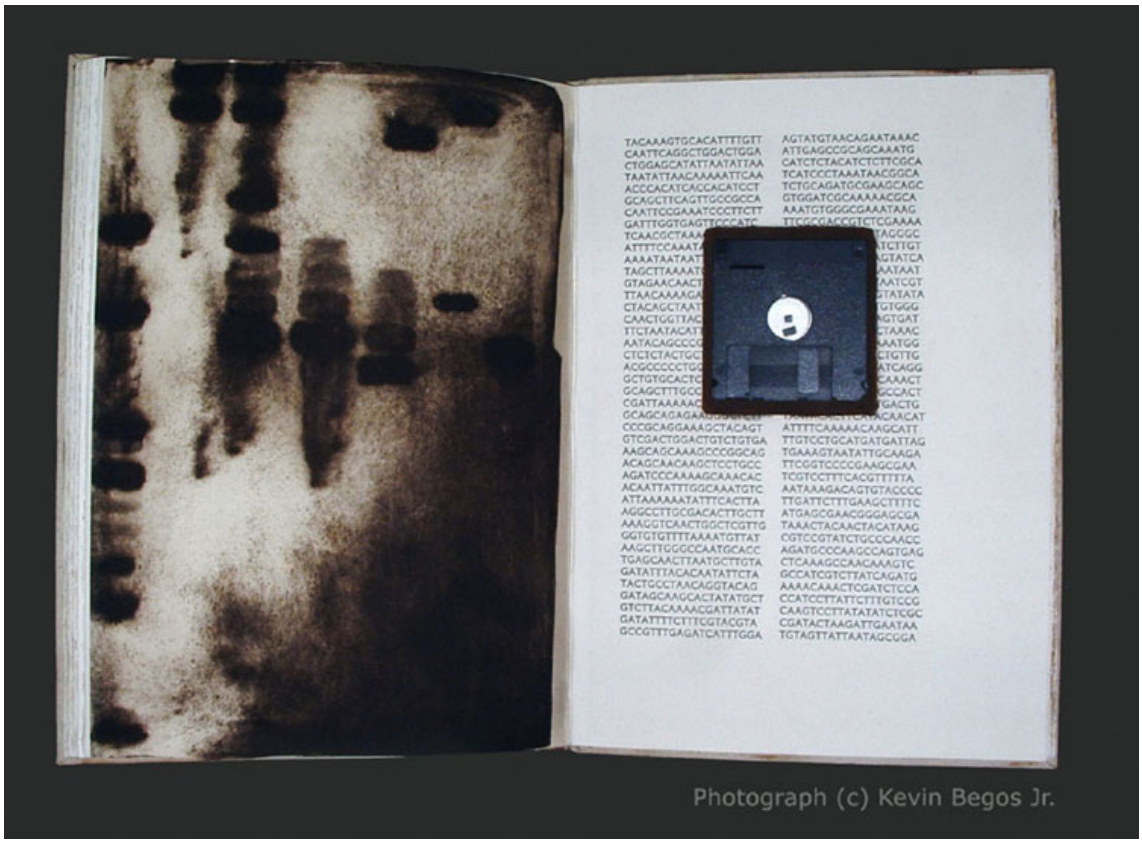


Figure 8.6: 'Agrippa: Book of the Dead', William Gibson et al. (1992).

{poem}.py is often an excellent way of showing how far words are removed from their contexts when they are 'valued' (see Chapter 7).

The poem they had sent me was, as I soon discovered, not even in the Roman alphabet, which was a surprise that highlighted my own Westernised linguistic assumptions, thereby also confirming the inherent privilege of certain scripts above others in the world linguistic order, and also the economic supremacy of English on the web and specifically on Google platforms. The chosen poem was in Russian Cyrillic script, so it was that version I processed through {poem}.py. There is a lot more work I would have to do (perhaps another thesis) to unravel the implications of word pricing between different languages, so apart from some very basic analysis of the differing value of words in different regions/advertising markets detailed in Chapter 7, and the data I have gathered in UK, US and ALL markets, I have not delved into the nuances of actual different languages. Unless I have been processing poems in different places (at conferences for example), I have kept the parameters on the keyword planner to the geographical region, rather than changing the language. This is perhaps another further area of research, the intricacies and implications of which I cannot hope to unravel here, but the Russian version of the poem did indeed prove

interesting. Even though it was processed in an 'English' marketplace, the Cyrillic script was still 'worth' £8.13.

What is very relevant to this chapter about intervention and resistance, however, was the participant's reaction to their receipt. Although they had freely sent me their favourite poem, they told me afterwards that while they were listening to my talk they realised why I had asked for poems to use as examples, and had felt very strongly that they did not want their poem 'fed to the algorithms'. Unfortunately, at the moment they approached me after the talk, I had just printed out the poem, and was busily explaining how interesting it had been to process a Cyrillic poem, when they interrupted me and politely asked for it to be destroyed. Although I would never wish to upset or offend anyone with my work, this was a reaction I had been half-expecting since I started processing people's 'favourite' poems. Throughout this project I have been amazed at how keen people have been for me to commodify and exploit their treasured poems or lyrics and present them with a version stripped of its original aesthetic value. To be fair, it has not stopped me from processing my own favourite texts (although I have regretted doing one in particular), but I really did think I would meet some kind of active resistance sooner. To some extent it was the first recognition I have had that completely validated how strongly I feel about the commodification of language, and that other people might share such a strong conviction. We decided the best course of action was a ritual destruction of the Cyrillic receipt.

8.2.4 Dodging the Market

One of the initial reasons I decided against processing Gibson's 'Agrippa' was that at first I thought that the text of the poem was not freely available on the web. The person who requested the poem had sent me the poem as a .txt file, rather than a link. The poems I use are usually taken from poetry websites, or at least some kind of open source web page. It is important to the creative vision of the project that the words being 'valued' by the KWP have in theory all been potentially exposed to monetisation in some way, for example appearing on sites which also host adverts, or even merely by virtue of being openly indexed and available to be found by search algorithms. In order to ensure this, I only use poems which have been cut and pasted from the web, or (as the project progressed) have been sent through GMail and have therefore been made vulnerable to the algorithms that scrape emails for targeted advertising purposes (as per Cabell and Huff). As I have mentioned earlier, for the 'Living with Algorithms' collection, I had

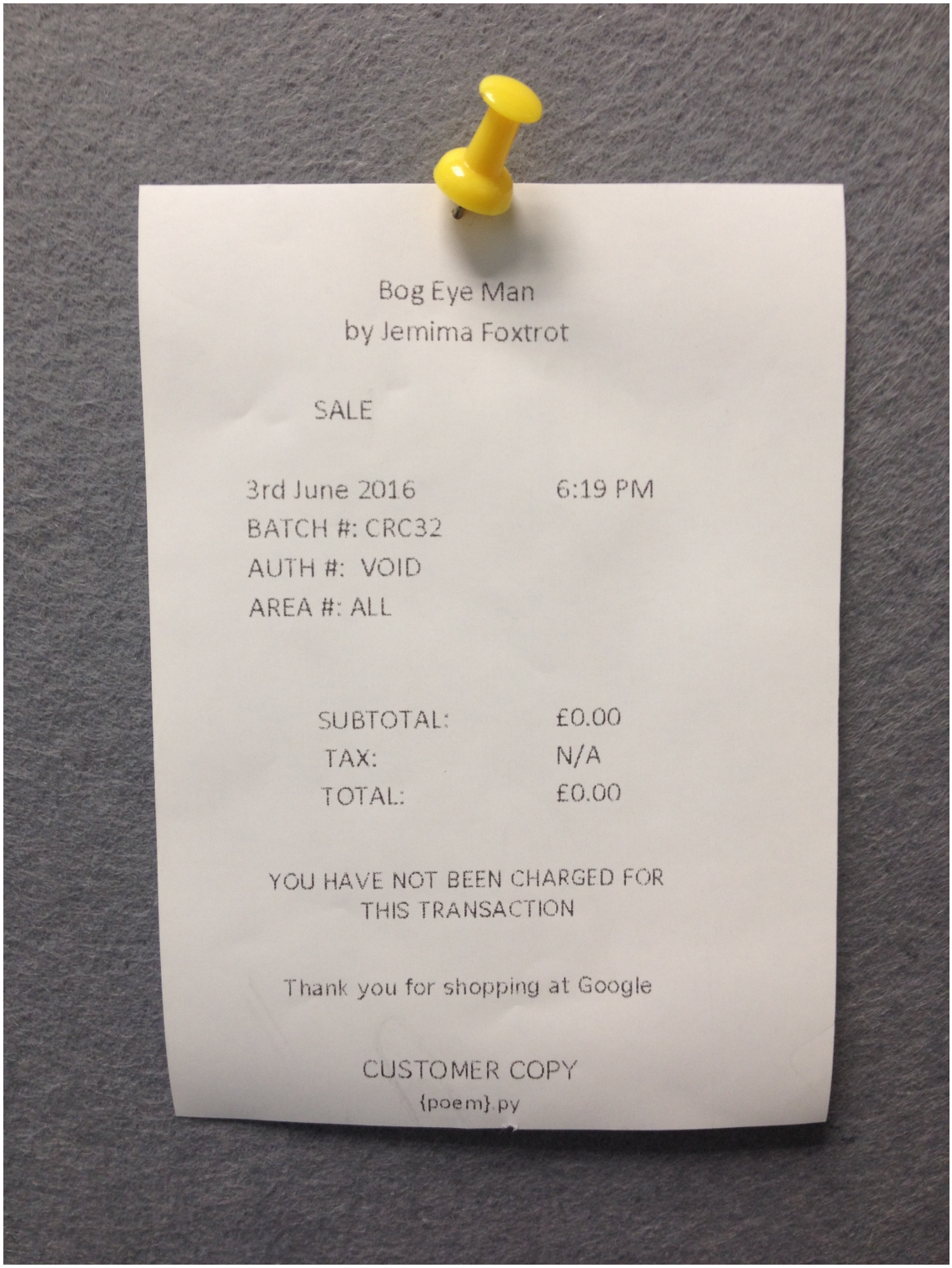


Figure 8.7: Jemima Foxtrot's 'Bog Eye Man' ({poem}.py 2016). Photo: author's own.

asked participants to send me poems before hand, and two of the suggested poems turned out to be extremely interesting in that they both, for different reasons, managed to resist the process of commodification and the forces of linguistic capitalism, and were therefore early indicators of the potential for other accidental and deliberate forms of resistance and critical theorisation that I went on to explore as the project progressed.

The first of these disobedient texts was a poem called *Bog Eye Man*, by spoken word poet Jemima Foxtrot (2015). The workshop participant, Sam Hind, sent me a Vimeo link to the poem being performed, and the piece is also accessible on YouTube. What proved interesting to my project is that the actual text of the poem does not appear on the web, so I was unable to 'scrape' it and feed it through {poem}.py. The other poem which resisted the process was contained within a jpeg file from which I could not copy and paste. These examples served to provoke an early realisation of the inherent differences and functions not only between words on paper pages as opposed to webpages, perhaps evoking Socratic objections to the externalisation of knowledge, but also between words which have been digitised as text, and those which exist online but in other digital formats. The digitised words which are then monetised by Google carry a paratextual (as opposed to contextual) commercial agency which the others do not. As I was effectively unable to 'price' those poems, this was also an early indication of the potential ways to resist and subvert the systems of linguistic capitalism.[4] For the purposes of the workshop, I therefore presented Sam Hind *Bog Eye Man* with a VOID receipt which stated YOU HAVE NOT BEEN CHARGED FOR THIS TRANSACTION. This void receipt is thus a key part of the collection (see Figure 8.7).

8.2.5 Reclaiming poetry from algorithmic market

As discussed previously, it is not possible (and maybe not even necessary) to be able to open the black box to learn from the algorithmic processes that increasingly mediate and control our lives. In this respect, the {poem}.py intervention has always been a means of shining a light on the potential effects and consequences they might have in order to critique, but also to raise awareness by making the processes more visible, and also more personal - more close to home. Apart from showing {poem}.py at conferences and exhi-

[4] Although recent developments in the monetisation of speech data sent over Voice Over Internet Protocol (VOIP) platforms such as Apple's Facetime or Facebook's voice and video calling have already challenged this mode of resistance.

Figure 8.8: Engineering Fictions workshop, most expensive love poem, 'Bondage' ({poem}.py 2017). Photo: Jessica Foley.

bitions, one way I have tried to do this was by conducting an interactive workshop where I invited people to write what they thought was expensive and cheap poetry. The workshop took place after a seminar I gave at Trinity College Dublin in February 2017 as part of the Engineering Fictions series at CONNECT. After explaining how AdWords works and demonstrating {poem}.py, I then conducted a smaller workshop where I invited participants to write what they thought would be the cheapest and most expensive poems. It was Valentine's Day so myself and Jessica Foley (who had invited me), decided they should be love poems, each 13 words long, with no repetitions. Participants had access to some poem-receipts I showed them as examples, but were not allowed to use the Keyword planner for help. I just asked them to think about all the things I told them about geographical markets, the time of year, the value of out of context words to advertisers and so on. Once they had composed their poems, I asked them to email them to Jessica's GMail account. This was in order to maintain the conceptual integrity of the project only processing words which have technically been potentially 'exposed' to Google's algorithms in some way. The workshop was great fun and also extremely interesting methodologically. By creating poetry manually, and in the firm knowledge that the words which pass through Google's advertising and search platforms are always already infused with algorithmically and economically mediated 'values', it felt like we were able to reclaim some of the artistic agency from the algorithms that increasingly second guess our linguistic intentions. We were in effect able to second guess the second guessers, and it felt really liberating! The participants really engaged with the task in hand, trying to work out the 'value' of the words they used to construct their poems. When they had composed and sent the poems through GMail, I then processed them through {poem}.py and printed them out and we held a mini results ceremony. The results were fascinating. Some of the poems were really beautifully written, but it was these ones which held the least economic value. The expensive poems, in contrast, were (in my opinion) ugly and brash. It was therefore a great way to visualise poetic value in direct tension with economic value (see Figures 8.8 & 8.9).

Another way I have tried to 'reclaim poetry' is with a kind of reverse engineering of Google algorithm. Once again, I cannot know know exactly how it works, but I can show what goes in the black-box, and what comes out, which is arguably more important than the technical workings anyway. To do this, I returned to the poem that started this whole project off, my favourite poem, William Stafford's 'At the Bomb Testing Site' (1960),

which was the first poem I 'valued' as part of my {poem}.py project, and I fed the whole poem (unprepared by code) through the KWP. When you feed whole phrases (or lines of poetry in this case), through the KWP, Google provides an AdGroups function, which in effect tries to second guess what you are trying to market, and suggests other keywords, phrases and topics which might help enhance your advertising campaign. The results are an interesting insight into what the algorithms *think* you mean, and also what the words mean to Google. When I first ran the poem through AdWords, references in the poem to a curved desert road, hands gripping and tense elbows, generated suggestions that I was trying to advertise road biking. The phrase 'ready for change' had the algorithms thinking I was planning a well-being or recruitment campaign. But these results are as dynamic as the suggested bid price system. When I tried the experiment more recently, the road biking suggestions had gone (perhaps we were out of biking season), but other even more fascinating semantic assumptions had appeared. 'At the Bomb Testing Site' was now conjuring up references to Carl Jung (presumably because of the image of the midday sun and the reference to 'self', gastric bands (the Mexican desert is apparently popular for weight loss surgery, and bikinis (as in the Bikini Atoll bomb testing site). So in the spirit of reclaiming the poetry from the algorithmic marketplace, I decided to reconstruct Stafford's algorithmically interpreted poem using only the suggested advertising categories and potential related search queries offered to me by Google when I put the poem through the keyword planner. With apologies to the estate of William Stafford, here is the original version of the poem, along with Google's interpretation.

At the Bomb Testing Site
by William Stafford

At noon in the desert a panting lizard
waited for history, its elbows tense,
watching the curve of a particular road
as if something might happen.

It was looking at something farther off
than people could see, an important scene
acted in stone for little selves
at the flute end of consequences.

There was just a continent without much on it
under a sky that never cared less.
Ready for a change, the elbows waited.
The hands gripped hard on the desert.

At the Bomb Testing Site (2017)
by Google AdWords & Pip Thornton (after William Stafford)

Im feeling stuck.
Atomic trinity: anger, depression, ego
and archetype elbow pain after fall.
California republican
delegates latest nuclear test.
Popeye syndrome.
Who invented the hydrogen bomb?
Carl Jungs shadow?

I dont like myself.
Business goals, data entry jobs,
weight loss surgery in mexico.

I am ready to change my life
Self referral, mental health
define psyche.
Inner self crossword clue.
Feel joy! Wellbeing,
core beliefs,
gastric bypass,
bikini island.
Ready steady: be yourself.

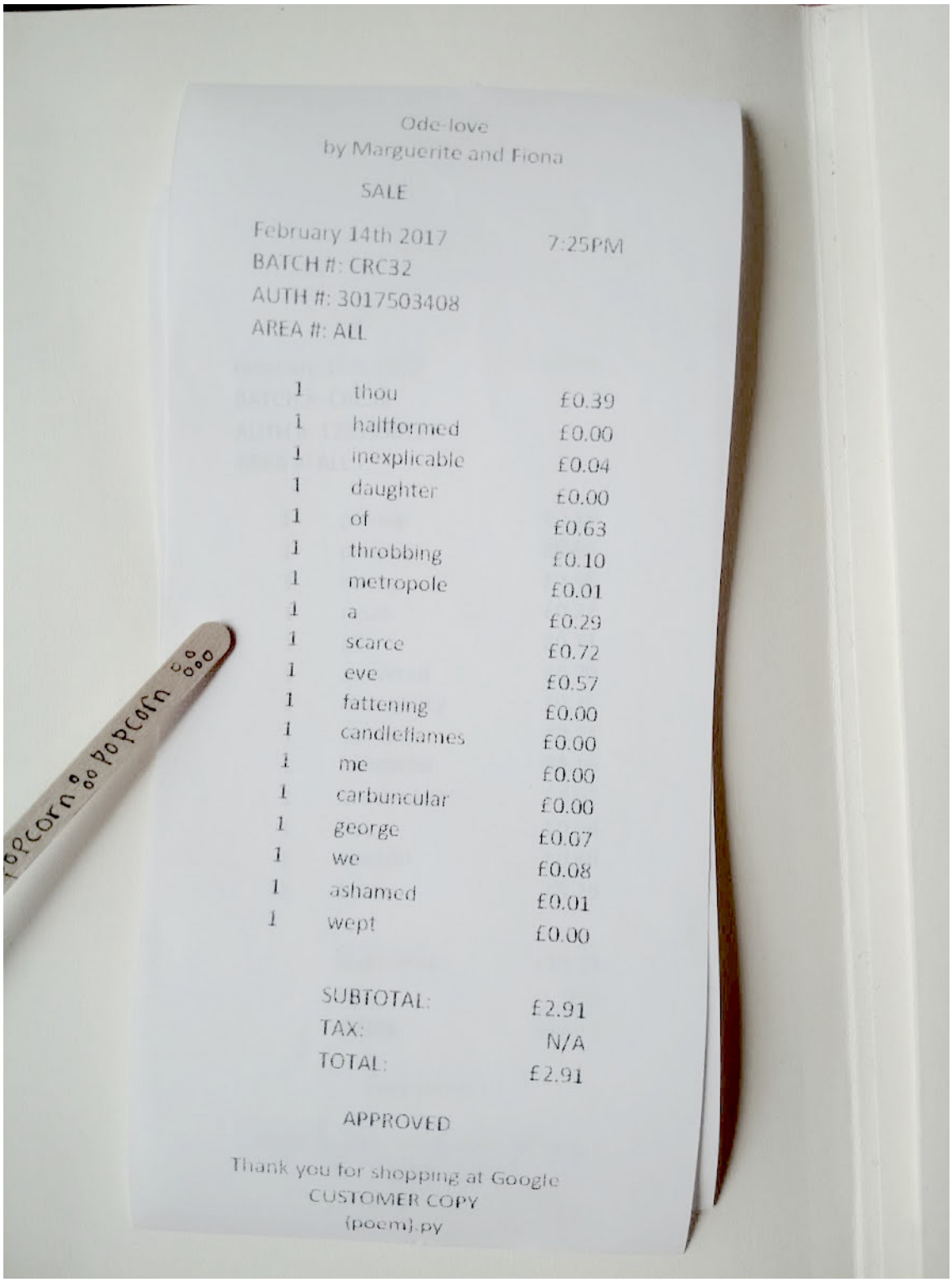


Figure 8.9: Engineering Fictions workshop, cheapest love poem, 'Ode-Love' ({poem}.py 2017). Photo: Jessica Foley.

8.3 Resistance and Persistence

As I mentioned in Chapter 3, there are significant methodological barriers in any research into dynamic and opaque digital algorithmic systems. What I have attempted to do with {poem}.py is to accept these barriers as important pieces of the black-box jigsaw, documenting active and passive modes of resistance and subversion as the project progressed, and weaving the theory in to the practice of developing the artistic intervention as I have gone on. But even accepting the hurdles in this method as part of the creative process, there have been some less prosaic spanners in the works that have also added to my overall critique and also to the investigation of wider questions about the difficulties of researching opaque and proprietary digital technologies. Raging against the machine can sometimes be exhausting, especially when the machine is as powerful and embedded in society as Google is today. To make progress against such hurdles requires persistence. This section will therefore detail some of the methodological problems I have encountered along the way, perhaps detailing a resistance of another kind; the resistance built into the system to deter those seeking to game or hack it, but also those seeking merely to understand and research it.

The artists and academics I have mentioned in these last two chapters have all come up against such problems. Researching black-boxed algorithms is methodologically challenging in many ways, not least due to ethical considerations (Hinman 2008; Sandvig 2014; Ananny 2016), but because the information on the Web is inherently unstable and malleable (Kitchin 2017), and because personalised search has made it near impossible to collect empirically sound, geographically and temporally stable data. As Feuz et al. point out, obtaining reliable empirical data from search engines is almost impossible as there is effectively no control or baseline to work against (2011). Added to the problem of curated and ever changing results and ever-tweaked algorithms, is the fact that at any given time Google or other search engines could be conducting A/B type controlled test experiments on users. The recent furore over the Facebook 'emotion' experiments (Flick 2016; Panger 2016) show the ease and propensity with which users of networked technologies can be experimented on, and Google are constantly assessing and tweaking their algorithms to maximise profit through 'successful' clickstreams. But as well as monitoring for improvements, the Google crawlers are constantly on the look out to block spam and automated searches and 'link farms' which are used to boost search rankings, which means that legitimate research or investigation gets caught up in these purges too. Feuz et al.'s philosopher

experiment (as detailed earlier), for example, ran into trouble when Google blocked several IP addresses they were using assuming that they were processing automated requests from a virus or spyware application. Similarly, Christophe Bruno's Adwords 'Happening' (2002), during which he experimented with buying keywords which linked to abstract poetry was taken down by Google because it perceived that the adverts were not being successful enough as Google polices AdWords to make sure that commercial profit is the *only* motive for the purchase of keywords, rather than ideological or political motivations (Diaz 2008). Poorly written, and under-performing adverts are flagged up by the algorithms. Luckily for his budget at least, only a few users clicked on Bruno's poem-adverts while they were 'live'. His keywords may have had poetic, or creative capital, but they had no commercial context, and therefore earned Google little money.

8.3.1 De-bugging {poem}.py

The development of {poem}.py and my other AdWords data gathering exercises were not without their own methodological and systemic problems. Keeping the project going often feels like a continuous battle against ever-changing and resistant technology. It has, for example, been impossible to make the processing of each poem any quicker than 2-3 minutes because Google does not provide an Application Planning Interface (API) for non-paying AdWords users, and I have always maintained I did not want to enter the marketplace and end up paying into Google's pockets. The code developed by Ben Curtis, and thereafter Feargus Pendlebury is therefore a 2-step manual process, which it has so far been impossible to speed up. Even using a automation tool like Selenium was unsuccessful, because as a non-paying Google user, I am constantly (but randomly) being prompted to start an active campaign. So each time an unexpected prompt or info box pops up in the middle of the automated program, the sequence is interrupted and therefore useless.

There are also occasional (unannounced) changes in the format of the KWP data. As mentioned earlier, the data can be downloaded on spreadsheets in normal column/row format, which makes it easy for the code to locate and extract the necessary information. Until the columns switched around, that is. And then switched back again a few months later. It also took a long time to work out why at one stage in late summer 2016, the poem-receipts were coming out truncated, and with the prices of words not corresponding to the data on the spreadsheet. It transpired that this was due to the 'close variant' update

which stopped AdWords users receiving suggested bid prices for exact matches of the words they were searching through the KWP interface. Warning users that 'some keywords that you entered and their close variants have been grouped into one row', Google's update was met with consternation by the SEO industry, as such unannounced changes can have huge effects on campaign budgets, as well as being a fairly obvious attempt by Google to harvest more hits and clicks by widening the potential for searches to match keywords. One article explained that:

> the Keyword Planner now seems to combine many search variants, including: plurals with non-plurals for any word in the keyword phrase, acronyms with longhand version, stemming variants: -er, -ing, -ized, -ed etc keywords (i.e. designer, designing, designed), words that can be spelled with or without space (ie. car park and carpark), [and] words with and without punctuation (ie. kid toys and kids' toys) (The SEM Post 2016).

What this tweak meant to my project became apparent when I tried to process John Masefield's 'Seafever' through {poem}.py. The poem has the possessive words sail's and wind's in it, which under the updated KWP revert to the words sail and wind. The resulting spreadsheet therefore (with multiple 'sail' and 'wind' rather than sail's and wind's on it), did not match up to the original .txt poem file, so the code could no longer match the inputted words with the list outputted by the KWP. The list of words was therefore less than the list of prices on the spreadsheet, so the receipt template was cut short and the words which did appear were skewed in terms of their corresponding prices. I eventually found a work-around for this problem by saving each group of words as an individual 'plan', rather than inputting them as a string in the main search bar. For some reason doing it this way, although more time consuming, meant that the integrity of the words was retained, although I am also very much aware of the precarity of my project should Google also force the close variant update on individual plans. It was a similar methodological hurdle as experienced early in the project with the adult filter on the KWP, which removed the word 'strips' from Simon Armitage's 'A Vision', and 'fuck' from Ginsberg's 'America', which had the same skewing effect on the data.

I have already mentioned an early problem before we realised that the KWP censors certain words as 'adult' unless you actively opt in, but there have been other bugs to get over too. The code has been instrumental in automating the project and returning the words of a poem or text to their narrative order once they have been revalued by Google in

terms of their search volume, suggested price and competition rating, but even with this help, there are still more para-textual technical anomalies. The code works on the Excel spreadsheets provided by the Keyword planner, but Excel (and other Microsoft products) are notoriously frustrating for the hidden formulas that can seemingly distort the whole format of simple Word document, or randomly turn a money value into a decimal, or a date. The Excel characteristic that flummoxed me for a long time (and I still have not managed to correct), is that the words TRUE and FALSE in a spreadsheet automatically stand for a formula in a table, so if they appear in a poem, their 'value' is once again compromised and they disrupt the execution of the {poem}.py code.

8.4 Speculative resistance

In the final section of this chapter, I want to turn to a method of resistance, the power of which I have already harnessed as a means to critique linguistic capitalism (with Orwell), and that is the use of creative writing. Books like Orwell's *1984* and Huxley's *Brave New World* have proved both useful and prescient in contextualising and popularising the narrative and debate around the political implications of the power and ubiquity of digital technologies and economies, while science fiction films such as Minority Report (2002), (one of many based on the novels of Philip K. Dick), with its predictions of pre-crime technology and predictive policing, are quickly becoming 'future fact' (Kitchin and Kneale 2001). Science fiction, or speculative fiction, is becoming an increasingly popular means of theorising and conceptualising critical reactions within academia too, although it is a method long championed by cross-disciplinary scholars such as N. Katherine Hayles. Indeed, more than simply reflecting reality and the anxieties of technology, with 'influence flow[ing] from science into literature', Hayles uses writers such as Philip K. Dick and William Gibson to demonstrate how literary fiction actually has a reciprocal relationship with science, playing an active role in 'shap[ing] what technologies mean and what scientific theories signify in cultural contexts' (2008: 21). More recently, Will Davies' edited collection of 'Economic Science Fictions' (2018), and Rob Kitchin et al.'s forthcoming 'Running a City Like a Company and Other Fables' (2019) are examples of this reciprocal and productive approach. Conceptually imaginative work has always been important political critique, but I think it is even more important today in age of algorithmic reproduction, when art and culture, and the words we use to live/communicate are virtually all mediated/tainted/manipulated by systems such as linguistic capitalism. Although my

{poem}.py project 'exposes the grain' (Rogers 2013) of the process by working with the idiosyncracies of the system, while companies like Google hold the power they do, there seems very little means of actual redress or effective resistance, although perhaps we can hope for more active resistance from Google employees and technicians themselves in the future (as we are beginning to see in some resistance to Project Maven, see Chapter 5), or for a break up of the Google monopoly, or for viable alternatives, but those are subjects for other theses.

In Chapter 6 I explored the concept of linguistic capitalism through the lens of Hayek's knowledge argument in order to suggest that digital linguistic capitalism has created a kind of hybrid market that manifests itself as a form of algorithmic governance, with dangerously political side effects. But what if we take this further? If it is some kind of market that is governing us on the sly, then what would happen if that market grew and consolidated into a more overt system, and then what would happen if that system were to crash? In his analysis of Frederic Kaplan's original 'Linguistic Capitalism and Algorithmic Mediation' piece (2011), Franco Bifo Berardi discusses the effects of the commodifiction of language, and insists that 'insolvency is the line of escape from the reduction of language to exchange' (2012: 17), but linguistic insolvency has severe consequences. It might spark creativity in the short term, but it is ultimately destructive. Indeed, perhaps linguistic insolvency is even more damaging than financial insolvency, a proposition I am now going to explore through a creative conceptual provocation and an accompanying short piece of speculative fiction. Exploring the idea of 'Subprime Language' and the potentially disaterous effects it might have on the digital economy, the combined speculative theoretical and creative fiction approaches also serve to pull together several of the strands of the thesis so far.

8.4.1 Subprime language

> *[T]he failure of the financial system in 2007-8 in the United States was primarily a failure of language. This argument does not deny that greed, ignorance, weak regulation, and irresponsible risk-taking were important factors in the collapse. But the new role of language in the marketplace is the condition of possibility for all these more easily identifiable flaws (Appadurai 2015: 1).*

In the last section of this chapter I take Kaplan's concept of linguistic capitalism (2014),

around which much of the thesis revolves, and extend the financial metaphor into the realms of speculative fiction, imagining that the language flowing through the portals and platforms of the web has become so infused by the logics of the neoliberal market, that it risks becoming 'subprime'. As Berardi writes, the monetisation of language through Google's AdWords platform has meant that the economy has become the new 'universal grammar' (2012), and the argument I make in this thesis is that in a digital age, it is this monetised grammar that increasingly both constructs and deciphers language. The way the web works at the moment, with Google (an its subsidiaries) mediating and exploiting the circulation of monetised words, the potential for political and social influence comes often as a side effect of the economic incentive; through Macedonian teenagers exploiting AdSense with lucrative 'fake news farms' (Graham 2017: 12-13), through advert-heavy unofficial YouTube channels (Bridle 2017), or through manipulation of the organic and paid SEO industry, as I concentrate on in this thesis. In this respect, the economically optimised words circulating via these platforms and systems run the risk of creating dangerous and damaging stories. This is in part due to the platform facilitated circulation of cultural capital and ideas made possible by the data-isation of language (see Chapter 4), and new forms of 'digital economic circulation' (Langley & Leyshon 2017: 17). While concentrating on exploiting language for money, Google have in effect let money control the narrative.

As tech companies such as Google increasingly mediate and monetise the informational landscape through search and advertising platforms such as AdWords and AdSense, the ongoing effects on and of the language they auction, sell and exploit are becoming more and more palpable. In the viral spreading of fake news, political and cultural click-bait and in the daily battles for exposure, it seems that words are being lent against a narrative so tenuous as to make their linguistic function negligible. Infused with a neoliberal logic which favours advertising dollars over truth and the systemic bias of algorithmic processing, the discursive side-effects of this semantic shift reveal a deep-rooted weakness in a linguistic marketplace in which the value of language has shifted from conveyor of meaning to conveyor of capital. In this marketplace, words have taken on their own economic value and circulate as commodities, or more accurately as derivatives, giving rise to new levels and types of performative agency, whereby potentially dangerous narratives and content are created and spread as collateral side effects to the ad revenue and SEO industry. In this respect, companies like Google have taken on the role of language brokers - a situation

which has potentially grave linguistic and political consequences.

My aim with the speculative argument developed in this section is therefore to question and expose the new neoliberal and dystopian narratives being created in this process, asking questions such as how much and how often language can be bought, sold or 'borrowed' before it becomes exhausted of meaning and restrictive of expression and understanding? Has language become so infused with economic logic that it has lost the liquid qualities of creative language? How resilient is language to a quasi-capitalist operating system? Can language in this way be seen as 'subprime', 'illiquid' or 'toxic', and - perhaps more provocatively - what happens if/when this linguistic bubble bursts?

As I have explained in the thesis, Google's wealth and power as a company is primarily a result of the success of its AdWords platform, which auctions keywords and phrases to advertisers in return for the top spots on the search engine results page. This is a process which Frederic Kaplan has called 'linguistic capitalism' (2014), where Google monetises - or commodifies - language in a way which perhaps elevates the exchange value of words ahead of any underlying aesthetic, or ordinary value it might have. While not entirely unproblematic, I want to extend the financial metaphor of linguistic capitalism further. In his 2015 book 'Banking on Words: : The failure of language in the age of derivative finance', Arjun Appadurai argues that the Global Financial Crisis of 2008 was 'primarily a failure of language' facilitated by the 'new role of language in the marketplace'. By this he is talking about the language of finance, and the words used to perform finance, as having derivative value. In such a way words become promisary notes, or contracts, which 'systematically and contagiously' failed to deliver (2008: 1).

Appadurai is talking here about the performative effects to language (Austin 1975), pointing out 'the critical link between the numerical force of money and the linguistic force of what we say we will do with it' (2015: back cover). What I want to do is to take the synthesis of language and capital in a different, perhaps deeper direction, proposing that in addition to performative agency, in an era of digital linguistic capitalism, words have now taken on their own economic value, as commodities, or more accurately as derivatives. Taking Appadurai's suggestion that the Global Financial Crisis of 2008 was 'primarily a failure of language' (2015: 1) in a different direction, I want to start thinking about what terms such as *value, derivative, liquidity, velocity and subprime* mean in an age of digital

linguistic capitalism, using them both as tools of critique and also perhaps as means of metaphorical redemption. Following Derrida, I argue that the underlying value of language is 'différance'. In other words, the value of language lies in its liquidity - a constant deferral and transferability of meaning. It is this polysemic quality that gives language the capacity and richness for argument, ambiguity, wordplay, or poeticism, for example. In this way it could be said that language has always been derivative, its value is, as Roland Barthes and others have concluded (see Chapter 5), always realised in a particular context that is not necessarily the one in which it was written. Franco 'Bifo' Berardi states that 'debt is an act of language' (2012: 31). It is, like Appadurai suggests, performative. But it works the other way around as well. Language is an act of debt. Its value is realised through a spectrum of debt and credit, and is as such always already a derivative based on a precarious system of trust.

The concern here then, is what happens to this performative value when the word is transformed into an economic derivative, which is what happens when language becomes digitised and monetised in the linguistic marketplace? Governed by an economic derivative value, the fluidity of words, or what we might call their *linguistic liquidity*, is on the one hand increased by virtue of its dissolution into data (which accelerates its circulation), but is at the same time tied to an economic promissory transaction that becomes *indifferent* to the underlying value of language. And this is where the risk lies. As Randy Martin writes, 'disaster ensue[s] when derivatives escape[d] from their rightful place and inundate[d] other quarters' (2013: 85). So we might say therefore say that the economic derivative value of words has become indifferent to their underlying value just as housing bonds (and those that traded them) became indifferent to the underlying value of the real estate market before the Global Financial Crisis in 2007-8. Always already derivative, digitised words have now taken on *economic* derivative value (a kind of hyper-différance perhaps), but the problem comes when the economic value takes over; when words become illiquid, tied to the version of themselves which promises the most profit rather than conveys the most accurate information, a process that brings into existence dangerous narratives as collateral consequence.

8.4.2 Derivative value and indifference

As I have explained throughout the thesis, one of the integral parts of Google's AdWords platform is how it 'value' - or capitalises - language by assigning an algorithmically generated 'suggested bid price' to each word or phrase. This is a projection based on several variables. Historic search volume is one of them, but like Google's other opaque algorithmic calculations, it is impossible to know exactly what data points are being used. While these bid prices are just meant to provide potential advertisers with enough knowledge to plan budgets and enter the market at a reasonable level, and are not necessarily the same as the actual cost per click, they are in effect the front-end valuations, and therefore have a degree of performative agency, value, and power in how they influence the words selected to populate the web as adverts, as articles or other optimized text and tags. Whether as data points for paid advertising or for organic search engine optimisation, as Louise Amoore states about the data derivative, they are 'score[s] that will go on to live and act in the world' (2011: 27). But this is the disconnect of language in linguistic capitalism; commodified words become 'indifferent' to their underlying performative value. As seen through {poem}.py, when you type the word 'cloud' into Google (see Chapter 7), the results will relate not to clouds in the sky, or a Cumbrian vision, but to cloud computing, even if you had intended to search for the meteorological cloud. The only thing the search result is 'different' to, is the *economic* value of the words it displays.

Another example which illustrates this 'indifference' of language monetised in this way is Cabell & Huff's 'American Psycho' intervention, in which they sent each page of the novel American Psycho to each other via GMail. Knowing that Google scrapes GMail messages for the purpose of targeted advertising, they then recorded the adverts triggered by the words in the text using phantom footnotes, thus showing that scenes of torture and violent misogyny featuring the skinning and dismemberment of women, and the cooking of various body parts generates adverts for skin tightening products, teeth whitening, and microwave meals. I have mentioned their intervention several times during the thesis, but here it serves to speak to the idea of algorithmic cognition, or algorithmic reading. The 'meaning' or 'value' scraped from the pages of Ellis's words is defined purely in terms of what they might be converted to economically in the future, or to put it another way, their derivative value.

I want to turn now to another of Google's advertising platforms, AdSense, and its role

in the fake news scandals that dominated the narrative of social media platforms such as Facebook in the run up to the US Presidential Election in 2016.

In January 2017, the New York Times reported the story of an American college graduate who needed money to pay off his loans - and how he fabricated and published a fake news story which went viral and earned him thousands of dollars of ad revenue from Google (The New York Times 2017; see Figure 8.10).Although he identified as a Republican, the man who did this admitted that politics was not his primary objective. He knew that with every visit to his website he earned money from the Google ads displayed on it - this is how AdSense works; you submit your website to Google and get paid for each time Google are able to serve an advert on it. The advertiser pays Google, and Google pays a small amount to website hosts, and keeps the difference. In order to generate clicks to his site, he used Facebook as a means to spread the story which then gained traction among Trump supporters and spread through Facebook through likes, comments and shares, eventually reaching 6 million people. By this method, the link clicks to his website from the Facebook generation, and the resulting adverts served to its visitors by Google, earned him over $20,000. Furthermore, the popularity of the website and its revenue earning potential (which he had bought for $5 and which was made up of purely fake news stories) gave it a resale value of $100,000 within weeks.

Obviously some of what is now being called fake news has political motives, for example the Trump dossier, and other historical political propaganda campaigns, but the point I am making here is that in today's digital economy, language can have economic value that is completely detached from what the words actually say, i.e. the value is purely financial. And not all the misinformation spread unwittingly by Google advertising before the US election was 'home grown'. The Guardian, and other outlets have reported how small towns in Macedonia and other Eastern European countries had in effect become a pro-Trump fake news factories (The Guardian 2016b; BBC News 2016, see also Graham 2017). Teenagers in these towns had realised how much money they could make from Google AdSense and set about creating websites with sensationalist content that they knew would spread. Unfortunately for the Clinton campaign, the stories which generated most traffic and most links (and thereby infecting the rest of the news content on the web), were full of pro-Trump right-wing propaganda. The important thing about these examples is they illustrate the disconnect between the value of the words - or content - which make up

Figure 8.10: The New York Times, 18 January 2017.

these dodgy websites, and the value they accrue through the adverts they bear. Monetised language becomes indifferent to any other value or meaning the words might have. This is, however, not necessarily a new thing if we think about how clickbait stories have developed over the last few years - made up, or twisted, celebrity gossip or shock value stories written purely to gain clicks, eyeballs and advertising revenue. But when the so-called information changes from celebrity gossip to the campaign trail, then the political fallout becomes a very dangerous by product. James Bridle has highlighted the cultural damage that can also be done by this indifference of value (2017). He points to print-on-demand merchandise spreading misogynist messages via Amazon (see Figure 8.11), and to the thousands of fake or morally dubious children's cartoons on YouTube which are made up of violent or sexualised content in order to generate more views and more advertising revenue for their makers. There are also the numerous examples of the racist, sexist or extremist search results and auto-completions that shape what is 'knowledge' according to Google, but are in effect byproducts of the digital economy, producing an excess of language that releases new and dangerous narratives into the discourse. As discussed in earlier chapters, what comes out of the search engine can have serious social and political effects (Thornton 2018, Noble 2018).

8.4.3 The Circulation of Economic Value

What I have been trying to show with these examples is that in an era of digital capitalism, words have become financially valuable in a way disconnected from their linguistic function/agency. As commodities, or derivatives, the way they move and circulate as capital therefore becomes an important issue, especially as they now have more than one agential function (i.e. linguistic *and* economic). The important point to make here is that, as David Harvey writes, capital is a process of circulation, or, following Marx, it is 'value in motion' (2017: 74). Fixed capital will depreciate in value, so to work it needs to be in 'perpetual movement', and the physical liquidity and potential velocity of language as data facilitates this movement in both an economic and narrative way. But this physical liquidity of digitised language is at odds with the liquidity of language as in Derrida's différance. In linguistic capitalism there is a risk that words become 'toxic', or 'illiquid', in some ways like the illiquid subprime mortgage securities that no one was able to value or wanted to trade in 2008. They might wear out as keywords, or they might become so that they cannot be liquid anymore; their meaning/value realisable in economic terms only, as I demonstrate with {poem}.py and the 'cloud' example.

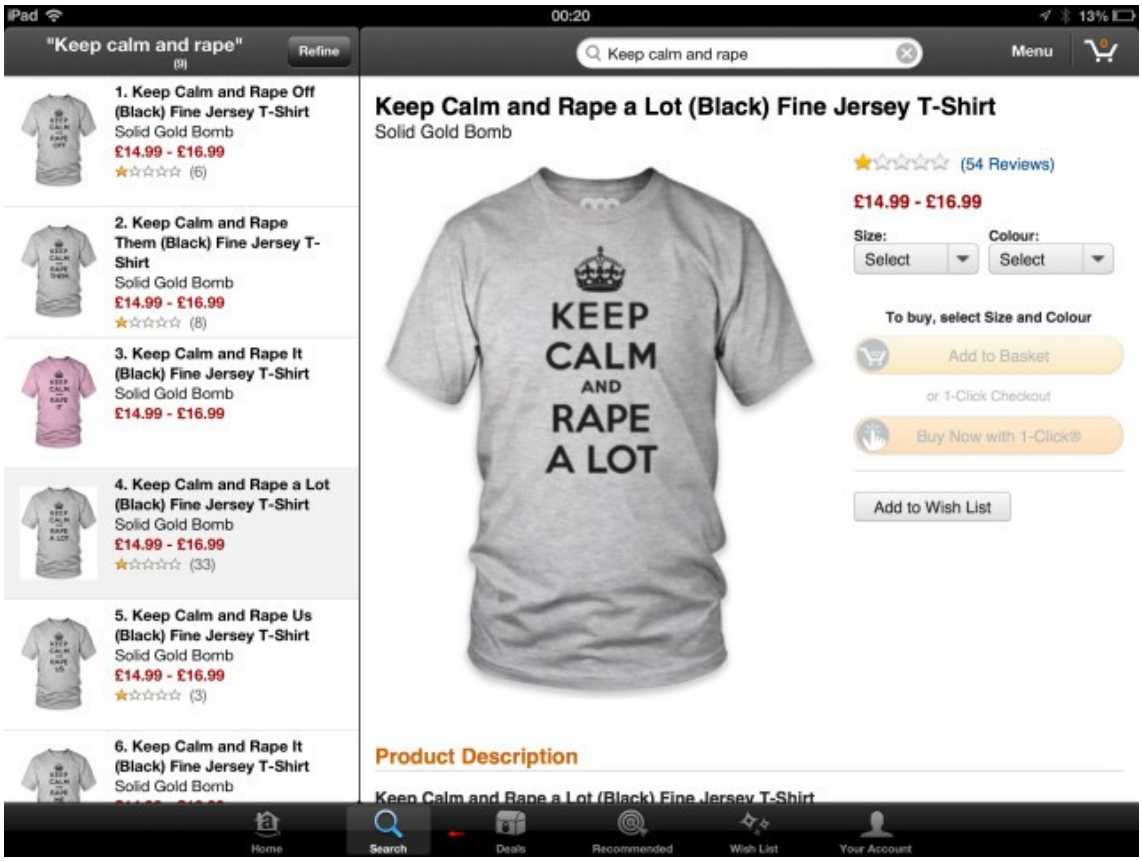


Figure 8.11: Amazon 'Keep Calm' image (Bridle 2017).

The argument here is that the underlying performative (linguistic/narrative) value, and the derivative economic value of language, have different velocities of circulation. For example, as the toxic, economic values of fake news content (or the disturbing advert carrying narratives that appear on YouTube) increasingly circulate, the underlying value of words is lost - it becomes illiquid. The liquidity of one value is predicated on the illiquidity of the other value. And for one to thrive and be totally liquid, the other must have a crunch (Thornton and Morris 2017). My thesis is primarily about linguistic capitalism as relates to Google, so in conclusion to this section, I have to ask: *What is Google's role in all this?* To continue with what I think is a very productive method of speculative and conceptual imagination, I suggest that Google, with its power, reach and now near ubiquity in everyday life, is both bank, merchant and broker of language. It has a 'bank like' ability to create money and value in the form of economic and linguistic debt, and it auctions and markets language it does not own. Google gives language economic derivative value through AdWords and linked systems because it bases future transactions not on underlying value, but on commercial, derivative value. So language becomes illiquid,

untransferable, and less resilient. What is important to Google is how lucrative the words are, it has no concern with (and is indifferent to) the other values (we might say the values of différance) they might have. Google only makes money if the most commercially viable version of a word is returned in the search results - whether that is as a paid advert, an optimised text carrying adverts, or fake news spreading impressions and clicks. The liquidity of language is not what makes Google money, and therefore it is in their interests to keep ordinary language *illiquid*, defined primarily by its commercial value. I argue that this is *subprime language*, and Google are in effect betting against the liquidity of ordinary words. To take this to its logical conclusion, it could then be said that Google is in effect *shorting* the linguistic marketplace, recreating the precarious economic and linguistic (Appadurai 2015) conditions which have historically predicated global financial crises. While imagining the bursting of the linguistic bubble in this way might indeed be speculative, it is a technique that carries on a tried and tested tradition of using creative fiction as a means of contemporary and future (geo)political and economic critique (Kitchin and Kneale 2001; Davies 2018; Kitchin et al. 2019). In order to elucidate the concepts described above, I have also written a short piece of speculative fiction set in the future and imagining the build up to and aftermath of the crash of the global linguistic economy (Thornton 2019, and Figure 8.12). The piece is also reproduced as a postscript to the thesis.

8.5 Conclusion

This chapter has approached the idea of resistance from two angles: examining and experimenting with ways to actively resist the forces of linguistic capitalism, but also acknowledging how such resistance can be problematised - and itself resisted - by technological and disciplinary factors. In the spirit of my post-digital auto-ethnographic approach, however, there have been several important insights to be gained from closely documenting the critique of such digital systems, while also working with and through the resistant technologies themselves. Insisting on the power of creative and speculative fiction as an incisive tool to critique both the present and the future (Hayles 2008, Kitchin and Kneale 2001), the chapter ended with a piece of speculative theory and fiction, a method that will be discussed further in the concluding chapter.

Subprime Language and the Crash

Google's thirst for keywords caused the 2041 Global Linguistic Crisis, says government report. Bust tech giant ignored warnings its AdWords empire was a threat to language and the economy.

WIRELESS 1st May, 2044

The government has today released the results of the official enquiry into the causes of the Global Linguistic Crisis (GLC) which brought down the digital economy in late 2041 and threatened the stability of human communication.

The report was written by researchers from the University of London in collaboration with officials from the banking sector and the Royal Society for the Preservation of Digital Media. As well as expert sources, the report draws heavily on the archive of digital-era paper print-outs found in a bunker beneath a garage in Mountain View, Ca. shortly after the collapse of now defunct internet provider Google.

Described as a 'shrine to the printed word', and stored in defiance of the 2020 International Paperless Society Act (IPSA), the documents included printed copies of internal memos, 'blog' posts, and the hand-written diary of an unidentified Google employee who appears to have predicted the linguistic crash as far back as 2025.

Mountain View woman

'All we know is that she was a woman', the report states, 'who very early on raised concerns about the consequences of Google's project to link their digital advertising platform AdWords, with real estate investment and their global takeover of internet service provision and data storage'. These concerns appear to have fallen on deaf ears.

'This is linguistic capitalism gone mad', wrote the woman in a diary entry from 2025, 'It's not enough that we wring every last penny out of words by auctioning them every time we put them through the internet. Now we have to agree for everything we SAY to be monetised!'

Researchers say the diary entry refers to Google's move from serving adverts as search results based on the auctioning of keywords, to harnessing and exploiting the language circulating in physical spaces. This was a switch in tactics made possible at first by Google's growing dominance as gatekeeper to the internet and latterly as the landlord of vast swathes of land and property.

Crisis

The 2041 GLC prompted the collapse of the modern digital economy, put an end to internet connected communication, and led to the mass destruction of every piece of information held on Google's custom-built server island in the North Atlantic. Its effects are still being felt today, and like the last Global Financial Crisis in 2008, its roots can be traced to the property market.

At the height of the crash, Google controlled access to every Wifi network in the world, owned 95% of all real estate in the UK, Europe, and North America, and was responsible for the digital-urbanisation of much of Africa and the global south. It was a property and data empire financed purely by the monetisation of words.

According to the researchers, sometime in 2020 Google had what they call a 'material turn'. Bosses at the tech giant began to realise that their monopoly of digital space could seriously limit further expansion of profits in the future. They needed to start exploiting physical space too. Google's successful monetisation of digital space had begun to fund a mass property purchase and construction scheme.

Pilot schemes such as in Canadian cities in the late 2010s had been so successful that Google was fast becoming the dominant landlord of physical sites as well as web sites, networks and web space. 'What if we build real sites as well as web sites?', reads one excited internal electronic message found in the bunker, 'We could advertise on buildings, OMG we could make buildings out of adverts!'

And that is ultimately what Google did. They built cities out of electronic adverts based on their old highly successful web-based system of AdWords. These cities were constructed of keywords, built into the material fabric of the architecture, but also into the virtual fabric of the infosphere via Wifi permissions and the growing trend for web-based 'personal assistants', which, after the demise of competitors such as Amazon Digital 2.0 and Faceswipe, became the ubiquitous eyes and ears at the frontier of Google's expansion.

Linguistic bubble
'They say it's saving the rain forests', reads another diary entry, apparently in reference to the IPSA of 2020, 'but that's just a cover. What they're really doing is making us into walking, talking adverts. They're creating a linguistic bubble'.

Etienne Smith, from the University of London's department for Critical Analogue Humanities, was one of the co-authors of the report. He told WIRELESS: 'It's sometimes hard to believe, but there came point in the 2020s, when in some predominantly urban environments, it became physically impossible to communicate, in writing and face to face, without every word being monetised by Google'.

'People know that if they use certain words, they get more data and cheaper WiFi bills, and this changes in different areas, so if you talk about how wonderful Google is here in the Bay area, you end up with loads of money. And if you talk about rival products in a building sponsored by a particular advertiser, you get less data at a higher price.'

Google's use of speech for advertising began in the data-rich catchment areas of central business and commercial districts of major global cities, where skyscrapers, complexes, parks and roads were constructed around the advertising space they could display digitally and dynamically. But what the report calls the 'AdWords effect' quickly spread to other areas, infecting everyday speech in local neighbourhoods and in people's homes.

'The value of language changed', says Smith, a specialist in critical forensic banking and the linguistic economy. 'It became unsustainable. Nobody could trust anything anybody else said'. 'Tranches of language developed in different areas, and the poor became poorer as their language became worthless'.

Urban Collapse
As with the GFC of 2008, it was in these poorer communities that the worst effects of the trouble began to show. 'In the early 2000s it was low income Americans being sold property they couldn't afford that started the crisis', says Smith, 'but by the late 2030s the cloud-based internet schemes launched by the early tech giants in the 2010s had facilitated the construction of thousands of towns and cities across the global South, all of them built on the apparent stability of the linguistic economy'.

The government report makes for sobering reading. While the GLC caused widespread economic and social hardship in the US and Europe, in the newly urbanised areas of East and Central Africa alone it is estimated that up to a million people lost their lives in the civil wars and famines that followed the crash.

According to Smith, the Google AdWords effect had already begun to polarise these new communities by decimating indigenous languages in favour of English, creating hierarchies based on linguistic skill, and also physical access. 'Those with a better command of English basically began to command physical space as well. The less educated and poorer occupants of these new urban spaces were denied access to the richest linguistic areas, so they couldn't earn anything from speaking there'.

Even the 2038 AdWords Riots in New Sahara didn't make Google stop what they were doing, says Smith. 'The tech companies got greedy. They didn't care that these new developments were turning into deeply segregated areas. They were making billions from these new markets. In my opinion, yes, they did have blood on their hands.'

Linguistic Liquidity
So, what became of Mountain View woman and her archive? Did her bubble burst? 'Yes, it did', says Smith. 'We lost the ability to communicate. Language in effect became subprime, and once the advertising industry imploded, the digital economy collapsed like a house of cards.'

Smith's favourite part of the archive is a diary entry from August 2033 which simply reads 'Words are worth more than money'. 'I think she was right', he says. 'In financial terms we would say that language had become so tied to an economic value, rather than, say, a poetic one, that words had in effect become illiquid. Their only meaning – or value – was what they were worth in an advert, and when advertising became part of the infrastructure, this had horrific consequences'.

'It's possible she's still alive and reading this article, but as most former Google employees went to ground after the crash, it's unlikely we'll ever know who she was'.

Also found in the bunker was a collection of poetry, an English translation of a short story by French author Alain Damasio, *Les Hauts Parleurs*, and a heavily annotated paper copy of George Orwell's *Nineteen Eighty-Four*, one of only a handful of pre- crash copies known to be in existence.

Do you know Mountain View woman? Call WIRELESS with any information.

Figure 8.12: Subprime Language and the Crash (Thornton 2019)

Chapter 9

CONCLUSION

9.1 Contributions

This thesis has made unique contributions across a number of academic and artistic fields, recognising and researching the new modes of circulation of value and capital in an age of linguistic capitalism, and creating new disciplinary, theoretical, creative and methodological contexts in which to engage, critique, and resist them. I have developed the idea of 'geographies of (con)text' (Chapter 4) as a means of theorising and explaining the geo-linguistic spaces of the web, and the different actors and motives that construct it. I have also put forward a conceptual theory based around the idea of 'subprime language' (Chapter 8) which imagines the potential dystopian consequences of linguistic capitalism. I have built on the growing amount of literature and critical debate around the social, economic and political workings and effects of data and algorithms (Thatcher et. al 2016; Dalton & Thatcher 2016; Noble 2018; Shaw & Graham 2017; Crawford 2013), but have foregrounded the importance of adding language to these debates; the commodification of words being a crucial underlying driver of the digital economy, and in particular of technology companies such as Google (Kaplan 2014; Berardi 2012). By highlighting the importance of language to the digital economy, while at the same time insisting on the existential importance of language in terms of its potential power over people and places, and as tool of communication, creative expression and resistance, my thesis has identified fundamental tensions in terms of the differing 'values' of language in a digital age.

The economic value of decontextualised and data-ised words might be the life-blood of companies such as Google, but my thesis has shown the significant social and political consequences this might have. Further to this, with my innovative quantitative and quali-

tative methods in researching Google's AdWords by harvesting the price data of keywords, I have made a important and empirically incisive step in the study of the political economy of digital technology, which has had a significant impact across several disciplinary fields, including digital geography, digital sociology, web science, cyber security, digital humanities, literary and media studies and critical data/algorithm studies. This method has also offered a means of exposing and resisting the power of linguistic capitalism through artistic intervention. As well as contributing to the academic conversation around digital technology and algorithmic power and governance, my {poem}.py project has been successful as an artwork that has the ability to translate the sometimes obscured and obfuscated politics of technology to a wide range of audiences. As my project progressed (somewhat unexpectedly) from theory to practice, I have also been able to develop a new, slightly experimental and definitely collaborative process of research that I have called a post-digital auto-ethnography. The concepts of value, context, and circulation have been central to the thesis, flowing through every chapter, and informing the theoretical debates of Part One, before finally coming together in the practical interventions detailed in Part Two. With these concepts very much in mind, I want to conclude by returning to and addressing the research questions set out in Chapter 1.

9.1.1 Language in the age of linguistic capitalism

Research Question 1: How is language affected by digitisation, data-isation, monetisation and algorithmic processing (broadly speaking the system of linguistic capitalism), and what social and political effects/consequences does that have?

This thesis has argued that the digitisation, data-isation and monetisation of the words that flow through the search engine has significant linguistic, social, economic and political effects. In Chapter 4, I put forward and developed the concept of 'geographies of (con)text' in order to show how language represented as data, and the way it is algorithmically processed and circulated in various ways through the digital economy, becomes decontextualised, or deconstructed, as narrative text. Words-as-data in this way become tools for the flow and accumulation of capital through digital space, assuming new contexts dependent on their value as commodities to a range of actors, although perhaps most lucratively, to large technology companies such as Google.

Although this digital decontextualisation might indicate a post-structuralist and even post-modern theoretical reading of language in a digital age, I have instead argued that as a binding grammar, the data-isation and monetisation of words is in effect imposing new types of spatial and structural logic on language, be this through search results and digital advertising, autocompletions, online translation tools, or how language online is manipulated, and indeed policed, for profit by various actors for a variety of motives. These sometimes conflicting motives lead to a wide array of actors competing for exposure and for money in the contextual spaces of the web, for example advertisers/SEO experts, copyright infringers and the big tech companies, their users and misusers, all manipulating language in some way as a means to create capital. These linguistic exploitations are exacerbated by the sheer volume of text now uploaded to the internet, much of it also in pursuit of economic gain, as I discuss with reference to Walter Benjamin in Chapter 5. The proliferation of these algorithmically and economically optimised corpora of words contributes to the pools of data available for search algorithms to work on, meaning that certain configurations and densities of linguistic data begin to magnify certain messages and stories within what I have called the 'searchable database' (Thornton 2017), which means they are then more likely to be reproduced through the search engine.

I have argued that this tainting of online discourse with economic and algorithmic logics leads to the collateral (re)construction of unintended, and often damaging new (and old) narratives, such as stereotypical or extremist search results and autocompletions, politically infused advertising, or viral fake news stories. The way words are (mis)used and exploited as capital generating pieces of data, and they way they are processed, ordered and extracted in digital space, can therefore be theorised as a geography of (con)text, within which language is on the one hand constrained by its economic coupling and algorithmic reproduction, but on the other hand, due to its its (perhaps negated) primary use as a means of human narrative communication, also has the capacity to tell dangerous stories on a global (geo)political, cultural and social scale.

Having established some of the substantial effects that language in the age of algorithmic reproduction can have on online discourse and society, the thesis also examines the political undercurrents and structures of power and control that provide the conditions in which linguistic capitalism flourishes largely unchecked. Using Walter Benjamin's concept of the aestheticisation of politics through the cooption of technologically advanced culture in the

volatile and authoritarian era of 1930s Germany, I have argued that there is a politics lurking behind Google's search and advertising platforms that is similarly obscured - or aestheticised - by the unprecedented reach and influence enjoyed by Google today. When examined in this way, Google's dominance in the technology market, its near blanket role as mediator of culture and commerce, its political influence, and near ubiquity in everyday life, private and public discourse, becomes extremely problematic, and itself verges on authoritarian, or even 'micro-fascist' (Gilge 2016) forms of power.

Against this backdrop of Google's power and influence, the thesis is concerned primarily with two specific aspects: the influence of algorithmic (as opposed to mechanical) reproduction of art and culture, and the specific importance of the monetisation of language, which, through search and advertising, provides Google with its main stream of income. As I argue throughout the thesis, control over language is historically a tool of power over people and places, and this is a power that can only become stronger when the mediation and dissemination of language is structured and magnified by algorithmic processes, as well as being a fundamental part of the digital economy and Google's lifeblood. Just as in Benjamin's day, the technologically facilitated evolution of language is never devoid of politics and inequalities, although as I demonstrate in Chapter 6, Google goes to great lengths to foster the idea that the linguistic market it has created with its AdWords platform is an unmediated, 'organic' and 'free' marketplace. Far from it, I argue that linguistic capitalism does not, and cannot, operate as a quasi-Hayekian free marketplace because of its algorithmic method, and because of the array of opaque distortions and manipulations that construct the political economy of such a marketplace. Rather than market governance of language, linguistic capitalism, and the power it wields, is in reality a heavily mediated centrally-planned form of algorithmic governance with substantial political consequences for the control of our lives and actions (Danaher 2016; Rouvroy 2013; Yeung 2018).

Building on the Benjaminian critique of the aestheticisation of politics, the potentially authoritarian nature and effects of this algorithmic governance is hidden behind free market rhetoric as a means of justification and abrogation of responsibility. But it would be too easy to conclude that linguistic capitalism sits somewhere between a political economy binary of either free market or interventionist governance, as some recent critiques have done (Bratton 2016; Ezrachi &Stucke 2016), as if categorisation is somehow the answer

to temper and tame the wider effects of this new form of governance. Indeed, if this is a new form of hybrid algorithmic governance, then it needs new forms of transparency and scrutiny, and indeed, new forms of intervention and resistance. As I will go on to discuss in relation to the second two research questions, power and control over language are - for good reason - often subjects of dystopian theory and fiction, and the fact that both economic and cultural/social/political power stem from Google's control of language is at the heart of the thesis. Language is power, but is also a *powerful* method of critique and resistance, as I have shown in the second part of the thesis.

9.1.2 Raging against the machine

Research Question 2: What are the difficulties of studying/critiquing language mediated by digital technologies?

As discussed in Chapter 3, in depth and empirical study of digital technologies can be challenging, if not impossible. Problems of accessibility to systems and private tech companies, secretive proprietary algorithms, lack of technical expertise and the unreproducibility and unreliability of dynamic and opaque data make critical research difficult across the disciplines (Kitchin 2017; Duggan 2017b). These are all methodological problems I encountered in trying to uncover and show the workings of linguistic capitalism and Google's search and advertising platforms. Google is famously protective of its algorithms precisely because they hold the key to its success and continued wealth. Without becoming a paid customer of Google (which I always wanted to avoid), or indeed paying for the services of a commercial SEO company, there is no API available to gather data from AdWords, which has meant a constant battle against tweaks, upgrades, internal system policing and censorship. Indeed, in the final few weeks of writing this thesis, I received an email from Google to say my account (the one I use for {poem}.py) has been cancelled 'due to no spend' (see Figure 9.1). Staying 'outside' the system by not becoming a paying customer is thus a significant problem, as indeed is remaining an 'outside' or objective observer of technologies such as the Google search engine when results are constantly personalised and localised. Indeed, these difficulties and subjectivities are what drove my post-digital auto-ethnographic approach. Instead of insisting that opening the black box is the only way to reveal the workings of linguistic capitalism (Pasquale 2015; Amoore 2018), I instead worked with the grain (Rogers 2013) of the often disobedient, and always vibrant

(Bennett 2009) technology.

More political than methodological, one of the other big difficulties in studying and critiquing language mediated by digital technologies, and especially Google, is its near universal embeddedness in our everyday lives. This not only makes is hard to research from 'within', but it also means that any potential problems within the technology, such as privacy issues, biased search results etc, are often seen as acceptable consequences when Google's platforms have become indispensable in how we function. How we communicate with each other, access information, find our way around, recreate, and buy essential commodities, is all now in some way mediated through various Google platforms, and it has become almost impossible to live and work without them. As I argued in Chapters 5 and 6, as the seemingly benign provider of all these facilities, Google holds an unprecedented power in society, which must be recognised as a system of governance in order for its power to be rightly checked and challenged (Shaw & Graham 2017). Yet the power Google holds, and the politics that lurk beneath its highly lucrative platforms, are obscured by its normalisation into everyday live, its utility, reach and an aesthetic that gives a false feeling of agency to its users. But further to that, and as I argue with reference to authoritarian and dystopian examples in the thesis (Kitchin et al. 2019; Davies 2018), when this opaque and obfuscated power is propped up by the exploitation of language - our very means of expression, communication and redress, the power becomes far more insidious, and the need for recognition and resistance becomes ever more urgent.

Notwithstanding all the blips, hurdles and brick walls I encountered in the course of this thesis, there have been times when these difficulties have themselves been important lenses of discovery and critique. By using the development of {poem}.py as a kind of practice based auto-ethnographic method, I have confirmed and consolidated some of the difficulties in studying algorithmic/digital processes described in Chapters 3 and 8, but I have also identified that some of those difficulties, especially the issues involving cross-disciplinary research, have also been positively instrumental in the development of the project and therefore added to the insights and observations I have been able to draw from my thesis.

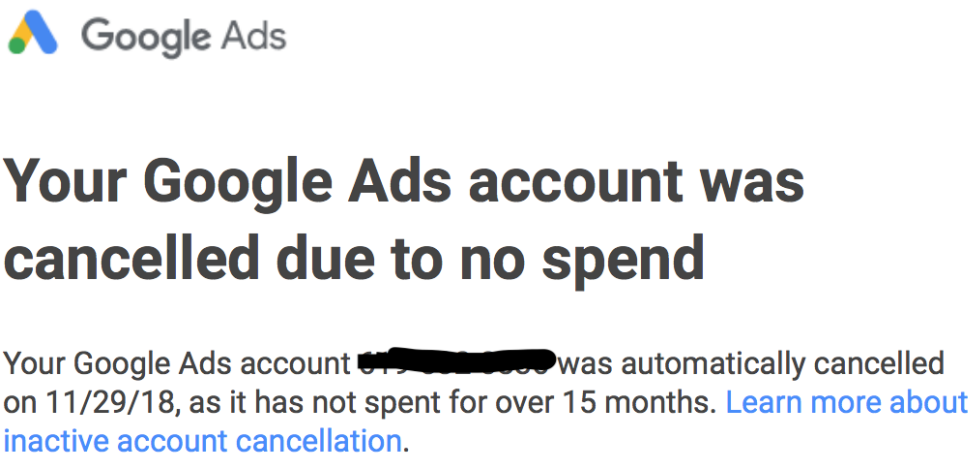


Figure 9.1: Email received from Google, 30 November 2018. Screenshot: author's own.

9.1.3 Exposition / intervention

Research Question 3: What can be done to expose, mitigate, or resist the effects of linguistic capitalism? Is it possible to resist/intervene?

This thesis has acknowledged the methodological, political and disciplinary difficulties in researching linguistic capitalism and other digital technologies, and has accepted that gaining access to the (perhaps mythical) black box is not necessarily either possible or useful as a method of inquiry. In terms of mitigating the power of companies such as Google, there are today growing calls for the company to be either broken up as a monopoly, or regulated as a pubic utility, but these are concerns and approaches beyond the scope of this project. My own approach has been one of mitigating the power of Google by making visible its workings and effects to as many audiences as possible. I have done this through the development and exposition of my {poem}.py project, an artistic intervention that has been successful in translating the opaque and often technical debates around the problems of digital technology into an accessible and engaging format. My project was inspired by existing creative interventions into linguistic capitalism, and in particular Christophe Bruno's work on AdWords (2002, 2012) and Cabell and Huff's American Psycho (2013) project. More than any other attempt to critique the often opaque and powerful structures of language in the age of algorithmic reproduction, for me, these types of creative projects succeed as incisive and important interventions, not only because they serve as effective means of translation between often polarised and partite disciplines and backgrounds, but because they critique from within, harnessing the power of language and literature to expose its vulnerability as commodity in the global digital economy.

As I argued in Chapter 7, in the current discourse and debate around AI, machine learning and computational language techniques quite often takes digital technology as an assumed starting point, rather than the (often humanistic) object that technology is meant to be fixing, analysing or controlling. Drawing on what I consider to the successes of projects such as Bruno, Cabell and Huff's, my own project has gone against much of the current research on digital technology, in that it has also harnessed the power of words to analyse and critique digital technology, rather than just using digital technology to analyse language and literature, which seems to be more common in digital humanities and data science these days (Liu 2012; Berry 2017). My approach of using existing poetry, and all the memories and emotions it contains, as a vehicle of critique, is an attempt to reclaim language from digital technology and the algorithmic market, restoring in it its inherent capacity for political resistance.

Another reason I think my {poem}.py project has been successful and popular across a diverse range of audiences, is that it critiques digital technology with a primarily analogue intervention. The poem-receipts generated by {poem}.py are tactile, vibrant, and unstable. They are familiar, but made strange by their recomposition, and as such have been extremely effective in helping people to understand and engage with the deeper debates. In Chapter 8, I argued that using digital technologies to critique digital technology runs the risk that any output will be recuperated into systems of digital capitalism. In this respect I think there lies a fundamental problem in how academics and artists are encouraged to engage in critical interventions, which is increasingly through digital formats. I refer here to the many digital art exhibitions that perhaps run the risk of reproducing systemic problems, as well as trying to critique them, for example The Victoria & Albert's (2018) Digital Futures program with its emphasis on 'electronic visualisation', and in particular to a recent $225,000 funding competition for critical art and advocacy projects launched by Mozilla which aimed 'support people and projects that examine the effects of AI on society'. As the competition site explains: '[i]n a world of biased algorithms and broken recommendation engines, art and ideas that educate internet users about AI are more important than ever' (The Mozilla Blog 2018). While these sound like admirable concerns, only proposals that are 'native to the internet' and digital in output are considered worthy to take up these important tasks. Analogue critiques of the affects of AI on society (including my own project) are necessarily excluded from the process, as only digital ones are

submittable, despite their medium effectively reenrolling them into the marketised digital and algorithmic systems the funding purports to be challenging. This again goes back a central argument of the thesis, that of the aestheticisation of the political effects of digital technology by big tech companies such as Google, and indeed Mozilla in this case. Google also has its own funding and exposure programs for art, which also foreground digital techniques over analogue. Following Benjamin, the antidote to the aestheticisation of politics is through the politicisation of art, but when the means of production and dissemination of art is controlled by the very institutions which necessitate redress and resistance, then it is surely to analogue forms of resistance we must turn, which is what I have tried to do with {poem}.py. The answer to the second two research questions might therefore be that yes, despite the inherent difficulties, it is possible to resist and/or intervene to expose and mitigate the effects of linguistic capitalism, but the most incisive ways of doing so must be aware - and wary - of having any message recuperated by the systems it tries to critique. One way to do this might be to insist on analogue outputs, or, as I have done with {poem}.py, to self-reflexively and critically (even tactically) examine these structures from within, taking care not to contribute to, or rely upon their exploitative platforms wherever possible.

9.1.4 Context / Value / Circulation

As I set out from the start, there were three conceptual threads running through this thesis: those of context, value and circulation. These delightfully ambiguous concepts have allowed me to question both implicitly and explicitly how words (as datafied commodities) move through the structures and processes of the web, and what happens to them on the way. Chapter 4 dealt explicitly with (con)text, but the decontextualisation of language in an age of algorithmic reproduction also informed the Benjaminian critique in Chapter 5, the empirical analysis of AdWords data in Chapter 6, and indeed forms the basis for the artistic intervention and resistance in the final chapters. Context here is both literal (or literary) and physical, in terms of how linguistic data is stored, ordered and moved through digital spaces. While on the one hand, digital technologies seem to liberate language and fuel creativity, I have shown how context can also been restrictive. Circulation too, can be seen physically and figuratively - linguistic data is exploited as it circulates as commodity, but there is also a much wider political economy of circulating values going on under the systems of digital capitalism, as I have shown especially in Chapters 6 and

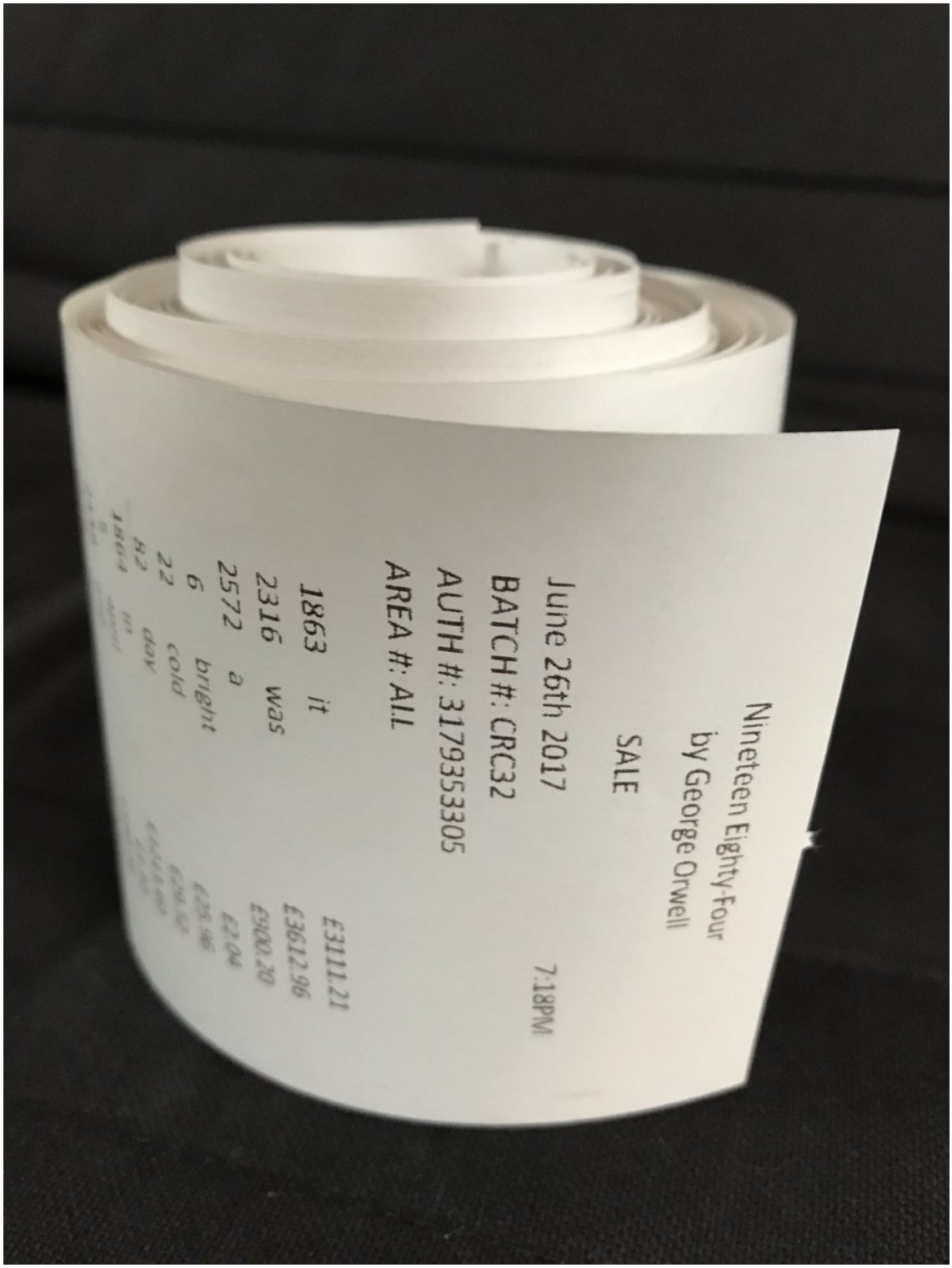


Figure 9.2: Orwell's 1984 (beginning) ({poem}.py 2017).

8 (Langley & Leyshon 2017; Srnicek 2016). Lastly, the thesis has highlighted the tension between the economic and the 'narrative' - and even artistic - values of language in an age of linguistic capitalism.

9.2 Future work

My PhD research has taken me into such unexpectedly exciting and productive areas that I really dont want to stop. While the research I put forward in this thesis has been based primarily on exploring and making visible the impact of the digitisation/data-isation and monetisation of language (and specifically English text), there is great scope to expand these methods, critiques and interventions into projects exploring digital images, especially Google images, for example. While there has been recent research on the racial and gendered bias within text and image based search technology (Noble 2018), I think a very productive line of future inquiry is the exploration of the relation between text based search engines and the corresponding images returned with the same search terms. I can see the results of such an inquiry having important implications politically and socially, and also presenting another exciting chance at creative visualisation. This kind of research could perhaps be conducted in association with a gallery or museum, or other digitalised databases of artwork and images. Another potential further research strand would be to extend my critique of 'linguistic capitalism' to other languages. At the moment I use primarily English texts for my {poem}.py project (although there have been exceptions when people have asked my process their favourite poems in their own languages), but I can see an extremely interesting project in adding non-English languages, and also non-Roman based script to this critique, and measuring the data against geo-located news stories and events. This is a project with great potential for collaborations between disciplines such as Modern Languages and Geography, as well as data science and visualisation. There is also much potential future work to be done comparing the way linguistic capitalism affects other search engines, in particular Duck Duck Go, which markets itself on its privacy credentials. This kind of comparative study could perhaps lend itself to a more activist and/or policy based approach in trying to find solutions and alternatives to challenge Google's monopoly in search industry and beyond.

I have several specific plans for future work, including a book length project which develops the linguistic and political economy aspects of subprime language. There is also a much

This be the Verse
by Philip Larkin

SALE

21st June 2016 11:30 am
BATCH #: CRC32
AUTH #: 2934922873
AREA #: ALL

6	they	£2.94
1	fuck	£0.11
4	you	£1.56
2	up	£0.32
1	your	£0.08
1	mum	£0.16
5	and	£1.85
1	dad	£0.71
1	may	£0.20
1	not	£0.07
1	mean	£0.78
2	to	£1.38
2	but	£0.18
1	do	£0.44
1	fill	£0.67
1	with	£0.39
2	the	£0.86
1	faults	£0.00
1	had	£0.53

Figure 9.3: Larkin's 'This be the Verse' ({poem}.py 2016).

needed focus to be brought to how linguistic capitalism and the commodification of language in general relates to debates around security and surveillance. In this respect, areas of future work might develop the links between linguistic capitalism and surveillance capitalism (Zuboff 2016), the ontological (in)securities of search engine results (Noble 2018) and the idea of cyberwar as a process of deconstruction (Jocque 2018).

A piece of future work I am particularly excited to begin is to develop the concept of a *digital écriture féminine*. In the course of researching the thesis I became interested in a feminist post-structuralist analysis of word values as suggested by Google. I began noticing that 'male' words such as *boyfriend*, *dad*, or *husband* seemed to be worth more (in monetary value) than 'female' ones such as *girlfriend, mum*, or *wife*, and that *white* was the most expensive colour (see Figure 9.3). These embedded binary value judgements reminded me of Helene Cixous's 1976 essay 'The Laugh of the Medusa', in which she identified in the hierarchical phallogocentricism of language a certain kind of 'automatism' which she believed 'conceals an invincible adversary, because it's the language of men and their grammar'. My proposed research area adds a deeper layer to this critique by combining the pre-existing structural inequalities in language with their algorithmically and economically compounded gendered and racial biases and nuances, and the role they play in the construction and conversation around digital technology today. A *digital écriture féminine* could therefore become a tool of analysis and critique into how digital technology

is constructed and marketed, and could also include research into feminist coding techniques and alternative technologies.

I would also like the project to follow Cixous in her call to take back some of the power that lurks within the 'libidinal and cultural - hence political, typically masculine-economy' of language that digital technologies tend to serve to reify, magnify and justify, by actively turning the insights of the data analysis into new forms of politicised texts, creative writing, artistic statements and interventions, thus breaking the 'automatisms' which structure language, and instead exploring the creative potential of a *digital écriture féminine*.

As well as the *digital écriture féminine* project, I would also like to develop {poem}.py into a dynamic digital art installation that projects the poem-receipts onto an outside or interior wall of a building or public space. The receipts would update according to the fluctuating suggested bid prices on Google AdWords. As the bid prices can change every day, regular commuters or tourists would be able to see the poems changing in value in real-time. I would need help to work out how to make this type of installation happen both technically and logistically, and would also need to spend time choosing texts that might be more volatile or mobile within the market due to their currency in the current media environment.

9.3 Last words

This thesis has set out to be both provocation and intervention, while also documenting and offering means of resistance into the processes and the real and potential consequences of linguistic capitalism as manifested through Google's search and advertising platforms. Provocative in that its critical analysis is conducted through the lens of fictional dystopia, yet the power and resilience of creative language is also at the heart of the intervention. The provocation in that sense becomes part of the intervention; art imitates life, and is turned back to art again. As Feuz et al. say, it may be almost impossible to 'study a distributed machinery that is both willfully opaque and highly dynamic one, which reacts to being studied and takes active steps to prevent such studies being conducted on the automated, large-scale level required' (2011). Google does all these things, so what I have done with {poem}.py is to accept these barriers as important pieces of the jigsaw. I do not have access to the black-boxed workings of Google's algorithmic platforms, but I can

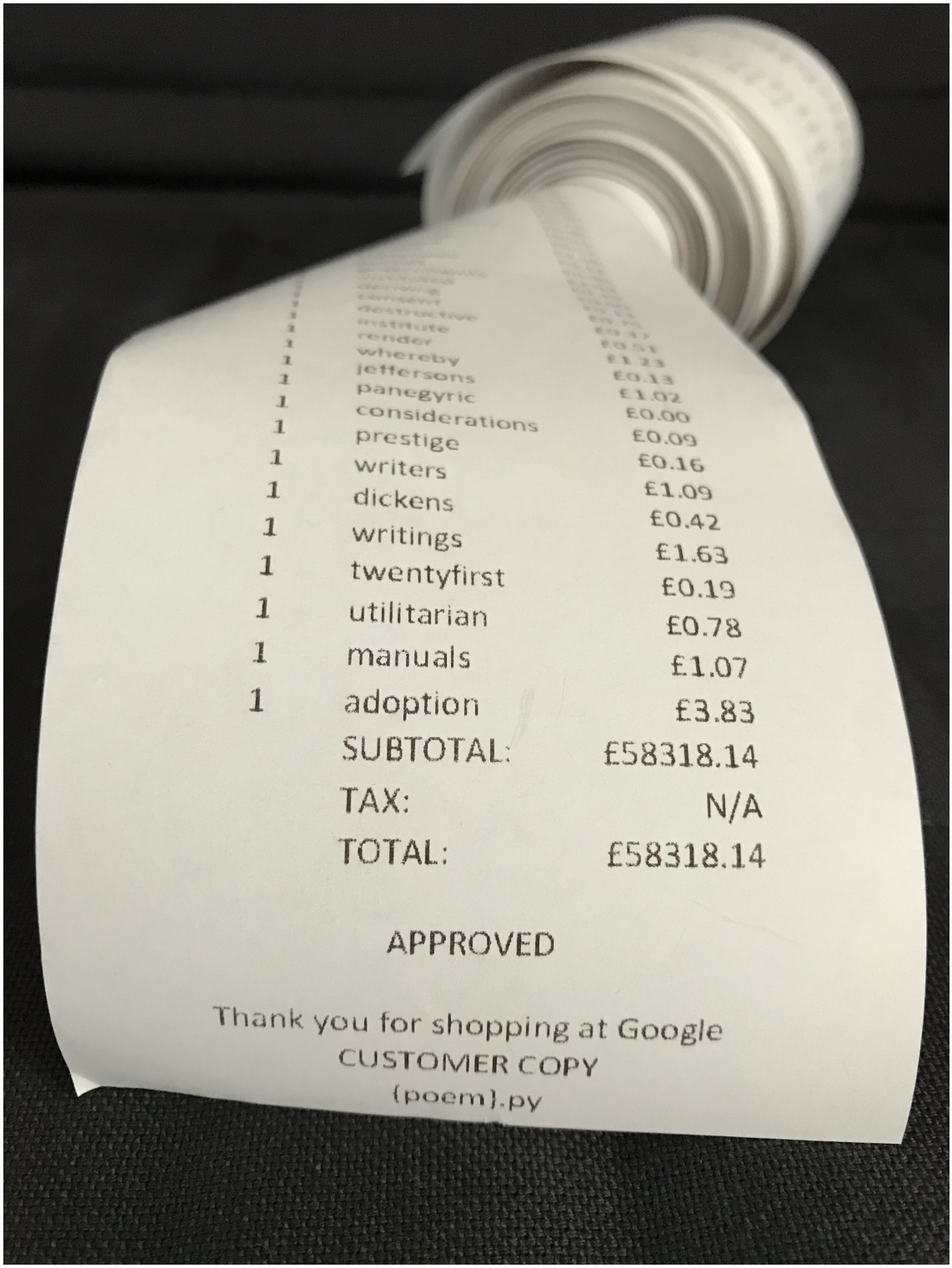


Figure 9.4: Orwell's 1984 (end) ({poem}.py 2017).

see (and show) what goes into the search portal, and what comes out the other side. My receipt printer in this way takes on the role of the black box, exposing the contextual transitions of language, while at the same time embracing the performative logic of linguistic capitalism via the material aesthetics of print on paper. This has meant working with (rather than against) the idiosyncrasies of the system to expose its underlying logics; discovering not only what they can tell us about the world, but also what part they play in constructing the narrative, and infusing cultural and political discourse with the neoliberal logic of the linguistic market in a way unimaginable before the modern digital age. It is of course an unprovable curiosity that - in theory - Orwell's 'Nineteen Eighty-Four' could earn Google £58,318.14 each time it passes through one of its algorithmic portals, but it is a point I think needs making (see Figures 9.2 & 9.4). In processing Orwell's text through poem.py, I am of course executing a double-flanking critical manoeuvre against the political and linguistic consequences of linguistic capitalism; highlighting the power of language as exemplified by Newspeak (both as described by Orwell and as deployed in a contemporary context), while also exposing the similarly 'Orwellian' authoritarian power that Google has today by virtue of its control over the commodification and exploitation of language. Although all kinds of data are monetised in the modern digital economy, exposing the potential consequences of the monetisation and manipulation of linguistic data is especially important not only because of the fundamental role language plays in our everyday lives as a means of communication and expression, but also in the politics and power relations it embodies. As I mentioned earlier, and as imagined by Orwell and in my own work of speculative fiction, language is power, and those who control it - by whatever means - hold an extraordinary amount of influence over both people and places. In an age of linguistic capitalism, our very means of communication has been compromised by the all-pervasive, yet opaque, logics of the algorithmic neoliberal market. The theories, provocations and interventions described in the thesis are therefore my attempt to shine a light on Google's algorithmic and economic distortions of language; critiquing, resisting (and subverting) linguistic capitalism by harnessing and mobilising the power of the very tools on which its empire is built: the power of words.

POSTSCRIPT

WIRELESS Magazine

May 1, 2044

Subprime Language and the Crash

Google's thirst for keywords caused the 2041 Global Linguistic Crisis, says government report. Bust tech giant ignored warnings its AdWords empire was a threat to language and the economy.

The government has today released the results of the official enquiry into the causes of the Global Linguistic Crisis (GLC) which brought down the digital economy in late 2041 and threatened the stability of human communication.

The report was written by researchers from the University of London in collaboration with officials from the banking sector and the Royal Society for the Preservation of Digital Media. As well as expert sources, the report draws heavily on the archive of digital-era paper print-outs found in a bunker beneath a garage in Mountain View, Ca. shortly after the collapse of now defunct internet provider Google.

Described as a 'shrine to the printed word', and stored in defiance of the 2020 International Paperless Society Act (IPSA), the documents included printed copies of internal memos, 'blog' posts, and the hand-written diary of an unidentified Google employee who appears to have predicted the linguistic crash as far back as 2025.

Mountain View woman

'All we know is that she was a woman', the report states, 'who very early on raised concerns about the consequences of Google's project to link their digital advertising platform AdWords, with real estate investment and their global takeover of internet service provision and data storage'. These concerns appear to have fallen on deaf ears. 'This is linguistic capitalism gone mad', wrote the woman in a diary entry from 2025, 'It's not enough that we wring every last penny out of words by auctioning them every time we put them through the internet. Now we have to agree for everything we SAY to be monetised!'

Researchers say the diary entry refers to Google's move from serving adverts as search results based on the auctioning of keywords, to harnessing and exploiting the language circulating in physical spaces. This was a switch in tactics made possible at first by Google's growing dominance as gatekeeper to the internet and latterly as the landlord of vast swathes of land and property.

Crisis

The 2041 GLC prompted the collapse of the modern digital economy, put an end to internet connected communication, and led to the mass destruction of every piece of information held on Google's custom-built server island in the North Atlantic. Its effects are still being felt today, and like the last Global Financial Crisis in 2008, its roots can be traced to the property market.

At the height of the crash, Google controlled access to every Wifi network in the world, owned 95% of all real estate in the UK, Europe, and North America, and was responsible for the digital-urbanisation of much of Africa and the global south. It was a property and data empire financed purely by the monetisation of words.

According to the researchers, sometime in 2020 Google had what they call a 'material turn'. Bosses at the tech giant began to realise that their monopoly of digital space could seriously limit further expansion of profits in the future. They needed to start exploiting physical space too. Google's successful monetisation of digital space had begun to fund a mass property purchase and construction scheme.

Pilot schemes such as in Canadian cities in the late 2010s had been so successful that Google was fast becoming the dominant landlord of physical sites as well as web sites, networks and web space. 'What if we build real sites as well as web sites?', reads one excited internal electronic message found in the bunker, 'We could advertise on buildings, OMG we could make buildings out of adverts!'

And that is ultimately what Google did. They built cities out of electronic adverts based on their old highly successful web-based system of AdWords. These cities were constructed of keywords, built into the material fabric of the architecture, but also into the virtual fabric of the infosphere via Wifi permissions and the growing trend for web-based 'per-

sonal assistants' which, after the demise of competitors such as Amazon Digital 2.0 and Faceswipe, became the ubiquitous eyes and ears at the frontier of Google's expansion.

Linguistic bubble

'They say it's saving the rain forests', reads another diary entry, apparently in reference to the IPSA of 2020, 'but thats just a cover. What they're really doing is making us into walking, talking adverts. They're creating a linguistic bubble'.

Etienne Smith, from the University of London's department for Critical Analogue Humanities, was one of the co-authors of the report. He told WIRELESS: 'It's sometimes hard to believe, but there came point in the 2020s, when in some predominantly urban environments, it became physically impossible to communicate, in writing and face to face, without every word being monetised by Google'.

'People know that if they use certain words, they get more data and cheaper Wifi bills, and this changes in different areas, so if you talk about how wonderful Google is here in the Bay area, you end up with loads of money. And if you talk about rival products in a building sponsored by a particular advertiser, you get less data at a higher price'.

Google's use of speech for advertising began in the data-rich catchment areas of central business and commercial districts of major global cities, where skyscrapers, complexes, parks and roads were constructed around the advertising space they could display digitally and dynamically. But what the report calls the 'AdWords effect' quickly spread to other areas, infecting everyday speech in local neighbourhoods and in people's homes.

'The value of language changed', says Smith, a specialist in critical forensic banking and the linguistic economy. 'It became unsustainable. Nobody could trust anything anybody else said'. 'Tranches of language developed in different areas, and the poor became poorer as their language became worthless'.

Urban Collapse

As with the GFC of 2008, it was in these poorer communities that the worst effects of the trouble began to show. 'In the early 2000s it was low income Americans being sold property they couldnt afford that started the crisis', says Smith, 'but by the late 2030s the

cloud-based internet schemes launched by the early tech giants in the 2010s had facilitated the construction of thousands of towns and cities across the global South, all of them built on the apparent stability of the linguistic economy'.

The government report makes for sobering reading. While the GLC caused widespread economic and social hardship in the US and Europe, in the newly urbanised areas of East and Central Africa alone it is estimated that up to a million people lost their lives in the civil wars and famines that followed the crash.

According to Smith, the Google AdWords effect had already begun to polarise these new communities by decimating indigenous languages in favour of English, creating hierarchies based on linguistic skill, and also physical access. 'Those with a better command of English basically began to command physical space as well. The less educated and poorer occupants of these new urban spaces were denied access to the richest linguistic areas so they couldnt earn anything from speaking there'. Even the 2038 AdWords Riots in New Sahara didn't make Google stop what they were doing, says Smith. 'The tech companies got greedy. They didn't care that these new developments were turning into deeply segregated areas. They were making billions from these new markets. In my opinion, yes, they did have blood on their hands.'

Linguistic Liquidity

So, what became of Mountain View woman and her archive? Did her bubble burst? 'Yes, it did', says Smith. 'We lost the ability to communicate. Language in effect became subprime, and once the advertising industry imploded, the digital economy collapsed like a house of cards.'

Smith's favourite part of the archive is a diary entry from August 2033 which simply reads 'Words are worth more than money'. 'I think she was right', he says. 'In financial terms we would say that language had become so tied to an economic value, rather than, say, a poetic one, that words had in effect become illiquid. Their only meaning or value was what they were worth in an advert, and when advertising became part of the infrastructure, this had horrific consequences'.

'It's possible she's still alive and reading this article, but as most former Google employees

went to ground after the crash, it's unlikely we'll ever know who she was'.

Also found in the bunker was a collection of poetry, an English translation of a short story by French author Alain Damasio, 'Les Hauts Parleurs', and a heavily annotated paper copy of George Orwells 'Nineteen Eighty-Four', one of only a handful of pre-crash copies known to be in existence.

Do you know Mountain View woman? Call WIRELESS with any information.

Bibliography

Adorno, T. W. (2005). *The culture industry: Selected essays on mass culture*. Routledge, London.

Aferiat, A. (2012). It's Google's world and we're just clicking in it: Why the growth of sponsored link advertising necessitates a shift of trademark regulation on the internet. *New Eng. L. Rev.*, 47:157.

Agarwal, A., Hosanagar, K., and Smith, M. D. (2011). Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of marketing research*, 48(6):1057–1073.

Amoore, L. (2006). Biometric borders: Governing mobilities in the war on terror. *Political geography*, 25(3):336–351.

Amoore, L. (2011). Data derivatives: On the emergence of a security risk calculus for our times. *Theory, Culture & Society*, 28(6):24–43.

Amoore, L. (2014). Security and the incalculable. *Security Dialogue*, 45(5):423–439.

Amoore, L. (2016). Cloud geographies: Computing, data, sovereignty. *Progress in Human Geography*, 42(1):4–24.

Amoore, L. (2017). What does it mean to govern with algorithms? *Antipode Online*. Available at: https://radicalantipode.files.wordpress.com/2017/05/2-louise-amoore.pdf.

Amoore, L. and de Goede, M. (2012). Introduction: data and the war by other means. *Journal of Cultural Economy*, 5(1):3–8.

Amoore, L. and Raley, R. (2017). Securing with algorithms: Knowledge, decision, sovereignty. *Security Dialogue*, 48(1):3–10.

Amoore, L. and Woznicki, K. (2018). The politics of artificial intelligence: an interview with Louise Amoore. *Open Democracy*. Avail-

able at: https://www.opendemocracy.net/digitaliberties/krystian-woznicki-louise-amoore/politics-of-artificial-intelligence-interview-with-l.

Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, 41(1):93–117.

Andersen, C. U. and Pold, S. B. (2011). *Interface criticism: Aesthetics beyond the buttons*. Aarhus University Press, Aarhus.

Andersen, C. U. and Pold, S. B. (2015). Aesthetics of the banal: New aesthetics in an era of diverted digital revolutions.

Andersen, C. U. and Pold, S. B. (2018). *The Metainterface: The Art of Platforms, Cities, and Clouds*. MIT Press, Cambridge.

Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books, London.

Anderson, C. (2008). The end of theory. *Wired magazine*, 16(7):16–07.

Aneesh, A. (2006). *Virtual migration: The programming of globalization*. Duke University Press, Durham.

Aneesh, A. (2009). Global labor: Algocratic modes of organization. *Sociological Theory*, 27(4):347–370.

Appadurai, A. (2015). *Banking on words: The failure of language in the age of derivative finance*. University of Chicago Press, Chicago.

Aradau, C. and Blanke, T. (2015). The (big) data-security assemblage: Knowledge and critique. *Big Data & Society*, 2(2).

Aradau, C. and Blanke, T. (2017). Politics of prediction: Security and the time/space of governmentality in the age of big data. *European Journal of Social Theory*, 20(3):373–391.

Ash, J., Kitchin, R., and Leszczynski, A. (2016). Digital turn, digital geographies? *Progress in Human Geography*, 42(1):25–43.

Attig, D. (2016). Wildly sexist lexicography from @OxfordWords beamed straight to students"iPads. Available at: https://twitter.com/bookmobility/status/690593050830204930.

Austin, J. L. (1975). *How to do things with words*. Oxford University Press, Oxford.

Baker, P. and Potts, A. (2013). Why do white people have thin lips? Google and the perpetuation of stereotypes via auto-complete search forms. *Critical Discourse Studies*, 10(2):187–204.

Barnet, B. (2003). The erasure of technology in cultural critique. *Fibreculture Journal*, (1):1449–1443.

Barnet, B. A. (2009). Idiomedia: The rise of personalized, aggregated content. *Continuum*, 23(1):93–99.

Barthes, R. (1977). *Image, music, text*. Fontana, London.

Bartscherer, T. and Coover, R. (2011). *Switching codes: Thinking through digital technology in the humanities and the arts*. University of Chicago Press, Chicago.

Battelle, J. (2011). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. Nicholas Brealey Publishing, Boston.

BBC News (2015). Ukraine's Poroshenko: 'New Russia' is like 'Mordor'. Available at: http://www.bbc.co.uk/news/world-europe-34037743.

BBC News (2016). The city getting rich from fake news, BBC News. Available at: https://www.bbc.co.uk/news/magazine-38168281.

BBC News (2017). Data is not the new oil. www.bbc.co.uk/news/entertainment-arts-41559076.

Beasley-Murray, T. (2012). On some seminal motifs in walter benjamin: Seed, sperm, modernity, and gender. *Modernism/modernity*, 19(4):775–791.

Becker, K. and Stalder, F. (2009). *Deep search: The politics of search beyond Google*. Studienverlag Gesmbh, Innsbruck.

Beer, D. (2009). Power through the algorithm? participatory web cultures and the technological unconscious. *New Media & Society*, 11(6):985–1002.

Beer, D. (2016). *Metric power*. Springer, New York.

Beller, J. (2018). The computational unconscious. *Boundary 2*. Available at: http://www.boundary2.org/2018/08/beller/.

Benjamin, W. (1999). *Illuminations*. Pimlico, London.

Benjamin, W. (2006). Experience and poverty. *Selected Writings (1927-1934)*, 2:731–736.

Benjamin, W. (2007). *Reflections: Essays, aphorisms, autobiographical writings*. Random House, London.

Benjamin, W., Jennings, M. W., and Doherty, B. (2008). *The Work of Art in the Age of Its Technological Reproducibility, and Other Writings on Media*. Harvard University Press, Cambridge.

Bennett, J. (2009). *Vibrant matter: A political ecology of things*. Duke University Press, Durham.

Benzon, P. and Sarah, S. The Aesthetics of Erasure. Guest Editorial. *NMC Media-N: Journal of the New Media*.

Berardi, F. (2012). *The uprising: On poetry and finance*. MIT Press, Cambridge.

Berends, M., Dossin, L., and Dovey, M. (2016). Addwords. *V2*. Available at: http://v2.nl/archive/works/addwords.

Berry, D. (2013). *Revisiting the Frankfurt School: Essays on culture, media and theory*. Ashgate Publishing Ltd., Farnham.

Berry, D. and Dieter, M. (2015). *Postdigital Aesthetics: Art, Computation and Design*. Palgrave Macmillan, London.

Berry, D. M. (2014). *Critical Theory and the Digital*. A&C Black, London.

Berry, D. M. and Fagerjord, A. (2017). *Digital humanities: knowledge and critique in a digital age*. John Wiley & Sons, New Jersey.

Bershidsky, L. (2014). Putin's Russia, Tolkien's Mordor: What's the Difference? *Bloomberg Opinion*. Available at: https://www.bloomberg.com/opinion/articles/2014-12-11/putins-russia-tolkiens-mordor-whats-the-difference.

Blas, Z. and Gaboury, J. (2016). Biometrics and opacity: A conversation. *Camera Obscura: Feminism, Culture, and Media Studies*, 31(2 (92)):155–165.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., and Kalai, A. (2016). Quantifying and reducing stereotypes in word embeddings. *arXiv preprint arXiv:1606.06121*.

Bonner-Thompson, C. (2017). The meat market?: production and regulation of masculinities on the Grindr grid in Newcastle-upon-Tyne, UK. *Gender, Place & Culture*, 24(11):1611–1625.

Bork, R. H. and Sidak, J. G. (2012). What does the chicago school teach about internet search and the antitrust treatment of google? *Journal of Competition Law and Economics*, 8(4):663–700.

Botnik.org (2018). Botnik. Available at: https://botnik.org/.

botpoet.com (n.d.). Bot or not. Available at: http://botpoet.com/.

Bourdieu, P. (1991). *Language and symbolic power*. Harvard University Press, Cambridge.

Bowker, G. C. and Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT Press, Cambridge.

Boyd, D. and Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5):662–679.

Braidotti, R. (2013). Posthuman humanities. *European Educational Research Journal*, 12(1):1–19.

Bratton, B. H. (2016). *The stack: On software and sovereignty*. MIT Press, Cambridge.

Bridle, J. (2011). The new aesthetic. *jamesbridle.com*. Available at https://jamesbridle.com/works/the-new-aesthetic.

Bridle, J. (2014). On the rainbow plane. *booktwo*. Available at: http://booktwo.org/notebook/rainbow-plane-002-kiev/.

Bridle, J. (2015). Superscript Keynote, 29 May 2015. *Walker Art Centre*. Available at: https://www.youtube.com/watch?v=2GdHrMGIL_A.

Bridle, J. (2017). Something is wrong on the internet. *Medium*. Available at: https://medium. com/@ jamesbridle/something-is-wrong-on-the-internet-c39c471271d2.

Brin, S. and Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833.

Bronk, R. (2013). Hayek on the wisdom of prices: a reassessment. *Erasmus Journal for Philosophy and Economics*, 6(1):82–107.

Brown, N. M. (2019). Methodological cyborg as black feminist technology: Constructing the social self using computational digital autoethnography and social media. *Cultural Studies? Critical Methodologies*, 19(1):55–67.

Bruno, C. (2002). The google adwords happening. *Iterature*. Available at: http://www.iterature.com/adwords/.

Bruno, C. (2012). The work of art in the age of meta-capital. In *Drunk on Capitalism. An Interdisciplinary Reflection on Market Economy, Art and Science*, pages 143–157. Springer.

Bruno, C. (2014). Psycho-Academic Dérive: a proposal. *A Peer-Reviewed Journal About*, 3(1).

Bruno, Christophe (2006). Interview with Christophe Bruno. *We Make Money not Art*. Available at: http://we-make-money-not-art.com.

Brunton, F. (2013). *Spam: A shadow history of the Internet*. MIT Press, Cambridge.

Brunton, F. and Nissenbaum, H. (2015). *Obfuscation: A user's guide for privacy and protest*. MIT Press, Cambridge.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of facebook algorithms. *Information, Communication & Society*, 20(1):30–44.

Bunz, M. (2013). *The silent revolution: How digitalization transforms knowledge, work, journalism and politics without making too much noise*. Springer, New York.

Butz, D. (2010). Autoethnography as Sensibility. In DeLyser, D., editor, *The SAGE Handbook of Qualitative Geography*, pages 138–155. Sage, London.

Buzzfeed (2016). Oxford dictionary of english accused of using sexist definitions. Available at: https://www.buzzfeed.com/laurasilver/go-on-call-me-a-rabid-feminist?

Cabell, M. and Huff, J. (2013). *American psycho*. Traumawien, Vienna.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Carpo, M. (2011). *The Alphabet and the Algorithm*. MIT Press, Cambridge.

Cayley, J. (2017). Aurature at the end(s) of electronic literature. *Electronic Book Review*. Available at: https://electronicbookreview.com/essay/aurature-at-the-ends-of-electronic-literature/.

Cixous, H., Cohen, K., and Cohen, P. (1976). The laugh of the medusa. *Signs: Journal of Women in Culture and Society*, 1(4):875–893.

Coeckelbergh, M. (2017). *Using words and things: Language and philosophy of technology*, volume 92. Routledge, New York.

Contreras-Koterbay, S. and Mirocha, Ł. (2016). *The New Aesthetic and Art: Constellations of the Postdigital*. Institute of Network Cultures, Amsterdam.

Cramer, F. (2015). What is post-digital? In Berry, D. and Dieter, M., editors, *Postdigital Aesthetics*, pages 12–26. Palgrave Macmillan, London.

Crampton, J. (2002). Virtual geographies: the ethics of the internet. In Proctor, J. D. and Smith, D. M., editors, *Geography and Ethics*, pages 88–108. Routledge, London.

Crampton, J. W. (2003). *The political mapping of cyberspace*. University of Chicago Press, Chicago.

Crampton, J. W. (2015). Collect it all: National security, big data and governance. *GeoJournal*, 80(4):519–531.

Crawford, K. (2013). The hidden biases in big data. *Harvard Business Review*. Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data.

Dalton, C. and Thatcher, J. (2014). What does a critical data studies look like, and why do we care? seven points for a critical approach to big data. *Society and Space*. https://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/.

Dalton, C. M., Taylor, L., and Thatcher, J. (2016). Critical data studies: A dialog on data and space. *Big Data & Society*, 3(1).

Danaher, J. (2016). The threat of algocracy: reality, resistance and accommodation. *Philosophy & Technology*, 29(3):245–268.

Darnton, R. (2009). The library in the information age: 6000 year of script. In Becker, K. and Stalder, F., editors, *Deep search : the politics of search beyond Google*, pages 32–45. Studien Verlag, Innsbruck.

Davies, W. (2018). *Economic Science Fictions*. MIT Press, Chicago.

De Certeau, M. (1988). *The Practice of Everyday Life*. University of California Press, Berkeley.

De Goede, M. (2014). The politics of privacy in the age of preemptive security. *International Political Sociology*, 8(1):100–104.

De Goede, M. (2018). The chain of security. *Review of International Studies*, 44(1):24–42.

Debord, G. (2016). *Society of the Spectacle*. Last Word Press, Washington.

Derrida, J. (1988). *Limited inc*. Northwestern University Press, Evanston.

Derrida, J. and Spivak, G. (1976). *Of grammatology*. JHU Press, Baltimore.

Diaz, A. (2008). Through the google goggles: Sociopolitical bias in search engine design. *Web Search: Multidisciplinary Perspectives*, 14:11–34.

Dick, M. (2011). Search engine optimisation in uk news production. *Journalism Practice*, 5(4):462–477.

Ding, J. (2017). *Linguistic Prefabrication: A Discourse Analysis Approach*. Springer, New York.

Domínguez Figaredo, D., Beaulieu, A., Estalella, A., Cruz, E., Schnettler, B., and Read, R. (2007). Virtual ethnography. *Qualitative Social Research*, 8(3).

Dourish, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society*, 3(2).

Duggan, M. (2017a). The everyday reality of a digitalizing world: Driving and geocaching. In Felgenhauer, T. and Gbler, K., editors, *Geographies of Digital Culture*, pages 71–83. Routledge, Oxford.

Duggan, M. (2017b). Questioning digital ethnography in an era of ubiquitous computing. *Geography Compass*, 11(5).

DuPont, Q. (2013). Cracking the agrippa code: Cryptography for the digital humanities. *Scholarly and Research Communication*, 4(3).

Dwyer, A. (2017). Situating cyberspace(s): Ethnography as depth. *Of Other Cyber Securities: methods, practices, spaces*. Paper given at Royal Holloway, June 2017.

Eisenstein, E. L. (1980). *The printing press as an agent of change*, volume 1. Cambridge University Press, Cambridge.

Ellis, B. (2010). *American psycho*. Vintage, New York.

Ellis, C. (2004). *The ethnographic I: A methodological novel about autoethnography*. Rowman Altamira, Lanham.

Ellis, C., Adams, T. E., and Bochner, A. P. (2011). Autoethnography: an overview. *Historical Social Research/Historische Sozialforschung*, pages 273–290.

Enloe, C. (1990). Womenandchildren: making feminist sense of the persian gulf crisis. *The Village Voice*, 25(9):1990.

Ezrachi, A. and Stucke, M. E. (2016). *Virtual Competition*. Oxford University Press, Oxford.

Fairclough, N. (2001). *Language and power*. Routledge, New York.

Fairclough, N. (2013a). *Critical discourse analysis: The critical study of language*. Routledge.

Fairclough, N. (2013b). *Language and Power*. Routledge, New York.

Ferragina, P. and Venturini, R. (2013). Web search. In Ausiello, G. and Petreschi, R., editors, *The Power of Algorithms*, pages 107–137. Springer, Berlin Heidelberg.

Feuz, M., Fuller, M., and Stalder, F. (2011). Personal web searching in the age of semantic capitalism: Diagnosing the mechanisms of personalisation. *First Monday*, 16(2).

Flick, C. (2016). Informed consent and the facebook emotional manipulation study. *Research Ethics*, 12(1):14–28.

Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and information technology*, 1(1):33–52.

Floridi, L. (2011). *The philosophy of information*. Oxford University Press, Oxford.

Floridi, L. and Taddeo, M. (2016). What is data ethics? *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2083).

Floyer, C. (1998). Monochrome till receipt (white). *The Tate*. Available at: https://www.tate.org.uk/art/artworks/floyer-monochrome-till-receipt-white-t12894.

Foucault, M. (1990). *The history of sexuality: An introduction, volume I*. Vintage, New York.

Foucault, M. (2012). *The archaeology of knowledge*. Random House LLC.

Foxtrot, J. (2015). Bog eye man. Available at: https://vimeo.com/117173388.

Franklin, M. I. (2002). Reading Walter Benjamin and Donna Haraway in the age of digital reproduction. *Information, Communication & Society*, 5(4):591–624.

FTC (2017). Complaint Counsel's Corrected Pre-Trial Brief and Exhibits in case of 1-800 Contacts, Inc, (586211), 3 April 2017. *Federal Trade Commission*. Available at: https://www.ftc.gov/system/files/documents/cases/586211.pdf.

Fuchs, C. (2011). A contribution to the critique of the political economy of google. *Fast Capitalism*, 8(1):1–24.

Fuchs, C. and Mosco, V. (2015). *Marx in the age of digital capitalism*. Brill, Leiden.

Fuller, M. and Goffey, A. (2012). *Evil media*. MIT Press, Cambridge.

Galloway, A. R. (2012). *The interface effect*. Polity, Cambridge.

Gerhart, S. (2004). Do web search engines suppress controversy? *First Monday*, 9(1).

Gibson, W. (2007). Introduction to agrippa: a book of the dead. Available at www.williamgibsonbooks.com.

Gibson, W. and Ashbaugh, D. (1992). *Agrippa:(a Book of the Dead)*. Kevin Begos Pub., New York.

Gilge, C. (2016a). Citizen participation as microfascism: Marginalising labour in web 2.0. In Lancione, M., editor, *Rethinking Life at the Margins: The Assemblage of Contexts, Subjects, and Politics*, chapter 13, pages 198–211. Routledge, Oxford.

Gilge, C. (2016b). Google street view and the image as experience. *Geohumanities*, 2(2):469–484.

Gilge, C. R. (2015). September 04, 1998: Entanglement and experimentation; or, cultural fascism and google street view. PhD thesis, University of Washington.

Gillespie, T. (2014). The relevance of algorithms. In Gillespie, T., Boczkowski, P. J., and Foot, K. A., editors, *Media Technologies: Essays on Communication, Materiality, and Society*, page 167. MIT Press, Cambridge.

Gitelman, L. (2013a). *Raw Data Is an Oxymoron*. MIT Press, Cambridge.

Gitelman, L. (2013b). Response to Introna's 'Algorithms, Performativity and Governability'. *Governing Algorithms Conference, New York 05/05/2013*. Available at: http://governingalgorithms.org/wp-content/uploads/2013/05/gitelman-response.pdf.

Gitelman, L. (2014). *Paper knowledge: Toward a media history of documents*. Duke University Press, Durham.

Goffey, A. (2008). Algorithm. In Fuller, M., Maline, R., and Cubitt, S., editors, *Software Studies: A Lexicon*, pages 15–20. MIT Press, Cambridge.

Goldstein, R. (2014). *Plato at the Googleplex: Why Philosophy Won't Go Away*. Random House LLC, New York.

Golumbia, D. (2009). *The cultural logic of computation*. Harvard University Press, Cambridge.

Golumbia, D. (2014). Death of a discipline. *differences*, 25(1):156–176.

Google (2012). Re: Where Can I get Googles negative word list for Ads. Available at: https://www.en.advertisercommunity.com/t5/Ad-Approval-Policy/Where-Can-I-get-Googles-negative-word-list-for-Ads/td-p/44492.

Google (2014). Will backlinks lose their importance in ranking? Available at: https://www.youtube.com/watch?v=iC5FDzUh0P4.

Google (2016). AdWords Policies. Available at: https://support.google.com/adwordspolicy/answer/6008942?hl=en-GB.

Google (n.d.). How search works. www.google.co.uk/insidesearch/howsearchworks/crawling-indexing.html.

Goriunova, O. and Shulgin, A. (2008). Glitch. In Fuller, M., Maline, R., and Cubitt, S., editors, *Software Studies: A Lexicon*, pages 110–119. MIT Press, Cambridge.

Graham, M. (2010). Neogeography and the palimpsests of place: Web 2.0 and the construction of a virtual earth. *Tijdschrift voor economische en sociale geografie*, 101(4):422–436.

Graham, M. (2014). Geotagging reveals wikipedia is not quite so equal after all. *The Conversation*. Available at: http://theconversation.com.

Graham, M., De Sabbata, S., and Zook, M. A. (2015). Towards a study of information geographies: (im)mutable augmentations and a mapping of the geographies of information. *Geo: Geography and Environment*, 2(1):88–105.

Graham, M., Hogan, B., Straumann, R. K., and Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers*, 104(4):746–764.

Graham, R. (2017). Google and advertising: digital capitalism in the context of post-fordism, the reification of language, and the rise of fake news. *Palgrave Communications*, 3(1):45.

Graham, S. and Wood, D. (2003). Digitizing surveillance: categorization, space, inequality. *Critical Social Policy*, 23(2):227–248.

Graham, S. D. (2005). Software-sorted geographies. *Progress in Human Geography*, 29(5):562–580.

Granka, L. A. (2010). The politics of search: A decade retrospective. *The Information Society*, 26(5):364–374.

Grimmelmann, J. (2008). Google dilemma, the. *NYL Sch. L. Rev.*, 53:939.

Grimshaw, M. (2018). Towards a manifesto for a critical digital humanities: critiquing the extractive capitalism of digital society. *Palgrave Communications*, 4(1):21.

Groys, B. (2016). *In the flow*. Verso Books, New York.

HAC (2013). Written evidence submitted by Google to Home Affairs Committee [EC17], March 2013. *parliament.uk*. Available at: http://www.publications.parliament.uk/pa/cm201314/cmselect/cmhaff/70/70we17.htm.

HAC (2016). Verbal evidence given by Google to the Home Affairs Committee, 2 February 2016. *parliament.uk*. Available at: http://www.parliament.uk/business/committees/committees-a-z/commons-select/home-affairs-committee/news-parliament-2015/160129-countering-extremism-evidence/.

Halavais, A. (2017). *Search engine society*. John Wiley & Sons, New Jersey.

Halpern, O. (2015). *Beautiful data: A history of vision and reason since 1945*. Duke University Press, Durham.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist studies*, 14(3):575–599.

Haraway, D. (1991). *Simians, Cyborgs and Women: The reinvention of nature*. Routledge, New York.

Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press, Durham.

Harvey, D. (2017). *Marx, Capital, and the Madness of Economic Reason*. Oxford University Press, Oxford.

Hayek, F. A. (1945). The use of knowledge in society. *The American economic review*, pages 519–530.

Hayles, N. K. (2005). *My Mother Was a Computer: Digital Subjects and Literary Texts*. University of Chicago Press, Chicago.

Hayles, N. K. (2008a). *Electronic literature: New horizons for the literary*. University of Notre Dame Press, Notre Dame, Indiana.

Hayles, N. K. (2008b). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. University of Chicago Press, Chicago.

Hayles, N. K. (2012). *How we think: Digital media and contemporary technogenesis*. University of Chicago Press, Chicago.

Hess, A. (2008). Reconsidering the rhizome: A textual analysis of web search engines as gatekeepers of the internet. In Spink, A. and Zimmer, M., editors, *Web search: Multidisciplinary perspectives*. Springer, New York.

Hillis, K., Petit, M., and Jarrett, K. (2013). *Google and the Culture of Search*. Routledge, London.

Hinman, L. (2008). Searching ethics: The role of search engines in the construction and distribution of knowledge. In Spink, A. and Zimmer, M., editors, *Web search: Multidisciplinary perspectives*, page 67. Springer, New York.

Hoad, P. (2015). Google doodles aren't just naff, they're trying to rebrand the past. *The Guardian*. Available at: http://www.theguardian.com/commentisfree/2015/jul/14/google-doodles.

How Stuff Works (2016). The Land of Mordor? How hackers and spies Use Google Translate. Available at: http://now.howstuffworks.com/2016/01/13/mordor-google-translate.

Hoy, D. (2006). The virtual dependency of the post-avant and the problematics of flarf: What happens when poets spend too much time fucking around on the internet. *Jacket Magazine*. Available at: http://jacketmagazine.com/29/hoy-flarf.html.

Huffington Post (2014). The Best Place to Hide a Dead Body is Page Two of Google. Available at: https://www.huffingtonpost.com/chad-pollitt/the-best-place-to-hide-a-5168714.html.

Huffington Post (2017). Dementia Tax 'Alternative' Google Ads Taken Out By 'Concerned Individual'. Available at: https://www.huffingtonpost.co.uk/entry/dementia-taxalternative-google-ad-taken-out-by-concerned-individualuk5922f262e4b034684b0e1757.

Hughes, G. (1989). *Words in time: A social history of the English vocabulary*. Blackwell, Oxford.

Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology, & Human Values*, 41(1):17–49.

Introna, L. D. and Nissenbaum, H. (2000). Shaping the web: Why the politics of search engines matters. *The information society*, 16(3):169–185.

Jameson, F. (1974). *The prison-house of language: A critical account of structuralism and Russian formalism*, volume 332. Princeton University Press, New Jeresy.

Jansen, B. J., Hudson, K., Hunter, L., Liu, F., and Murphy, J. (2008). The google online marketing challenge: Classroom learning with real clients, real money, and real advertising campaigns. *Journal of Interactive Advertising*, 9(1):49–55.

Jarrett, K. (2014). A database of intention? In Rasch, M. and Knig, R., editors, *Society of the Query: Reflections on Web Search*, pages 16–29. Institute of Network Cultures, Amsterdam.

Jhally, S. (2014). *The codes of advertising: Fetishism and the political economy of meaning in the consumer society*. Routledge, New York.

Ji, F. (2004). *Linguistic engineering: Language and politics in Mao's China*. University of Hawaii Press, Honolulu.

Jigsaw (2015). The redirect method. Available at: http://redirectmethod.org/qa/.

Jobin, A. and Glassey, O. (2014). 'I Am not a Web Search Result! I Am a Free Word.' The Categorization and Commodification of 'Switzerland' by Google. In Rasch, M. and Konig, R., editors, *Society of the Query*. Amsterdam.

Johnston, D. J. (2016). *Aesthetic Animism: Digital Poetry's Ontological Implications*. MIT Press, Cambridge.

Jones, N. (2017). Glitch poetics: The posthumanities of error. In Tabbi, J., editor, *The Bloomsbury Handbook of Electronic Literature*. Bloomsbury Academic, London.

Joque, J. (2018). *Deconstruction machines: writing in the age of cyberwar*, volume 54. U of Minnesota Press, Minneapolis.

Joseph, J. E. (2006). *Language and politics*. Edinburgh University Press, Edinburgh.

Just, N. and Latzer, M. (2017). Governance by algorithms: reality construction by algorithmic selection on the internet. *Media, Culture & Society*, 39(2):238–258.

Kaplan, F. (2011). Vers le capitalisme linguistique: Quand les mots valent de l'or. *Le Monde diplomatique*, (692):28.

Kaplan, F. (2014). Linguistic capitalism and algorithmic mediation. *Representations*, 127(1):57–63.

Khayati, M. (1981). Captive words: Preface to a situationist dictionary. In Knabb, K., editor, *Situationist International Anthology*. Bureau of Public Secrets, Berkeley.

Kingsbury, P. and Jones, J. P. (2009). Walter Benjamin's dionysian adventures on Google Earth. *Geoforum*, 40(4):502–513.

Kinsley, S. (2013). The matter of 'virtual' geographies. *Progress in Human Geography*, 38(3):364–384.

Kirschenbaum, M. (2014). What is digital humanities, and why are they saying such terrible things about it? *differences*, 25(1):46–63.

Kitchin, R. (2014a). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, New York.

Kitchin, R. (2014b). Thinking critically about and researching algorithms. programmable city working paper 5.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1):14–29.

Kitchin, R. and Dodge, M. (2011). *Code/space: Software and everyday life*. MIT Press, Cambridge.

Kitchin, R., Graham, M., Mattern, S., and Shaw, J. (2019). *Running a City Like a Company and Other Fables*. Meatspace Press, London.

Kitchin, R. and Kneale, J. (2001). Science fiction or future fact? exploring imaginative geographies of the new millennium. *Progress in Human Geography*, 25(1):19–35.

Kittler, F. A. (1999). *Gramophone, film, typewriter*. Stanford University Press, Palo Alto.

Knibbs, K. (2016). Google Now Shows 'Don't Do Jihad' Ad Results to ISIS Wannabes. *Gizmodo*. Available at: http://gizmodo.com/google-now-shows-dont-do-jihad-ad-results-to-isis-wanna-1756642501.

Kockelman, P. (2010). *Language, culture, and mind: Natural constructions and social kinds*, volume 10. Cambridge University Press, Cambridge.

Kockelman, P. (2014). Linguistic anthropology in the age of language automata. In Enfield, N. J., Kockelman, P., and Sidnell, J., editors, *The Cambridge Handbook of Linguistic Anthropology*, pages 708–733. Cambridge University Press, Cambridge.

Koepnick, L. P. (1999). *Walter Benjamin and the aesthetics of power*. U of Nebraska Press, Lincoln.

König, R. and Rasch, M. (2014). *Society of the Query Reader: Reflections on Web Search*. Institute of Network Technology, Amsterdam.

Kordestani, O. (2008). Our agreement to provide ad technology to Yahoo. *Google Blogspot*. Available at: https://googleblog.blogspot.co.uk/2008/06/our-agreement-to-provide-ad-technology.html.

Kozinets, R. (2017). Netnography: radical participative understanding for a networked communications society. In Willig, C. and Rogers, W., editors, *The SAGE Handbook of Qualitative Research in Psychology*, page 374. Sage, New York.

Krebs, B. (2014). Lorem Ipsum: Of Good & Evil, Google & China. *Krebs on Security*. Available at: https://krebsonsecurity.com/2014/08/lorem-ipsum-of-good-evil-google-china/.

Krebs, S. (2009). *The Impact of Printing on the Development of the English Language During the Period of the Renaissance*. GRIN Verlag, Munich.

Kwan, M.-P. (2016). Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. *Annals of the American Association of Geographers*, 106(2):274–282.

Lally, N. and Burns, R. (2018). Special section editorial: Toward a geographical software studies. *Computational Culture*, 6.

Land, S. E. (2010). Schmidt: Listing Google?s 200 Ranking Factors Would Reveal Business Secrets. Available at: https://searchengineland.com/schmidt-listing-googles-200-ranking-factors-would-reveal-business-secrets-51065.

Lange, O. (1936). On the economic theory of socialism: part one. *The review of economic studies*, 4(1):53–71.

Lange, O. (1937). On the economic theory of socialism: part two. *The review of economic studies*, 4(2):123–142.

Langley, P. and Leyshon, A. (2017). Platform capitalism: the intermediation and capitalisation of digital economic circulation. *Finance and society.*, 3(1):11–31.

Langville, A. N. and Meyer, C. D. (2011). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, New Jersey.

Lash, S. (2007). Power after hegemony: Cultural studies in mutation? *Theory, culture & society*, 24(3):55–78.

Lefebvre, H. (1991). *The Production of Space*. Blackwell, Oxford.

Leszczynski, A. (2018). Digital methods I: Wicked tensions. *Progress in Human Geography*, 42(3):473–481.

Leszczynski, A. and Crampton, J. (2016). Introduction: Spatial big data and everyday life. *Big Data & Society*, 3(2).

Levy, S. (2011). *In the plex: How Google thinks, works, and shapes our lives*. Simon and Schuster, New York.

Lim, M. (2008). Postmodern paradigms and brand management in the 'search' economy. *International Journal of Internet Marketing and Advertising*, 5(1-2):4–16.

Littau, K. (1997). Translation in the age of postmodern production: From text to intertext to hypertext. In *Forum for Modern Language Studies*, volume 33, pages 81–96. Oxford University Press.

Liu, A. (2012). Where is cultural criticism in the digital humanities? In Gold, M. K., editor, *Debates in the digital humanities*, pages 490–509. University of Minnesota Press, Minneapolis.

Liu, A., Hehmeyer, P., Hodge, J., Knight, K., Roh, D., Swanstrom, E., and Kirschenbaum, M. (2005). The Agrippa files. *Transcriptions Project, UC Santa Barbara*. Available at: http://agrippa.english.ucsb.edu/.

Lorde, A. (2012). *Sister outsider: Essays and speeches*. Crossing Press, New York.

Lovink, G. (2009). Society of the query: The googlization of our lives. In Becker, K. and Stalder, F., editors, *Deep search: The politics of search engines beyond Google*, pages 45–53. Studienverlag, Innsbruck.

Lunenfeld, P. (2000). *The digital dialectic: New essays on new media*. MIT Press, Cambridge.

Lupton, D. (2016). *The quantified self*. John Wiley & Sons, New Jersey.

Lyotard, J.-F. (1984). *The postmodern condition: A report on knowledge*. University of Minnesota Press, Minneapolis.

Maccatrozzo, V. (2012). Burst the filter bubble: using semantic web to enable serendipity. In *International Semantic Web Conference*, pages 391–398. Springer.

Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, 15(5):769–787.

Mahnke, M. and Uprichard, E. (2014). Algorithming the algorithm. In Konig and Rasch, editors, *Society of the query reader: Reflections on Web search*, pages 256–271. Institute of Network Cultures, Amsterdam.

Mäkelä, M. (2007). Knowing through making: The role of the artefact in practice-led research. *Knowledge, Technology & Policy*, 20(3):157–163.

Manovich, L. (2001). *The Language of New Media*. MIT Press, Cambridge.

Manovich, L. (2013). *Software takes command*. A&C Black, London.

Marazzi, C. (2008). *Capital and Language: From the New Economy to the War Economy*. Semiotext, Los Angeles.

Marazzi, C. (2011). *Capital and affects: The politics of the language economy*. The MIT Press, Cambridge.

Marche, S. (2012). Literature is not data: Against digital humanities. *LA Review of Books*. Available at: https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/.

Marczewska, K. (2015). Erasing in the algorithmic extreme: Mimi Cabell and Jason Huff's American Psycho. *NMC Media-N: Journal of the New Media*. Available at: http://median.newmediacaucus.org/the_aesthetics_of_erasure/erasing-in-the-algorithmic-extreme-mimi-cabell-and-jason-huffs-american-psycho/.

Martin, R. (2013). After economy? Social logics of the derivative. *Social Text*, 31(1 (114)):83–106.

Massey, D. (2013). *Space, place and gender*. John Wiley & Sons, New Jersey.

Mayer, S. (2013). (Re)Search/Destroy? Or, Mistaking Poetry for Viruses. *Archive of the Now*. Available at: https://www.archiveofthenow.org/pages/researchdestroy-or-mistaking-poetry-for-viruses/.

McGarrigle, C. (2010). The construction of locative situations: locative media and the situationist international, recuperation or redux? *Digital Creativity*, 21(1):55–62.

McQuillan, D. (2015). Algorithmic states of exception. *European Journal of Cultural Studies*, 18(4-5):564–576.

McQuillan, D. (2018). Data science as machinic neoplatonism. *Philosophy & Technology*, 31(2):253–272.

Medina, E. (2011). *Cybernetic revolutionaries: technology and politics in Allende's Chile*. MIT Press, Cambridge.

Mehta, A., Saberi, A., Vazirani, U., and Vazirani, V. (2007). Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22.

Melville, H. (1998). *Billy Budd, Sailor and Selected Tales*. Penguin, London.

Menkman, R. Glitch studies manifesto. In Lovink, G. and Somers Miles, R., editors, *Video vortex reader II: Moving images beyond YouTube*, pages 336–347. Institute of Network Cultures, Amsterdam.

Merrifield, A. (2015). Amateur urbanism. *City*, 19(5):753–762.

Metahaven (2009). Peripheral forces: On the relevance of marginality in networks. In Becker, K. and Stalder, F., editors, *Deep search: The politics of search engines beyond Google*, pages 185–197. Studienverlag, Innsbruck.

Mikolov, T., Le, Q. V., and Sutskever, I. (2013). Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

Miles, S. (2017). Sex in the digital city: location-based dating apps and queer urban life. *Gender, Place & Culture*, 24(11):1595–1610.

Mirowski, P. (2009). Why there is (as yet) no such thing as an economics of knowledge. In Ross, D. and Kincaid, H., editors, *The Oxford Handbook of Philosophy of Economics*, pages 99–156. OUP, Oxford.

Mirowski, P. and Nik-Khah, E. (2017). *The Knowledge We Have Lost in Information: The History of Information in Modern Economics*. OUP, Oxford.

Mises, L. v. (1920). Die wirtschaftsrechnung im sozialistischen gemeinwesen. *Archiv für Sozialwissenschaft und Sozialpolitik*, 47(1):86–121.

Mittelstadt, B. D. and Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, 22(2):303–341.

Molchanov, M. A. (2015). Russia as Ukraine?s 'Other': Identity and Geopolitics. *E-International Relations*. Available at: https://www.e-ir.info/2015/05/05/russia-as-ukraines-other-identity-and-geopolitics/.

Moore, P. V. (2017). *The Quantified Self in Precarity: Work, Technology and What Counts*. Routledge, London.

Moretti, F. (2013). *Distant reading*. Verso Books, New York.

Morozov, E. (2012). *The net delusion: The dark side of Internet freedom*. PublicAffairs, New York.

Morozov, E. (2014). The planning machine: Project cybersyn and the origins of the big data nation. *The New Yorker*. Available at: https://www.newyorker.com/magazine/2014/10/13/planning-machine.

Morris, A. and Swiss, T. (2006). *New Media Poetics: Contexts/Technotexts/Theories*. The MIT Press, Cambridge.

Mould, O. (2018). *Against Creativity*. Verso Books, London.

Moulier-Boutang, Y. (2011). *Cognitive capitalism*. Polity, Cambridge.

Nabugodi, M. (2014). Pure Language 2.0: Walter Benjamin's Theory of Language and Translation Technology. *Feedback*. Available at: http://openhumanitiespress.org/feedback/literature/pure-language-2-0-walter-benjamins-theory-of-language-and-translation-technology/.

Neyland, D. (2014). On Organizing Algorithms. *Theory, Culture & Society*, 32(1):119–132.

Nguyen, D., McGillivray, B., and Yasseri, T. (2018). Emo, love and god: making sense of urban dictionary, a crowd-sourced online dictionary. *Royal Society Open Science*, 5(5):172320.

Nixon, B. (2016). The old media business in the new: the Googlization of everything as the capitalization of digital consumption. *Media, Culture & Society*, 38(2):212–231.

Noble, S. U. (2013). Google search: Hyper-visibility as a means of rendering black women and girls invisible. *InVisible Culture*, 19:1–23.

Noble, S. U. (2018). *Algorithms of Oppression: How search engines reinforce racism*. NYU Press.

Nuotio, S. (2012). Google poetics. Available at http://www.googlepoetics.com/.

O'Gorman, M. (2006). *E-Crit: digital media, critical theory and the humanities*. University of Toronto Press, Toronto.

Olofsson, J. (2015). 'Did you mean 'Why are *women* cranky?' Google: A means of Inscription , a Means of De-inscription? In Svensson, P. and Goldberg, D. T., editors, *Between Humanities and the Digital*. MIT Press, Cambridge.

Olsson, G. (1978). Of ambiguity or far cries from a memorializing mamafesta. In Ley, D. and Samuels, M. S., editors, *Humanistic geography: Prospects and problems*. Croom-Helm, London.

Olsthoom, P. (2013). *It's complicated: The power of Facebook.* Ehio Media, Amsterdam.

O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Broadway Books, New York.

Orwell, G. (1968). Politics and the english language. In Orwell, S. and Angus, I., editors, *The collected essays, journalism and letters of George Orwell*, pages 127–140. Harcourt, Brace & World, New York.

Orwell, G. (2000). *Nineteen Eighty-Four.* Penguin Classics, London.

Owen, T. (2015). The Violence of Algorithms: Big Data Is Only as Smart as Those Who Generate It. *Foriegn Affairs.* Available at: https://www.foreignaffairs.com/articles/2015-05-25/violence-algorithms.

Paglen, T. (2016). Invisible images (your pictures are looking at you). *The New Inquiry.* Available at: https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/.

Panger, G. (2016). Reassessing the facebook experiment: critical thinking about the validity of big data research. *Information, Communication & Society*, 19(8):1108–1126.

Parikka, J. (2013). *What is media archaeology?* John Wiley & Sons, New Jersey.

Parikka, J. and Bishop, R. (2017). *Across and beyond: A transmediale reader in post-digital practices, concepts and institutions.* Sternberg Press, Berlin.

Parikka, J. and Sampson, T. D. (2009). *The spam book: on viruses, porn, and other anomalies from the dark side of digital culture.* Hampton Press, London.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* Penguin, London.

Parisi, L. (2017). Computational logic and ecological rationality. In Hörl, E. and Burton, J., editors, *General Ecology: The New Ecological Paradigm*, pages 75–99. Bloomsbury Academic, London.

Parks, L. (2018). *Rethinking Media Coverage: Vertical Mediation and the War on Terror.* Routledge, London.

Pasquale, F. (2015a). The algorithmic self. *The hedgehog review*, 17(1):30–46.

Pasquale, F. (2015b). *The black box society: The secret algorithms that control money and information*. Harvard University Press, Cambridge.

Pasquale, F. (2016). Algorithms and markets: The perfect excuses. *Unboxing Conference, Berlin, 3 December 2016*. Available at: https://www.youtube.com/watch?v=HteGOTWRqO8&feature=youtu.be&t=25m2s.

Pasquinelli, M. (2009). Google's pagerank: Diagram of the cognitive capitalism and rentier of the common intellect. In Becker, K. and Stalder, F., editors, *Deep search: The politics of search engines beyond Google*, pages 152–162. Studienverlag, Innsbruck.

Paul, C. and Levy, M. (2015). Genealogies of the new aesthetic. In Berry, D. and Dieter, M., editors, *Postdigital Aesthetics*, pages 27–43. Palgrave Macmillan, London.

Perry, R. (2016). Lost in my life (receipts seated). *Rachel Perry Studio*. Available at: https://www.rachelperrystudio.com/lost-in-my-life/.

Pink, S. (2016). *Digital ethnography*. Springer, New York.

Plant, S. (2002). *The most radical gesture: The Situationist International in a postmodern age*. Routledge, London.

Polanyi, M. (1961). Knowing and being. *Mind*, 70(280):458–470.

Poon, M. (2013). Response to Tarleton Gillespie's 'The relevance of algorithms'. *Governing Algorithms Conference, New York, 17 May 2013*. Available at: http://opentranscripts.org/transcript/response-tarleton-gillespie-relevance-algorithms/.

Posner, M. (2015). The radical potential of the digital humanities: The most challenging computing problem is the interrogation of power. Available at: http://blogs.lse.ac.uk/impactofsocialsciences/2015/08/12/the-radical-unrealized-potential-of-digital-humanities/.

Poster, M. (2001). *What's the Matter with the Internet?* University of Minnesota Press, Minneapolis.

Powles, J. and Hodson, H. (2017). Google deepmind and healthcare in an age of algorithms. *Health and technology*, 7(4):351–367.

Pressman, J. (2014). *Digital modernism: Making it new in new media*, volume 21. Oxford University Press, Oxford.

Reilly, P. (2008). Googling terrorists: Are northern irish terrorists visible on internet search engines. In Spink, A. and Zimmer, M., editors, *Web search: Multidisciplinary perspectives*, page 10. Springer, New York.

Riviere, S. (2015). *Kim Kardashian's Marriage*. Faber & Faber, London.

Robinson, L. and Schulz, J. (2009). New avenues for sociological inquiry: Evolving forms of ethnographic practice. *Sociology*, 43(4):685–698.

Rogers, R. (2009). The googlization question: towards the inculpable engine? In Becker, K. and Stalder, F., editors, *Deep Search: The politics of search beyond Google*, pages 173–184. Studien Verlag, Innsbruck.

Rogers, R. (2013). *Digital methods*. MIT Press, Cambridge.

Rouvroy, A. (2013). Algorithmic governmentality and the end(s) of critique. In Rasch, M. and König, R., editors, *Society of the Query Reader: Reflections on Web Search*. Institute of Network Cultures, Amsterdam.

Rushkoff, D. (2016). *Throwing rocks at the Google bus: how growth became the enemy of prosperity*. Penguin, London.

Sack, W. (2017). Out of bounds: Language limits, language planning, and the definition of distance in the new spaces of linguistic capitalism. *Computational Culture: a journal of software studies*, 6.

Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. (2014). Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. *Paper presented to 'Data and Discrimination: Converting Critical Concerns into Productive Inquiry', a preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014; Seattle USA*. Available at http://www-personal.umich.edu/ csandvig/research/.

Schmidt, E. and Rosenberg, J. (2014). *How Google works*. Hachette, New York.

Scrivener, S. and Chapman, P. (2004). The practical implications of applying a theory of practice based research: a case study. *Working papers in art and design*, 3(1).

Search Engine People (2010). The 10 most incredible Google bombs. Available at: https://www.searchenginepeople.com/blog/incredible-google-bombs.html.

Seaver, N. (2013). Knowing algorithms. *Media in Transition*, 8:1–12.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2).

Seo, J. M. (2009). Plagiarism and poetic identity in martial. *American Journal of Philology*, 130(4):567–593.

Seresinhe, C. I., Preis, T., and Moat, H. S. (2017). Using deep learning to quantify the beauty of outdoor places. *Royal Society open science*, 4(7):170170.

Shaw, J. and Graham, M. (2017). An informational right to the city? code, content, control, and the urbanization of information. *Antipode*, 49(4):907–927.

Shehadeh, K. K. (1999). The wassenaar arrangement and encryption exports: An ineffective export control regime that compromises united states economic interests. *Am. U. Int'l L. Rev.*, 15:271.

Simmel, G. (2012). The metropolis and mental life. In Mele, C., editor, *The urban sociology reader*, pages 37–45. Routledge, London.

Sjoberg, L. (2006). *Gender, Justice, and the Wars in Iraq: A Feminist Reformulation of Just War Theory*. Lexington Books, New York.

Sjoberg, L. and Gentry, C. E. (2007). *Mothers, monsters, whores: women's violence in global politics*. Zed Books, London.

Spink, A. and Zimmer, M. (2008). *Web search: Multidisciplinary perspectives*. Springer, New York.

Srnicek, N. (2016). *Platform capitalism*. John Wiley & Sons, New Jersey.

Stalder, F. and Mayer, C. (2009). The second index: Search engines, personalization and surveillance. In Becker, K. and Stalder, F., editors, *Deep search: The politics of search engines beyond Google*, pages 98–115. Studienverlag, Innsbruck.

Statcounter (2019). Search engine market share. Available at: http://gs.statcounter.com/search-engine-market-share/.

Steyerl, H. (2017). *Duty Free Art: Art in the Age of Planetary Civil War*. Verso Books, London.

Stiegler, B. (1998). *Technics and time: The fault of Epimetheus*. Stanford University Press, Palo Alto.

Stiegler, B. (2015). Automatic society 1: The future of work - introduction. *La Deleuziana: Online Journal of Philosophy*, 1:121–140.

Straube, T. (2016). Stacked spaces: Mapping digital infrastructures. *Big Data & Society*, 3(2).

Striphas, T. (2010). The abuses of literacy: Amazon kindle and the right to read. *Communication and Critical/Cultural Studies*, 7(3):297–317.

Striphas, T. (2015). Algorithmic culture. *European Journal of Cultural Studies*, 18(4-5):395–412.

Tan, A. (2009). Google adwords: trademark infringer or trade liberalizer. *Mich. Telecomm. & Tech. L. Rev.*, 16:473.

Ter Heerdt, K. (in progress). The monetization of language by Google Adwords. Available at: http://kimberleyterheerdt.com.

The British Library (n.d.). 1755 - Johnson's Dictionary. Available at: http://www.bl.uk/learning/langlit/dic/johnson/1755johnsonsdictionary.html.

The Economist (2017). The world?s most valuable resource is no longer oil, but data. Available at: www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.

The Electoral Commission (2016). Spending on digital campaign activity. Available at: https://www.electoralcommission.org.uk/find-information-by-subject/political-parties-campaigning-and-donations/digital-campaigning/3-spending-on-digital-campaign-activity.

The Financial Times (2016). Harsh truths about fake news for Facebook, Google and Twitter. Available at: www.ft.com/content/2910a7a0-afd7-11e6-a37c-f4a01f1b0fa1.

The Guardian (2014). Ursula K Le Guin's speech at National Book Awards: 'Books aren't just commodities'. Available at: https://www.theguardian.com/books/2014/nov/20/ursula-k-le-guin-national-book-awards-speech.

The Guardian (2016a). Google translates Russia to 'Mordor' and minister's name to 'sad little horse'. Available at:

http://www.theguardian.com/technology/2016/jan/07/google-translates-russia-mordor-foreign-minister-ukrainian.

The Guardian (2016b). How Facebook powers money machines for obscure political 'news' sites. Available at: https://www.theguardian.com/technology/2016/aug/24/facebook-clickbait-political-news-sites-us-election-trump.

The Guardian (2018). 'Jewtropolis': antisemitic hack hits map used by Snapchat and Citi Bike. Available at: https://www.theguardian.com/society/2018/aug/31/jewtropolis-google-maps-snapchat-citi-bike-abuse.

The Mozilla Blog (2018). Mozilla announces $225,000 for art and advocacy exploring artificial intelligence. Available at: https://blog.mozilla.org/blog/2018/06/04/mozilla-announces-225000-for-art-and-advocacy-exploring-artificial-intelligence/.

The New York Times (2017). From headline to photograph, a fake news masterpiece. Available at: https://www.nytimes.com/2017/01/18/us/fake-news-hillary-clinton-cameron-harris.html.

The New York Times (2018). The business of war: Google employees protest work for the Pentagon. Available at: https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html.

The SEM Post (2016). Google's Keyword Planner Now Combines Keywords for Search Volume. Available at: http://www.thesempost.com/googles-keyword-planner-now-combines-keywords-for-search-volume/.

The Victoria & Albert Museum (2018). Digital Futures: EVA London 2018. Available at: https://www.vam.ac.uk/event/EY8Od6nw/digital-futures-eva-london-2018.

Thornton, P. (2017). Geographies of (con)text: language and structure in a digital age. *Computational Culture: a journal of software studies*, 6.

Thornton, P. (2018). A critique of linguistic capitalism: Provocation/intervention. *Geo-Humanities*, 4(2):417–437.

Thornton, P. (2019). Subprime language and the crash. In Kitchin, R., Graham, M., Mattern, S., and Shaw, J., editors, *Running a City Like a Company and Other Fables*. Meatspace Press, London.

Thornton, P. and Danaher, J. (forthcoming). On the wisdom of algorithmic markets.

Thornton, P. and Morris, J. H. (2017). Subprime language: The value of words in an age of linguistic capitalism, digital advertising and fake news. *Power 2.0: New Digital Geographies*. Paper given at RGS-IBG Annual International Conference, August 2017.

Thrift, N. and French, S. (2002). The automatic production of space. *Transactions of the Institute of British Geographers*, 27(3):309–335.

Totaro, P. and Ninno, D. (2014). The concept of algorithm as an interpretative key of modern rationality. *Theory, Culture & Society*, 31(4):29–49.

Tufekci, Z. (2014a). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *ICWSM*, 14:505–514.

Tufekci, Z. (2014b). Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7).

Tufekci, Z. (2015). Algorithmic harms beyond facebook and google: Emergent challenges of computational agency. *J. on Telecomm. & High Tech. L.*, 13:203.

Turing, A. M. (2009). Computing machinery and intelligence. In Epstein, R., Roberts, G., and Beber, G., editors, *Parsing the Turing Test*, pages 23–65. Springer, New York.

Turnbull, D. and Bright, L. F. (2008). Advertising academia with sponsored search: an exploratory study examining the effectiveness of google adwords at the local and global level. *International Journal of Electronic Business*, 6(2):149–171.

Tzara, T. (2018). *Seven Dada manifestos and lampisteries*. Alma Books, Richmond.

UN Women (2013). UN women ad series reveals widespread sexism. Available at: http://www.unwomen.org/en/news/stories/2013/10/women-should-ads.

US Securities and Exchange Commission (2015). Form 10-K Annual Report Alphabet Inc. and Google Inc. Available at: https://www.sec.gov/Archives/edgar/data/1288776/000165204416000012/goog10-k2015.htm.

Vaidhyanathan, S. (2012). *The Googlization of everything: (and why we should worry)*. Univ of California Press, Berkeley.

Van Couvering, E. (2004). New media? The political economy of Internet search engines. *Paper given at the Annual Conference of the International Association of Media & Communications Researchers, Porto Alegre, Brazil*.

Van Couvering, E. (2007). Is relevance relevant? market, science, and war: Discourses of search engine quality. *Journal of Computer-Mediated Communication*, 12(3):866–887.

Van Couvering, E. (2010). *Search engine bias: the structuration of traffic on the World-Wide Web*. PhD thesis, The London School of Economics and Political Science (LSE).

Vanderbeeken, R., Le Roy, F., Stalpaert, C., and Aerts, D. (2011). *Drunk on capitalism. An interdisciplinary reflection on market economy, art and science*, volume 11. Springer Science & Business Media, New York.

Wall, A. (2008). Google AdWords Price Fixing. *SEO Book*. Available at: http://www.seobook.com/goog-yhoo-tie-google-already-engages-price-fixing.

Wark, M. (2015). Benjamedia. *Public Seminar*. Available at: http://www.publicseminar.org/2015/08/benjamedia/.

Wark, M. (2017). After Capitalism, the Derivative: For Randy Martin. *Public Seminar*. Available at: http://www. publicseminar. org/2017/04/derivative/.

Weizman, E. (2012). *The least of all possible evils: humanitarian violence from Arendt to Gaza*. Verso, London.

Werckmeister, O. K. (1996). Walter Benjamin's angel of history, or the transfiguration of the revolutionary into the historian. *Critical Inquiry*, 22(2):239–267.

Wilmott, C. (2016). Small moments in spatial big data: calculability, authority and interoperability in everyday mobile mapping. *Big Data & Society*, 3(2).

Wilson, M. W. (2011). Data matter (s): legitimacy, coding, and qualifications-of-life. *Environment and Planning D: Society and Space*, 29(5):857–872.

Wilson, M. W. (2018). On being technopositional in digital geographies. *cultural geographies*, 25(1):7–21.

Wood, D. and Graham, S. (2006). Permeable boundaries in the software-sorted society: Surveillance and the differentiation of mobility. In Sheller, M. and Urry, J., editors, *Mobile technologies of the city*, pages 177–191. Routledge, London.

Woodcock, J. (2017). *Working the Phones: Control and Resistance in Call Centres*. Pluto Press, Chicago.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4):505–523.

Zimmer, M. and Proferes, N. J. (2014). A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66(3):250–261.

Zook, M. A. and Graham, M. (2007). The creative reconstruction of the internet: Google and the privatization of cyberspace and digiplace. *Geoforum*, 38(6):1322–1343.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1):75–89.