English Language and History: Geographical Representations of Poverty in Historical Newspapers

Book Section

How to cite:


For guidance on citations see FAQs.

© 2019 Ian N. Gregory; 2019 Laura L. Paterson

Version: Accepted Manuscript

Link(s) to article on publisher’s website:

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
English Language and History: Geographical representations of poverty in Historical Newspapers

Ian N. Gregory and Laura L Paterson

INTRODUCTION

The use of computational approaches in history is not new (Boonstra et al 2004). However, until fairly recently, their use has been restricted to a relatively small number of fields such as economic history and historical demography. These are fields that make extensive use of quantitative sources, typically in tabular form, that have long been well-suited to analysis using a computer. Despite the potential benefits of quantitative sources, the overwhelming majority of historians use textual sources and thus, computational approaches have traditionally had little penetration into mainstream history. Over the past decade or so this has changed rapidly. The mass digitisation and dissemination of historical source material such as Early English Books Online (EEBO), Eighteenth Century Collections Online (ECCO), the British Library Newspaper Collections, The Times Digital Archive, Project Gutenberg, and many others have meant that large amounts of historical texts are now available in digital form (Hitchcock 2013). Thus, historians now have access to an unprecedented volume of digitised source material. However, the computational techniques used to analyse such materials remain rooted in traditional close reading with the digital simply providing ease of access and the use of keyword searching. There is thus a requirement to adopt and adapt the digital approaches available from other fields, and develop new techniques where appropriate, such that scale of the resources available to us can be
exploited in a way that is impossible for close reading alone. This challenge is not unique to history; it has occurred across humanities disciples, providing a major impetus to the field of digital humanities (Kirschenbaum 2010; Schreibman et al. 2014; Terras 2016) where a major theme is the use of digital approaches to the analysis of texts in disciplines including English Literature (Moretti 2013; Jockers 2013) and History (Atkinson & Gregory, 2017; Pumfrey et al. 2012).

One of the key, and under-represented, opportunities within digital humanities is the ability to bring together methods, approaches, and topics from a range of disciplines and use them to conducted genuinely interdisciplinary research. In this chapter we bring together methods from corpus linguistics and geographical information science to conduct a piece of analysis relevant to both historians and geographers. We provide a case study of how UK place-names are used in relation to the term poverty in The Times newspaper between 1940 and 2010. We use a new technique, Geographical Text Analysis (Donaldson et al. 2017; Gregory & Donaldson 2016; Porter et al. 2015), to demonstrate how corpora – and particularly the corpus linguistics tool of concordance – can be used in combination with methods from geographical information science (GISc) to foreground, visualise, and analyse explicit mentions of place within texts. (For good introductions to GISc and the closely related field of geographical information systems (GIS) see, for example: Chrisman 2002; Heywood et al. 2006; Longley et al. 2015.) As such, we present an innovative approach to the study of long-term geographical change from very large digital collections that could be applied in many other contexts.

BACKGROUND: THE USE OF NEWSPAPER CORPORA
Some of the largest corpora available to historians and the academic community more broadly come from newspapers. Part I of the British Library’s 19th Century Newspaper collection on its own contains, where available, continuous runs for the whole century of forty-seven newspapers from across Britain and Ireland (see: https://www.gale.com/uk/c/british-library-newspapers-part-i). It consists of almost two million pages and at least thirty billion words of text. Historians have realised the potential value of these sources and accepted the need for new approaches to their study (Mussell 2012; Nicholson 2010). These approaches particularly focus on the use of corpus linguistics to allow a combination of quantitative and qualitative explorations of the language used within the text (Adolphs 2006; McEnery & Hardie 2011). A simple example of this is provided by Nicholson (2012) who explores the changing frequency of words such as ‘America’ and ‘Germany’ and other words that they are associated with these countries using the British Library Nineteenth Century Newspaper collection. These approaches have also been adopted by scholars exploring modern newspapers to research, for example, themes associated with minority groups (Baker et al. 2014), immigration (Gabrielatos & Baker 2008), and political correctness (Johnson et al 2003).

There are two recurring issues in the use of historical newspapers within historical research. From an applied perspective, all newspapers have political, editorial, and other biases which may change over time (Brake & Demoor 2009). The way that they represent the word is thus not impartial. Instead, exploring the biases in the themes and places that newspapers cover, and how these biases were represented, becomes an area of active research interest (Hardie & McEnery 2009). The second, more pragmatic one, is the issue of OCR (optical character recognition) errors. Bulk digitisation of texts is most cost-effectively conducted by first scanning the page and then using OCR software to convert the pixels on
the image into the alpha-numeric characters that a computer can understand as text. The difficulty is that, while quick, OCR is an error-prone process and a significant number of words are likely to have one or more characters that have been incorrectly identified by the software. This, in turn, means that a computer will not recognise the word in, for example, a keyword search. Historic newspapers are particularly prone to errors of this sort as the quality of both the paper and the printing is likely to be relatively poor (Holley 2009). Using the British Library Nineteenth Century Newspaper Collection, Tanner et al. (2009) estimated that only 83.6% of characters were accurate leading to a word accuracy of only 78.0% which declined to 63.4% for words that started with a capital letter. The impact of these errors has led to significant criticism (Hitchcock 2013; Putman 2016). However, detailed research suggests that, from a corpus linguistics perspective, poor OCR may not be as significant an issue as has been suggested (Joulain-Jay, 2017).

In this study, we explore *The Times* from the 1940s to the 2000s. *The Times* is widely regarded as Britain’s newspaper of record. It has been published since 1785 making it the longest continuously-published newspaper in the world. It has always been published in London and is generally regarded as being aimed at an educated, middle class market. The newspaper has been digitised for the period from 1785-2012 as *The Times Digital Archive* (see: https://www.gale.com/uk/c/the-times-digital-archive). We have access to the underlying text, hosted in CQPweb (Hardie 2012) at Lancaster University, under licence from Gale Cengage.

**ISSUES AND RESEARCH IN THE STUDY OF POVERTY**

The study of the geographies associated with poverty has a long history stretching back at least to Charles Booth’s maps of poverty in Victorian London (see:
https://booth.lse.ac.uk). Much geographical research on poverty has been quantitative, based on identifying poverty and inequality from sources such as the census or household surveys (Dorling 2013; Thomas & Dorling 2011; Townsend 1979). There have also been a number of studies which have explored changing geographical patterns of poverty over time, in London (Dorling et al. 2000), England & Wales (Gregory 2009), and Britain as a whole (Dorling 1997). Working at different scales, these studies have all shown that patterns of poverty are very strongly entrenched. With very few exceptions, poor areas remain poor relative to other places, while rich ones remain relatively rich. These studies are all based on quantitative representations of how unequal or impoverished areas are. Poverty is, however, a notoriously difficult concept to define and quantitative measures, such as unemployment, low-skilled work, and overcrowded housing, are well known to be surrogates for what is a complex and multi-faceted phenomenon.

In this chapter, we are interested not in poverty as defined quantitatively, but instead as it is represented in textual sources, namely newspaper corpora. Our particular interest is in whether the temporal stability which geographical studies have found in quantitative poverty are also found in newspaper representations of impoverished places. In other words, we want to find out which places The Times newspaper associates with poverty and investigate whether these places remain constant over time. To this end, we trace the relationship between the use of place-names and the term poverty in The Times from the 1940s to the start of the twenty first century. Whilst we have chosen to demonstrate our method by focusing on a single corpus query – we use <*>poverty*>\(^1\) to locate terms such as poverty, anti-poverty, poverty-stricken, etc. – larger scale and more robust Geographical Text

\(^1\) The asterisk in the query term denotes ‘zero or more characters’, so this particular query will return instances of poverty, poverty-laden, etc.
Analysis depends on the analysis of a larger number of search terms to build a picture of how place is used discursively in relation to a given topic. This large-scale Geographical Text Analysis is exemplified by Paterson and Gregory’s (2018) analysis of discourses of poverty and place in the Guardian and Daily Mail newspapers between 2010 and 2015.

For the present chapter, we use seven decades from The Times: the 1940s, 1950s, 1960s, 1970s, 1980s, 1990s, and 2000s. In interrogating how place is used during different periods in British history, we can trace how notions of poverty move in space as well as in ideological, social, political, and cultural realms. For example, our 1940s corpus covers the end of the Second World War, rationing, and austerity, as well as the founding of the NHS and the beginnings of the welfare state. In 1950 Joseph Rowntree published his third report on poverty, which suggested that post-war poverty rates had declined dramatically, although reanalysis has indicated that this was an overstatement of the case (c.f. Piachaud and Webb 2004:45). At the end of the 1950s, in 1957, Prime Minister Harold Macmillan claimed that most Britons had ‘never had it so good’, but the apparent prosperity of the decade was followed by wage freezes in the 1960s and rising inflation rates. Piachaud and Webb (2004:50) note that poverty was ‘rediscovered in the 1960s’, a decade which included the founding of the Child Poverty Action Group. Poverty rates decreased in the 1970s but by the end of the decade they were set to rise dramatically. The 1980s corpus represents The Times’ response to Thatcherism and deindustrialisation, a period of time when poverty rates continued to rise. The first Breadline Britain survey was conducted in this decade (Mack and Lansley 1985), which estimated that 7.5 million people in Britain were living in relative poverty (1985:182). There was a recession in the early 1990s, but in the years that followed there was decreasing unemployment and the introduction of the minimum wage in 1999. Finally, the 2000s corpus
includes initial economic prosperity followed by the downturn of the global economic crash in 2008.

All of these historical events and measurements have a relationship to poverty in both its economic and social conceptualisations. As a crude measure of media interest in poverty we can compare how often the term is used across our seven corpora (see Table 1, below). Mentions of poverty decline from a normalised frequency of 6.09 tokens per million words (pmw) in the 1940s to 5.30 pwn in the 1950s. There is an increase in mentions throughout the 1960s and 1970s with a plateau in the 1980s corpus. Mentions of poverty decline in the 1990s before reaching their peak in the 2000s corpus. However, far from prompting conclusions about poverty in Britain during this time span, these figures do not provide information about a) what The Times was saying about poverty in each of these decades, nor b) whether the poverty written about was geographically located. The 1980s corpus, for example, contains reports about the Ethiopian famine alongside references to London and Philadelphia – the locations of the Live Aid charity concerts held to raise funds for famine-relief in 1985. Thus, the poverty referenced in these cases is not located within the UK despite the use of the UK place-name 'London'.

[Table 1: Hits for <*poverty*> in The Times corpora]

In order to understand which locations The Times associated with poverty (in the UK or elsewhere) we can turn to Geographical Text Analysis. What GTA can tell us is how, or indeed whether, place was significant in The Times’ representations of poverty during each of our chosen decades. Specifically, we pose the following research questions:
1. Which UK places are associated with poverty by *The Times* and are these associations stable over time?

2. Why are particular places associated with poverty?

3. How is the relationship between poverty and place characterised in each decade?

**RESEARCH METHODS: GEOGRAPHICAL TEXT ANALYSIS**

Geographical Text Analysis involves combining methods from corpus linguistics and GISc. For the uninitiated, corpus linguistics involves the computationally-aided analysis of large bodies of texts. It facilitates the analysis of language at a scale beyond that possible using manual analysis alone and many currently-available corpora, especially those which have been constructed using web crawling technology, contain billions of words. The corpora we are using here contain a combined total of 3,838,432,279 words split across seven decades (Table 1). Corpus linguistics tools include the calculation of word frequency lists, keywords (those words which occur at a statistically higher rate in a given corpus compared to a reference corpus), collocations (the relationship between words), and concordances (all examples of a given query term in its immediate co-text) (see, for example, Adolphs 2006; McEnery & Hardie 2012). GTA draws particularly on concordances – the case study presented here uses a form of geoparsing known as *concordance geoparsing* – but this is not the only corpus tool that can be useful for GTA. As an example, Paterson and Gregory (2018) used collocates and contrastive collocation analysis (see below) to illuminate trends in discourses of poverty which illustrated traces of underlying ideologies, such as flawed consumerism (Bauman 2005) and the deserving/underserving poor (Katz 2013).

The tools of corpus analysis are often used in triangulation with other methods of linguistic analysis. For example, there is an established tradition of combining corpus analysis
with discourse analysis (see Baker 2006). Increasingly too, corpora are being used beyond the realms of linguistics within the wider social sciences and the humanities (see http://cass.lancs.ac.uk/ for examples of such initiatives). By drawing on geographical methods of analysis, this chapter is an example of the expansion of corpus linguistics tools beyond the boundaries of the discipline of linguistics. In particular, we combine corpus linguistics with methods from GISc to implement a GTA. This requires us first to convert the unstructured text from a corpus into a structure that can be used within GIS, and then to use techniques from both GISc and corpus linguistics to explore the geographies found within the corpus.

A major challenge for GTA is thus how to convert unstructured text into the tabular form that GIS software requires. This is implemented by making what we term the place-name co-occurrence (PNC) the basic unit of analysis within GTA. PNCs consists of the query term, the place-name, the co-text that surrounds them, and the point location (the geographical coordinates) that can be used to map that place-name. As shown in Figure 1, it is possible for a single concordance line to produce multiple PNCs; there are two occurrences of ‘London’ within the 15 word span of the node ‘poverty’, and so these represent two separate PNCs. For examples of studies in the spatial humanities using the concept of PNCs see Gregory & Donaldson (2016) and Donaldson et al. (2017). Analysis using PNCs assumes that co-occurrence of a query node and a place-name is indicative of a relationship between the two. At its most basic, this relationship can be expressed in the Firthian sense that ‘you shall know a word by the company it keeps’ (Firth 1957: 11). Murrieta-Flores et al. (2015: 300) note that proximity of place-name mentions to a query node does not guarantee this link, nevertheless, they come to the conclusion that work on collocations ‘seems to suggest that simple proximity, with no explicit connection, if repeated consistently, is indeed enough to create an implicit link between two meanings in the mind of speakers and hearers’.
Creating PNCs from large corpora requires a tool known as a geoparser. A geoparser first identifies likely place-names using named entity recognition (NER) techniques and then uses a gazetteer to assign geographical coordinates to all occurrences of place-names (Grover et al. 2010). Place-names are, however, complicated things that are difficult to process in an entirely automated manner. To raise standards of accuracy in both identifying place-names and in allocating them to an appropriate co-ordinate, we use a process called concordance geoparsing in which only the immediate co-text around the search-term is geoparsed allowing for easier error checking and correcting (Rupp et al. 2014). The nature of concordance geoparsing is such that it effectively returns PNCs as the results include the search-term, place-name, co-ordinate, and surrounding co-text. One critical choice in defining a PNC is setting the maximum span between the search-term and the place-name for the concordance to be included. In this case, we chose a span of 15 words/tokens (punctuation counts as one token) either side of the query node. Whilst we have used a smaller span of 10 words elsewhere (Paterson & Gregory 2018), we chose to increase the span here to account for the fact that there were some OCR (optical character recognition) errors in the 1940s corpora as they had been scanned and converted to a digital format.\(^2\)

\(^2\) As a result of poor OCR and other corpus formatting issues, we were not able to include texts from a small subset of data for each of the following years: 1971, 1983, 1986, 1988, 2000, 2001, 2005-2007.
Figure 2 shows five examples of concordance lines that will generate PNCs using the +/-15 token span. Once the PNCs are generated, they are downloaded as a plain text file (including their metadata, such as their text ID, date of publication, etc.). This process can be repeated for multiple queries, although here we focus on just one query (<poverty*>) to demonstrate the method. The first concordance line in Figure 2 generated two PNCs (one for Liverpool and one for Broadstairs). By contrast, the final example in Figure 2 generates one PNC (London) as we did not analyses the texts down to street level; thus ‘London Bridge Street’ as not considered to be its own PNC. Once the PNCs have been generated, the resulting files were then opened in flat-file database software (in this case Microsoft Excel) to be cleaned.

Cleaning involved, for example, removing erroneous hits from our results (i.e. ‘hull’ referring to the hull of a ship as opposed to Kingston-Upon-Hull in Yorkshire). It may also involve disambiguating between different places with the same name, such as ‘Newcastle’, which could refer to Newcastle-upon-Tyne or Newcastle-under-Lyme. We also removed references to poverty that comprised part of TV and/or radio listings, as well as more figurative uses of poverty, such as a ‘poverty of titles’ which referred to not having suitable music to play, and the description of low-quality rugby (‘poverty play’) and cricket (‘poverty of wickets’). We were concerned with the relationship between poverty and place in the UK only, so we removed references to other countries, such as America and France. Additionally, we removed country-level references, such as Scotland, Wales, etc., in order to focus more closely on local areas within the UK. Once the data was clean, we uploaded the final database files into database management and cartographic software tool ArcGIS to analyse, and ultimately visualise, how poverty and place are used within The Times newspaper from 1940-2010.
SAMPLE ANALYSIS: THE GEOGRAPHY OF POVERTY IN THE TIMES 1940-2010

Table 2 shows the distribution of hits for the query \(<*poverty*>\) (henceforth expressed as \emph{poverty}) by decade. Column 2 of the table suggests that \emph{The Times}' interest in poverty more than doubled between the 1950s, when there were 5.3 hits pmw, and the 1970s with 12.4 hits pmw. Since the 1970s, however, the interest has remained roughly constant. Column 4 shows that overall there is a link between poverty and place. It shows that over the entire period, for every thousand instances of \emph{poverty}, 28.7 of them had a place-name within 15 tokens. In other words, poverty is associated with place to a certain extent. Interestingly, unlike the overall instances of \emph{poverty} in the corpora, there does not seem to be a clear pattern of change over time in the link between poverty and place. This suggests that whatever caused interest in poverty to rise over time, it was not directly linked to geography, at least not geography as revealed by the explicit use of place-names. Column 5 shows that when normalised by the overall size of the relevant corpus, interest in poverty and place, as revealed by \emph{poverty} PNCs, rose through the 1950s and 1960s to peak in the 1980s before dropping back. Poverty was thus far more closely associated with place in the 1980s than in other decades. The 1940s are also interesting, but also exceptional; although there was a relatively low interest in poverty overall in the 1940s, it appears to be far more associated with geography than any other decade.

[Table 2: \(<*poverty*>\) by decade]

This does not, of course, tell us anything about where \emph{The Times} associates with poverty, but rather that it links poverty and place to a degree that varies over time. However,
as the geoparsing process provides us with a coordinate for each place-name within a PNC, we can conduct spatial analysis to explore the geographies of poverty and place. The coordinates form spatial data that allow us to read the PNC into GIS. The easy way to map this data would then be to use a dot map where each coordinate is represented by a point.

[Figure 3a and b 1940a and 1980s dot maps]

Figure 3 shows the pattern of PNCs in the 1940s and the 1980s. It shows that in the 1940s poverty instances are only found in London and Newcastle, indeed almost all of these are in London. By the 1980s the pattern is much more dispersed, particularly across England. This type of mapping is not, however, a particularly effective solution for visualising the PNCs. Dot maps are difficult for the human eye to interpret, particularly as large numbers of points may be at exactly the same location and thus disappear under each other, as in Figure 3a in particular. For this reason, Figure 4, which shows all of the poverty PNCs in The Times corpora, also uses a process called density smoothing to simplify and summarise the pattern. Density smoothing effectively measures the number of point locations, in this case PNCs, which are near to each location on the study area (Lloyd 2007). The more PNCs there are near the location, and the nearer they are, the higher the value at that location. A critical factor in implementing this is how we define ‘near’. This is done using a bandwidth that means that the weighting given to each PNC declines with its distance from the location. The choice of bandwidth can be termed quantitatively based on the distribution of points around the study area (Fotheringham et al. 2000: 149). To be consistent, and allow comparisons between decades, we have normalised all of the densities using z-scores. A z-score of 0.0 shows the mean density, 1.0 is one standard deviation above the mean, and so on. This allows us to
identify clusters based on the idea that values with a z-score of above 1.96 are significant at 
p<0.05 and those above 2.58 are significant at p<0.01 although, for various statistical reasons, 
these thresholds should be seen as exploratory rather than as formal measures of statistical 
significance.

[Figure 4: Density smoothed map of all poverty instances]

Figure 4 thus highlights clearly that The Times strongly associates poverty with London; 
the only place on the map that lies above the z>1.96 threshold. Of the 1,131 PNCs across all 
corpora, 53.1% are within Greater London suggesting that The Times is far more focussed on 
poverty in London than anywhere else. This figure, in itself, highlights one of the advantages 
of the spatial analytic approach over a conventional corpus linguistic approach: while corpus 
linguistics would be able to show that there were more co-occurrences between, for example 
poverty and ‘London’, than other place-names, it would be unable to automatically include 
named places within and around London, such as Westminster, Tower Hamlets, and Islington 
within this calculation. Liverpool and Manchester, with 41 and 40 PNCs respectively, have 
smaller clusters and there is also raised interest in poverty in many other urban places such 
as Belfast, Glasgow, Edinburgh, Newcastle, and so on.

It would be possible to produce maps similar to Figure 4 for each decade, however, 
the similarities and differences between these maps are difficult to present in paper form, 
particularly, to compare over time. Instead, Figure 5 divides the country into three major 
macro-areas: first, Greater London; second, the South which includes all counties outside of 
Greater London that are south of the arbitrary line running from the Severn Estuary in the 
west to the Wash in the east, which is extensively used to define the UK’s north-south divide;
and third, the North: all counties north of the Severn-Wash line including those in Scotland, Wales and Northern Ireland. The North is then further sub-divided into places in the historical counties of Lancashire and Yorkshire – which include many major industrial cities including Liverpool, Manchester, Leeds, and Sheffield – and ‘Other North’ which is the rest of the UK north of the Severn-Wash line.

[Figure 5: Macro-geographies]

Figure 5 does not include the 1940s as this decade is a strange anomaly – all but one of its 95 PNCs are in London. For other decades, however, the coverage of poverty is dominated by either London and/or the North, whilst the South is consistently not associated with poverty. For most decades, London attracts more coverage than the North, typically over fifty percent of the total PNCs. The exceptions to this are the 1950s and – marginally – the 1980s. In the 1950s this exception is driven by a spike in interest in poverty in Lancashire and Yorkshire. The 1980s appear to be a pivotal decade in interest in poverty in the North: PNCs in Lancashire and Yorkshire rise from being around ten percent of the total to being around 20%. At the same time, interest in poverty elsewhere, in the North, which had been rising steadily since the 1950s, peaks at 24.4% of PNCs, from where it declines in subsequent decades. This combination accounts for instances of poverty in the North narrowly exceeding interest in poverty in London in the 1980s.

This type of spatial analysis is very good at identifying patterns. However, these patterns pose questions to the researcher to then try to explain them. The major questions that emerge from the above are, why is poverty only associated with London in the 1940s? What is it overall about poverty in London after the 1940s that attracts attention? How does
what is said about London compare with other parts of the country? What is it about Lancashire and Yorkshire, in particular, in the 1950s that attracts attention? And, why does interest in Lancashire and Yorkshire increase in the 1980s?

To begin to answer these questions we need to move away from GIS and back to corpus linguistics. There are two approaches to answering these questions: the first is qualitative and simply involves exploring the corpus to perform close analysis of the concordance lines in which the place-names co-occur with the query node within the set span of +/-15 words. The second is more quantitative and involves contrasting the co-text of one set of PNCs with a second set.

When there are relatively small numbers of PNCs, the qualitative approach is usually sufficient on its own. To understand what it was about poverty in London in the 1940s we needed to explore 94 PNCs. 82 of these PNCs refer to ‘London’ itself, the others are places such as Whitechapel, Stepney, Bethnal Green and Westminster. Exploring their concordance lines reveals that the majority are concerned with charity appeals during the War to help people in impoverished parts of London. The concentration of these is perhaps somewhat misleading as some of these appeals are repeated multiple times. For example, an appeal by the South London Mission which included example (1) was repeated eighteen times through November and December 1940.

(1) Surrounded by docks and high tenements, in this largest area of unbroken poverty in all London, the world knows something of what our people in South London are enduring in the Battle for Freedom. Surely this year, in brave hearted South London Father Christmas is needed more than ever. Your Christmas gift will help them hold on until Victory
The other interesting thing is that only nine of the PNCs are in the post-War period and these are again largely associated with charitable appeals. This shows that, in the period while the welfare state was being created, there was little, if any, consideration, at least in *The Times*, of the geographical aspects of poverty, at least as revealed by our single search-term. Exploring the negative – why does *The Times* ignore poverty everywhere else in the country in the 1940s? – is, of course, more difficult.

When there are larger numbers of PNCs a more quantitative approach is required as simply looking at concordance lines is insufficient. To explore the interest in poverty in London in comparison to other places from the 1950s onwards means exploring 507 PNCs in London and 529 elsewhere. The question thus becomes, what is it about PNCs in London that is different for those from elsewhere? To explore the similarities and differences between how poverty is related to place in different geographical locations, we use a specialist type of contrastive concordance analysis. To do this we compared the co-text element of the PNCs for London with the co-text element of the PNCs for the rest of the country in order to identify what words are associated more with one set of PNCs than the other. We call the resulting sets of words PNC keywords.

[Table 3: PNC keywords, London vs not London 1950s onwards]

In Table 3, log-likelihood, one of a range of possible statistics, has been used to identify statistically significant PNC keywords for poverty from the 1950s onward. Close analysis of the London PNC keywords reveals that a lot of poverty charities are based on London (‘group’, ‘child’, ‘action’, ‘against’, ‘Peabody’, and some instances of ‘war’). There is also some discussion of, or comparisons with, poverty in the past or in history (‘Victorian’, some
instances of ‘war’ in senses such as ‘since the war’ or ‘before the war’). ‘Greatest’ is often used in the sense of a comparison with London having both the greatest affluence and the worst poverty (2).

(2) ...while London has both the greatest affluence and also some of the worst pockets of poverty. The North West and London have most districts falling within the country’s 150 most deprived districts.

(3) A mention in your will can bring food and shelter to tiny hungry children and to old people. - WAR ON WANT Campaign Against World Poverty. 9 Madeley Road. London.

Another interesting adjective is ‘genteel’, as ‘genteel poverty’ seems to have been a condition frequently referred to in London. Beyond this, ‘homelessness’ is a key theme and poverty in London also seems to be associated with ‘old’ people (3). Attempts to alleviate poverty are discussed in the sense of ‘Supplementary’ benefit, ‘evidence’ of poverty and the need to ‘deal’ with poverty, and there is also ‘press’ attention.

Beyond the capital, politics seems much more important, particularly in discussions of debates in parliament where politicians representing non-London constituencies are frequently quoted (‘Lab’ [in the sense of the Labour party], ‘MP’, ‘Mr’, some instances of ‘Sir’). There is also considerable discussions about poverty in the letters pages. Letters either discuss poverty occurring outside London or have been written by people outside London (other instances of ‘Sir’, ‘Times’, ‘September’). In example (4) we have two PNCs generated because the letter writer lived in Houghton (Hampshire) but was making a derogatory comment about people from Devon.

(4) Sir, On a recent visit to the library in a small market town in Devon I found a large volume entitled Poverty in the careers section. Yours faithfully, SUSAN ELKS, The Willows, Houghton
Using examples like this we can also begin to see the types of stance taken in *The Times* towards poverty; although (4) is not an example of text authored under the newspaper’s masthead by an in-house journalist, the letter was selected for print by an editorial gatekeeper. Although it is not sufficient to assume that this automatically implies *The Times*’ endorsement of the sentiment expressed in (4), we can at least take this as an example of the types of letters concerning poverty that *The Times* saw fit to print.

There are also more PNC keywords directly associated with poverty than are found in London, for example ‘grown’ up in poverty in a place, the ‘effects’ of poverty or unemployment in a place, the poverty ‘trap’, and ‘primary’ poverty (‘primary’ is also sometimes used in the context of primary schools). Housing reveals itself in a different way to London, as elsewhere poverty is associated with ‘estate(s)’ and also with ‘town(s)’. Poverty outside London is also associated with ‘industrial’ areas or sectors (5). There is little on what can be done about poverty except in some aspects of ‘planning’ and in actions by the ‘church.’

(5) Textile mills cling to the sides of the *Calder* valley, providing an industrial base on which the prosperity or the poverty of the area depends. Rugged, wind-swept hillsides provide a foot-hold for small farms whose occupants, more out of tradition than anything else, tend to lean towards the political right.

Our third question concerned what it was about poverty in Lancashire and Yorkshire that attracted attention in the 1950s. Here qualitative analysis is again sufficient as there are only 73 poverty PNCs for the country as whole in the 1950s. Of these, 22 were in Lancashire and Yorkshire, almost all of which – 17 – were in York. The main reason for this is that 1951 saw the publication of Seebohm Rowntree’s final report on poverty in York *Poverty and the Welfare State* (Rowntree & Lavers, 1951) (the earlier ones had been published in 1901 and
1936) which, as well as attracting interest in its own right, also seems to have made York the example place for poverty in this decade.

The 1980’s interest in Lancashire and Yorkshire is more diverse both in terms of geography and theme. There are 42 PNCs in this region over the decade and they cluster around Liverpool, Manchester, Bradford, and Humberside in particular. The co-text of these PNCs are mainly concerned with levels of poverty in these places, sometimes in relation to the riots that occurred in the early 1980s, especially in Liverpool, with links between poverty and race being another prominent issue that emerges from the 1960s onwards. There are also strong hints about the impact of industrial decline in these areas although poverty is frequently associated with the industrial era as well as the 1980s. An example of this is:

(6) ...when Britain was the workshop of the world and when the ports were full of ships and ship-building, cities like London, Glasgow, Liverpool, Manchester and Newcastle were notorious for their poverty, their overcrowded slums, their terrifying juvenile mortality, crime, prostitution, disease and destitution, as well as for the existence of a savage urban underclass.

It is clear from the above that the geographies of poverty reporting undergo major changes over time, particularly with an increased focus on Lancashire and Yorkshire and a decline in interest elsewhere in the North over the same period. To explore this in more detail we needed to both lower the scale of the investigation to remove, for example, the impact of a single city, such as York in the 1950s, having such a marked influence on broader regions, and to analyse whether these changes are statistically significant. Doing this involves returning to spatial analysis where we can use a technique called spatial segregation analysis which addresses both issues (Diggle et al. 2005; Rowlingson, 2015). Spatial segregation
analysis involves comparing two point patterns, thus avoiding aggregation, to see whether they appear to have statistically significantly different distributions and, if so, to identify where on the map they are different.

Spatial segregation analysis was used in two different analyses: the first compared the 1980s with other decades to test whether the 1980s appeared to be exceptional, the second compared the period before the 1980s with the 1980s and later decades to see whether the 1980s appeared to represent a shift in the pattern. In both cases the 1940s were excluded as they have an exceptional geography.

[Figure 6 – 1980s segregation analysis]

The overall results comparing the 1980s PNCs with other decades shows that the two patterns are not significantly different (p=0.11). Figure 6, however, explores the more local patterns identifying the contour lines where the PNCs in an area are significantly more likely to come either from the 1980s or from other decades. It shows that the north-west of England was significantly more associated with poverty in the 1980s than it was in other decades while the south-east had significantly less in the 1980s, although only at the p<0.05 level. Interestingly, the p=0.50 contour, the level at which there is an even possibility of the PNC being from either decade, runs almost perfectly along the Severn-Wash line. This suggests that the 1980s shows an increased north-south divide in The Times’ reporting of poverty compared to other decades. The more northerly p=0.50 contours, such as the one that runs through northern and eastern Scotland, can safely be discounted as they are based on very small numbers.

[Figure 7 – 1980s onwards segregation analysis]

Figure 7 repeats the segregation analysis but this time compares the decades from the 1950s to the 1970s with the 1980s and later. In this case there is a statistically significant
difference between the two sets of PNCs at 0<0.01. The local pattern, shown in Figure 7, is in many ways similar to the pattern shown in Figure 6. The north-west of England became significantly more associated with poverty after 1980 while areas that become less associated with poverty in the later decades are again in the south and east, albeit this line is now somewhat further north than in Figure 6. The area where interest in poverty has dropped the most, in other words where it was higher in the pre-1980s period, was the Midlands, while there was no statistically significant change in the number of PNCs in London. Overall, these two maps do suggest that the 1980s was a pivotal decade when *The Times*’ interest in poverty shifted towards the north-west of England and away from the south and, in particular, the Midlands.

**CONCLUSIONS**

Early examples of Geographical Text Analysis using sizable corpora, of which this chapter is one example, demonstrate the fruitfulness in considering place in an analysis of language. Simultaneously, they show how analysis of textual data can inform analysis of geographical patterns. The preceding sections have demonstrated that, not only does the relationship between poverty and place fluctuate temporally in *The Times*, particularly in relation to the wider social context, economic trends, and political policies, but that different parts of the UK are associated with poverty in different ways. London, although it dominates the PNCs, seems to be the location of anti-poverty charities (see the data from the 1940s, and references to the Child Poverty Action Group), whereas northern cities, tend to be associated with the relationship between poverty and industrial decline. Poverty in the north also appears to be constructed as a state of being: people grow up in poverty, are in the poverty ‘trap’, etc. We have drawn these conclusions based on a single search term, but rather than
this being a limiting factor, we argue that, now these relationships between poverty and place are apparent, they can form the basis of hypotheses about the geography of *The Times’ coverage of poverty.

We acknowledge that there are limitations to the geoparsing process; whilst the concordance geoparser can account for certain spelling variations of place-names, the generation of PNCs is, nevertheless, restricted to explicit mentions of identifiable locations. We cannot, at the present state of the art, computationally analyse more general references to place which potentially have a set of coordinates, such as ‘the city’, or ‘the village’, nor can the geoparser generate PNCs for more general geographical areas, such as ‘the south’. Given a small text, the former can be added to the analysis manually, using a combination of geoparsing and close reading. However, at the scale of large corpora, this is not feasible. Furthermore, there is currently no systematic way to geoparse references to place which are relative, such as deictic terms like ‘here’ or ‘over there’. Future work to test our findings could, in the first instance, involve performing GTA on additional search terms associated with poverty, and could also involve more close reading of texts to uncover deictic references.

[8,890 words]

**FURTHER READING**

This article presents a critical reflection on the use of a large digital archive of Victorian newspapers to explore the discourse around a topic in nineteenth century history, namely public health.

Donaldson, C., Gregory, I.N. and Taylor, J.E. 2017. Locating the beautiful, picturesque, sublime and majestic: Spatially analysing the application of aesthetic terminology in descriptions of the English Lake District. *Journal of Historical Geography* 56, pp. 43-60

This article uses Geographical Text Analysis to explore the way that landscape aesthetics were represented prior to 1900 using a corpus of writing about the English Lake District

Hitchcock, T. 2013. Confronting the digital: or how academic history writing lost the plot. *Cultural and Social History* 10, pp.9-23

A highly opinionated article that critiques the ways that historians are making use of the digitised sources that are available to them.


A detailed guide to the use of Geographical Text Analysis using modern newspapers to study poverty.


This is a pioneering text on the use of corpus linguistics in historical research. It focuses on the changing use of the language of science and experimentation in the Early Modern period.
REFERENCES


Hitchcock, T. 2013. Confronting the digital: or how academic history writing lost the plot. Cultural and Social History 10, pp.9-23


Kirschenbaum, M.G. 2010, What is Digital Humanities and what’s it doing in English departments? ADE Bulletin, 150, pp. 55-61


Nicholson, B. 2012. Counting Culture; or, How to read Victorian newspapers from a distance. *Journal of Victorian Culture*, 17, pp. 238-246


ACKNOWLEDGEMENTS

This work was part-funded by the ESRC Centre for Corpus Approaches to Social Science (CASS): Grant number ES/K002155/1. We would also like to acknowledge the contributions of Andrew Hardie, Deputy Director for CASS, who has maintained the geoparser and helped to facilitate our analysis of The Times corpora.
BIOGRAPHICAL NOTES

Ian Gregory is Professor of Digital Humanities at Lancaster University. His research centres on the use of digital geographical approaches across the humanities and social sciences covering fields ranging from modern poverty to 19th century public health to literary history.

Laura L Paterson is a Lecturer in English Language and Applied Linguistics at The Open University, UK. Her work involves using corpus linguistics and critical discourse analysis to focuses on UK poverty and audience response to poverty porn. She has also written on discourses of marriage and epicene pronouns.
<table>
<thead>
<tr>
<th></th>
<th>Hits</th>
<th>Freq.</th>
<th>No. of texts</th>
<th>Avg. hits per text</th>
<th>Total corpus word count</th>
<th>Total corpus texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940s</td>
<td>1,768</td>
<td>6.09</td>
<td>1,153</td>
<td>1.53</td>
<td>290,448,326</td>
<td>3,100</td>
</tr>
<tr>
<td>1950s</td>
<td>2,230</td>
<td>5.30</td>
<td>1,337</td>
<td>1.66</td>
<td>420,978,622</td>
<td>3,087</td>
</tr>
<tr>
<td>1960s</td>
<td>4,359</td>
<td>9.39</td>
<td>1,956</td>
<td>2.23</td>
<td>519,523,147</td>
<td>3,098</td>
</tr>
<tr>
<td>1970s</td>
<td>6,129</td>
<td>12.75</td>
<td>2,095</td>
<td>2.93</td>
<td>480,659,269</td>
<td>2,789</td>
</tr>
<tr>
<td>1980s</td>
<td>6,651</td>
<td>12.38</td>
<td>2,373</td>
<td>2.80</td>
<td>537,031,328</td>
<td>3,090</td>
</tr>
<tr>
<td>1990s</td>
<td>8,614</td>
<td>11.23</td>
<td>2,610</td>
<td>3.30</td>
<td>767,266,910</td>
<td>3,125</td>
</tr>
<tr>
<td>2000s</td>
<td>14,500</td>
<td>17.63</td>
<td>2,913</td>
<td>4.98</td>
<td>822,524,677</td>
<td>3,122</td>
</tr>
</tbody>
</table>

Table 1: Hits for <poverty> in The Times corpora

<table>
<thead>
<tr>
<th>(1) Query hits</th>
<th>(2) Freq. per million</th>
<th>(3) PNC poverty +/- 15</th>
<th>(4) Normalised PNCs per query hits (per thousand hits)</th>
<th>(5) Normalised PNCs pmw (whole corpus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940s</td>
<td>1,768</td>
<td>6.09</td>
<td>95</td>
<td>53.73</td>
</tr>
<tr>
<td>1950s</td>
<td>2,230</td>
<td>5.30</td>
<td>73</td>
<td>32.74</td>
</tr>
<tr>
<td>1960s</td>
<td>4,359</td>
<td>9.39</td>
<td>121</td>
<td>27.76</td>
</tr>
<tr>
<td>1970s</td>
<td>5,952</td>
<td>12.38</td>
<td>157</td>
<td>26.38</td>
</tr>
<tr>
<td>1980s</td>
<td>5,977</td>
<td>11.13</td>
<td>242</td>
<td>40.49</td>
</tr>
<tr>
<td>1990s</td>
<td>8,614</td>
<td>11.23</td>
<td>215</td>
<td>24.96</td>
</tr>
<tr>
<td>2000s</td>
<td>10,408</td>
<td>12.65</td>
<td>232</td>
<td>22.29</td>
</tr>
<tr>
<td>TOTAL</td>
<td>39,308</td>
<td>10.15</td>
<td>1135</td>
<td>28.74</td>
</tr>
</tbody>
</table>
Table 2: Instances of `<poverty>` in the Times corpus by decade (pmw).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>PNC Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>group, child, Victorian, I, war, action, want, against, tiny, genteel, wishes, deduction, by, homelessness, big, suburbs, price, press, greatest, Peabody, evidence, administer, deal, society, link, hall, an, supplementary, rights, some, familiar, may, little, old, much</td>
</tr>
<tr>
<td>Not London</td>
<td>lab, Mr, sir, industrial, September, shopping, had, said, incomes, this, would, MP, rather, grown, effects, escape, measure, imagination, trap, when, times, been, primary, another, planning, asked, town, first, left, right, into, church, estate, not</td>
</tr>
</tbody>
</table>

Table 3: PNC Keywords derived from a comparison of the PNCs for Greater London with those for elsewhere from the 1950s onwards. Ranked by log-likelihood (p<.01, Italics p<.05. Observed ≥ 5). Punctuation, place-names, and numbers excluded.
Surrounded by docks and high, crowded poverty tenements, in this largest area of unbroken poverty, in all London, the world knows something of what our people in South London, poverty-stricken Liverpool, ' and compares his life then to his life today in Broadstairs 140, if he came to that conclusion, in the next session."

"Quite simply, poverty kills," he said. Deaths in the first year of life in Easterhouse. Many of the buildings on the London list are there not through inherent poverty but are victims of an unsympathetic owner or a slow-moving planning machine.

A safe and healthy place to live and work; anti-poverty strategies and 32 London Bridge Street community development that will foster an inclusive London SE 140.

Figure 2: Example concordance lines for <poverty*> in The Times 1990s corpus
Figure 3: Dot maps showing the pattern of PNCs in the 1940s and 1980s
Figure 4: Density smoothed map of all <poverty> PNCs from the Times, 1940-2009.
Figure 5: Macro-geographies of poverty instances over time. 1950s onward.
Figure 6: Spatial segregation analysis comparing the locations of PNCs in the 1980s with those from other decades from the 1950s on. PNCs and densities are for the 1980s.
Figure 7: Spatial segregation analysis comparing the locations of PNCs from the 1980s onwards with those from the 1950s to 1970s. PNCs and densities are for the 1980s onwards.