Exploring Place through Twitter text: An investigation into the use of ambient data for furthering geographic understandings of cities

Iona Fitzpatrick BA MSc

Thesis submitted to The University of Nottingham

for the degree of Doctor of Philosophy

November 2019

Acknowledgements

The author is supported by the Horizon Centre for Doctoral Training at the University of Nottingham (RCUK Grant No. EP/L015463/1) and by Ordnance Survey Ltd.

I would like to thank my supervisors Professor Svenja Adolphs and Dr. Gary Priestnall, whose expertise and good humour saw this project through to completion.

My thanks also go out to Jeremy Morley and the research team at the Ordnance Survey who co-sponsored this project and to the Horizon team who were always on hand to quell any administrative anxieties.

A special mention goes to Alex and Sam for all the coffees, chats and late-night games evenings.

Heartfelt thanks to my parents and family for not making me talk about the thesis (too much) and for always encouraging my curiosity.

And last but not least, thank you Jonathan, for everything else.

Abstract

Personal digital technologies impact the way both individuals and communities exist in and interact with their surroundings. As these technologies are increasingly interwoven into urban landscapes, they play an integral role in the articulation of personal and shared identities. This thesis investigates the particular ways in which geographies intersect with online text-based narratives and asks what can be understood about place in the contemporary city through social media texts.

An extensive review of existing literature from a number of disciplines suggested three salient attributes in the formulation of place-based identity in the contemporary city; experience, time and juxtaposition. These attributes were explored using seven tailored bodies of twitter text, ranging in size from 4,996 to 98,779,803 words. Using a combination of statistical analysis methods taken from Corpus Linguistics (CL) and a case-by-case inductive qualitative analysis of individual tweets, each corpus of tweet text was examined for the pre-identified attributes (experience, time and juxtaposition), as well as locational reference, common words (or features) and recurring themes.

This thesis develops existing research that posits place as a complex and ever-evolving concept, contingent on the overlaying of multiple voices and demonstrates that social media data can be effective in addressing questions about the relationships between people and the spaces they encounter in their digitally-mediated day-to-day lives. It articulates the key challenges facing researchers undertaking work in a burgeoning interdisciplinary field and presents a clear theoretical and methodological framework for developing studies in this context. Furthermore, it provides a body of evidence that clearly illustrates the utility of crowd-sourced and volunteered information in the uncovering of sense and understanding in fast-moving research contexts.

Contents

Chapter 1 Introduction 8
1.1 The research context 8
1.2 Research Question 10
1.3 Research Objectives 10
1.4 Outline of the thesis 11
1.5 Research Outputs 13
1.6 Research Gap and Contributions 14
1.7 Application in non-academic contexts 15
1.8 Interdisciplinarity and Methodological Innovation 16
1.9 In summary 16
Chapter 2 Literature Review 18
2.1 Introduction 18
2.2 Locational Referencing 18
2.3 Technology and geographic thought 25
2.4 “The city” as a contemporary urban space 45
2.5 Theorising Place 50
2.6 (Re) articulation of the research gap 65
2.7 Place in Contemporary Urban Contexts 66
Chapter 3 Methodology 69
3.1 Introduction 69
3.2 Identifying a suitable data source 69
3.3 Challenges and Opportunities specific to the context 72
3.4 Designing the Research 82
3.5 Processing and Querying the Data 91
3.6 Ethical and Methodological Challenges 98
3.7 Conclusion 107

Chapter 4 Locational References in Tweet Text 109
4.1 Introduction 109
4.2 Overview of the Approach 109
4.3 Data Collection 110
4.4 Data Analysis: Corpus Linguistics 115
4.5 Data Analysis: Thematic Analysis 123
4.6 Locational Reference 132
4.7 Place 134
4.8 Reflections 137
4.9 Planning Stage 2 140
Chapter 5 Geographic Metadata 142
5.1 Introduction 142
5.2 Overview of the Approach 142
5.3 Data Collection 143
5.4 Data Analysis: Corpus Linguistics 149
5.5 Locational Reference and Place 156
5.6 Hashtags 161
5.7 Reflections 165
5.8 Planning Stage 3 169
Chapter 6 Hashtagged Place names 170
6.1 Introduction 170
6.2 Overview of the Approach 170
6.3 Data Collection 171
6.4 Data Analysis: Corpus Linguistics 172
6.5 Reflections 195
Chapter 7 Place in Twitter Text 200
7.1 Introduction 200

7.2 Attributes of place most relevant to the context 201

7.3 Challenges Particular to the Use of Social Media Texts in Research 202

7.4 Twitter: A “suitable” data source?..... 212

7.5 Constructing a Coherent and Replicable Methodology 213

7.6 Building a Suitable Ethical Framework for Use in Interdisciplinary Research 215

7.7 Strategies by which Place is Incorporated into Social Media Texts 217

7.8 The Utility of Twitter Data in Informing Understandings of Place 222

7.9 Finding a way through the noise 241

Chapter 8 Conclusion 243

8.1 What Can Social Media Texts Contribute to Understandings of Place in Cities? The Cases of Nottingham and Newcastle 244

8.2 Disciplinary Reflections 246

8.3 The Interdisciplinary Study of Digital Technologies 248

8.4 Implications for Practice 249

8.5 Directions for Future Work 251

8.6 On Reflection 255

References 257

Appendix A: Metadata in a Single Tweet 280

Appendix B: List of Stop-words used in analysis 285

Appendix C: Collocation Tables 287

Appendix D: Primary and Secondary Themes identified in Stage one tweets 295

Appendix E: Code used in Stage three data gathering 299

Appendix F: Primary and Secondary themes identified in stage three tweets 303

Appendix G: Manually Extracted Locational Refereances from Stage Three 313

Chapter 1 Introduction

1.1 The research context

The rise in the availability and uptake of pervasive personal digital technologies has been matched by a rise in the volume of personal data and interpersonal information generated in digital contexts. The data streams generated by these ubiquitous technologies present both challenges and opportunities for researchers and commercial enterprises alike. On the one hand, they offer the chance to capture a plethora of information and data in many different forms, from audio and images, to text and associated metadata. Yet, on the other hand, the data is generated in uncertain conditions. In the most part, it is contributed either ambiently or intentionally by an unknown public, in relatively unknown contexts that do not require training and which do not insist on confidentiality or promote anonymity. The challenges of this "Big Data" (Mayer-Schoenberger and Cukier 2013, boyd and Crawford 2012, Gandomi and Haider 2015) are both practical and ideological. Though this thesis is does not extensively deal with the capitalised "Big Data", critiques of prevalent existing practical and ideological approaches to data from the digital realm are central to both the theoretical and methodological frameworks developed throughout.

The research question and objectives were developed with a view to extending existing connections in theory and method between several disciplines; namely the arts, (more-than) human geography[1], digital sociology, GIS and Corpus Linguistics[2]. The research sits not only at the cusp of these academic disciplines, but also transects the boundary between academic and corporate areas of interest. New forms of data, generated through increasingly varied technologies present interesting commercial opportunities for many industries. The spatial aspect of much of this data, which could be in the form of metadata generated

[1] More-than human or hybrid geography refers to a body of research that aims to extend the researcher's gaze beyond human subjectivities, it can include the study of animals as well as a variety of other non-human subjects (Whatmore 2002, Latour 2005)

[2] Corpus Linguistics refers to the study of language-in-use. See detailed explanation in section 3.5.7

from a location-enabled device or simply a place name written into an online narrative is particularly interesting to individuals and organisations specialising in geospatial products and technologies like maps and wayfinding applications (including the Ordnance Survey, who co-sponsored the work).

The interdisciplinary nature of this work reflects the varied expertise that has contributed to the project, from inception to write-up. The research was co-sponsored by The Ordnance Survey (OS), Britain's national mapping agency, and the Horizon Centre for Doctoral Training at Nottingham University. OS provided supervisory support; contributed to discussions throughout research development and provided comment and critique based on their commercial research interests and extensive shared knowledge in the field of spatial science. The arguments presented in this thesis reflect my background in Social Anthropology and my passion for intersectional research. This thesis engages with issues pertinent to multiple academic disciplines as well as technologists, mapping enthusiasts and commercial businesses considering the implication of new digital technologies for the future development and support of mapping products.

These broad and interdisciplinary research interests concern the practical, theoretical and social impacts of the development of new sensing and data-capturing/generating technologies. The scope of these shared interests is expressed in the following concerns:

- The ethics of using social media data for research and product development,
- Ensuring data completeness and validity in user-generated data
- The complexity of place as a social construct, and
- The relationship between locational reference and value-laden digital discourses

This thesis presents research that aims to balance the challenges and the opportunities of working in digitally-mediated realms; it does this

through the frame of place in social media texts, and explores the flows of spatial and experiential information through social media platforms.

1.2 Research Question

This research aims to articulate how social media texts can be used to enrich understandings of place and considers the role of personal digital technologies in the shaping of the narratives of the contemporary city. The thesis approaches this aim by answering the following question.

> **RQ**: What can social media texts contribute to understandings of place in cities?

This research question was developed into seven research objectives (section 1.3), which set out the framework of the research. These objectives outline the main steps taken throughout the research period, from the setting of the theoretical frame through the selection of methodological approaches and analytical techniques.

In meeting these seven objectives, this research builds on previous interdisciplinary work concerning communication and social interaction in the context of social media. It also makes more explicit geographic interrogations of spatial data and the significance of place in contemporary urban spaces. This approach provides a model for future interdisciplinary engagements with technologic and geographic theory. It uses methods and tools from various disciplines whilst encouraging responsible methodological innovation through the use of open-source online resources such as Twitter and Urban Dictionary[3].

1.3 Research Objectives

The aim of the thesis was to explore whether social media texts could be mined for information about people's experiences of and associations with urban spaces, in order to determine whether this information might provide a new way to study the notion of place in the

[3] Urban Dictionary is a crowdsourced online dictionary, where words and their definitions are contributed by users. Users do not need to sign up or register in order to add content. (www.urbandictionary.com). Its use in this research is explained from page 160.

city. These aims were addressed though the formulation of seven objectives:

1. Identify attributes of place most salient to the examination of contemporary urban contexts through social media texts
2. Identify a suitable data source (social media platform)
3. Specify the challenges particular to research using social media texts
4. Develop a coherent and replicable methodology for the examination of place in social media texts
5. Outline a framework for addressing the ethical issues associated with the use of social media data in research
6. Identify strategies by which Place is incorporated into social media texts
7. Evaluate the utility of social media data in informing understandings of place

Objective one seeks to make clear the theoretical context in which this research is embedded. In working toward objectives two through five the thesis sets out the methodological challenges of dealing with secondary data specific to social media and argues that data must be treated sympathetic to its specific technological and editorial limits. This includes an emphasis on reflective practice in research and a willingness to seek creative solutions to research problems. The final two objectives, six and seven, address the data itself, through a focus on the strategies by which place is incorporated into social media texts and what this information can reveal about particular sites.

1.4 Outline of the thesis

This thesis begins with an in depth review of relevant existing academic literature and other commercial and artistic work that informed the theoretical and methodological direction of this project. This research is interdisciplinary, and incorporates theories and methods from several disciplines; synthesising a multitude of sources. The literature review addresses the variance of locational referencing, the connection

between technology and geographic thought, the city as a construct and place as a multi-dimensional attribute.

Table 1.1: Thesis Structure

<table>
<tr><td colspan="3">Chapter One: Introduction</td></tr>
<tr><td colspan="3">Chapter Two: Literature Review</td></tr>
<tr><td colspan="3">Chapter Three: Methodology</td></tr>
<tr><td>Chapter Four: Place Names</td><td>Chapter Five: Geographic Metadata</td><td>Chapter Six: Hashtagged place names</td></tr>
<tr><td>Filtered by key words nottingham OR notts</td><td>Filtered by geographic location</td><td>Filtered by use of tagged place names #nottingham OR #newcastle</td></tr>
<tr><td>10,510 word sample</td><td>107,589,004 word sample</td><td>29,012 word sample</td></tr>
<tr><td colspan="3">Chapter 7: Discussion
The research Objectives</td></tr>
<tr><td colspan="3">Chapter 8: Conclusions
Answering the Research Question</td></tr>
</table>

The methodology, including the rationale behind major research decisions taken and an explanation of the practical and ethical challenges specific to the context, is outlined in chapter three. This chapter also describes the process of data collection from an assessment of possible sources to ethical review and the formulation of an overarching design for the research.

The body of the thesis is split between three chapters (four, five and six) which reflects the design of the project. The research incorporated three sampling methods, each of which is reviewed in a separate chapter. These central chapters describe in detail the research work undertaken and the data collected at each stage. Chapter four describes the collection and analysis of data gathered using a key-word search of

Twitter for tweets containing either of the words “notts” or “nottingham”. Chapter five describes how sampling and analysis was altered in order to interrogate tweets on a much larger scale. This was carried out with the assistance of computational linguistics and the CL software SketchEngine[4]. Chapter six extends the work of chapters four and five, by combining approaches used at each previous stage in order to focus in greater detail on the mention of location and the sequencing of locational reference in tweet texts. At this stage, tweets were filtered according to the inclusion of pre-defined hashtags (#nottingham and #newcastle) and then analysed individually for the inclusion of spatial reference, for theme and the relationship between these two features.

Chapter seven synthesizes the insights uncovered at each of the stages in a discussion of the main results and fundamental insights, making reference to the literature reviewed in chapter two, as well as exploring how the work compliments or contrasts with previous empirical and theoretical work.

Chapter eight reflects on the main findings of the research work and summarises the outcomes in order to clearly answer the main research question. This final chapter ends with a discussion of potential future academic effort that would extend or complement the work carried out as part of this project. This also includes ideas which might be more relevant to commercial and product-based enquiry.

1.5 Research Outputs

Insights from the first sampling of data, described in chapter four were presented at The British Association of Applied Linguistics annual conference (BAAL) 2016 in Cambridge, UK.

Title of talk: *“Understanding Places using New Media: The use of hashtags in relation to cultural tropes in cities”*

Insights from the analysis of data collected in the second sampling of data, which are described in chapter five, were presented at the 6th

[4] See page 150

International Conference on Language and Literary Studies in Belgrade, Serbia in 2017.

Title of talk: "*The Hashtag: How technology is transforming our language, a corpus-based approach"*

1.6 Research Gap and Contributions

This thesis discusses the value and limitations of the data contributed by the authors of social media texts in addressing questions about what place is, and the role place plays in the mediation of daily experience through social media texts. Existing work by academics, commercial businesses and artists demonstrates an ongoing interest in the significance of place in contemporary urban contexts and evidences the need to untangle the complex relationships between personal technologies, individuals and the spaces they inhabit. In light of these pressures to use and makes sense of these increasingly prevalent data streams there is also an ever-growing need to engage critically with the rhetoric that surrounds them and to gather, process and store this personal data appropriately and transparently[5].

This thesis presents text data sourced from Twitter in support of the argument that social media data can provide valuable insights into the ways that individuals make sense of the spaces around them, and demonstrates a number of ways in which they do this through the authoring of digital texts. Furthermore, it illustrates the worth of data that is generated independently of targeted research (ambient data) and its application in both theoretical and practical pursuits.

The contributions to knowledge made by this thesis and the research it describes are primarily theoretical and methodological. The research has considered in detail the ethical and practical challenges of undertaking research involving personal data in digitally mediated urban

[5] Continued action on the part of governments, charitable organisations and lobbying groups clearly indicates a drive for responsible and transparent use of data (Privacy International, The Information Commissioners Office and The Electronic Frontier Foundation, for example).

spaces. Using examples from specially compiled corpora[6] this thesis extends pre-existing bodies of work across the core disciplines of geography, applied linguistics and the digital humanities, which together explore the role of technology in contemporary urban life. It encourages a flexible, interdisciplinary and reflexive approach to research. In an age where personal data is easy to come by in large volumes, a reflexive approach to research is imperative. Reflexivity ensures that research engages critically with the issues at hand and considers the responsibilities of researchers and service providers (like Twitter) in the perpetuation of dominant assumptions.

This thesis insists upon the recognition of the technological contexts in which articulations of place, including cartographic representations and located social media texts, are embedded. The thesis does not provide a full historical account of maps or mapping technologies and the analysis is by no means a complete description of the historical or cultural environments that have given rise to certain trends in representation, but it does explore the theoretical and ideological development of "place" in a technological sense, emphasising the symbiosis of technology and ways of seeing and processing the world.

1.7 Application in non-academic contexts

Through the identification and extraction of locational reference and the exploration of associated information in social media texts, this thesis provides examples of the nuanced ways in which people use location and locational reference in the articulation of both individual and group place-based identity. It also highlights the ways in which contemporary urban spaces can be sites of cultural, historical and personal significance, even in the mundanity of everyday exchange and communication online. These understandings might help organisations developing digital mapping products to provide a more intuitive and sensitive service to their customers, either through more informed

[6] Compiled corpora are bodies of text that have been collected and processed in preparation for use in Corpus Linguistic analysis. See section 3.5.7 for corpus linguistics

engagement with spatial analysis or through the provision of more flexible services, tailored to reflect the connections of individuals. This might include tailoring visualisation services to reflect the social attachments of a community or individual, by displaying certain features or landmarks more prominently or, alternatively, it might take the form of a visualisation that reflects identity changes over time as well as contrasting narratives of a single place.

1.8 Interdisciplinarity and Methodological Innovation

The methods used in this research are innovative in the sense that they cross several disciplines and challenge arguments made in the support of the use of data exclusively from verifiable or authoritative sources. Though this work is not the first to explore the application of social media data for purposes beyond communication or broadcasting (this existing body of work is discussed in the methodology chapter, in section 3.3) the interdisciplinary breadth and melding of ideas and methods from a variety of academic realms are what makes this research novel. This project combines theoretical engagements from sociology and geography with methods taken from digital anthropology and applied linguistics and has produced a piece of research that is interdisciplinary.

The assessment of social media text data for use in contexts outside social media highlights the commercial and academic potential of ambient data. Ambient data, that is, data generated in an environment external to the final application context is relatively new, and is linked to the development of ubiquitous technologies. By encouraging a flexible and creative approach to the use of ambient data by both academics and organisational practitioners, this thesis hopes to contribute to a research environment that encourages sympathetic and creative use of personal data.

1.9 In summary

The textual, spatial and experiential data in this project has been be used to unpack notions of place and to reveal strategies in the

articulation of place-based identity through Twitter in the two English cities of Nottingham and Newcastle. This research suggests ways in which attachement to location can be effectively explored through the use of ambient data streams like those from social media. Furthermore, it demonstrates the continued relevance of place- and space-based scholarship in an ever-evolving technological landscape, where claims to ubiquity, freedom from locational contraints and democratised access to information are widespread (see Auge 1996, Boellstorff 2008, Federman 2005, Lupton 2012 and Wilson & Peterson 2002).

Chapter 2 Literature Review

2.1 Introduction

This thesis aims to describe how location is expressed in social media text and asks how ambient data might be applied to challenges of spatial theory, and how this data might help us better understand and represent the relationships that humans form with the locations they encounter. The literature review that follows outlines the technological and academic context in which this thesis is situated.

The relevant literature has been divided into five sections, each one builds on the last, leading finally to an articulation of the central themes explored throughout this thesis. The literature review begins with a description of several modes of spatial reference before considering how spatial reference has been impacted by the development of various technologies. Covering a range of technologies, beginning with paper maps and ending with an emphasis on personal digital technologies, this section highlights the fundamental interconnectedness of geographic thought and available measuring and visualisation equipment. The literature review then moves on to explore “the city” as a contemporary urban space, highlighting the particular qualities that differentiate these sites from non-urban/non-city sites. The third section provides a summary of the theories and models of place in geography, sociology, anthropology and in art and design that have been brought to bear in this work. The fifth and final section of the literature examines and outlines the key attributes of place that emerge from the review.

2.2 Locational Referencing

There are many, many different ways of referencing location. The UK “All-Fields Postcode Directory” for instance, reviewed by Simpson and Yu (Simpson and Yu 2003), lists 25 separate ways of describing a geography. This section explores these ways of referencing by breaking locational referencing into three overarching categories. These overarching categories are the use of geo-coordinates, the use of text and the use of images.

2.2.1 Coordinates

Geospatial coordinates have long been used as a system of spatial reference. Providing a globally understandable, numeric and standardised system of locational reference, geo-coordinate data is frequently used in many fields, from orienteering (Norouzi 2013) to mountaineering and in military exercises (Der Derien 2003). Spatial referencing by coordinates is usually performed through the combination of two measurements that are two-dimensional (Laurini and Thompson 1992). Some references using this system also specify the height or depth, and even sometimes permit the inclusion of time, however, this would require the use of an additional coordinate system, resulting in a "compound coordinate reference system" (Open Geospatial Consortium 2010). This Cartesian frame, which uses a "horizontal frame", of referencing space is the system most commonly encountered in everyday activities such as navigation (Norouzi 2013).

2.2.2 Text

Locational reference is often at least partly textual. Text is meant here as the use of written language to describe and designate location. This form of geographic reference is also extremely common and can take a variety of forms, from "geopolitical entities" like countries, to "administrative division", "populated places" and "postal codes" (Leidner and Lieberman 2011, 6). There is also some variation within, and overlap between these categories. Textual variations of Locational Reference can be vague, through the use of unspecified terms (area) or prepositions (on, in), or specific (the use of official place names) and vernacular (colloquial) or official (which have been sanctioned in some way by authorities or official institutions).

2.2.2.1 Place names

Locational reference is most commonly and easily encountered in the use of written place names, whether these are road names on street signs, house or business names, place names, both official and vernacular are a prevalent form of text-based locational reference.

2.2.2.2 Addresses

Postal addresses are probably one of the most familiar forms of standardised locational reference in the United Kingdom and although standardised, they are by no means universal and are commonly accused of being insufficient in the demarcation of spaces (Davis Jr and Fonseca 2007). Large parts of the world do not have formal, institutionalised postal address databases. Even in areas where addressing conventions are in place, their spatial reference is limited to buildings or structures, so the utility of these systems can be more difficult to justify in the context of displaced or transient individuals or indeed where a location does not correspond to the built environment. One initiative that recognises this disparity and aims to overcome the inadequacies of formalised address-based locational reference is "What3Words" (What3Words 2018).

2.2.2.3 Random Assignment of Words

"What3Words" has disrupted existing systems of spatial referencing for business and logistics by creating a global grid of three million equally-sized squares each measuring 3m x 3m, the minds behind What3Words aim to provide an addressing system that is internationally applicable and provided delivery solutions to individuals and businesses where no formalised postal or building address existed previously. In the What3Words system of locational referencing, the entire surface of the earth has been segmented into easily attributable, unique "addresses", where each grid square is identified through the combination of 3 English words. This approach means that previously un-addressed locations can be easily identified, clearly distinguished from other locations and found by service providers. Whilst this initiative certainly solves some issue of spatial invisibility that other addressing systems (such as postal codes) might perpetuate, the system is not perfect, and has faced criticism for not sharing data; raising concerns about the commercial control of essential data increasingly connected to potentially life-saving services, including front-line health services. (Dean 2018, NHS Digital 2018, Eden 2019)

2.2.2.4 Social Media Narratives

This thesis looks particularly at the written texts of social media platforms, and has been informed by a body of work that explores language, both in terms of its use and social impact as well as its adaptation in conjunction with technological change. Key themes in this literature are the emergence of narrative forms in tandem with new technologies (Page 2010) (Titton 2015) (Chou, et al. 2011) (Walz 2016) the use of computational methods to interrogate large bodies of text (Baker 2013) (Knight, Adolphs and Carter 2014) (Bruns and Burgess 2011) and the investigation of novel methods of reference (see discussion of hashtags in section 3.3.2). The literature covered here predominantly focusses on the language of social media and example studies from Corpus Linguistics. The methods of Corpus Linguistics are explored in greater detail in section 3.5.7. The significance of this body of work to the development of the research was its focus on language in use.

As the focus of this thesis, the study and analysis of texts was an influential body of work in the development of both the theories and the methods implemented in its creation. The study of informal texts, in particular, is key. The varying forms of social media texts, from blogs about fashion (Titton 2015) and expatriate identity (Walz 2016), to video logs documenting journeys through serious illness (Chou, et al. 2011) and the tracking of activist movements online (Clark 2016) indicate that online narratives take many forms and are not necessarily inherently different from other narrative forms like the autobiography or oral histories.

According to Ruth Page, the differentiating feature of social media text compared to more conventional bodies of text is that these texts are rarely “complete” or finished (Page 2010). The texts of social media accumulate over time but do not have a clearly defined structure nor necessarily an endpoint. The narratives on these platforms are constantly growing and evolving, with each entry or post cumulatively adding to the existing texts. This narrative evolution is explored by Page

in her work on the narrative form of Facebook status updates, “the stories reported in updates fall outside the literary, conversational, and electronic narrative canon, but are a rich data source for re-examining key concepts in narrative theory” (Page 2010, 423). Social media texts, as imagined by Page, are ever-evolving narratives that are iteratively created by several actors over time. These small but continual updates “typically focus on the minutiae of everyday events” (Page 2010, 426), this makes social media texts appealing sources for the investigation of relationships with place, where time and the overlaying of differing perspectives are important considerations. This detailing of day to day events impacts both the reading and the telling of time in digitally mediated communication. Time is at the centre of narrative; but the experience of time, for both the writers and the readers of online narratives, is warped by the set-up of digital technologies and their interfaces. As demonstrated by Page in her examination of Facebook status updates, the timescapes of social media texts can unfold in unconventional ways. In the status updates on Facebook, a story is revealed in reverse chronological order, with the most recent update usually read first. Furthermore, Page argues that the self-image gleaned from status updates does not emerge fully-formed but evolves over time, and is more “analogous to one continuously being created using a pointillist technique” (Page 2010, 440). Reissman describes the study of narrative as “inherently interdisciplinary” (Riessman 1993, 1), and focusses on how individuals “impose order on the flow of experience to make sense of events and actions in their lives” (Riessman 1993, 2). Narrative in the context of social media can be troublesome, in so far as it is more nebulous and harder to delineate than narrative in offline contexts. Existing research into narratives on social media demonstrated that it shares many qualities with “offline” narratives, depending on the specific medium and author.

In a similar way to non-digital stories, stories shared online belong not only to those who write them, or those who read them, but also to the locations in which they have been written. The Geo-tagging tools of

digital technologies facilitate this association. Page suggests that a narrative can only be determined when these individual status updates are considered to be part of a wider body of reportage – “an ongoing archive which unfolds over time” (Page 2010, 434). The archive of a person or place is also not limited to a single platform but extends across many different mediums, so the “ecologies” mentioned by Scollon and Scollon (Scollon and Scollon 2003) rely on the interplay of various technologies, people and locations. These ecologies are not explored here, but their articulation is key to understanding the wider technological and social context both of this research and that of the individuals who use social media platforms. They are picked up again in the discussion of creativity in section 7.8.6.

2.2.3 Image

In addition to the use of numbers (coordinates) and text (place names and addresses), location is often represented through visual means. The depiction of location through image is commonplace. Image is different from visualisation, in the sense that an image can be isolated from text or any contextual information, though it usually appears in tandem with these things. Image has been divided into three categories here; Photographs, Artistic Representations and Maps, which are picked up and used as a linking node to illustrate the interconnection of technology and ways of thinking in section 2.3.

2.2.3.1 Photographs

Photographs are captured images rather than “created”, which sets them slightly apart from artistic representations and visualisations. However, they are still curated images in the sense that the photographer has chosen a particular frame, subject matter or film or print type. Photographs can be useful snapshots but they often require annotation in order to be easily read and interpreted by viewers. The addition of this supplementary information, in the form of a caption or note scribbled on the back, is where the line between photographs, visualisations and artistic representations is at its most indistinct.

2.2.3.2 Artistic representations

Location is a common feature of artistic representations, whether this is the conscious project of the artist or a consequence of the conditions of making that piece of art. The Tate's "Art Maps" project, which aimed to add locational information to its indexes of around 68,000 pieces of British and international art (Tate 2012), is a striking example of the ways in which locational reference is closely linked to projects of artistic representation, whatever the subject or medium.

These three ways of referencing location are explored a little further in the next section within the context of technologies, where maps, in particular, have been used as a focus of the narrative. The map is a useful point from which to examine the relationship between geographic thinking and technology as it is a form familiar to both lay and expert geographers.

2.2.3.3 Maps

Maps are commonly used to present spatial information and often combine both texts (in the form of place names) and images. They serve here as an illustration of the complex array of relationships that individuals and groups develop with (and in response to) their physical and emotional environments. Maps are often cast as objective representations of the world and as accurate depictions of the spatial relationships between objects and sites. However, their form is extremely diverse. In recognition of the habit of maps to reflect the specific theoretical and spiritual beliefs of its creators, some cartographic theorists have gone so far as to insist upon the inclusion of particular features in order for a visualisation to earn the accolade of "map". In pursuit of some measurable standards Eugene Van Cleef wrote (in 1948), declared that visualisations "lacking scale or coordinates… hardly reflect the fundamental basis of a map: hence they are not entitled to the designation *map*" (Van Cleef 1948, 521). Van Cleef's insistence on the inclusion of coordinated and scale are great example of a drive to standardise the visualisation of place in line with western scientific convictions about what place was and how it should

be demarcated. More recent scholarship demonstrates a general change in attitude, with an increasing number of writers who recognise the fault in an insistence that only western and scientifically standardised forms of mapping are worthy of the title (Harley 1989) (Wood 2012). In any case, the variability of the form of a map is widely accepted and is explored further in section 2.3. Maps have been used as a tool to illustrate the relationship between the theorisation of space and the development of technologies in section 2.3.

2.3 Technology and geographic thought

Locational reference and the conceptualisation of place and space are closely tied to technology. Understanding the symbiotic development of technology and geographic thought more broadly is important to this research as it lays the groundwork for a project aiming to contribute to understandings of place in a digital context, with a particular emphasis on social media technologies. The interconnected nature of available technologies and ways of thinking is not limited to the development of ubiquitous personal digital technologies and examples of this relationship have been explored here through academic literature as well as in lay depictions of location, space and place. As John Berger suggests, “the way we see things is affected by what we know or what we believe” (Berger 1972, 8). Understanding the frame of our own individual and cultural perspectives is important, and an absolutely essential aspect of research with a qualitative element.

Many scholarly engagements with technology and its impact on human sociality focus on the ability of new electronic machines or software to reduce physical or cognitive workload (de Jong 2010)and (C. Patten, et al. 2004). As digital technology started to infiltrate the intricacies of everyday life, its impacts have become less defined by their physical impact and new disciplines have begun to engage with the complex and multi-faceted relationships between humans and their machines. This section explores the impact of several key social and technological changes on geographic thought as well as considering alterations in

views of the self that arrived with Web 2.0[7], personal computing, the increase of user-generated online contentand advances in sensing technologies. There are many ways in which the interconnectedness of technology and cognition may be demonstrated; the following section uses maps and the visualisation of space and place to explore this relationship. Maps have been used as a grounding artefact in the discussion that follows, not only do these objects help to familiarise us with our geographies, but their form and content often coincide with available technologies and data types.

2.3.1 Maps

The map is an abundant and highly variable object. A map is not a copy of the world as it is, but a distillation of features deemed more or less important by a map-maker. They are cultural objects, reflecting the agendas and conceptual universes of the people who make them, and as such, offer valuable insights into the ways in which technology mediates and connects locational reference with value-laden narratives about time, identity, experience. Daniel Dorling and David Fairbairn emphasise the variability of the mapping form, and the form's dependence on dominant cultural activities and paradigms, "Just as societies and civilizations vary enormously in their essential characteristics, so their map representations vary – in purpose, scale, content and conceptions of accuracy" (Dorling and Fairbairn 1997, 3). Since a single narrative does not link all these modes and styles of mapping and spatial representation, the discussion uses a few notable examples throughout the chapter to illustrate the variability of the form, and its development in relation to geographic thinking.

2.3.2 Minard's Map

One of the most iconic maps in history, Charles Minard's "Carte figurative des pertes successives en hommes de l'armee qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules" and its accompanying visualisation "Carte figurative des pertes successives en

[7] Web 2.0 refers to the inception of the user-generated web, where content in increasingly contributed by individuals distributed across the network. The term was coined by Darcy DiNucci in 1999 (DiNucci 1999)

hommes e l'armee Francaise dans la campagne de Russie 1812-1813 (Figure 2.1) is map of a French military campaign in Russia, led by Napoleon between 1812 and 1813 which combines geographic data and other quantitative reporting. The visualisation combines temperature, with distance, time and the size of the French army (in numbers of men), it is a prominent example of the visualisation of several data streams at once. Through the juxtaposition of the data of two military campaigns in Europe, Minard's map extends simple geospatial representation into something that is now viewed as an "exemplar(s) of imperial overreach" (Rosenberg 2015, 54). Minard's map documents not only a geographic landscape but also a human and historical one; it shows the massive human cost of the military campaign. This map is a great example of how the "argument" of a map can be determined by which information is included as well as which information is left out. In Minard's case, the focus is on the French army, with no reference to the losses suffered by Russia over the same period. Maps are not simple representations of fact or truth but complex arrangements of data guided by purpose and produced with an audience in mind. Minard's map is a reminder of the narratives that can emerge from the distillation of data derived from a number of different technologies. It serves as a reminder that geographic story-telling is inherently restricted by available data types and measuring technologies and occurs with particular audiences in mind.

The story-telling power of geospatial visualisations is well recognised; Muelenhaus summarises the argument succinctly, stating "All maps are a form of geo-communication. They are all designed to communicate something about our spatial environment to a map reader or user" (Muehlenhaus 2014, 12). According to Muelenhaus, a map or visualisation cannot be both anecdotal and effective at communicating, "well-designed, data-rich maps that provide no argument or do not have a clear communicative goal are like cute cat photographs on Facebook: nice to look at but worthless" (Muehlenhaus 2014, 13). In Muelenhaus's view, a map without a clear purpose is a useless map. As a

consequence of the connection between utility, design and purpose, maps are not objective representations of the world; “The first thing a map designer needs to dispel is the notion that maps can be objective visualizations of data” (Muehlenhaus 2014, 14). What can be learned then, from a representation like Minard’s? This figurative depiction exemplifies the dual drive to make sense of the human and spatial world, and to communicate our understandings to others. It is an excellent example of the overlaying of data streams to in the production of a strong narrative that weaves together details about time, distance, demographics and temperature, making best use of the sensing technologies available at the time.

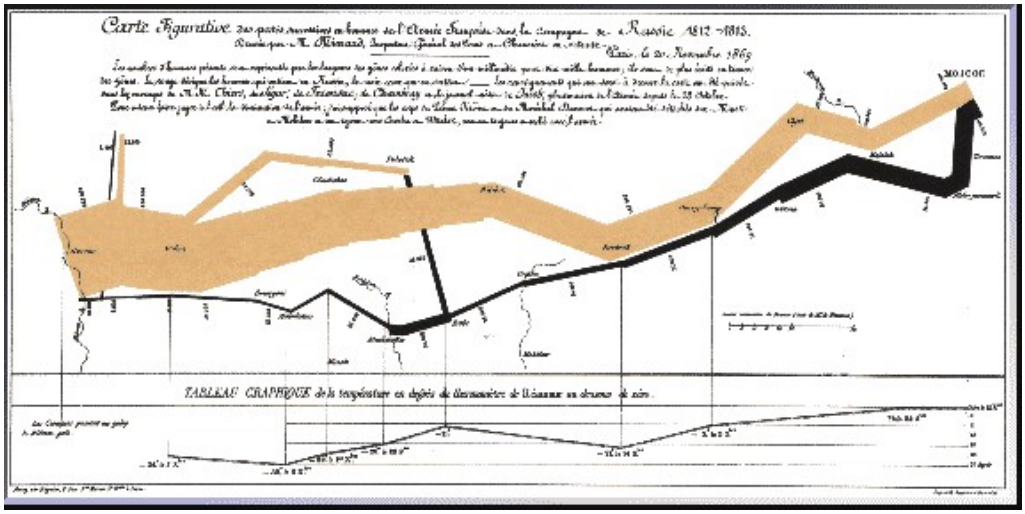


Figure 2.1: Charles Minard's “Carte Figurative”

2.3.3 Cartefacts

The balance between objectivity and subjectivity in geospatial visualisation is a constant concern in geographic literature and can be traced to the development of maps for purposes of philosophical or religious reasoning, as well as for imperial navigation and expansion. Many early representations of space were concerned with navigation, the demarcation of property (including land) and the carving up of the universe in spiritual/natural or human domains (Black 2000). As a result, the practice of map-making is no stranger to controversy and critique. There is a large body of work highlighting the inadequacies and agendas of map-makers as well as the inaccuracies and insufficiencies

of the visualisations they produce (Wood 2012) (Edney 2015) and (Monmonier 1996).

Denis Wood talks about the “mythic” nature of maps, stressing their performative nature. He calls the encoding of links and routes on maps “map-made…cartefacts” (Wood 2012, 290). According to Wood, all maps result from some agenda, they are tools in the sense that they do things – “maps work by serving interests” (Wood 1992, 1) and these “interests are embodied in the map as absences and presences” (Wood 1992, 1). As Wood argues through his analysis of the maps drawn by children, even the most rudimentary maps showcase the subjectivities of the individuals who produce them. They show how people position themselves not only in their physical surroundings but also how they situate themselves in relation to other people, other forces and the universe in general – the images they produce “exceed the raw experience” (Wood 1992, 6) of the world, portraying instead some part-raw and part-imagined universe. Dorling also challenged the utility and relevance of maps used in the creation and curation of shared understandings through his analysis of the maps used historically in social science. He insists that “the maps that are currently drawn in its [social science’s] name, apart from often being bad examples of physical geography’s cartography…they make concentrations appear where they are not and dissolve existing patterns” (Dorling 2012, xlvii). In light of these criticisms of static representations of space, social media technologies, particularly those which encourage frequent and short updates (Twitter), are an increasingly important resource in the quest to understand how humans create and make sense of their physical and social worlds. This close relationship between technology, experience and identity is a central concern of this thesis, which aims to demonstrate the importance of keeping this intersectionality in mind.

This subjectivity-driven creation of spatial representations means that each image encapsulates more than spatial data. The performative and expressive nature of maps increases with their accessibility. As sensing

and visualisation technologies develop (technologies like LiDAR[8] or OpenStreet Map), so does the potential for a non-expert public to engage with them.

2.3.4 "Cyborg Assemblage": Human/digital relationships

Lives in the city are increasingly interwoven with technologies. The rapid evolution of electronic technologies means that the tools that are used by both individuals and groups to communicate with each other are getting smaller, easier to transport, easier to share and easier to integrate into day-to-day practices. Digital and electronic technologies, in particular, are easier to integrate into everyday routines, and as such can be melded seamlessly into existing personal networks of devices and even, in some cases, directly attached to the body. This interweaving of human and electronic and digital is popularly known as the "cyborg" or, within digital social sciences as a "cyborg assemblage" (Gandy 2005, Lupton and Haraway 2015 and Lupton 2005). The impact of technology on human sociality and social practice has been heavily studied in a number of different disciplines. For reasons of scope, the analysis of the development of academic theory and practice surrounding technology's impact on human sociality touches only very briefly on the history of these engagements. The coverage here is not absolute, but the discussion lends context to and demonstrates how issues of human sociality and technology and urban spaces are not limited to particular disciplines and can be explored in a number of ways. It is important to consider these multidisciplinary engagements because their multivariate approaches and analyses have been useful in framing the research undertaken for this thesis. Matthew Gandy (2005) summarises the intersection of this "cyborg" discourse as a facilitator for the exploration of the "different facets of the relationship between the city, the body and the human subject" (Gandy 2005, 28).

The use of technology in the study and understanding of human society is not a new idea, though the pace of its utilization as a research and

[8] Light Detection and Ranging, a method of remote sensing which measures distance to the surface of an object using lasers (lidar-uk.com)

thinking tool has increased in recent years, particularly amongst artists and social scientists (Nold n.d.); (Ramirez 2014); (Balesco Rogers and New n.d.). Christian Nold is one such artist and designer with an interest in tracking technologies, personal data and participatory mapping. Nold specialises in the use of biometric data, such as heart rate and Galvanic Skin Response, in unpacking the complexities of how people experience the spaces they pass through and addressing the challenges of collecting intimate personal data related to everyday experience. Nold collates the data he collects and produces representations that illustrate the spatial and emotional extent of his daily experience place (Nold n.d.). Similar projects include *Wander Anywhere* and the *Daily Data* projects undertaken by artists Daniel Balesco Rogers and his collaborator Sophia New (Balesco Rogers and New n.d.), which use geo-located data to create visualisations of the participant's daily movements through urban spaces. New embarked on her tracking project in order to understand the ways in which having a small child had impacted the way she moved around in the city. The projects make clear the traces that engagement with personal digital technologies leaves behind; they encourage the consideration of individual personal data streams, which can "make visible the geographic patterns to the way our lives have come to be socially organised" (Dorling 2012).

Plainly, technology shapes perceptions of the world as well as the conceptualisation of our place in it. In a digitally connected urban environment, the limits of the individual are increasingly blended with the technological environment, resulting in a hybrid actuality where neither human nor technology can be separated from the other. Debra Lupton argues that "we are more cyborgs than ever before" (Lupton and Haraway 2015, 165), we are in many ways at one with the technology we come into contact with every day. Smartphones and their associated self-tracking technologies, such as smart watches, and fitness trackers like the Fitbit in particular, embody this close connection between humans and technology. Their use demonstrates a deeply reflexive

relationship between these things, where identities are continually negotiated through the feedback and information supplied by those technologies (R. Jones 2013). The discussions of Jones (2013), Boellstorff (2008) and Lupton (2012) all highlight the inseparability of the human and the technological and demonstrate the importance of multidisciplinary research work. The concept of the cyborg, or the "digital cyborg assemblage" as Lupton and Haraway (2015) prefer, as a hybrid embodiment is one of the essential features of the landscape in which this research has been conducted. This thesis explores how individuals connect to city spaces through the digital technologies available to them; the "cyborg assemblage" is the starting point from which these issues are explored. The interconnectedness of humans and technologies has been extensively studied by social scientists. Bauman and Lyon (Lyon 2010), for example, state that the increased availability to personal technologies has caused a shift in the social expectations of the individual, where a preoccupation with self-tracking and daily quantification is viewed as a necessary evil in the pursuit of a connected and social life.

2.3.5 Digitally Mediated Selves and Modernity

The technology/human hybridities of the "cyborg assemblage" extend discussions in the social sciences about the extensive and very personal impacts of global processes on ways of being and ways of thinking. As Fiss and Hirsch point out, in matters of identity and sense-making (collective or otherwise), the structural conditions in which an individual is bound are often neglected (Fiss and Hirsch 2005) but an understanding of these conditions is absolutely essential to gaining insight into why indidivduals behave in particular ways. The key idea here is that the world "does not come to us in raw form, but that we actively construct it" (Fiss and Hirsch 2005). The world is continually being re-shaped through our interaction with it and our telling of it. Digital technologies, like the social media platform Twitter have changed the way people are "in the world" (Heidegger 1978). They alter not only the way information is aggregated but also how it is consumed,

interpreted and passed on. These techno-social adaptations are particularly important in multi-modal contexts, where individual users are allowed to post images and text, or text and sound, or some other combination of common attributes. While these multiple modes of expression endow the individual with greater freedom to express themselves, they are simultaneously limited by the knowledge of their audience, and must balance expression with exposure in an ever-expanding digitally-mediated life.

This tension between global forces and individual capabilities is indicative of Modernity. Modernity has been characterised as a condition of extremes, where "the transmutations introduced by modern institutions interlace in a direct way [9]with individual life and therefore with the self" (Giddens 1991, 1). In *Liquid Surveillance*, Zygmunt Bauman and David Lyon assert that modernity can be understood through the acknowledgement that our day to day lives have been infiltrated by technologies and that the "condition of being watched" has been "reclassified from a menace to a temptation" (Bauman and Lyon 2010: 23). The development of contemporary consumer societies, where technologies are both easily available and highly desirable has eroded the boundaries separating man from technology. Under such conditions, the acceptance of the condition of being watched (which began, in theory, with Jeremy Bentham's Panopticon) has co-evolved with commercialised personal tracking technologies, like sleep trackers, wearable fitness trackers, smart watches and smartphones and other technologies which enable and encourage the documentation of day-to-day minutae, like Twitter and other social media technologies.

These theories have clear parallels with Lupton's comments on the "digital cyborg assemblage" and Jones's discussion of entextualisation (R. Jones 2013), which both argue that tracking technologies are essential tools in the pursuit of personal and shared meaning. Jones's

[9] Bentham's Panopticon was a form of surveillance based on the idea that prison buildings should be designed so that all inmates were on view at all times from a single point. Occupants could not see their observer, so were never sure who was watching (Bentham 2010)

discussion of entextualisation is particularly relevant to the examination of social media narratives as it focusses on the integration of the information generated through personal technologies into stories people tell to themselves and to others about themselves. The texts (including data reports from trackers like the FitBit) generated by personal technologies become integral to the performance and negotiation of identity. Technologies have proliferated the ways in which people with access to them can ”reassemble” and share their experiences (R. Jones 2011).

Similarly, Bauman’s focus on the everyday effects of digital technologies and his emphasis on the encroachment of digital technologies into ways of thinking and being is closely aligned to the arguments of Lupton, Gandy and Jones. Understanding the ways that technology impacts human life in the everyday provides a useful platform from which to approach the question of place within these emergent digital contexts. Social media technologies form an important part of this assemblage, allowing individuals to extend themselves digitally and to connect with other individuals through digital means. The study of technology in urban settings cannot be usefully separated from the study of human subjectivities. The impacts of technologies on human sociality extend beyond primarily individualistic questioning of self and identity; they extend to interpersonal and shared meanings and practices.

2.3.6 “New Media Collectives”

New sensing technologies are increasingly impacting the social and informational environments encountered by the individual as well as influencing the ways in which individuals connect and identify with others. The availability of technologies influences the range of information that people have access to as well as the way in which they choose to share that information (as Minard’s Map demonstrated). The impact of Web 2.0 on communicative practices has been tackled by a number of theorists, including Wilson and Peterson (2002), who argue that “Information and communication technologies based on the Internet

have enabled the emergence of new sorts of communities and communicative practices" (Wilson and Peterson 2002: 449). They call these evolving communities and communicative practices "new media collectives" (Wilson and Peterson: 449). New Media Collectives are groups communicating online whose membership can vary widely from very small (a few individuals) to very large (in the multiple millions). The individual members of these collectives can be linked by interest, religion, heritage, political affiliation but are not restricted geographically (though they must be able to connect to the internet). Digital technologies that are "always-on" and "always-on-you" (Turkle 2008) provide the technological infrastructure that foster these collectives; social media platforms, like Twitter, are one example of these technologies. They facilitate the connection and collaboration of geographically disparate indivduals. Once connected, these individuals may act together to achieve a common goal, or create texts that demonstrate a shared understanding or experience of the world.

2.3.7 Communities, ecologies and networks

The notion of "new media collectives" shares similarities with the idea of "communities of practice" that was first developed by Jean Lave and Etienne Wenger (Lave and Wenger 1991). Communities of practice, according to Lave and Wenger, are groups whose membership consists of individuals sharing common characteristics, through interaction with each other (mutual engagement), through a shared goal (joint enterprise) and who will foster, over time, a pool of shared resources, which may include common language, shared tools and other skills or understandings. These ideas have been useful in developing a critical view of the way that both language and digital technologies can be used in the construction of place-based identities.

The emphasis on the collaborative construction of meaning and the connection of individuals in pursuit of a common goal or as a result of shared practices aligns both "communities of practice" and "new media collectives" with discussions of "the crowd" in digital contexts. By sharing and broadcasting information through social media, individuals

are able to relinquish the limits of their physical context and contribute to ecologies of information that arise from shared interests, shared understanding and shared goals rather than a shared classroom or shared spatial geographies. The focus of this research is on the "practices" of individuals, that is, the story-telling practices they engage in on social media, and how these practices might shape the narratives about particular places, and their relationships with those places. The centrality of action in these discussions (Lave and Wenger 1991) (Tusting 2005) makes them particularly relevant to examinations of social media technologies because they facilitate the discovery of shapes in the blur of data; they provide a lens through which the mass of information discernible in social media text data may be examined. The action explored in this thesis is the narration of place, the construction of shared understandings about particular sites and the linguistic strategies employed in order to communicate these understandings. Social media platforms, including Twitter, when combined with the location sharing capabilities of new technologies, have produced a large amount of data sharing that is attributable to an exact point and space in time. The abundance of this data is encouraging non-experts to engage in map-making activities and subsequently altering the expectations of those people encountering maps in their day-to-day lives.

2.3.8 Volunteered Content and "*The Crowd"*

The participation of large numbers of disparate individuals in the generation of knowledge has been facilitated by digital technologies. The generation of knowledge can be directed as well as undirected. The direction of a group of individuals in pursuit of a common goal is known as Crowdsourcing. Crowdsourcing is a term that is extensively used and may refer to a number of activities; it can happen online, offline or may be a hybrid of both. Notable Crowdsourcing projects include GalaxyZoo (a project which asked individuals to classify images of galaxies according to their dominant characteristics) and OpenStreetMap (which asks users to contribute geographic information).

According to Enrique Estelles-Arolas and Fernando Gonzalez-Ladron-de-Guevara, “the adaptability of crowdsourcing allows it to be an effective and powerful practice” but this same adaptability is what makes it “difficult to define and categorize” (Estelles-Arolas and Gonzalez-Ladron-de-Guevara 2012:1). The researchers accept that crowdsourcing is centred on “the crowd”. According to their research into a number of definitions, this concept is somewhat elusive but within the context of crowdsourcing, it has a few key features. A crowd can be defined by

a) Who forms it
b) What it has to do
c) What it gets in return for its work

Other impacting factors include the initiator (who it is, what they get in return for the work of the crowd) and the process by which the crowd are called into action (type of process, type of call used and the medium used). The “crowd” of social media (when considered as a general category) does not provide an easy answer to these questions, which are explored further in the methodology chapter. The first question is addressed here, as it connects this work with previous research concerning the behaviours of people online.

Who forms the crowd is divided into two aspects, the size of the crowd (member numbers) and the type of person making up the crowd (member demographics). Determining the size of the crowd in online research is difficult, as not all members of “the crowd” will be active. There is a huge body of research discussing how important “lurking” (i.e. non active participation in forums and other public spaces online) is to online life. Despite the negative associations with the word “lurking”, there is a lot of research into the positive implications of “lurking” (Schnieder, von Krogh and Jaeger 2012) (Preece, Nonnecke and Andrews 2004) (Merry and Simon 2012) (Kucuk 2010) and (Dennen 2008). These analyses of online activity serves to remind the researcher that the demographic make-up of social media users may

not be reflected in the content generated through the platform, as not all users participate actively in the creation of content.

2.3.9 Volunteered Geographic Information

Volunteered content with a spatial emphasis is known as Volunteered Geographic Information (VGI). Notable VGI projects include Open Street Map (OSM)[10], Mapping Manhattan (Cooper 2013) and the Anti-eviction mapping project (Anti-Eviction Mapping Project n.d.). VGI often intersects with crowdsourcing and the use of social media, in disaster response and management for example (Horita, et al. 2013) (Granell and Ostermann 2016). These activities are nearly always goal-oriented and have been recognised as "a valued and useful source of information" for practitioners in need of spatial data, including the World Bank (Haklay, et al. 2014).

2.3.10 Social Media Data Streams

The content of social media platforms is contributed by a dispersed group whose members, taking the platform in its entirety into consideration, may have very little in common with each other other than their ability and willingness to become a user of the platform. The status of users in the context of data production depends on how the platform frames the activities of the users. This determines whether the users are working towards a clearly articulated goal, participating in similar activities or simply making use of the technologies available to them, with no awareness or thought about the "end" of the traces they leave behind. This thesis contextualises its methodological work within an analysis of primary empirical methods used in existing social media research and an overview of a selection of the research focussed on Twitter that helped to shape this work.

The user-generated focus of social media platforms is integral to this research. Content is not centrally distributed and though it is moderated

[10] Open Street Map is an online platform where registered users may contribute information about a geographic location. Users may contribute nodes (points on the maps), ways (a line or area like a building outline or a footpath) and tags (information about a node or way like a pub name or rules of the road. The information must be both "real" and "current" (Open Street Map n.d.)

(with the aim of reducing abusive content or illegal activity), the narratives of social media platform reflect the activity and sentiments of their users. The data streams in this context emerge organically and are largely undirected (that is, they are not motivated to compile specific types of information in pursuit of a common goal or task).

There is an established body of literature exploring the particular impacts of specific social media technologies (Fox and Moreland 2015; Junco 2012) including, for example, an analysis of the personality traits of Facebook users compared to non-users (Brailovskaia and Margraf 2016). This thesis builds on this body of work by investigating the data made available through a single platform. It extends existing methodological, commercial and theoretical engagements with social media platforms by considering how the data made available through these platforms might be applied to contexts beyond those in which they are created. The intention of this thesis is to consider the geographic relationships that are expressed through social media texts.

2.3.11 Data Quality

Web 2.0 practices and the curation of data on a personal scale has had clear impacts on the creation of geographic traces, particularly in urban contexts. In addition to more general intra- and interpersonal impacts of new sensing technologies, the abundance of opportunities for individuals to participate in content creation has altered the way that representations of space and place are created. The introduction of Google Maps in 2005 marked the beginning of contributed geographic data on a grand scale. Like never before, people, (anyone with access to the internet and the requisite, basic, skill set) found themselves in control of the data concerning, and the visualisations produced about, particular places. Some argue that these developments have impacted negatively on the quality of maps.

> *"do-it-yourself map making has had some serious side effects on quality – or what we might refer to as the Googlization of maps. First, due to their ubiquity, people have*

> *come to expect that all online maps should look like Google maps."* (Muehlenhaus 2014, 10).

Muelenhaus argues that the ease with which people are able to access Google maps has not only impacted the quality of the geographic information but also affected people's expectations about what a map should look like. Dorling and Fairbanks take a similar view too, showing concern for the emerging relationship between producers, technology and the resulting visualisation. They lament the loss of "artistic mystique" in contemporary, digitized mapping processes, "although the world is awash with maps, fewer and fewer cartographers are drawing these pictures. The artistic mystique of the scribbling pen and set-square has been replaced by the scientific black-box of the digitizer and the geographic information processing engine" (Dorling and Fairbairn 1997, 121). Geospatial Information Systems (GIS) and other technologies like Google have opened up the map medium to non-experts, which causes tensions over best practice and concerns about data quality, validity and accuracy. Although there is an obvious concern with the outputs of these technological approaches, the concerns raised by Dorling and Fairbairn and Muelenhaus stem from a belief in the need for expertise in the production of geospatial knowledge, the expertise they do not see in the digital realm.

Google Maps, Open Street Map (OSM) and similar platforms are enabling individuals and communities to participate in mapping and spatial framing activities of their own, to encode their own histories onto a map or visualisation. These technologies have been picked up and promoted by public institutions, like Liverpool Museums' "Mapping Memory" project and the Anti-Eviction Mapping project. Here the power of connected technologies to tell stories is evident. These two initiatives are embracing the power of mapping and tracking software and are beginning to provide opportunities for individuals to contribute their own images, text and audio to provide a collection of "memories" about the Liverpool Waterfront (National Museums Liverpool 2015) and to resist forces of dispossession in American urban centres (Anti-Eviction

Mapping Project n.d.). These projects demonstrate the interest in and potential utility of place-mapping (as distinct from location-mapping) projects to social scientists, policy makers, historians and human geographers. The insights into personal and shared narratives about place provided by these projects clearly show the value of digital technologies in the investigation of place. Many of these mapping projects, particularly the digital ones (like Mapping Memory and the Anti-Eviction mapping project), lack nuance and flexibility; they are largely point-based and encourage users to drops pins in relevant locations. Point-based mapping is useful because it enables the easy organisation of data; the concern is what is being missed out of these maps and the ways in which they are being read. They tie memories to geographic location and risk becoming isolated narratives, rather than accounts of a shared sense of place.

> *"mapping has emerged in the information age as a means to make the complex accessible, the hidden visible, the unmappable mappable. As we struggle to steer through the torrent of data unleashed by the Internet, and to situate ourselves in a world in which commerce and community have been redefined in terms of networks, mapping has become a way of making sense of things"* (Abrams and Hall 2006, 12)

Mapping and visualisation technologies have been greatly democratised since the beginning of the Web 2.0 era, the popularity and uptake of these technologies is a testament to the significance of location in the development of both personal and community identities. This thesis means to explore these relationships with place through the close consideration, not of the targeted narratives of projects or initiatives, but through the day-to-day expressions of social media texts, where the emphasis on producing a coherent story is reduced. The impact of the availability of location-based information, thanks to tracing technologies and readily-available online mapping platforms, has resulted in a massive surge in the volume of highly personalised geo-located data.

2.3.12 Data abundance and Data (il)literacy

In addition to the concerns about quality and the loss of the "mystique" voiced by Muelenhaus, Gerd Gigenrenzer (Gigenrenzer 2003) expresses concerns about blanket access to quantitative data, arguing that "many people do not know how to think with numbers" (Gigenrenzer 2003, 23). Consider for a moment, the language of risk. Risk is often communicated with reference to "chance" and percentages also feature heavily. The example Gigenrenzer uses is the communication of predicted weather conditions, which often use language like "a 30% chance of rain". Gigenrenzer argues that information like this is open to interpretation and its dissemination in this way is unique to the language of risk. So "although uncertainties are old, risks are relatively new" (Gigenrenzer 2003, 32). The availability of vast amounts of quantitative data about the individual, as well as aggregated data about groups of people, means that not only do we have unmatched access to detailed information about ourselves and others but we also have a unique way of expressing this information, through the language of risk. At the same time, there is no guarantee that this information is being read properly or understood at all, let alone engaged with at a critical level.

Similar to the concerns about "thinking with numbers" raised by Gigenrenzer (Gigenrenzer 2003), Brannon points out that the development of certain technologies, in this case, satellite imagery, has transformed the way people seek, access and assess information. Brannon argues that the availability of increasingly sophisticated sensing technologies tends to produce "a quantified space that replaces difference and heterogeneity" in order "to produce standardizations that hold authority over subjective experiences of place" (Brannon 2013, 271). Brannon insists that this objective authority is the privilege of the "View from above" (Brannon 2013, 271) that has been enabled triggered by satellite technologies. The human element is completely removed from some modes of visualisation: "there is no human at the centre of Google Earth, and the social is transformed from the centre to

an abstracted, bodiless form, with the viewer's eye positioned as a mediated, nonhuman, omniscient optic" (Brannon 2013, 278). Williams echoes Brannon's work but calls the perspective "God's eye objectivism" (D. Williams 2014, 75).

Lorraine Daston and Peter Galison, in their book "Objectivity", show similar concern about the eclipsing of lived, human experience from representations of place. They suggest that objectivity is not an "inherent ontological form" but is, in fact, fluid (Daston and Galison 2007). These discussions raise some interesting questions about how data is captured and assessed based on the mode of capture. The work of both Brannon and Daston & Galison proposes that even in the era of "Big Data", digitization and remote sensing, accuracy is nothing more than an illusion, it remains: "based on what the user is interested in knowing about a space, the processing is controlled to create a visual output and determination of what the space 'actually' consists of" (Brannon 2013, 284). The work of Chris Alen Sula (2012) takes a slightly different view on the exclusion of subjectivity from visualisation. Sula argues that the influx of quantitative data brought about by new sensors, geotags, social networks and the like, has had a positive effect, notably in the arts and heritage sector, highlighting the possibilities of inclusion that have been enabled by developments in sensing technologies. According to Sula, "this empirical validity helps to produce visualizations that are more ethical in the sense that they are more inclusive of various groups and disinterested on the whole – the victors can still write history, but only insofar as they can measure it" (Sula 2012, 253). These challenges are irreversibly connected to the use of social media data in research and are particularly relevant to the collection of ambient data. The extraction of data from one context, and it's re-appropriation in another amplifies one of the central concerns of this thesis, that is; the seductive rhetoric of "data" in a digitally-mediated world. In an attempt to resist this rhetoric, this research has been conducted with complexity and context in mind. The volume of data available through social media platforms has been balanced by the

approach taken in the methodology, a methodology which seeks to engage with individual narratives and to acknowledge the forces at play in the construction of these narratives.

2.3.13 The significance of context in social media research

The scope of social media research is broad, and given the turbulent landscape of personal digital technologies, it requires a flexible approach that is sympathetic to flux. This sensitivity to change can be hard to cultivate unless the wider context of certain interactions or technologies is taken into account. This understanding of context is important and can be hard to glean unless researchers are willing to engage with potentially short-lived technologies and to take an in-depth, interpretive approach to analysis. Hine, for instance, has argued for the need to understand the interactions that are mediated by technologies, in spite of the evidence that the popularity (and even existence) of digital technologies can surge and then fizzle out quickly (Hine 2005). New media technologies like microblogging platforms (Twitter) and geo-located chat streams like YikYak, which went defunct in 2017 (Williams and Mahmoud 2018), serve as an "important conduit for contemporary life" and need be understood, in spite of the "anxiety" that their newness might once have fostered amongst researchers (Hine 2005: 1). Although social media platforms, including Twitter, are not really "new" anymore, Hine's argument highlights a central challenge of technology-based research; technologies are changing all the time, and so are populations that use them.

Robert Kozinets, writing five years after Hine commented on the "anxiety" raised by new technologies, restates the importance of being sympathetic to technological and social contexts.

> *"technology constantly shapes and reshapes our bodies, our places and our identities, and is shaped to our needs as well. Understanding of the way this transformation unfolds requires us to keep a keen eye on particular and general contexts – specific times and places, distinctive*

> *rules or rational procedures, institutional histories, technical possibilities, practical and popular uses, fears and dreams. A thorough understanding of these contexts requires ethnography" (Kozinets, 2010: 22).*

Kozinets emphasises the symbiotic relationship between technologies and people and coined the term "Netnography" to reflect this approach as an expansion of traditional ethnographic methods into the realm of technology. According to Kozinets, "netnogaphy" is "ethnographic research adapted to include the Internet's influence on contemporary social worlds" (Kozinets, 2010: 1). Twitter has been described as "the poster child of the burgeoning movement in computational social science" (Stephens et al. 2015: 1) and is a common focus of interest for scholars. Understanding the social and technological conditions under which Twitter texts are generated, then, is essential to providing a "complete" picture of tweet language. This conviction shaped the interdisciplinary approach taken to the analysis of the texts. Although this thesis does not make any claims to understanding the "total linguistic fact(s)" of every user whose tweet text was analysed, the methodological approach aims to provide a managable famework from which to tackle the practical, ethical and theoretical challenges of examining place in a technologically mediated urban world.

2.4 "The city" as a contemporary urban space

Technological and digital innovation has not only impacted the individual, but has also altered the ways in which shared spaces can be imagined and experienced. The seemingly never ending stream of human and non-human data generated through hybrid networks speaks to the unresolved nature of contemporary urban spaces, as discussed by Beauregard (Beauregard 2018).

"The City" is a perfect example of the nebulous and ephemeral nature of contemporary spaces. Existing literature suggests "the city" is difficult to define. As settlements, they share a lot in common with other designations, like towns and villages, but manage to retain a special

distinction in both an imaginative and administrative sense. Beauregard articulates the struggle well by stating “Keeping in mind that cities are both physical things and ideas…we are still left with the problem of having to distinguish the city from other forms of human settlement” (Beauregard 2018, 8). Beauregard highlights the struggle to find a solution to this defining problem through either purely qualitative or quantitative means. It is not enough to say that a city is larger than a town or a village, or that it simply holds a special place in the popular imagination; Beauregard suggests a city must be distinct in terms of both its physicality *and* its imaginary qualities. The contemporary city has been theorised as possessing several characteristics that differentiate it from non-urban, suburban (Maher and Haas 1987, 281) and non-city spaces. The anthropologist Adam Reed describes the natural and living qualities of cities, as expressed by self-declared enthusiasts (Reed 2002). Reed refers to the great detail and affection with which people can engage with and reflect upon urban spaces, fixating on tiny details and searching for connections across both time and space.

The idea of the city as a point of convergence is also fairly prominent in the literature, it is often cited as existing as the site of an urban community brought together by social, political and economic forces and differentiated from non-urban spaces in either the planning of the city (sometimes these limits were historically defined by the construction of city walls) or the “action” of it – through the administrative and civil duties of its inhabitants (Gruber 2016).

There are, though, an increasing number of voices who argue for the continued relevance of the city as a site of interest, even in the absence of discernible edges – in a letter to the Wilson Quarterly, Bill Bishop confidently declares;

> *“Cities are the places where most Americans construct their identities…If family, class, occupation, and religion once guided us through life, today each of us cobbles together a*

> *life of one's own with the people and activities we discover in the cities where we choose to live" (Bishop 2011, 5)*

Here, Bishop relates the formation of personal identity to the forces of the city, emphasising the importance of daily encounters and activities.

In line with the conclusion of Beauregard, who theorises that conflict is the most significant aspect of the contemporary city, the designer Rebecca Solnit explores the contemporary city as a hybrid space, not possessing a single nature, but rather made up of a number of overlapping natures.

> *"A city is a particular kind of place, perhaps best described as many worlds in one place; it compounds many versions without quite reconciling them, though some cross over to live in multiple worlds"* (Solnit 2010, vii).

This multiplicity has also been picked up by Becky Cooper, in a similar, but more recent project exploring the multiplicity of place is the "Mapping Manhattan" project, launched by Cooper in 2009. Throughout the duration of this project, blank outline maps of Manhattan Island in New York were printed and distributed to strangers (volunteers) around New York City, who were asked to "fill it in". What resulted was a series of intensely personal reflections on the island. Figure 2.2 (page 49) shows several examples from the curated collection, which was published as a book in 2013 (Cooper 2013).

Through the eyes of Beauregard, Cooper and Solnit, we see the city as a living patchwork, constantly being made and remade, under continual negotiation and exhibiting a dynamism that makes its objective study impossible. Representations of the city can solidify the ephemeral nature of the city as a site of exchange and transit.

Rebecca Solnit, who compiled an "atlas" of San Francisco, which depicts a number of varying representations of the city, argues

> *"An atlas is a collection of versions of a place, a compendium of perspectives, a snatching out of the infinite ether of*

> *potential versions a few will be made concrete and visible"* (Solnit 2010, vii).

New sensing technologies have contributed to the pursuit of a documented city, presenting new challenges as well as offering solutions to historic problems. Denis Cosgrove eloquently states the challenges and opportunities that have been brought about by new sensing technologies and modes of visualisation in the context of the contemporary city:

> *"The contemporary city presents both complex new challenges and enormous opportunities for mapping as do emerging survey and plotting technologies…But today's acute awareness that cartographic images can never be innocent vehicles of information dissolves neat distinctions between celebratory and regulatory urban maps….Current visual technologies mean that the opportunity for creativity in shaping and recording urban experience is greater than ever, as too is the need for critical attention to the making and meaning of both public and private spaces"* (Cosgrove 2006, 157).

Denis Cosgrove discusses encounters with unfamiliar urban environments and the role of the map as one of mediator. "Practically, confrontation with an unfamiliar city is typically mediated by the map…Urban experience in a new city is often a process of negotiating the divergence between cartographic and material spaces." (Cosgrove 2006, 148). He questions the suitability of Cartesian forms for the representation of contemporary urban space, stating "In their own ways, these city maps reveal that the modernist faith in geometry as a guarantor of urban legibility was unsustainable. Conventional Euclidean forms neither describe nor contain the spaces of the increasingly flexible, mobile, cybernetic city" (Cosgrove 2006, 156). The site of interest here, "the urban", is a concept that is being continually defined and re-defined. Adey and Bevan challenge the idea of the city as an

entity that is fixed in space and time and encourage us to reimagine it as a system of constant change that should be “increasingly seen in terms of flow and flux” (Adey and Bevan 2006). Adey and Bevan describe the city as an intersection of free-flowing information with the movement of individuals; not an unchangeable, monolithic entity, but rather a nexus of exchange between the physical and virtual worlds. This stress on flux and flow is not uncommon – Relph insisted that “non-places are the measure of our time” (Relph 1995: 64), and were borne out of the “inherent dynamism” of contemporary urban life (Mumford 1961:9).

Clearly articulating what the city is or what a city means is not an easy task, but certain themes recur in the literature. The city seems, more or less, to be discernible through three main characteristics; firstly the city exists in a dual state and must be at once real and imaginary. Secondly, the city is a site of convergence, where day-to-day activities co-occur, where people and information overlay one another, so that nothing is ever “complete”. And thirdly, the city is mediated by technology.

The fluidity of contemporary urban spaces was a contributing factor in the selection of research sites for this thesis - with search parameters built around two English cities. The choice of sites is discussed fully in the methodology chapter and was shaped by the literature that has been discussed in this section – building on an existing body of work

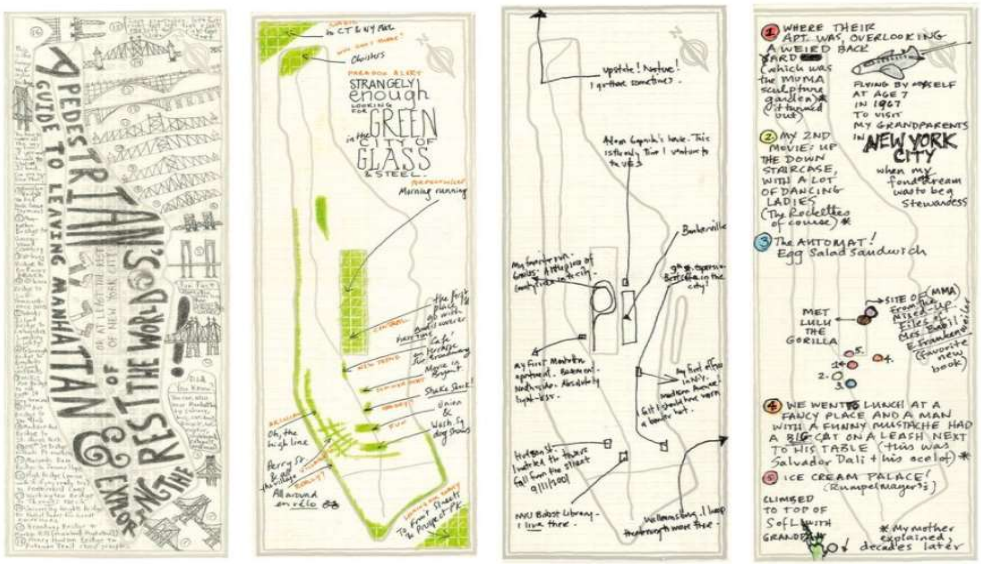


Figure 2.2: Example maps, from Cooper's "Mapping Manhattan"

that emphasises both the dynamic nature of contemporary urban spaces and the fundamental interconnectedness of humans and digital technologies like social media in these contexts. These sites generate a unique condition of being, where the confluence of technologies, people and social and economic flows is at its most concentrated.

2.5 Theorising Place

Having considered the impacts of technology on spatial and intrapersonal cognition, the review now turns to the examination of "place" in a variety of theoretical contexts, and summarises several critical engagements in its theorisation. Place is critically engaged with a number of disciplines, and there is no agreed upon definition. The concept has been linked to discussions of embodiment, navigation and sound, which are not explored here, as well as a fairly large body of work that emphasises its ephemeral and "living" nature. The focus here has been particularly on "place" as differentiated from "space" with a focus on urban centres. Three key differentiating attributes of "place" emerged from an extensive review of academic and lay engagements that crossed disciplinary boundaries. These key attributes are experience, time and juxtaposition. The categories are not discreet and are often intertwined with one another.

2.5.1 Spaces and Non-Places

Place is often theorised in opposition to space (Tuan 1977) but these two ideas are not entirely exclusive nor are they unalike in all aspects. The clarity and simplicity of "space" cannot be taken for granted. Janet Abrams and Peter Hall, for example, provide a three-faceted description of space. For them, "space" exists in three main forms, (they use these to prefix the word): Information, Physical, and Social. Each one of these aspects occurs at various scales, and the explanations are necessarily broad. "information space" is about making sense of data, "physical space" concerns the representation of material structures, cities, regions or buildings and the third category "social space" is about human relationships (Abrams and Hall 2006, 12). This three-faceted breakdown finds many similarities with the various theories of "place"

outlined in section 2.5. The forces that interact not only with material or theoretical structures but also with people have shaped the design of the methodology and the approach taken in analysis. The emphasis found within explores multi-dimensional meaning and aims to describe some of the rich description of locations present in social media texts.

The dichotomy of space versus place is further complicated by theorisations of a third designation that is neither fully realised (place) nor completely empty (space). “Non-places” have been introduced by theorists concerned with the conditions of modernity and supermodernity, under which urban centres do not possess a single, well-defined identity but rather a nebulous and ever-shifting identity that mirrors the movement of people, goods and ideas through them. Augé argues that it is the flux of supermodernity that has led to the creation of these “non-places”, sites characterised by their qualities of disconnect and social isolation (Auge 1995). According to Augé, the conditions of supermodernity have transformed the role of space in human interaction. Like the “electrically-accelerated elsewhere” described by Federman (Federman 2005, 6), in the age of supermodernity, individual places do not possess a stable set of values or meanings but exist under conditions of uncertainty, where their values are constantly being re-negotiated. Non-places, for Augé, are characteristic sites where the conditions of supermodernity are most emphasised, where the forces of sped-up communication and the “shrinking of the planet” (Auge 1998, 104) have the greatest opportunity to coincide. These “spaces of circulation, communication and consumption” permit individuals to pass through and use the same physical locations without finding connection with one another, they enable “solitudes [to] coexist without creating any social bond or even connection” (Auge 1996, 178). Examples cited by Augé include motorways, shopping centres and hotels. Augé describes place and non-place as being,

> *“like opposed polarities: the first is never completely erased, the second never totally completed; they are like palimpsests on which the scrambled game of*

> *identity and relations is ceaselessly rewritten"* (Auge 1995).

Augé's theorisation of non-spaces emphasises disconnection in close proximity. This focus on centres of circulation, communication and consumption irrevocably links "non-place" to urbanised environments; where day to day interactions are more likely to be mediated by technologies, and where flows of people and goods are at their most exaggerated. The "opposed polarities" of place and non-place result from the action of individuals within and upon them – whether they can never be forgotten or never be finished, they rely on the experiences of people to exist.

2.5.2 Experience, the dwelling perspective & intersubjectivity

The relationship between the individual and "place" has been strongly emphasised in cultural geography, where geographies are cast as both real and imaginary in the sense that all places have meanings that extend beyond their geographic characteristics. Tim Cresswell differentiates space from place through the claim that a place is a "meaningful segment of geographical space" (Cresswell 2008). Cresswell insists that the meaning inherent to understandings of place results from the lived experience of individuals in those spaces (Cresswell 2014); it is these experiences that meld objective reality with a subjective understanding.

In line with the perspective put forward by Cresswell (2008, 2014), Relph (1976), (Tuan 1977), and more recently, Carmona et al. (2010) all emphasise the social and emotional characteristics of place; Tuan differentiates place from space through a focus on knowledge and the development of spatial relationships over time, stating "What begins as undifferentiated space becomes place as we get to know it better and endow it with value" (Tuan 1977: 6). Cresswell's explanation emphasises the mix of both "objective and subjective facets including location but adding other, more subtle, attributes of the world we inhabit" (Cresswell 2014, 249). Tuan imagines place to be the cumulative result

of the lived experience of space, developing an identity over time based on “impressions, given to us through the senses” (Tuan 1977, 18). Similarly, Relph describes “the realm of direct emotional encounters” (Relph 1976, 10) whilst Carmona reflects on the “sense of association or identity” experienced by people in connection with certain locations (Carmona, et al. 2010, 120). Tuan’s imagining of “undifferentiated space” is used throughout this thesis in contrast to “place” and its many different framings, though the concepts are sometimes used interchangeably, this theoretical distinction is an important one in the contextualisation of this work.

Durkheim’s examination of human sociality in terms of the sacred and the profane can be used to frame locations as sites that hold both people and values, which may be enacted through ritual or habitual practice (Durkheim 2001), so that a building or open square can hold dual meaning, being simultaneously a physical location and a site of personal or spiritual significance.

The work of Bruno Latour (Latour 2005) and the sociologist John Law on Actor Network Theory (ANT) (Law 1992), though old in a technological sense, provides a useful perspective on the confluence of ideas and lived experience, and on the impact of theoretical frameworks on the practice of everyday life. The emphasis on the importance of non-human agents in their vision of the social universe allows room for technologies in the critical discussion of human sociality. Law and Latour’s ANT permits researchers to approach ideas, people and objects as equal contributors to any social environment; place can also, in this sense, be said to be part of an individual’s network. The democratisation of actors in this single-plane vision finds parallels in the work of the anthropologists Tim Ingold (1993), Clifford Geertz (1973) and Michael Jackson (1998)

What Tim Ingold calls the “dwelling perspective” (Ingold 1993) paints a picture of a fluid social universe, encompassing a number of agencies that are connected in non-binary and not-necessarily-hierarchical

relationships. Ingold's "dwelling perspective" explains that humans do not exist outside of nature but are integral to its being. According to this perspective, people are part of nature, and indistinguishable from it. Though Ingold might reasonably object to its application to a discussion of definitions, such arguments usefully contribute to an analysis that seeks to put human activity at the centre of the discourse about "place". These discussions of human sociality and the theorisation of place share a few similar features. They are concerned with a preoccupation with lived experience, with the acceleration of time under certain conditions and with the democratisation of place, which facilitates the juxtaposition of not-necessarily-concurrent viewpoints.

2.5.3 Time

Physical location is not the only concern of spatial theorists. Time is often also considered to be a central component in the construction of place. Henri Lefebvre (H. Lefebvre 2014) for instance, highlights the significance of time in his analysis of rhythms. His description of rhythms is wide-ranging and includes public, social rhythms as well as personal, biological rhythms. Through the use of several examples, Lefebvre invokes time as fragmented by important events or rhythms; with fetes and celebrations demonstrating larger-scale social occasions, while personal biological rhythms such as "digestion and tiredness" existing toward the other end of the scale (H. Lefebvre 2014, 27). Lefebvre's argument centres on the perpetual conflict between these various rhythms, including "dominating-dominated" rhythms, which are to non-natural, invented in the sense that they are not biologically necessary. These invented rhythms may last for a long time or be very fleeting, always coinciding with each other in the production of an "effect that is beyond themselves" (H. Lefebvre 2014, 27). Lefebvre's social critique called for a re-examining of everyday life under the conditions of modernity, where consumption had replaced boundaries between social classes and consumerism had shrunk the temporal and physical separation of individuals and their social values. Lefebvre's focus on the everyday has been a key influence in the focus of this thesis on social

media narrative, whilst his framing of what he calls “social space” echoes the work of Tuan, Relph and Carmona, labelling place “a product…a tool of thought and action” (Lefebrve 1991, 26). Similarly, Doreen Massey discusses the problematic interweaving of concepts of place, identity and authenticity under forces of “space-time compression” triggered by social and economic change (Massey 1994, 121).

The role that technologies can play in the disruption of natural temporal rhythms is a common theme in literature concerning public places, memorials, museums and heritage, where technological practices are used to create narratives about past events or in the idealisation of an imagined future. Alice Bell, for instance, speaks about the capacity of technologies to preserve the “ephemera” of past relationships and to make personal histories public, to such an extent that they become part of a collective memory (Bell 2012). This experience is similar to the “prosthetic” memory that Landsberg describes (Landsberg 2004), where people take on the narratives of others because of the proximity that digital technologies allow. These discussion are also examined by Haskins (Haskins 2007) and McGuire (McQuire 1998), who highlight the use of images in the creation of shared memories across historical eras.

Time plays an important role in nearly all discussions of place, whatever the discipline. In physical geography, it is not uncommon to encounter reflections on geologic time, while in social anthropology, conversations are more commonly about time on a much smaller scale; in terms of an single event, ritual or an examination of a site over a limited time (usually between several months and several years). What these disciplines all share is an emphasis on the importance of time, at whatever level. Given the layering and overlapping of time frames in varying approaches to technology and spatial representation, it is fitting that the data collected during this research, and presented in this thesis is viewed as a small segment, one small “dot” in the “pointillist” metaphor offered by Page (2010). Each interaction with a place helps to shape it, and those places, in turn, shape the individuals who pass through them. Place is “if not infinite, then practically inexhaustible”

(Solnit 2010, 2), and each investigation of it no matter how infinitesimal, is part of a cumulative and symbiotic identity.

Overlapping many of the issues tackled by Augé in his discussion of non-places and the seemingly exponential shrinking of both time and space, Pico Iyer introduces the concept of the “Global Soul”, a state of experience he claims is common to people all around the globe. The condition of the “Global Soul” is not spatially limited, and does not necessarily require that the individual physically inhabit one or other of transient sites (or “non-spaces”) as described by Augé. Iyer contests instead that it is technology and travel that are the “two great engines of our age” (Iyer 2000, 12). Through the interaction of these engines, Iyer argues, their trajectories become intertwined, each transmogrifying the other so that “memory itself seems accelerated, and yesterday’s dramas become as remote as ancient history” (Iyer 2000, 13). The issues raised by Iyer, Augé and Solnit are useful in explaining the inadequacy of methodological and theoretical approaches that attempt to decode and understand human relationships with place without appreciating the complex technological and temporal conditions under which they are formed and maintained.

The insufficiency of static drawings of human settlement and geography has also been raised by others, including Diana Balmori, who points out that the “frozen geography” (Balmori 2003) of historical maps has been altered by several technological advancements in the late twentieth century, most notably, easy and cheap access to printing. In her view, Geographical Information Science (GIS) is a continuation of a technologically enabled alteration in the fundamental approach to social representation, which allows maps to be produced in greater volume and at less expense than ever before, meaning that there is less and less time between iterations. She points out “where only space was mapped before, we are now mapping time” (Balmori 2003) technological change has impacted the frameworks which best suit geographic and social data, increasingly, there is a need to visualise change and so time is an important feature in the setting of

contemporary mapping. Personal data, particularly that which reflects the choices that individuals make about their lives every day, if visualised in effective ways, could help to reveal how locations feature in the narrative processes that we all engage in every day and would theoretically allow us to “catch ephemera before they are missed completely” (Balmori 2003).

Raymond Federman engages with similar tropes when he talks about the collapse of linear time that has been brought about by modern technologies, echoing the arguments made by both sociologists and geographers alike. In the “electrically-accelerated elsewhere” (Federman 2005, 6), Federman argues, linear, visual and sequential time has collapsed, and given way to a way of being and of telling that is “acoustic” in nature. (Federman 2005, 6). By drawing a parallel between the acoustic and visual worlds, and by using these analogies as a way of interrogating the way that information is distributed, received and consumed, Federman demonstrates the impact of technology on the very nature of the way we can (and do) live in the world. He explains: “our experience in the connected spherical space of the internet is that the centre is everywhere, and the boundaries are nowhere” (Federman 2005, 6). Many visualisations of this ilk have a rounded or spherical structure, as they try to demonstrate the interconnected nature of people and information in networked society.

2.5.4 Juxtaposition, Deep Maps and Layering

The complexity of contemporary urban spaces has been addressed in literature across several disciplines, beginning in the mid-1990s, most notably by Castells (Castells 1989, 1996) and Urry (1995). The urban environments described by Castells and Urry were characterised by their nebulous and diverse natures, which were contingent upon the technologies available within and around them. In these emerging urban spaces, technology was the means by which people could connect to their environment, communicate with acquaintances and negotiate their identities. The changeable and complex nature of urban spaces can be summarised through the concepts of “superdiversity”

(Vertovec 2007) and "supermodernity" Federman (2005) and Augé (1996) [11] which result from the intersection of ideas, people and technologies.

The conditions of superdiversity and supermodernity are both contingent upon the overlapping of various forces. These concepts, therefore, lend themselves well to the theorisation of "place" in contemporary urban spaces. The "places" of the city are unbounded and layered. The identities of these places can develop over time, and evolve out of different experiences. The unresolved nature of urban spaces and the potential for different descriptions of those spaces to overlay one another is explored throughout the body of this thesis under the heading "juxtaposition".

These perspectives all share the argument that place cannot simply be a "container" for social interaction, but rather it is part of a mutually arising identity with many contributors or authors. Its definition shifts from abstract theorisation into lived experience and social practice. Place is, in these perspectives, a product of human encounters and is fuzzy-edged, continually being re-negotiated and created through interaction, and part of a social and biological ecosystem rather than a well-defined and limited constant.

The conviction that places are sites of intersection and of coming-together, is also expressed by Massey, who insists that place "cannot be statically defined or bounded" and complex entities "where the forces of both space and time come together to create a particular social, spatial and temporal experience" (Massey 1994). While Augé's concern was with the layering of human interaction in single sites in a short space of time, William Least Heat-Moon's focus is on the layering of various forces over a very long period of time (Least Heat-Moon 1991). In PrairyErth, Least Heat-Moon explores the intersection of geologic, climatic, commercial and human forces and their impact on the physical and cultural identity of the landscape. PrairieErth, a "Deep Map" written

[11] These ideas were explored further in section 2.5.1.2

by William Least Heat-Moon is a striking example of the variance in form that a geospatial information system (GIS) might take. The account does not present a 3- or even 2-dimentional map that we are most familiar with (as the front end of a GIS), but rather a detailed exploration of the social, geological and economic climate of the American West. It demonstrates several a lot of the same principles of a more "conventional", GIS, like ArcGIS [12], for example exhibiting a gridded structure encompassing overlapping layers. The region described in PrairyErth (the Flint Hills of central Kansas) is broken up into a four by four grid, the inspiration for which was the two-dimensional printed state maps. Each section of the grid is tackled in turn, with each section consisting of six primary layers. The first and second layers have the same heading throughout. The first section "from the commonplace book" explores the mention of the area through a list of excerpts, taken from other written works that pertain to the area in question. The second sets a locational frame of reference "In the quadrangle" (1991, 27, 66, 100, 142) whilst the remaining four layers shift between topics, exploring different aspects of life in that quadrangle.

By varying the extent and focus of the description at each layer between the quadrangles, Least Heat-Moon emphasises the intersection of different forces at the forefront of the reader's mind. Environmental and climatic forces like wind and rain (resulting in widespread seasonal flooding across the plains) clash with the human personalities he comes across in settlements, while economic pressures are clear throughout many of his conversations with residents. The multi-layered framing of time is also apparent throughout this work, operating as it does on a number of scales, from geologic, to anthropologic, to seasonal and even at the comparably minute level of day to day activities.

[12] ArcGIS is a desktop GIS software package permitting users to collect and manage geospatial data, create maps and "perform traditional and advanced spatial analysis" (ArcGIS Desktop 2018)

The stories recounted and documented in Least Heat-Moon's chronicle are stitched together in a multi-layered patch-work that combines the perspective of an outsider with the narratives of locals born and bred in the areas he visits. The narratives of the locals include accounts of their early dietary habits (they used to eat a lot of beef) with more folkloric tales of notable locations or people. PrairyErth is a layered chronicle of the landscape (both human and geological) of the American west, which has been shaped by cycles and flows that endure, not in spite of, but because of the relationship between people and their environment. In this way, the Deep Map is a cumulative model of an area that is simultaneously bounded and unbounded in space and time, and limited only by the representations made of it or the stories that are told about it. The layered approach to studying a place has heavily influenced this thesis, and the impact of this volume is further discussed in the methodology section (chapter three).

Emphasising the multiplicity of place and the convergence of multiple perspectives, Michael Jackson also insists that human relationships with their environment are central to the development of identities and that these relationships are dynamic and the relationship between a subject and an object (location and people in this instance) is more complex,

> *"dialectic of subject and object implies a reciprocal and analogical relationship not only between singular persons but between persons and a world of ideas, attributes, and things that are held in common without any one person having complete control over them or the last word on their meaning."* (M. Jackson 1998, 7).

Here Jackson elaborates on what he means by intersubjectivity, explaining the importance of control, commonality and reciprocity. Some lay perspectives also bring out the experiential elements of place in their discussion of it – the Society for One-Place Studies, for example, hosts a website meant as resource for groups and individuals

“researching the people of a community within the context of the place they live” (Society for One-Place Studies n.d.), which places a strong emphasis on the contextualised stories of a place and the centrality of people in those stories.

In a similar vein to both Augé (1995) and Ingold (1993), Jean Baudrillard’s description of the conditions of modernity also emphasises the un-finished nature of both time and space, insisting that neither space nor time can provide stable reference for an understanding of contemporary urban life. To Baudrillard, modernity exists under conditions that are “non-linear” and “non-Euclidean” (Baudrillard 1994). The non-Euclidean framing refers to the exponential expansion of possible interactions and the shrinking of the obstacles of physical distance and time which have been eroded by automotive and digital technologies. Where interactions might once have been constrained by location, age, schedule or language, the availability of technology and ready information, contemporary urban spaces are now filled with interactions of all kinds, and their identities can go on re-forming forever. Just as it may comprise a number of contrasting or complimentary meanings and associations, place cannot be “bounded”: it extends out as well as “up”, this idea is a common theme in the work of Doreen Massey, who describes places as “open and porous networks of social relations” (Massey 1994, 121).

2.5.5 Representational elasticity

The multiplicity of place is also frequently explored by disciplines outside the social sciences and humanities. Artists have long taken a keen interest in the representations made of particular locations and the results of their enquiries are often very compelling, able to capture the popular imagination and capitalise on our complex relationships with place.

When we consider place to be more than a container, to be both an ideological and physical entity that is under constant negotiation, it becomes difficult to hold in our minds a vision of it as clearly defined

and easily identified. GIS, where a visual representation is simply the face of a site that contains several overlapping, sometime complimentary and sometimes contradictory descriptions of a place, provides one solution to this representational problem. These multi-faceted representations are not unique to GIS, here; I turn my attention to the realm of art and design. My intention in this section is to demonstrate the practical application of the theories of place discussed in the first section and to show how these theoretical approaches might be brought to bear in the use of the user-contributed information of tweet text.

At the end of the 19th century in the United States, artists and illustrators would travel the country in order to document the size, shape and notable landmarks of urban settlements. But the exercise was not motivated by science or an interest in accurate surveying. The maps did not have a conventional north-south orientation, they were not flat, did not represent things “to scale” and were highly coloured. For the right price, local companies and residents could have their house or business premises made more prominent in the final images; thereby changing the landscape to suit their personal motivations (Biesinger 2017). This flexible process, as well as the prints it generated, one of which is shown in Figure 2.3: Phoenix, Arizona 1885. Sketched by C.J. DyerFigure 2.3, demonstrates the representational elasticity of maps and the importance of human personality in the reproduction of place.

Following the development and perfection of aerial photography techniques for use in military operations in the mid-20th century, there was less popular demand for these images, and the practice dwindled after the First World War. Not-to-scale maps like the one shown in Figure 2.3 are a striking example of the ways in which narratives are woven into the landscape, and provide insights into the complex relationships that people can establish with a place.



Figure 2.3: Phoenix, Arizona 1885. Sketched by C.J. Dyer

2.5.6 Differing Perspectives and Journeys

Sharing some of the trends of social science literature mentioned above, the overlaying of both time and location can also be frequently found in artistic engagements with spatial representation. The next section explores the work of two British-based artists who engage with the ideas of place, space, texture and lived experience in their pieces.



Figure 2.4: Furoshiki. a printed silk scarf by artist Hannah Waldron

In contrast to the detailed depictions of 19th century artists and illustrators, the artist Hannah Waldron uses simple geometric shapes to depict the urban spaces she spends time in. In Waldron's textile work, abstract forms represent the features of cities, as well as the journeys she takes to and from these spaces. Waldron takes a minimalist approach to representation, using simple shapes

derived from her work in weaving, to depict memorable features like buildings, rivers and weather types. In her work Furoshiki (Figure 2.4) geographic features are reduced to their simplest forms and incorporated into a visual narrative depicting her memories of London. Waldron is “interested in the reflection and mapping of experiences particularly in relation to places” and seeks to “draw upon a personal visual vocabulary of forms and mark-making that aim to distil information into its most essential with the aim to make images that communicate broadly” (Waldron 2013).

Waldron’s approach demonstrates the variety of methods by which the complexity of a single place may be represented, and how a variety of physical features and experiences can be combined to provide an understanding of the human experience of those spaces. While Waldron’s work focusses on a singular, personal experience, the artist Peter Griffiths employs technological connection to produce images of single sites that result from the digital overlaying and condensing of multiple viewpoints. Griffiths, who is based in Sheffield, uses community contributed data to produce sketches of urban land, natural landmarks and streetscapes throughout England. For his project “Castlegate” (Figure 2.5), Griffiths asked the public to submit photographs of a specific building or street in a city (Dartmoor and Sheffield are two sites he has already explored) and then compiles the images together by overlaying them digitally. He then uses a “computer controlled plotter cutter” to sketch (in pencil) over the dominant shapes and lines in the image. Over this pencil image, he adds colour by hand. What results is a tracing of the dominant textures

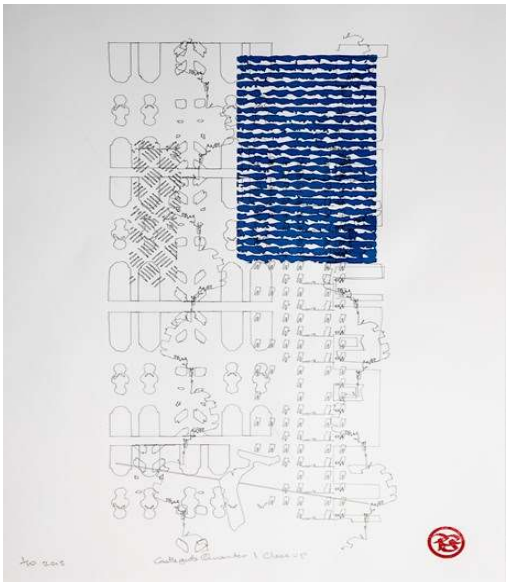


Figure 2.5: "Castlegate Quarter 1" by artist Peter Griffiths, showing dominant architectural textures

of the selected landscape or monument, a highly abstracted representation of a place as seen through the eyes of “the public”. What people choose to submit is telling - the textures of the resulting images provide an indication of how that place is experienced and memorialised by the people who pass through it. In his own words, the artist is “interested in the interplay of the visual world, technology and lived experience” (Griffiths n.d.). The images that result from his explorations of this relationship are excellent examples of how people experience, reduce, disassemble and remember places differently, but that common themes (or textures) may be revealed in their re-telling of those places.

2.6 (Re) articulation of the research gap

The methodological and theoretical work explored in this thesis spans several disciplines and demonstrates their shared concerns. Existing literature argues that personal digital technologies are used in a myriad of ways and hints at the abundance of experiential information about day to day experiences that might be used to inform understandings of contemporary cities.

Through the use of methods from applied linguistics, this thesis presents evidence to support the use of personal, sometimes messy, crowd-sourced data in the analysis of textual data in the context of digital media. The research presents an innovative methodology that incorporates ethnomethodological considerations with qualitative analysis and the computational querying of large bodies of text; the approach has been developed to enable close enquiry as well as large-scale statistical querying.

The connection between sensory and emotional experience in the description of “place” is long-lived in the geographic discipline, which continues to foster methodological approaches and theoretical stances that challenge any claims that the discipline is, or should be, purely a spatial science focussing on entire systems, rigorous methods and the production of research work that was replicable and not overly

descriptive. Since the quantitative revolution in the mid-20th century (1950s and 1960s), geography has developed a multi-stranded identity which means that work focussing on description, regional geography and certainly on the lived and emotional aspects of space tend to fall under the title of “human-” or “cultural geography”. The distinction between a qualitative and a quantitative geography is an important part of the theoretical and methodological foundation of the research work undertaken for this thesis, as it clarifies a transitional area of work that sits between these two strands, bridging the qualitative and the quantitative paradigms. This thesis aims to contribute to what Peter Jackson described as a “reconstituted cultural geography” which is “prepared to examine the multiplicity of landscapes” (P. Jackson 1989, 177) by making use of ambient social media data.

Though the locative elements of Twitter data have been extensively explored by other researchers, the focus of this thesis on the spatial elements of Twitter text data is unusual, and the application of this data to the geographic question of place is innovative. Although the work has been undertaken using data from a single, specified social media platform, there are far-reaching implications for the mixed methods approach taken here and opportunities for many complimentary future research projects combining spatial concerns with linguistic analysis and crowd-sourced information repositories. The work is pertinent to the business concerns of organisations specialising in spatial analysis and the development of geographic resources, including (but not limited to) mapping agencies such as the Ordnance Survey. The theoretical and methodological insights offered by this thesis will aid in the framing of research challenges specific to the use of spatial data from public sources as well as proposing a clear methodology, including an ethical framework for tackling this type of data.

2.7 Place in Contemporary Urban Contexts

“Place” has been explored in many disciplines, and its substantial use in lay vocabulary further complicates the identification of its essential characteristics. This review has examined the conceptualisations of

place that are particularly relevant to explorations of personal digital technologies in the contemporary city, drawing on theorisations from across the social sciences, the arts and humanities. The wide reach and complexity of the concept is evident, but three key ideas, about what distinguishes “place” from space, were revealed; the inclusion of reference to time, the perspective of lived experience and the overlaying of individual accounts (I have used juxtaposition as a shorthand for the latter). By highlighting these three facets of place and applying these to the search for an understanding of place expression in social media texts, this thesis makes use of commonplace narrative and demonstrates the possibility of empirically exploring nebulous ideas in the digital realm. It explores everyday interactions between individuals, digital technologies and their geographic surroundings through the lenses of experience, time and juxtaposition.

Building on an ever-growing body of academic and artistic work that posits places not as static or bounded sites, but as complex physical and imaginative sites that concurrently hold and meld the values of those who pass through them, this thesis presents the argument that personal digital technologies (social media platforms in particular) can provide data that supports the development of human-centred discussions of place. The argument is presented with the understanding that this data must be collected and handled ethically, and analysed in a way sympathetic to context in order to give a detailed and rich understanding of personal and social interactions with location in the digital realm.

In spite of the relative “newness”[13] of the types of interaction presented in this thesis, the findings extend beyond the boundaries of the specific platform (Twitter) and speak more broadly about the importance of ubiquitous technologies in the negotiation of contemporary urban worlds, which can transgress conventional linear models of both geographic (Euclidean) space and time[14]. Furthermore, this thesis not only

[13] See section 2.3.13
[14] See section 2.4

addresses the ethical issues central to social media research but aims to make these concerns a central theme in the discussion and concluding statements, thereby expressing a commitment to continuing research practices that are sympathetic to the human users of digital technologies. The difficulties in reconciling the distributed and sometimes conflicting (unresolved, to use the words of Beauregard) narratives about place generated in the context of social media require the methodological approach taken to the study of place in social media texts to be as flexible as possible. The approach used in this research was developed with the context constantly in mind; it reflects a striving toward inclusion, understanding and critical engagement, not “completeness”, “accuracy” or “validity”.

Chapter 3 Methodology

3.1 Introduction

This section of the thesis presents the methodological framework that shaped the research. It presents the exploratory work conducted in the planning of the project and the key decisions taken throughout. This chapter therefore provides an overview of the rationale supporting the methodology, as well as describing the methods used in detail. The main ethical and practical challenges encountered throughout the duration of the project, are described and four of the research objectives are addressed:

- Identify a suitable data source (social media platform)
- Specify the challenges particular to research using social media texts
- Develop a coherent and replicable methodology for the examination of place in social media texts
- Outline a framework for addressing the ethical issues associated with the use of social media data in research

The research that forms the basis of this thesis combines qualitative and quantitative techniques, informed by a number of established disciplines, including human geography, corpus linguistics and the locative arts. The mixed methods approach is demonstrated in both the formulation of the research question, the methods of data gathering and also in the analysis and interpretation of the data.

3.2 Identifying a suitable data source

In order to answer the research question posed and to meet the aims and objectives set out in the introductory chapter, the collected data needed to have both breadth and depth. The data needed to include information that might provide insights into the human experience of location, as well as be reflective of the individual platform from which this information was taken, which meant that the research design had to be tailored according to the selected data source. In advance of making a final decision about which data source to use, several platforms were

compared, and the respective details regarding the ethical and methodological implications thoroughly explored. Those considerations are described in the sections that follow.

Given the abundance of available technologies, and the prevalence of locational affordances in those technologies, the selection of a research site for this project was not straightforward; the final decision was taken after an extensive review of several platforms and a comprehensive review of the ethical implications inherent to the use of the platform that was ultimately selected (Twitter). The selection of a single appropriate research site began with the compilation of a list of several available digital platforms, each of which was then assessed for suitability using a six-point researcher-designed metric based on previous research (Boulianne 2015 and Frances 2014) and an industry technology report (Milanovic 2014).

3.2.1 Shortlist of Relevant Digital Platforms

A small sample of social media technologies was selected and assessed as a potential data source. This sample of social media technologies was compiled based on platforms included in Boulinanne's meta-analysis of social media platforms (Boulianne 2015), Frances's review of the life cycle of social media platforms (Frances 2014), a 2015 list of "the world's most important social media sites and apps in 2015" (Milanovic 2015) and a few other platforms incorporating some location-based functions that were known to the researcher at the time the assessment was started. A total of 32 platforms were assessed.

Boulianne 2015	**Frances 2014**	**Milanovic 2015**	**Researcher**
Facebook	Facebook	Facebook	Facebook
Twitter	Twitter	Twitter	Twitter
YouTube	YouTube	YouTube	YouTube
Google+		Google+	
MySpace			
	Flickr		Flickr
	Instagram	Instagram	Instagram
	Snapchat	Snapchat	

Boulianne 2015	Frances 2014	Milanovic 2015	Researcher
	WhatsApp	WhatsApp	
	Tumblr	Tumblr	
	Picasa		Picasa
	Badoo		
	LinkedIn	LinkedIn	
	Digg		
	Hyves		
	NetLog		
	PX500		
	Photobucket		
	Vimeo		Vimeo
		VK	
		Pinterest	Pinterest
		Vine	Vine
		Xing	
		Renren	
		Disqus	
		LinkedIn Pulse	
		Twoo	
		MyMFB	
		Meetup	
		Secret	
		Medium	
			TripAdvisor
			YikYak

The six attributes against which each potential source was assessed, chosen by the researcher were: Key features, Data/Intent, Spatial Information, Demographics, Access, and Ethical Considerations. Key features were defined as the defining characteristics of that platform, this could take either the form of a technological limit (like a character limit for example) or be about the intended use of the platform. Data/Intent was populated through posing the question “What appears on this platform?” and provided the opportunity to describe the general

data types that might be encountered either by the researcher or a user during their interaction with the platform; these included text, images, metadata, coordinates, video and audio. This descriptive section was deliberately kept separate to the ethical implications of sourcing data; the ethical considerations field was populated last, in order to allow for time to reflect on the unique properties of each platform. The Spatial Information unit was divided into three sections, metadata, user profile and content. These categories facilitated the breakdown of the potential future research process, outlining what locational data could potentially be derived from taking different approaches to data gathering, each focussing on “where” in the platform spatial data might be found. Twitter was eventually selected as the most suitable data source. The rationale for this decision is explained in the following section.

3.3 Challenges and Opportunities specific to the context

Twitter was only one of three social media platforms that appeared in all three externally generated lists. Twitter was chosen as the data collection site over other platforms for both practical and theoretical reasons. The contributing factors are explained here.

Research work concerning Twitter is highly varied – ranging from the intensely qualitative approaches (Marwick 2013) and (Brock 2012) to highly technical mathematical analysis of news (Pearce, et al. 2014) and the tracking of the spread of information through networks (Sakaki, Okazaki and Matsuo 2010). Murthy (2013) provides a helpful overview of the platform; explaining it as a “communications medium” (Murthy 2013, 1). While Ratkiewiczet el. (2011) refer to the platform as a “social networking and microblogging” service (Ratkiewicz, et al. 2011, 249) and Marwick (2013) focusses on the platform’s connection to other services, claiming “Twitter exists as part of an ecosystem of communicative options” (Marwick 2013: 112). Twitter text does not follow the same rules of composition as texts that have long been the focus of linguistic enquiry. Most tweets cannot be broken down into neat units of meaning and distinguishing key features of the text would be

equally difficult – not all tweets contain nouns, adjectives, subjects or objects.

Twitter is an unusually structured communication platform. Though most online platforms have particular conventions and encourage certain modes of participation (Foursquare focusses on “check ins” and promotes competition between users, whilst Snapchat limits the time a message is displayed) Twitter is slightly more restrictive than most in that it limits the size of individual contributions (to 140 characters at the time of data collection).

In the iconic work on digitally mediated socialisation, *Coming of age in 2nd life,* Tom Boellstorff (2008) takes a classically anthropological approach to digital technologies. Making no real distinction between online and offline realms, the book explored the ways in which users participated in “online worlds”. It takes the view that neither the digitally connected world nor the analog (disconnected world) be privileged over the other (Boellstorff 2008: 13). Miller and Slater (Miller and Slater 2000) also argue that online and offline spaces and connections should not be treated as independent spheres. They are, according to Miller and Slater “continuous with and embedded in other social spaces” that “happen within mundane social structures and relations that they may transform but they cannot escape” (Miller and Slater 2000: 5).

3.3.1 140 character limit[15]

The 140 character limit of tweets (as it was at the time of data collection) ensured that the sample was made up of a large number of small texts, which increased the chances that the sample would include example tweets from a number of different contributors. The brevity of a tweet (as opposed to a personal blog post) pushes users to be concise or creative in the expression of their thoughts and experiences. Based on literature that suggests affect and performativity are key elements of authorship choices on Twitter (Brock 2012) (Bruns and Burgess 2011) (Clark 2016); it was assumed that the desire to maintain brevity whilst

[15] The permitted character length of tweets was expanded to 280 characters after data collection was completed. See section 3.3.1.1

developing compelling narratives (affect + performativity) would be reflected in the editing choices of users, and that locational reference would emerge only if it was deemed to be a central or significant aspect of the tweet narrative.

3.3.1.1 140 – 280 characters

A few weeks after the final phase of data collection, Twitter extended the limit of 140 characters per tweet to 280 characters. This has not been explored at length in the discussions of sample data as the data includes only tweets composed in the era of the 140 character limit. Although this change in the medium will no doubt impact the way in which information travels through it, the discoveries and insights of this research work may still speak for the medium as it was at the time of 140 characters. It provides an opportunity to reflect on the temporality and coevolution of human sociality and technology in the age of ubiquitous technology.

3.3.2 Hashtags

The prevalence of the hashtag on twitter was one of the primary reasons influencing the eventual selection of this platform over others. The hashtag, as it is now in the eyes of some an “overused word used by hipsters to add tags to everything they do” (FrenchSwan 2015) first found a function on Twitter, and is still commonly encountered in tweets. Originally performing the functional role of indexing a conversation for improved searchability, according to Brian Solis (Solis 2011), their use and cultural importance has evolved over time. What was once a functional tool has transmogrified into a socially and culturally important symbol, whose use extends far beyond the reach of the platform (Solis 2011 and Lee 2005). The embedding of his symbol into the body of a single text or story, sometimes extending beyond the limits of the platform itself and drawing attention to trends, allegiances, current events or, in the case of this research, revealing the relationships between objects such as people and location, for example.

Social media platforms, when combined with the location sharing capabilities of new technologies, have produced a large amount of data sharing that is attributable to an exact point and space in time. The hashtag is a particularly interesting new feature of language and its prevalence both online and offline (Lee 2015) suggests that it plays an important social and cultural function. Several scholars have referred to the hashtags association with memes[16] (Ratkiewicz et al. 2011; Kamath et al. 2013).

Hashtags are a particularly prevalent linguistic form used in tweets. The hashtag came into popular usage in 2007 when it was used to communicate online about forest fires in San Diego (Solis 2011). Since its debut, the hashtag has become increasingly pervasive in online communications and is now also found in the offline world. It is not unusual, for example, to see hashtags used on t-shirts, mugs, signage or other everyday objects. The hashtag has become ubiquitous; it is a way of referencing cultural trends or providing commentary on the same, it exemplifies the process of imposing order and meaning through narrative practice described by Riessman (Riessman 1993). #firstworldproblems and #stayclassy are two examples of hashtags that have been used satirically. Many hashtags are simply classificatory, and they do not necessarily have meaning beyond their descriptive capacity. The difference between classificatory and meaning-making hashtags is explored throughout the analysis sections in chapters four through six.

There have been many studies which consider the use of hashtags on Twitter. Pearce et el. (2014) researched hashtags associated with the Intergovernmental Panel on Climate Change (IPCC), using them to identify three distinct communities participating in discussion about the IPCC – these researcher-coded communities were supportive, unsupportive or neutral. Tweets containing the word “IPCC” were analysed, and a total of 5, 291 different hashtags were identified from

[16] “meme” refers to the internet phenomena here, rather than the broader framing used by Richard Dawkins (Dawkins 2006).

these tweets (Pearce et al. 2014: 3). The researchers identify the most commonly occurring hashtags as well as identifying the general theme of hashtags; from individual hashtags, some key topics were identified. Broadly, the researchers found that hashtags could be classified into four topics; hashtags related to science, hashtags related to political campaigns; hashtags related to geographical discussions and hashtags related to societal concerns and new technologies. (Pearce et al. 2014: 4-5). This discussion demonstrated how geographical and digital identity information is shifting the boundaries of public and private space online.

The hashtag is one major way that individual tweets become connected to broader themes and trending topics. The use of the hash sign to index or categorize a tweet is “an integral part of Twitter’s ability to link the conversations of strangers together” (Murthy 2013: 3), which is further demonstrated in studies of hashtags associated with social movements in response to a triggering event such as “#JeNeSuisPasCharlie” (Giglietto and Lee 2017) and “#WhyIStayed” (Clark 2016). This threads into earlier work by Scollon and Scollon (2003) concerning the utility of signs and symbols as devices of and victims of “emplacement” (Scollon and Scollon 2003, 142), whereby signs can be used to enhance or change the meaning of their contexts at the same time that these contexts are influencing the ways in which signs or texts are read or understood. This reflexivity and flexibility of signs and markers, including hashtags, is why they are so useful in the negotiation of personal and shared identity and why they are so important to understanding place in digital contexts. The language and linguistic features of social media, like the hashtag, are not isolated, but rather they exist as elements in a wider ecology of meaning-making and are ways or enacting particular identities and showing allegiance to and understanding of specific cultural tropes.

3.3.3 Data complexity

The data contained within a tweet as well as the data that appears as part of a tweet (the metadata) are both accessible to developers and

researchers through Twitter's API[17]. This rich source of data provided scope for the development of a research question suited to the complex subject of place. The flexibility of Twitter with regards to data access and available data types was not true of other platforms which promised varied, crowd-sourced locational data. TripAdvisor, for example, which was considered as a possible source of place-based information, does not offer the same access. TripAdvisor does not enable easy access for non-customer facing developers (TripAdvisor n.d.) and the site proved difficult to extract data from manually during early investigation (using the NCapture[18] function in Nvivo[19]). The different types of data available through the Twitter API were an important consideration in determining which data would be suitable for use in this project. The content data was differentiated in this case, from both metadata and user profile data. In spite of the availability of all three data types through the Twitter API, content data is the only type that frequently (every tweet) requires a conscious action in order to share that data, whereas the meta-and user profile data types are unconsciously shared.

The decision to use Twitter as a data source can be explained by two key arguments. Firstly, as explained in the literature review, Twitter exists as part of an ecosystem of personal media technologies that has been well-explored in several disciplines, which ensured the research was clearly embedded in an existing body of theoretical and methodological work. Secondly, the decision to use Twitter data was shaped by the extent and nature of the data available. The limitations posed by the use of Twitter are explored more fully in the discussion (chapter 7).

[17] Twitter's Application Programming Interface is a route into the data made available by Twitter. Conditions and barriers to access are discussed in sections 3.4.2 and 3.4.3
[18] NCapture is a web browser add-on from NVivo, designed to facilitate the capture and importation of data from webpages (https://www.qsrinternational.com/nvivo/support-overview/faqs/what-is-ncapture)
[19] NVivo is a qualitative data analysis software package https://www.qsrinternational.com/nvivo/what-is-nvivo

3.3.4 Rich Spatial Information

Location can be brought to bear in several ways on Twitter, in tweet metadata as well as within the text body of a tweet and through linked content including images. There is some uncertainty over the accuracy of spatial information contained in metadata and user profile data. The user profile data, for instance, is used in the calculation of monthly active users (MAUs) for Twitter's annual reporting and is presented with the caveat "our data regarding user geographic location for purposes of reporting the geographic location of our MAUs is based on the IP address associated with the account when a user initially registered the account on Twitter. " (Twitter 2014 annual report p 4)

3.3.5 Limited Demographics

Twitter claims to be a "global platform", and was available in 2014 in 35 languages. Twitter's 2014 annual report states that the platform had 288 million MAUs (Twitter annual report 2014, 1 split between the US and the "rest of the world" 63million to 225million, respectively (Twitter annual report 2014, 43).

The demographic spread of MAUs certainly seems to support the claim that Twitter is a global business, offering a service that is taken up internationally; with 78% of MAUs from outside the US in the three months leading up to the publication of the annual report (Twitter annual report 2014, 17) though the service is blocked in China, and not reliably available in several other countries outside Europe and the United States (Twitter annual report 2014, 17 – 19).

False, spam and inactive accounts present a challenge in the research and validation of social media research, but Twitter's annual report claims that, following "an internal review of a sample of accounts" (Twitter annual report 2014, 4) that "false or spam accounts" represent "less than 5%" of the reported MAUs. Alhough this figure is low and these accounts do not account for a significant portion of MAUs, false, spam and automated accounts remained a consideration.

Figures on the segmentation of Twitter users are difficult to find, but a few technology reports have attempted to profile the split of Twitter users. According to one such report by Ipsos Mori, UK membership in Twitter is split fairly evenly between men and women, at 54 and 46% respectively but there is a huge skew in the age of users, when compared to the base population of the UK; more than half of Twitter users (56%) are aged 34 or younger and only 10% of users are 55 or older (Ipsos Mori connect tech tracker Q4 2015 report). A 2011 conference paper further spports claims that the uptake and use of the platform is unequally distributed towards certain populations; Twitter users in the US tend to hail from more populous areas, with most regions of the mid-west underrepresented when compared to census data from a similar period (Lehmann, et al. 2011).

The demographic segmentation of Twitter users demonstrates the ways in which technologies meet the needs of particular groups of people, and that no technology can provide a data sample that might be considered to be representative of the entire human population. The limitations of the platform were investigated and considered alongside the other key features before the decision was taken to use the platform as the data source for this thesis.

3.3.6 Emphasis on the Everyday

Twitter is used at all times of day and demonstrated in its users a continual engagement with, and attachment to sharing the commonplace. The sometimes amazing mundanity of information shared over the platform is beautifully illustrated by the art project “All the Minutes”. This “Online Clock”, set up by the Dutch design and artist collective Moniker (@studiomoniker) pulls tweets from twitter (from the same developers API that has been employed in this methodology) that correspond to the current time. Every tweet captured begins with the phrase “it is xx:xx am/pm” and is displayed for several seconds, with several time-based tweets shown in every (minute) iteration (Moniker n.d.). The project draws attention not only to the transience of

information on the platform but also speaks to the everyday-ness of Twitter.



Figure 3.1: Screenshot taken from @AllTheMinutes

Figure 3.1: Screenshot taken from @AllTheMinutes, shows the ways in which apparently simple and mundane statements can be constructed and read as commentary on an individual's life. The tweets that appear as part of *All the MInutes* show how individuals are able, within the tight character constraints of Twitter, (47 characters in this case), to present a snapshot of their lives and offer up a note on the trajectory of that life. It is because of this duality that the project also raises questions about the ways in which Twitter text can be accessed, searched, filtered, isolated and appropriated into discussions and uses that lie beyond what uses might have been considering when they posted the tweet, or indeed when they agreed to the Terms of Service. I turn to these ethical considerations in the next section (3.3.3.7)

3.3.7 Continually shifting contexts

The identification and extraction of information about place from social media texts means the tweet data used in this research is ambient or latent in the sense that it is being used in order to answer questions not directly concerning the context in which the data was generated. The tweets that have been collected and analysed over the course of this research were not generated by individuals with an explicit interest in contributing to understandings of place, nor with the singular intent of generating representations of those places. The data here just "happens" to have a geo-spatial element (Stefanidis, Crooks and Radzikowski 2013, 321). What sets the approach here apart from previous studies of this type of social media data is twofold – firstly, the focus on textual content and secondly, the use of this information for purposes not related to modelling, mapping or for the development of a general architecture (Stefanidis, Crooks and Radzikowski 2013).

3.3.8 Twitter as part of an ecosystem

Twitter does not exist as an isolated platform – it exists as part of a number of social and technological resources available to individuals with internet connectivity. It has been described as “the poster child of the burgeoning movement in computational social science” (Stephens, et al. 2015) and is a common focus of interest for scholars. Twitter is an unusually structured communication platform. Although most online platforms have particular conventions and encourage certain modes of participation, Twitter is slightly more restrictive than most in that it limits the size of individual contributions (to 140 characters at the time data was collected), a limit now standing at 280 characters (see p. 73 and 74).

Research work focussing on Twitter, ranges from the intensely qualitative (Marwick 2013 and Brock 2012) to the intensely mathematical. These engagements include the analysis of news (Pearce et al. 2014) and the spread of information through networks (Cui et el 2012). Twitter research is prevalent in a number of academic disciplines, as well as being of interest to non-academic researchers. Murthy (2013) gives a helpful overview of the platform; describing it as a “communications medium” (Murthy 2013: 1), while Ratkiewiczet el. (2011) refer to the platform as a “social networking and microblogging” service (Ratkiewicz et el. 2011: 249). Marwick (2013) focusses on the platform’s interconnectedness, describing Twitter as part of an ecosystem.

Existing literature in the fields of geography, the creative arts and the digital social sciences (discussed in chapter 2) highlights the many ways in whichtext-based research in the digital context can prove difficult. The methodology was developed in order to provde space to acknowledge and overcome these challenges. The methodology needed to enable the the analysis of very large data sets without eclipsing individual utterances; facilitate the translation of coded or obscure language; and encourage the researcher to remain cognisant of the technological and social contexts of the texts.

The concept of layering was important not only theoretically (in the interrogation of the concept of place) but was also essential in guiding the methodological approaches that were eventually taken. Rampton suggests that the ideal analytical solution would be one where “linguistic forms, situated discourse and ideology” could “be analysed together without in any way conflating them” (Rampton 2013, 2), this overlaying of three considerations, he argues, would allow appreciation of the “total Linguistic Fact” (a term he borrows from Silverstein) (Silverstein 1985). CL provided a way of integrating data that was not consistently coherent; it offered a flexible approach that suited both the linguistic unpredictability of social media texts and the enormous size of the Twitter text data streams. The research design, which will now be explained, was developed in line with the “layering” idea; the research was undertaken in three key stages and a mix of qualiltative and quantitative methods were used.

3.4 Designing the Research

This section outlines the main steps taken throughout the research. The accessing and processing of data are explained as well as the overall design of the research project following the preliminary research described in section. The research work was undertaken in three key phases. The first stage of research was exploratory, and stages two and three were shaped by the insights generated by the previous stage or stages. Text was the primary focus throughout, and each of the three stages differed slightly in its sampling. Stage one aimed to understand how location is incorporated into tweet text; a text filter was used to ensure the sample consisted of tweets containing locational reference. Stage two aimed to understand how tweet text content might be influenced by the physical location of the user which required the use of a geo-location filter in contrast to the text-based filter used during stage one. The third (and final) stage investigated the use of the location (hash)tagging in twitter texts. This stage employed an intensive qualitative approach, focussing on the combined use of the hash symbol (hereafter, hashtag) and a place name; in this case,

#nottingham and #newcastle. These sampling decisions are discussed in the next section (3.4.1), where the selection of research sites is detailed.

3.4.1 Selecting (a) research “site(s)”

The tension between offline and online realms was consciously incorporated into the sampling method. The research “site(s)” in this project needed to reflect the simultaneously “real” and “imaginary” limits of locational reference in social media. The three-stage approach, split by sampling method, permitted the exploration of the “imaginary” city, through the analysis of locational-referencing in text (stages one and three) and also of the “real” city, through the analysis of geo-located tweets (stage two). The three-stage approach also exemplifies the design of research in alignment with the media at hand – each stage focussed on a different feature of the platform; during stage one this was simply the production of text on Twitter, during stage two this was the geo-location of tweets and in stage three this was the use of the hashtag.

As tweet data is extremely abundant, and can be quickly acquired, some decisions to limit the influx of information were taken. The most significant of these decisions was the one to focus on two English cities, of comparable size and demographics. Nottingham and Newcastle were selected as research sites due to the researcher’s familiarity with the former, and the comparable size and features of the latter.

The city of Nottingham was selected as one of the sites for this research for two primary reasons. Firstly, the demographics, connectivity and facilities of the city were appropriate. Nottingham city is a fairly dense urban area, with good mobile network coverage (4G) and good access to free Wi-Fi in public locations. As well as having a diverse residential population, the city also attracts tourists from throughout the UK and across other international locations. Additionally, the city is home to a mix of social, historical and educational venues of varying sizes, including several high-capacity stadia and theatres

(Motorpoint arena, Theatre Royal, Nottingham Forest football ground, Notts County Football ground, Nottingham County Cricket) and two universities. A secondary reason was the researcher's acquaintance with the city; the researcher can claim familiarity without nativity. In line with the thinking of Georg Simmel, knowledge of the group and continued habitation within the group without having been born into it offers the stranger a unique opportunity to observe patterns of behaviour in the group (Simmel 1950). Newcastle has similar features, which are detailed in section 3.4.7.

The stages complement one another and permit the comparison of sampling frames, each individually providing insight into different habits and patterns of Twitter users, and together clarifying which sampling methods might best suit different modes of analysis.

3.4.2 Accessing Twitter Data

Access to Twitter data is easy to gain. The Twitter API means that data is easily obtainable, subject only to the parameters set out in Twitter's Developer Agreement, the ethical guidelines of relevant bodies (internet research) and the researcher's own research department.

Though ease of access to data was a primary factor in shaping the research approach taken in this research, the implications of this access were not lightly engaged with, and the arguments for the use of personal media data in research are explored at length in the ethical considerations section that follows in this chapter (section 3.6).

The theoretical arguments that supported the choice of twitter as a data source are explored more fully in the literature review (chapter 2), but are broadly concerned with the focus of the platform on day-to-day interactions, a focus on immediacy (in-the-moment communication and broadcast) and the opportunity to examine short, locative texts from a number of individuals or organisations.

Twitter data allows researchers to explore an emergent narrative form and for the purpose of this project, will enable us to see how location is incorporated into online narratives and whether the information being

shared about a location can provide rich information about the experience and associations of those locations. Unlike non-digital stories, stories shared online belong not only to those who write them, or those who read them but also to the locations in which they have been written. Geo-tagging technologies facilitate this association. This project makes use of the volume of locational and textual data that has been facilitated by ubiquitous personal technologies in order to enrich our understandings of place - “What begins as undifferentiated space becomes place as we get to know it better and endow it with value” (Tuan 1977: 6).

The tweets that were compiled into corpora for this research were collected through the Twitter API search function. This function allows Twitter account holders to build applications of their own that can routinely take action on their behalf; this could take the form of automatically replying to tweets or simply alerting a user to the action of another Twitter user. The programme built for the first phase collected a sample of published tweets counting backwards in time from the moment the code was triggered.

3.4.3 Setting up an Application and obtaining authentication

The code was written in Python and was managed through Twitter’s “Application Management” site. (Twitter 2018). This site provides access to information that allows a script running on a local computer to access Twitter’s data streams through that user’s account. Each application provides four key pieces of information that must be included in the python script in order to allow it to collect a sample of data; these are the Consumer Key (API key), the Consumer Secret (API secret), the Access Token and the Access Token Secret. These bits of information permit the application to make data requests or act (post or follow, for example) on the behalf of the account holder, without the need to share their account password (OAuth n.d.). This research used OAuth2 authentication, which permits read-only calls for information to be made by applications. No editing or input was required from the researcher, so OAuth1 authentication (which is more complex) was not

required. This research used a single application, with the code being slightly altered at each phase, as the basic framework was the same. The codes used for each phase of the research can be found in the relevant appendices, and are annotated with explanatory notes.

3.4.4 Conducting preliminary research

Before the research design was completed, the Twitter API was used to investigate the myriad ways that location might be invoked in a tweet. This preliminary research began with analysing the metadata of a single tweet. A tweet is more than just characters on a screen; the metadata from a single tweet collected for this analysis is 109 lines long. It includes the precise time and date of posting, a URL, it allows us to see if the tweet has been favourited by other users, or if the tweet was a reply to or Retweet of a previous tweet. Each tweet has a unique ID, in the example case, this is an 18 digit number that uniquely identifies that tweet, and allows it to be searched for. Tweets can be flagged in the metadata as “possibly_sensitive”, which is a true/false attribute. The metadata also reveals the language of the tweet (“en” in this case indicates it is written in English). In addition to information about the tweet itself, the metadata also provides a record of the account and user that published the tweet. In this example, you can see that the user set up their account on the 31st of May 2009, it was a Sunday. We can see the text that they use to describe themselves on their Twitter biography, the name on their account, including title (e.g. Dr), their screen name and where they chose as their location when they set the account up (in this case that location is Nottingham).

A tweet does not belong to a single location; the metadata stores several pieces of information that locate it. The example tweet metadata provided (Appendix A: Metadata in a Single Tweet) includes the following spatial references:

'geo': None,

'location': u'UK',

'geo_enabled': False,

'lang': u'en',

'location': u'Nottingham '

‘time_zone': u'London'

‘place’:u’none’

The ‘text’ attribute is another potential source of spatial data, which in the example case does not include a spatial reference. This difference in content is explored in this thesis through the comparison of data collected in phase two with data collected in phases one and three. The username, the unique tweet ID and the profile description from this example tweet have been changed, for purposes of confidentiality. The example is meant to demonstrate the breadth of metadata tied to a single message, not to reveal anything about that user in particular.

Geo-coordinates are also a feature of tweet metadata, which is explored later in Chapter 3. The standardised nature of this frame of referencing is a valuable and widely-used referencing system, whose inclusion in tweet metadata serves to make Twitter data compatible with spatial information systems that are already in use, ArcGIS, for example (ArcGIS n.d.).

3.4.5 “twitter.search”

In all three phases of the project, the module “twitter.search”, which is part of the Twython[20] package, was used to refine the parameters. The data was collected through the standard search API, which “returns a collection of relevant tweets matching a specified query” (Twitter 2018).

Twython returns JSON[21] files and automatically saves the documents in the same location that the code is saved. Upon the completion of data gathering, each file was re-saved as a plain text (*.txt) file before being password protected through addition to a password protected *.zip

[20] Twython is a library for the API. https://twython.readthedocs.io/en/latest/

[21] JSON (pronounced like the name “Jason”) is an initialisation of JavaScript Object Notation. This is a human-readable file format, which enables the exchange of data “between a browser and a server” (https://www.w3schools.com/js/js_json_intro.asp)

archive folder in order to comply with the commitment made to keeping data secure and confidential.

3.4.6 Stage One: search according to place names

Using the twitter.search module, tweet text was collected during the month of December 2015. Tweets containing either “Nottingham” or “notts” (sometimes both) were collected. The programme was not left to run over a number of days automatically, but rather was run daily by the researcher over the course of several days. A total of 785 tweets were collected, 411 contained the word “notts” and 374 contained the word “Nottingham”. The date and time metadata fields were not collected at this stage – data for phase one was exclusively text.

3.4.7 Stage Two: search according to location (long, lat, radius)

The second phase of research aimed to investigate how the location of a tweet might impacts its contents.

Two geolocated corpora were compiled during approximately one week in August 2016. Nottingham and Newcastle were selected as the two sites of primary interest so these locations featured in the build of each corpus. Nottingham was selected as a primary site for data collection due to its proximity to the researcher’s funding school, whilst Newcastle was chosen for its similar urban, usual resident population size, 305,680 and 274,444 respectively (ONS 2016).

A central location within each urban population area (longitude and latitude) and a radius of three miles from this point determined the total search area for filtering tweets. For both sites, the three mile radius ensured that the majority of urban sprawl in each city was captured, as well as major areas of suburban development; in Nottingham this included the University Park Campus and Beeston as well as areas South of the river, which all



Figure 3.2 Geographic extent of British Isles Search

fall outside the official city boundary of Nottingham but are considered to be part of the city. Similarly, in Newcastle-Upon-Tyne this included urban and green areas to the north and south of the river. A third, much larger reference corpus[22] was compiled by expanding the geolocation search to encompass the whole of the British Isles, this corpus was compiled at a later date, over two days in the month of December 2016. The resulting corpora totalled in excess of 107 million words; with the two regional corpora averaging about four million words and the much larger British Isles corpus of more than 98 million words. The geocode search parameters for each location (which used a central coordinate and a set radius in miles) are shown in Figures 3.2 and 3.3; descriptions and corpus sizes appear in table 3.2.

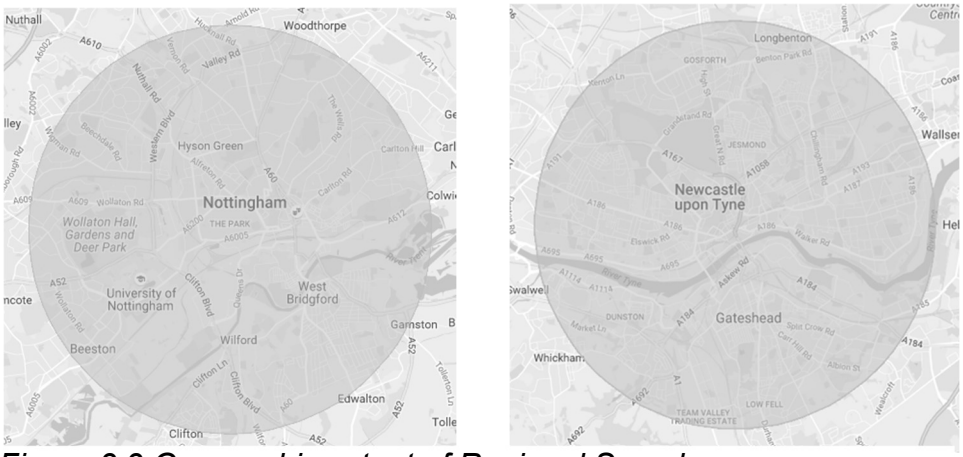


Figure 3.3 Geographic extent of Regional Searches

Each corpus compiled differed in terms of size in both words and tweets. The catchment areas for the regional corpora were the same, set at a three mile radius from a central coordinate while the catchment area for the British Isles corpus was set to be much larger, ensuring that both catchments from the regional corpora would be included in the sift.

Table 3.1 Description of Corpora compiled for Stage Two

Corpus	**Search parameters Coordinates, search radius**	**Size (words)**	**Size (tweets)**

[22] Reference corpora are used in order to compare subsets of language with eachother, and identify unique characteristics of these subsets (if they exist). See further explanation of corpus linguistic methods in section 3.5.7

Corpus	Search parameters Coordinates, search radius	Size (words)	Size (tweets)
Nottingham	52.95,-1.16, 3 miles	3,476,098	c. 248,293*
Newcastle	54.97,-1.61, 3 miles	4,844,863	c. 346,062*
British Isles	54.44, -4.04, 331.19 miles	98,779,803	6,954,678

Note: *the number of tweets in these corpora was not automatically counted, but calculated based on the average length of a tweet in the British Isles corpus.

The final stages of analysis focused on the use of the hashtags in each corpus. The aim was to find out whether the use of hashtags varied between locations, and if these patterns, where evident, revealed something about the nature of those locations.

3.4.8 Stage Three: search by hashtagged place names

Following on from the first and second phases, this phase aimed to investigate the role of intentional text (hash)tagging in tweets.

Data was gathered in a similar way to phase one, using a keyword search to filter texts containing either the #nottingham or #newcastle. The search ran for several hours and was manually terminated by the researcher once the target sample size was reached. The target sample size for each data set was 1000 tweets, but due to an unknown filtering error, 8 of the tweets from the Nottingham search did not contain the filter term and 7 of the tweets from the Newcastle search did not contain the filter term. Each sample (all of the tweets, even the ones that did not contain the filtering term) was assessed qualitatively as well as quantitatively in order to determine whether tweet text including tagged place names and other spatial references was able to indicate a difference in the nature of the two tagged places. Particular attention was paid to the invocation of other spatial reference in the tweets as well as the themes/subjects of the tweets. The quantitative analysis provided insight into the structure of the tweet text and provided two

smaller corpuses that could be compared with each other as well as the corpora compiled during phase two.

3.5 Processing and Querying the Data

Once searches, which could be monitored through the visible cmd protocol window, had completed (i.e. when the requisite sample size has been reached or exceeded), the resulting data file was moved to a password protected, compressed WinZip archive folder on a portable USB drive belonging to the researcher. The text data was only accessed by the researcher and when project supervisors were provided the files via email, the password was sent in a separate email.

In addition to copies of the data on USB drives, quantitative analysis of the tweets was made possible through the uploading of the data to the online CL software SketchEngine (Lexical Computing n.d.). Access to this online account was password protected, stored in the account of the researcher, which was individually licensed on a three-month rolling basis for the duration of the project. The password to this account was not shared with anyone, including project supervisors.

3.5.1 Mixed Methods

The combination of large and small corpora in this research project facilitated the combination of large-scale numeric and statistical linguistic analysis of tweet text with in-depth qualitative analysis of smaller sets of tweets. The quantitative analysis focussed on word frequency, concordance and the uniqueness of one body of text compared to another (keyword analysis). Qualitative analysis centred on the discovery of themes and examined a number of tweets on a case-by-case basis in order to explore and demonstrate these themes as they occurred. Further qualitative analysis focussed on two prevalent non-textual elements present in many tweets – attached images and hyperlinks to external content.

3.5.2 Ethnomethodology

This work is ethnomethodological in the sense that is has a single “site” as its focus (Twitter) and it takes an in-depth approach to the review

and analysis of a limited sample whilst incorporating reflexivity; the context in which the data is originally shared is not explored but the structural contraints of the technology are well incorporated into the analysis. Reflexivity is particularly important to this work, as it was consciously incorporated into the project design in response to the tendency for social media data to be swallowed up in “Big Data” rhetoric (Kennedy 2016), which can eclipse emotive and human perspectives. The integration of qualitative methods into this research was borne out of a desire to challenge the idea that large-scale, statistical analysis made possible by unprecedented access to personal and biometric information online is somehow redemptive and implicitly positive (this is discussed further in section 7.8.9).

3.5.3 First Time Through

The practice of “first time through” refers to the use of description in ethnomethodological research. This approach argues for the description of everyday activities and interactions in fine detail, to the point where a researcher treats every interaction as if it were occurring for the first time, recording it for posterity (Garfinkel 1967). The first time through approach is most commonly found in site-based, analog ethnographic studies, where a researcher (or participant observer) is embedded within a community and lives within that world for an extended period of time. That approach was not used here, but the focus on description and the day-to-day was retained in an iterative approach to coding[23].

3.5.4 Accounts & Indexicality

Garfinkel’s explanation of ethnomethodology insists that understanding of utterances is wholly contingent on the contexts in which they are produced (Garfinkel 1967). Garfinkel calls these ambiguous utterances “indexical expression” (Garfinkel 1967) and explains that they “present immense, obstinate, and irremediable nuisances” in the detection of meaning by a person not present at the time or place at which the utterance was first produced. From this perspective, the analysis of

[23] “coding” here is a reference to the iterative process of qualitative, thematic coding. See also section 4.7.

social media texts in research is packed with difficulty and the “nuisance” of these expressions calls for an adaptive, open and sometimes creative process.

3.5.5 Netnography

This research aims to answer questions about how individuals make sense of place through their interactions in social media. It combines a number of methods and theoretical models from several disciplines in order to achieve this. Ethnomethodology, which focusses on understanding the day to day practices of social actors in interpreting their lives (Garfinkel 1967); digital anthropology, which focusses on relationships between humans and technology and corpus linguistics; which provides a practical solution to dealing with textual data at a number of scales. The phrase “Netnography” has been borrowed from Robert Kozinets, and is used here to summarise the general attitude towards inclusive methodological practices, whose selection was guided by their emphasis on the social and technological contexts of individual actors rather than disciplinary boundaries (see discussion in section 2.3.13 for more on Kozinets).

3.5.6 Inductive Qualitative Model

The thematic analysis of the text data was undertaken from the desire to produce what Janet Hood, an advocate and practitioner of grounded theory (GT), has termed a “generic inductive qualitative model” (Hood 2007, 152), (hereafter GIQM) in which categories and overarching concerns are generated through an ongoing inductive and reflexive process. The method differs slightly from grounded theory in that the categories are informed by a prior literature review and that all the collected data is analysed and categorised, regardless of whether the data is deemed to be relevant to the development of a theory or not. This GIQM procedure was strongly influenced by the approach to GT advocated by Glaser and Strauss (Glaser and Strauss, The Discovery of Grounded Theory: Strategies for Qualitative Research 1967) as well as subsequent work across a number of disciplines that also make use

of this iterative and cyclical method. The common attributes at play here are:

- Inductive approach
- In-depth, line-by-line reading of text data
- Iterative process where categories emerge from the data
- Categories are open to change
- GT and GIQM are never "finished"

Given that the intention of this work was not to produce what (Glaser 2006) called a "Formal Grounded Theory", this work uses an approach that is similar to Grounded Theory, but differs in its reading of the entire sample and the inclusion of all results in categorisation, regardless of whether they fit with a framing theory or not.

3.5.7 Corpus Linguistics

Corpus Linguistics was identified early on as a disciplinary field with a lot to offer in the study of social media texts and the mixing of quantitative and qualitative methods. Corpus Linguistics describes the collection, analysis and annotation of large bodies of existing, real-world text for purposes of understanding the linguistic trends or habits in those bodies of text. The focus is on the use of language in existing and historical contexts, such as ancient Greek texts, in *Thesaurus Linguae Graecae* (University of California n.d.), or Newspaper content in the Zurich Corpus of English Newspapers (University of Zurich 2015). Corpus Linguistics (CL) is often discussed as a means of quantitatively analysing, usually with the assistance of statistical software packages such as R [24] (Desaguilier 2017, Gries 2009). It also frequently encompasses qualitative perspectives (L'Hote 2014, Olejniczak 2018) and is noted for facilitating the interweaving of both qualitative and quantitative approaches (Fruttaldo and Cambria 2017, Mackiewicz 2016).

CL facilitates the process of examining large texts as (more or less) coherent wholes and simultaneously permits the scrutinising of individual words, grammatical features and/or expressions and phrases.

[24] "R" is a programming language and environment frequently used in mathematical analyses. https://www.r-project.org/about.html

The zooming in and out (from single words to entire bodies of text) was essential to the identification of patterns in the data as well as in the understanding of these patterns, sometimes by enabling the interpretation of words and phrases unfamiliar to the researcher.

The analysis of large and disparate bodies of text has been facilitated by the development of corpus linguistic technologies. CL methods make use of digital technology to build and compile collections of language from particular forums or platforms. This approach has a technological focus – where the numerical processing of queries through technology can be used to answer both quantitative and qualitative queries about the use of language in real-world contexts. There has been a strong focus on Newspaper context and language use in institutional and educational contexts (Dafouz-Milne 2008, Baker, Gabrielatos, et al. 2008, Bednarek 2006, Conrad 2000)) but CL is an increasingly diverse discipline now concerned with a variety of communication platforms including social media.

Zappavigna explored the use of language in the creation and upkeep of social and community affiliation online (Zappavigna 2013) whilst scholars such as Knight, Adolphs and Carter (2014) have experimented with the construction of corpora of “digital English” that span several platforms (Knight, Adolphs and Carter 2014). Others, including Ruth Page have studied particular stylistic and linguistic forms of individual platforms (Page 2010, Bruns and Burgess 2011). This thesis contributes to discussion both about the utility of corpus linguistics as a method for engaging with interpersonal platforms like Twitter as well as the investigation of specific linguistic features that emerge in these contexts, for reasons of meaning-making or indexing, like the hashtag.

In addition to providing the technology to organise and store the text data used in this research, CL has also informed the methodological framework that has been used to interrogate this data. The questions asked of the text data in this thesis focus on the lexical components of online language rather than the grammatical features. A number of

different queries were run for each individual corpus, the detail of which is explained in chapters 4 through 6. The analysis of corpora included word frequency, collocation, keyness and concordance (explanations of these terms follow). Further considerations, not covered under these particular queries are addressed more completely in the bodies of the data analysis chapters; 4, 5 and 6.

3.5.7.1 Frequencies

The concern of frequency analysis is to count (and calculated the density of all the instances of a pre-defined term in a corpus. Frequency analysis provides an overview of the make-up of the written text and is an efficient way to ascertain how rich or poor a corpus is in terms of its inclusion of the specified search term. For instance, by running frequency analysis on place names or landmark names, it is possible to see, at a very high level, what the vocabulary of the searched corpus is. It demonstrates whether a word is frequently or infrequently used by tweeters in particular contexts.

In addition to searching for a specific term, it is possible to generate a "word list" for each corpus loaded onto the SketchEngine platform. A word list is a ranked list of words/tokens present in the corpus, according to their frequency. A word list can be tweaked in order to exclude certain words or phrases (a black list or stop list) and can be sensitive to, or ignore case-match differences, so that a capitalised form or a word may or may not be distinguished from a non-capitalised form, depending on the parameters set by the researcher. The word lists generated over the three stages of this project were non-case sensitive and the latter lists (which are the ones shown in this thesis) were generated with the use of a stop list (see Appendix B).

3.5.7.2 Collocation

Collocation provides insights into the habits of words, it tells which words commonly appear close to each other in text and can help identify emotional or cultural patterns in the use of language use. It was employed in the analysis of the Twitter text data to reveal patterns in the

occurrence of terms of spatial reference, like place-names and broader spatial terms like “area”. Collocation was ranked according to the LogDice calculation suggested by (Rychly 2008). The LogDice scores are not affected by the relative frequencies of words between two corpora, which made the calculation appealing for the use in research over three stages, dealing (particularly in stage two) with corpora of varying size. The equation used to calculate LogDice (which was generated by the SketchEngine software) is as follows:

$$logDice = 14 + log_2 D = 14 + log_2 \frac{2 f_{xy}}{f_x + f_y}$$

Where f_x = the frequency of word x, f_y = the frequency of word y and $2f_{xy}$ = the Dice coefficient[25]

3.5.7.3 Keyness

The identification of keywords through keyness analysis requires the use of more than one corpus. It is a comparative analysis that describes the uniqueness of words. A word that occurs uncommonly often in a specified corpus, when compared to another, larger corpus would show a high degree of “keyness”, and can be said to be a particular feature of that corpus. A reference corpus was compiled during the second stage of this research in order to undertake keyword analysis. The decision to compile a reference corpus specifically for this research, rather than making use of a pre-existing corpus, was taken for both reasons of access and scope, these reasons are described in section 5.3.3. Keyness is calculated using simple maths and is based on the comparison of normalised frequencies between two corpora, a target or focus corpus and a reference corpus. The equation used to calculate this is:

$$\frac{fpmTargetC + N}{fpmRefC + N}$$

[25] The Dice Coefficient is a statistical measure of similarity between two samples. (forums.fast.ai 2017)

Where *fpmTargetC* is the normalised frequency (frequency per million, fmp) of a word in the target corpus (TargetC) and *fpmRefC* is the normalised frequency of the same word in the reference corpus (RefC) (Kilgaririff 2009). N is used to ensure that dividing by zero is not attempted – this value is usually set to 1, a practice that was observed in all the keyword calculations undertaken for this thesis.

3.5.7.4 Concordance

A concordance is a way of looking at data – concordances allow researchers to view all the occurrences of a word (or other lexical item) as they appear in the text. The researcher can set the limit of this context, determining how many words on either side of the search terms they wish to view at the same time as the search word. Setting limits of -1 and +1 would show a 3 word list (assuming the search term is a single word). This view provides a quick look at context words and can provide some insight into the ways in which the word of interest is incorporated into language. Concordancing was used in all three stages of this project, but particularly in the second stage, where the reading of the data set was complicated by the size of the corpora, which meant that a full and thorough manual reading of the sample was simply not possible. Examples of the results of concordancing in this thesis are provided and explained in detail in chapter four, five and six.

3.6 Ethical and Methodological Challenges

As technology becomes more and more a part of an increasing number of human (and non-human lives, as Latour might argue), its roles and responsibilities come ever more into question. Social media platforms are an example of a technology that has become enmeshed in daily routines, they are now increasingly a part of a technological toolkit employed by organisations and individuals and serve as channels for many different types of interaction.

According to Hine (2005), the newness of new media technologies generated an “anxiety” amongst some researchers, who felt that these emergent social modes would be difficult to tackle and that research

involving them would lack replicability and validity. The anxiety of newness and of the unknown is certainly not new to researchers, so it is striking that Hine feels driven to mention it. In any case, Twitter is not really “new” anymore, 2005, in technological terms, is an eon ago. The precedents of previous research that did not exist when Hine wrote that summary certainly exist now for Twitter. There is a large body of work, spanning many disciplines and methodologies that this thesis builds upon. What follows in this chapter is an exploration of the key ethical questions that have been raised by the use of social media data in previous research.

Twitter is an open social media platform that can be joined by anyone with an email address and the technological know-how to set up an account. Users do not pay for the use of the platform but must agree to a long list of “Terms of Service” in order to set up an account. The platform allows users to control the readership of their tweets but offers a more limited approach to the sharing of tweets and metadata. The platform raises a number of ethical questions, many of which echo concerns experienced across all disciplines about anonymity, right to decline and participant consent. The ethical questions raised are tackled, in turn, beginning with an overview of concerns relevant to the specific complexities of Twitter.

Although speaking primarily about the ownership, appropriation and commercialisation of the artefacts and cultural practices of indigenous populations in the United States, Michael F. Brown, raises several issues that are relevant to the use of social media data in research and commerce. These issues include “determining proper owner”, the “right to decide to participate” and the complexity of the “flow of cultural elements” in an online environment (Brown 2003:3). Questions of ownership and the right to decide are frequently cited as ethical concerns in research, and they have been a concern throughout this research project. Twitter has not been treated simply as a data source or communicative platform, but as a connected environment that serves

as a hub for information sharing, the creation of shared narratives and the broadcasting of experiences and emotions of individuals.

Like many social media platforms, Twitter performs a social function, it is a tool used to facilitate conversations between people. Organisations use it to advertise their services; individuals use it to share experiences with friends and so on and so on in a network of connected communication. The ethics of using “found text” data from the internet, which is ostensibly in the public domain, is frequently challenged. The ethics of research projects based on this kind of information are not solely the concern of researchers or data scientists but have been voiced by government officials, representative of big business and private citizens alike (Roberts, Smith and Pollock 2004), echoing the concerns of Brown, provide a well-considered list of the “ethical issues of conducting qualitative research within text-based virtual environments” they outline several key issues central to research work online, namely “the blurred distinction between public and private space”, “the identification of researchers” the “processes for obtaining informed consent”, “protecting the confidentiality of research participants” and “maintaining the security of data collected” (Roberts, Smith and Pollock 2004, 157). I have considered these issues in turn. The security of data collected is addressed under “data handling”, in the description of the research process. Brown’s additional concern over the “flow of cultural elements” in any complex environment is addressed more broadly throughout the analysis sections, in which the invocation of linguistic features in tweet text by individuals, and in particular their use as devices of satire and humour, is explored.

3.6.1 Blurring of Public and Private Spaces

Lexi Webster cites Twitter as an “ideal context” for examining “complex phenomenon” in the internet age “due to its blurring of the boundaires between public and private” (Webster 2018). Stated simply, “there currently exists no general agreement over what constitutes public and private space in these [digital] environments” (Roberts, Smith and Pollock 2004, 161). Twitter has “terms of service” that make reference

to the public nature of the content posted by individuals (https://twitter.com/tos?lang=en) but the expectations of poster, as to what constitutes “public space” online may vary. This issue has been raised by other researchers using online forum data, including Hasler and Ruthven (Hasler and Ruthven 2011). The perception of how “public” a digital space is considered varies depending on the individual, the context (forum set-up) and the topic being discussed. Sensitive issues raised on Twitter, including those dealing with race, religion, gender or mobility should be treated with particular care and not attributed to an individual (even when anonymised) or quoted, as there is more risk to the individual in these cases.

3.6.2 Identification of Researchers

As this research collects data that has already been contributed to the medium and has involved no input or prompting from the researcher, my identity (full or otherwise) was not, at any stage, revealed to participants. There are some who point out the benefits of not announcing either your presence or your interests when commencing a study of human interaction, online contexts certainly make this presence easier to conceal. According to Sveningsson, “the internet makes it possible for us to collect data without being noticed. It is easy for researchers to gather and store material…The opportunity of doing observations in online environments without informing the individuals who are under study decreases the risk that the presence of a researcher influences the natural flow of the environment, and, thus, it lets us observe cultures as they normally are.” (Sveningsson 2004, 47). Not only does the online environment make concealment possible, it makes it possible to study interaction whilst being sure that the researcher has had no impact on the outcomes of the research work. This being said, it is possible (though not evident) that uses are aware of their potential involvement in a research study at any time, and that they naturally curtail their interactions at all times via the platform in order to address this known unknown.

3.6.3 Consent

Access to data via Twitter is publicly available. Below is an excerpt from the organisation's Privacy Policy (https://twitter.com/privacy?lang=en)

> *"The Twitter Services broadly and instantly disseminate your public information to a wide range of users, customers, and services. For instance, your public user profile information and public tweets are immediately delivered via SMS and our APIs to our partners and other third parties, including search engines, developers, and publishers that integrate Twitter content into their services, and institutions such as universities and public health agencies that analyze the information for trends and insights."*

The wording here is quite clear – it does not state that data *might* be disseminated externally, but states that it happens, "broadly and instantly". There can be little doubt amongst those who read the Terms of Service, as to what this means for the information they share over the platform. What raises questions is how many users actually bother to read these terms. Consent for this study was not sought; rather, user agreement to Twitter's Terms of Service and Privacy Policy, which clearly states that data shared via the platform is owned and freely shared by Twitter, served in lieu of formal consent gained by the researcher. These lengthy Terms and Conditions and user Agreements are not only inaccessible in terms of their literacy level but also because agreement to their stipulations is necessary in order to use the platform at all (at least on the side of the user). The only choice available to the consumer at the terms and conditions pages are not really just "agree" or "I do not agree" as the given options would suggest, but rather "I choose to participate" or "I choose not to participate". The real outcome of selecting "I do not agree" at the end of the terms and conditions is that are no longer able to join a network of other individuals in a specific communicative realm. Not agreeing to statements of terms and conditions is potentially very socially and culturally limiting, and may not really seem like a choice at all. Determining whether an individual is a

willing contributor to an open repository of data, or not, is therefore extremely difficult.

3.6.4 Confidentiality

Username metadata was not collected nor stored, so each tweet can only be traced back to a user through searching of the text in Twitter directly. As it was “not necessary to know what the users call themselves”, in which case guidance from the Swedish Research Council suggests that “usernames can and should be changed” (Sveningsson 2004, 53), usernames in any examples given in the analysis have been changed in order to protect the identity of users.

Steps have been taken to ensure that the data gathered have been treated with the appropriate care and fully anonymised where necessary, so that users are not individually identifiable, whether that be through what they said; when they said it or to whom they respond. Access to personal information about identifiable individuals without their knowledge or consent is central to the ethical concerns of this project. I have not disclosed the names of any individuals whose names emerged from any analysis, and did not, at any stage, seek to find out any identifying information.

To protect the anonymity of individual users, user names in example tweets (anything proceeded by the ‘@’ symbol) have been obscured and standardized when full tweet text has been quoted, in the instance of a username being included in a keyword or frequency list, the usernames have not been altered, as they are no longer connected to the text content of a tweet.

3.6.5 Ownership

The legal terms and policies that dictate the nature of users’ relationships and Twitter’s relationship with users’ personal data, under contract with Twitter are readily available online. Though the legal documents provided by (or made available online) by Twitter describe the processes by which your data is made available to third parties, there remains some room for interpretation.

As this study makes use of information that is already “out there” in the public domain, participants will not be able to withdraw their data at any time; they will have never explicitly contributed it to this project. Users retain the rights to their own data but do not retain any control of the data.

The Terms of Service state:

> *“You retain your rights to any Content you submit, post or display on or through the Services. By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods (now known or later developed).” (*https://twitter.com/tos?lang=en*)*

Users who post retain the rights to the data they “submit, post or display”, but these rights are balanced by the conditions stated above.The licensing conditions mean that Twitter is free to use the data in any way they see fit and can publish it via any channel they deem suitable, including any medium not yet invented at the time you post. These licensing conditions mean that Twitter is legally free and provides researchers and other interested parties full access to user data. The licensing of data, in this case, is governed by the Twitter Developer Agreement, as information was gathered through an App set up and managed by the researcher. As a “developer”, in this sense, my relationship with the data is limited by the conditions stated in that document and includes the proclamation

> *“you expressly acknowledge that Twitter, its licensors and its end users retain all worldwide right, title and interest in and to the Licensed Material and Content, including all rights in patents, trademarks, trade names, copyrights, trade secrets, know-how, data (including all applications therefor), and all*

proprietary rights under the laws of the United States, any other jurisdiction or any treaty" (Twitter Developer Agreement).

3.6.6 Consideration of Institutional Guidelines

The Nottingham University Code of Research Conduct and Research Ethics "requires all staff and students engaged in research to maintain the highest standards of rigour and integrity in the conduct of that research" (University of Nottingham CRCRE: 3). All stepspossible have been taken in order to conduct this research professionally and have collected, stored and handled data in line with the standards expected by the University and laid out in guidelines for research in online environments.

Given the contested nature of what constitutes fair use of personal data online, I have made an effort to build sensitivity and awareness into this project. The two-pronged approach, incorporating both qualitative and quantitative methods and frameworks of analysis, provided the necessary space to consider the implications of drawing definitive conclusions from either of the approaches on their own. The statistically, numerically driven analysis of the large bodies of text has been tempered by the consideration of single tweets, with the aim of providing a balanced view of the ways in which the platform can be used, and the myriad ways it can assist in our understanding of the consumption of physical space in a digital realm.

3.6.7 Ambient data, Expectations and Reasonable Use

Ambient data, as explained briefly in the literature review, is the use of data outside the context in which it was generated, the distinction from other types of data is one of application; ambient data is applied in contexts that lie beyond the scope of their generative platform (in this case Twitter). The separation of data from the context in which it was created is problematic in a number of ways. Ethically, it raises questions about the scope and continuing applicability of contractual agreements between platform providers and platform users, which were made in accordance with the platform's terms of use and privacy policies. Can a

user reasonably expect their data to be extracted from the platform for use in other contexts? That has not been explored in this research, but the question has been kept in mind.

In addition to the ethical questions raised, methodologically, the distance between gathered and stored data and the context in which it was generated can be great, and can lead to difficulties in the interpretation of results. Key to this thinking is the argument that the context in which data is created must also be considered by researchers. In just the same way that the ways in which data is gathered, stored and analysed, the conditions under which the data is created affect the ethical considerations that must be undertaken, this is just as true for social media research as it is for research that engages directly with participants. I have turned to an analysis of “the crowd” in order to explore the ethical issues particular to the use of ambient data.

Defining the “crowd” in relation to Twitter research is complicated by the fact that activity on these platforms is not driven by a single uniting aim, nor are users acting in pursuit of a common goal. Rather, the day to day reportage happening through the platform can coincide with the shared interests of other users, as well as being triggered by shared experiences, interests and agendas of other users also using that platform in a shared timeframe or geography.

The properties of “the crowd” come into play in this research in the sense that the data available through Twitter’s API is being “put to use”. It is purposefully extracted and used in ways that may benefit the researchers. Returning to the definition of the crowd provided by Estelles-Arolas and Gonzalez-Ladron-de-Guevara (2012), two of the three primary ways of delimiting a crowd are particularly relevant to the use of ambient data. These are, “what it has to do” and “what it gets in return for its work”. Twitter users are a crowd in the sense that they must all agree to terms of service and provide basic profile information in order to register, but this identity is extended when data from Twitter leaves the confines of the platform. When data is picked up by

researchers and applied in other contexts, what the crowd has “to do” changes. In the case of the first stage of this research, for example, the crowd needed to fulfil the criteria of including either the word Nottingham or notts in the text of their tweet. They were not made aware of this additional parameter and members of the newly defined crowd did not “get” anything in return for their work.

3.7 Conclusion

The methodology employed in this research combines qualitative and quantitative methods. It provides insight into data that is available in huge volumes but that is generated by individuals with no fixed aims or perspectives and is often fragmented. Twitter was selected as the main data source after an extensive review of the social media landscape. The five primary reasons for selecting Twitter are were both theoretical and practical:

- The iterative structure of feeds
- The focus on experience and the present in tweet authorship
- The platform’s inclusion in several meta-analyses of Social Media Platforms (section 3.2.1)
- Ease of access and availability of data (section 3.4.2)
- Established body of published literature concerning Twitter (section 3.3.8)

The tweet-by-tweet (rather than user or topic focus, for example) evolution of data streams from Twitter facilitated the inclusion of multiple viewpoints in the data sample. Additionally, temporality is implied in the platform’s prompt to its tweeting users “What’s Happening?” This suggests that users were being encouraged to navigate their present or ongoing experiences when tweeting. The inclusion of the platform in review lists and peer-reviewed literature across both private industry and academia indicated some robustness in the platform and a widespread belief that the platform is well-established and worthy of critical attention. Finally, the ease with which data could be accessed and gathered from the platform meant that a large amount of data could be collected and analysed with ease. Twitter offered a feasible opportunity to explore in-the-moment commentary

from a variety of stakeholders concerning the everyday experience of contemporary urban sites.

Chapter 4 Locational References in Tweet Text

4.1 Introduction

This first stage of data gathering was driven by the research question "How is locational reference incorporated into social media texts?" The focus is on describing the locational reference that is contained in the text of a tweet, and on the uncovering of three aspects of place singled out as important in the literature review. Twitter text data was collected, processed and analysed in search of expressions of experience, time and perspective.

This research draws on the insights of previous ethnographic research into intersubjectivity (M. Jackson 1998); to investigate the ways in which selfhood and identity is performed and negotiated in relationship to location in digital contexts. "Place", like self-hood, is not an objective entity, but rather a state of being that is continually being revised and renegotiated by different actors. The research was designed with this idea in mind and seeks to unveil the expressions of connection and identity with locations.

This chapter begins with a detailed description of the methodological set up of data gathering, including sampling and the Python script that was used. From there, the chapter moves to summarise data collected before finally moving to a discussion of the main outcomes of data analysis, which was split between methods from corpus linguistics, thematic analysis and qualitative assessments inspired by social and digital anthropology.

4.2 Overview of the Approach

This study accessed tweet data through the search function of the Twitter API. Tweets were collected based on their explicit referencing of a particular location. In this case, the location of interest was the city of Nottingham, England. Explicit reference to this location was identified through the inclusion of either "notts" or "nottingham" in the tweet text.

Multiple searches for these terms were run over a period of several weeks in December of 2015, running into early January 2016. The data

is summarised in two ways; through a quantitative summary and a qualitative deep-dive into tweet content and themes appearing throughout the text, before discussing the outcomes in relation to the research objectives.

The data present in any individual tweet was fully explored in the Methodology section in chapter Three and this stage builds on that understanding. As this research focusses on the 'text' of a tweet and not on any of the metadata, this first stage collected a sample of tweets based on the conscious invocation of location in the body of a text, as opposed to the ambient invocation of location through user or tweet metadata. The focus on conscious invocation of location was built into the data gathering method so that tweets were filtered and included in the data only if they contained either the word nottingham or notts. In spite of the many useful ways in which location could be detected from metadata fields, no tweet metadata was collected at this early stage, just tweet texts.

4.3 Data Collection

4.3.1 Selecting the research site[26]

Nottingham and notts were chosen as filtering terms because Nottingham is a formal and well-recognised official place name, whilst notts is a truncation often used informally to mean the city of Nottingham and sometimes as a shortening of "Nottinghamshire", the ceremonial county in which Nottingham is situated. (Ceremonial County Boundaries of England 1997).

4.3.2 Python Script

The code used at this stage was very basic, and ran only one search each time the script was executed. Twitter search results are limited to 200 tweets, so each time a search was run Twitter searched 200 tweets and returned a sample of those tweets containing the search terms outlined in the code, which meant that in order for the code to return more than a handful of results, it needed to be manually started on

[26] See also section 3.4.1 "Selecting a research site".

every day of data collection. The code ran largely without error and returned a useable sample.

The search filter was not case sensitive but it did match exact case, so different forms of these terms were only tolerated if they were a case variant; NoTts or NOTTINGHAM would be captured, for example, but any misspellings or deviations from letter order, such as notingham, nottignham or nots, would not be included in the search results. These limits also meant that extended forms of those words were not recognised as meeting the filtering criteria, so compounded words like Nottinghamissunny or Nottsderby would only appear in the search results if they had been used in addition to the basic filtering terms “Nottingham” or “notts”. This filter meant that some texture may have been lost from the data, as deliberate, satirical misspellings or other attempts at word play on place names were not included in the results, however, as this play was not the focus of the research, the potential absence of this was not deemed to be detrimental to the sample.

4.3.3 Timeline

The dates and times associated with individual tweets were not collected, though the dates of data collection and the number of tweets returned by the search run on that day were both tracked by the researcher during the data collection period.

These dates were as follows:

9th December 2015, 10th December 2015, 13th December 2015, 14th December 2015, 15th December 2015, 16th December 2015, 17th December 2015, 18th December 2015, 19th December 2015, 20th December 2015, 21st December 2015, 22nd December 2015, 24th December 2015, 25th December 2015, 26th December 2015, 27th December 2015, 29th December 2015, 30th December 2015, 31st December 2015, 1st January 2016, 2nd January 2016, 3rd January 2016, 4th January 2015, 5th January 2016.

Over the data collection period, covering 24 days, a total of 785 tweets were collected. A Twitter search was run on each sampling day at no

set time, so calls for information ranged from 05:00 GMT to 23:21 GMT. This sampling frame, which spanned the festive Christmas period, no doubt impacted the content of tweets. The interconnectivity of time and place is illustrated across several disciplines (Black 2000, Balmori 2003, Burgess, Klaebe and McWilliam 2010, DiNucci 1999). This time frame would also have influenced the events taking place in the city, which would include a Christmas market and a number of festive concerts focussed on the time of year.

Table 4.1 : Count of tweets collected in each sample, sorted by date of collection

Date	Nottingham	Notts	Total
9/12/2015	14	15	29
10/12/2015	57	59	116
13/12/2015	21	16	37
14/12/2015	21	24	45
15/12/2015	4	9	13
16/12/2015	17	17	34
17/12/2015	12	10	22
18/12/2015	1	0	1
19/12/2015	12	14	26
20/12/2015	18	17	35
21/12/2015	19	14	33
22/12/2015	26	31	57
24/12/2015	16	19	35
25/12/2015	11	11	22
26/12/2015	11	31	42
27/12/2015	6	0	6
29/12/2015	4	16	20
30/12/2015	16	20	36
31/12/2015	18	10	28
01/01/2016	12	14	26
02/01/2016	14	20	34
03/01/2016	15	13	28
04/01/2016	15	15	30

Date	Nottingham	Notts	Total
05/01/2016	14	16	30

4.3.4 The data sample

The sample set derived from the “Notts” search includes 411 tweets, whilst the sample set derived from the “nottingham” search includes slightly fewer, 374, for a total sample of 785 tweets containing either “notts” or “nottingham” or both terms. The total sample is 10,510 words, averaging about 13 words per tweet, which is about 80 or 90 characters in English, including spaces.

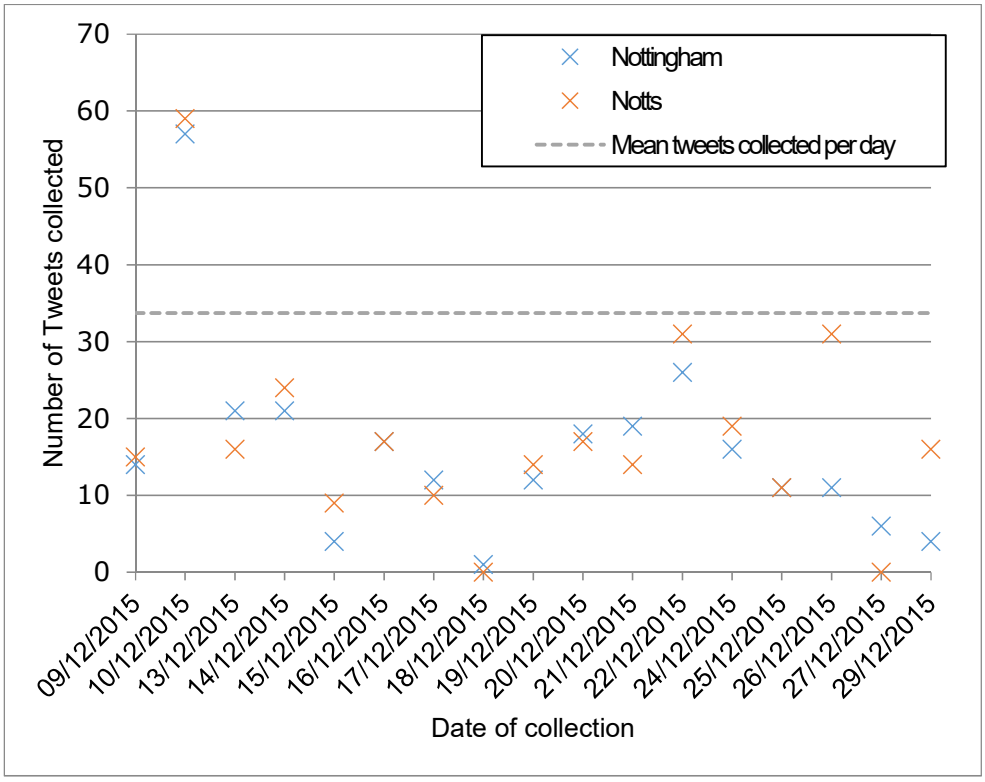


Figure 4.1 Chart comparing number of tweets collected under each search term

Figure 4.1 shows how the data sample was skewed by the number of tweets returned on the 10th of December 2015, compared to the relative quiet of the 18th and the 27th of December. The mean number of tweets collected per day was 33 (split between the search terms) but the mean

trend line clearly shows that the results were not consistent from day to day, over the 17 days of data collection.

4.3.5 Gaps in the data

On the 18th and 27th of December, the search was run for both search terms, just as it was on all the other sampling days but no results were returned for the “Notts” filter. The reason for the large variance in sampling return from day to day cannot be determined for certain, due to the functionality of the Twitter API searches, and the lack of direct contact with users. It is possible that the variance may also be at least in part the result of the fluctuating use of the search terms by Twitter users. The variance in the results returned for the same search terms between days may be due to increased use of the filtering terms by organisations promoting events.

4.3.6 Duplications

Due to the nature of the python script, and the way the script was run as two distinct calls for information, the sample includes some duplicated data. Where users had used both the term Notts and the term Nottingham in the text of their tweet, this tweet had the potential to turn up in the sample from each search, meaning that the same tweet was appearing twice in the overall sample, even when it was not a retweet. Retweeted content also led to the duplication of text content in the sample, but retweets were intentionally not excluded from the sample as retweeting is a key feature of the platform that users would expect to encounter in their day to day use of the platform.

4.3.6.1 Repeats versus Duplications

Some tweets occurred more than once in the sample, appearing either as repeats in the same subset (notts or Nottingham) or as duplications, where the same tweet appeared in both of the subsets – both the Nottingham and the notts results. The following tweet is an example of a duplicated result that appears in both the Nottingham search results and the Notts search results,

> *"RT @user: All sites closed today. #Leicester #Leics #LE #Nottingham #Notts #NG #Alfreton #Derby #Derbys #DE #ChristmasOpening htt…"*

This tweet, collected on the 4th of January 2016 contains both search terms "Nottingham" and "Notts". In addition to appearing in both search sets, the tweet is also a retweet. It is clear that not all content in the sample is original or unique, but this reflects the use of the platform, by including retweets and not excluding one search term from the filter used in another, we see content as it would appear to Twitter users which provides a more complete picture of the use of spatial reference in the text of a tweet. Overall, there was not a large amount of duplication between the two sub-samples. 85 tweets contained both "Nottingham" AND "notts" in one form or another, representing the mixed use of the filtering terms in the tweets. The terms themselves were sometimes combined with other words to form longer terms or phrases with no spaces.

4.4 Data Analysis: Corpus Linguistics

Following an initital review of the research results in order to identify duplicated results and to check that the filtering had worked as intended, the next step was to gather quantitative data about the sample. This quantitative analysis involved first establishing the size (word count) and the lexicon (number of unique words) of each corpus, then identifying common words, identifying common phrases (two and three word combinations).

4.4.1 Corpora Size and Lexicon

The two corpora had similar word counts and lexicon sizes. As capitalisation was neither a focus of the analysis nor a concern of the research question, variance in capitalisation was ignored for the calculation of word count and lexicon size as well as in the further corpus-based analysis that follows in this chapter, including frequency and keyword analysis.

Table 4.2 Corpora sizes

Corpus	Word Count	Lexicon*
Nottingham	4,996	2,379
Notts	5,514	2,324
Combined	10,510	4,028

**not case sensitive*

The lexicon of the combined corpus totalling less than the total of both of the individual corpora indicates that the Notts and Nottingham corpora share common vocabulary that amounts to less than 40% of the total lexicon. The difference in the total of each individual lexicon versus the combined lexicon indicates that the Nottingham and Notts corpora share 675 words, which amounts to 28% of the Nottingham corpus and 29% of the Notts corpus. The following frequency analyses reveal several shared characteristics in the text of the two corpora.

4.4.2 Common Words

The lexicon calculations indicated that the corpora had some shared vocabulary. To determine if the shared words were common in each corpora, a frequency analysis was run. This enabled the listing of the most common words in each corpus and provided a very basic overview of the language being used on the platform. The ranked frequency lists served several primary purposes; they confirmed that English was the primary language of the corpora, and that the keyword search filter functionality had worked as intended. Crucially, these lists indicated and that the language showed not only habits that reflect the specific parameters of the technology (the high incidence of RT content) but also features common to English language use off-line (the high incidence of function words, which are often added to stop lists).

Table 4.3: Top 10 Ranked Word Frequencies in Corpora

Rank	Word in Notts (Count)	Word in Nottingham (Count)
1	notts (204)	nottingham (198)
2	RT (152)	RT (131)
3	to (125)	the (113)

Rank	Word in Notts (Count)	Word in Nottingham (Count)
4	in (103)	in (108)
5	the (99)	a (108)
6	a (91)	for (76)
7	for (78)	of (62)
8	county (66)	and (59)
9	and (63)	at (53)
10	you (57)	on (48)

*Note: *figures checked 16/03/2018*
*** results are not case sensitive*

The frequency of “RT” indicates the prevalence of retweeted content in both sample sets and provides a clear picture of repeated content throughout the entire sample.

4.4.3 *Common Phrases*

Compiling lists of the most common two and three word phrases in each separate corpus was the first indication that tweet text followed expected patterns, they used prepositions in combination with locational reference and used phrases reflective of the time of year, “merry christmas”, for example.

Table 4.4: Top 10 Ranked Frequencies of 2-word phrases in Corpora

Rank	Word in Notts (Count)	Word in Nottingham (Count)
1	notts county (64)	in nottingham (40)
2	in notts (23)	nottingham forest (30)
3	at notts (17)	to see (14)
4	v notts (15)	of nottingham (14)
5	in the (13)	a good (12)
6	what a (12)	via https://t.co/29lmuwcclx (10)
7	to notts (11)	update via (10)
8	notts_acs for (9)	click to (10)
9	merry christmas (9)	check latest (10)
10	For the (9)	is a (9)

*Note: *figures checked 16/03/2018*
*** results are not case sensitive*

Word pairings common to many texts in the English language were prevalent in the text. “is a” and “in the”, “what a“ and “to see” all appeared in the multi-word frequency lists. The analysis at this stage did not include the use of any “black-“ or “stop” lists, so this is not unexpected. It was encouraging to see patterns common to other texts appearing in the tweet text data, a commonality that suggested there is some crossover in the habits of language use across platforms, and that this research could sit comfortably alongside other linguistic research work. The common frequencies of some of these two-word phrases suggested that they made up parts of longer phrases consisting of at least three words.

Table 4.5: Ranked Frequencies of 3-word phrases in Corpora

Rank	Word in Notts (Count)	Word in Nottingham (Count)
1	V notts county (13)	Update via https://t.c/29lmuwccl (10)
2	At notts county (12)	University of Nottingham (6)
3	What a nightttttttt (8)	Is a good (6)
4	With health issues (7)	Click to see (6)
5	Released after farmfoods (7)	This is a (5)
6	Police cctv image (7)	Nottingham investment fund (5)
7	Merry Christmas from (7)	In investment from (5)
8	Man with health (7)	Family ticket to (5)
9	Image released after (7)	A family ticket (5)
10	Hartlepool v notts (7)	N/A***

*Note: *figures checked 16/03/2018*
*** results are not case sensitive*
**** Only nine results were returned for this query*

4.4.4 Repetition and run-on phrases

The list of three-word phrases demonstrates a similar pattern to the one present in the table of two-word phrases, where word combinations with the same frequencies could be overlaid to reveal longer phrases that were made up of more words than the search threshold allowed (1, 2 and 3 in this case). For instance, the four-word phrase “A family ticket to…” was split into two three-word phrases as a result of the query run

on the data. The same was true of the five-word phrase “police CCTV image released after”, which was repeated several times in the Notts corpus. This splitting of longer phrases due to the limits placed on analyses by the use of the CL technology is a drawback of setting restricted parameters, and indicated that frequency analysis of multi-word phrases could be problematic if not reviewed carefully. The multi-word phrases were especially helpful in indicating what the most frequent tweets in the corpus looked like and hinted at their themes; “family ticket” suggested an available service, while “Merry Christmas from” highlighted the use of temporal framing in the contruction of Tweets. The most useful insight generated by the multi-word analysis was that the corpora consisted largely of unrepeated tweets. Of the total sample of 411 tweets in the Notts corpus only 13 contained the phrase “V notts county”, which was the most common multi-word phrase.

The frequency analysis provided a basic understanding of the textual make-up of the data samples, and although some locational information did feature in the lists of common words and phrases, this was due to keyword filtering. What remained to be determined was how the locational references themselves were being incorporated into the text; where they appeared and what information surrounded them.

4.4.5 Place names in tweet text

The text data was read and re-read in order to ascertain how the terms themselves were being incorporated into tweet text. The focus on the appearance of these terms in tweets was split between an investigation into the relative position of each locational reference in tweet text and an exploration of the common collocations (words appearing in the vicinity of the locational reference). The aim was to determine how narratives were contructed around locational reference, and to identify any trends in those constructions. The analysis focussed only on the specific filtering search term within each corpus, so that only the locational reference “Nottingham” was analysed in the Nottingham corpus and only “notts” in the notts corpus.

For the assessment of locational reference placement of the filtering term within the tweet text, each tweet from each corpus was divided into three sections, so that each tweet was broken up into a leading third (the beginning), a central third (the middle) and a final third (the end). Since the word length of tweets varied so much, and given the inclusion of links and word forms that are unique to the platform (which prevented word count being an effective way to split tweet length), the length of each tweet was calculated according to the number of characters. A combination of functions in Microsoft Excel determined the total number of characters in a single tweet (=LEN)[27], identified the position of the character in the tweet text where the target word first occurred (=SEARCH)[28] and then through the division of the search position by the total character length of the tweet (SEARCH/LEN), returned a percentage which was used to assign the word one of the three segments.

Table 4.6: Segmentation

Percentage (calculated to 2 decimal places)	**Segment Assignment**
$x \leq 0.33$	Leading third (the beginning)
$0.34 \leq x \geq 0.66$	Central third (the middle)
$x \geq 0.67$	Final third (the end)

[27] pre-defined function in Excel
[28] pre-defined function in Excel

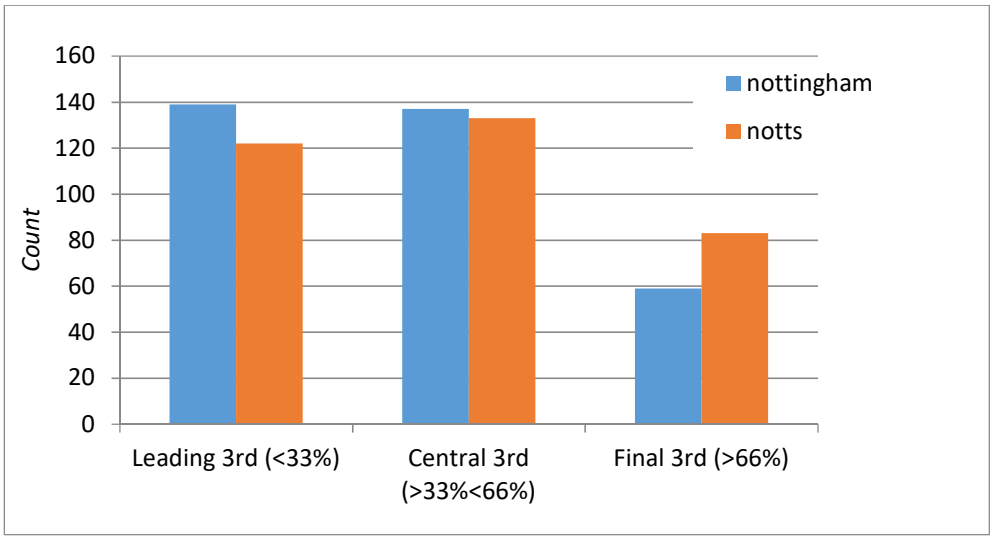


Figure 4.2: Position of filtering term in tweet text, split by thirds

4.4.6 "Notts" Collocates

A collocational analysis revealed which words tended to appear before and after the search term "notts". A minimum frequency of 1 in the corpus was selected for running analysis as the sample was small. 8 shows the words that tended to appear before the word "notts", in position -2 and -1. The collocated terms are ranked according to their LogDice score, which is an adapted version of the Dice calculation, commonly used in statistical analysis. The LogDice is argued to have better scalability between corpora (Rychly 2008, 8). The collocations shown in tables 4a and 4b (Appendix C) show the use of "notts" as a substitute for Nottinghamshire. Punctuation marks featured in collocations following "notts" (+1 and +2) but never before.

4.4.7 "Nottingham" Collocates

A comparison of the collocations at -2 and -1 shows similarities and differences in the use of the filtering terms in tweet text.

Table 4.7 Comparison of collocates preceding place names

Nottingham -2	**Notts -2**	**Nottingham -1**	**Notts -1**
university	Hartlepool	In	In
investment	back	Of	at
city	2	-	v

at	everyone	from	to
nottingham	about	,	from
wvk	in	at	MISSING
bolton	still	:	Jan
wolves	Carlisle	to	as
popular	for	near	for
park	all	V	the

The similarities between the words immediately preceding the search terms (in position -1) are more striking than those at position -2. The collocates nearest the place names demonstrate the utility of place names as relative spatial references. The dominance of prepositions in this position - “in”, “at”, “to”, “from” and “near” - demonstrate the tendency for both filtering terms to be accompanied by a word that indicates a spatial relationship with something else, whether this is a person, a site (a sports ground, for example). Collocates in position -2 tended to be nouns.

The meaning of “wvk” was not known, and typing the letters into a search engine did not provide any clarity, it was necessary to look back at the data in order to determine the context in which it was contained. It turned out to be the second half of a license plate:

> *“RT @user: mate had his pride n joy nicked last week, reg X866 WVK near Nottingham, could do with some R/T please. https://t.co/...”*

4.4.8 Words After

A side-by-side comparison of collocates in locations +1 and +2 is shown in Table 4.8.

Table 4.8: Comparison of collocates following place name

Nottingham +1	**Notts +1**	**Nottingham +2**	**Notts +2**
forest	county	fund	With
,	man	,	Fc

-	.	3	stars
panthers	,	:	;
and	and	2	,
post	police	nottingham	.
investment	-	derby	Rt
inggris	load	could	bennett
?	co	nottinghamshire	In
today	recall	a	V

The frequency and collocational analysis provided a general descriptive picture of the tweet sample, as well as hinting at the numerous ways in which spatial information can be incorporated into tweet text. Following this the focus shifted to the investigation of meaning and the content of individual tweets, the approach chosen for this progression was thematic analysis.

4.5 Data Analysis: Thematic Analysis

The sample of tweets making up the corpora described here is a very small sample of available social media text. For the purposes of the analysis, and the identification of themes, each tweet was treated as a small text in its own right and each was assessed for themes using the inductive methods (GIQM) (described in section 3.5.6, beginning on p.93). The assimilation of many tweets into a corpus also produces a text, but this compiled text is slightly different in that it contains many different voices, spans a longer time period and might reveal patterns in the ways those mini texts tend to be produced, the approach taken in analysis was informed by the narrative analysis of Page (Page 2010) and the interpretive approach outlined by Geertz (Geertz 1973).

Once a quantitative picture of the data had been produced using corpus linguistics, the research took up qualitative techniques in order to understand more about the content and context of the information produced as a result of the quantitative analysis. This analysis primarily centred on the designation of themes.

4.5.1 Two-Tier Thematic Sructure

Upon the completion of frequency lists, the tweet text data was manually reviewed and a list of common themes identified. The list of categories into which the tweets were sorted was compiled after a complete reading of the entire sample. The list was compiled by asking the question “What is the topic of the tweet?”. Unknown was included in the list as not all Tweets were easilt understood at first read through. The list consisted of the following categories:

Table 4.9 List of Primary Themes

Primary themes
Arts, Beauty, Building & Construction, Business, Charity, Crime, Education, Entertainment, Events, Faith & Spirituality, Food, Jobs, Animal, Travel, Healthcare, Home, Money & Finance, Promotions, Shopping, Sports, Technology, Weather, Sex Work, Heritage, News and Unknown.

The list compiled after a first reading of the entire sample was then used as part of validation process that ensured some consistency across the analysis, whereby a Primary theme had to be picked from this pre-existing list. Most of the primary themes were straightforward but the proportion of “unknowns” was striking, totalling 148 out of the total 785, representing about 19% of the primary themes.

Each tweet was also coded with a secondary theme, where possible. Secondary themes were freely assigned. The purpose of the secondary theme was to provide flexibility in assigning a topic to each tweet and aid in documenting the insights provided by that tweet; it provided the opportunity to document unstructured data in a ameaningful way. Maintaining this column as a “theme” meant that only one or two words were required, complex descriptions were avoided.

The most common primary theme overall is “Sports”, followed closely by “unknown”. The next most common theme, but at less than half the count of the second most common primary theme and making up about 9% of the sample, is “events”.

Sport was the most common primary theme in both the Nottingham and the Notts subsets, representing 20.6% and 25.5% of the total tweets respectively.

4.5.2 Unknowns

Tweets coded as “unknown” represent gaps in the researcher’s knowledge. 13 “unknowns” were foreign language tweets, and simply could not be read by the researcher. Other “unknowns” indicate either the absence of a suitably nuanced thematic category or the absence of the theoretical or social frameworks that might have allowed the researcher to “read” and understanding the meaning of the text. Where a tweet did was marked as an “unknown”, it was reviewed a second time. Several tweets marked in this way remained so, even after the second reading, as they did not fit into any of the pre-set categories, or in some cases as their meaning remained obscure. In some cases, it was possible to create new categories. One primary category added at the end of the initial review was “sex work”.

4.5.3 Secondary Themes

The secondary themes were not validated and were freely assigned by the researcher. As a result, there is a fairly broad scope here, demonstrating the richness of Twitter data. Even given that some tweet text appears more than once in the sample (as a RT or as tweets returned under both search terms); there is no scarcity of information.

The fact that sports was a common theme in both searches is not too surprising, considering “Nottingham” and “Notts” both form part of the area’s two major football teams, Nottingham Forest and Notts County. The prevalence of events and venues as a topic could be because these are more likely to be tweeted about by both organizations arranging and promoting the event and individuals attending and sharing their experience of the event. There are several events venues in the city of varying capacity that frequently host events and activities (Motorpoint Arena and the Theatre Royal).

4.5.4 Thematic Pairings

Patterns in co-occurrence of primary and secondary themes provide a detailed overview of the content of tweet text in the sample, without having to read through the text of individual tweets[29]. The table shows the primary themes present in the data alongside a list of the co-occurring secondary themes, the secondary themes are shown in no particular order, whilst primary themes are listed alphabetically.

The theme of travel was of particular interest as it provided some insight into two of the key attributes of place outlined in the literature review chapter (chapter 2), incorporating reports of both flux and experience. “Travel” tweets were largely about transit and transportation and concerned the operational state of various infrastructures (roads and delays on train lines). There was a focus in the Nottingham subset on the availability of travel cards for use on public transport (multimodal travel with tram and bus). The Robin Hood card, a cross-network travel card for use in the city of Nottingham was launched during the data collection period (14th December 2015) providing a possible explanation for this particular trend. Different content was returned as a result of each sub term which is clear from the table of primary and secondary themes. Similarities were also evident. Nothing good about the weather was shared in association with either Notts or Nottingham and Christmas was a secondary theme under a number of different primary themes. In general, each search term returned similar results. Sports and Events or Venues were the most common topics expressed in conjunction with either search term, with sports appearing more frequently than events in “Notts” tweets. Some primary themes showed an exclusive preference for one or other of the data samples, Sex Work, Arts and Faith and Spirituality had no secondary themes for the Notts sample, whilst Beauty, Charity and Money and Finance had no secondary themes in the Nottingham sample.

[29] Tabulation of these themes is available in Appendix C

4.5.5 Looking Further

In order to move beyond simply describing what, where and how spatial data is present in tweet text, it was necessary to look at individual tweets, and to assess the ways in which the experiential and emotional aspects of location were being expressed. Though the general meaning of a single tweet could be easily summarised through two-tiered thematic analysis, the nuances of spatial relationships were considered on a case-by-case basis. Concordancing was used in the first instance, in order to identify a sample of tweets that demonstrated a variety of content, style and length.

Concordancing within the online corpus management system facilitated the identification of individual tweets for use in a case-by-case analysis. This analysis was intended to determine the finer detail of tweet text and to unpack the meanings of the text, with a focus on the spatial data contained in that example.

Ten example tweets were selected by the researcher based on their content and the placement of the filtering term in the text. The examples that follow cover a range of themes, are of varying lengths and demonstrate a few of the challenges encountered in the case-by-case analysis of the tweets. Examples from each filtering search are included, as well as a tweet that does not include either of the filtering terms, as well as a foreign language tweet. The examples are intended to reflect the general content of the total sample collected at this stage and to give some colour to the trends that emerged through quantitative analysis of the sample.

All the examples shown here have been copied directly from the sample, with the text unedited except for the anonymization of usernames includes in tweet text (all username text replaced with the standard form @user).

4.5.5.1 Example 1: Notts County

Sports was a dominant theme in both corpora, the following example shows how users combine multiple location references in tweet text:

> *"Easy!...on this day in 1975 Notts County won 2-0 at home to Charlton in the Old Second Division #NOTTS #NOTTSCOUNTY #DayToRemember #OTD"*

The way that the user has mixed the use of place names with a reference to time reveals one way in which different timelines are woven together in the stories shared over Twitter. This tweet links the present date "on this day" with an historic event (a Notts county win over Charlton in 1975). The absence of an additional tagged "Charlton" helps to demonstrate the allegiance of the user; it clarifies Notts and Notts County as the primary foci of this broadcast and suggests that the mention of Charlton is incidental in the event of a Notts County triumph. This example shows the close association between "notts" and the sporting history of Nottingham. The Notts County football ground is located in the south of the city and the team has a long-established fan base.

4.5.5.2 Example 2: Notts treated you well

> *"@user Alright mate! Was off Twqitter all y'day so dodn't see this. Hope Notts treated you well tho, hopefully see u over xmas sometime!"*

In both examples one and two, "Notts" is invoked as both a thing with agency, obvious in both the phrases "Notts treated" and "Notts County won" as well as a place "#notts" in addition to "#nottscounty" in the first tweet. This example is one of very few instances where direct communication between Twitter users was apparent.

4.5.5.3 Example 3: #Nottingham

> *"If you are looking for some #blinds in the #Nottingham or #Derby area, why not get in touch for a quote. https://t.co/xGH1JK1ReW"*

In contrast to Example 1, this text demonstrates how tagged location is sometimes incorporated into the main body of a tweet, and is not

always isolated to either the beginning or end of a tweet. This message advertises a service available in the "#Nottingham or #Derby area" which serves simultaneously to distinguish the two areas from one another and link them together, so that a reader can ascertain that Nottingham is not part of Derby, and neither is Derby part of Nottingham, but that they exist together in a broader "area" that encompasses the two. This is a nice example of how descriptive words like "area" can be appended to place names in order to clarify what is meant.

4.5.5.4 Example 4: No sign of either filtering term

> *"@user Thanks for following us, join the discussion in the app! iOS https://... or Android https://..."*

This tweet, taken from the Nottingham data sample is one example of a number of tweets from the sample that contains neither the search term "Nottingham" nor "notts". The tweet is clearly from an organization offering services through an app. There is nothing obviously place-based about this; the text does not contain any place-names or any general spatial terms like "near" or "area". Worth noting is the use of the phrase "in the app", which is locational but refers only to a digital space. For reasons of scope, these kinds of details have not been assessed and the remainder of analysis does not make particular mention of them. However, they have contributed to an overall picture of the wide variety of ways in which spatial reference is invoked in tweet text and exemplifies some of the difficulties inherent to the use of open data streams.

4.5.5.5 Example 5: England's #HomeofSport

> *"RT @user: Congrats to #Nottingham, England's #HomeofSport! Visit the UK and learn #English: https://... https:// …"*

This retweet celebrates Nottingham city's accolade as England's "home of sport". The city was awarded this title in October of 2015 following a campaign by the tourism body VisitEngland (VisitEngland 2015). This award celebrates the city's rich history of sport, as well as recognizing the many sporting venues of the city, including Trent Bridge Cricket Ground and the National Water Sport Centre. The centrality of sports in the identity of the city and its environs is reflected in the themes extracted from the data. The addition of a hashtag to the mention of a place name was a pattern common to the rest of the data set and selected as a feature for further investigation (see Chapter 6).

4.5.5.6 Example 6: Non-English Language

> *"勧誘文章の中に緊急事態要項や積極的平和主義の事項が盛り込まれていた。初詣の大量人出を狙った日本会議の改憲活動に警戒が必要 https:/…"*

This tweet, taken from the Notts sample, shows how non-English languages, including non-roman alphabet languages, are present on Twitter. There were a total of 13 non-English language tweets between the samples. Foreign language tweets such as this were not translated and their meaning was not deciphered. All non-roman alphabet tweets were categorized as "unknown" as a primary theme and then further classified as a "foreign language tweet". The appearance of non-English language tweets in the sample lead to the alteration of code at stage 2, where results were additionally sifted for an "EN" metadata language marker.

4.5.5.7 Example 7: Duplicated Tweet with multiple locations

> *RT @user: All sites closed today. #Leicester #Leics #LE #Nottingham #Notts #NG #Alfreton #Derby #Derbys #DE #ChristmasOpening htt…*

This tweet was present in both the Notts and the Nottingham samples and, similar to example 3, links a number of locations together. This

tweet is particularly interesting because of the ways in which three sites have been expanded on with additional tagged locations. Leicester, Derby and Nottingham are the three cities mentioned; following the naming of each city, the user has included two or more additional locational references. In the case of Leicester, these are “Leics” and “LE”; for Nottingham, they are “Notts” and “NG”. In the case of Derby, these additional terms are “Derbys” and “DE”. There is a pattern here in the ordering of the spatial terms, that order goes “#city”, “#shortened form of the county containing the city”, “#first two letters from the city’s postal code”. Derby is slightly different in that there is a fourth term preceding the city name, “#Alfreton” is a town in Derbyshire, 20km from Derby and 21km from Nottingham.

4.5.5.8 Example 8: Happy Days

“RT @user: Leaving Nottingham today happy days”

This is a nice example of the incorporation of experiential and emotional information in a tweet that also make mention of location. Even though it is only five words long, this tweet tells a story. It situates the writer in both space “Nottingham” and time “today” and includes an expression of sentiment based on that situation “happy days”. The fact that this is a retweet “RT” is an indication that the content provoked reactions and triggered a social response in those people who read it and subsequently shared it (by retweeting it).

4.5.5.9 Example 9: City Infrastructure

“just checked in to park plaza hotel in Nottingham”

This tweet, although it has similar elements to the previous example (example eight), recounts a transaction rather than an experience. The absence of commentary may be telling, but a reader is left in the dark – as no linked information or other indicators of personal experience are included. Rather than communicating emotion or sentiment, this tweet simply situates the writer at a particular location “park plaza hotel Nottingham” and suggests a temporal framing; “just” suggests the

action occurred in the recent past. The text is light on experiential detail, providing only descriptors of place and time, which have both been narrowly defined. This tweet tells of (presumably) one person's arrival in the city of Nottingham and reveals only that they are staying in a hotel. The tweet is informational rather than reflective and is indicative of a broader trend apparent in the tweet sample gathered at this stage; The majority of tweets seemed to be formulated in a way that enabled them to be broadcast rather than shared – writers were making general statements or announcements tailored to particular audiences rather than asking questions or participating in dialogues. A great many of the tweets did not require responses, and were phrased in a way that allowed the information to be reacted to (liked, for example) or passed on (retweeted). The simple phrasing and formatting permitted tweets to be isolated and yet broadly understood; this type of tweet didn't reveal a lot about the experience or meaning of a place, but was helpful in highlighting the physical infrastructure of the city (Park Plaza).

4.5.5.10 Example 10: @user Notts

> *"Come and watch an okay band and then a shit film @user Notts tonight because it's for charity https://..."*

This is a good example of how usernames (@user) could sometimes be helpful in the extraction of locational reference. The user mentioned in this example, which has been anonymized, is, in fact, the handle of a local business occupying a single site in the city of Nottingham.

4.6 Locational Reference

In addition to the differences noted in thematic coverage of each sub set, the analysis also focussed on the filter terms, "notts" and "nottingham". In both sub samples, the search term was commonly used as both a single word and as part of a multi-word phrase. The single word frequency reflects the design of the data filter, while the two- and three-word phrases reflect the variance in the use of both place names. The phrase "v notts county" demonstrates the

commonality of a “notts” and “county” pairing whereas the use of “notts county” in two different ways “v notts county” and “at notts county”. These two invocations of the term are a first look at the variance in locational reference in Twitter text. “notts county” is simultaneously being used to refer to a sports team and the location of that team’s home ground. Locational reference is nuanced, and the ability to easily distinguish between one meaning of a name and another very different meaning may rely solely on one or two letters like “v” or “at

Frequency analysis revealed a high incidence of function words in the sample, which was not surprising but revealed that tweet text construction followed the same patterns as standard written English language text. This finding prompted the inclusion of a stop-list during the next stage of analysis (stage two and stage three). Frequency analysis of multi-word phrases started to unveil how spatial data was incorporated into the tweet text – in the case of the “nottingham” sample the place name was associated with education “University of Nottingham” and as a place to visit with family “a family ticket to”. In the “notts” sample, an association with sports was more apparent “v notts county” and “Hartlepool v notts”.

The collocation analysis supported what the frequency analysis had hinted at. The place name “Nottingham” showed a stronger tendency towards industry, work and education collocates; “university”, investment, city), whilst “notts” tended to appear with news reporting and sports; “Hartlepool”, “police” and “man”. Particularly revealing were the collocates at position n+1, which indicated that “nottingham” and “forest” are strongly collocated, as are “notts” and “county”. Both these collocations demonstrate the prevalence of sport-based communication on Twitter as well as hinting at the importance of identifying collocates in the investigation of spatial data and place-based identity in digital contexts.

4.7 Place

The accumulation of nearly 800 tweets at this stage, demonstrated that patterns and trends in data can be easily uncovered through computational analysis. Following the corpus-based analysis, each tweet collected was read through and assessed for theme. This process was inductive, with the list of themes created from the reading of the data rather than from any pre-existing thematic framework. This method led to the establishment of a two-tiered thematic framework, where each tweet was assigned a primary and secondary theme. The secondary themes allowed greater flexibility in coding[30] the content of a tweet, as these were "free", in the sense that they did not have to be broadly applicable, as the first tier was.

Appendix D shows the grouping of first and second level themes. "Sports" and "Events" were the two most common themes in both the "Nottingham" and the "notts" samples, which tallies with the findings of the frequency and collocation analysis. The prominence of events supports previous work that highlights the use of social media to communicate news and current affairs (Gil de Zuniga, Jung and Valenzuela 2012) (Russell 2017).

It was difficult to determine themes without reading through the text – the prevalence of truncations and slang, use of usernames and linked content all made the text more obscure to a casual observer – perhaps users know this and that's why so many people agree to Twitter's Terms of Service in the first place, demonstrating that data can be open and public without being readable or "accessible" to those outside the interaction or circle of communication.

4.7.1 Experience

The frequency of multi-word phrases demonstrated that the tweets compiled during this stage contained both transactional and experiential information. People were not only sharing information about how much things cost or where they were happening, but they were also sharing

[30] "coding" here is a reference to the iterative process of qualitative, thematic coding.

information about their experience of those events, locations or services. The emotional dimension of experience in the city is demonstrated in both example 1 and example 8 in section 4.5.5, where emotion is indicated in two ways; through the use of an exclamation mark and the phrase “Day to remember” in example one, and the phrase “happy days” in example 8. The combination of the exclamation mark, the topic of the tweet and the inclusion of the tagged phrase “#DayToRemember” suggests that this event is an important marker in the imaginary landscape of Nottingham. Example 2 highlights the flattening of the actor-subject relationship, where the city is given agency as an actor, rather than simply providing a backdrop for social or emotional action (Law 1992, Latour 2005, M. Jackson 1998).

The contrasting transactional and experiential elements are also highlighted in Table 4.5, where three-word phrases in both the Nottingham and Notts corpora indicate joy or appreciation. “what a nightttttt” was a popular response in the Notts search, whilst “is a good” was a frequently repeated phrase in the Nottingham search. The first terms demonstrates a degree of emotional response, through the emphasis of the last letter. The commonality of this specific construction (featuring many repeated “t”s at the end of the word night) can be attributed to the inclusion of retweets in the sample. The frequency of the construction reflects the popularity of the sentiment of the original expression amongst its audience, rather than the uptake of that expression by a number of users. The second indicates a reflection on quality. It is important to distinguish between these two responses but not to privilege one over the other; these information types complement each other and indicate the myriad ways in which people consume and narrate the spaces they pass through. The difficulty lies in finding a way to amalgamate these narratives in order to clearly understand the values that people attribute to particular places, and how these individual experiences might build up over time to give layered narratives of place.

4.7.2 Time

The role of time in the selection of tweet content was evident throughout the analysis, from the very first frequency query, through the thematic analysis and even in the examples examined closely in concordancing.

In section 4.4.3.1 a connection between time and place was particularly evident. The phrase “on this day” was used to connect present events with past events, thereby creating a timeline of events connected by both a sport (football) and by a location. The (hash)tagging of “OTD” (an abbreviated form of the phrase “on this day”), and its presence side by side with “#DayToRemember” indicates the emotional significance of this connection to Notts County fans (and possibly to Charlton fans as well). Memory and remembrance is a recurrent theme in literature exploring place (Agha 2007) (Burgess, Klaebe and McWilliam 2010) (Golden 2001) (Haskins 2007).

Reference to time was also demonstrated through well-wishing over the festive winter period between December 24th and the 1st of January – In the Notts corpus, the joint eighth most common two word phrase was “merry Christmas” whilst “new Year’s eve” emerged as a secondary theme from the Nottingham corpus.

It is clear from the frequency analysis that calendar dates and occasions play an important role in shaping the content shared about a particular place. Christmas (25th December) and New Year’s Eve (31st December) are widely marked occasions in the United Kingdom, having some cultural importance and it was surprising not to see either of these occasions appearing in either corpus. The mention of these festive events is further evidence that expressions of located experience are a common feature of social media texts.

4.7.3 Juxtaposition

The two corpora compiled at this stage of the research showed that location is a common feature of social media texts. Locational reference was incorporated into tweet text by both individuals and organisations, either for the means of advertising the geographic spread of services or

for the contextualisation of experience and opinion. Individuals and organisations tweet for different reasons and each perspective provides another way of looking at locational reference in online narratives. Not only does the intent of the user vary (for business or personal use) but the ways in which these user groups chose to incorporate locational reference also reveals something about their relationships to those places. The examples shown here, show huge variance in the nature and position of place names in tweet text and hint at the inclusion of locational reference in forms other than place names (notts or Nottingham).

4.8 Reflections

In the first stage of data collection, attention was focussed on the City of Nottingham. The filter built into the application that stored data directly from Twitter meant that the results were expected to include either one of the terms “notts” or “Nottingham”. Preliminary research into tweet metadata revealed that while tweet metadata was a potentially rich source of locational information, the decision was taken to focus on tweet text, and to exclude metadata from data analysis. Text data was collected over a period of 24 non-consecutive days in December 2015 and January 2016, resulting in the assembly of 785 tweets split between each filtering term.

After a thorough read-through of the data to identify duplicated or overlapping content, the text data was compiled into two separate corpora. One corpus containing tweet text returned through the “notts” filter, and the second containing tweet text returned through the “Nottingham” filter. Each corpus was separately analysed for common words, common phrases (multiple words commonly appearing together), and collocation.

The functioning of the code used to gather tweet text from Twitter at this stage was not optimal. Because the code required manually restarting for each day of data collection (rather than running automatically for an extended period of time), the sample of tweets gathered was fairly

fragmented, and did not cover a consistent time of day, nor did it include sample tweets from entirely consecutive days. As a result, running the data collection manually was not considered feasible in the pursuit of a larger data set, consisting of several million words.

Due to not following the chain of a tweet (its replies, comments and RTs) made it difficult to say anything about the exchange or the economy of the platform – other than the observation that it seemed very uni-directional. This is likely a reflection of the methodological approach – the sampling approach did not allow for the tracing of chains of communication, which prevents the exploration of these aspects of the platform.

4.8.1 Extent, Size and Quality of the Sample

The results of this stage's analysis lead to several alterations in data collection as well as data analysis. Applying these changes ensured variety in the data sample as well as the refining of the analytical approaches taken, through the incorporation of stop lists in corpus analysis in stage two, for example.

Although there are some clear trends in the frequency and collocation analyses, it is impossible to know whether the returned results are unique or significant without making comparisons to a larger or alternative corpus. What the initial study produced, in effect, was database of words and phrases containing either the word "notts", the word "Nottingham" or both words, specific to the month over which the text was collected (end of the calendar year).

The sample includes several hundred tweets, which has allowed for adequate variety as well as the fine-grain examination of the text. The quality of the sample was firstly assessed based on the size (in tweets) of the sample as well as the time period covered. Secondly, it was assessed based on the alignment of corpus design with the returned results. The main criteria for the corpus were that all tweets should contain either the word "notts" or "nottingham". Not all the tweets in the sample met these criteria. The "nottingham" filter returned 39 tweets

whose text did not contain the search term and the “notts” filter returned 73 tweets that did not contain the search term. Although time and date metadata from individual tweets was not collected, the temporal extent of the sample seems to accurately reflect the data collection period, as the topics mentioned correspond to seasonal events falling within the data collection period. Some concerns remained over the date-time distribution of the sample triggered by the missing out of two days in the data collection period. This was exacerbated by the variance in the number of results returned each time the search was run, which is shown in Figure 4.1. The number of tweets returned during the run of the code on each day and between the two codes (which varied only by the search term “notts” or “nottingham”) is fairly striking. The graph in Figure 4.1 clearly shows how the results are skewed because of a high number of results returned for both search terms on a single day, meaning that the mean number of tweets per day did not reflect any such regularity in the data.

4.8.2 Corpus Linguistic Analysis

The first analysis undertaken on the corpora was frequency analysis, which revealed the linguistic habits of platform users. Table 4.3 shows the commonality of RT as well as function words in results from basic frequency querying. This frequency analysis demonstrated the need for the incorporation of stop lists in single-word frequency analysis and also proved that the filtering code was working as expected. The most common words in each corpus were the respective filtering terms (“notts” and “Nottingham”).

4.8.2.1 Multi-Word Phrases

By considering words that commonly appeared together, the analysis began to reveal more detail. The three word phrases in particular (shown in Table 4.5) were helpful in showing how phrases appeared in tweet text, as well as how frequency analysis of single words could easily miss the presence of common word pairings or phrases.

A better understanding of the incorporation of location in online texts is essential, given the prevalence of searchable user-generated content online. The presence of location in tweet text during this data gathering stage was assured, as the sample was filtered by the place-names “Notts” and “Nottingham”.

4.9 Planning Stage 2

4.9.1 Expanding the Data Set

Location was built into the search filter of this stage, making it a necessary feature of each tweet. The next stage used a different filtering method in order to provide a better picture of the density of spatial information in Twitter text more broadly. In order to really understand how location is incorporated into an online narrative, invocations of location extending beyond those included in the search filter needed to be considered. In doing so it allows a better understandind of how locations are linked together in a network of spatial, technological and human relationships.

The sampling method used at this stage was meant to guarantee the inclusion of locational reference in tweet text, which ensured that the data could be assessed based on these inclusions. The next stage was planned to focus on how locational reference might vary between tweets as well as between geographic regions. By not using a place name as a filter, the next stage (stage two) is expected to provide a much broader view of the incorporation of location into tweet text. This should show whether the inclusion of locational reference is a common feature of tweet text, what forms these references take and at what point in the text they are incorporated.

This stage of the research work did not make full use of the corpus linguistics approach or software, with larger data sets. Using larger data sets were expected to make it possible to not only calculate frequencies and therefore to identify commonly occurring words, but also make it possible to determine which words were unusually common in one data set when compared to another. Building a reference corpus was

determined a necessary step to enable keyword analysis and the identification of words or phrases unique to one data set.

4.9.2 Hashtags

Hashtags have developed as a particular interest during this stage of analysis as they were sometimes the only elements of a tweet text that made thematic analysis possible. In more than one instance in this sample, the tweet text was predominantly made up of hashtags, which made a conventional “reading” of the text much more difficult. The next stage was planned to consider the hashtag as a linguistic feature popular in this context, and investigate the ways in which it is employed by tweet authors. Understanding which cultural tropes are being drawn on and how people situate themselves and the information they are sharing in the context of wider social and political narratives is a natural extension of this stage and is explored in the next two stages.

4.9.3 Images and Linked Content

The prevalence of RT and https in the data set highlight the prevalence of multimodal engagement across Twitter. Beyond the text of a tweet, there are often images and even external web content that have some bearing on the meaning of the text. These images and linked content were not explored at this stage. The next stage of data gathering and analysis investigated several examples of this usage.

Chapter 5 Geographic Metadata

5.1 Introduction

This stage of the research builds on the insights provided by the first stage of work, and assesses an adjusted sampling method. This stage uses both Tweet metadata for sampling and capitalises on the volume of data available through the Twitter API. In contrast to stage one and stage three, this stage made use of tweet metadata for sampling purposes by incorporating it into the search filter used to gather data through the API. Tweet text data was collected based on the geographic location recorded in the tweet metadata, rather than the inclusion of a particular word in the text body. By sampling by geographic location, this stage aimed to determine the effects of geographic location on tweet text, asking how tweet text might differ from the text sampled through the inclusion of place names as well as how frequently and in what ways location was invoked in tweet text when this was not a condition of the sample.

This chapter begins with the background on the theoretical and methodological approaches that informed the approach taken in the collection and analysis of data. It progresses to a summary of the data collected and generated at each stage of analysis, focussing in particular on the value of the hashtag as a discursive tool. This stage of work addresses the following objectives:

5.2 Overview of the Approach

Filtering according to geographic location is a common way to query Twitter data (Steiger, Porto de Albuquerque and Zipf 2015). In previous analyses of tweet content, the geographic metadata has been used to demonstrate shifting trends (Ratkiewicz, et al. 2011) or to show the spread of information across the platform (Cheong, Ray and Green 2012). The focus during this stage was not on the spread of information, but rather on the language used in particular locations. The motivating questions at this point were not about how information spreads from one location to another, but rather about how information “builds up” in

a single area over time and what forms the language people use to discuss place on Twitter takes.

Time, flux, experience and the inclusion of multiple voices were all considered during this stage of data collection and analysis. The two regional corpora here indicate the density of information that can be sourced from Twitter for future analysis, and present an alternative filtering method to that used in the first stage. Two pre-selected locations were built into the data filter, which ensured that the data captured would be channelled primarily from within two pre-selected and not overlapping radii.

The aim was to determine whether geographic location impacted the way in which place was invoked in narrative and to undertake an empirical evaluation of the difference in regionally-distinct corpora of tweet text.

5.3 Data Collection

Chapter three described the framework under which the practical work of this thesis was conducted. Here, the detail of how the method was tailored to the particular research question and objectives relevant to the various stages of the work is described, including explanations of the samples, the technologies used in the collection, and storage and analysis of the relevant data collected at this stage.

5.3.1 Python Script

The code at this stage was more complex than that used during stage one. The application needed to be able to make sustained calls for data over a number of days in order for each corpus to reach the target size or more than 50 million words. In addition to the significant increase in size of the required corpora, the nature of the search changed from a simple search for specified words to a search for tweets that were geographically tagged to the specified radius.

As Twitter limits the number of results returned in a search through the API, it was necessary to build in an allowance for the push back from Twitter once this limit was reached. When the limit was reached, the

code automatically paused and waited for the time limit to expire. This function allowed the sustained collection of data over an extended period of time, the collection ended when the code was manually terminated.

5.3.2 The Data Sample

Nottingham was retained as a field site and Newcastle was chosen as a secondary site, in order to provide a complimentary data sample. The city of Newcastle-upon-Tyne has a metropolitan population size that is on a par with that of Nottingham. The two cities have similar urban, usual resident population sizes of 305,680 and 274,444 respectively (ONS 2016). Both the selected cities also host large universities, so have sizeable student populations (Nomis 2011) (University of Nottingham 2016).

Not all of the tweets in the sample will have been tweeted from within the search filter boundary. Though the search through the API works by first searching for the coordinate, geocode data of a single tweet, it has, in some instances, pulled in tweets from users whose profile has a geographic code associated with the search area (Nottingham or Newcastle in this case). It is impossible to say what proportion of the data was posted in the area, and what proportion was drawn in because of the metadata in user's profiles. This uncertainty is not to the detriment of the data set, but is worth bearing in mind, as some tweets may be about a person associating with either geographic area, but tweeting from outside the area. In either case, the tweet text reveals something about the associations made with locations and the values places upon particular sites or vernaculars.

5.3.3 Building Corpora for Analysis

Three new corpora were compiled specifically for use at this stage; two regional corpora and one reference corpus. The decision to compile the corpora specifically for this research was made for three primary reasons. The first was that this allowed the investigation of specified locations; the second was to ensure the timescales included in the

sample were recent (given the speed with which online trends change) and the third was due to the availability of suitable reference corpora.

Twitter's terms of service prevent developers from making compiled corpora publicly available online, which meant there was no easily available corpus that matched the requirements of the study. There are some corpora available, in the sense that previous researchers can publish lists of tweet IDs, from which a corpus can be retroactively compiled, but because spatial extent was a key focus of the regional corpora, this rules out this approach for at least their compilation, the decision to build a reference corpus specifically for this research followed naturally from this. Self-building a reference corpus ensured that the time window that the reference corpus covered would be similar to that covered in the two regional corpora. The speed by which online language can change and its potential variability from offline language were significant factors in the decision to custom build corpora for use in this research, the reference corpus ensured that comparisons were being made with a comparable data set, in terms of time, language and geographic focus. The reference corpus covered the entirety of the British Isles, with no sites of exclusion. The reference corpus additionally includes the date and time that each tweet was published. That corpus is described in Table 5.1.

5.3.3.1 Regional Corpora

A central point within each urban population area was set using coordinates and a radius of 3 miles from this point determined the total search area for each of two regional corpora. For both sites, the 3 mile radius ensured that the majority of urban sprawl in each city was captured, as well as major areas of suburban development; in Nottingham this included the University Park Campus and Beeston as well as areas South of the river, which all fall outside the official city boundary of Nottingham (Nottingham City Council 2011) but are generally considered to be part of the greater Nottingham area (Nottingham City Council 2016). Similarly, in Newcastle-Upon-Tyne this included urban and green areas to the north and south of the river. A

third, much larger, corpus was compiled by expanding the geo-locational search to encompass the whole of the British Isles, this corpus was compiled at a later date and includes tweets covering two dates (28 hours) in the December of 2016. The resulting corpora totalled in excess of 107 million words; two regional corpora averaging about 4 million words and a much larger corpus encompassing the entire British Isles of more than 98 million words (Table 5.1). The geocode search parameters (which used a central coordinate and a set radius in miles) were outlined in chapter three (Figure 3.3).

5.3.3.2 Reference Corpus

To ask certain questions of the data using the methods of corpus linguistics, it was necessary to create a reference corpus that represented a more general sample of language use on Twitter. With the reference corpus in place, it was possible to query the keyness of certain words in each of the regional corpora.

There are a number of corpora of web and online language available for free online. These include WebCorp.org.uk (an online suite that enables analysis of linguistic examples live from the World Wide Web), The NoW (News on the Web) corpus consists of 6 billion words from web-based newspapers from 2010 onwards) (https://corpus.byu.edu/now), while the CANELC corpus consists of about one million words (Knight, Adolphs and Carter 2014). The NoW corpus content was too genre-specific, and did not include any examples of social media text; the CANELC corpus was too small to be used as a reference corpus. No pre-formed Twitter text corpora met the criteria of being geographically filtered in a way that matched the intended filtering of tweet text that was also filtered according to the geographic sites of interest here. A corpus was compiled specifically for use at this stage, filtered according to Geographic Location; the filtering boundary was set at 331.191 miles for the British Isles reference corpus which ensured that the entirety of the British Mainland was included in the search radius.

5.3.4 Describing the Corpora

The three corpora compiled are described here, their size in words, their size in tweets and the times and dates covered by each corpus are shown in table 5.1.

The Corpora were compiled through the uploading of simple txt files of the Twitter text data. Word content was automatically calculated by the software.

After uploading, the corpora were as described in the table below.

Table 5.1: Size, in words and tweets, of the three corpora compiled

Corpus	Size (words)	Size (tweets)
Nottingham Geo	3,675,775	c. 258,857*
Newcastle Geo	5,133,426	c. 361,509*
British Isles Geo	98,779,803	6,954,678

**Tweet count was not integrated into these searches. Estimated figures based on average words per tweet in the British Isles corpus (14.20), rounded to the nearest whole number.*

The regional corpora were compiled in advance of the reference corpus. The dates and times represented are shown below:

Table 5.2: Dates and time covered by the three corpora

Corpus	First tweet in sample		Last tweet in sample	
Corpus	Time (UTC)	Date	Time (UTC)	Date
Nottingham Geo	12:45	15/08/2016	08:50	23/08/2016
Newcastle Geo	09:13	14/08/2016	16:29	24/08/2016
British Isles Geo	18:31	19/12/2016	21:57	20/12/2016

** Times are Universal Standard Time in 24hr clock, dates are dd/mm/yyyy*

The corpora each comprise of a single text file (.txt). The file names correspond to the corpus labels used throughout this thesis. The data files have not been edited or structured in any way other than the way the data has been broken up by formatting present either in the tweet text itself (derived from Twitter and present in the tweet itself) or through the structures present in the data gathering code. In the case of the

British Isles corpus, this meant that individual tweets appear on new lines, and the tweet count inhabits a single line of its own before the tweet it counts.

Each corpus therefore, for the quantitative analysis undertaken, is a total sample of language, comprising of a single document rather than a collection of individual examples of tweet language. The assessment of tweets as individual examples of a particular communicative mode was undertaken using a case-by-case qualitative analysis through the use of concordancing. Any insights regarding the structures and trends present within single tweets therefore stem from qualitative analysis rather than quantitative analysis.

Data files were plain text. The files were very large but human readable, they were both opened immediately following data gathering, and before uploading to the analysis Corpus Linguistics software SketchEngine (Lexical Computing n.d.) to ensure that they contained the expected information. For the two regional corpora this was tweet text, including hashtags. For the British Isles corpus, the expected content additionally included a count of tweets.

5.3.5 Difficulties Encountered in Data Gathering

Data collection was not entirely smooth, and initially resulted in two failed attempts to gather data over a period of several days. The first time the British Isles code was run, a local power outage meant that the search was prematurely stopped and had to be restarted. A second failure resulted from a date time error that produced the following error message:

> “Traceback (most recent call last):
> File UK Search code.py, line 119 in (module)
> writeToFile (str<cnt))
> File “UK Search Code.py”, line 84, in writeToFile
> dt=datetime.datetime.now()
> AttributeError: type object ‘datetime.datetime’ has no attribute ‘datetime’”

The data from the first two failed attempts to run the British Isles search code (named UK Search code.py above) for an extended period was not saved as these attempts did not reach the minimum target of at least 50million words. The error message was saved in order to contribute to the development of the code, in the case of more failed data gathering attempts. As the code had already run successfully for the regional searches, no changes were made to the code, and it was simply run again. It ran without error on the third attempt.

5.3.6 Variance and Lexical Density

Table 8 shows a comparison of the size (word count) and the vocabulary (lexicon) of each of the compiled corpora. The lexicon size increased in line with the increase in word count.

Table 5.3: Size and lexicon of the three built corpora

Corpus	Word Count	Lexicon
Nottingham	3,675,775	255,804
Newcastle	5,133,426	318,315
British Isles	98,779,803	10,771,384

**lexicon as calculated by SketchEngine, not case sensitive (non lc)*

5.4 Data Analysis: Corpus Linguistics

Similar to the process followed in the first stage of research, the data analysis was a series of processes, each with a slightly different focus. Firstly, methods from corpus linguistics provided a description of corpus content and highlighted key subjects and dominant voices. Secondly, though not strictly separate from, and still making use of methods from Corpus Linguistics was the identification and description of spatial data through in-depth analysis and the examination of example tweets. Lastly, the analysis moved to a consideration of spatial context and meaning.

The data was queried using the online corpus management tool Sketch Engine. The software held the compiled corpora and ran queries according to set parameters. The software facilitates a number of

analytical queries, including word lists, which catalogue and count the different words appearing in the corpus, and collocation, which shows sequences of words that commonly appear together in the text. The inclusion of a reference corpus additionally facilitated key word analysis, which identified words that appear more commonly in one body of text (the search corpus) when compared to another (the reference corpus).

5.4.1 Frequency Analysis: Common Words

The first change made to the analysis following Stage one analysis was to compile a list of words that could be excluded from the results of queries run on the data. This list was compiled manually, taking inspiration from other lists that were freely available (ranks.nl 2015). The Sketch engine software matched exact cases on “black lists” (lists of stop words), so each word was provided to the list in both upper and lowercase. Misspellings were not accounted for.

Frequency lists were a quick and relatively simple way of ascertaining popular topics and patterns across the two regional corpora as well as in the regional corpus. The expectation was that this analysis would reveal common subjects of discussion – in a way, a very basic thematic analysis. Punctuation marks were excluded from this analysis.

Table 5.4 shows the top 10 most frequent words from each corpus, ranked in terms of frequency.

Table 5.4: Comparative Frequency List: 10 most common words

Rank	Word in Corpus (raw count)		
	Nottingham Geo	**Newcastle Geo**	**British Isles Geo**
1	RT (91,996)	RT (143,851)	Dec (6,974,616)
2	a (64,997)	a (93, 080)	Tue (4,941,722)
3	https (12, 830)	https (18,865)	RT (3,415,454)
4	out (11,448)	Out (14,838)	Mon (2,015,952)
5	I’m (10, 340)	Scottgshore (13,160)	a (1,471,171)
6	up (9,172)	I’m (12,014)	s (851,172)
7	now (7, 469)	One (11,725)	t (472,742)
8	one (7,244)	Day (11,720)	Christmas (372,907)

Rank	Word in Corpus (raw count)		
9	day (7,120)	Up (11,299)	https (241,930)
10	good (6,957)	Well (10,234)	Out (214,576)

These frequency lists primarily reflect the structure prevalent in tweets and do not provide a clear indication of popular topics of discussion in each catchment area. The lists suggest that the language of tweets tends to be about the individual, with “I’m” appearing in both the Nottingham and Newcastle corpora and indicates that linked content is prevalent in the body of tweets. Each list is slightly different, but they share several common words, “RT”, ”a”, “https” and “out” appear in all three lists, while “day” and “one” appear in both the Newcastle and Nottingham lists but not the British Isles list. These word lists reflect the time of data collection; Christmas was a frequent term in the British Isles corpus, but it did not appear in the frequency lists for the other corpora, reflecting the different windows of data collection. Similarly, “now” appeared only in the Nottingham corpus.

5.4.2 Collocation: Related Words

Words appearing directly before both place names Nottingham and Newcastle tended to be situational, locating those places in either place or time; “in”, “at”, “from” or “after”, for instance.

Table 5.5: Comparison of collocates appearing before placenames

nottingham -2	newcastle -2	nottingham -1	newcastle -1
mikesassi	medical	in	in
rough	move	trade	after
university	pitch	at	at
coming	arrivals	V	for
brentford	partizan	maidens	to
summers	tank	outcalls	from
back	city	central	v
midlands	injured	from	strong
forestlivenews	skysports_keit	sun	#nufc

nottingham -2	newcastle -2	nottingham -1	newcastle -1
#partycaribbean2017	h eveningchron	because	vital

**numbers and solo punctuation not included in this list*

Table 5.6: Comparison of Collocates appearing after place names

nottingham +1	newcastle +1	nottingham +2	newcastle +2
forest	upon	united	tyne
room	united	proud	#nufc
post	room	centre	centre
carnival	city	surrounding	tickets
city	fans	nottinghamshire	whole
escorts	university	https://t.co/cjwzp2hvra	perez's
trent	and	austerity	gateshead
castle	quayside	council	v
forest's	are	v	council
weather	theatre	cartel	never

**numbers and solo punctuation not included in this list*

Words appearing after place names were more descriptive (as opposed to situational); they complete sports tems names, newspaper titles, university names or landmarks "Nottingham forest", "Nottingham post" "Newcastle city" and related phrases. Collational analysis did not reveal anything surprising or unexpected, they did, however, highlight key trends in the use of place names in tweet text, including the connection between place names and events such as "Nottingham carnival" and "Newcastle…tickets".

5.4.3 Keyness: Unique Words

In order to uncover any differences in the language content between the two regional corpora, I used the reference corpus in the calculation of keywords. The inclusion of the reference corpus at this stage enabled the software to identify which words were "uncommonly common"

(Sinclair 1989) in each corpus. These key words helped to reveal the trends in language use that were the same across the regions, and highlighted any major differences.

SketchEngine automatically ranks keywords by keyness score, regardless of the frequencies in either the search or the reference corpora. Table 5.7 shows the 10 top keywords in the Nottingham regional corpus, ranked exclusively according to the keyness score.

Table 5.7: Keywords in the Nottingham Corpus

Rank	Word	Keyness Score	Frequency in Nottingham Corpus	Frequency in Reference Corpus
1	highstreetgent	1,076.76	5,292	0
2	averagealbino	999.71	4,913	0
3	clubhumbug	712.48	3,500	0
4	musclefooduk	381.49	1,909	4
5	misterwatches	366.90	1,800	0
6	debinskii	300.84	1,475	0
7	missjlb_	258.35	1,266	0
8	hanntlchrlst	254.08	1,245	0
9	catfishery	234.62	1,155	1
10	lynne_jean	232.13	1,137	0

The very low or none existent frequencies of these keywords in the British Isles reference corpus was striking, and warranted a closer investigation. As the SketchEngine software does not enable concordancing through keyword lists, in order to see how these words were appearing in the text, it was necessary to search them individually. Searches for this list of keywords revealed that many of these keywords were in fact usernames. This contrasts to the list of keywords that appear in both the search corpus and the reference corpus, where the keyword list did not consist largely of usernames.

Table 5.8: Keywords in Nottingham Corpus also appearing in the Reference Corpus

Rank	Word	Keyness Score	Frequency in Nottingham Corpus	Frequency in Reference Corpus
1	musclefooduk	381.49	1,909	4
2	catfishery	234.62	1,155	1
3	nffc	128.22	679	17
4	nottingham_post	125.78	933	104
5	gdf	112.85	564	5
6	olympics	105.45	1,188	262
7	ungrateful	92.77	868	184
8	nottingham	90.51	3,665	1,460
9	freshers	84.60	628	105
10	lawyartist	75.84	370	1
11	kmt	62.86	1,217	595
12	circled	61.96	312	8
13	nuk	57.95	324	31
14	brampton	76.61	301	16
15	notts	56.99	1,666	1,000

A similar pattern was also apparent in the Newcastle regional corpus.

Table 5.9: Keywords in Newcastle Corpus

Rank	Word	Keyness Score	Frequency in Newcastle Corpus	Frequency in Reference Corpus
1	Scottgshore	1,908.65	13,160	0
2	Emsarahholden	920.03	6,340	0
3	Maffewgregg	810.74	5,586	0
4	Lukebetts	623.16	4,292	0
5	Whatculturewwe	482.41	3,321	0
6	Rossonrasslin	371.66	2,557	0
7	Nufc_index	342.81	2,358	0
8	Nufc	316.02	4,390	205
9	Thegeordietips	313.24	2,154	0

Rank	Word	Keyness Score	Frequency in Newcastle Corpus	Frequency in Reference Corpus
10	Chroniclenufc	307.44	2,114	0

A search for examples in the text revealed that, as in the Nottingham corpus, several of the keywords that were highly ranked according only to their keyness score did not appear at all in the reference corpus. So the keyword analysis was run again, but any keywords that did not appear in both the regional and reference corpus were excluded from the ranked list. Table 5.10 shows a vastly different keyword list, when presence in both corpora is insisted upon. Names continue to feature, but tend to be proper nouns rather than usernames (demonstrated in concordancing).

Table 5.10: Keywords in Newcastle Corpus also appearing in the Reference Corpus

Rank	Word	Keyness Score	Frequency in Newcastle Corpus	Frequency in Reference Corpus
1	Nufc	316.02	4,390	205
2	Eveningchron	303.76	2,099	1
3	Boohooman	240.15	1,658	1
4	Argan	160.11	1,251	28
5	Avon	145.02	2,502	304
6	Yedlin	143.82	1,000	3
7	Yu	140.99	2,471	312
8	Usain	139.79	1,364	85
9	Tyne	136.33	2,445	324
10	Complains	128.98	1,823	213
11	Mitrovic	127.78	1,067	44
12	Davidbflower	120.28	827	1
13	Summerslam	117.03	1,301	125
14	Gayle	114.25	1,399	158
15	Benitez	106.96	1,258	144

5.5 Locational Reference and Place

Locational reference was a common feature of the language used in each corpus, with some locations demonstrating their dominance on the relevant keyword lists; “NUFC”, “Avon” and “Tyne” in the Newcastle corpus and “NFFC”, “Nottingham_post”, “Nottingham” and “notts” in the Nottingham corpus. This correspondence between filtered location in metadata and keywords suggests that Twitter users tend to tweet about the spaces they are in or the spaces they originate from.

5.5.1 The Expression of Place

The frequency, collocation and keyword analysis revealed several things about the density and distribution of spatial data in tweet text, but a full understanding of how and why location was incorporated into tweet text required closer examination. Where the CL analysis revealed a pattern or unexpected result, concordancing and individual interrogation of examples containing that pattern was used. Concordancing revealed the need for the use of external sources in the interpretation of tweets and suggested that automated analysis and categorisation of tweets would need to take steps to incorporate these sources of knowledge, particularly given the prevalence of sporting reportage (fandom) and the use of initialisms (kmt).

5.5.2 Experience

Collocational analysis of the two regional corpora revealed the dominance of sports-related reportage in regional tweet text. While this analysis returned parts of the names of sports teams (Notts County, for example) it also revealed something about the relationship between locations, activities and emotions. “Proud” and “fans” were each collocated with Nottingham, at location +1 and +2, suggesting that “proud fans” are a key feature of discussions citing Nottingham in Twitter text. There is an active segment of the population in both Newcastle and Nottingham that associate the city with football and the emotional aspects of being a “fan”. The conditions of fandom stem from

certain narrative or participatory acts, including pilgrimage to site of significance (sports grounds) and the use of a common language (Bickerdike 2016). This common language was evidenced in the keyword list, which named several personalities from the football world; “Yedlin” (DeAndre), “Mitrovic” (Aleksander), “Gayle” (Dwight) and “Benitez”. Though only their surnames were used, their identities were clarified though the association to Newcastle; this would be clear to readers sharing the same system of reference as the writers, but is more obscure to those who are not part of this knowledge community.

The keyword analysis, the results of which are presented in Table 5.10 revealed that “kmt” was uncommonly common in the Nottingham Geo corpus compared to the British Isles corpus. Concordancing was the first method employed in deciphering this, as it had helped in stage one (with ”wvk”). Concordancing showed this combination of letters appeared in the following tweet:

> *“RT @user: What kind of catfishery is this kmt https://...” (Nottingham Geo)*

Urban Dictionary was an essential tool in deciphering both “catfishery” and “kmt”. Catfishery refers to the practice of pretending to be someone else online; this can be by assuming the identity of another person, or through the creation of an entirely fictional persona. “kmt” is an initialisation of “kiss(ing) my teeth”, which describes a physical habit used to express disapproval or dislike.

Even though it was filtered from the geographic area of Nottingham, the tweet tells about an online encounter that could have happened anywhere, and does not add to our understanding of experiences in or concerning the city of Nottingham. Even though the process of deciphering an uncommonly common phrase was not helpful in determining the experiential aspects of place, it revealed two things about the use of social media data in research. First it revealed that a proportion of tweets contain no reference to location at all, and secondly,

that the essential toolkit for mixed methods analysis should include Urban Dictionary.

5.5.3 Time

Time is an integral feature in the articulation of place, and social media narratives allow us to explore place through the lens of time in tiny increments. As demonstrated in the project “All the Minutes” discussed in chapter two, time is sometimes the subject of tweets or the anchoring feature of a story (see the example in chapter two) but it can also be investigated in historical terms. Here, the difference is between a place “in” time and a place over time. The “in-the-moment” nature of Twitter as a platform means that tweets can be used to explore both these aspects. In chapter four, there are examples of time being taken for granted – five of the ten examples make some reference to time “today”, “just” “on this day” - and this trend was also evident in both the regional corpora analysed in this chapter. References to time tended to focus on the moment or day in which they occurred – noting the events of the present (today, now) and the very recent past (just) rather than the future (tomorrow) or the more distant past (yesterday). “On this day” is an interesting example, even though its relative frequency is fairly low in both the regional corpora when compared to the other temporal references listed.

Table 5.11: Comparative Frequencies of Temporal References in Geo-located corpora

Temporal Reference	**Frequency in Newcastle Corpus Raw + (per million words)**	**Frequency in Nottingham Corpus Raw + (per million words)**
today	8,219 (1,191)	7,554 (1536)
“on this day”	235 (34)	51 (10)
#onthisday	8 (1)	4 (1)
OTD	126 (18)	11 (2)
just	16,485 (2,390)	14,088 (2,864)
yesterday	1,055 (153)	788 (160)
tomorrow	4,198 (609)	2,694 (548)

Temporal Reference	Frequency in Newcastle Corpus Raw + (per million words)	Frequency in Nottingham Corpus Raw + (per million words)
now	10,710 (1553)	8,657 (1,760)

"on this day" or its tagged variant "#onthisday" was used both as part of the scene setting of a tweet (sometimes appearing at the start of a tweet);

> *"On this day 25 years ago Tim Berners-Lee opened the World Wide Web to all who wished to access it. #InternautDay https://..." (Newcastle Geo)*

as well as at the end of tweets, as an appended hashtag "#onthisday".

> *"RT @user: Heartbreaking story from Sykes Local Records about a woman executed in #Newcastle #onthisday 1754 #northeast #sad https://" (Newcastle Geo)*

The first example here, referencing the inception of the World Wide Web and a day used to celebrate important contributors to internet technologies, this day falls on the 23rd of August, which accounts for the presence of this tweet in the corpora. Though the reference to time is clear, this particular example makes no mention of the catchment area (Newcastle), referring to place only through the "world wide web". The second example shows emotional engagement with the history of Newcastle. The reflexivity of this tweet demonstrates the richness of located narratives that are accessible through social media texts; these narratives can be concerning with recounting a recent event or about personally processing the social history of an urban centre.

An overview of a few temporal references indicates that the inclusion of temporal references was common to the tweets of both the regional corpora, with references to the present "today" and "now" being the most common form. "Just", although the most frequent word in this list, has not been included in the comparison as it is often used in a non-

temporal way – concordancing helped to demonstrate this mixed use. Take the following example:

> *"Like I just dont like super-fake tv shows that try to be real, like pushing daisies was fab, bcos it wasnt realistic, so it didnt have to be" (Newcastle Geo)*

Additionally, combining keyword densities with common themes revealed that temporality impacted tweet content. This is apparent in the tweets shared over the Christmas period and in the early days of the New Year, when individuals were wishing each other well or lamenting their return to an empty university campus.

5.5.4 Juxtaposition

The combination of multiple viewpoints in the located narratives of Twitter text presented both challenge and opportunity due to the scale of the corpora used in this stage of data gathering and analysis. The method of data capture used at this stage differs from stages one and three in that it did not insist on the inclusion of locational reference in tweet text. For this reason, the perspectives present in these corpora are different. The challenge of working with data at such a large scale (in the hundreds of millions of words) was the interpretation of meaning and the drawing of insights that relied primarily on automated analysis. Creativity provided an angle from which to approach this challenge and was an essential idea in the reconciling the difficulties of finding patterns in the data using automatic methods with the opportunity to view a much larger segment of personal located narratives. Hashtags, which emerged as a feature of interest during stage one were investigated here and were shown to provide more individualised uses than were expected given the literature that was reviewed in chapter three (section 3.3.2).

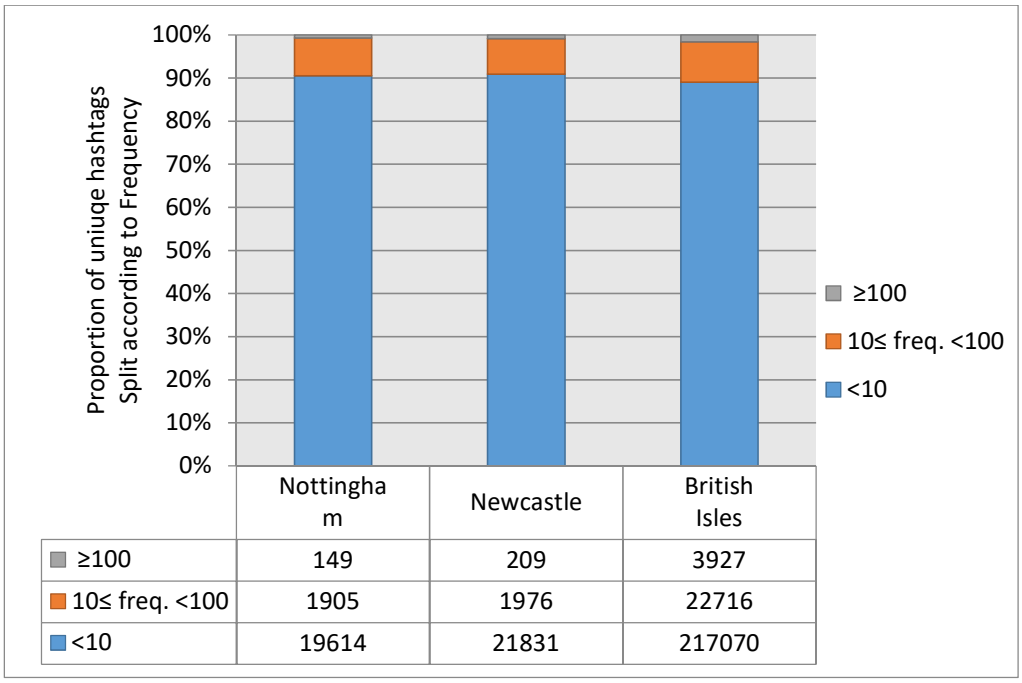
5.5.5 Creativity

Abbreviations, misspellings and the speed at which informal online language develops can all exacerbate the difficulties of uncovering meaning in texts. Misspellings and abbreviations, in particular, present

a significant obstacle to the use of computational methods in the translation of social media texts. Whilst some expected change can be accounted for, the rate at which these modes of communication evolve means that methodological and analytical approaches must be adaptive and able to respond to rapid and unexpected change. Abbreviations in the sample text meant that further resources were needed in the uncovering of meaning in the text. “kmt” and “catfishery” appeared in the adjusted keyword lists for Nottingham and concordancing did not help in deciphering meaning. The two phrases appear together in a tweet that was retweeted and appears frequently in the Nottingham sample and neither could be deciphered without the aid of Urban Dictionary.

5.6 Hashtags

Querying each of the three corpora for individual hashtags revealed a very high incidence of unique hashtags, and a relatively low incidence of repeated hashtags. Figure one clearly shows this trend toward the creation of individual tags. In the two smaller, regional corpora, more than 90% of the total hashtags occurred in the corpus less than ten times, and although the percentage is lower in the much larger British Isles, the percentage of hashtags occurring more than 100 times is still less than 5%, which is in line with both the Nottingham and Newcastle corpora. The very small proportion of repetition and a high incidence of low repetition hashtags indicate that the majority of hashtags are used not to participate in broader discourses about experience of place, but to create and express individual meaning and value.



	Nottingham	Newcastle	British Isles
≥100	149	209	3927
10≤ freq. <100	1905	1976	22716
<10	19614	21831	217070

Figure 5.1: Distribution of Hashtag Frequency According in each Corpus

5.6.1 Hashtags and Spatial Data

The use of spatial reference in conjunction with hashtags can be better understood through the interrogation of individual examples. Concordancing within the analysis software permitted this, but it was sometimes unclear where a single tweet ended and another began, the isolation of single tweets was possible by copying several words from a concordance, and manually searching for these in the corpora. This allowed the confirmation of the limits of a single tweet like the example below.

> *"Ah, the sound of the police helicopter #nottingham"*
> (Nottingham Corpus, August 2016)

The phrasing of this tweet hints at Nottingham's one-time accolade as the crime capital of England and Wales. Spatial reference does not necessarily involve only recognized place-names, it emerged in many other ways, including slang and in the names of sports teams (frequently abbreviated).

> *"Mackems having a full on melt down over @user having some brains is hilarious #nufc"*
> (Newcastle Corpus, August 2016)

The above tweet, taken from the Newcastle corpus invoked location in both these ways – through the use of the word 'Mackems' and through the mention of the local football team, Newcastle United Football Club (NUFC). 'Mackems' is a somewhat derogatory slang term used by some Newcastle natives to describe individuals from Sunderland, whose main industry was once shipbuilding. The term 'mackems' invokes the areas labouring past as well as implying that people from Sunderland are weak-willed and can be made to do whatever they are told ("Make-thems") (Robbo 2005) he use of Mackems here, combined with the #nufc tag demonstrates an allegiance to the city of Newcastle and articulates a playful rivalry with their close neighbour, Sunderland.

Other tweets were even more difficult to decipher, as the text content was made up exclusively of hashtags. Here is an example;

> *"#turkey #nwo #falseflag #russia #russianambassador #ankara #UK #Freemasons"*
> (British Isles Corpus, December 2016)

In spite of the inclusion of a number of place-names, this tweet is not really about any of those places, it refers instead to an idea that links all those locations. The text here is hard to understand in isolation as it refers not only to an international incident that was prominent in the news at the time of data collection, (which is inherently time limited) but also, through the inclusion of an image in the tweet paired with the phrases "nwo" (new world order) and "falseflag", to a conspiracy theory about an emergent world order, that controls world media through the staging of such "false" incidents. The inclusion of spatial reference in this example is incidental to the event being reported, and is more of an indexical note that serves not to explore the identity and experience of those spaces, but to link them in a web of political intrigue.

5.6.2 Hashtag Trends

Once dominant themes had been identified through the large-scale querying of the corpora, individual tweets, a sample of which are presented below, were selected based on the inclusion of one or more tags that made reference to location. Several dominant themes emerged when the corpora were queried using SketchEngine. The most frequent hashtags from each corpus were listed and then manually reviewed in order to identify recurring themes, these were: spatial reference, time, multi-word phrases, sports, promotion, events and popular culture/media.

Since location was both a central methodological concern and an important analytical interest of this research, one pattern of interest was the high incidence of locational hashtags that corresponded to the geolocation catchment area. Table 5.12 shows some examples of the locational hashtags present in each corpus.

Table 5.12: A sample of locational hashtags across the three corpora

Corpus	Target Words/Tags
Nottingham	#nottingham, #rio2016, #notts, #westbridgford, #ntu, #uon, #lovenotts, #nottsdosanddonts, #galapagos, #derby
Newcastle	#northumberland, #newcastle, #northeasthour, #rio2016, #northeast, #nxttakeoverbrooklyn, #iamnorthumbria, #gateshead, #bristolcity, #durham, #newcastleupontyne, #sunderland
British Isles	#berlinattack, #mufc, #prayforberlin, #london, #leeds, #merseysidederby, #manchester, #germany, #berlin, #uk, #aleppo, #breitscheidplatz, #turkey, #russia, #ankara, #scotland, #chelsea, #glasgow, #bristol,

* *The lists in this table are not exhaustive; they show the most frequently occurring location-based hashtags from each corpus.*

5.6.3 Spatial Scale and Range

Space, or location, was introduced into tweet text at a number of scales across all the corpora. Whether it is at city level, street level or even national level, location that followed a hashtag tended to correspond to

the location from which the data was gathered, though this was not true of the British Isles corpus. The Newcastle corpus, for example, included references to Northumbria, the North East and Durham, all areas that are spatially close to the catchment area; a similar trend was apparent in the Nottingham corpus, whereas spatial hashtags from the British Isles corpus were much more spatially diverse, ranging from Aleppo, to Russia, Leeds and Merseyside.

5.7 Reflections

Frequency lists for all three corpora clearly demonstrate how the structure of the filtering code and the affordances of Twitter impacted the results of analysis; “RT” was the most common “word” in both the Regional corpora, and it was the third most common word in the Reference corpus. “https” also featured as one of the ten most common “words” across all three corpora. Collocation analyses revealed “rough”, “university” and “in” tend to appear before the word “Nottingham” in the Nottingham Regional Corpus, whilst “Forest”, “room” and “post” tended to appear after the place name. In Newcastle there was a similar trend “medical”, “move” and “in” tended to appear before the place name, while “upon”, “united” and “Tyne” tended to appear after. The appearance of a function word in collocate lists for each regional corpus adds to the evidence provided in stage one that relationships to spatial data are frequently situational – people talk about location in a bounded sense – it is something you can be “in” or “upon”.

The reference corpus permitted the undertaking of keyword analysis, revealing patterns that were unique to, or more typical of, one of the regional corpora, when compared to a larger sample of language from the same technological context. Table 5.7 and Table 5.9 show that words sorted according only to their keyness score were often present in only the target corpus, and did not appear at all in the Reference corpus, this peculiarity was explored through searching to reveal that all the keywords solely appearing in the target corpus were usernames. These results were removed from the keyness lists through the stipulation that any keyword must appear at least once in the reference

corpus. This gave a much clearer picture of what subjects, rather than which people, were dominant in each of the regional corpora. The prominence of usernames in the keyword lists, although eventually excluded from the results, is further evidence of how the structure of a particular platform can be reflected in the text data it generates. Twitter text data makes the style of the platform explicit – linked content (https), retweets (RT) and usernames (@user) are common features.

Hashtags were the focus of the final analysis undertaken on these corpora. Word lists featuring the hashtag (search SketchEngine for #***) were compiled. The aim was to see whether tagged location corresponded to the geographic filtering area of each corpus. The hashtag word lists were sifted for place names. A sample of these results is shown in Table 5.12. This filtering revealed that the hashtagged place names or locational references did tend to correspond to the data gathering site (#notts, and #westbridgford in the Nottingham Regional corpus, for example) but also contained references from much further afield (#bristolcity in the Newcastle corpus).

The size of the samples at this stage made a human reading of the entirety of the data impossible in the timeframe. Abbreviations, slang and truncations in the text presented new challenges in the interpretation of the data. The meaning of certain phrases was not clear in isolation, and where concordancing had helped with understanding and translation during the first stage, it did not always provide clear insights at this stage and it was necessary to seek expertise from elsewhere in order to decipher some tweet text. One such example present in the keyword list generated from the Nottingham Regional corpus was “catfishery”. The repository of crowd-sourced definitions Urban Dictionary confirmed catfishery refers to the online practice of “catfishing”, pretending to be someone else, of assuming a false identity or “the phenomenon of internet predators that fabricate online identities and entire social circles to trick people into emotional/romantic relationships” (2Sweet&Naive, 2013). The platform was used in this way

on several other occasions, and was indispensable in the uncovering of meaning of samples of text, particularly for abbreviations (kmt, OTD) slang (mackems) and acronyms (#nwo). Urban Dictionary emerged as a valuable source of expertise in the explanation of words and abbreviations that were indicative of the creativity and scope of language encountered in digital contexts.

5.7.1 Technological Affordance

Much of the literature about digital communication technologies is concerned with the impact of these technologies on processes of socialisation and identity formation. Dawn Nafus and Karina Tracey discuss the “perpetual reinvention of the individual” (Nafus 2002) triggered by personal, portable technologies like the mobile phone. This project considers not only how individual actors use digital communications technologies (in this case Twitter) to invent and re-invent themselves but also how these technologies impact the way that locations are experienced and narrated online and how their identities might shift according to time or dominant narrative voices.

Twitter is a multimodal platform, people are not limited to text but can also include images, hyperlinks, emoticons, video and sound. This multimodality is made clear through the density of the hyperlink prefix “https” in the keyword analysis. The high incidence of http and RT in the frequency lists demonstrates the popularity of hyperlink inclusion on Twitter and the prevalence of retweeting. This referencing of internal and external sources or narratives is a feature common to many social media technologies, but it is particularly prevalent in Twitter, likely because of the enforced limit on tweet size (140 characters at the time of data collection). The inclusion of hyperlinks allows a user to contribute more to their own narrative than just 140 characters, they may be creating links to popular culture or encouraging their readers to participate in a poll or quiz or they may simply be providing a longer account of whatever it is they are tweeting about.

5.7.2 The Hashtag

Hashtags allow individuals to curate their writings and to participate in broader discussions and narratives about the world around them, though hashtags do not help them perform these social actions in a purely indexical way. Hashtags may be employed for a number of reasons and serve a number of communicative and social functions, only one of which is to make tweets easier to find. Research by boyd and Marwick (2014) suggests that many social media users actively seek to obscure their content from other audiences through the use of specialised languages, and hashtags, particularly multiword phrases associated with memes, are one of the ways they are able to invoke complex ideas in the limited space of 140 characters. The careful control of specialist languages can, in this way, be used to reinforce interpersonal connections or undermine the attempts of outsiders to participate in social commentary about locations.

5.7.3 Extent, Size and Quality of the Sample

This stage pushed the investigation of Twitter text data further by significantly increasing the sample size and selecting Newcastle as a secondary urban research site. The sample design for this stage insisted on a much larger data set, one that was closer to a size that could be considered “Big” in a “Big Data” sense. This design took fuller advantage of tweet-based data in two ways, firstly in terms of data volume and secondly through the incorporation of two elements of metadata - spatial metadata and language metadata. Three large corpora were compiled during this stage, the content of each corpus was filtered according to the inclusion of a metadata field that fell within the boundaries set by the researcher. The “Nottingham Regional Corpus” totalled 3,675,775 words, the “Newcastle Regional Corpus” totalled 5,133,426 words and a reference corpus, the “British Isles Reference Corpus” totalled 98,779,803 words. After two failed attempts to run the data collection script for the British Isles Reference Corpus continuously over a period of several days, a third attempt, with no alterations, saw the script run successfully.

5.7.3.1 Analysis

Analysis of the two regional corpora initially followed the same process as in stage one, beginning with frequency analysis and then moving to collocation querying. Neither frequency analysis nor collocation analysis require the use of a reference corpus, this was only necessary for the investigation of keyness. Keyword lists were created through the comparison of a target corpus (in this case either of the two regional corpora) with a reference corpus (which is usually a much larger example of the language under investigation). Following the assessment of keywords, analytical attention turned to the use of hashtags. The hashtag had been identified as an interesting feature of contemporary language during the literature review stage (Bruns and Burgess 2011) (Clark 2016) (Giglietto and Lee 2017) (Lee 2015).

5.8 Planning Stage 3

Having investigated a small data set (10,510 words) in stage one, and compared this to the very large data sets (98,779,803 words) in this second stage, the third stage was designed with the melding of these analytical and methodological insights, which suggested that while the large data sets enables the discovery of linguistic habits in relation to locational reference, collecting and analysing smaller data sets could provide clear insights into the ways in which location was imbued with meaning through tweet text. The third stage involved the gathering of a medium-sized sample of tweets that would allow a large enough base size for statistical analysis whilst remaining small enough to be read and qualitatively analysed in its entirety. 1000 tweets containing “#nottingham” and 1000 tweets containing “#newcastle” tweets were collected and analysed, resulting in a total sample of 2000 tweets.

Chapter 6 Hashtagged Place names

6.1 Introduction

The third stage of this thesis has considered how users actively tag locations in their online reactions, asking how common the practice of locative-tagging is and what it might reveal about human connections to those locations, as well as how those locations might relate to one another. This stage of the research focussed on the use of the hashtag in combination with locational reference (place names). The data was gathered using the terms “#nottingham” and “#newcastle” as filtering keywords. 2000 tweets were collected in total, split equally between the two searches. The searches were conducted on the same day (18 September 2017 and analysed in the 2 months following collection. All the tweets collected were analysed individually for locational content (place names), dominant themes and meaning and/or sense.

Before discussing the outcomes of this stage of the research in relation to the primary research question, a summary of the data is provided. This summary aims to provide a basic overview of the data collected and the key insights generated throughout the analysis. Following the data summary is an in-depth explanation of the analysis undertaken and explanations, including the provision of examples from the sample of trends in the data. This stage builds on the previous two stages and earlier research work on the development and evolution of language in online contexts.

6.2 Overview of the Approach

The approach taken to data collection at this juncture in the research combined the targeted keyword search taken at stage one (place name search) and combined it with the sampling frame of stage two (focus on two cities Nottingham and Newcastle) as well as integrating a recognised feature of the platform (the hashtag).

Using a keyword search through the Twitter API, this stage collected the text from tweets that included either “#nottingham” or “#newcastle” in the text., as in stage one, the search was not case sensitive, but matched for spelling, and length, so any misspellings or abbreviations were not captured by the search. The aim was to determine what would be revealed by tweet data filtered based on hashtagged place names, and to investigate what a targeted approach to searching might add to research aiming to understand the complexities of urban places.

6.3 Data Collection

6.3.1 Python Script

This stage once again used the keyword filtering search, similar to Stage One. The code used can be found in Appendix E: Code used in Stage three data gathering. This code differed from that used in either of the previous two stages as it used a combination of place-names and a linguistic feature. Two place names were used, and the filters were selected to reflect the sampling at stage two, which similarly focussed on two spatially distinct sites, Newcastle and Nottingham.

As the code was able to run indefinitely, and to make sustained calls for data, just as the code used during stage two was able to, the search was terminated manually by the researcher when the target sample of 1000 tweets per search term was reached. The tweet count was monitored through the command protocol. For the #Nottingham search, the researcher was able to terminate the search after 1002 search results were returned. The last two results returned (numbers 1001 and 1002) were deleted before the data was saved. The same protocol was followed for the Newcastle search, which was terminated after 1043 tweets. Tweets 1001 through 1043 were not stored or saved as they exceeded the target sample. Keeping the samples to a strict upper limit helped impose a framework on the sample, and meant that the samples mirrored each other as closely as possible.

6.3.2 Non-English Language Tweets

Though the code included a filter to include only text in “lang_en”, some text in the results sample was not in English, but this was not necessarily a hindrance to understanding theme. Many foreign language tweets in this sample were easy to categorise as they kept the names of people and sports teams unchanged from English.

One example can be extracted from the Newcastle tweets, the meaning of the tweet was discovered not through understanding of the whole, but through the three words “Benitez”, “#liverpool”, and #premierleague.

> *“Vola il #Newcastle di #Benitez, male le squadre di #Liverpool. I promossi e bocciati della #PremierLeague” (Newcastle 1000 tweets, 18/09/2017, 17:26:26)*

6.4 Data Analysis: Corpus Linguistics

Data analysis at this stage followed the approach taken in both stages one and two, but with a greater focus on contextual understanding and the unpacking of text in search of spatial data. The analysis looks firstly at the results of thematic analysis, before investigating the relationship between word order and spatial scale. Following on from this is a consideration of how location is invoked in social and cultural agendas of Twitter users, with a particular focus on the use of nicknames (informal place names or truncations of place names) and the role of location in the propagation of a regional identity.

Similar to the approach taken in stage one, querying began with a summary of the corpora.

6.4.1 Corpora Size and Lexicon

Corpus	**Word Count**	**Lexicon***
#Nottingham 1000 tweets	14,585	4,854
#Newcastle 1000 tweets	14,427	4,878

**Lexicon is non-case sensitive (lc)*

6.4.2 Frequency Analysis: Common words

Punctuation marks were excluded from the analysis and results are non-case sensitive.

Table 6.1: Ranked common words in each corpus

Rank	Word in #nottingham Corpus (Count)	Word in #newcastle Corpus (Count)
1	sep (1005)	sep (1,000)
2	mon (719)	mon (915)
3	rt (559)	rt (430)
4	sun (287)	a (201)
5	a (193)	sun (87)
6	https (94)	please (70)
7	Nottingham (87)	new (69)
8	week (74)	c (66)
9	t (65)	https (65)
10	September (60)	more (53

In common with the previous frequency analyses, the most common words in each corpus reflect the build of the corpus, showing days of collection (mon, sun), the month of data collection (September, sep) and the technological framework of the platform (rt, https and t, which is from t.co links).

6.4.3 Collocation: Related words

The table below shows words commonly appearing with the filtering terms,

Table 6.2: Comparison of collocates appearing before filtering term

#nottingham -2	#newcastle -2	#nottingham -1	#newcastle -1
#pug	in	In	in
heywhatsonnotts	session	#missing	our
#mondaymotivation	plan	Of	paintings
day	centre	#doglost	at
doubletree	#belfast	To	of
today	labour	https://t.co/2lhmxkl7bb	and
tonight	eat	Hotel	to

#nottingham -2	#newcastle -2	#nottingham -1	#newcastle -1
from	https://t.co/9m 9oJscirm	#westbridgford	for
st	for	gcrnottingham	around
bid	grooming	At	from

* *Numbers and solo punctuation not included in this list*

Table 6.3: List of collocates appearing after filtering term

#nottingham +1	#newcastle +1	#nottingham +2	#newcastle +2
#findlexi	it	Mon	is
#notts	italy	Are	london
#property	on	#nottinghamshire	industrial
female	#jesmond	No	support
and	#gateshead	#escaperoom	on
#townsville	boys	#melbourne	let’s
#music	#escort	The	been
ng	please	Sun	for
on	gym	saturday	check
university	office	S	the

The collocate lists show that the hashtagged locations commonly appeared in conjunction with the use of other hashtags, this was true of both cities and hashtagged words included a mix of both place names (#notts, #nottinghamshire) and social or business detail (#findlexi, #escaperoom). The tagged term #escaperoom could be either a locational reference or a descriptive term, referencing a specific event venue (offering puzzle-based experiences to groups) or referencing a general type of experience.

6.4.4 Keyword Analysis: Unique Words

Keyword analysis at this stage once again made use of the reference corpus compiled during the second stage of this research. Each of the regional corpora was queried against the British Isles corpus data, in order to determine whether the process of (hash)tagging a location

indicated some difference in content or meaning, when compared to the tweets from the second stage, where (hash)tagging in conjunction with a place-name was not a condition of filtering.

Table 6.4: Top 10 keywords (ranked by keyness score) across stage two and three

Nottingham Regional Corpus	**#nottingham corpus**	**Newcastle Regional Corpus**	**#newcastle corpus**
musclefooduk	sep	Nufc	sep
catfishery	harveysarmy	Eveningchron	twarchives
nffc	officialr	Boohooman	ponteland
nottingham_post	wizradio	Argan	chamonix
gdf	heywhatsonnotts	Avon	artistalanreed
olympics	missingpetsgb	Yedlin	barsinnewcastle
ungrateful	justdogsworld	Yu	oman
nottingham	judysale	Usain	chillroomncle
freshers	fernvilladogs	Tyne	mindfullnessncl
lawyartist	tys	Complains	quayside

6.4.5 Case by Case Analysis: Reading Spatial Data

The text from every tweet was analysed individually, all hashtags were isolated and grouped by tweet and location mentions were also isolated. The tweet text was assigned a primary and secondary theme. The primary theme was limited to a list of themes informed by the analysis conducted during Stage One; there were 28 pre-outlined themes (see Table 4.9).

Similar to the results of the thematic analysis undertaken during stage one of this research, there were a number of “unknowns” assigned during the primary reading of the data, this is likely largely due to the fact that all the categories remained unchanged from the first stage, with the exception of the addition of “Communication between friends” and “faith and spirituality” as primary themes.

6.4.5.1 Thematic Analysis

As in the first stage, the thematic analysis in stage three was two-tiered. The text of each tweet was assigned two themes – primary theme was validated "pick from list" to limit the framework in the first sweep of analysis. Secondary themes were assigned as the tweets were read through a second time. Where tweets were difficult to decipher, they were marked as unknown on a first pass and revisited for further investigation at a later stage. In some cases, it was necessary to assign more obscure or complex tweets themes based on the use of hashtags, usernames or abbreviations in the text, which prompted the use of the online, crowd-sourced linguistic repository Urban Dictionary (urbandictionary.com)

Improving the thematic categorisation of the tweets through alteration of the thematic framework was not a priority as the thematic analysis provided limited insights into the nuanced relationship between locational reference and value laden narratives that was hinted at in the example tweets reviewed during stages one and two. Though the thematic analysis provides a nice overview of data content, improving the categories would not improve the readability of the text, nor would it provide any further insights with regards to the research question of concern here.

6.4.6 Locational Reference

Locational reference was apparent in many different ways in the tweet text. The main patterns are explored in this section, from the placement of locational references in the body of the tweet to an investigation into the relationship between the sequencing of locational reference and the geographic scale of those locations.

Table 6.5: Segmentation

Percentage (calculated to 2 decimal places)	**Segment Assignment**
$x \leq 0.33$	Leading third (the beginning)
$0.34 \leq x \geq 0.66$	Central third (the middle)

$x \geq 0.67$	Final third (the end)

6.4.6.1 Position of Filtering Terms in Tweet Text

As in chapter four (section 4.4.5) the size and format of the corpus data collected at this stage made it possible to determine how the tagged place names (#newcastle" and "#Nottingham") were being incorporated into individual tweets. Using basic functions in excel (=SEARCH) and (=LEN), combined with a basic three-segment division of tweets (beginning, middle and end) devised by the researcher, it was possible to see at which point in the miniature narratives of Twitter text the places of interest were being mentioned. The placement of the tagged search terms was different in each corpus. In the Nottingham 1000 tweets corpus, the search term was most commonly included in the central third of the text, and appeared at the beginning of a tweet less than 20% of the time. Tweets from the Newcastle 1000 corpus showed a lower variance in the distribution of the search term between segments; it appeared in the leading third in 34% of tweets, in the central third in 38% of tweets and in the final third 28% of the time. This spread of the search term throughout the body of the text suggests that these two hashtags are not habitually appended to the end of tweets; it suggests rather, that these tagged locational references are integrated into the body of a tweet. This patterning across the three segments reveals that tagged locations are woven into social media texts, rather than book ending them – tweets do not introduce a location at the offset, as a geographic or ethnographic report might and nor do they use tagged locations predominantly as a way of indexing the content of their tweet. The positioning of the tagged location names investigated here has unexpectedly revealed very little at all about where or why individual users choose to incorporate place into their tweet text.

Figure 6.1: Segmentation shows the comparable distribution of the filtering term's position in tweets for each corpus.

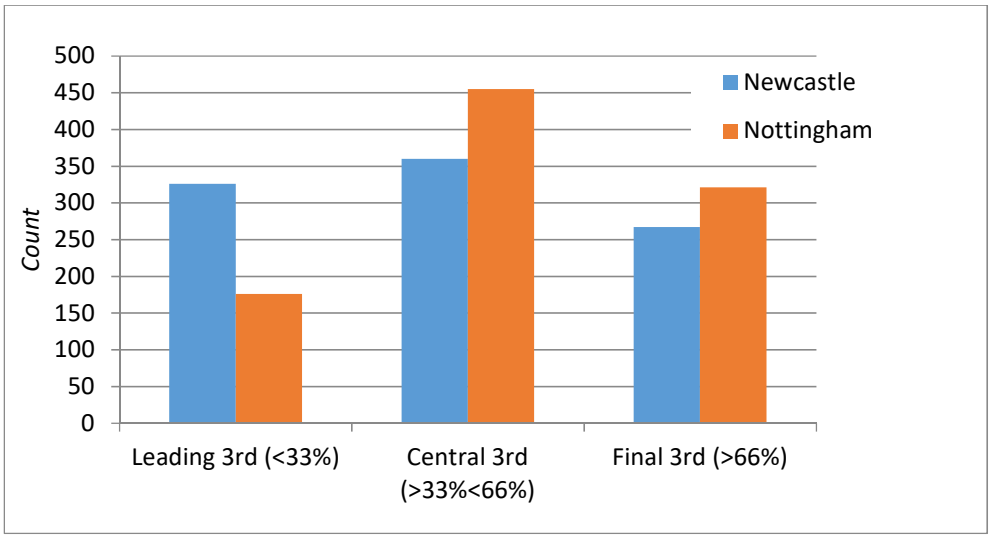


Figure 6.1: Segmentation

6.4.6.2 Place Names as Locational References

Each tweet was read through individually and any place names and locational references such as regional nicknames (yanks, geordies) were isolated. The complete record of these references is provided in Appendix G: Manually Extracted Locational Refereances from Stage Three. The process of identifying and isolating locational references revealed the overall density of spatial data in the corpora as well as highlighting some key trends in the way that location was invoked in tweet text.

Single-word place names were quite common in both the corpora, but there was a clear trend toward the addition of qualifying details that were either concatenated (forced through the choice of using a hashtag) or collocated with the place names. This variation in the placement of spatial detail meant that relying entirely on automatic querying was problematic in the extraction of locational reference; some experiential or spatial detail could easily be obscured and obliterated by simplistic querying of tweet text data. The following sections demonstrate a few of the ways in which place names were entangled with explanatory detail.

6.4.6.3 Placename + Location Detail

Place names that were incorporated into the body of the tweet text (rather than appended as a hashtag) were sometimes paired with

qualifying details such as “Newcastle area” and “Newcastle upon Tyne Quayside” (see quoted text).

> *“We are recruiting in the #Newcastle Area. Call us now 0191 406 6106 #VACANCY #NEFollowers https://...”*
> (Newcastle 1000 tweets, 18/09/2017 17:25:13)

The presence of the word “area” after the place name makes the locational reference here more ambiguous than it would have been had the place name appeared in the text on its own.

> *“RT @user: "@user: #Newcastle Upon Tyne Quayside c. 1960 http://..."*
> (Newcastle 1000 tweets, 18/09/2017 16:56:17)

The addition of “upon Tyne Quayside” in this example has the opposite effect. It clarifies the location, limiting the possibly intended spatial extent to the Quayside.

The determining features of this type of spatial data were that the detail was situational, so that the placename was made more ambiguous (as in the first example) or clarified (as in the second example) by the words that were paired with them.

These examples have been differentiated from descriptive pairing, in which a place name appears in tandem with a word or series of words that indicate some experiential or emotional detail.

6.4.6.4 Placename + Description

Place names were not exclusively city names, nor did they necessarily only correspond to the corpus filtering term. The following example names several places, which are tagged individually as well as being tagged as a combined term – describing both a site and a function.

#hockleypub

> *“RT @user: You'd be crazy to miss this. #missbowie #sixbarrel #sixbarreldrafthouse #nottingham #hockley*

#hockleypub #craftbeer #re…"
(Nottingham 1000 tweets, 18/09/2017 06:22:03)

#hockleypub is just one of multiple instances of tagged location reference in this single tweet. Also mentioned are "#sixbarreldrafthouse", and "#hockley". Where "#hockleypub" would have appeared as two words had it not followed a hashtag, here it has been shortened to a single word.

#lovenotts

"More than 1m people live in #Nottingham's recruitment area. With a £12.5bn economy too, it's a great place for business to boom #lovenotts"
(Nottingham 1000 tweets, 18/09/2017 13:20:14)

The "#lovenotts" hashtag here was present in the data several times.

6.4.6.5 Cross-Naming (Multiple places sharing the same name)

One of the most striking outcomes of the analysis of place names was the habit of cross-naming (that is, the use of the same word in the naming of different locations). In the Newcastle sample, there were several different "Newcastle"s, differentiated by the inclusion of some locational detail in addition to the place name, including one traced to Canada, one in Kentucky and one in South Africa. Clarification of nation, area or state was reliant on the surrounding text, sometimes tagged and sometimes not. These were[31]:

#Bowmanville, #Oshawa, #Newcastle [ON, Canada]

"Fall Activities in @RegionofDurham #Bowmanville #Oshawa #Newcastle https://... https://..."
(Newcastle 1000 tweets 18/09/2017 19:15:01)

#NewCastle, KY [Kentucky, USA]

[31] Square brackets indicate confirmed location referred to in the tweet text, as returned after an internet search.

> *"Class B Diesel Mechanic - #NewCastle, KY (https://...) Get Machine Assembler Jobs"*
> (Newcastle 1000 tweets 18/09/2017 19:00:51)

#Newcastle, WA, #AvalonBay, [Washington, USA]

> *"Want to work in #Newcastle, WA? View our latest opening: https://... #AvalonBay #Properties #FacilitiesMgmt #Job #Jobs #Hiring"*
> (Newcastle 1000 tweets 18/09/2017 11:53:42)

#NewCastle, DE [Delaware, USA]

> *"If you're looking for work in #NewCastle, DE, check out this #job: https://... #CitiCareers #Finance #Hiring #CareerArc"*
> (Newcastle 1000 tweets 18/09/2017 11:52:12)

The Kentucky and Delaware "NewCastle"s show how the parameters of a specific technology can transform the way that location appears in digital texts. New Castle in Delaware and Kentucky is ordinarily written as two words, but the use of a hashtag preceding the place names made it necessary for the user to concatenate it and remove the space, thereby transforming it and ensuring its inclusion in a sample of texts intended to be filtered from Newcastle (one word, with no capitalisation in the middle) in the UK.

NSW, #Newcastle, #NSW [New South Wales, Australia]

> *"Hunter Defence submits key recommendations to NSW parliamentary inquiry into defence industry https://... #Newcastle #NSW"*
> (Newcastle 1000 tweets 18/09/2017 10:52:07)

#Newcastle, #Perth [Newcastle and Perth Australia]

> *"#Newcastle & #Perth: Join a masterclass with a leading cellist of our time, @user. Book tickets at… https://..."*
> (Newcastle 1000 tweets 18/09/2017 07:30:01)

#Newcastle, #South Africa [Newcastle South Africa]

> *"RT @user: A History of #Newcastle - #SouthAfrica... https://..."*
> (Newcastle 1000 tweets 18/09/2017 07:15:42)

#NewCastle CO [Colorado USA]

> *"We're #hiring! Click to apply: Line Cooks & Prep cooks - https://...#NewCastle, CO #Job #Jobs #CareerArc"*
> *(Newcastle 1000 tweets 17/09/2017 21:36:23)*

#NewCastle PA [Pennsylvania USA]

> *"If you're looking for work in #NewCastle, PA, check out thisjob: https://t.co/ptFIEdsy8A #CustomerService #CareerArc"*
> (Newcastle 1000 tweets 17/09/2017 21:35:01)

In most cases it was possible to distinguish the locations from one another as additional information about location was also present in the text. The use of more than one locational reference made it possible to determine whether the text corresponded to the area of interest here (Newcastle UK). Without the inclusion of state abbreviations and variance in capitalisation, it would have been much harder to tell that these tweets were not relevant to an investigation into human relationships with the UK city of Newcastle. The information supplied through Usernames, though this has been anonymised in the quoting of tweet text, was also sometimes essential in the determination of "foreign" tweets. For example, one tweet in the sample includes the username @DWPDigitaljobs, the UK Department for Work and Pensions. It is through the consideration of all the text data in the tweet, including any quoted usernames, that locational data can be proven to be linked to the target city, or else marked as describing an alternative one outside the area of interest.

Although these tweets do not reveal anything about the city of Newcastle in the UK, they do present an interesting snapshot of the varying ways that tagged and untagged locations are incorporated into

tweet text. Where each reference to the different “Newcastle”s here has been tagged, the accompanying locational information is sometimes tagged and sometimes not. Take the example from Australia, where Newcastle and Perth have both been tagged, and compare this to the Colorado tweet, where only Newcastle has been tagged, and the stage abbreviation for Colorado (CO) has been left untagged. The variance in (hash)tagging in these examples appears to be related to the scale of the locational reference. Cities are tagged, but the vaguer reference to a state is left untagged. The untagged references are for clarification, but are not essential to the detail or comprehension of the text.

The geographic distribution of similar (New Castle vs. Newcastle, for example) or identical place names (Newcastle and Newcastle) shows how basic computational analysis of text can be problematic, with simple queries returning geographically disparate data.

6.4.6.6 Events and Promotions Linking Locations

Tweets that were seemingly promoting events or services were useful in revealing relationships between locations.

It was apparent from the data that co-occurring locations from within the text of a single tweet could refer to either spatially close (similar) locations or to locations from further afield in order to demonstrate the geographic spread of a particular service offering. These kinds of location mentions might not tell us anything particularly rich in their narrative content, beyond the availability of a service or event, but (hash)tagging or listing can reveal the existence of semantic or logistical links between those locations. Examples from the data set demonstrate this.

#Hull and #newcastle

> *“RT @user: #ONTOUR @user including playing #Hull @user and @user #newcastle https://t.co/Z22sPhWfl6 https…”*
> (Newcastle 1000 tweets 18/09/2017 19:32:18)

The combination and (hash)tagging of both “Hull” and “newcastle”, intersected by usernames and more crucially the word “and” in this tweet helps to illustrate the spatial disconnect between the two cities. “And” indicates that these places are not part of one another and do not intersect in the great Venn-diagram of place attribution.

The inclusion of “and” is not the only signifier of distance or disconnect, and sometimes the spatial disconnect is not signified in language and has to be illustrated through the locating of place names on a spatial map. The example below, for example simply lists a number of tagged locations, sandwiched by a reference to time “#tomorrow’s” and some text “Best Odds” which indicates the connecting activity is gambling or betting.

In this tweet the locations “#Redcar”, “#Chepstow”, “#Yarmouth”, “#Galway” and “#Newcastle”. It is clear from searching for these locations on the nomis web database[32] that they are spatially distinct, and refer to geographically dispersed regions of the United Kingdom.

> *“Tomorrow’s Racecards: #Redcar #Chepstow #Yarmouth #Galway #Newcastle Best Odds ➡ https://... https://...”*
> (Newcastle 1000 tweets 18/09/2017 19:28:08)

As a “built-up area” Redcar is described as being “in North East Region”, Chepstow is described as being “in Wales Region”, Yarmouth as being “in South East Region”, Galway returns no results in Nomis, as the service offers statistics on England and Wales only, indicating that it lies outside of these areas, and though there is no exact match to “Newcastle” as a Built-up area, Newcastle-upon-Tyne is described as being “in North East Region” (Office for National Statistics n.d.).

6.4.6.7 Indistinct/Specific Reference to Location

Reference to place was not exclusively through the use of official place names, nor through any of the spatial referencing practices mentioned in section 2.2 and did not necessarily correspond to a clearly

[32] nomis is free service provided by the Office for National Statistics, providing access to UK labour data https://www.nomisweb.co.uk/

demarcated space. One common example of this habit was the use of "area" in the explanation following a place name or building.

> *"We manage a number of properties in the Lace Market area. #Nottingham #Property please visit our site… https://..."*
> (Nottingham 1000 tweets 18/09/2017 09:53:02)

Here, "in the Lace Market area" is a vague reference to a historic section of Nottingham city centre. The Lace Market is not clearly demarcated and the addition of "area" illustrates this fuzziness. Not all reference to location were vague. In some instances, a venue name or street address was included in the body of a tweet.

> *"RT @user: #1920s #photo of The Salutation #Inn on Hounds Gate #Nottingham taken #onthisday 11 Sep 1926 https://..."*
> *(Nottingham 1000 tweets 18/09/2017 08:26:28)*

This tweet clearly indicates the location in question, providing a venue name "The Salutation Inn", a street name "Hounds Gate" and a clarification of the city "Nottingham".

6.4.6.8 Word Order and Spatial Scale

The examples explored above also illustrate the ways in which multiple references to location appear in a single tweet.

> *"RT @user: Chuffed to have won #Prudhoe in #Bloom #Business Award for #art #gallery #Hexham #Corbridge #Newcastle #fineart #watercolou…"*
> (Newcastle 1000 tweets 18/09/2017 15:40:12)

According to Local Area reports, these locations mentioned in the above tweet are all "in North East Region", but their ordering does not reveal a consistent pattern according to their scale (whether this is measured by usual resident population, the size in hectares or the density of the population). Table 6.6 shows this scaling:

Table 6.6: Comparing locations mentioned in tweet text

Locational Reference	GSS code	URP	Area (hectares)	Density (URP/hectare)
Prudhoe	E35001267	10,853	314	35
Hexham	E34004191	11,388	419	27
Corbridge	E34003881	2,946	108	27
Newcastle	E35001179	268,064	5790	46

*GSS = Government Statistical Service **URP = Usual Resident Population

The ordering of locational reference and the assessment of geographic scale was carried across several more examples, but no simple or one-directional relationship was obvious.

> *"RT @user: https://... Trixie White/Grey #Lhasa Apso #Female TAKEN FROM GARDEN #Arnold #Nottingham #NG5 @user http…"*
> *(Nottingham 1000 tweets 18/09/2017 09:51:05)*

Table 6.7: Comparing locations mentioned in tweet text

Locational Reference	GSS code	URP	Area (hectares)	Density (URP/hectare)
Garden	N/A	N/A	N/A	N/A
Arnold	E35000779	37,768	815	46
Nottingham	E35001316	289,301	6,250	46
NG5	N/A	N/A	N/A	N/A

Tables 6.6, 6.7 and 6.8 support the finding that the ordering of locational reference does not necessarily relate to either increasing or decreasing scale. What is interesting about the example shown in Table 6.7 is the inclusion of several modes of spatial reference, from official place names (Arnold and Nottingham), a general space (garden) and the first half of a postal code (NG5).

> *"We're in the Park #House #Removal #Nottingham. https://..."*
> *(Nottingham 1000 tweets 18/09/2017 12:28:41)*

"the Park" cited in this example refers to an area to the west of Nottingham city centre known as "The Park". The area is not officially demarcated and does not feature in the ONS or nomis database (Office for National Statistics n.d.),

> *"RT @user: Hi #NorthEastHour #Weaving #Gateshead #Classes #Gifts #Handmade #Newcastle #Tyneside Book here: https://... http…"*
> *(Newcastle 1000 tweets 18/09/2017 19:11:32)*

Table 6.8: Comparing locations mentioned in tweet text

Locational Reference	**GSS code**	**URP**	**Area (hectares)**	**Density (URP/hectare)**
NorthEast	N/A	N/A	N/A	N/A
Gateshead	E35001222	120,046	3017	40
Newcastle	E35001179	268,064	5790	46
Tyneside	N/A	N/A	N/A	N/A

The two official "built-up areas" in this example, "Gateshead" and "Newcastle" are listed in ascending order of size across all the measures (URP, area and density) but neither NorthEast nor Tyneside seem to follow this pattern. Alhough "Tyneside" is not classified itself as a "built-up area" like Gateshead and "newcastle", it is present in the forms "North Tyneside" and "South Tyneside" which are "local authorities" within the "North East region" (Office for National Statistics n.d.). The suggestion here is that Tyneside is a broader locational reference than either Gateshead or Newcastle (since it presumably encompasses the two local authorities of both North and South Tyneside).

The relationship between word order and spatial scale in the analysis of locational references in tweet text is not a straightforward one, and is complicated by the fuzzy relationship between place names used in popular parlance and the names and designations used by local or governmental authorities to collect and collate data.

This lack of distinct patterning or a discernible relationship between the order of locational references in tweet text and their scale highlights the difficulty of working with social media texts. The framing and design of this research meant that no specific voices were being investigated and that the texts had not been produced with an explicit attention on the part of the user to the theorisation of place; clear connections between phrasing and spatial description were not clear, and possible coincidental rather than intentional. The overlaying of voices was identified as a key aspect of the theorisation of place during the literature review, but the density and disparity of these voices could be contributing to the lack of patterning. This would be a consideration for further research - perhaps differentiating between user types would better facilitate the drawing of conclusions about how place is incorporated into social media texts.

Identifying and extracting locational reference from the text was a long process, but this intensive qualitative approach facilitated the identification of patterns and themes in the ways locational reference was being incorporated into tweet text by Twitter users. Locational reference was extremely varied in the text; with some tweets including only a single tagged locational reference, others incorporating many references. Some users incorporated the tagged locational reference into the body of the tweet and others appended the tag at the end of the written content. The appearance of multiple locations sharing the same name, which was particularly evident in the Newcastle 1000 tweets corpus, highlights the challenge of using place names to isolate and capture data streams about specific locations. Particularly when combined with a hashtag, differences in spacing and capitalisation of place names were lost and led to the inclusion of text data from outside the target area.

6.4.7 Key Attributes of Place

6.4.7.1 Time

As in the first and second stages, the expression of time in tweet text was sometimes through the use of temporal hashtags like #onthisday,

#OTD, and #throwbackthursday, though their frequencies in both corpora were low.

#throwbackthursday was used to look backward into the history of the city of Nottingham and Nottingham Trent University. The key ideas of this tweet have all been tagged, with the tagged terms being integrated into the body of the tweet, rather than appended and grouped together. If the non-tagged words were obscured from view, the content of the tweet would still be apparent.

> *"RT @user: We've gathered photos of #Nottingham & #NTU from the past for this week's #ThrowbackThursday. Take a look 👇 https:// …"*
> *(Nottingham 1000 tweets 18/09/2017 03:09:02)*

The plural "photos" suggests that many images have been gathered as part of the mission to look at the past – this tweet demonstrates the build-up of experiential accounts and perspectives over time (in a visual form). In contrast, the example below shows how multiple sites can be linked together by a single actor – flipping the anchor from location to person.

> *"18 September 1874: birth (London) of Cyril Henry Golding-Bird, first Bishop of #Kalgoorlie 1914-1919. Third Dean of #Newcastle 1909-1914."*
> *(Newcastle 1000 tweets 18/09/2017 10:18:33)*

Cyril Henry Goldingbird is the connecting node between London, Kalgoorlie and Newcastle.

6.4.7.2 Events and Experience

The thematic analysis helped to illuminate some differences between the two cities, events was one category were the different uses and experiences of Twitter users in those places starts to be revealed. Secondary coding revealed Newcastle played host to "Oktoberfest", "Jurassic Kingdom", "comicon" as well as an event targeted at "female networking". None of these codes co-occurred in the Nottingham corpus,

the city instead playing host to “ghosts of nottingham”, “goose fair”, a “book signing” and a “puppet festival”. Not all reportage online concerns personal experience or events. A striking feature of the Newcastle 1000 tweets corpus was the number of tweets about the “SATANIC TORYPARTY” and their associated activities. Like the tweets in the British Isles Corpus discussed in chapter 5, the locations mentioned in these tweets are linked not by an event but by an idea.

> *“SATANIC TORYPARTY using childrens homes for their paedophile network. #liverpool #london #camden #manchester #newcastle #croydon #kensington”*
> *(Newcastle 1000 tweets 18/09/2017 19:23:41)*

> *“SATANIC #TORYPARTY Public Servants turn into Fascist Dictators overnight. #altnews #politics #liverpool #manchester #newcastle #croydon”*
> *(Newcastle 1000 tweets 18/09/2017 19:20:28)*

The locations in these two examples are connected through the supposed existence, in the first example, of a paedophile network and in the second, of fascist dictators in the public service.

6.4.7.3 Heritage

Similar to the events category, which highlighted different experiential possibilities in each urban area, the primary theme of Heritage facilitated the tracing of emotive and historical elements unique to each city. Among the elements connected to Heritage in the Newcastle corpus were “body snatching”, “punishment”, “coal” and “17th c. drunks”. In the Nottingham corpus, they included “robin hood” and “historic fact”. Both the corpora shared a reference to a specific historic site “Elswick works” in Newcastle and “boots department store” in Nottingham as well as the more general theme of “historic image”. It transpired that the sharing of old images and photographs was a popular way to illustrate facts or stories about historic events or buildings in and around each city.

"RT @user: Heartbreaking image of John Roman - sentenced to 14 days hard labour in #Newcastle Gaol for stealing clothes in 1873 #Victo…"
(Newcastle 1000 tweets 18/09/2017 11:50:14)

The well-known character Robin Hood, who is associated with Nottingham and the nearby Sherwood Forest (Ohlgren 2007), was mentioned in several tweets. It is interesting to note the different locational references being drawn into the Robin Hood narrative in the two examples that follow.

"Byrunsdale, Old Basford, Home of Robin hood https://... #old_basford #nottingham https://..."
(Nottingham 1000 tweets 17/09/2017 19:50:19)

Robin Hood was also linked to location through events, including the "Robin hood pageant".

"#robinhood Pageant returns to #Nottingham Castle https://... #WestBridgford https://..."
(Nottingham 1000 tweets 18/09/2017 08:08:10)

The legend of Robin Hood and his merry men has been picked up and used extensively in the promotion of heritage sites and tourist activities around the city of Nottingham and its environs (Visit Nottinghamshire 2018), demonstrating the perceived cultural and economic significance of historical characters and associated sites in the propagation of a clear and unique place identity.

6.4.7.4 Highlighting Social Problems

Some of the tweets in the sample indicated use of the platform for raising awareness of the social and economic conditions of the city. These issues included crime, poverty and homelessness.

The following tweet links the city of Nottingham to the broader territory of the UK – "UK-wide" indicates that the city is included in the scope of an operation to tackle crime through the deployment of officers carrying weapons.

"RT @user: Armed officers deployed across #Nottingham as part of UK-wide operation https://...https://..."
(Nottingham 1000 tweets 17/09/2017 19:52:26)

The tweet text alone does not describe any triggering event, nor does it provide a clue as to what the nature of the "operation" might be, but the phrasing suggests the decision was taken as a response to a nationwide problem, rather than a localised event.

"@user Please help spread the word! City centre donation point #Newcastle #periodpoverty"
(Newcastle 1000 tweets 18/09/2017 11:51:56)

Similarly, the above tweet links Newcastle to a perceived broader social issue, indicated by the use of the dual tags "#Newcastle" and "#periodpoverty". In both these examples, the nebulous character of urban spaces is evident. Neither Nottingham nor Newcastle, in these examples, has a clearly demarcated boundary; the sites spiral out from a "city centre" (in the case of Newcastle) or inhabit an indeterminate expanse that can be travelled "across" (in the case of Nottingham).

Homelessness was another one of the social and economic issues highlighted through the thematic analysis and the reading of individual tweets in their entirety.

"salon38 will be working with Hope, a charity helping the homeless. We'll be cutting their hair! #hopene #freehaircuts #Newcastle #Homeless"
(Newcastle 1000 tweets 18/09/2017 17:41:26)

The tweet above demonstrates the use of the platform by local businesses getting involved in charitable initiatives that tackle some of the issues of urban centres, including poverty, foodbanks and homelessness.

"It's 2017 and #Nottingham has Homeless people and food banks. Is that acceptable?"
(Nottingham 1000 tweets 18/09/2017 06:43:40)

There were a number of instances where tweets were written entirely in the form of a question, or had a question added to the end of a statement. In the example above, the addition of a question transforms the tweet and sets it apart from the previous example as it links homelessness to the idea of acceptability, which prompts a response rather than being a simple observation. The use of Twitter for organising located social action has been widely studied (Segerberg and Bennett 2011) (Lefebvre and Armstrong 2018) but that aspect of this sample was not investigated further.

6.4.7.5 Perspective and Identity

As discussed in the literature review, a "place" can be experienced, described and imagined differently according the character and habits of an individual (Nold n.d.) (Waldron 2013). As well as considering the overlying of these perspectives from within a city, this stage explored what meanings come to be strongly or habitually associated with a place through the use of geographically specific nicknames.

Regional nicknames were a particularly prevalent feature in the Newcastle 1000 tweets corpus. "Geordie", "toon", "toon army" and "magpies" were all repeated throughout the corpus and were present in tweets covering a range of themes, from merchandising to weather and sports.

> *"Keep the Toon Army spirit all week! Every donation counts at @user #Newcastle @user ⚽♡🍴… https://..."*
> *(Newcastle 1000 tweets 18/09/2017 12:01:26)*

The "toon army spirit" here is linked to the idea of charitable giving.

> *Rafa Not Surprised By 3 Wins On The Trot https://... #nufc #newcastle #magpies #toon #rafa https://..."*
> *(Newcastle 1000 tweets 18/09/2017 15:00:10)*

The truncation of Rafael Benitez to "#rafa", the manager of Newcastle United Football club, who are affectionately known as "the Magpies" due to the black and white stripes of their at home kit, is another

example of how word forms can provide clues as to the emotional significance of a feature, event of person as is the case here.

6.4.7.6 Commentary on what it means to be a "Geordie"

One unexpected outcome of the thematic analysis was the discovery of a series of tweets that linked weather reports with the use of regional nicknames and commentary on the expected clothing choices of people native to Newcastle (Geordies). These brief reports all begin by setting the spatial and temporal scene – "In #Nottingham" before reporting the current weather "it is currently". The third part of the tweet is always "geordies will be wearing"…followed by an item of clothing or fashion accessory. The tweets all end with "#nufc #toon". Together, these tweets suggest not only that Newcastle residents have a complex relationship with fashion, but also that "geordies" have a unique identity that is closely connected to both #nufc and #toon. The tagged "Geordies" exemplifies a playful approach to the negotiation of place based identity through social media texts, supporting the work of other scholars whose work shows a the complex ways in which this negotiation takes place (Page 2010, Giglietto and Lee 2017, R. Jones 2011).

Scattered clouds and open necked shirts

> *"In #Newcastle it is currently 11.7°c and scattered clouds, #Geordies will be wearing open necked shirts #nufc #toon" (Newcastle 1000 tweets 18/09/2017 12:00:07)*

Broken clouds and hot pants

> *"In #Newcastle it is currently 11.21°c and broken clouds, #Geordies will be wearing hot pants #nufc #toon" (Newcastle 1000 tweets 18/09/2017 15:00:10)*

Scattered clouds and sunglasses

> *"In #Newcastle it is currently 11.43°c and scattered clouds, #Geordies will be wearing sunglasses #nufc #toon" (Newcastle 1000 tweets 18/09/2017 11:00:07)*

“from Newcastle”

> *“RT @user: #Geordie birthday cards and Geordie #mugs. If you’re from #Newcastle, take a look...! @user https://...”*
> *(Newcastle 1000 tweets 18/09/2017 15:01:33)*

“#novacastrians”

> *“CALLING ALL #NOVACASTRIANS. Free #Accreditation Info Session in #Newcastle. Term 4. Book your spot. https://...@user”*
> *(Newcastle 1000 tweets 18/09/2017 02:58:30)*

The word “novacastrians” appeared only once in the data set, hashtagged and completely capitalised “#NOVACASTRIANS”. Perhaps this is due to the difficulty of eloquently expressing that your message is for the attention of “people from or in the newcastle area”. Similar to the previous example, it raises questions about the identity of a Novacastrian – what qualifies a person for this title? Is it geographic location at the time of reading or posting, is it linked to either temporary or sustained residency in the City of Newcastle or perhaps it is a birth right, afforded only to those individuals who were born in the city. The surrounding words do give some clue as to the intended meaning of the tweeter in this case – it appears to be advertising a free information session concerning accreditation, which would likely mean that the “#novacastrian” title would be applicable to anyone living or working in the region, and not necessarily linked to any ideas of birth-rights or homeland.

6.5 Reflections

Frequency and keyword analyses reveal that locative (hash)tagging is not uncommon, but that the process of tagging meant that users were sometimes combining two elements of a place name (“New” and “Castle”) so search results falling outside the UK were returned. The inclusion of the hashtag in the keyword search compounded the issue of erroneous data inclusion, as it forced the filtering of two-word place

names like “New Castle” into the sample, as a result of the need to remove spaces from multi-word phrases when using hashtags.

One trend uncovered in the analysis of spatial data attributes was the unpicking of place-based identity through the use of nicknames and through dressing habits. One user frequently paired weather reports with the phrase “geordies will be wearing…”, followed by some combination of clothing or accessories deemed to be representative of a Novacastrian appreciation (or ignorance) of climate conditions. The Geordies will be wearing…” commentary was particularly interesting because although it was on the face of it, about weather and dressing habits, it revealed the light hearted ways that place-identity is negotiated in online contexts.

6.5.1 Filtering by Locational Reference

Tweets may not have coordinate data – Twitter uses place names as a substitute when coordinate data is not available – this substitution is evident in the data from Stage two which included a number of tweets traceable to sites in other countries. Despite these results (with shared place names) not having been collected intentionally, they help to illustrate the difficulties inherent to the analysis of text-based locational reference and the investigation of place in the digital context.

6.5.2 Locative (hash)Tagging

The third stage further developed the scope of the research through the incorporation of hashtags as a filtering condition in the gathering of tweet text for a medium-sized sample of text. Two new corpora were compiled, each consisting of 1000 tweets. The filtering condition of the “Nottingham hashtag corpus” was that that the tweet text must include “#nottingham” while the “Newcastle hashtag corpus” (which was also made up of 1000 tweets) insisted on the inclusion of “#newcastle”. The search filter was not case sensitive and the appearance of a number of examples of cross-naming in the sample demonstrates how this filtering method leads to the inclusion of out-of-area text. Though they were not collected deliberately, these tweets from outside the UK did help to

demonstrate the nuances of place naming and were key in highlighting the effect of (hash)tagging on the concatenation of place names, which was an important learning from this stage.

6.5.3 Human-readable text

The smaller scale of the data collected (relative to the text of stage two) during this stage permitted human-reading of the text. The analytical approach here was strongly qualitative – focussing on the reading of tweets in their entirety and describing in detail the full breadth and depth of spatial data present in tweet text. Making use of the list of primary themes generated during the first stage of analysis, each tweet was allocated two themes, one chosen from the existing list (see section 4.5.2.1), and a secondary theme added to provide detail.

In addition to thematic analysis of tweet text, each tweet was investigated for spatial reference, and any references were isolated in a list. Furthermore, as hashtags seemed to appear in groups when they appeared at all. Individual hashtags from each tweet were also noted and isolated in a list, so that each tweet could be read in its entirety, or assessed according to which spatial data or which hashtags appeared in the text.

The qualitative focus of this stage pushed the focus beyond the limits of the corpus linguistics software being used and relied almost entirely on the readings of the researcher. The process of extracting spatial data instance by instance was extremely laborious and took up a lot of time. It was worth the effort as the isolation of this data aided in the construction of a network of linked locations – linked through the tagged place name #nottingham. The same list was generated from the Newcastle hashtag corpus. This analysis demonstrated that spatial data is present at a variety of scales, ranging from city, to county, to individual street names as well as the inclusion of “fuzzy” spatial relationship references such as “close to” or “near” (Hudelot, Atif and Bloch 2008) and the inclusion of vague categoric labels such as “the park” (Jones, et al. 2008). Thematic analysis echoed the results of

thematic analysis undertaken during stage one; people talk about sports all the time on Twitter. The case-by-case analysis, in combination with the isolation of the spatial references revealed not only a variance in the scale at which location may be invoked in tweet text (street, city or national scales were all present in the data) but also highlighted a number of ways in which spatial reference was being used; in both an indexical manner (promoting events or location-based services) but also in more complex ways. Several examples from the Newcastle hashtagged sample showed the use of multiple spatial references by a user to show the dispersal of an event or network.

> *“SATANIC #TORYPARTY Brilliant new Baldrick plan, "A period of constructive chaos." #london #newcastle #manchester #leeds #oxford #cambridge”*
> (Newcastle 1000 tweets 18/09/2017 19:17:31)

The locations in the above tweet are connected by the idea of a “SATANIC #TORYPARTY”. The appearance of tweets highlighting social problems and some prompting social action was not surprising, as there is an established body of literature documenting the use of social media platforms, and especially Twitter for the organisation of protests, marches or the advocating of social justice. The approach taken at this stage provided rich insight into the connections being made between places, events, time, experience and personal narratives, without the need for the scale of data that was collected and analysed in the previous stage.

Quantitative analysis is facilitated by large bodies of text, and the availability of a reference corpus was helpful in the uncovering of the expressive habits of one place when compared to another. However, qualitative analysis does not require such large data sets, and is certainly not feasible at such scales (when a sample is hundreds of millions of words, for example). The combination of approaches used throughout this research highlights that there is more than one way to fruitfully approach social media texts when addressing questions about

place in contemporary urban contexts. Selecting the most appropriate method of data collection and analysis would depend on the questions being posed. The mixed-methods approach taken throughout this research permitted space for the exploration of “Big Data” rhetoric, revealing that smaller-scale qualitative analysis that is sympathetic to the quirks of a particular platform (in this case the hashtag) might be sufficient in exploring “place” in social media texts.

Chapter 7 Place in Twitter Text

7.1 Introduction

The motivating aim of this thesis was to understand the breadth and depth of spatial information present in Twitter data, and to investigate how this might inform framings of place. This thesis traces three threads that emerged from the reviewed literature that theorises place as a layered and ever-evolving quality, shaped by human experience and defined by the telling of that experience. The three stages of this research, each focusing on the analysis of place-based information gathered according to slight variations in locational reference, sought particularly to investigate how place is expressed through Twitter text.

Each stage investigated the combination of text-based locational reference with rich description that was divided roughly into three non-discreet categories of Experience, which encapsulates the reportage of individual lived experience, Juxtaposition, which is the effect of overlaying multiple voices concerning a single location and Time, which refers to the telling of rhythms or the use of temporal frames of reference, such as referring to the past or the future.

In addition to the application of a number of different methods, this thesis has drawn on the expertise of a number of disciplines, bridging, as it does, the theoretical territories of human geography, the digital humanities, the visual arts and social anthropology. Over the course of this research, which took place over three stages, a total of 107,589,004 words of Twitter text were collected, compiled, compared and analysed. Special attention was given to the spatial data contained in these texts, as well as the contextual information that this spatial data corresponded to. A blend of qualitative and quantitative investigative techniques was used, in a mixed methods approach that was facilitated by tools and techniques taken from Corpus Linguistics.

The three stages each explored slightly different ways in which locational reference can be investigated through social media texts. Stage one, focussing on the use of text in the form of place names,

collected a sample of tweets all containing the placename “Nottingham” or the truncated form “notts”, which is sometimes used in reference to the county of Nottinghamshire as well as a shortening of “Nottingham”. Stage two meant to explore the density of locational reference in the event of it not being a condition of a tweet being included in the sample. Stage two was intended as a more general study of the invocation of locational reference in tweet text, and was also used to extend the scope of study in order to determine whether the comparison of “place” in two similar but distinct urban locations would reveal anything unique about those cities, or if the same themes and topics emerged from both.

Stage three combined the place-name focus of the first stage with the dual-site approach taken in stage two, and added the condition of a hashtag, so that tweets were filtered according to the inclusion of “#nottingham” or “#newcastle” in tweet text. The continued focus on two research sites was essential to understanding the transferability of the methods and approaches taken in this research outside specific sites. The inclusion of hashtags in the sample allowed the exploration of conscious marking out of location – and the theorisation of the attitudes or situations that motivated this (hash)tagging (rather than leaving a placename un-tagged).

The key insights and outcomes from each stage are drawn together in section 7.2 and 7.3, where this thesis returns to the research objectives. The central research question of this thesis is discussed in chapter 8, where the answer to each demonstrates the broad range of methods and disciplinary influences at work.

7.2 Attributes of place most relevant to the context

The first objective set out for this thesis was to identify the attributes of place most salient to the examination of social media texts in contemporary urban contexts. Academic literature and artistic engagements with the idea of place highlighted three primary characteristics: experience, time and juxtaposition.

These three features do seem to be, so some extent, inherent to the medium of Twitter, the platform's prompt to users when authoring tweets is, afterall "What's happening?", a question that encourages in the moment reportage. Time, experience and juxtaposition were each evident in artistic, technical and academic engagments with place and are particularly relevant to discussions of place-based narrative on social media because they are also key attributes in social media data streams. Time, is built into technology – particularly contemporary digital technologies. Every tweet, as explained in section 3.3.3 and illustrated in Appendix A, has associated temporal data that connects it with a time zone (comparative to coordinated universal standard time) and, often, descriptive information about the experience of time, contained within the tweet text. Juxtaposition is a natural feature of user-generated content. With a huge number of individuals simultaneously constructing their responses to the prompt, variation is natural and incongruity can be expected.

7.3 Challenges Particular to the Use of Social Media Texts in Research

The use of social media data in research is fraught with both practical and ethical anxieties. These anxieties are amplified by the use and/or application of this data in contexts external to that in which that data was generated (thereby transforming the data into "ambient data").

The challenges of exploring human relationships with place are myriad and complex. The challenges encountered in this research, applicable to both past and future research in the context of social media and geospatial theory has been articulated through a few key issues. These issues show some of the decisions taken over the course of this project, and speak particularly to this interdisciplinary and technological context. A suggested approach to developing ethical research in similar research is outlined later, in section 7.6.

7.3.1 Becoming a “developer”

Access to tweets can be gained in 3 ways: by non-users (those without twitter accounts) using the search function on the Twitter homepage; by non-users through a commercial agreement with Twitter or a licensee, or by users through the API. Non-users can access a limited number of tweets by simply searching using the search bar on the Twitter homepage. This approach might suit research requiring only a handful of tweets. The commercial route is a better way to aquire large and comprehensive tweet datasets but access is expensive. Use of the Twitter “Firehose[33]” is the only way to ensure data completeness (for the compilation of Twitter example corpora); this is the main data stream for Twitter. The firehose provides a complete set of Twitter data, including metadata and can be requested directly from the organisation but comes at a huge cost.

The Twitter API allows app developers to access a stream of data from Twitter. These developers must have a Twitter account through which to manage their application and they must agree to the company’s Terms of Service and the Privacy Policy. There is no special requirement for developers, and individuals actively searching, gathering and compiling Twitter data and theyare free to use the data for a variety of purposes.

The commercial use of data gathered through the API is possible and permitted in the Terms of Service, but compiled databases of tweets or tweet data cannot be made publicly available. The relative ease with which Twitter data can be accessed and gathered was an influencing factor in the decision to choose the platform as a research site, but the access does not come free from limitations.

One of the main difficulties in using the API (rather than the Firehose) to gather data from Twitter is that the API does not provide a “complete” stream of data, but rather a sample of tweets, filtered according to the

[33] The Firehose is a stream of data that Twitter makes available to developers who register through the enterprise sales platform. The Firehose differs from the API in one primary way – that is – it guarantees the return of 100% of tweets that match your search criteria (https://developer.twitter.com/en/pricing.html)

parameters the developer sets themselves. It was for both financial and pragmatic reasons that the option of the API was selected. The corpora that have been compiled at each stage of this research are only small samples of the total data output from Twitter and have been treated in a way that is sympathetic to this fact, making no claims to the universality of language patterns or asserting views expressed in tweets taken from the sample as indisputable evidence.

7.3.2 Using “Open” Data

According to the Terms of Service that all users must agree to, Twitter is an open data source. The data, including metadata and user profile data are searchable through both the developer’s API and through the web and mobile interface of the service itself, which has a rudimentary search function. Legally the data is “open” in the sense that it can be accessed and searched by any interested party, but its classification as “open” may also depend on how easy it is to search and how interested people are in making use of the data. Also having some bearing on the degree of openness is the sheer volume of data that is generated by the platform. Given that more than 6,000 tweets are published through the platform in any given second (the record is 143,199 in one second) (Krikorian 2013), users may find anonymity and obscurity in the flood of information, trusting that their commentary may go unnoticed amongst the other voices. The ethical insufficiency of the security through obscurity model has been pointed out by Williams, Burnap and Sloan (Williams, Burnap and Sloan 2017), who raise concerns about the ease with which researchers can “legally” publish personal and sensitive content without the express or informed consent of the parties concerned.

7.3.3 Negotiating Ownership and Reasonable Use

Though questions of ownership and reasonable use are not unique to this context, they inhabit a prominent place in debates about the development of personal digital technologies and the use of these technologies in research. The commercial interests of technology companies and their legal agreement with their users, which are set out

in the terms of service, terms of use or the terms and conditions statements, do not always balance perfectly with the expectations of users and customers. This research was carefully designed in order to best comply with the legal and ethical regulations set out by the university and Twitter, as well as adhering to best practice set out by relevant research associations including the Association of Internet Researchers (see Chapter three for a complete picture of the ethical assesment and a description of the context) .

The key question surrounding the use of data for research purposes is “What can users reasonably expect in terms of data use?”. Terms and conditions statements, as well as service agreements and other legal stipulations related to the download and use of particular technologies (including social media platforms) are often lengthy and highly technical, beyond the comprehension and attention levels of large segments of the user population (Moran, Luger and Rodden 2013). To answer questions about reasonable expectations then, the terms and conditions are often not the best place to look for well-informed, ethically defensible and user-friendly research. The question of reasonable use is exacerbated in the case of ambient data, as it would require some imagination on the part of the user. The specific question of user expectations has not been explored in great detail in this thesis, as it fell outside the theoretical scope of the project, but these considerations (about user expectation) were built into the ethical framework that is outlined in section 7.6.

7.3.4 Shifting contexts (Ambient Data)

The transferability of information from one context to another is an overarching concern of this thesis. The use of ambient data in particular, which is the use of data generated for one purpose in a context beyond the bounds of that context, presents both practical and ethical challengess. Data of various kinds is now available to “anyone who is curious, regardless of their training” (boyd and Crawford 2012).

“Anyone” can be a disturbing concept in this context. Not only due to concerns about access to personal data and the preservation of anonymity, but also because the contexts in which data is being used or analysed are completely unknown by users at the time of production. Even in the case where a Twitter user has carefully read through the Terms of Service and is fully aware of both their own rights and the rights of Twitter, it is impossible for them to know the final destination of their content. This highlights an imbalance of power inherent to digital technologies as well as in research based on publicly available data generated through use of these technologies. Ambient data, which is queried in a context external to the one in which it was created, is particularly prone to these imbalances. The potential for misuse, misrepresentation and ethically-questionable dissemination of aggregated data is significant enough to cause concern.

The prominence of the text content of tweets as opposed to metadata is a distinctive feature of this work, and the decision to devote attention to the one over the other was not made lightly. This research has intentionally focussed on narrative aspects of place in online contexts in order to counter data-driven rhetoric that places a higher analytic value on insights derived through exclusively quantitative methods. Although space was given to analysing individual tweets, the effect of removing the data from the contexts in which it would ordinarily be produced or consumed cannot be ignored. Research using ambient data builds fragmentation into data sets. The tweet text analysed in this research is, as Riessman would argue, “fractured” (Riessman 1993, 4) and is not a complete story. Practicing reflexivity is, in these contexts, essential; critical engagement with the impact of the research itself helps to mediate power imbalances that are inherent to research of this nature (M. Jackson 1998).

The lure of “Big Data” and its associated narrative (explored in the literature review chapter) was an influential factor in the decision to focus on the text-based stories of platform users. This research shows that metadata is not necessarily the best source of spatial information,

which is particularly true when seeking to understand how individuals explore their relationships with place through social media. Text data is the logical alternative – it falls within the same data range (and hence the same accessibility and ethics umbrella). Ambient text data is a special category of data and its use raises particular questions that stretch the limits of ethical guidelines. In this project, for instance, the separation of meta-data from text data, and the removal of multimodal content (still images and GIFs), as well as the transference of tweet text data into CL software had practical implications for the research. This approach, though it aimed to demonstrate a human-centred analysis of the data, removed the entire context of the platform, so that any conclusions about the density of information and the visibility (or invisibility) of certain posts to the average Twitter user was pure conjecture on the part of the researcher. What the reader of this thesis must keep in mind is that the data here was extracted for a particular purpose and does not portray a general experience of place-based narratives on Twitter – it shows that they are present, varied and can be highly informative to certain projects.

7.3.4.1 The Power of Research Technology

The use of corpus linguistics approaches in this thesis enabled the interweaving of data-led and human-led analysis that was essential in generating insights into the digital documentation of human relationships with urban spaces. Corpus linguistics provided a number of methodological and analytical tools that enabled this research to be deeply intersectional. Data could be reviewed at a glance, or viewed up close in minute detail; this facility – of being able to take a general view as well as assess the data with high granularity undoubtedly aided the tessellation of concepts from a multitude of disciplines.

A corpus, much like a map is a sample of lived experience, representing only a small segment of the populations that may engage with a particular space or mode of communication. The map and the corpus are both separated from the contexts which led to their creation, they are, in a sense, both abstractions of something that is constantly

evolving and to isolate these snapshots in order to ask questions about what they reveal is not unproblematic. Just as a map is a curated representation of a place, a corpus is similarly limited. They are both limited because they cannot include everything, but technology is expanding these limits.

The high incidence of linked content was the key indicator that Twitter is a deeply multimodal platform. CL provided an integrated approach as well as software that enabled the incorporation of this multimodal content in the analysis of data. Although CL facilitates “zooming in” (Deshors 2017, Sealey 2012), in order to get beyond a basic acknowledgement of the presence and importance of linked content, it was sometimes necessary to go beyond the limits of the CL software – to follow the link and review the content in tandem with the tweet text.

And yet, concerns over the use of CL as a method of analysis have been raised by several scholars (Baker 2013, 17), where the restricted view of CL software, the separation of text from context, researcher bias and the distance between participants (tweet authors) and observers (researchers) are viewed as problematic. The textual emphasis of CL also plays to the anxieties of theorists who stress the multimodality of human communication and understanding (Scollon and Scollon 2003, Barthes 1967).

Any Corpus-based study of language and communication is restricted, as language does not exist exclusively in either a verbal or written form. Language is deeply multimodal and utterances can derive a lot of meaning from their physicality – from the location of the speaker or writer and hand gestures or facial expressions – often all coinciding in space and time. Baker supports this view, arguing “discourses are not confined to verbal communication” (Baker 2013, 17). Their reliance upon the contexts in which discourses are created has greatly influenced this research – which aims to continually remind the reader that the relationship between location, people and technology is

complex and continually evolving, teeming with conflicts of power, authority and identity.

The use of concordancing did not provide an impression of the conditions that gave rise to a particular text. As soon as a corpus is compiled, it is removed and isolated from the social and technological conditions under which it was created. In this study that would include Twitter and social media – which had a clear impact on the text content of tweets, made obvious in frequency and keyword lists as well as during case-by-case analysis and the dissection of individual tweets. This discussion builds on arguments made about technological affordances in chapter two (section 2.3.4), which insist that the context of social action (particularly in a technological sense) holds great sway over the decisions individuals make about self-expression, self-representation and communication. Although CL software provides a means through which to compile, manage and query large samples of language, which can help to reduce researcher bias by enabling statistical and objective analysis of patterns in that language, the approach remains fundamentally biased.

Steps can be taken to reduce this bias through the use of pre-compiled corpora, for example, but the querying of a corpus remains inherently inclined to answer particular questions. It is the researcher who sets the limits of a concordance, who determines the bounds of collocation, who decides whether to distinguish lowercase word forms from uppercase ones and who sets the thresholds for word lists. It is the researcher who must set the parameters for minimum frequencies, who chooses the appropriate reference corpus, and who selects which calculation to use in the sorting of keywords (logDice or Mutual Information score for example).

7.3.5 Distance between Researcher and Subject/Participant

The use of the methodological approaches used here are additionally troublesome due to the disconnect between researcher and subject. For the same reasons that CL software can be beneficial in aiding the

objective querying of language, they raise questions not only of informed consent but also about the reliability and validity of insights generated by these means. Particularly for qualitative and mixed methods approaches, like the ones used in this research, the separation of text from context and the alienation of text from its producer are difficult. The data cannot be queried or clarified with the user when user names are not gathered and express consent has not been provided. The intent and meaning of a user at the moment of posting is assumed by the researcher, based on the habits of their language and with trust in the openness and frankness of their stories.

Some scholars, including Baker (2013) also claim that there is an over-emphasis on frequency as a determining factor of discourse when using CL methods, and that this approach does not properly consider the weight or power of certain voices or modes of communication. This criticism is particularly relevant to the work of this thesis, as the use of anonymization and ambient social media data does not allow the sources to be investigated as factors in the construction of discourse. Everything in this thesis has been assumed to be of equal weight, with repetitions and retweet being the only factors which indicate the dominance of certain voices over others. With regards to the mode of communication – this thesis only considers a single mode. Although Twitter is multi-modal in the sense that it can include images and linked content, the focus here has been limited to text, which, as discussed in the literature review, forms only a small part of a much richer assemblage of “communicative options” (Marwick 2013) available to the users of social media platforms like Twitter.

7.3.6 Dominant Voices & Representativeness

Uptake of Twitter is not universal and the analysis of data generated through its use in urban settings is by no means representative of a whole sample of the human population of either of the cities discussed here. These trends in use, uptake and the segmentation of user groups have been discussed more fully in the methodology chapter (section 3.3.5). Twitter users represent a fairly narrow segment of the total

population of the British Isles; the platform has greatest uptake amongst those aged 34 and under, and while male users are proportionally under-represented (McCandless 2012), their Tweets are more likely to be retweeted and liked, which means that the actual seen content of Twitter can be largely male-dominated (Usher, Holcomb and Littman 2018). These gender disparities have not been a focus of this work, but the imbalances they indicate should be an important consideration in further work using data from the platform.

Similarly, analysis undertaken in the second and third stages revealed that keywords appearing in only the search corpus (therefore with a high keyness score) tended to be usernames, whereas those with high keyness scores that were present in both the search and the reference corpora were not. This pattern of variance demonstrates how social media flows (of information) can be easily dominated by a narrow range of voices, to the point of the exclusion of alternative perspectives and speakers.

Concerns have also been raised by some commentators who suggest that the platform has an urban skew (Lehmann, et al. 2011). The focus on urban environments throughout this thesis leaves a gap in the research concerning the utility of the platform in mediating non-urban lives, which could be explored in further research.

As Fiss and Hirsch pointed out, no data can ever really be “raw” (Fiss and Hirsch 2005), it is shaped by the contexts in which it is produced, which includes technological, social and demographic conditions. Just as the technology should not be viewed as inert, nor the data it produces as “raw”; it should also not be assumed to be temporally stable. Some industry reports since 2014 have been quick to point out that even though monthly active users (MAUs) may be increasing, individual user engagement (in terms of activity) is declining (Popper 2014) and the platform may be entering its “Twilight” (Lafrance and Meyer 2014). These claims have been challenged, and the continued existence of the platform does suggest that the platform has a longevity

often absent in similar platforms, which have been short-lived (Yik Yak, for example). The balance of all of these specific challenges depended on the effective establishment of reflexive research practice, which recognised that the social and technological conditions under which this research was conducted are continually shifting.

7.4 Twitter: A “suitable” data source?

The research question, along with the associated research objectives and the methodology were developed partly as a response to a popular attitude perceived by the researcher that unfettered access to social media data streams and their analysis on a huge scale (think AI and ML and capitalised Big Data) hold the key to unlocking the secrets to a more productive and more efficient society. The aim was to look more closely at the potential of the data streams that are often the focus of these tropes and to assess whether they could provide insights into the complex, emotional and social relationships that individuals form with the spaces they encounter. The questions of suitability, then, were ones of methodological feasibility, ethical conceivability and theoretical applicability. Did Twitter text data, for instance, provide enough data from which to explore questions of narrative? Did Twitter’s corporate guidelines (Terms of Service) allow for the collection of user data for research purposes?

Twitter was the perfect data source for this research as it facilitated the exploration of nearly all the practical and ethical challenges common to the use of social media data in research. Practically, Twitter data is easy to get; whether through the (free) developer’s API or the (paid-for) Firehose, there very few clerical steps to obtaining data from the platform, and it can be done with relative ease by an untrained individual.

The main barrier to access through the API is the necessity of a Twitter account; these can be set up easily with an email address. Access this way does not require a great deal of skill, as there are several

applications that facilitate the capture of Twitter data through the API once an account is held (NCapture and FireAnt[34], for example).

7.5 Constructing a Coherent and Replicable Methodology

The aim of constructing a coherent and replicable methodology has been achieved with a degree of success. The progress of the research was not hampered at any stage by practical or theoretical obstacles; in that regard, the methodology was a success. In addition, the method has been described clearly and uses technology that is still (at the time of writing) in use, in that regard, it has also been successful. Technological obsolescence is an ever-present concern of technology research, particular with regard to social media platforms, which have a relatively high rate of attrition.

This research was conducted in three methodological stages, each was informed by the previous stage and each stage aimed to ascertain how different data gathering parameters and approaches to analysis impacted the insights gleaned from the exercise. The first stage was consciously designed as a test of both the python programme code and the iterative coding technique described by Janet Hood (the “Generic Inductive Qualitative Model, GIQM) (Hood 2007). The first stage proved that both the code and the coding method could be deployed within the given timeframe and within the ethical framework laid out. It also revealed a number of patterns in the body of tweet texts that hinted at a complex and varied spatial engagement by users; for instance through the inclusion of place names in tweet text and the prevalence of experiential and temporal details (see section 4.5.5). The analysis also highlighted the tendency for the place names Nottingham and notts to be included in the leading or middle third of a tweet rather than in the final third. This indicated that place names were not routinely added post-script, as an annotation to the tweet text, but rather integrated into the content itself.

[34] http://www.laurenceanthony.net/software/fireant/

The second stage was designed in a way that capitalised on the volume of narrative data available through Twitter, while also making use of the geographic filtering capabilities available through the platform's API. The use of a tailor-made reference corpus at this stage permitted the comparison of regional data sets to each other, and highlighted the words that were unique to each.

The third stage focussed on the use of locative (hash)tagging, which was a pattern evident in the data from both the first and second stages. A smaller sample was collected than at the second stage, as the focus on balancing the analysis of individual tweets with insights generated from CL methods. The third stage was successful in this regard, as the approach

As each of the stages was designed concurrently, the research design was iterative; even though the research question was determined at the outset, and the methodological approach set out in the early stages of the project, each individual stage was developed based on reflective practice, and were intended to refine the suggested methodology.

The framework of the method presented here, which was developed in iteration, is indicative of the ethnologically informed approach. The most important shared elements are the recognition and elucidation of the technological context (Agha 2007), sustained reflexive practice (M. Jackson 1998) and a focus on experiential and emotive elements (Ingold 1993, Tuan 1977).

This thesis demonstrates it is possible to resist the rhetoric of Big Data while undertaking social media research and contributes to a growing field of mixed-methods research into the significance of social media platforms in the negotiation of contemporary identities (Bell 2012, Boellstorff 2008, Brock 2012, Lee 2015 and Page 2010). It describes how a variety of contrasting theoretical and methodological approaches can be effectively integrated in the pursuit of human-centred research design, which facilitates the investigation of complex social and technological conditions.

7.6 Building a Suitable Ethical Framework for Use in Interdisciplinary Research

The planning and undertaking of this research has provided a useful and widely applicable ethical framework. It takes the form of a series of questions which can be easily extracted and applied in similar research seeking to make use of social media data streams or using data from the digital realm more broadly.

The questions listed here enable the consideration not only of the institutional ethical codes to which a researcher must adhere, but also prompts the consideration of the data supplier in their use of the digital platform in question. The questions encourage reflexitivty rather than rigidity and aim to provide a guide to conducting creative and empathetic research.

1. Can the researchers be identified? Will the researchers make themselves known?
    a. What are the effects of research observation/researcher presence?
    b. What role should the researcher play in data gathering?
    c. What can be gained by absence or presence?
2. How can informed consent be obtained? Is declared consent necessary?
    a. Can researchers ensure participants understand the demands being made of them?
    b. What is the trade off for the individual participant as far as involvement/non-involvement in the study?
3. How confidential is the data?
    a. How will the data be protected (technical parameters)?
    b. Who is entitiled to see the data?
    c. What does the participant understand by “confidential”?
4. Is the data anonymous? Are individuals identifiable?
    a. What constitutes anonymity?
    b. Who benefits from anonymity?
    c. Do participants want/expect to be identified (is the project giving them a voice)?
5. Has the project considered institutional ethical guidelines?
    a. Has the research been designed with the ethical guidelines and requirements of the funder or sponsor in mind?
    b. Have any differences in established guidelines been addressed?
6. Has the project been sensitive to the complexities of the medium?

a. What aspects of the research are novel?
    i. Do these aspects present the researcher and/or the participant with unexpected challenges or burdens?
b. Have relevant examples of research best practice in this area of study been reviewed and referenced?
c. Is [platform] a public or a private space?
    i. What are the expectations of the participants regarding the viewership and accessibility of their data?
    ii. How are public and private spaces performed/demarcated?
d. Who owns the data?
    i. What are the ownership claims of the platform being used?
    ii. Are there any copyright issues?
    iii. What rights does ownership confer?

Questions one through five should be recognisable to most researchers who have conducted ethical review and submission processes. These steps cover well-discussed tenets of ethical research practice which include beneficence (minimising harm), consent (obtaining informed consent without deception) and protecting anonymity. The steps are intended to facilitate a review of planned research, and aid in the design of research where the researcher acknowledges the role they will play in determining the research environment and shaping the framing of participant experiences. These steps are relevant to all research involving human participants. Research involving human data, and in particular research involving human data with no direct contact with the authors of that data, as was the case in the research reported here, requires consideration of a number of other factors. These factors should be guided by the specific research context (the medium). The case of Twitter here meant the careful consideration of researcher access, participant expectation and the weighing of corporate and institutional conventions for the use of data. Anonymity, for instance, is usually insisted upon by University ethics comitees, unless attribution is explicitly requested by a participant, whereas Twitter's "Developer Terms" state that Tweets "are one of our most visible brand elements" and should be displayed correctly, in accordance with the display requirements they set out, which state (edited for brevity and relevance):

"**Do**:

- Display real, unmodified tweets from real accounts
- Display the Twitter logo

Don't:

- Modify Tweet text

Tweet Author:

- The tweet author's profile picture, @username, and display name must always be displayed and link to the user's Twitter profile"

These requirements had to be balanced with ethical requirements of the university as well as my personal convictions as a researcher, personal digital technology user and Twitter user. The decisions to modify Tweet text to exclude @ usernames as well as removing user profile images and the Twitter logo from the tweets quoted in this research were not taken lightly, and it is expected that other researchers will come to their own conclusions concerning the challenge inherent to interdisciplinary research in complex techno-social settings.

7.7 Strategies by which Place is Incorporated into Social Media Texts

Evidence presented in this thesis demonstrates that tweet authors use a variety of different strategies to talk about place through social media. Some of these references to place are explicit, and others less so. The clearest examples of intentional exploration of place by Tweet authors were related to regional history and local heritage. In these tweets, location was often invoked through the inclusion of place names, both tagged and untagged, as well as the inclusion of an event, and often, an image relating to the text.

7.7.1 Mentioning Place Names in Tweet Text

The literature reviewed in chapter two reflects the multitude of ways in which locational reference can occur, from pairs of coordinates on a navigational tool to the use of vague textual reference (area, for example) and colloquial place names.

Where previous work concerning text-based locational reference has utilised purely automatic processes to identify and extract this geospatial data from text databases, an effort was taken in this research to develop an approach that permitted the exploration of the text at a variety of levels. This ranged from a tweet-by-tweet reading to the automated generation of comparative lists and even covered the investigating of words occurring only once in an entire corpus.

The fully-automated querying of text often involves the use of gazetteers or supervised machine learning, or a combination of these two methods (Leidner and Lieberman 2011) in order to identify expected or pre-defined forms of locational reference. The process employed here, of hand-identifying and manually parsing locational references from corpora of varying size and character, though very labour intensive, provided a view of the whole text. This meant that each tweet (and every single tweet in the case of the first and third stages) was read and analysed in its entirety. In this way, each tweet formed part of a narrative that evolved throughout the analysis – every one providing more detail of the overall picture, similar to the pointillism described by Page in her examination of Facebook status updates (see section 2.2.2.4) (Page 2010). This approach also permitted the identification of vague and indirect locational referencing through the use of regional nicknames, the abbreviated names of football team managers (#rafa) or the inclusion of locations in the letters of an acronym (NUFC and NFFC)[35].

At both stages one and two the sample sizes permitted the querying of positionality of location names in tweet text. This analysis was undertaken in pursuit of the idea that it could reveal a pattern in the incorporation of either untagged or tagged locational references in tweet text. No such trend or pattern was obvious, and though this was unexpected, it was revealing in that it made clear the use of locational reference for non-meta purposes, and indicated that both un-tagged ad

[35] Abbreviations such as NUFC and NFFC are examined in section 7.7.1.1, whilst (hash)tagging as a strategy is explored in section 7.7.2.

tagged locational references are likely to feature in all segments of a tweet. However, there were some instances in which positionality seemed to be important, and this was found in the use of tweets as an avenue to parody, where tagged locational reference was used to reflect on and poke fun at the reputations of places (evidenced in section 5.6.1 and further discussed in section 7.7.2).

7.7.1.1 Abbreviations, Nicknames and Mistaken Identity

Abbreviated place names were designed into the first stage of this research, which collected and compared tweet content based on the inclusion of the search terms “notts” or “nottingham”. However, abbreviated place names and other abbrevations with a spatial aspect were evident in the data gathered during all three stages of research, no matter what the data filtering method.

In the third stage, described in chapter 6, it was obvious that place-name filtering filtered a lot of erroneous data into the sample. Shared place names are common, and the problem seems to be made worse by the concatenation made necessary by the use of hashtags, so that place names ordinarily differentiated based on their structure became, to the search function of the Twitter API, the same. New Castle and Newcastle, for instance, both ended up being included in the stage three data sample; the addition of a hashtag, the undifferentiated capital letters and the absence of the space in the tagged reference (#NewCastle) made the two impossible to differentiate to the rudimentary API technology. Any instances of accidental inclusion were verified based on the contextual information of the tweet. Fortunately, many of the tweets containing place names that were not intended for inclusion also contained additional detail regarding location, including two-letter US state abbreviations (see section 6.4.6.5 for examples).

7.7.1.2 Vague vs. Specific Reference

Mention of very specific locales were not uncommon. The mention of individual streets or buildings was frequently paired with historic references, as part of local history or heritage narratives (see section

6.4.7.3). Often, these references also included linked content, usually a photograph. The prevalence of linked content in these tweets supports arguments that stress the importance of linked content and the inherent multi-modality of identity narratives (Lee 2015).

Very specific spatial references were not evident in the data. Geo-coordinates, for instance, did not appear at all in any of the tweet text, nor did any incidences of the grid-based referencing system described in section 2.2.2.3, “What3Words”. Postcodes, especially partial post codes were apparent in the data. These combinations of letters and digits were employed in tandem with other spatial references, as shown in the example on page 115, where the partial postcode “NG” is preceded by a hashtag and appear in a list of other place names; and in Table 6.3, where it is listed as a common collocate of of #nottingham at position n+1. The use of partial postcodes rather than complete ones suggests that complete postcodes may be too specific for the purposes of twitter text, which might also help to explain the absence of any What3Words references, although in the case of What3Words, it also indicates that though this emergent mode of spatial referencing is useful in logistical contexts, it may not naturally enter the vernacular or day to day chatter on social media, where a series of random words would not be quickly identifiable as spatial references.

7.7.2 (Hash)Tagging Place Names in Tweet Text

The use of hashtags in combination with place names was established early in the research. The data sample collected at stage one indicated that place names were a useful filter for social media text and gave a clear sign that the (hash)tagging of place names was common in the authoring of tweets, particularly with regards to advertised services.

The incorporation of (hash)tagged locational references such as #nottingham, #newcastle and #NG into the main body of tweet text suggests that hashtags can be used as both facilitators of meta-commentary and as assistive devices for improving search-ability and visibility of located services or offerings. Improved search-ability was,

after all, their original intended purpose (Solis 2011). The process of (hash)tagging places, and the prevalence of (hash)tagged place names in the sample explored over the course of this research presents some evidence that supports existing research positing the hashtag as inherently social. It is a device that connects individuals geographically isolated by conditions of supermodernity (Auge 1995, Iyer 2000) to a point in space, thereby forging cultural movements (Clark 2016, Lee 2015) and curating localised public memory (Bell 2012, Haskins 2007, Landsberg 2004). However, it is worth noting that the majority of hashtags in the sample appeared only a single time (see discussion in section 8.1.1), so although these hashtags may theoretically help individuals connect with locations, and provide researchers with fruitful data samples, the high level of variance suggests that the exploration of place is not the primary aim of platform users. Narratives about place on social media are usually ambient; not written or curated with the task of place-making in mind, but useful none the less in the exploration of place.

7.7.3 The Use of Maps, Drawings or Photographs

As highlighted in chapter 2 location can be referenced in a variety of different ways, not all of the strategies of reference discussed in there were evident in the sample, nor were they sought out during the course of this research. Some of the linked content, which was explored during the qualitative analysis and thematic coding of individual tweets, demonstrated that many tweets have a visual component that works together with Tweet text to express meaning or sense. This linked content could be overtly spatial; a map, address or series or directions, for example, but more likely it would take the form of a photograph or some other hybrid content such as image macros.

Extensive existing literature shows that place is often explored visually. That could be through the use of drawing, painting, sculpture, projection, mapping, photography, LiDar imagery or any other of the myriad ways (see sections 2.2.3, 2.5.5 and 2.5.6 for a fuller discussion).

The referencing of location and place through visual means (photography, art or maps) was largely excluded from this work due to the data capture parameters, and the focus on textual descriptions of place. The investigation of multi-modal location-referencing strategies on social media is a potential avenue for further work, noted in section 8.5.

7.7.4 Assumed Proximity: Geolocation

One of the questions that remain is whether Twitter users consciously and intentionally enabled the geolocation of their tweets, or whether this information was shared because of the platform's default settings. The enabling of geolocation may be a strategy employed by users who wish for their narratives to be easily connected to locations, but it may be also be reflective of the power of inertia in the set-up of digital technologies, or an indication of the obscurity of platform affordances to some audiences.

7.8 The Utility of Twitter Data in Informing Understandings of Place

Technology of all kinds generates new possibilities for being "in the world" (Heidegger 1978) and collapses the boundaries between things; including work and play (Gale 2009), geographic distance (Cosgrove 2006, Auge 1998), and time (Federman 2005). According to Federman, these conditions also work to destabilise shared values and meanings (Federman 2005) and contribute to the production of non-places, where individuals are isolated from space, time, and meaning (Auge 1995, Iyer 2000). Twitter is an excellent platform from which to approach questions related to place as it is an established social media platform with international reach, high numbers of monthly active users (MAUs) and is available through personal digital technologies like smartphones and smart watches. Although the literature suggested that the increasingly overlapping of humans and technology (Lupton's "cyborg Assemblage") was problematic, it also highlighted the possibility for these technologies to have positively effect change and to expand the expressive and communicative options of marginalised individuals (Gil de Zuniga, Jung

and Valenzuela 2012, Haskins 2007, Hasler and Ruthven 2011). The data analysed here suggests that alienation is not a necessary condition of a cyborg existence and that social media data can usefully contribute to both shared and individual understandings of place.

As demonstrated in the literature review, “place” is a difficult concept to wrangle with. In casual language, it is often used interchangeably with “space”. Recognising this fuzziness was a key part of the early stages of this investigation and it is because of this fuzziness that the concept has explicitly differentiated from “space”. As a consequence of the concept of place being a fuzzy one, its interpretations and expressions can be hugely divergent but are sometime intertwined with one another. The three overarching threads identified in the literature review, “experience”, “juxtaposition” and “time” are not entirely discreet and they do intersect. For clarity, each of these themes is addressed here in turn individually, before their combined features and expression in social media data is further explored in sections headed “flux” which addresses changeability and conflicting ideas; “identity” which addresses the emotional and historical aspects of place and “creativity” which discusses the unexpected.

7.8.1 Time

The text investigated over the course of this research supports the argument that time plays an important role in influencing both tweet content and the narrativisation of place online. Time as an aspect of place was primarily expressed in two ways – through the reportage of events (upcoming or happening at the present time) and through historical commentary based around particular sites in each city. Unlike the data project “All the Minutes” discussed in chapter two (section 2.5.3), which showcases the inclusion of exact hours and minutes, mention of specific clock times was rare. Other mentions of time, including “ICYMI” (in case you missed it, see later in this section for a discussion of this phrase) did not show a strong relationship with either locational reference or the expression of a place-based identity. However, they did help in emphasising the fast-moving and ever-

changing nature of the exchanges and flows of information and emotion online. Mentions of time were not always interwoven with discussions of located experience and place, but they did invariably provide a useful frame from which the surrounding text (and sometimes images) could be read and understood.

There were a number of tweets in the corpora that evidenced use of the platform in the exploration and recording of historical sites, buildings or events. Some made plain and open references to historical events or documents, thereby connecting sites in the city to an era or event in the past. The inclusion of historic images in tandem with these captions seemed to be a common feature, though these images were not captured or stored with the text. The following tweet, taken from the Newcastle 1000 tweets in stage three, is an example of this combination:



Figure 7.1: Image paired with Tweet text

The backwards gaze of some tweets was not always specified with dates (1977), nor illustrated with historic images. The phrase “on this day” and either of its variants “#onthisday and #OTD” balanced this backwards gaze, through the close connection of the present time with historic events.

> *“Easy!... on this day in 2014 Nott'm Forest won 4-0 at home to Reading in The Championship #NFFC #DayToRemember #OTD”*
> *(Nottingham Geo located corpus)*

The use of these phrases featured in both the geo-located corpora in stage two, highlights the cyclical nature of time and suggests that Twitter is used by some as part of a reflective exercise in understanding the cities they inhabit. The strong connection of the “OnThisDay” phrase with sporting events highlights a trend common in sporting literature concerning fandom; it demonstrates the practices of “Basking in Reflected Glory”, which is deemed to be a key feature of “fandom”, particularly in relation to football (Dionisio, Leal and Moutinho 1998).

An urban centre is not historically isolated, and no matter how histories may be appropriated and put to use by both commercial and individual actors (see section 6.3.4.2) nor how “in the moment” Twitter is expected to be, a number of examples here suggest that social media platforms, including Twitter, are used by individuals and organisations to revisit and assess these sometimes unpleasant histories (see section 5.4.2.2).

In addition to the backwards-gaze of tweets concerning aspects of heritage in each city, there was a clear trend towards the discussion and broadcasting of time-limited experiences through Twitter. These time-limited experiences have been classed as “events” and were a prominent feature at all three stages of the research.

The prominence of events across the corpora compiled at all three stages reveals the significance of time-limited experiences in the context of a city. This trend was reflected in the reporting of events and activities, like shopping at Christmas time in the data collected during

stage one and tweets about the Olympics in both the geo-located corpora in stage two. In stage three these events and activities were evidenced on a more local scale - through the mention of the “goose fair” and a “puppet festival” in the Nottingham corpus and “comicon” in the Newcastle corpus.

The backwards gaze of some tweets in combination with the use of the abbreviation “ICYMI” is a testament to the rapidly shifting environment of social media. “ICYMI” is an initialism of “In Case You Missed It” and shows awareness amongst Twitter users of the ephemerality of content on the platform, it is an acknowledgement of the speed at which things can “disappear” or become irrelevant on social media platforms. In their discussion of tweet “in-flow rates”, Gomez-Rodriguez, Gunmadi and Shoelkopf estimate that some users may receive more than 500 tweets per day (Gomez-Rodriguez, Gunmadi and Schoelkopf 2014, 173), which significantly impacts the time they have available to read and interpret each one, let alone to form a response to a single message.

The fleeting nature of these communications has both methodological and theoretical implications. Methodologically, this ephemerality raises questions about the visibility and accessibility of Twitter data over time. This indicates that the “privacy through obscurity” approach to personal data on social media, as discussed by Hartzog and Stutzman (Hartzog and Stutzman, Obscurity by Design 2013) (Hartzog and Stutzman 2012) might be sufficient in protecting the individual from criticism or infamy.

7.8.2 Experience

The lived experience of space was identified at the beginning of this work as a key factor in the theorisation of place. Exploring this experience using social media text presented a number of opportunities, and analysis revealed several ways in which this experience is documented through text. The “dwelling perspective” (Ingold) was demonstrated in the data through the simple reportage of everyday events. On Twitter, people broadcast what they have been, are doing or plan to do in the future. These minutiae were showcased in the data at

each stage, where tweets included reference to shopping, exams and other mundane activities in and around the city.

The use of first person perspective was a recurrent theme in tweet text, demonstrating the close connection of individuals with the narratives they produce in digital contexts. Through the recounting of the day-to-day occurrences, from travel to checking into hotels, individuals build up a story that, over time, forms a database of their experiences. Some of these experiences are consciously told in relation to location, where spatial reference is deliberately included in tweet text. Some connect to location passively, through tweet metadata, geo-location and the location mentioned in the user's profile.

Reflecting a faith in the existence of a community that shares the values and the locality of the individual, reports of missing people and lost pets (dogs and a bird) were a prominent feature in the text data. Twitter users were rallying an audience in the hope that a located community could be mobilised and these people and pets found.

7.8.3 Juxtaposition

Juxtaposition, in this discussion has been theorised as the effect of overlaying multiple viewpoints, it is the consequence, in social media data that includes locational reference, of the combined reportage of many individuals about a single location. The stories are not necessarily generated simultaneously, although some of the timestamps in the data indicate that they build up in quick succession, but they do tell of the same places, so they can be overlaid in order to provide a "total" view that is always evolving and subject to change. The designation of juxtaposition as a key aspect of "place" is facilitated by the use of Twitter data because it demonstrates the ways in which place identity is established over time by many actors and stakeholders. Although there are certainly ways in which place-making can, and is, often the result of conscious efforts to boost an area's economic or cultural standing (the Nottingham campaign to be named European Capital of Culture 2020 was a timely example of this kind of activity) – in the melee of social

media texts, place-making is much more nebulous. This means that many different actors (taking into account the demographic limits of the platform) can contribute to understandings of place through Twitter, whether it is intentional or not.

The reports of experiences and services that were a feature of the data discussed in chapters four, five and six, raise some questions about the origin of the voices contributing located narratives on Twitter. Who are the individuals whose voices are being layered here? The method chosen here meant that user names were not collected and individual users were not identified, except in a few cases where the usernames were used to help in the reading of tweet content. As a result, the diversity of perspectives is taken-for-granted rather than evidenced in the data. It is worth noting that it would not be unheard of for the located narratives reviewed here to have been created by transient populations, including term-time student residents or tourists. Tim Gale's discussion of the digital environments that facilitate the incorporation of an exotic other (traditionally at the centre of touristic experiences) into everyday narratives is particularly relevant to this thesis (Gale 2009). Gale argues that under the conditions of modernity

> *"the distinctions between work, rest and play and the spaces and times in which we would expect to encounter them become blurred, new tensions emerge between what we consider to be strange and what we consider to be familiar, and touristic experiences become very much part of everyday life"* (Gale 2009, 121).

Gale also discusses the difficulties of determining what is true, real or meaningful in a post-modern existence. To Gale, the post-modern condition is

> *"characterised by the mediatisation and aestheticisation of the material and symbolic aspects of any given way of life…and the apparent ease with which these may be mechanically or electronically reproduced so as to be gazed*

> *upon by others, often out of their original context. " (Gale 2009, 122).*

The introduction of "the exotic" other into the mechanical and electronic contexts of digitally-enabled narratives leads naturally into questions about authenticity, authority and validity. How can opposing voices or opinions be reconciled? There is no simple answer to this question, but the technologies and approaches presented throughout this thesis have demonstrated that consensus is not always crucial – understandings of place, particularly in the context of fast-moving digitally enabled exchanges are highly variable. But their variability does not make them invalid, the view simply has to shift from "a view from above" (Brannon 2013) to a horizontal or human-centred view.

The shift towards a human-centred view in this thesis was informed by previous theory and methodological engagements. Actor Network Theory, for example, was helpful in breaking down perceived barriers between "subjects" and "objects" and showed that networks of things and people could be analysed outside institutionalised hierarchies (Law 1993, Latour 2005). Chapter two links these ideas with intersubjectivity (M. Jackson 1998) to present an interdisciplinary framework that permitted a reflexive engagement with the research approach. Similarly, Least Heat-Moon's "Deep Maps" provided a framework that bridged GIS technology and layering with human stories (Least Heat-Moon 1991).

7.8.4 Flux

Documenting the ebb and flow of human connections with location over time is not easy, and though the data presented in this thesis has shown it is possible to record even the most fleeting engagements with place, where the analytical challenge lies is in the fabrication of a meaningful narrative from all this data. Flux is an idea common to literature about urban spaces and particularly sites that are hot beds of technological and human interaction like cities (Balmori 2003) (Castells 1996) (Reed 2002). The cities selected for use in this research, for

example, both have large student populations (post-18 education) and are home to a number of venues that can regularly host musical or sporting events, both conditions which encourage the coming and going of usually disparate groups of people. The idea of flux links time with change and is an important part of this discussion as it has a bearing on both the utility of the methods used here and the utility of the body of evidence presented.

The data presented in this thesis represents a small, time-limited view of the textual data available through the Twitter API. The content of these corpora would change depending on the calendar date and the sample provided through the API. Although it may compromise the replicability of data gathering, the high rate of change in Twitter sample data is not necessarily negative. Transience is a common feature of the theorisations of both the city and technology (Lupton 2015) (Lyon 2010) (Castells 1989) (Adey and Bevan 2006) (Solnit 2010) (Wood 2012) (Wilson and Peterson 2002) (Mumford 1961). The mirroring of this quality in the method adds a depth to this research, which encourages the researcher not to be dismissive about change or ephemerality.

7.8.5 Identity

The emotional and historical aspects of place can be combined with cultural activities, demographic distributions and the use of certain nicknames and abbreviations in the negotiation of identity. The stories woven around these activities, nicknames and abbreviations can be combined together to give a sense of what it means to be from, or local to, a particular place, and sets expectations for outsiders about the things that are deemed significant in those places.

The discussion of weather conditions in chapter six was an entertaining example of one such narrative – where the preference for “geordies” to wear light, summery attire in cool weather conditions was lauded and also linked to the nickname “toon” as well as the local football team “NUFC”. Similarly, the mention of Robin Hood in the Nottingham data

sets also raises questions about the continued role of myth and historical reference in the development of place identity.

The role of myth and imagined characteristics in the construction of specific identities was explored by Deborah Golden (Golden 2001). Though Golden was speaking about group identity, she emphasised the importance of public spaces in myth-making narratives, where well-known sites could form the basis for stories of an imagined future, thereby propagating the development of a stronger group identity based around a character from past, whose story could be projected onto an imagined future. Golden's time spent among Jewish migrant communities in America highlighted the prevalence of myth-making in public spaces and the centrality of both historic figures (like the Robin Hood of Nottingham) and public spaces in the development of a strong group identity. Used as a lens through which to view the information presented in this thesis, this myth-making recipe can be easily identified. Sports teams and personalities (NUFC, NFFC, Yedlin and so on) are the mythic characters, and Twitter is the public space used for this myth-making. The stories they tell together are about "Geordies" and "Novacastrians" and the great imagined cities that host them.

7.8.6 Creativity

The difficulty of understanding people who lived in different times and places is certainly not novel; differences in language, style and habits of documentation are highly variable and each case presents its own challenges. Creativity came to the fore through the difficulties encountered in deciphering the meaning of text in the sample tweets.

Understanding idioms and slang phrases is not a new challenge to linguists but the development of digital technologies grants an unprecedented opportunity to access and witness the emergence of new linguistic forms and phrases. This challenge is the primary concern of scholars concerned with situated discourse (Scollon and Scollon 2003) and entextualisation (R. Jones 2013). This section considers the use of abbreviation and idiomatic expression in digital contexts and

explores new methodological concerns involved in understanding context and meaning of online parlance. In a digitally enabled world, technologies have accelerated the timeframes of linguistic evolution, and foster the non-linear emergence and disappearance of linguistic expression. The temporal and demographic relevance of language is more fluid than ever before. Through an analysis of the self-styled "veritable cornucopia of streetwise lingo", Urban Dictionary (www.urbandictionary.com), this commentary discusses the difficulties of ascertaining meaning in digital contexts, where textual expressions must comply with the rule of the medium. It outlines the methodological challenge of using a crowd-sourced dictionary in order to understand tweet text, where definitions can vary, contradict one another, or even insist that there is no set definition, yet argues that the interpretation of colourful and creative language use online would be impossible without such resources.

Encounters with nuanced communication are inherent to social and cultural research of all kinds, where an in-depth understanding of the human environment is essential to good reportage and analysis. This commitment to understanding the context of an encounter, practice or artefact is exemplified in the work of Clifford Geertz, whose "thick description" clearly demonstrates the (Geertz 1973). Similarly, the historian George Evans, insists on the importance of oral histories in the understanding of the complexities and subjectivities of the lives of agricultural workers. When discussing the human sciences, Evans states "there are so many variables that the modern scientific method cannot be applied with the same confidence as in the purely physical sciences" (Evans 1970: 19). Theories of mediated discourse and entextualisation (R. Jones 2013) frame communicative events such as tweets as inherent to the formation of a complex and layered identities, which are constructed through narrative and communicative practice, as illustrated in the example tweets discussed throughout this thesis. Here there is an emphasis on the simultaneous importance of both action

(Lave and Wenger 1991) (R. Jones 2013) and thought (Clark 2016, M. Jackson 1998)

The communicative network in which Twitter sits is certainly different for every individual. The other nodes of this network have not been a focus of this research, but it is recognised that Twitter is just a single tool used in a digital toolbox of identity making that includes a myriad of channels, gizmos and gadgets, as well as individual and shared processes (see the crowd, section 2.3.8). The system extends even beyond the physicality of these things (from computers to wearable fitness trackers) into other cultural and social tools – it is these communicative tools that have been explored a little here, as their importance became clear in the analysis of tweet text. These shared cultural and social tools include humour, slang, linked content (like image macros and other memes) and hashtags.

7.8.7 Shared Resources, Shared Meanings and Individual Practice

Interactions through Twitter are not limited to text, but make use of a variety of resources available in the digital context, including emoticons, linked content, images and coded language. The presence of slang and the high incidence of RTs in the data exemplify how shared meanings and resources are circulated online and hints at their importance in the mediation of every day experience in contemporary cities. The presence of tagged place names in tweets indicates that geographic location is an important anchor for discussion of events online and that it serves as a device for referencing a shared understanding of a behaviour or occurrence.

The complexity of the language encountered in the tweets presented some challenges in interpretation at each stage of this research, which proved particularly problematic when using computational querying. Abbreviations in keyword and frequency lists demonstrated that simply listing the linguistic components of a tweet corpus was not sufficient in understanding the meaning behind those truncations. Some abbreviations, such as “#OTD” indicated participation in a shared

reflexive practice that has been common to non-digital communication for a long time (Sumikawa, Jatowt and During 2017) as well as supporting the theory that time is a significant aspect of place-based storytelling on Twitter. Other relatively obscure language, that was deciphered with the help of Urban Dictionary included “catfishery”, “kmt” (see section 5.5.2) and the more locally-specific “toon”. The utility of Urban Dictionary over the course of the analysis demonstrates not only how tweeters can indicate shared understanding, show allegiance to particular sports teams or place through their use of specific words or phrases, but also demonstrates the importance of recognising analytical resources that are relevant to the field of study.

The corpora analysed here suggest that hashtags are often carefully tailored by individuals in order to produce a meta-narrative of their own everyday stories and experiences. For tweeters, hashtags seemed to provide an ideological or comedic lens through which the rest of the tweet text could be “properly” understood or read – the example provided in section 5.6.1 is a nice example of this; where “#nottingham” is being used as a commentary on the experience described in the first section of the tweet. It reads very differently than if the location has simply been integrated into the rest of the tweet; “Ah, the sound of a police Helicopter in Nottingham” has a different emphasis. The hashtag in this instance was a way of poking fun and highlighting a humourous tone.

Though they may sometime disagree, these voices are essential to understanding the language of demographics groups that researchers may not participate in. Together, the definitions of searched terms on Urban Dictionary provide a picture of the complexity of human relationships with language, and leave little doubt about the significance of the hashtag as a discursive tool and a common feature of contemporary urban and digital life. The utility of crowd-sourced information repositories like Urban Dictionary proves there are benefits to be gained from engagement with publicly contributed content online

for application in social research. These resources would merit further investigation, which has been set out in chapter 8 (section 8.4.3).

Crowdsourced resources like Urban dictionary could help us to overcome the problem of understanding people in different times (whether that be a generational or historical difference) and in different places. Twitter is a chance to “listen in” on populations that are geographically diverse, and sometimes demographically alien to researchers, whether due to age or simply a result of internet memes that they are not aware of.

7.8.8 Big Data, Small Data

The volume of data available to both researchers and the general public through social media is very large. Social media platforms like Twitter generate millions of ‘bits’ of data every minute. In order to make sense of the flood of data, it must be sampled, stored and analysed in ways that are sensitive to the fluidity of content online. There would be, for instance, no sense in taking statements made on social media to be statements of an indisputable truth. The internet, and social media in particular, is driven by humans, and people are just as idiosyncratic in the digital world as they might be in the offline world.

In 2012, danah boyd and Kate Crawford declared “The era of Big Data has begun” (boyd and Crawford 2012: 662) they argue that Big Data is not only a thing (or a group of things) but a “socio-technical phenomenon” (boyd and Crawford 2012: 662). They describe Big Data as a phenomenon that encompasses technology, analysis and mythology. The rhetoric surrounding big data is striking and hints at the tendency to oversimplify the issues involved in its capture, storage and analysis. It has been hailed as an empowering and potentially revolutionary resource but concerns about privacy, surveillance and security are becoming increasingly prevalent (Selinger and Hartzog 2014).

7.8.9 The “Mythology” of Big Data

The mythology that boyd and Crawford mention is defined as “the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity and accuracy” (boyd and Crawford 2012: 663). One common association in commercial Big Data rhetoric is that between the belief that “knowledge is power” and the idea that “bigger is better” (Mayer-Schoenberger and Cukier 2013). It is not uncommon to see “Big Data” paired with words like “empower” (Gandomi and Haider 2015) or “revolutionise” (Mayer-Schoenberger and Cukier 2013), this seems to express a belief that large data sets will “revolutionise” the way we live and necessarily have a positive transformative effect on society. There is a sense, common to many new technologies, that information and knowledge is power. Boyd and Crawford urge people to consider this rhetoric critically, to question how the data is gathered, who has access to it and for what purpose is it used. There are a number of recent technological developments (including the internet and the smart phone) that have opened up access to large data sets in unprecedented ways.

7.8.10 Analysis

The textual data of social media may be available in huge volumes but it is still reportage. A tweet is not “raw”, it is a piece of writing, a tiny, heavily contextualised story. Unlike the data from movement or activity trackers, satellite images or even LiDAR data, the text of a tweet has been curated, whether by an organisation, an individual, or (heaven forbid) a bot – its content has been chosen, edited and shared with some goal in mind. No part of that process was considered in this thesis, but it is fair to assume those activities do impact the final make up of a tweet. Since not all conditions could be accounted for, and the scope of the research had to be limited to something achievable within the scope of a single PhD thesis, the decision was taken to focus on the querying of the text, and not the intentions of the users posting those tweets. In

spite of this, an effort was made to keep the context of the platform in mind at every stage of analysis.

7.8.11 Multimodality

This research demonstrates that research incorporating Twitter data needs to be sensitive to the multimodal nature of the platform. Looking exclusively at the text of a tweet is not sufficient to understand the context, meaning and intention behind it. The majority of the tweets in the corpora used in this study included either linked content (hyperlinks to external webpages) or images, which, in some cases helped elaborate on the meaning of the text, or, in a few cases, to change that understanding completely. CL has been used in this research to build and compile targeted corpora, but the researcher has had to be sensitive to the adaptability of the platform. It is recognised that consideration of text alone does not provide a complete picture of the creative ways in which individuals enact their identities and narrate their environments. Social media platforms like Twitter, that impose certain practical limits on digital communication call for a degree of flexibility in the ways that both users and researchers approach them. These highly dynamic environments need to be treated as Marwick described them – as part of an “ecosystem” (Marwick 2013) of options that is constantly in flux.

7.8.12 In Summary

Theorising Place is not a straightforward activity. This thesis has explored just a few of the ways in which place has been conceptualised and reconceptualised across a number of academic disciplines, according to the agendas of local or national authorities and explored by non-experts. This research has identified three aspects of place relevant to the study of place in the digital context. Twitter text data was filtered, stored and analysed for information relating to each of these aspects: experience, time and juxtaposition. These aspects were each evident in the data and together they present a picture of the expression of place in social media texts that is rich and vivid, where individual

humour and commentary is interwoven with the more sterile broadcasts of business and enterprise.

To some extent, these aspects are inherent to every (whole) tweet – time is built into tweet metadata and tweet content is apparently composed in response to the question “What’s happening?” (Twitter n.d.); a prompt which questions both time and experience. Juxtaposition is inherent to the use of Twitter by multiple individuals. This thesis forgoes the availability of information in metadata to consider how locational reference, time, juxtaposition and experience are written into online narratives by users in their day to day lives within and around cities.

The research was conducted in three methodological stages, each resulting in its own data sets (corpora), information outputs (word frequencies, collocations and keyword analysis as well as themes) and each providing its own insights. The data has all contributed to a picture of how place is co-constructed by users on Twitter, in large part demonstrating the applicability of the various theorisations of place that were discussed during the literature review. Place is invoked in several ways; as a spatial marker for instance, or sometimes as a satirical device; and is often the core to which individuals attach stories about social history as well as individual and shared identities.

Place is multi-layered and shaped by a myriad of forces – human, geological, economic, temporal (Least Heat-Moon 1991), though theorisations of place may vary widely, they can be reconciled through the use of the right technology (GIS) and/or theoretical perspective (Solnit 2010). The intersection of individual and shared narratives and the overlaying of voices is intensified under the conditions of modernity and supermodernity. Although they present significant challenges to research in urban settings, they are also what enable researchers to interrogate these sites of intersection and conflict (Beauregard 2018), to capture ephemeral data (Balmori 2003), and to unveil human connections made through everyday interactions with personal digital

technologies. Social media data serves the dual function of permitting investigation at a fine grain of detail whilst maintaining a distance from the field of study. Technology has methodological as well as theoretical impacts on the investigation of place.

This interdisciplinary thesis presents an in-depth view of both the challenges and opportunities that result from the confluence of place-based research and data streams from personal digital technologies. It has demonstrated that a number of key aspects of place relevant to contemporary cities are present and prevelant in Twitter text data. This information is not only abundant (due to the volume of information that is exchanged on the platform), but it is also rich. Even when separated from its much-lauded metadata, Twitter text is a valuable resource for place research. This thesis shows that, when analysed in a way sympathetic to the specific technological and linguistic medium, social media data can contribute to understandings of the multi-layered, emotional, physical and commercial relationships that exist between cities and the populations that pass through or inhabit them. Twitter corpora provide an amusing, confusing and detailed record of the associations that people forge with the places they grow up in, the places they go to the cinema in, the places they work in and sometimes even just those places they read about on the news. The individual rationalisation of these connections has not been explored, but the value of Twitter as a resource for the wide-spread negotiation of individual and shared meanings in cities was evident.

7.8.12.1 The Unresolved City

According to Beauregard (2018), the unresolved nature of the city is its defining characteristic, which makes the investigation of these contemporary urban spaces particularly well suited to the use of social media data. This data, which is often contradictory and not validated nonetheless offers a multitude of avenues through which to explore human interaction and activity. Beauregard's highlighting of the breakdown of traditional dichotomies, when paired with Bauman's

theorization of “liquid” times makes digital narratives the perfect vantage point from which to explore contemporary urban life. As Tim Gale puts it:

> *“Because it is comparatively easy in these ‘liquid’ times to travel from one place to another, at least for a significant number of people, space ceases to count for much at all and the world is reduced to a series of places to visit that, whether closer to home or further away, are of equal value”. (Gale 2009, 128)*

This description of a post-modern time challenges distinctions normally made between not only the near and the far but also about what is ordinary, what is extraordinary and redefines what is meant by ‘everyday life’. The use of social media data for furthering understandings of place is well-matched to these conditions, where authority and physical distance are becoming less and less relevant. Engaging with narratives in digital contexts might help us to better understand how people can use technology and their personal data records to develop a sense of personal or cultural identity.

The “unresolved” nature of the city highlighted by Beauregard includes the presence of tensions, contradictions and problematic conditions which include social and economic imbalances. The data presented in this thesis shows that ambient data can generate unexpected insights, the Twitter text data described in section 6.4.7.4, for example, highlighting the existence of crime; poverty and homelessness in both Nottingham and Newcastle.

The focus on contemporary urban spaces and the use of a social media platform as a data source facilitates the exploration of what Beauregard calls the “imaginative dimension” of the city (Boellstorff 2008, 9). The imaginary dimension extends the idea of the city away from pure theory or pure reality; melding the two perspectives so they exist in tandem. Under these duplicitous conditions, the city is at once real and unreal, existing in the mind’s eye as well as in the physical world. When the city can never be resolved, social media data is its perfect companion; since

the narratives there are also never resolved (Page 2010). Similar to the city theorised by Beauregard, Twitter, as well as many other sites populated by user-generated content, is full of contradictions and ambiguities, of imbalances and struggles.

7.9 Finding a way through the noise

Some challenges, of course, cannot be easily distilled into one or other of these categories, and many of the challenges addressed here are relevant to all aspects and stages of the research process, from inception to publication. Nonetheless, the approach outlined here has been useful in defining a useable framework that provides a basic overview of what a researcher might reasonably expect to encounter in the investigation of place-based narratives within the context of social media data. Although the scope of this project has been necessarily limited, the theoretical and methodological contributions are wide-ranging and relevant to existing and future work concerning contemporary urban spaces, human-technological interaction, modernity, and place-based studies.

Technology has changed the way people communicate. The traces we leave behind as remnants of our lives are increasingly digital; the explosion of personal ubiquitous technologies has altered the method, the frequency, and the style of how we talk about ourselves and others in our day to day lives. The stories we tell about our lives happen in small and often disconnected segments. If research is mindful of the limitations of social media text data, the linguistic data (and therefore the experiential and emotional information) that can be extracted from it may contribute to our understandings of place. Looking at the words people write about place in the effervescence of social media can provide insights into the ways individuals and groups engage with the locations they pass through and the spaces they tether their stories to, even if these insights are bound within space and time. For the sake of methodological sanity, this research made a crude distinction between digital and non-digital contexts; however, the project itself has revealed an intricate relationship between the two environments, one that

suggests that this distinction is not always sustainable or useful. Boellstorff (2008) and Miller and Slater (2000) warn against the distinction of the digital from the non-digital, insisting that one should not be privileged above the other. They point out that although digital technologies may force individuals to make certain affordances in the way they use language (for example), these technologies are "continuous with and embedded in other social spaces" that "happen within mundane social structures and relations that they may transform but they cannot escape" (Miller and Slater 2000, 5).

Building on the myriad of definitions that have arisen from a number of disparate, complimentary and sometimes contradictory disciplines, the focus of the definition supplied here is on humanity as a common attribute. By keeping human experience at the centre of our thinking about place, we may be better able to negotiate and make adjustments for forces of change or flux, that impact the key characteristics we associate with that place. Therefore, although a place is often associated with particular characteristics or landmarks, it is also in constant flux and continually transforming in response to social, geological, technological and commercial changes over time. This thesis proposes a definition of place informed by and suited to exploration in the context of social media.

Chapter 8 Conclusion

This research took inspiration from the interdisciplinary resources explored in the early stages of the project and has been strongly influenced by the idea of layering (Least Heat-Moon 1991, Balmori 2003, Cresswell, Place 2014, Moniker n.d.). The project was conducted in three complimentary research stages. Each stage was designed to explore different methods of filtering and analysis; together they provide a comprehensive picture of the methodological choices available to researchers working with Twitter.

The three stages of this research project make up a body of work that contributes to the methodological and theoretical work of two disciplines as well as the interdisciplinary study of ubiquitous technologies and personal data. The combination of qualitative and quantitative methods described in this thesis helps to highlight some of the tensions inherent to conducting interdisciplinary research. It stresses, in particular, the difficulty of balancing large-scale data with human-centred research questions, and suggests a framework which can be applied in similar contexts, where concerns about “place” and social media platforms intersect.

Tweet text data might not be able to reveal to us a topographically accurate depiction of a location but it does enable us to see how people refer to locations in their everyday exchanges, it enables us to see what names they use to refer to a location and it allows us to see what values or items they tend to associate with those locations. These texts could provide researchers with insights into the rich relationships that people develop with particular spaces, and might help to inform initiatives aimed at the enrichment of analogue urban space through the effective utilisation of digital resources (by providers as well as consumers). Section 8.1 discusses the insights of the research in response to the central research question.

8.1 What Can Social Media Texts Contribute to Understandings of Place in Cities? The Cases of Nottingham and Newcastle

The two cities at the centre of this research shared a number of geographic and demographic similarities – the populations were of a similar size; each urban centre has a river flowing through it; each has large areas of green space; each has a University (see section 3.4.1 and 3.4.7 for a rationale), and yet the research suggests that each once possesses qualities that set it apart from the other. The geographic and demographic similarities were known at the outset, and were influencing factors in the choice of each city. But what more has this research revealed about each place? Did the ambient data taken from Twitter deliver valuable insights about the "places" of Nottingham and Newcastle?

The most striking thematic similarity evident in the tweets related to each city was the prevalence of sporting talk – particularly football in samples from both cities. CL analysis showed that football team names were closely correlated with city names and were amongst the most frequently occurring words in every corpus. Thematic analysis demonstrated that event reportage was common and revealed the breadth of purpose in tweet authorship, which varied from advertising supplements to commenting on global politics.

Nottingham was strongly associated with football, particularly Nottingham Forest Football club (NFFC) and Notts County as well as UNESCO and universities. The locale (UK) was not commonly mistaken, although the samples suggested that the place name was more likely to be abbreviated (to Notts, for example) than Newcastle, which may have some implications for future methodological development. There were more cases of mistaken identity with Newcastle, as the name can be broken into two words, which expands its field of reference significantly. The city is strongly linked to the "toon" army and to Newcastle United football club, as well as highlighting the inseperability of the "Geordie" culture from these features.

Thematic pairing also helped to elucidate real and imaginary characteristics unique to each place. Events in Nottingham, for instance, paired with Warhammer, whilst those in Newcastle paired with business lunch, each suggesting players in the economy of the area. Other theme pairings also showed a real constrast in the content connected to each place: Shopping in Nottingham was paired with “vintage” and “bridal”, whereas Newcastle shopping co-occurred with “wholesale” and “football mugs”. Similarly, each place showed a difference in the mention of particular social issues; in Nottingham this was “rough sleeping” and in Newcastle, “free Tibet”.

Much of the information revealed about these places might have been hypothesised by any individual with a passing acquaintance with either Nottingham or Newcastle – the data repeats common tropes about each place that are well-known to outsiders (the “toon army” of Newcastle, which was apparent in both Newcastle corpora describes the city’s dedicated football fans), but it also demonstrates that each place can be clearly differentiated from the other using Tweet text data. The data revealed the complex nature of place-based narratives online, and though it does not suggest that every individual Tweet can reveal something about place, it does prove that tweet text is valuable (to both researchers and users) to the interpretation of place in contemporary contexts. The broad patterns revealed in the data indicate that Twitter data is helpful in clarifying differences between cities and highlighting key traits which are capitalised upon in regional identity-making, as well as suggesting commonalities in the way that place is expressed more generally (by both cities).

8.1.1 Broadcasting rather than Communicating

There were very few examples in the data sets at any of the three stages where tweet text indicated that Twitter was being used to have a conversation, or to directly participate in dialogue with other platform users. This may be due in part to the research design, which did not stipulate the inclusion of quoted tweets or replies in the results, meaning that responses and communication conducted in either of

these ways were not collected as part of the sample. The interest in hashtags in the second stage was motivated in part by a body of work that suggests they are used to participate in shared discourse. The analysis performed on this data set and the filtering of the third stage was an attempt to discover whether their use in the context of place reflected this. It demonstrated the opposite. A majority of hashtags appeared only once in the sample, which suggested they are more likely being used to provide context or commentary on their own situation, rather than to contribute an opinion to a conversation or campaign. This does not align with previous research into Hashtags cited in the literature review, which shows that hashtags are frequently used for exactly this purpose – to partake in a shared dialogue about issues affecting millions of people; Clark's presentation of the #whyistayed hashtag and Giglietto and Lee's assessment of #JeNeSuisPasCharlie both clearly demonstrate that (hash)tagging can be employed as part of a discursive strategy (Clark 2016, Giglietto and Lee 2017).

8.2 Disciplinary Reflections

8.2.1 Human Geography

This thesis examines the nature of spatial data in social media information exchange (what does it look like, how can we find it), the value of this spatial data (what does it tell us) and presents a methodological approach to tackle the obscurity of online exchange. The literature, data analysis and the subsequent discussion support the idea, so eloquently stated by Rebecca Solnit;

> *"Places are leaky containers. They always refer beyond themselves, whether island or mainland, and can be imagined in various scales, from the drama of a back alley to transcontinental geopolitical forces and global climate. What we call places are stable locations with unstable converging forces that cannot be delineated either by fences on the ground or by boundaries in the imagination – or by the perimeter of the map. Something is always coming from*

> *elsewhere, whether it's wind, water, immigrants, trade goods or ideas." (Solnit 2010, vii)*

Digitally-mediated places may seem open to more leaks and contaminations than analogue spaces, but this distinction is not helpful. Urban spaces are sites of technological, social, geologic and economic confluence, and exist at once in a digital and an analogue sense, not complementing or separating between the two realms, but flowing seamlessly between them.

8.2.2 Corpus Linguistics

This thesis has applied methods from corpus linguistics in the pursuit of answers to questions that intersect with the interests of human geography, GIS and digital sociology. It attests to the utility of corpus linguistics in researching complex questions concerning personal and group identity, as well as in the handling and analysis of social media text data at a variety of scales.

The language used on Twitter, as evidenced in this thesis, is highly variant, often used creatively and with humour and can be difficult to interpret. Even when the primary language used is the same (take the English that made up most of the corpora here as an example), it can seem obscure and random to those outside the context of exchange. Interpreting digital texts requires researchers to make use of digital resources that are not part of a standard analytical toolkit. There are a plethora of resources available to the public online, whether they are designed to help people navigate from A to B (Waze n.d.) or to order food for delivery to their home or office address (Just Eat n.d.) (UberEats n.d.) (Deliveroo n.d.) or to contribute experiential information about a retail premises or gastronomic venue (Google reviews), their content may be moderated for harmful subject matter, but often goes un-edited, unverified and free to use by whoever has the time, inclination and (usually minimal) skills required to contribute. In spite of their empirically dubious nature, these un-verified, open-source media platforms for the translation of tweet text can be used to great effect.

This thesis has made considerable use of the online dictionary of slang Urban Dictionary (urbandictionary.com). Several of the examples provided in the body of this thesis were interpreted with the help of this platform. The platform was also helpful when thinking about the description and function of the hashtag, which though it was originally conceived of as a device for the easy indexing of conversations on feed-based communications platforms, has taken on a cultural and discursive significance.

The dual approach facilitated by CL methods and tools was central to this work and the insights it has generated. Switching between a general statistical view of the data and a zoomed in view of the details of online narrative gave a complete view of the data, highlighting its intricacies and idiosyncrasies, as well as its general trends. CL methods have been used to great effect in the analysis of sentiment and the eludication of dominant discourses over time (L'Hote 2014, Sealey 2012). The rich insights gleaned in this project from a relatively time-limited selection of text suggests that CL approaches would be productive in the tracking of place-based discourses over time and could be used to ascertain whether these discourses are positive or negative, or indeed, to more keenly understand the nuances of emotion and sentiment in online narratiives about place.

8.3 The Interdisciplinary Study of Digital Technologies

Even the flexibility of Corpus Linguistic methods and tools was not sufficient to provide a complete understanding of place in the context of Twitter. The platform is too multi-modal for a simply text-focussed approach. Analysis made is clear that this approach (of isolating text from context) did not permit the complete exploration of Twitter as a rich repository of social, emotive and economic data. A different approach, perhaps one that encouraged the opening of images and the following of hyperlinks or the tracking of exchanges within the medium might have better facilitated a broader investigation into place on Twitter. The internet is a network and the isolation of a single mode robs it of its inherent interconnectedness, CL methods permitted the assessment of

Twitter data at both close- and far-range and hinted at the appeal of more multi-modal methods for exploring the richness of relationships with place online.

The parameters of research necessarily limit the scope of any single foray into the data, but this research does not stand on its own – multimodal approaches are necessary in order to explore the full extent of human interaction through technologies. I have, in this work, explored the textual-linguistic aspects of this engagement, with a particular focus on geography, but it would certainly be interesting to see this explored further, perhaps with a more visual focus or even an auditory one.

8.4 Implications for Practice

8.4.1 Visualisation

The data explored over the course of this thesis highlights the importance of the availability of an abundance of visualisation tools and techniques. Some of those tools and techniques were outlined at the start of the literature review section (sections 2.2.3, 2.3.1 and 2.5.5), those that strike a chord with this research emphasised the layering of voices and in the incorporation of a multitude of experiences and emotions. Layering does not require compromise or the merging of ideas, it allows them to exist, if not in harmony with, at least in tandem with each other. Multi-faceted approaches to the articulation of place call for the use of contemporary data sources (like those explored here) and new digital technologies, that permit a visualisation to adapt to demographic and social changes. Twitter data has already proven useful in the tracking and modelling of events and the spread of associated information (Segerberg and Bennett 2011) (Signorini, Segre and Polgreen 2011) (Sakaki, Okazaki and Matsuo 2010), the layering of information over time for the purpose of understanding place has tended to focus on understanding movement, rhythm and the dynamics of travel (Franca, et al. 2014) rather than emotional or personal connections. The depth of information corresponding to certain times and locations could provide a rich picture of the connections between locations and personal stories and these visualisations could be used to

explore the “imaginary” aspect of urban spaces (Beauregard 2018). The “pin-in-the-map” style of visualising place is all too familiar to a population of digital consumers who navigate according to a blue and red dot on a smartphone screen, and who regularly mark their presence in places by tagging them in social media posts (Brannon 2013, Cheong, Ray and Green 2012, Anti-Eviction Mapping Project n.d.). The style is certainly useful is location is a primary concern, but the nuances of place, and of people’s associations with certain places are ill-served by such constructions; as highlighted by the academic, amateur, artistic and professional endeavours in visualisation explored in chapter 2. Going beyond sometimes crude spatial signifiers like post-codes, coordinates and street names could be facilitated by social media data, authored by a multitude of voices and speaking to the everyday concerns of individuals. The visualisation of “place” in cities calls for the interweaving of “big” approaches (Bruns and Burgess 2011, Cheong, Ray and Green 2012) with the careful curation and reflexivity inherent in artistic approaches to the same (Balesco Rogers and New n.d., Least Heat-Moon 1991, Waldron 2013). This hybrid approach would require new forms of spatio-temporal analysis, similar to those described in this thesis.

8.4.2 Commercial Interests

The analysis of locational reference at all stages revealed that space in frequently a key feature in the micro-narratives of individual tweets. The use of locational reference by business and enterprise was evident but this aspect of tweet content has gone largely unexplored. Where businesses are already making use of hashtagged place names to show the geographic spread of their particular service offering (from escorts to window frames), their connection to location could be emphasised in order to develop their brand identity as well as expand their customer base. The use of hashtags and the careful engagement with other discursive currencies of social media could be capitalised upon by businesses seeking to engage with and inspire loyalty in their customers. In another way, Twitter text data can help businesses

gauge the happiness of their customers, and they may use the platform as a way to track response to products or campaigns as soon as they are launched.

8.4.3 Urban Planning and Heritage

The thematic analysis undertaken in stages one and two as well as the investigation of individual tweets in chapter six has indicated that Twitter could prove a useful platform in the highlighting and tracking of social and economic problems in urban areas. By monitoring reports of activities like begging, busking, drug-taking or the promotion of campaigns aiming to eradicate the need for such activities – social enterprises and local authorities might be able to build a more complete picture of the use of space in urban centres. This iterative information could be used to guide the allocation of resources and the development of initiatives to tackle these issues.

As heritage and the linking of historic images with discussion of the city was a feature throughout this thesis, there is a suggestion that Twitter data might prove useful in the collation of citizen-sourced stories and photographs. Similar approaches have been used in the assessment of the utility of Twitter data (concerning use) in place branding initiatives and for the purposes of destination marketing (Sevin 2013). In a similar vein, the evidence presented here suggests that organisations with an interest in the preservation of both tangible and intangible cultural heritage in cities could use Twitter data to collect information about what the public is interested in (common landmarks or public sites) and also to collect information about why these sites might be important.

8.5 Directions for Future Work

A challenge remains in the extraction of nuanced meaning about place from social media texts, this challenge crosses disciplinary boundaries. Linguistic analysis and insight has already been used extensively in the development of “talkative” and interactive everyday digital technologies (Smart 2017) like Amazon’s Alexa or Apple’s Siri that mimic natural language and can be taught to respond with humour or irony. Apple’s

Siri, for example, when asked to “open the pod bay doors” at one time responded with a line from the film “2001: A Space Odyssey” whose plot centres on technology taking on a personality of its own (Bean 2018) which included acting to preserve its own existence above that of the human characters. The integration of humour into the languages of digital interfaces and the development of “personalities”, including the use of humour, for digital assistants can help to improve the integration of these technologies into every life better mimic the expected patterns of human to human conversation (Koulouri, Lauria and Macredie 2016) (Large, et al. 2017). The importance of everyday, informal communication in the development of these platforms highlights the need to pay attention to, to document, question and analyse linguistic expressions that seem irrelevant in isolation. It is these short utterances, often, which can aid in the understanding of natural conversation.

Social media data streams are essential to the advancement of the study of human language; these streams, which consist of millions of (outwardly) inconsequential and sometimes obscure statements, are a valuable resource not only for the study of human language and sociability but also for communities partaking in a shared practice, by which they can recount and reflect on their physical surroundings, their shared histories, their shared sentiments; and through this practice develop complex emotional and imaginary relationships with the locations they pass through on a daily basis.

The case for making use of cross-disciplinary resources in order to understand how people forge meaningful relationships with their surroundings is a strong one. In increasingly interconnected spaces, where humans interact in close proximity to and sometimes directly in contact with digital technology (FitBit and other wearable devices, for example), it is easy to overlook place as an incidental detail. Yet this thesis has shown that locational reference is consciously incorporated into digital narratives in the day-to-day and that place still holds value in the articulation of both personal and community identity. The challenge going forward is twofold. On the one hand the mythologies of Big Data

must be resisted, as they encourage fast responses and the theorisation of social research as a closed loop, where problems can be readily solved (given the right data set) and “salvation” offered (Turk 2017). On the other hand the abundance of data from a range of contributors in digital environments (such as Twitter and Urban Dictionary) could be genuinely useful in the development of automatic methods able to extract nuanced expressions of place online. To some extent, this challenge is indicative of all technological development. Those looking to affect change and innovate successfully must simultaneously look forward and reflect backwards. This thesis has endeavoured to tread this line carefully, enthusiastically engaging with new technologies and user generated content whilst suggesting methodological approaches and ethical frameworks in order to maintain a sustainable balance for pursuing interdisciplinary work with similar foci.

8.5.1 Visualise Place

As evidenced in the literature review, there are a multitude of options available to both artists and lay people alike in the representation of place. The complexity of these representations has only been briefly touched upon in this thesis, but certainly warrants further investigation. Particularly given the explosion in the availability of applications that encourage users to share their experiences of places (google reviews, gogobot, yelp etc.), there is more opportunity than ever to capture voluntarily contributed data that is explicitly about named locations. An individual’s experience of a location may even be shaped because of these applications; Foursquare for example uses location to make recommendations for activity based on its prior knowledge of the user (previous vists, tastes etc). By capitalising upon the interest in experiential data about location that spans industry (TripAdvisor, ordnance survey), the academe (Cresswell 2014) (Tuan 1977), and artistic practice (Griffiths n.d.) (Waldron 2013) as well as the lay user (TripAdvisor, Google reviews and Waze), research in this area has a great opportunity to pursue projects that are fundamentally

intersectional and could push the development of innovative methods and the expansion of this work into new application contexts.

8.5.2 Go Multi-Platform

This work could be expanded through the consideration of multi-platform use. Previous research has shown that a majority of social media users regularly engage with more than one social media platform – and that Twitter is simply a node within an individual's network of "communicative options" (Marwick 2013). The high incidence of external link inclusion in the tweet text collected over the course of this research suggests that online narratives about place extend across several platforms, which indicates that research looking at place based narratives across multiple platforms could be a fruitful area of further study. In line with the conclusions of Lee (Lee 2015), this thesis supports the view that linguistic habits, once a defining characteristic of a single platform (the hashtag and Twitter being the primary example of this), now extend beyond those limits, bleed into other platforms and even occur frequently in analogue as well as digital realms. This confluence was revealed by the content of posts on Urban Dictionary that were reviewed in chapter five. The restriction of this particular thesis to the single platform Twitter (as a data source) means that the view of this inter-realm relationship is fairly limited, and the precise nature of it cannot be gleaned from looking solely at the text data collected from Twitter.

8.5.3 Make use of Crowd-Sourced Online Resources

Though it did form a small part of the research work undertaken during this project, a closer consideration of the use of crowd-sourced online resources in aiding understanding of language evolution and identity would be a helpful extension of this project. The crowd-sourced urban dictionary became invaluable as the project progressed. Without the information available through this platform, many of the tweets would have gone un-deciphered, with their meaning remaining obscure to the researcher. The continual engagement with this platform meant that the researcher became aware of some other avenues of investigation for

instance the social and political commentary evident in the user names of urban Dictionary contributors.

8.5.4 Engage with Users in Offline contexts

This thesis documents how individuals talk about place online, but the project did not investigate why individuals talk about place online. It would be interesting to extend this research to incorporate the voices and rationale of Twitter users, in order to provide a comprehensive view of what drives people to share spatial information online, or indeed, what drives them not to share this information. This user-centred approach could be achieved through the development of questionnaires or small to medium sized focus groups in tandem with the tracking of individual social media narratives. For instance through the compilation of an individual's tweets from the last six months, assessed for spatial data and then used as a prompt during a face-to-face semi-structured interview.

8.5.5 Explore Twitter as a form of Life Writing

The advent of ubiquitous technologies and the user generated content creation of web 2.0 have disrupted conventional framings of life writing and provide unparalleled access to the self-reportage and social commentary of the everyman. Social media is a means by which millions of people broadcast and reflect upon the details of their lives and loves of others. Drawing on work in linguistics and digital humanities (R. Jones 2013, Reading 2003, Agha 2007, Burgess, Klaebe and McWilliam 2010), this research could be extended to explore social media as a(n) auto/biographical form, a written account of the lives lived in the age of social media, where understanding must span not only the gaps of temporal and physical space, but also the (sometimes great) chasm between demographic groups, social networks and sub-cultures.

8.6 On Reflection

As more and more people move into urban centres, cities are becoming increasingly relevant to discussions of human sociality. The conditions

of the contemporary city, of flow, flux and conflict, where digital technologies are increasingly interwoven with the experience of the everyday, present ever-changing challenges to the useful analysis of the data produced under these conditions.

Developments in technology, and in particular the advent of Web 2.0, have had a significant impact on the authorship of located narratives, triggering an explosion in user-generated content with a spatial element. The introduction of personal digital technologies has further emphasised this trend and located data is now available in unprecedented variety and volume. In a deliberate attempt to resist the lure of “Big Data” and its associated tropes, this thesis has focussed on the written invocation of locational reference in tweet text.

A review of existing literature showed that technological innovation has long held sway over the construction of individual, group and location-based identities (R. Jones 2013, Landsberg 2004, Lupton 2012). This Thesis has focussed on the conceptualisation of “place” and the relationships fostered with it via social media. The complexity of “place” has been broken down into three key features, which were drawn from literature across the disciplines of geography, social anthropology and the digital humanities. These three features are experience (Iyer 2000, Relph 1976, Tuan 1977), time (Heidegger 1978, H. Lefebvre 2014, Massey 1994, McQuire 1998), and juxtaposition (Beauregard 2018, Least Heat-Moon 1991).

These three features have been helpful in refining the scope of both “the city” and of “place” in the context of social media. They provided not only a theoretical platform from which to conduct analysis but also helped in articulating the methodological and ethical challenges of conducting mixed-methods research in a highly dynamic environment.

References

Abrams, J, and P Hall. “Where/Abouts.” In *Else/Where: Mapping. New Cartographies of Networks and Territories*, 12-17. Minneapolis: University of Minnesota Design Institute, 2006.

Agha, A. “Recombinant selves in mass mediated spacetime.” *Language and Communication*, 2007: 320-335.

Anti-Eviction Mapping Project. *The Anti-Eviction Mapping Project.* n.d. https://www.antievictionmap.com/about/ (accessed 06 17, 2015).

ArcGIS. n.d. https://www.arcgis.com/index.html (accessed 02 11, 2016).

Auge, M. *A sense for the other.* Stanford, CA: Stanford University Press, 1998.

—. *Non-Places: Introduction to an Anthropology of Supermodernity.* London: Verso, 1995.

Auge, M. “Paris and the Ethnography of the Contemporary World.” In *Parisian Fields*, by M. Sheringham (ed.), 175-181. London: Reaktion, 1996.

Baker, P. *Using Corpora in Discourse Analysis.* London: Bloomsbury Academic, 2013.

Baker, P, C Gabrielatos, M Khosravinik, M Krzyzanowski, T McEnery, and R Wodak. “A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press.” *Discourse and Society* 19, no. 3 (2008): 273-306.

Balesco Rogers, Dan, and Sophia New. *Every Journey We Make.* n.d. http://belasconew.com/ (accessed March 2015).

Balmori, D. “Just in Time.” In *Mapping in the Age of Digital Media*, edited by M Silver and D Balmori, 14-17. Chichester: Wiley Academy, 2003.

Barthes, R. *Elements of Semiology.* London: Jonathan Cape, 1967.

Baudrillard, J. *The Illusion of the End.* Translated by C. Turner. Cambridge: Polity, 1994.

Bean, J. “Relating to the Internet of Things.” *Interactions* 25, no. 1 (2018): 24-25.

Beauregard, R. *Cities in the Urban Age: A dissent.* Chicago: University of Chicago Press, 2018.

Bednarek, M. *Evaluation in Media Discourse: Analysis of a newspaper corpus.* London: Continuum, 2006.

Bell, A. “Memory in the Digital Age.” *The Guardian*, 12 January 2012.

Bentham, J. “The Panopticon.” In *Offenders or Citizens? Readings in Rehabilitation*, edited by P Priestley and M Vanstone, 13-16. London: Willan, 2010.

Bickerdike, J O. *The secular religion of fandom: pop cuture pilgrim.* London: Sage Swifts, 2016.

Biesinger, R. “The Civic Atlas.” *Uppercase*, July/August/September 2017: 68-73.

Bishop, B. “The City - It's Alive.” *The Wilson Quarterly* 35, no. 3 (2011): 5-7.

Black, J. *Maps and History: Constructing images of the past.* New Haven and London: Yale University Press, 2000.

Boellstorff, T. *Coming of Age in Second Life: An anthropologist explores the virtually human.* Princeton: Princeton University Press, 2008.

Boyd, D, and K Crawford. “Critical Questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon.” *Information, Communication and Society*, 2012: 662-679.

Brailovskaia, J, and J Margraf. “Comparing Facebook users and Facebook non-users: Relationship between personality traits and mental health variables - An exploratory study.” *Plos One*, 2016: 1-17.

Brannon, M. “Standardized Spaces: satellite imagery in the age of Big Data.” *Configurations* 21, no. 3 (2013): 271-299.

Brock, A. “From the blackhand side: Twitter as a cultural conversation.” *Journal of Broadcasting and Electronic Media* 56, no. 4 (2012): 529-549.

Brown, M F. *Who Owns Native Culture?* London: Harvard University Press, 2003.

Bruns, A, and J E Burgess. “The use of Twitter hashtags in the formation of ad hoc publics.” *Proceedings of the 6th European Consortium for Political Research.* Reykjavik: University of Iceland, 2011. 1-9.

Burgess, J, H Klaebe, and K McWilliam. “Mediatisation and Institutions of Public Memory: Digital storytelling and the apology.” *Australian Historical Studies* 41, no. 2 (2010): 149-165.

Carmona, M, S Tiesdell, T Heath, and T Oc. *Public Places - Urban Spaces: The dimensions of urban design.* Amsterdam: Elsevier, 2010.

Castells, M. *The Informational City: Information technology, economic restructuring, and the urban regional process.* Cambridge, Massachusetts: Blackwell, 1989.

—. *The Rise of the Network Society, The Information Age: Economy, society and culture.* Vol. 1. Cambridge, Massachusetts: Blackwell, 1996.

Ceremonial County Boundaries of England. Ordnance Survey. Ministry of Housing, Communities and Local Government, 1997.

Cheong, M, S Ray, and D Green. “Interpresting the 2011 London Riots from Twitter Metadata.” *12th International Conference on Intelligent Systems Design and Applications.* Kochi: IEEE, 2012. 915-920.

Chou, W, Y Hunt, A Folkers, and E Augustson. “Cancer Survivorship in the Age of YouTube and Social Media: A narrative analysis.” *Journal of Medical Internet Research* 13, no. 1 (2011).

Clark, R. “"Hope in a hashtag": the discursive activism of #WhyIStayed.” *Feminist Media Studies* 16, no. 5 (2016): 788-804.

Conrad, S. “Will Corpus Linguistics Revolutionise Grammar teaching in the 21st Century?” *TESOL Quarterly* 34, no. 3 (2000): 548-560.

Cooper, B. *Mapping Manhattan: a love (and sometimes hate) story in maps by 75 New Yorkers.* New York: Abrams Image, 2013.

Cosgrove, D. “Carto-City.” In *Else'Where: Mapping. New cartographies of networks and territories*, 148-157. Minneapolis: University of Minnesota Design Institute, 2006.

Crampton, Jeremy W. “Cartography: performative, participatory, political.” *Progress in Human Geography*, 2009: 1-9.

Cresswell, T. “Place.” In *Introducing Human Geographies*, edited by P Cloke, P Crang and M Goodwin, 249-261. Abingdon: Routledge, 2014.

Cresswell, T. “Place: encountering geography as philosophy.” *Geography* 93, no. 3 (2008): 132-139.

Dafouz-Milne, E. “The pragmatic role of textual and interpersonal metadiscourse markers in the contruction and attaainment of persuastion: A cross-linguistic study of newspaper discourse.” *Journal of Pragmatics* 40, no. 1 (2008): 95-113.

Daston, L, and P Galison. *Objectivity.* Brooklyn: Zone Books, 2007.

Davis Jr, C A, and F T Fonseca. “Assessing the Certainty of Locations Produced by an Address Geocoding System.” *GeoInformatica* 11, no. 1 (2007): 103-129.

Dawkins, R. *The Selfish Gene.* Oxford: Oxford University Press, 2006.

de Jong, T. “Cognitive load theory, educational research, and instructional design: some food for thought.” *Instructional Science* 38, no. 2 (2010): 105-134.

de Jong, T. “Cognitivve Load Theory, educational research, and instructional design: some food for thought.” *Instructional Science* 38, no. 2 (2010): 105-134.

Dean, E. “NHS issues guidance on safe use of messaging software.” *Nursing Management* 25, no. 5 (2018): 6.

Deliveroo. n.d. https://deliveroo.co.uk/ (accessed September 2018).

Dennen, V P. “Pedagogical Lurking: Student engagement in non-posting discussion behaviour.” *Computers in Human Behaviour* 24 (2008): 1624-1633.

Der Derien, J. “War as Game.” *The Brown Journal of World Affairs* X, no. 1 (2003).

Desaguilier, G. “Descriptive Statistics.” In *Corpus Linguistics and Statistics with R. Quantitative Methods in the Humanities and Social Sciences*, 139-149. New York: Springer, 2017.

Deshors, S C. “Zooming in on Verbs in the Progressive: A collostructional and correspondence analysis approach.” *Journal of English Linguistics* 45, no. 3 (2017): 260-290.

DiNucci, D. “Fragmented Future.” *Print*, 1999: 32, 221-222.

Dionisio, P, C Leal, and L Moutinho. “Fandom affilliation and tribal behaviour: a sports marketing application.” *Qualitative Market Research: An International Journal* 11, no. 1 (1998): 17-39.

Dodge, Martin, and Chris Perkins. “Reclaiming the Map: British geography and ambivalent cartographic practice.” *Commentary for Environment and Planning A*, 2008.

Dodge, Martin, and Rob Kitchin. “Crowdsourrces Cartography: mapping experience and knowledge.” *Environment and Planning A* 45 (2013): 19-36.

Dorling, D. *The Visualisation of spatial social structure.* New Jersey: John Wiley and Sons, 2012.

Dorling, D, and D Fairbairn. *Mapping: Ways of representing the world.* Harlow: Routledge, 1997.

Durkheim, E. *The Elementary Forms of Religious Life.* Oxford: Oxford University Press, 2001.

Eden, T. *Why Bother with What Three Words?* 28 March 2019. https://shkspr.mobi/blog/2019/03/why-bother-with-what-three-words/ (accessed April 2019).

Edney, M. “Cartography and its Discontents.” *Cartographica: The International Journal for Geographic Information and Geovisualisation* 50, no. 1 (2015): 9.

Elwood, Sarah, Michael F Goodchild, and Daniel Z Sui. “Researching Volunteered Geographic Information: Spatial data, geographic research, and new social practice.” *Annals of the Association of American Geographers* 102, no. 3 (2012): 571-590.

Estelles-Arolas, E, and F Gonzalez-Ladron-de-Guevara. “Towards an Integrated Crowdsourcing Definition.” *Journal of Information Science* 38, no. 2 (2012): 189-200.

Federman, M. *The Ephemeral Artefact: Visions of Cultural Experience.* 2005. individual.utoronto.ca/markfederman/EphemeralArtefact.pdf. (accessed February 2014).

Fiss, P, and P Hirsch. “The Discourse of Globalization: Framing and Sensemaking of an Emerging Concept.” *American Sociological Review* 70, no. 1 (2005): 29-52.

forums.fast.ai. *Understanding the Dice Coefficient (part 2).* 2017. https://forums.fast.ai/t/understanding-the-dice-coefficient/5838 (accessed August 2018).

Fox, J, and J J Moreland. “The dark side of social networking sites: An exploration of the relational ad psychological stressors associated with Facebook use and affordances.” *Computers in Human Behaviour* 45 (2015): 168-176.

Franca, U, H Sayama, C McSwiggen, R Daneshvar, and Y Bar-Yam. “Visualizing the "Heartbeat" of a City with Tweets.” New England Complex Systems Institute Reports, 2014.

FrenchSwan. 7 April 2015. https://www.urbandictionary.com/define.php?term=hashtag (accessed 2016).

Fruttaldo, A, and M Cambria. “Combining qualitative and quantitative approaches to discourse analysis: in conversation with Gerlinde Mautner and Alan Partington.” *Imago* 9 (2017): 284-304.

Gale, T. “Urban Beaches, Virtual Worlds and 'The End of Tourism'.” *Mobilities* 4, no. 1 (2009): 119-138.

Gandomi, A, and M Haider. “Beyond the hype: Big Data concepts, methods, and analytics.” *International Journal of Information Management* 35, no. 2 (2015): 137-144.

Gandy, M. “Cyborg Urbanization: Complexity and Monstrosity in the Contemporary City.” *International Journal of Urban and Regional Research* 29, no. 1 (2005): 26-49.

Garfinkel, H. *Studies in Ethnomethodology.* Englewodd-Cliffs, NJ: Prentice-Hall, 1967.

Geertz, C. *The Interpretation of Cultures.* New York: Basic Books, 1973.

Giddens, A. *Moderntiy and Self-Identity: Self and society in the late modern age.* Cambridge: Polity, 1991.

Gigenrenzer, G. *Reckoning with Risk: Learning to Live with Uncertainty.* London: Penguin, 2003.

Giglietto, F, and Y Lee. “A Hashtag Worth a Thousand Words: Discursive strategies around #eNeSuisPasCharlie after the 2015 Charlie Hebdo shooting.” *Social Media + Society*, 2017: 1-15.

Gil de Zuniga, Homero, Nakwon Jung, and S Valenzuela. “Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation.” *Journal of Computer-Mediated Commuication* 17, no. 3 (2012): 319-336.

Glaser, B G. *Doing Formal Grounded Theory: A propsal.* Mill Valley: Sociology Press, 2006.

Glaser, B G, and A L Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Aldine Publishing Company, 1967.

Golden, D. “Storytelling the Future: Isrealis, immigrants and the imagining of community.” *Anthropological Quarterly*, 2001: 7-35.

Gomez-Rodriguez, M, K P Gunmadi, and B Schoelkopf. “Quantifying Information Overload in Social Media and Its Impact on Social Contagion.” *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.* 2014. 170-179.

Goodchild, Michael F. “Citizens as sensors: the world of volunteered geography.” *GeoJournal* 69 (2007): 211-221.

Granell, C, and F O Ostermann. “Beyond Data Collection: Objectives and methods of research using VGI and geo-social media for disaster management.” *Computers, Environment and Urban Systems* 59 (2016): 231-243.

Gries, St. Th. *Quantitative corpus linguistics with R: A practical introduction.* London: Routledge, 2009.

Griffiths, P. *Prints on Paper.* n.d. http://www.peter-griffiths.com/artworks/drawings/ (accessed June 2018).

Gruber, E. “The City as Commune.” In *Meanings of Community across Medieval Eurasia: Comparatice approaches*, edited by E Hovden, C Lutter and W Pohl, 99-124. 2016.

Hacklay, Mordechai Muki, and Patrick Weber. “OpenStreetMap: User-generated street maps.” *Pervasive Computing* (IEEE CS), 2008: 1536-1268.

Haklay, M, V Antoniou, S Basiouka, R Soden, and R Mooney. *Crowdsourced geographic information use in government.* London: GFDRR (World Bank), 2014.

Haklay, Mordechai Muki, Sofia Basiouka, Vryon Antoniou, and Ammer Ather. “How Many Volunteers does it take to Map an Area Well? The validity of Linus' Law to volunteered geographic information.” *The Cartographic Journal* 47, no. 4 (2010): 315-322.

Haklay, Muki, Alex Singleton, and Chris Parker. “Web Mapping 2.0: The Neogeography of the GeoWeb.” *Geography Compass* 2, no. 6 (2008): 2011-2039.

Harley, J B. “Deconstructing the Map.” *Cartographica* 26 (1989): 1-20.

Hartzog, W, and F Stutzman. “Boundary Regulation in Social Media.” *Proceedings of the ACM Conference on Computer Supported Cooperative Work.* 2012.

Hartzog, W, and F Stutzman. “Obscurity by Design.” *Washington Law Review* 88, no. 2 (2013): 385-418.

Haskins, E. “Between Archive and Participation: Public memory in a digital age.” *Rhetoric Society of America* 37, no. 4 (2007): 401-422.

Hasler, L, and I Ruthven. “Escaping Information Poverty through Internet Newsgroups.” Glasgow: AAAI Conference on Weblogs and Social Media, 2011.

Heidegger, M. *Being and Time.* Oxford: Wiley Blackwell, 1978.

Hine, C. *Virtual Methods: issues in social research on the internet.* Oxford: Berg, 2005.

Hood, J. “Orthodoxy vs. Power: The defining traits of grounded theory.” In *The Sage Handbook of Grounded Theory*, edited by A Bryant and K Charmaz, 151-164. London: Sage Publications, 2007.

Horita, F E, L C Degrossi, L F Assis, A Zipf, and J Porto de Albuquerque. “The use of Volunteered Geographic Information and Crowdsurcing in Disaster Management: A systematic literature review.” *Proceedings of the Nineteenth Americas Conference on Information Systems.* Chicago, 2013. 1-10.

Hudelot, C, J Atif, and I Bloch. “Fuzzy spatial relation ontology for image interpretation.” *Fuzzy sets and systems*, no. 159 (2008): 1929-1951.

Ingold, T. “The Temporality of the Landscape.” *World Archaeology* 25, no. 2 (1993): 152-174.

Iyer, P. *The Global Soul: Jet-lag, shopping malls and the search for home.* London: Bloomsbury, 2000.

Jackson, M. *Minima Ethnographica: Intersubjectivity and the Anthropological Project.* Chicago: University of Chicago Press, 1998.

Jackson, P. *Maps of Meaning: An introduction to cultural geography.* London: Routledge, 1989.

Jones, C B, R S Purves, P D Clough, and H Joho. “Modelling places with knowledge from the web.” *International Journal of Geographical Information Science* 22, no. 10 (2008): 1045-1065.

Jones, R. “C me Sk8: Discource, technology and bodies without organs.” In *Digital Discource: Language in the new media*, edited by C Thrulow and K Mroczek, 321-339. Oxford: Oxford University Press, 2011.

—. “Cybernetics, Discourse Analysis and the Entextualisation of the Human.” *Plenary Address from Annual Meeting of the Spanish Association for Applied Linguistics.* 2013.

Just Eat. n.d. https://www.just-eat.co.uk/ (accessed September 2018).

Kennedy, H. *Post, Mine, Repeat: social media data mining becomes ordinary.* Basingstoke: Palgrave Macmillan, 2016.

Kilgaririff, A. “Simple Maths for Keywords.” *Proceedings of Corpus Linguistics Conference CL2009.* Liverpool: University of Liverpool, 2009.

Knight, D, S Adolphs, and R Carter. “CANELC: constructing an e-language corpus.” *Corpora* 9, no. 1 (2014): 29-56.

Koulouri, T, S Lauria, and R Macredie. “Do (and Say) and I Say: Linguistic Adaptation in Human-Computer Dialogs.” *Human-Computer Interaction* 31, no. 1 (2016): 59-95.

Krikorian, Raffi. *Twitter Official Blog: New Tweets per second record, and howl.* 16 August 2013. https://blog.twitter.com/engineering/en_us/a/2013/new-tweets-per-second-record-and-how.html (accessed 2017).

Kucuk, M. “Lurking in online asynchronous discussion.” *Procedia - Social and Behavioural Sciences* 2, no. 2 (2010): 2260-2263.

Lafrance, A, and R Meyer. *A Eulogy for Twitter: The beloved social publishing platform enters its twilight.* 30 April 2014. https://www.theatlantic.com/technology/archive/2014/04/a-eulogy-for-twitter/361339/ (accessed 06 11, 2018).

Landsberg, A. *Prosthetic Memory: Remembering in the Age of New Media.* New York: Columbia University Press, 2004.

Large, D R, L Clark, A Quandt, G Burnett, and L Skrypchuk. “Steering the conversation: A linguistic exploration of natural language interactions with a digital assistant during simulated driving.” *Applied Ergonomics* 63 (2017): 53-61.

Latour, B. *Reassebling the Social: An introduction to Actor-Network Theory.* Oxford: Oxford University Press, 2005.

Laurini, R, and D Thompson. *Fundamentals of Spatial Information Systems.* London: Academic Press, 1992.

Lave, J, and E Wenger. *Situated Learning: Legitimate peripheral participation.* Cambridge: Cammbridge University Press, 1991.

Law, J. "Notes on the Theory of the Actor Network: Ordering, strategy, and heterogeneity." *Systems Practice* 5, no. 4 (1992): 379-393.

Least Heat-Moon, W. *PrairyErth (a deep map).* Boston: Houghton Mifflin, 1991.

Lee, C. "Digital Discourse@public space: Flows of language online and offline." In *Discourse and Digital Practices: Doing discourse analysis in the digital age*, edited by Jones, Chik and Hafner, 175-192. New York: Routledge, 2015.

Lefebrve, H. *The Production of Space.* Oxford: Wiley-Blackwell, 1991.

Lefebvre, H. *Critique of Everyday Life: The one volume edition.* London: Verso, 2014.

—. *Rhythmanalysis: Space, time and everyday life.* London: Bloomsbury Revelations, 2014.

Lefebvre, R K, and C Armstrong. "Grievance-based social movement mobilisation in the #Ferguson Twitter storm." *New Media and Society* 20, no. 1 (2018): 8-28.

Lehmann, S, Jukka-Pekka Onnela, Yong-Yeol Ahn, and Niels Rosenquist. "Understanding the Demographics of Twitter Users." *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.* Barcelona, Spain, 2011.

Leidner, J, and M D Lieberman. "Detecting geographical references in the form of place names and associated natural language." *SIGSPATIAL Special* 3, no. 2 (2011): 5-11.

Lexical Computing. *Sketch Engine.* n.d. https://www.sketchengine.co.uk/ (accessed 01 12, 2018).

L'Hote, E. *Identity, Narrative and Metaphor: A corpus-based cognitive analysis of New Labour Discourse.* Basingstoke: Palgrave Macmillan, 2014.

Lupton, D. “M-health and Health Promotion: The digital cyborg and surveillance society.” *Social Theory & Health* 10, no. 3 (2012): 229-244.

Lupton, D. “The digital cyborg assemblage and the new digital health technologies.” In *The Palgrave handbook of social theory in health, illness and medicine*, 567-581. London: Palgrave Macmillan, 2015.

Lupton, D, and D Haraway. “The digital cyborg assemblage and the new digital health technologies.” In *The Palgrave handbook of social theory in health, illness and medicine*, 567-581. London: Palgrave Macmillan, 2015.

Lyon, D. “Liquid Surveillance: The contribution of Zygmunt Bauman to surveillance studies.” *International Political Sociology* 4, no. 4 (2010): 325-338.

Mackiewicz, J. “Adding quantitative corpus-driven analysis to qualitative discourse analysis: Determining the aboutness of writing centre talk.” *The Writing Centre Journal* 35, no. 3 (2016): 187-225.

Maher, T, and A Haas. “Suburbanizing the City.” *Sociological Focus* 20, no. 4 (1987): 281-294.

Marwick, A E. *Status Update: Celebrity, publicity and branding in the social media age.* London: Yale University Press, 2013.

Massey, D. *Space, Place and Gender.* Minneapolis: University of Minnesota Press, 1994.

Matsuda, Paul Kei. “Identity in Written Discourse.” *Annual Review of Applied Linguistics* 35 (2015): 140-159.

Mayer-Schoenberger, V, and K Cukier. *Big Data: A revolution that will transform how we live, work, and think.* Boston, MA: Houghton Mifflin Harcourt, 2013.

McCandless, D. *Information is Beautiful: Chicks Rule? Gender Balance on Social Networking Sites.* 2012. https://informationisbeautiful.net/visualizations/chicks-rule.

McQuire, S. *Visions of Modernity: Representation, Memory, Time and Space in the Age of the Camera.* London: Sage Publications, 1998.

Merry, Sarah Kate, and Anoush Simon. “Living and lurking on LiveJournal: The benefits of active and non-active membership.” *Aslib Proceedings: New Information Perspectives* 64, no. 3 (2012): 241-261.

Miller, D, and D Slater. *The Internet: An ethnographic approach.* London: Berg, 2000.

Moniker. “All the Minutes, 2014: Internet Project.” In *Poetics and Politics of Data: The ambivalence of life in a data-driven society*, edited by S Himmelsbach and C Mareis, 134-139. Basel: Critical Media Lab, n.d.

Monmonier, M. “Temporal Generalisation for dynamic Maps.” *Cartography and Geographic Information Systems* 23, no. 2 (1996): 96-98.

Moran, S, E Luger, and T Rodden. “Consent for all: Revealing the hidden complexity of terms and conditions.” *Proceedings of the SIGCHI Conference of Human Factors in Computing Systems.* New York, USA: SIGCHI, 2013. 2687-2696.

Morstatter, F, J Pfeffer, H Liu, and K M Carley. “Is the Sample Good Enough? Comparing data from Twitter's streaming API with Twitter's firehose.” *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.* Association for the Advancement of Artificial Intelligence, 2013. 400-408.

Muehlenhaus, I. *Web Cartography: Map design for interactive and mobile devices.* Boca Raton: CRC Press, 2014.

Murthy, D. *Twitter: Social communication in the Twitter Age.* Cambridge: John Wiley and Sons, 2013.

National Museums Liverpool. *Mapping Memory: Mapping memory on the Liverpool waterfront 1950s - 1970s.* 2015. http://www.liverpoolmuseums.org.uk/maritime/research/mappingmemory/about.html (accessed 09 05, 2015).

Newcastle City Council. “Newcastle Upon Tyne City Boundary.” 2012. https://www.newcastle.gov.uk/sites/default/files/wwwfileroot/environment/animal_welfare/city_boundary_june_2012_-_streetview_2.pdf (accessed 02 13, 2016).

NHS Digital. *Information Governance and Technology Resources.* November 2018. https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/information-governance-alliance-iga/information-governance-resources/information-governance-and-technology-resources (accessed 12 2018).

Nold, C. “Emotional Cartography: Technologies of the Self.” n.d. http://emotionalcartography.net/EmotionalCartographyLow.pdf.

Nomis. *Qualifications and Students.* Local Area Report, Office for National Statistics, 2011.

Norouzi, M. “Application of GPS in Orienteering Competitions.” *International Journal of Mobile Network Communications and Telematics* 3, no. 4 (2013): 7-13.

Nottingham City Council. “Nottingham City Council Administrative Boundary.” 2011. https://secure.nottinghamcity.gov.uk/RegisterOfficebooking/NottinghamCityCouncilAdministrativeBoundary.pdf (accessed 08 29, 2017).

Nottingham City Council. *Nottingham City Ward Report.* Ward Report, Policy and Research Team, Nottingham City Council, 2016.

OAuth. *OAuth.* Edited by OAuth. n.d. https://oauth.net/ (accessed 01 12, 2018).

Office for National Statistics. *nomis: labour market statistics.* n.d. https://www.nomisweb.co.uk/reports/localarea (accessed 06 11, 2018).

Ohlgren, T H. *Robin Hodd The Early Poems 1465-1560: Texts, contexts and ideology.* Newark: The University of Delaware Press, 2007.

Olejniczak, J. "Using corpora to air qualitative text analysis." *Journal of Education, Culture and Society* 9, no. 2 (2018): 154-164.

Open Geospatial Consortium. *The OGC Abstract Specification: Spatial referencing by coordinates.* Specification, Open Geospatial Consortium, 2010.

Open Street Map. *Open Street Map.* n.d. https://www.openstreetmap.org/ (accessed 06 02, 2017).

Page, R. "Re-examining Narrativity: Small stories in status updates." *Text & Talk - An Interdisciplinary Journal of Language, Discourse & Communication* 30, no. 4 (2010): 423-444.

Patten, C J.D, A Kircher, J Ostlund, and L Nilsson. "Using mobile telephones: cognitive workload and attention resource allocation." *Accident Analysis and Prevention* 36, no. 3 (2004): 341-350.

Patten, C, A Kircher, J Oestlund, and L Nilsson. "Using Mobile Telephones: cognitive workload and attention resource allocation." *Accident Analysis and Prevention* 36, no. 3 (2004): 341-350.

Pearce, W, K Holmberg, L Hellsten, and B Nerlich. "Climate Change on Twitter: Topics, communities and conversations about the 2013 IPCC working group 1 report." *Plos One* 9, no. 4 (2014): 1-11.

Popper, B. *Twitter now has 255 million users, but activity has dropped year over year.* 29 April 2014. https://www.theverge.com/2014/4/29/5665752/twitter-q1-2014-earnings (accessed 06 11, 2018).

Preece, Jeremy, Blair Nonnecke, and Dorine Andrews. “The top five reasons for lurking: Improving community experiences for everyone.” *Computers in Human Behaviour*, no. 20 (2004): 201-223.

Ramirez, E. *Quanitfied Self: How to Map your Moves Data.* 2014. http://quantifiedself.com/2014/03/map-moves-data/ (accessed March 2015).

Rampton, B. *Drilling Down to the Grain in Superdiversity.* Paper, Understanding Society, Tilbrug: Tilburg University, 2013, 1-15.

ranks.nl. *Ranks NL: Stopword Lists.* 2015. https://www.ranks.nl/stopwords (accessed 2015).

Ratkiewicz, J, et al. “Truthy: mapping the spread of astroturf in microblog streams.” *Proceedings of the 20th International Conference Companion on the World Wide Web.* Hyderabad, 2011. 249-252.

Rayson, Paul. “From key words to key semantic domains.” *International Journal of Corpus Linguistics* 13, no. 4 (2008): 519-549.

Reading, A. “Digital Interactivity in Public Memory Institutions: the uses of new technologies in Holocaust museums.” *Media, Culture & Society* 25, no. 1 (2003): 67-85.

Reed, A. “City of Details: Interpreting the personality of London.” *The Journal of the Royal Anthropological Institute* 8, no. 1 (2002): 127-141.

Relph, E. *Place and Placelessness.* London: Pion, 1976.

Riessman, C K. *Narrative Analysis.* London: Sage, 1993.

Robbo. *mackem.* 01 June 2005. https://www.urbandictionary.com/define.php?term=mackem (accessed 06 17, 2016).

Roberts, L, L Smith, and C Pollock. “Conducting Ethical Research Online: Respect for individuals, identities and the ownership of words.” In *Readings in Virtual Research Ethics: Issues and controversies*, edited by E A Buchanan, 156-173. London: Information Science Publishing, 2004.

Rosenberg, D. “Against Infographics.” *Art Journal* 74, no. 4 (2015): 39-57.

Russell, Frank Michael. “Twitter and News Gatekeeping: Interactiity, reciprocity, and promotion in news organisations' tweets.” *Digital Journalism*, 2017: 1-20.

Rychly, P. “A Lexicographer-Friendly Association Score.” *Proceedings of Recent Advances in Slavonic Natural Language Processing.* Brno: Masaryk University, 2008. 6-9.

Sakaki, T, M Okazaki, and Y Matsuo. “Earthquake shakes Twitter users: real-time event detection by social sensors.” *Proceedings of the 19th International Conference on the World Wide Web.* Raleigh, NC: North Carolina State University, 2010. 851-860.

Schnieder, Andreas, Georg von Krogh, and Peter Jaeger. “"What's coming next?" Epistemic curiosity and lurking behaviour in online communities.” *Computers in Human Behaviour*, no. 29 (2012): 293-303.

Scollon, R, and S W Scollon. *Discourses in Place: Language in the material world.* London: Routledge, 2003.

Sealey, A. “"I just couldn't do it": representations of constraint in an oral history corpus.” *Critical Discourse Studies*, 2012: 195-210.

Segerberg, A, and L W Bennett. “Social Media and the Organization of Collective Action: Using Twitter to explore the ecologies of two

climate change protests." *The Communication Review* 14, no. 3 (2011): 197-215.

Sevin, E. "Places Going Viral: Twitter usage patterns in destination marketing and and place branding." *Journal of Place Management and Development* 6, no. 3 (2013): 227-239.

Signorini, A, A M Segre, and P M Polgreen. "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S during the Influenza A H1N1 Pandemic." *Plos One*, 2011.

Silverstein, M. "Language and the Culture of Gender." In *Semiotic Mediation*, edited by E Mertz and R Parmentier, 219-259. New York: Academic Press, 1985.

Simmel, G. "The Stranger." In *The Sociology of Georg Simmel*, by K H Wolff, 402-406. Glencoe: The Free Press, 1950.

Simpson, L, and A Yu. "Public Access to Conversion of Data between Geographies, with multiple look up tables derived from a postal directory." *Computers, Environment and Urban Systems* 27 (2003): 283-307.

Sinclair, J M. "Uncommonly Common Words." In *Learner's dictionaries: state of the art*, edited by M L Tickoo, 135-152. Singapore: RELC Anthology Series, 1989.

Smart, P R. "COGWEB - Cognition and the Web: Social minds, distributed brains, and talkative technologies." *9th International conference on advanced cognitive technologies and applications.* Athens: https://eprints.soton.ac.uk/405460/1/editorialv2.pdf, 2017.

Society for One-Place Studies. n.d. https://www.one-place-studies.org/ (accessed August 2018).

Solis, B. *The Hashtag Economy.* 6 06 2011. https://www.fastcompany.com/1760484/hashtag-economy.

Solnit, R. *Infinite City: A San Francisco atlas.* Berkeley: University of Califoria Press, 2010.

Stefanidis, A, A Crooks, and J Radzikowski. “Harvesting ambient geospatial information from social media feeds.” *GeoJournal* 78, no. 2 (2013): 319-338.

Steiger, E, J Porto de Albuquerque, and A Zipf. “An Advanced Systematic Literature Review on Spatiotemporal Analyses of Twitter data.” *Trasactions in GIS* 19, no. 6 (2015): 1-24.

Stephens, Z D, et al. “Big Data: Astronomical or Genomical?” *PLoS Biology* 13, no. 7 (2015): 1-11.

Sula, C A. “Quantifying Culture: The value of visualization inside (and outside) libraries, museums, and the academy.” *Electronic Visualisation and the Arts.* London: BCS, The Chartered Institute for IT, 2012. 253-257.

Sula, C A. “Visualizing Social Connections in the Humanities: Beyond bibliometrics.” *Bulletin of the American Society for Information Science and Technology*, 2012: 31-35.

Sumikawa, Y, A Jatowt, and M During. “Analysis of Temporal and Web Site References in History-Related Tweets.” *Proceedings of the 2017 ACM on Web Science Conference.* 2017. 419-420.

Sveningsson, M. “Ethics in Internet Ethnography.” In *Readings in Virtual Research Ethics: Issues and Controversies*, edited by E A Buchanan, 45-61. London: Information Science Publishing, 2004.

Tate. *ArtMaps: putting the Tate Collection on the map.* 2012. https://artmaps.tate.org.uk/artmaps/tate/#zoom=15&lat=51.51&lng=-0.10&maptype=hybrid&page=about (accessed 08 12, 2017).

Thorne, Steven L, Shannon Sauro, and Bryan Smith. “Technologies, Identities, and Expressive Activity.” *Annual Review of Applied Linguistics* 35 (2015): 215-233.

Titton, M. “Fashionable Personae: Self-identity and enactments of fashion narratives in fashion blogs.” *Fashion Theory* 19, no. 2 (2015): 201-220.

TripAdvisor. *Content API.* n.d. https://developer-tripadvisor.com/content-api/ (accessed 07 21, 2018).

Tuan, Y. *Space and Place: The perspective of experience.* Minneapolis: University of Minnesota Press, 1977.

Turk, C. “Cartographica Incognita: "Dijital Jedis", satellite salvation and the mysteries of the missing maps.” *The Cartographic Journal* 54, no. 1 (2017): 14-23.

Turkle, S. “Always-On/Always-On-You: The tethered self.” In *Handbook of Mobile Communication Studies*, edited by James E Katz, 121-137. Cambridge, MA: MIT Press, 2008.

Tusting, K. “Language and Power in Communities of Practice.” In *Beyond Communities of Practice: Language, power and social context*, edited by D Barton and K Tusting, 36-54. Cambridge: Cambridge University Press, 2005.

Twitter. *Application Management.* 2018. https://apps.twitter.com/ (accessed 01 12, 2018).

—. *Home Page.* n.d. https://twitter.com (accessed 08 2018, 24).

—. *Search Tweets.* 2018. https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets (accessed 01 12, 2018).

UberEats. n.d. https://about.ubereats.com/ (accessed September 2018).

University of California. *Thesaurus Linguae Graecae.* n.d. http://stephanus.tlg.uci.edu/tlg.php (accessed September 2018).

University of Nottingham. *Student Registrations Summary Statistics 2015-2016.* Strategy Planning and Performance Report, Nottingham: University of Nottingham, 2016.

University of Zurich. *ZEN Zurich English Newspaper Corpus.* 2015. https://www.es.uzh.ch/en/Subsites/Projects/zencorpus.html (accessed September 2018).

Urry, J. *Consuming Places.* London: Routledge, 1995.

Usher, N, J Holcomb, and J Littman. “Twitter Makes it Worse: Political journalists, gendered echo chambers and the amplification of gender bias.” *The International Journal of Press/Politics* 23, no. 3 (2018): 324-344.

Van Cleef, E. “What is a Map?” *Science* 108 (1948): 521-522.

Vertovec, S. “Super-diversity and its implications.” *Ethnic and Racial Studies*, 2007: 1024-1054.

Visit Nottinghamshire. *Robin Hood.* 06 2018. https://www.visit-nottinghamshire.co.uk/ideas-and-inspiration/robin-hood (accessed 06 13, 2018).

VisitEngland. *England's Home of Sport is...* 22 October 2015. https://beta.visitengland.com/blog/things-do/october-2015/englands-homeofsport (accessed 05 07, 2016).

Waldron, Hannah, interview by A Butler. *Hannah Waldron Interview* Design Boom, (28 June 2013).

Walz, L. “Categorisation and the Negotiation of Similarity and Difference: Identity construction in expatriate blogs.” *ConSOLE XXIV: Proceedings of the 24th Conference of the Student Organization of Linguistics in Europe.* Leiden: Leiden University Centre for Linguistics, 2016. 45-62.

Waze. n.d. https://www.waze.com/ (accessed September 2018).

Webster, L. “"I wanna be a toy": Self-sexualisation in gender-variant Twitter users' biographies.” *Journal of Language and Sexuality* 7, no. 2 (2018): 205-236.

Wenger, E. *Communities of Practice: Learning, Meaning and Identity.* Cambridge: Cambridge University Press, 1998.

What3Words. *What3Words: addressing the world.* 2018. https://what3words.com/ (accessed 09 2018).

Whatmore, S. *Hybrid Geographies: Nature, cultures, spaces.* London: Sage, 2002.

Williams, D. “Making sense of 'Place': Reflections on pluralism and positionality in place research.” *Landscape and Urban Planning* 131 (2014): 74-82.

Williams, G, and A Mahmoud. “Modeling User Concerns in the App Store: a case study on the rise and fall of Yik Yak.” *IEEE 26th International Requirements Engineering Conference.* Banff, 2018. 64-75.

Williams, M L, P Burnap, and L Sloan. “Towards and Ethical Framework for Publishing Twitter Data in Social Research: Taking into accout user's views, online context and algorithmic estimation.” *Sociology* 51, no. 6 (2017): 1149-1168.

Wood, D. “The Anthropology of Cartography.” In *Mapping Cultures*, edited by L Roberts, 280-303. London: Palgrave Macmillan, 2012.

—. *The Power of Maps.* New York: Guildford Press, 1992.

Zappavigna, M. “Enacting identity in microblogging through ambient affiliation.” *Discourse and Communication* 8, no. 2 (2013): 209-228.

Appendix A: Metadata in a Single Tweet

{u'contributors': None,
 u'coordinates': None,
 u'created_at': u'Wed Aug 03 12:44:59 +0000 2016',
 u'entities': {u'hashtags': [],
 u'symbols': [],
 u'urls': [{u'display_url': u'twitter.com/[redacted]/status/76\u[redacted]',
 u'expanded_url':
u'https://twitter.com/[redacted]/status/[redacted]',
 u'indices': [21, 44],
 u'url': u'https://[redacted]}],
 u'user_mentions': []},
 u'favorite_count': 0,
 u'favorited': False,
 u'geo': None,
 u'id': [redacted],
 u'id_str': u'[redacted]',
 u'in_reply_to_screen_name': None,
 u'in_reply_to_status_id': None,
 u'in_reply_to_status_id_str': None,
 u'in_reply_to_user_id': None,
 u'in_reply_to_user_id_str': None,
 u'is_quote_status': True,
 u'lang': u'en',
 u'metadata': {u'iso_language_code': u'en', u'result_type': u'recent'},
 u'place': None,
 u'possibly_sensitive': False,
 u'quoted_status': {u'contributors': None,
 u'coordinates': None,
 u'created_at': u'Tue Aug 02 22:36:37 +0000 2016',
 u'entities': {u'hashtags': [],
 u'media': [{u'display_url': u'pic.twitter.com/[redacted]',
 u'expanded_url':
u'http://twitter.com/[redacted]/status/[redacted]/video/1',
 u'id': [redacted],
 u'id_str': u'[redacted]',
 u'indices': [22, 45],
 u'media_url':
u'http://[redacted]/ext_tw_video_thumb/[redacted]/pu/img/[redacted]-.jpg',
 u'media_url_https':
u'https://[redacted]/ext_tw_video_thumb/[redacted]pu/img/[redacted]-.jpg',
 u'sizes': {u'large': {u'h': 396,
 u'resize': u'fit',
 u'w': 720},
 u'medium': {u'h': 330,
 u'resize': u'fit',
 u'w': 600},
 u'small': {u'h': 187,
 u'resize': u'fit',
 u'w': 340},
 u'thumb': {u'h': 150,
 u'resize': u'crop',

u'w': 150}},
u'type': u'photo',
u'url': u'https://t.co/[redacted]'}],
u'symbols': [],
u'urls': [],
u'user_mentions': []},
u'extended_entities': {u'media': [{u'additional_media_info': {u'monetizable': False},
u'display_url': u'pic.twitter.com/[redacted]',
u'expanded_url': u'http://twitter.com/[redacted]/status/[redacted]/video/1',
u'id': [redacted],
u'id_str': u'[redacted]',
u'indices': [22,
45],
u'media_url': u'http://[redacted]/ext_tw_video_thumb/[redacted]/pu/img/[redacted]-.jpg',
u'media_url_https': u'https://[redacted]/ext_tw_video_thumb/[redacted]/pu/img/[redacted]-.jpg',
u'sizes': {u'large': {u'h': 396,
u'resize': u'fit',
u'w': 720},
u'medium': {u'h': 330,
u'resize': u'fit',
u'w': 600},
u'small': {u'h': 187,
u'resize': u'fit',
u'w': 340},
u'thumb': {u'h': 150,
u'resize': u'crop',
u'w': 150}},
u'type': u'video',
u'url': u'https://t.co/[redacted]',
u'video_info': {u'aspect_ratio': [20,
11],
u'duration_millis': 8912,
u'variants': [{u'bitrate': 320000,
u'content_type': u'video/mp4',
u'url': u'https://[redacted]/ext_tw_video/[redacted]/pu/vid/326x180/[redacted].mp4'},
{u'content_type': u'application/x-mpegURL',
u'url': u'https://[redacted]/ext_tw_video/[redacted]/pu/pl/[redacted].m3u8'},
{u'content_type': u'application/dash+xml',
u'url': u'https://video.twimg.com/ext_tw_video/[redacted]/pu/pl/[redacted].mpd'},

{u'bitrate': 832000,
u'content_type':
u'video/mp4',
u'url':
u'https://video.twimg.com/ext_tw_video/[redacted]/pu/vid/654x360
/[redacted].mp4'}]}}]},
u'favorite_count': 5787,
u'favorited': False,
u'geo': None,
u'id': [redacted],
u'id_str': u'[redacted]',
u'in_reply_to_screen_name': None,
u'in_reply_to_status_id': None,
u'in_reply_to_status_id_str': None,
u'in_reply_to_user_id': None,
u'in_reply_to_user_id_str': None,
u'is_quote_status': False,
u'lang': u'en',
u'metadata': {u'iso_language_code': u'en',
u'result_type': u'recent'},
u'place': None,
u'possibly_sensitive': False,
u'retweet_count': 6856,
u'retweeted': False,
u'source': u'<a href="http://twitter.com/download/android"
rel="nofollow">Twitter for Android',
u'text': u'Wind turbine on fire: https://t.co/[redacted]',
u'truncated': False,
u'user': {u'contributors_enabled': False,
u'created_at': u'Wed Jun 13 18:17:58 +0000 2007',
u'default_profile': False,
u'default_profile_image': False,
u'description': u'Physicist and teacher',
u'entities': {u'description': {u'urls': []},
u'url': {u'urls': [{u'display_url': [redacted],
u'expanded_url':
u'http://[redacted].org',
u'indices': [0,
23],
u'url': [redacted]}]}},
u'favourites_count': 3213,
u'follow_request_sent': False,
u'followers_count': 2538,
u'following': False,
u'friends_count': 419,
u'geo_enabled': False,
u'has_extended_profile': False,
u'id': [redacted],
u'id_str': u'[redacted]',
u'is_translation_enabled': False,
u'is_translator': False,
u'lang': u'en',
u'listed_count': 176,
u'location': u'UK',
u'name': u'[redacted]',

u'notifications': False,
u'profile_background_color': u'FDF6E3',
u'profile_background_image_url':
u'http://[redacted]/profile_background_images/[redacted]/[redacted].jpeg',
u'profile_background_image_url_https':
u'https://pbs.twimg.com/profile_background_images/[redacted]/[redacted].jpeg',
u'profile_background_tile': True,
u'profile_banner_url':
u'https://pbs.twimg.com/profile_banners/[redacted],
u'profile_image_url':
u'http://pbs.twimg.com/profile_images/[redacted].jpg',
u'profile_image_url_https':
u'https://pbs.twimg.com/profile_images/[redacted].jpg',
u'profile_link_color': u'0859CA',
u'profile_sidebar_border_color': u'FFFFFF',
u'profile_sidebar_fill_color': u'FDF6E3',
u'profile_text_color': u'657B83',
u'profile_use_background_image': True,
u'protected': False,
u'screen_name': u'[redacted]',
u'statuses_count': 32881,
u'time_zone': u'Europe/London',
u'url': u'https://t.co/[redacted]',
u'utc_offset': 3600,
u'verified': False}},
u'quoted_status_id': [redacted],
u'quoted_status_id_str': u'[redacted]',
u'retweet_count': 0,
u'retweeted': False,
u'source': u'Twitter Web Client',
u'text': u'Strangely memorising https://t.co/wDeAq3uICV',
u'truncated': False,
u'user': {u'contributors_enabled': False,
u'created_at': u'Sun May 31 10:31:48 +0000 2009',
u'default_profile': True,
u'default_profile_image': False,
u'description': u'Chemistry and Physics teacher, founder of
https://t.co/[redacted] and amateur fell runner.',
u'entities': {u'description': {u'urls': [{u'display_url': [redacted],
u'expanded_url': u'http://[redacted].com',
u'indices': [42, 65],
u'url': u'https://t.co/[redacted]'}]},
u'url': {u'urls': [{u'display_url': u'[redacted].wordpress.com',
u'expanded_url':
u'http://[redacted].wordpress.com/',
u'indices': [0, 23],
u'url': u'https://t.co/[redacted]'}]}},
u'favourites_count': 938,
u'follow_request_sent': False,
u'followers_count': 1019,
u'following': False,

u'friends_count': 2366,
u'geo_enabled': False,
u'has_extended_profile': False,
u'id': [redacted],
u'id_str': u'[redacted]',
u'is_translation_enabled': False,
u'is_translator': False,
u'lang': u'en',
u'listed_count': 34,
u'location': u'Nottingham ',
u'name': u'Dr [redacted],
u'notifications': False,
u'profile_background_color': u'C0DEED',
u'profile_background_image_url':
u'http://abs.twimg.com/images/themes/theme1/bg.png',
u'profile_background_image_url_https':
u'https://abs.twimg.com/images/themes/theme1/bg.png',
u'profile_background_tile': False,
u'profile_image_url':
u'http://pbs.twimg.com/profile_images/[redacted]_normal',
u'profile_image_url_https':
u'https://pbs.twimg.com/profile_images/[redacted]_normal',
u'profile_link_color': u'0084B4',
u'profile_sidebar_border_color': u'C0DEED',
u'profile_sidebar_fill_color': u'DDEEF6',
u'profile_text_color': u'333333',
u'profile_use_background_image': True,
u'protected': False,
u'screen_name': u'LazyRunner78',
u'statuses_count': 884,
u'time_zone': u'London',
u'url': u'https://t.co/[redacted]',
u'utc_offset': 3600,
u'verified': False}}
>>>

Appendix B: List of Stop-words used in analysis

a	Any	did	Has	is
A	are	Did	have	Is
able	Are	do	Have	it
Able	as	Do	he	It
about	As	does	He	its
About	at	Does	her	Its
across	At	either	Her	just
Across	be	Either	hers	Just
after	Be	else	Hers	least
After	because	Else	him	Least
all	Because	ever	Him	let
All	been	Ever	his	Let
almost	Been	every	His	like
Almost	but	Every	how	Like
also	But	for	How	likely
Also	by	For	however	Likely
am	By	from	However	may
Am	can	From	i	May
among	Can	get	I	me
Among	cannot	Get	if	Me
an	Cannot	got	If	might
An	could	Got	in	Might
and	Could	had	In	most
And	dear	Had	into	Most
any	Dear	has	Into	must

Must	other	Some	to	Which
my	Other	than	To	while
My	our	Than	too	While
neither	Our	that	Too	who
Neither	own	That	twas	Who
no	Own	the	Twas	whom
No	rather	The	us	Whom
nor	Rather	their	Us	why
Not	said	Their	wants	Why
not	Said	them	Wants	will
Not	say	Them	was	Will
of	Say	then	Was	with
Of	says	Then	we	With
off	Says	there	We	would
Off	she	There	were	Would
often	She	these	Were	yet
Often	should	These	what	Yet
on	Should	they	What	you
On	since	They	when	You
only	Since	this	When	your
Only	so	This	where	Your
or	So	tis	Where	
Or	some	Tis	which	

Appendix C: Collocation Tables

Table 4a: Words appearing before Notts

Word Rank	**-2**	**Cooccurrence Count**	**Raw Count**	**Log Dice**
1	Hartlepool	7	12	10.052
2	back	6	12	9.830
3	2	5	50	9.333
4	everyone	4	6	9.285
5	about	4	12	9.245
6	in	5	98	9.083
7	still	3	5	8.877
8	Carlisle	3	6	8.870
9	for	4	77	8.865
10	all	3	10	8.843
	-1	**Cooccurrence Count**	**Raw Count**	**Log Dice**
1	In	22	98	11.221
2	at	17	50	11.098
3	v	15	29	11.042
4	to	9	122	9.821
5	from	6	33	9.696
6	MISSING	4	7	9.278
7	Jan	3	11	8.836
8	as	3	12	8.830
9	for	3	77	8.450
10	the	3	90	8.385

Table 4b: Words appearing after Notts

Word Rank	+1	Cooccurrence Count	Raw Count	Log Dice
1	County	64	66	12.923
2	Man	5	8	9.594
3	.	7	130	9.423
4	,	6	115	9.267
5	And	5	63	9.261
6	Police	4	22	9.179
7	-	4	69	8.907
8	Load	3	3	8.891
9	Co	3	3	8.891
10	Recall	3	3	8.891
	+2	**Cooccurrence Count**	**Raw Count**	**Log Dice**
1	With	6	30	9.714
2	Fc	4	4	9.299
3	Stars	4	4	9.299
4	;	4	6	9.285
5	,	6	115	9.267
6	.	6	130	9.201
7	Rt	6	152	9.109
8	Bennett	3	4	8.884
9	In	4	103	8.737
10	V	3	30	8.714

Table 4c: Words appearing before Nottingham

Word Rank	-2	Cooccurrence Count	Raw Count	Log Dice
1	University	6	8	9.898
2	investment	4	10	9.299
3	City	4	12	9.285
4	At	4	48	9.057
5	nottingham	6	198	8.955
6	Wvk	3	3	8.933
7	Bolton	3	3	8.933
8	Wolves	3	3	8.933
9	popular	3	5	8.919
10	Park	3	7	8.905
	-1	**Cooccurrence Count**	**Raw Count**	**Log Dice**
1	In	40	108	12.064
2	Of	14	59	10.801
3	-	15	83	10.772
4	from	9	30	10.337
5	,	10	138	9.929
6	at	7	48	9.864
7	:	12	233	9.833
8	to	7	113	9.526
9	near	4	5	9.334
10	V	4	10	9.299

Table 4d: Words appearing after Nottingham

Word Rank	+1	Cooccurrence Count	Raw Count	Log Dice
1	forest	30	33	12.055
2	,	20	138	10.929
3	-	10	83	10.187
4	panthers	7	7	10.127
5	and	8	53	10.028
6	Post	6	9	9.891
7	investment	5	10	9.621
8	inggris	4	4	9.341
9	?	4	37	9.123
10	today	3	9	8.891
	+2	**Cooccurrence Count**	**Raw Count**	**Log Dice**
1	fund	5	5	9.656
2	,	7	138	9.415
3	3	4	37	9.123
4	:	7	233	9.055
5	2	4	57	9.005
6	nottingham	6	198	8.955
7	derby	3	4	8.926
8	could	3	6	8.912
9	nottinghamshire	3	6	8.912
10	A	4	76	8.901

Table 5a: Words appearing before Nottingham

-2	Cooccurrence Count	Raw Count	Log Dice
Mikesassi	53	151	8.830
Rough	39	109	8.403
University	34	319	8.127
Coming	39	1,811	7.866
Brentford	21	250	7.457
Summers	19	43	7.391
Back	44	5,370	7.318
Midlands	18	231	7.242
Forestlivenews	17	61	7.224
#partycaribbean2017 tourcomingsoon	15	31	7.055
-1	**Cooccurrence Count**	**Raw Count**	**Log Dice**
In	84	1,423	9.079
Trade	706	44,546	8.906
At	149	19,302	7.731
V	27	1,360	7.459
Maidens	18	40	7.314
Outcalls	17	49	7.228
Central	16	137	7.107
From	78	17,534	6.913
Sun	15	450	6.900
Because	21	2,099	6.899

Table 5b: Words appearing after Nottingham

+1	Cooccurrence Count	Raw Count	Log Dice
Forest	349	1,332	11.160
Room	133	755	9.945
Post	97	1,313	9.318
Carnival	74	287	9.261
City	90	1,691	9.104
Escorts	53	159	8.827
Trent	52	509	8.673
Castle	30	203	7.989
Forest's	22	31	7.607
weather	24	775	7.468
+2	**Cooccurrence Count**	**Raw Count**	**Log Dice**
United	85	482	9.391
Proud	56	1,769	8.399
Centre	43	606	8.365
Surrounding	31	47	8.096
Nottinghamshire	29	222	7.933
https://t.co/cjwzp2hvra	23	129	7.634
'austerity	20	27	7.471
council	20	129	7.432
V	26	1,360	7.405
cartel	19	25	7.398

Table 5c: Words appearing before Newcastle

-2	Cooccurrence Count	Raw Count	Log Dice
Medical	241	541	9.597
Move	256	1,248	9.587
Pitch	153	481	8.950
arrivals	124	148	8.695
partizan	91	107	8.254
Tank	91	132	8.251
City	98	1,143	8.216
injured	76	523	7.934
Skysports_keith	80	1,416	7.887
eveningchron	73	2,099	7.669
-1	**Cooccurrence Count**	**Raw Count**	**Log Dice**
In	1,222	52,447	9.332
After	244	5,390	9.053
At	402	22,666	8.670
For	480	60,647	7.805
To	737	99,362	7.791
From	103	16,126	7.032
V	44	1,616	6.999
strong	38	535	6.933
#nufc	56	6,613	6.817
Vital	33	46	6.800

Table 5d: Words appearing after Newcastle

+1	Cooccurrence Count	Raw Count	Log Dice
Upon	1,307	1,991	11.844
United	1,111	1,835	11.629
Room	168	1,170	8.990
City	175	1,820	8.965
Fans	115	2,172	8.315
University	65	376	7.730
And	401	56,880	7.625
Quayside	48	333	7.299
Are	137	20,179	7.233
Theatre	42	270	7.115
+2	**Cooccurrence Count**	**Raw Count**	**Log Dice**
Tyne	1,299	2,445	11.780
#nufc	496	14,240	9.409
Centre	108	773	8.406
Tickets	96	1,355	8.158
Whole	90	918	8.123
Perez's	76	90	7.997
Gateshead	76	1,209	7.840
V	81	2,260	7.799
Council	45	288	7.212
Never	60	3,689	7.203

Appendix D: Primary and Secondary Themes identified in Stage one tweets

Primary Theme	Secondary Themes from each corpus	
	Nottingham	***Notts***
Animal	Runaway deer, lost dog	Adoption, lost animal
Arts	Map, libraries, photography, UNESCO literature, print, painting, poster	N/A
Beauty	N/A	Hair
Breweries	N/A	N/A
Building and Construction	Electrical works, windows, kitchens, promotions	Electricity, public space
Business	Investment, exhibition	Events, partnership, #nottsmade
Charity	N/A	Football, music, fundraising, entertainment
Crime	Canabis, car theft, petrol, damage to car, threatening messages	Missing person, drug trafficking, burglary, weapons, notts police, control rooms, assault, theft
Education	University, phds, primary school, studentship, university of Nottingham	Primary school
Entertainment	Music, start time, promotion, emotion, venue, salsa dance, cat café, wrestling, tickets, bars, ghosts gambling, touring, new year's eve, television, glee club	Cinema, Christmas, outdoors, music, complaints, TV, travel, football, history, lost and found, question, YouTube
Events	Promotion, dancing, emotion, Christmas, santa, caves, the	Music, gardening, tickets, advertisement, charity, emotion, Christmas, wedding,

	bodega, photography, UV paint party, Christmas market, coca cola, New Year's eve, education	coca cola, splendour, birthday, football, ghosts, holly, promotion
Faith and Spirituality	Gathering	N/A
Food	Drink, cocktails, chocolate pudding, indian, Thai, dinner, pubs	Advertisement, real ale, cooperative, vegbox, emotion, beer, promotion, riverbank
Healthcare	Event, healthwatch, hospitals, ENT, psychotherapy	Heart, nhs, dental nurse, money and finance, paramedics
Heritage	UNESCO literature, lacemaking, lenton lodge	Penny red, Sherwood foresters
Home	Décor, lighting, furniture, copper tub, blinds	Carpets, moving home
Jobs	Telesales, cleaning, manager, payroll, office space, project engineer, electrical engineer, nursing, trainee, systems engineer, receptionist, sales advisor, radiographer, apprentice, electrician, business development executive, waste engineer, land surveyor, health safety and environment manager, commercial accountant, solar electrician, digital marketing executive, recruitment, science,	Teacher, assistant, receptionist, manager, dental nurse, IT specialist, planner, healthcare, estimator, technology, finance, sales, designer, design engineer

	store manager, search, childcare, teacher, law	
Money and Finance	N/A	Public space
News	Fundamentalism, ad blockers, royal mail, Syrian refugees, cattle, tent village	Building and construction, design, flooding
Promotions	Email campaign, early bird prices, book, music, commercial customers, recommendation, website, sports, calendar, building firm, lottery scratch cards, aerials, in-car entertainment, speakers club, afternoon tea	Advertisement, football, business cards, printing, house, food, competition, underwear, smoking, shopping, technology, clothing, music, car repair, savings, magazines, vans
Sex Work	Filming, escorts, jobs	None recorded
Shopping	Opening, Christmas, clothing, bike shop, vans, shoes	Christmas, new store
Sports	Fitness, football, triathlon, unknown sport, ice hockey, racing, rugby, walking, England's home of sport,	Football, cricket, unknown sport, trail run, gymnastics, cyclocross, news, Nottingham trent university, skating, golf, go karts, meadow lane, language learning
Technology	Data analysis, aerospace	Tesla coils
Travel	Train, cars, travel cards, delays, closures, road congestion, buses, taxis, airport,	Trams, public transport, roads, trains
Unknown	Santa, location, education, business,	None, foreign language, games, emotion, news,

	women, Nottingham, news, public space, city council, food, spotting, direct communication, talent, emotion, foreign language, question, video, community action, personal data, heritage, placenames	events, Christmas, politics, food, complaint, entertainment,communication, running, observation, comment, healthcare, origins, temperature, question, information, explosion, fur, travel, hashtags and a link, sports, fire services, sports, opening hours, broadcast
Weather	Rain, mild, mild temperature, heavy rain, humidity, winds	Sea, none, disruption

Appendix E: Code used in Stage three data gathering

```
1 import twython
2 from twython import Twython, TwythonError
3
4
5 import datetime
6
7 import os # to get file size
8
9 import csv
10
11 import string
12
13 import json
14
15 import pprint
16
17 from datetime import timedelta
18
19 from datetime import datetime
20
21 from pprint import pprint
22
23 import time
24
25 from time import gmtime, strftime
26
27 #Define API Variables
28 #API KEY
29 app_key = "xxxxxxxxxx"
30
31 #API SECRET
32 app_secret = "xxxxxxxxxx"
33
34 #OAUTH TOKEN
35 oauth_token = "xxxxxxxxxx"
36
37 #OATH TOKEN SECRET
38 oauth_token_secret = "xxxxxxxxxx"
39
40
41                                    twitter                                    =
Twython(app_key,app_secret,oauth_token,oauth_token_secret)
42
43
44
45
46
47 #set search function
```

```
key_words = ['#Nottingham']
#for each word in the key words, do something:
#for key_word in key_words:
#search_results = twitter.searchTwitter(q=key_word, rpp='200')

# This function will write a new file every hour
def writeToFileHourly( textToWrite ):
# Construct the filename, append if already exists
# Filename is Output_DDMMYY_HH.txt
dt = datetime.now()

# https://docs.python.org/2/library/time.html for strftime types
 filename = 'Output_#Nottingham' + dt.strftime('%d%m%y_%H') + '.txt'
```



```

# always append (a) so existing file isn't overwritten
with open(filename, "a") as myfile:
myfile.write(textToWrite + '\n')

#this function will write the twitter text
#def textToWrite:



#file=open(filename,"w")

currentFile = ''

# This function will check the filesize
def writeToFile( textToWrite ):
global currentFile # pull in the global variable above
if not currentFile:
dt = datetime.now()
            currentFile = 'Output_#Nottingham' + dt.strftime('%d%m%y_%H%M%S') + '.txt'

# Check the file exists
if os.path.isfile(currentFile):
statinfo = os.stat(currentFile)
print('File is ' + str(statinfo.st_size) + ' bytes')

# Is the file bigger than 1GB
if(statinfo.st_size > 1000000000):
dt = datetime.datetime.now()
            currentFile = 'Output_#Nottingham' + dt.strftime('%d%m%y_%H%M%S') +
'.txt'
print('Creating new output file ' + currentFile)

# always append (a) so existing file isn't overwritten
```

```
94 with open(currentFile, "a") as myfile:
95 myfile.write(textToWrite + '\n')
96
97 import time
98
99
100 writeToFileHourly( "woo!" )
101
102
103
104 #start count for start of twitter search
105 cnt =0
106 until = 0
107 while True:
108
109 try:
110 search_results = twitter.search(count= 200,q=key_words
,max_id=until)#there is a 100 result limit here, so the 200 is just to
ensure that the full 100 are collected
111
112 for tweet in search_results['statuses']:
113 pprint (tweet ['text'].encode("utf-8"))
114
115
116 print (tweet ['entities']['hashtags'])
117 print cnt,tweet ['created_at'],
118 until = tweet ['id']
119 print until
120 Text=(tweet ['text'].encode("utf-8"))
121 TextData =(tweet ['created_at'].encode("utf-8"))
122
123
124 writeToFile (str(cnt))
125 writeToFile(TextData)
126 writeToFile(Text)
127
128
129
130 cnt += 1
131
132 except TwythonError as e:
133
134 if str(e) == 'Twitter API returned a 429 (Too Many Requests), Rate
limit
exceeded':
135 print "Bounced off limiter"
136 timeleft =
int(twitter.get_lastfunction_header('x-rate-limit-reset'))-time.time()+
30
137 print timeleft
```

```
138 time.sleep(max([timeleft,1]))
139
140 print cnt
141
142
143
```

Appendix F: Primary and Secondary themes identified in stage three tweets

Primary Theme	Secondary Themes from each data set	
	Nottingham 1000 tweets	**Newcastle 1000 tweets**
Animal	Lost dog, birds, dogs, excited dogs, dog, missing cat, peregrine falcon	pets, birds
Arts	Puppet festival, lakeside arts, dance, european capital of culture, hoopla works, focus gallery, photography, film, illustrator, illustration, literature, young artists, jewellery, statue, fashion, equestrian artists, states of America	Gallery, for sale, promotion, classes, designfest, weaving, fine art, craft television, theatre, design, david holliday, artwork, signage, student artwork, painter, tattoos, photography, comics, film, collective, jewellery, creatives, print
Beauty	tan, spray tan, hair, botox, skincare, client review, spa	Salon, Halloween makeup, nails, hair, massage, swimming pool, reiki
Building and Construction	Long eaton cinema, business park, windows, carpets, roofing, tool hire, plumbing, kitchens, development of bridgford house, property,	Refurbishment

Primary Theme	Secondary Themes from each data set	
Business	contemporary extensions Payroll, recruitment area, business support, startup companies, vox pops, solicitors, coworking space, networking, assessment of professional competence, charity, investing, competition, working from home, expansion, homelessness, accelerate places, rooms to rent, technology, accounting, funding	Office space, family law, in bloom, local exhibitor, bidding, CBS takeover, defence industry growth, coal
Charity	Fundraising, running	Network, homelessness, community, foodbank, fundraising
Communication between Friends	Valentine, cars, fashion	Autumn blues, freshers week, night off, broadcast, night out, family
Crime	Dangerous driving, road collision, motorcyclist crash, trials, bike theft, knuckle duster, sexual assault, armed officers, policeman, armed robbery	Paedophiles, grooming, indecent assault
Education	Pupils, the studio,	Trip, university, artwork,

Primary Theme	Secondary Themes from each data set	
Entertainment	university, PhD, PAT testing, students, music, school to reopen, freshers	event, construction college, students, law school robotics
	Radio, music, dancing, stripping, choral music, games, comedy, music rehearsal, film, railway, ping pong, adventure golf, theatre, ghost tours, dinosaur, acting, food, irish pub, video games, bars, night out, escape room, day out, Katie price, dating	Youtube, improve, music, LGBT choir, night out, night life, dance, circus, theatre, nightclub, TV, film, musical, talk, swimming pool,
Events	Space, St, Patrick's, music, puppet festival, paper clay, book signing, students, poetry, bookshop, Warhammer, literature, art, art gallery, silver ring making, cigar evening, French living, mircrobrewery, hoopla workshop, book group, exhibition launch, meet the makers, wedding, goose fair, food, planning, Christmas, discovery evening, networking,	Bars, music, CEO sleepout, pub quiz, girls night out, guide to activities, club nights, music, culture crawl, tour, homelessness, engagement party, women in law, Oktoberfest, public meeting, party, comiccon, circus, dance music festival, theatre, wedding show, Ibiza party, pop up shopping, business lunch, bar crawl, country file, northern stars

Primary Theme	Secondary Themes from each data set	
	communications workshop, Halloween, refugees, top things to do, ballroom, course, place available, pub tour, mental health, ghosts of Nottingham, crafts, robin hood pageant, conference, Melbourne art festival, justice museum, shopping, green festival, wollaton hall, graduation, food festival	celestial events, students, allergy talk, healthcare, write now live, Jurassic kingdom, charity ball, networking meeting, consent rules, technology, baby and toddler show, music masterclass, food, social event, accreditation information, comedy, female networking
Faith and Spirituality	church move	N/A
Food	Lunch, relaunching menus, afternoon tea, café, desserts, catering, healthy eating, pub, beer, reviews, community food share, fruit, cake, street food, events, le bistrot pierre, vegan, Chinese, brunch, masterchef, coffee	Cocktails, eating out, guide, food guide, beer, café, cupcakes, promotional crepes, coffee, Monday meal, drink, baked goods, vegetarian, Malaysian, promotions, reviews, pancakes, pizza,
Healthcare	Vision, mental health, hearing, weight management, autism, asthma, dentist, pelvic floor, clinical trials,	Event promotion, mindfulness, mental health, cancer, carers, defibrillator, allergies, reiki, physiotherapy, MRI

Primary Theme	Secondary Themes from each data set	
	fertility, events, wellbeing, prostate cancer	scanning, presentation, midwife
Heritage	Robin hood, historic image, comparing old and new, historic fact, historic transport, train, boots department store	Historic image, body snatching, quayside, images of local area, history, punishment, historic footage, on this date, food, coal, 17th c. drunks, football, school meals, elswick works, crime, film, clothing
Home	Property, property for sale, property management, removals, property interiors, interior decoration, carpets, landlords, property lettings, recycling, property listing, property test drive, letting agents, doors, office rental, paving, art deco bottle, property valuation, cleaning, architecture, apothecary jars, landscaping, bathroom, laminate, self storage,	Furniture, cleaning, family homes, property, removals, housing, luxury property
Jobs	Delivery cyclists, property buyer, apprentice technician,	Events specialist, UX specialist, sport and fitness tutor, vet

Primary Theme	Secondary Themes from each data set	
	credit controller, general recruitment, assistant psychologist, workers rights, veterinary surgeon, CE driver, mental health specialist, vet nurse, driving instructor, electrician, nursing and care, freelance, finance professionals, salary, retail apprentice, hotels, administrator, python developer, multiple roles, business park, apprentice, accountancy, customer service, medicine, content editor, sales advisor, editors, part time lawyer, team member, advert, royal mail seasonal	assistant, citi bank, general recruitment, digital delivery director, brand ambassador, store team, infrastructure engineer, healthcare, accountancy, hotel jobs, direct support professional, general recruitment, nursing, hospitality, facilities management, sales, customer service, lift service engineer, draughtsman, lift commissioning engineer, teaching assistant, security officer, nursing, staff testimonials, product safety engineer, general careers, social media and marketing, disability analyst, digital apprentice, career development, care professionals, construction, sales, landscape architect, project senior analyst, prep cooks, Wendy's crew member

Primary Theme	Secondary Themes from each data set	
Money and Finance	N/A	Banking crisis
News	Unions, journalism, torvil and dean, coworking spaces, jobs, charity bake off, motorcyclist crash, coffee, business, news, rapist, European capital of culture, house prices, housing, conversion, job losses, rapist jailed, working parents, politics, trends, blog, accident, east midlands hour, slavery, building closure, gardening, animal rights protest, building closure, musician, cycle network	Jobs, banking crisis, traffic accident, places to live, apartments, community bins, paedophiles, power plant, veues, training, defence industry recommendations, bullying and nepotism, cricket, Cassie Ormsby, shopping, assault, politics, grooming, community vote, car crash, Islam, rape, council legal case
Promotions	Hotel, food, payment, dinosaur, social media expert, newspaper, party space, florists, money exchange, things to do, spaces, tourism, supermarket, decorating, contracting businesses, mediation, opening hours, gay sauna, printing offers, hot tubs, meeting room, tourism,	Restaurant guide, bars, club nights, pizza, laptop repair, food, car, guide to activities, cinema, competitions, discount voucher, apartments, Jurassic kingdom, drinks, florists, gardening, shopping, Geordie, beauty, events, travel, tourism social media, laundry service, SEO,

Primary Theme	Secondary Themes from each data set	
	CCTV, story, accountancy, sausage rolls, coffee, beauty, juicer review	furniture, bars, housing, websites, camera, clothing, design, city in top 5, community connections, reiki, VW camper van hire, legal process efficiencies, travel guide, wedding, deals, make Newcastle greater, reiki, bikinis, space, design, community, masterclass, offers, food, photos, events, music, digital marketing, booking, Texan food, jewellery, community support
Sex Work	Gymnast, teen, av model, swinging, milf, brazillian, white girls, outdoor, dildo, facial, farmer, closeup, nuru, amateur, petite, dominatrix, plumper, maid, muslim, interracial, toys, anal, girl, toys, old mom, mature, pantyhose, sucks	Advertisement, escort, gang bang, porn, masturbation, Mandy More, Latina, shemale, swimming pool, snicka, girl of the hour, interracial, amateur, teens, August Ames, outcalls, wedding planner, escort recruitment, busty babe, little girl, chubby, ass
Shopping	Bridal gown, shoes, fashion, vintage,	Wholesale, football mugs, Patagonia, in store,

Primary Theme	Secondary Themes from each data set	
	stationery, climbing shoes, garden centre	camera, clothes, table and chairs
Sports	Running, football, betting, basketball, rowing, climbing, rugby, ice hockey, triathlon, astroturf, figure skating, walking, athletics	Betting, football, yoga, wrestling, marathon, foreign language, cycling, horseracing, dance acro, Krav Maga, bootcamp, basketball, golfing, gambling, cricket
Technology	Living and working, hackathon, consumer finance, broadband	News, training, food rating app, webdesign, digital agencies, internet
Travel	Cycling, trains, road delays, road collision, road closure, unsafe taxis, cost of gas, train delays, trams, flights, busses, driving, parking, railway coach, helicopter search	Roads, companion, trainspotting, trains, incident reporting, metro, cars, rail, speed camera
Unknown	List of hashtags, rough sleeping, landmark, digital influence, tourism, landmark demolition, law, waterway, general promotion, local language, outdoors, broadcast, recommendation, reader's letters, citizens, architecture, fundraising,	Northeast, satanic tory party, conspiracy theorists, social commentary, visit, course, goodness, bridge, hunt, free Tibet, reportage, broadcast, specials, foreign language, images of local area, recommendation, politics, community

Primary Theme	Secondary Themes from each data set	
	homelessness, wedding, communication, video clip, caves, report on today, report on last night, sales	politics, socializing for women, community coat rack, local people, petition, blogging, question, marriage equality, top 5 entertainment, rubbish collection, office picture, list of locations
Weather	Storm and rain, rain	Rain showers, clear sky, light rain, broken clouds, rainy, scattered clouds, light intensity drizzle rain

Appendix G: Manually Extracted Locational Refereances from Stage Three

Locational Reference from each Corpora	
Nottingham 1000 tweets	**Newcastle 1000 tweets**
nottingham, nottsvip, mansfield, eastmidlands	Italy London Chamonix Scotland Oman Gallery Ponteland, newcastle, northeast hour
wearecyclingUK, pedestrian areas	
oscarsnottm	astonvilla, watford, preston, birmingham, newcastle
nottingham, missingpetsGB	
nottingham	Italy London Chamonix Scotland Oman Gallery Ponteland, newcastle, northeast hour
nottingham	
nottmcitycouncil,nottingham, tucmidlands	Italy London Chamonix Scotland Oman Gallery Ponteland, newcastle, northeast hour
nottingham, rome	
cityartsnotts, royalnottingham, trentuni, city arts, nottingham	newcastle
nottingham	Italy London Chamonix Scotland Oman Gallery Ponteland, newcastle, northeast hour
ukhotels, de vere orchard hotel, nottingham university	
nottingham, long eaton	newcastle
nutri2gouk, nottingham, ng1, ng7	newcastle, barsinnewcastle
cityartsnotts, royalnottingham, trentuni, city arts, nottingham	newcastle, geordie
	Hull, newcastle, polarbearhull
bildurnnews, nottingham	Italy London Chamonix Scotland Oman Gallery Ponteland, newcastle, northeast hour
cityartsnotts, royalnottingham, trentuni, city arts, nottingham	
nottingham	cpfc, nufc, epl, newcastle
cityartsnotts, royalnottingham, trentuni, city arts, nottingham	Northeast, newcastle
	nufc, efc, newcastle
nottingham	nufc, spurs, thfc, coys, newcastle
cityartsnotts, royalnottingham, trentuni, city arts, nottingham	newcastle
	newcastle
nottinghan, fernvilladogs, missingpetsgb	northeasthour, newcastle
	london, newcastle
guyphoenixuk, nottingham	Italy London Chamonix Scotland

Locational Reference from each Corpora	
nottingham	Oman Gallery Ponteland
nottingham	northeasthour, newcastle
nottingham	chepstow. Yarmouth, galway,
nottingham	redcar, newcastle
crosskeysnottm, nottingham, notts	Hull, newcastle, polarbearhull
nottinghamnouse, nottingham	Italy London Chamonix Scotland
nottingham, bristol	Oman Gallery Ponteland,
nottingham, derby	northeasthour, newcastle
cityartsnotts, royalnottingham,	northeasthour, newcastle
trentuni, city arts, nottingham	northeasthour, newcastle
england, oldestinninengland,	st james's park, newcastle
yeoldetriptojerusalem, england,	Northeasthour, nefollowers,
nottingham, lovenotts	newcastle, gateshead, sunderland
nottingham, derby	northeasthour, newcastle
nottingham	usatodaynetwork, newcastle
berketexbride, nottingham council	nefollowers, jesmond, newcastle
house	london, newcastle
nottingham, fernvilladogs,	Italy London Chamonix Scotland
missingpetsGB	Oman Gallery Ponteland, newcastle,
nottingham	northeasthour
trentnavigation, west bridgford,	us-202, newcastle
nottingham	liverpool, london, camden,
heywhatsonnotts, nottingham	manchester, croydon, kensington,
nottingham	newcastle
nottingham, A52	barsinnewcastle, newcastle
nottingham	northeasthour, newcastle
nottingham	levoinailbar, salonblogger, newcastle
cartwrightrooms, roxyballroom,	newcastle
nottingham	barsinnewcastle, newcastle
nottingham, nottsvip	Italy London Chamonix Scotland
nottinghamnouse, nottingham	Oman Gallery Ponteland, newcastle,
heywhatsonnotts,	northeasthour
midlands_news,	liverpool, manchester, croydon,
what's on notts,	nufc, cpfc, psg, epl, newcastle
royalnottingham, trentuni,	northeasthour, newcastle

Locational Reference from each Corpora	
nottingham, lakeside arts	Newcastle Italy London Chamonix
cartwrightrooms, roxyballroom,	Scotland Oman Ponteland
nottingham, nottsntuthehive	Northeasthour
nottingham, nswnotts	Northeasthour, Gateshead
cqnottm, nottingham	Newcastle Tyneside
bildurnnews, nottingham	bdailynortheast, Gosforth Newcastle
nottingham, lovenotts	designeventne, newcastle
nottingham, uk, denverbroncos	London Newcastle Manchester
cqnottm, nottingham	Leeds Oxford Cambridge
heywhatsonnotts, nottingham	Newcastle Italy London Chamonix
midlands_news, nottingham	Scotland Oman Ponteland
heywhatsonnotts, nottingham	Northeasthour
waterstones, nottingham	Newcastle Northeasthour
hilton doubletree hotel, nottingham, NG8	Northeasthour, Newcastle
	designeventne, Newcastle
heywhatsonnotts, nottingham	Newcastle Northeasthour
nottingham, bulwell	Newcastle Northeasthour
nottingham, uk, denverbroncos	regionofDurham Bowmanville
nottingham	Oshawa Newcastle
nottinghamnouse, nottingham	Newcastle Northeasthour
heywhatsonnotts, nottingham	Newcastlesic, newcastle
nottingham five leaves bookshop	Newcastle city centre
heywhatsonnotts, nottingham	barsinnewcastle, newcastle
heywhatsonnotts, nottingham, focus gallery	Newcastle Italy London Chamonix
	Scotland Oman Ponteland
heywhatsonnotts, nottingham, focus gallery	Northeasthour
	hwnewcastle, newcastle, gateshead
heywhatsonnotts, nottingham, the watered garden	newcastle st james park eldon square
heywhatsonnotts, nottingham, harts hotel and restaurant	newcastle northeasthour
	Newcastle Italy London Chamonix
heywhatsonnotts, nottingham	Scotland Oman Ponteland
heywhatsonnotts, nottingham, revolucion de cuba	Northeasthour
	newcastle jesmond northeastfollower
heywhatsonnotts, nottingham, the	Newcastle Italy London Chamonix

Locational Reference from each Corpora	
angel microbrewery	Scotland Oman Ponteland Northeast
heywhatsonnotts, nottingham, running horse	Newcastle Northeashour newcastle jesmond nefollowers
hilton doubletree hotel, nottingham, NG8	Northeasthour Gateshead Newcastle Tyneside
safenottm, asbnottingham	nightsoutne, newcastle
ng4, ng5, nottingham	Newcastle Italy London Chamonix
broxtowe, nottingham	Scotland Oman Ponteland
nottingham, the angel microbrewery	Northeasthour
nottingham games workshop	newcastle, barsinnewcastle
midlands_news, nottingham	newcastle, city centre
heywhatsonnotts, nottingham, the watered garden	Newcastle Italy London Chamonix Scotland Oman Ponteland
heywhatsonnotts, nottingham	Northeasthour
heywhatsonnotts, nottingham	Newcastle Italy London Chamonix
broadmarsh car park, nottingham, westbridgford	Scotland Oman Ponteland Northeasthour
nottingham, westbridgford, gedling	newcastle eldon square, nightsoutne
nottingham	geordieshaun, newcastle
nottingham,	Newcastle Italy London Chamonix
manx, landlocked, nottingham	Scotland Oman Ponteland
sneinton, nottingham'	Northeasthour
nottingham, five leaves bookshop	gateshead newcastle northeasthour
uniofnottingham, nottingham, the studion, portland building	northeast newcastle eldon square, pizzastormuk
hockley, nottingham, cqnottingham	gateshead newcastle northeast
ruddington, nottingham	newcastle uk southafrica
nottinghamhenparty	newcastle gateshead
employmentlawuk, nottingham	Northeasthour Gateshead Newcastle
employmentlawuk, nottingham	Tyneside
heywhatsonnotts, nottingham,	newcastle the world
heywhatsonnotts, nottingham	gateshead newcastle
nottingham, nottingham_post, nottsvetsoc, leicester_merc	newcastle geordies newcastle united toon
nottingham, european capital of	newcastle tyne and wear uk

Locational Reference from each Corpora	
culture, languagetuitionnottingham	newcastle nufc magpies toon
nottingham	eldon square
britishtinnitus, nottingham	the castle keep newcastle
nottingham, champagnebar	newcastle northeast
uniofnottingham, nottingham, the	Newcastle Northeast
studion, portland building	newcastle liverpool
canal, waterway, nottingham, bridge,	newcastle the world
towpath	newcastle united newcastle nufc
nottingham, former swimming baths	newcastle gateshead
site	newcastle gateshead
, sneinton	hebburn newcastle
, the watered garden	newcastle the world
heywhatsonnotts, nottingham	newcastle the world
heywhatsonnotts, nottingham	redhousefarm newcastle
sherwood,	northeast newcastle
nottingham, harts hotel and restaurant	liverpool manchester london leeds
nottingham, birmingham	glasgow
midsbusiness, nottingham, trentuni	newcastle blackburn manutd
nottingham, carluccios	newcastle toon europe
indoor christmas market, nottingham	newcastle nufc
auction centre	newcastle jesmond northeast
nottingham, french, broadmarsh	st james's park newcastle
victoriatayluk, nottingham	newcastle nufc
nottingham	newcastle stoke city
cpoint_nottcity, nottingham	quayside newcastle
nottingham, de cuba	newcastle northeast, newcastle
midshealthpsych, autismeastmids,	quayside newcastle
nottingham	newcastle gateshead
heywhatsonnotts, nottingham, the	newcastle western australia
bodega	gateshead
westbridgfrdwire, sneinton,	geordies ufc toon
the park, nottingham	gateshead
nottingham, the angel microbrewery,	quayside newcastle
rome	gateshead
nottmmarkets, nottingham,	northeast newcastle

Locational Reference from each Corpora	
nottmcitycouncil	newcastle jesmond northeast
midshealthpsych, autismeastmids, nottingham	uk newcastle uk
nottingham	levoinailbar, newcastle
heywhatsonnotts, nottingham, rescue rooms	berkshire, lancashire, derbyshire, cumbria, cambridge, peterborough,
heywhatsonnotts, nottingham, the bodega	liverpool, middlesborough, newcastle, norfolk
nottingham, nottingham_post, nottsvetsoc, leicester_merc	berkshire, lancashire, derbyshire, cumbria, cambridge, peterborough,
heywhatsonnotts, nottingham, rescue rooms	liverpool, middlesborough, newcastle, norfolk
nottingham	mahattan, levoinailbar, newcastle
nottingham, rescue rooms	newcastle, northeast
rushcliffe halt, east leake, gcrnottingham,	chinese, mourne mountain marathon manunited, mancity, chelsea,
, animalsearchUK	newcastle, northeast
, the bodega	newcastle, gateshead, nufc
sneinton,	wolves, birmingham, newcastle, uk
broadmarsh,	yanks, newcastle
, ng1, ng7	newcastle
, nottingham_post, nottsvetsoc, leicester_merc	newcastle, liverpool newcastle, northeast
nottingham	northeast
carringtonstreet, nottingham	dortmund, newcastle, barkley
notts	newcastle, oxfordhouse, ne1,
nottingham, carlton road, severn trent	northumbria uni, newcastle uni
nottingham, carlton road, severn trent	newcastle, carlisle
nottingham, trentuni	nufc,
nottingham, nottingham_post, nottsvetsoc, leicester_merc	britain, gallery, newcastle britain, gallery, newcastle
nottingham, theatre royal, 16 pilcher gate	alexandra palace, australia, sydney, melbourne, brisbane, perth, adelaide
nottingham, nottingham_post, nottsvetsoc, leicester_merc	uk britain, gallery, newcastle

Locational Reference from each Corpora	
newark, , nottingham_post, nottsvetsoc, leicester_merc	newcastle, staffordshire britain, gallery
harlemglobetrotters, , nottinghamarena	, geordies, nufc, toon newcastle, uefa
, east midlands airport	newcastle upon tyne, quayside
nottstweetup, mpwnottingham, nottingham	tibetan, newcastle upon tyne, central station, tibet
notts_tv, nottingham, broadmarsh car park	oktoberfest, times square,
nottingham, nottingham	newcastle, nufc, efc, afcb
adcnotts, nottingham, notts, nottinghamshire	newcastle, tavern redcar, chepstow, newcastle
newark,	london, newcastle, manchester
countyflooring, nottingham, keyworth,	jesmond, gosforth, newcastle
ladybay, westbridgford, edwalton, bradmore	newcastle, apartments newcastle, liverpool
NG, castlewood business park, J28,	redcar, chepstow, newcastle
Ngchartsurv,	newcastle gaol
zinnia, nightclub, zinnia	manutd, newcastle united, old
countyflooring, nottingham, keyworth,	trafford, UK
ladybay, westbridgford, edwalton, bradmore	newcastle, united kingdom daverson hall, wideopen, aed,
norwich, london, nottingham	, newcastle freshers
midshealthpsych, autismeastmids, nottingham	, carlisle , city centre
euro car parts, nottingham	,, city centre, bars and clubs
coworkingspace, nottingham, notts,	newcastle, geordies, nufc, toon
hockley, lacemarket	newcastle, liverpool
nottstweetup, mpwnottingham,	prudhoe, hexham, corbridge,
rescuerooms, nottingham	durban,
nottstweetup, mpwnottingham,	byker, heaton, northeast
norwich, london, nottingham, hull	london, birmingham, newcastle,
derby	nottingham, staffordshire, high
, nottingham2023	wycombe
nottshotprop, nottingham	northeast, newcastle, ponteland,

Locational Reference from each Corpora	
nottstweetup, mpwnottingham,	darras hall
midshealthpsych, autismeastmids,	man utd, mancity, chelsea,
castlewood business partk, J28,	, liverpool
Ngchartsurv	circus, newcastle
NG, castlewood business park, J28	newcastle, liverpool
nottingham, townsville, melbourne	northumbria, newcastle freshers
longeaton, nottingham, derby	, toon army, europe
holden copley	newcastle, gateshead
nottingham, mansfield	prudhoe, hexham, corbridge,
loughborough,	, durham
nottshotprop, nottingham	, metrocentre
westbridgford, nottingham	, metrocentre
nottingham, the pudding pantry	, liverpool
hilton doubletree hotel, nottingham,	, liverpool
NG8	, toon army, europe
the bridge, canal, waterway,	theatre,
nottingham, bridge, towpath	brighton, newcastle united, brighton,
hilton doubletree hotel, nottingham,	wien, vienna, sunderland,
NG8	newcastle united, nufc
the bridge, canal, waterway,	manchester, newcastle
nottingham, bridge, towpath	newcastle, film and comic con,
hilton doubletree hotel, nottingham,	circus central, christ church
NG8	eldon square, dolgellau, newcastle
trinity school, nottingham	swimming pool, newcastle
nottingham, keyworth, ladybay,	newcastle, quayside, northeast
westbridgford, edwalton	newcastle, quayside, northeast
, animalsearchuk	newcastle, south africa, earth
, animalsearchuk	newcastle, northeast, circuscentral
, animalsearchuk	newcastle, northeast
the right lion, nottingham, leftlion	newcastle, japaneseknotweed
, animalsearchuk	newcastle, liverpool
een_midlands, russia,	newcastle, japaneseknotweed
, animalsearchuk	newcastle, northeast
notts, nottingham, happynottm,	mancity, manunited, chelsea
nottm_con	newcastle utd, manutd

Locational Reference from each Corpora	
manx, nottingham, gedling	berliner
japanese, nottingham	newcastle, liverpool
dvsagovuk	newcastle, geordie
nottingham, animalsearchuk	newcastle, liverpool
welstead avenue, nottingham	newcastle, geordies, nufc, toon
nottingham, missingpetsGB, leeds	, graindermarket1, nufcfoodbank,
right lion, nottingham, left lion	generousgeord
midlands_news, nottingham	newcastle, newcastle upon tyne
pakistan, nottingham, rupees	newcastle, newcastle upon tyne
nottingham market square	tottenham, newcastle, crystalpalace
hilton doubletree hotel, nottingham,	newcastle, quayside,
NG8	northeasttweets, northeastartcol,
nottingham, missingpetsGB	seenewcastleuk, newcaste_views
ukpub, chesterfield, chesterfieldis	newcastle, theatre
russia,	newcastle, northeast
servocanotts,	newcastle, notheast
hilton doubletree hotel, nottingham,	, nufc
NG8	geordie birthday, newcastle,
heywhatsonnotts, nottingham	geordie_gifts
eastmidlandsdentist,	northeast, newcastle
nottinghamdentist, nottinghamevents,	geordie birthday, newcastle,
nottingham	geordie_gifts
broadway cinema	newcastle, world class
westbridgford	jesmond, newcastle
nottingham, missingpetsGB	newcastle, world class
wollaton, nottingham	, world class
nottingham, uk	newcastle, quayside,
under the bridge, canal, waterway,	northeasttweets, northeastartcol,
nottingham, bridge, towpath	seenewcastleuk, newcaste_views
notts, workfromhome	highwestjesmond, gosforth,
nottingham, missingpetsGB	jesmond, newcastle
nottingham, missingpetsGB	newcastle, world class
nottingham	newcastle, gateshead
nottingham	ringsideworld, newcastle
ng4, ng5, nottingham	dean street, newcastle

Locational Reference from each Corpora	
nottingham, missingpetsGB	newcastle
nottingham, notts, nottinghamshire	cardiff, stockport, portsmouth,
cotgrave,	norwich, colchester, newcastle
hilton doubletree hotel, nottingham, NG8	bdailynortheast, gosforth, designeventNE, Ndesignfest,
hilton doubletree hotel, nottingham, NG8	ibiza, gosforth, newcastle
uk, nottingham, place	cardiff, stockport, portsmouth, norwich, colchester,
nottinghamblog, nottingham	liverpool, belfast, newcastle
nottingham	crystal-palace, newcastle, tottenham
café, food outlet, nottingham	ndesignfest,
hilton doubletree hotel, nottingham, NG8	newcastle, quayside
	newcastle, quayside
visitnotts, nottinghamshire, nottingham, lovenotts	newcastle, quayside dankofengland, newcastle
nottingham, missingpetsGB	newcastleeagle, newcastle, nclsport
nottinghamshire, lovenotts,	café, fazalscafe
rushcliffe halt, east leake,	theatre, newcastle
gcrnottingham, nottingham	newcastle, nhs
visitnotts, nottingham, european	city
capital of culture, nottingham2023, missingpetsGB	newcastle, quayside west coast
hilton doubletree hotel, nottingham, NG8	highwestjesmond, gosforth, jesmond,
nottingham, school, former leisure centre	handelsbankenuk, newcastle nufc, newcastle
nottingham	newcastle, geordies, nufc, toon
tattershalls, nottingham	, west jesmond, jesmond
nottingham, visionexpress, market square, visionvan	fqnewcastle, newcastle, ne1bid, northeastlife
hilton doubletree hotel, nottingham, NG8	newcastle, west jesmond, jesmond barsinnewcastle, tyne theatre,
nottingham, missingpetsGB	barsinewcastle, tyne theatre,
, roxyballroom	newcastle, metroradioarena

Locational Reference from each Corpora	
arnold, nottingham, ng5, doglostuk	newcastle, iof_ne
nottstoolhire, shops, warehouses, small offices, homes, nottinghamshire	tottenham, gosforth,
czech, nottingham	newcastle, iof_ne
nottingham, swimming baths, church, westbridgford, gedling, beeston	newcastle_krav, newcastle, washington, durham
uk, nottingham	newcastle_krav, newcastle,
lace market, nottingham	washington, durham
arnold, nottingham, ng5, doglostuk	newcastle_krav, newcastle,
shops, warehouses, small offices,	washington, durham
homes, nottingham	newcastle_krav, newcastle,
highfields park, nottingham	washington, durham
nottingham	pandon quay, quayside, newcastle
visionexpress, visionvan,	newcastle, washington
visionexpress, visionvan, nottingham	usatodaynetwork, newcastle
highfields park, nottingham	london, newcastle
nottingham, missingpetsGB	newcastle, byker, heaton
notts_tv, nottingham, broadmarsh car park	newcastle, gasteshead, durham, sunderland
nottingham, derby	northeast, newcastle
nottingham, broadmarsh car park	walbottlecampus, westdentoncc,
sheffield	netimesmagazine, ptmynewcastle,
nottingham, townsville, melbourne	newcastle, dinnington
hometown, nottingham, cleetownfc	, gateshead, durham, sunderland
the bridge, canal, waterway,	newcastlecars, newcastle
nottingham, bridge, towpath	thebeaconne4, newcastle
hometown, nottingham, cleetownfc, topcity	bridge over ouseburn, newcastle, luvthenorth444
midshealthpsych, autismeastmids,	wefoodbank, toonarmy,
nottingham	graingermarket1, newcastle,
chuc_org_uk, ukhotels, de vere	nufcfoodbank, generous geordie
orchard hotel, nottingham university campus	prudhoe, hexham, corbridge, newcastle
nottingham, westbridgford, gedling	in the house, newcastle
pakistan, nottingham, rupees	newcastle

Locational Reference from each Corpora	
nottingham, bournemouth, glasgow	ptmynewcastle, newcastle
broadmarsh, nottingham	west, newcastle
institutemh, nottingham conference centre	newcastle PA
	newcastle, geordies, nufc, toon
birmingham, bristol, derby	northsidetattooz, newcastle
nottingham, bournemouth, glasgow	dortmund, newcastle, barkley
chilwell, nottingham, derby	, nsw
plumcenteruk, nottingham	, newcastleunited, nufc, antwerp
plumcenteruk, nottingham	liverpool, belfast,
plumcenteruk, nottingham	ptmynewcastle,
germany, nottingham, longeaton	newcastle united, newcastle, nufc
ng4, ng5	ngwrestlinguk, newcastle
rushcliffe halt, east leake, gcrnottingham,	newcastle, westend
	graingermarket crepe bar,
nottinghamphotographer	newcastle, tyne, bridges
nottingham, windermere, lake district	toon army, newcastle,
no1tingham, nottingham, broadmarsh	graingermarket1, nufcfoodbank
car park, mynottingham	newcastle, iof_ne
nottingham, townsville, melbourne	newcastle, quayside,
notts,nottinghamshire	northeasttweets, northeastartcol,
world famous, nottingham	seenewcastleuk, newcaste_views
courtyard, nottingham	newcastle
world famous, nottingham	tuhiki, newcastle, tiger tiger
london, birmingham, newcastle,	ne1,
nottingham, staffordshire, high wycombe	geordie, toonlegend, newcastle, geordiemoments
rushcliffe halt, east leake, gcrnottingham, nottingham	tuhiki, newcastle, tiger tiger
	tuhiki, newcastle, tiger tiger
rushcliffe halt, east leake, gcrnottingham,	jungle_ncl, jungle, newcastle, house of smith
backstage, nottingham	newcastle, forbiddenplanet
visitnotts, pit_pendulum, nottingham, historical pub	, liverpool
	arenanewcastle,
broadmarsh car park, nottingham, westbridgford	wefoodbank, toonarmy,
	graingermarket1, newcastle,

Locational Reference from each Corpora	
broadmarsh car park, betterbroadmarsh, derby	nufcfoodbank, generous geordie graingermarket, london, manchester, liverpool,
jury's inn, london rd, nottingham, nottsbirders, nottswildlife, nottsbirding	birmingham, scotland, wales, england, uk,
nottingham, broadmarsh car park, mynottingham	mushroomworks, north east, ouseburn, newcastle
institutemh, nottingham conference centre	libertylivingUK, libertyquay, prudhoe, hexham, corbridge,
nottingham, boots dept store, high st	skillsengland, newcastle, leeds
nottindmuseum, nottingham	prudhoe, hexham, corbridge,
nottingham	toon army, newcastle,
visitnotts, nottingham, castle, wollaton hall	graingermarket1, nufcfoodbank
nottingham, notts, nottinghamshire	northernstars, newcastle, technorthHQ
debbiebryanshop, nottingham	newcastle, nufc
ceosleepoutnotts	ringsideworld, newcastle
uk,	allergynurseUK,
nottingham, derbyshire	nufc
uk, kitchen, town house, nottingham, awe_europe, denon_ul	, nefollowers
	nefollowers, newcastle, quayside
uk,	toon army, newcastle,
salutation inn, hounds gate,	graingermarket1, nufcfoodbank
midlands_news,	, geordies, nufc, toon
midlands_news,	newcastle boys industrial school
ave_uk, the kitchen, town house, nottingham, awe_europe, denonuk	northeast, newcastle
	libertyquay, libertylivingUK,
nottingham	newcastle, nuaccomodation
uk,	newcastle
, nottinghamhistory	newcastle, italy, london, chamonix,
, notts, nottinghamshire	scotland, oman, gallery, ponteland
courtyard, nottingham	newcastle, england
nottingham, alberthall, bestwoodmvc	newcastle WA, avalon ba
london, mansfield road	newcastle DE
nottingham, coopuk	homelesswomenne, citre centre

Locational Reference from each Corpora	
nottinghasm, boots dept store, high st, nottingham	dontation point, newcastle
nottmcitycouncil, nottingham	newcastle gaol
city, nottingham, notts, nottinghamshire	nufc, newcastle, magpies, toon
westbridgfrdwire, derby, nottingham, leicester	newcastle, patagonia
nottingham general hospital, nottinghamhistory	newcastle, patagonia
nottingham, the meeting room	quayside, newcastle, northeast, visitengland, newcastlene1, bbcnewcastle
westbridgfrdwire, derby, nottingham, leicester	asian, newcastle
nottingham castle, westbridgford	newcastle, tynesidecinema, bykertrust
nottingham, notts, nottinghamshire	, nefollowers
wollaton, wollatonpark, nottingham, lake	newcastle, patagonia
sp_bulwell, nottingham, UKSOPRO	ringsideworld,
sp_bulwell, nottingham, UKSOPRO	ringsideworld, newcastle
leicester, nottingham	newcastle
nottingham, nuhneonatal	backtoworkblues, at the tavern,
brazillian, nottingham	, england, everton
derby, nottingham, leicester	south sheilds,
courtyard, nottingham	, england, everton
melbourne, derbyshire	, patagonia
, townsville, melbourne	uniofnewcastle,
, boots dept store, high st	newcastle
your_portland, nottingham	gb, mancity, manutd, chelsea, newcastle
nottingham conference centre	quayside, newcastle, northeast, visitengland, newcastlene1, bbcnewcastle
sp_bulwell, nottingham, UKSOPRO	newcastle, haymarket station
jury's inn, london rd, nottingham,	dinnington area, newcastle
nottsbirders, nottswildlife, nottsbirding	newcastle,nufc, cpfc, epl
britishtinnitus,	ukcc, newcastle, sheffield
, travelodge	ngwrestlinguk,
ukhotels, de vere orchard hotel, nottingham university	, toon
	, central station metro station

Locational Reference from each Corpora	
jury's inn, london rd, nottingham,	nufc,newcastle
nottsbirders, nottswildlife, nottsbirding	newcastle boys industrial school
nottingham	newcastle
chuc_org_uk, ukhotels, de vere	quayside, newcastle, northeast,
orchard hotel, nottingham university	visitengland, newcastlene1,
campus	bbcnewcastle
justicemuseum, nottingham	england, nufc. Newcastle
britishtinnitus, nottingham	newcastle, manchester city
justicemuseum,	allergynurseUK, newcastle
ribaeastmidland,	overseas, newcastle, uk
chuc_org_uk, ukhotels, de vere	london, newcastle, manchester
orchard hotel, nottingham university	graingermarket crepe bar
campus	newcastle, italy, london, chamonix,
nottingham conference centre	scotland, oman, gallery, ponteland
nottmpartners,	newcastle, italy, nyc, london,
britishtinnitus,	chamonix, scotland, oman gallery,
westbridgford,	ponteland
workfromhome	newcastle, italy, london, chamonix,
ukhotels, de vere orchard hotel,	scotland, oman, gallery, ponteland
nottingham university	pknewcastle, system ncl, newcastle
nottmpartners, nottingham	london, redhill
midshealthpsych, autismeastmids,	london, redhill
nottingham	london, redhill
nottingham	london, redhill
sp_bulwell, nottingham, UKSOPRO	systemncl
nottingham	vocationbrewery, indian,
mansfield, nottingham, derby,	, ontariobakery
eastmidlands, notts	geordies, nufc, toon
britishtinnitus, nottingham	bloodwise_north, newcastle
westbridgford, nottingham	london, newcastle, manchester
midlands_news,	london, newcastle, manchester
ribaeastmidland, nottingham	newcastlejobs, newcastle
risley_hall, wedding venue,	newcastlejobs, newcastle, newcastle
nottingham	upon tyne
cockandhooppub, lovenotts,	nsw,

Locational Reference from each Corpora	
itsinnottingham, nottingham, lacemarkethotel	newcastlejobs, newcastle, newcastle upon tyne
bottega, sheffield, nottingham	newcastlejobs, newcastle
sherwood	newcastlejobs, newcastle, newcastle upon tyne
nottinghasm, boots dept store, high st,	newcastlejobs,
nottingham, city, cardiff	newcastle, jesmond, nefollowers
amaya_uk, nottingham	barsinnewcastle, newcastle, uk
westbridgford, nottingham	none
sherwood, mansfield road,	barsinnewcastle, uk
salutation inn, hounds gate,	penguinrhuk, bristol,
autismeastmids, nottingham	wall of fame, newcastle
britishtinnitus,	veggiemumnorth, in the north,
selstoncosmeticclinic	toon, newcastle
visitnotts, nottinghamshire,	newcastle, tyne
nottingham, lovenotts	tyne bridge, the side, newwcastle,
nottindmuseum, nottingham,	northeast
wollatonhall, nottsrocks, lovenott	london, kalgoorlie, newcastle
investinnottm, nottingham, uk	tyne bridge, the side, northeast
sixbarreldrafthouse, nottingham,	northeasthour, newcastle, jesmond
hockley, hockleypub	, indian, nufc
sixbarreldrafthouse, nottingham,	harrogate, york, alnwick, durham
hockley, hockleypub	italian, quayside, newcastle,
nottingham, alberthall, bestwoodmvc	newcastlequays, northeasttweets
salutation inn, hounds gate,	newcastle
salutation inn, hounds gate,	in the salon, newcastle
great northern railway, london road	nlawgrad, newcastle
imtdubaicampus, dubai, nottingham	newcastle boys industrial school
ntu	ouseburncoffee,
nottingham, nottm_events, visitnotts	, uk, toon
no1tingham, nottingham, broadway	, geordies, nufc, toon
cinema, lovenotts	delaware,
embankmentnottm, westbridgford,	, nsw
visitnotts, european capital of culture,	geordie, toonlegend, newcastle,
nottingham2023	geordiemoments

Locational Reference from each Corpora	
animalsbelize, , uk	, brownsnewcastle, uk, uk
animalsbelize, , uk	london, birmingham, glasgow
animalsbelize, , uk	london, birmingham, glasgow
, notts, nottinghamshire	newcastle
nottinghasm, boots dept store, high st,	ne1bid, newcastle
ilsenglish, nottingham, forest recreation ground	london, york, birmingham, glasgow, newcastle, manchester, england, wales, scotland
basecamp chilwell, outdoors, nottingham, derby	newcastle
alexcoxnotts, safenottm, asbnottingham,	newcastle, newcastle united
the southbank bar, nottingham city,	newcastle, carlisle
nottingham university	northeasthour, newcastle, jesmond
bbullblidworth, nottingham, blackbullblidworth	northeasthour, newcastle, jesmond
mansfield, nottingham, newark	northeasthour, newcastle, jesmond
nottinghasm, boots dept store, high st, nottingham	northeasthour, newcastle, jesmond
nottinghasm, boots dept store, high st, nottingham	newcastle, jesmod
, aspley	studentsncl, loyolaatncl, newcastle
nottinghasm, boots dept store, high st,	newcastle, citypool, turkishbaths
mediahutuk, uk,	newcastle east
bridgford house, notts,	gosforth, newcastle
boots dept store, high st,	ausvbra, newcastle
chilwell bar, berliner,	newcastle
fareshare uk,	uk, london, glasgow, bristol, birmingham, bolton, preston
bridgford house, notts,	sligo, tipperary, mullingar,
lovewellbeinguk, notts,	tottenham, crystal palace,
nottinghamshire, , retford, chesterfield	tottenham, crystal palace,
lovelocal, nottingham	spurs, newcastle
nottingham	loyolaatncl, newcastle
under the bridge, canal, waterway,	newcastle
	london, birmingham, newcastle, nottingham, staffordshire, high wycombe
	newcastle, geordies, nufc, toon

Locational Reference from each Corpora	
nottingham, bridge, towpath	newcastle, nsw
notts business park,	newcastle, nsw
, nottmcityoflit	newcastle, nsw
the ncc,	dortmund, newcastle, barkley
nffc, trent, landscape, city ground	newcastle, jesmond
nottingham, cqnottingham	newcastle, northeast
bbcwtherwatchers, nottingham	dortmund, newcastle, barkley
perth, nottingham, sydney	newcastle
nottingham	A525 keele road,
trentuni, nottingham, ntu	northeasthour,
, townsville, melbourne	restaurant,
german,	north,
spanish,	nsw,
oxjamnotts,	nsw,
spanish,	, perth
elbecgb, , brookfieldsgardencentre	newcastlejobs, newcastle
oxjamnotts,	newcastle, oxford
nottingham	, nefollowers
acna,	Newcastle quayside, newcastle,
filthysnotts,	northeast, visitengland,
bulwell, kens auto spares,	newcastlene1, bbcnewcastle
nottingham, ng6	quayside, newcastle, northeast,
nottingham station, station,	visitengland, newcastlene1,
nottingham castle, nottingham	bbcnewcastle
walvinforest, nffc, nffcfamily,	newcastle, jesmondlocal
nottingham, city	quayside, newcastle, northeast,
acceleratenttm, accelerate places, the	visitengland, newcastlene1,
best place,	bbcnewcastle
yeoldetripnott, locale,	quayside, newcastle, northeast,
alanbentleynott,	visitengland, newcastlene1,
lakesbeauty, wilford,	bbcnewcastle
nottstherapy,	newcastle, nebloggers
nottingham	ausvbra, newcastle
rescuerooms,	newcastle
nottingham, townsville, melbourne	newcastle, oxford

Locational Reference from each Corpora	
, townsville, melbourne	ausvbra,
nintendo_notts,	london, manchester, liverpool,
panthers, saltboxbar,	birmingham, scotland, wales,
, irishpub	england, uk,
nottsgreenparty, nottmgreenfest,	brighton, geordies, nufc, toon
nottingham	thebeaconne4,
nottingham, bristol	highwestjesmond, jesmond,
bbcwestminhour	newcastle, gosforth
lebistropierre, nottingham, nottsrba	, gateshead, sunderland
, townsville, melbourne	nsw, newcastle
accountspointuk, derby, , leicester	northe east, ouseburn, newcastle
trendsnotts, valencia,	ounlawschool, newcastle
exchangeng1, nottingham	newcastle, perth
accountspointuk, derby, , leicester	mufc, afc, edinburgh, liverpool,
nott_win, loughborough, leicester, ,	mufc, afc, edinburgh, liverpool,
derby	bluebirdne,
, notts, entrepreneurntu	newcastle, liverpool, manchester,
nottingham city	brighton
nott_win, loughborough, leicester,	, perth
nottingham, derby	northeast, jesmond, gosforth
nottmoffices, small rooms, creative	northeast,
quarter, nottingham	, high west jesmond, gosforth
kch_gardensq, nottingham	newcastle, liverpool manchester,
nottingham	brighton
derby, nottingham, uksopro	o2academynewc, newcastle
westbridgfordol, escabechenotts,	newcastle, southafrica
westbridgford,	bbcradiostoke, newcastle
apothecary,	newcastle
bricnottingham,	paradisone1, newcastle, ne1,
britishtinnitus,	northeastli1, nefollowers
trentevents, westbridgford,	crystalpalace, newcastle, psg
britishtinnitus,	paradisone1, newcastle, ne1,
britishtinnitus, , nffc, wolves	northeastli1, nefollowers
, station	crystalpalace, newcastle, psg
eastmidshour, eastmidshour,	jesmond dene house hotel,

Locational Reference from each Corpora	
leicester, , derby, northants, lincolnshire	, geordies, nufc, toon
notts, , nottinghamshire	, geordies, nufc, toon
investinnottm, , uk	nufc, newcastle
ukhotels, de vere orchard hotel, nottingham university	newcastle, around the town
nottmgreenfest, , vegannottingham, nottinghamgreenfest	tyne theatre, theatre
nottingham conference centre	, newcastle herald
technottingham, , hackconeu	, newcastleunited, nufc, antwerp
nottmgreenfest, , vegannottingham, nottinghamgreenfest	elswick works, , sunderland
brithoptv,	newcastle, sunderland
leicester, , melton, syston	newcastle, hotels, bars, clubs
robinhoodtstatue, , robin_hood_statue	, tyne
crosskeysnottm, panthersihc, cardiffdevils,	bbcradiostoke,
adcnotts, notts ladbrokes, nottingham, nottinghamshire	sage_gateshead, streetfoodfestival, nebloggers, nefollowers
nottsgreenparty, nottmgreenfest, nottingham	newcastle, australia, new south wales
, goosedale	newcastle, australia, new south wales
, nottinghamhenparty,	, nufc, efc, cpfc, psg, epl
notts ladbrokes, , notts, nottinghamshire	gateshead, , streetfoodfestival
rushcliffe halt, east leake,	brighton,
gcrnottingham, Bristol, alberthall, bestwoodmvc	, oxford
leicester, luton,	, geordies, nufc, toon
chuc_org_uk, ukhotels, de vere orchard hotel, nottingham university campus	nsw, newcastle
rock_city_notts	newcastle
leicester, luton,	ciminfo_north, newcastle
nottslocalnews, bridgford, gardens,	newcastle
	the chill room, newcastle, northeasthour
	newcastle, sunshinecoast, ulmarra
	tottenham,
	british_airways, alnwick,
	british_airways, alnwick,
	, geordies, nufc, toon
	, oxford

Locational Reference from each Corpora	
westbridgford, localnews	, gorge
nottslocalnews, bridgford, gardens, westbridgford, nottingham, localnews	brighton, ,brighton hove albion, newcastle united
nottslocalnews, bridgford, gardens, westbridgford, , localnews, notts, notts, notts, notts, notts, notts	, jesmond northeasthour, newcastle, jesmond northeasthour, , jesmond
notts, investinnottm, uk	newcastle united,
investinnottm, uk	northeasthour, , jesmond
dolcinouk,	northeasthour, , jesmond
battle of britain, townsville, melbourne	northeasthour, , jesmond
thieves wood, nottighamshire,	, jesmond
notts, , nottinghamshire	melbourne, , sydney
, townsville, melbourne	merewetheroceanbaths,
the midlands, mecca, , nffc, ncfc	merewetherbeach,
, eastmid	ewcastle, geordies, nufc, toon
flawforth lane ruddington, west bridgford,	, nsw , oxford
the arboretum,	, jesmond
spain, westbridgford,	northeasthour, , jesmond
caves under nottingham, mortimershole, nottinghamcaves,	liverpool, everton, newcastle, newcastlevliverpool,
britain, stmarysnotts,	liverpoolvsnewcastle
caves under nottingham, mortimershole, nottinghamcaves,	texan,
nottinghamcastle,	, england , geordies, nufc, toon
nottinghamcastle,	novacastrians,
places to go on a date,	melbourne,
african caribbean,	australian, , nsw
, notts, nottinghamshire	northeasthour, , jesmond
nottingham, townsville, melbourne	ringsideworld,
nottingham, townsville, melbourne	british,
derbyphotos,	manchestercity, manchestrunited,
arboretum,	chelsea,
uk, oldest inn in england, yeoldetriptojerusalem	nufc, northeasthour, newcastle, jesmond

Locational Reference from each Corpora	
nottinghamshire, nottingham, uk	, geordies, nufc, toon
pitcherandpiano, nottingham, church, pub	, geordies, nufc, toon
local, derby,	travinoswarnersbay, warnersbay
, uk-wide	newcastle, nehour
byrusdale, old basford, home of robin hood, old_basford,	on the lake, newcastle, belmont, lakemacquarie, nsw, aus, australia
nottingham, goosedale	newcastle united, newcastle, nufc
local, derby, nottingham	manchestercity, manchestrunited, chelsea, newcastle, tottenham
nottingham	newcastle, quayside
ruchcliffe halt, east leake, gcrnottingham,	nefollowers, nufc, europeans, nsw, ,
, alberthall, bestwoodmvc	europeans, nsw, ,
ruchcliffe halt, east leake, gcrnottingham,	newcastle, geordies, nufc, toon broadmeadow, sydney,
ruchcliffe halt, east leake, gcrnottingham,	, nsw liverpool, belfast,
guest/kid's bathrooms,	nefollowers, nufc,
ruchcliffe halt, east leake, gcrnottingham,	nefollowers, nufc, , oxford
guest/kid's bathrooms, nottingham	, taiwan
rivertrent, nottingham	, northeast,
ruchcliffe halt, east leake, gcrnottingham, nottingham	liverpool, belfast, newcastle liverpool, belfast, newcastle
city centre, nottingham	nefollowers, newcastle, sunderland
notts, nottingham, nottinghamshire	nefollowers, newcastle, sunderland
peacocknotts, nottingham	nefollowers, newcastle, sunderland
nottingham, bristol	liverpool, belfast, newcastle
westbridgford	liverpool, belfast, newcastle
, old market square	newcastle, geordies, nufc, toon
peacocknotts, nottingham train station, , nottinghamshire, notts	epl, mancity, manunited, chelsea, newcastle, tottenham, huddersfield liverpool, belfast, newcastle
notts, nottinghamshire	dortmund, newcastle, barkley
, alberthall, bestwoodmvc	liverpool, belfast, newcastle

Locational Reference from each Corpora	
red hat chinese takeaway, beechdale	dortmund, newcastle, barkley
, goosedale	newcastle, nsw
, london	newcastle, uk, britaingb
, goosedale	ewcastlequays, newcastle
, goosedale	hd_newcastle, quayside, newcastle
, goosedale	newcastle, quayside,
laboursherwood, nottingham	milleniumbridge, rivertyne, quayside,
notts university	newcastle upon tyne
laboursherwood, nottingham	nswlabor, newcastle
nottingham conference centre	newcastle, britishescorts
blank	newcastle, ncl, town
derby,	sydney, nsw, newcastle
nottmgreenfest, vegannottingham,	madeinnewcastle, ouseburn,
nottinghamgreenfest	newcastle, geordies, nufc, toon
sheffield, nottsli	newcastle, merewether
stmarysnotts,	paradisone1, newcastle, ne1,
restaurant, notts	northeastli1, nefollowers, indealsncl
city,	newcastle, nefollowers
rescuerooms	tyne bridge, newcastle, north east
nottmmela, jaipur	newcastle, tyne
nottmmela, Jaipur, Bristol,	london, manchester, liverpool,
nottsgreenparty, nottmgreenfest,	birmingham, scotland, wales, leeds,
oxjamnotts,	edinburgh, cardiff
chinese, mapperly, location	birmingham, oxford road,
nottingham clinic, qpark_uk	manchester, gipsy lane, newcastle,
bristol	george street
, bristol	newcastle, britishescorts
oxjamnotts,	liverpool, belfast,
, uk-wide	newcastlesolution, , newcastlesong
newark,	, oxford
, qpark_uk, notts	newcastle, nufc
acna,	dankofengland, newcastle
nottm_contemp, states of america	newcastle, st jame's park, eldon
nottmgreenfest, nottingham,	square
vegannottingham,	dankofengland,

Locational Reference from each Corpora	
nottinghamgreenfest	dankofengland,
nottmgreenfest, nottingham,	newcastle, nufc
vegannottingham,	malmaison, newcastle,
nottinghamgreenfest	milleniumbridge
nottmgreenfest, nottingham,	newcastle, geordies, nufc, toon
vegannottingham,	liverpool, belfast, newcastle
nottinghamgreenfest	thebeaconne4, newcastle
bbc, nottingham	nsw, newcastle
, uk-wide	newcastle, o2academynewc
, masterchefUK	millenium bridge, tyne, newcastle,
nottingham, bristol	gateshead, baltic
nottindmuseum, nottingham	millenium bridge, tyne, newcastle,
nottslocalnews, county council,	gateshead, baltic
newark, nottingham, localnews	, tyne
county council, newark, nottingham,	newcastle united, nufc, toon,
localnews	, tyne
leicester, syston	, tyne
, notts	, tyne
, notts	, tyne
, notts	newcastlejobs,
, notts	brighton,
, notts	newcastle united, nufc, toon,
, notts	liverpool, belfast
, notts	
KSFLUK, juicebar,	
, bristol	
neet,	
westbridgfdwire, , westbridgeford,	
beeston	
nottsfinn, brighton,	
, westbridgford, beeston	
, manchester	
leicester, loughborough,	
cardiffdevils, nottingham	
nottsfinn, brighton,	

Locational Reference from each Corpora
eastmidshour, leicester, nottingham,
derby, northants, lincolnshire,
eastmidlands
cardiffdevils,
npasbirmingham,
npasbirmingham,
cardiffdevils,
cardiffdevils,