

Distant readings of the geographies in text corpora

Donaldson, Christopher; Murrieta-Flores, Patricia ; Gregory, Ian

Document Version
Peer reviewed version

Citation for published version (Harvard):
Donaldson, C, Murrieta-Flores, P & Gregory, I 2013, Distant readings of the geographies in text corpora: Mapping Norman Nicholson's poems and letters. in *Proceedings of the Third Conference on Digital Humanities in Luxembourg with a Special Focus on Reading Historical Sources in the Digital Age*. vol. 1681, (CVCE) University of Luxembourg, Luxembourg.

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:
Checked for eligibility: 25/01/2017

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Distant readings of the geographies in text corpora: Mapping Norman Nicholson's poems and letters

Christopher Donaldson*, Patricia Murrieta-Flores‡ and Ian N. Gregory†

*University of Birmingham

c.donaldson@bham.ac.uk

‡University of Chester

p.murrietaflores@chester.ac.uk

†Lancaster University

i.gregory@lancaster.ac.uk

Abstract

This short article summarises a preliminary study of the poetry and correspondence of the English poet Norman Nicholson (1914-1987), undertaken as part of Lancaster University's Spatial Humanities: Texts, GIS, Places project. In addition to offering a concise explanation of the Spatial Humanities project and the methods it employs, the article explains how working with GIS, visualisation techniques, and distant reading enriches our understanding both of the geography of Nicholson's poetry and of the spatial dimensions of the network of people with whom he exchanged letters throughout his career.

1 Introduction: Spatial Humanities: Texts, GIS, Places

Spatial Humanities: Texts, GIS, Places is a five-year European Research Council-funded research project. Our main ambition is to investigate how computational tools developed in the fields of geography and corpus linguistics can be adapted to enrich the study of literature, history and the arts. Within this context, our work primarily focuses on

developing techniques for analysing large corpora of books and newspapers using geo-visualisation technologies, primarily Geographic Information Systems (GIS).

Although they are often thought of as a type of computer mapping system, GIS are really a type of relational database. Conventionally, relational databases store data in tables wherein each row contains attribute information that corresponds to a specific entity, such as a person, place or event. In a GIS, this data includes spatial information, usually in the form of geographical coordinates, which allows for the relevant entity to be represented cartographically as a point, a line or polygon. (The latter shape is used when the spatial information refers to an area as opposed to a single precise location.) Broadly speaking, then, GIS are excellent resources for arranging, visualising and analysing the geographies of any dataset that contains spatial information or to which spatial information can be added (see, for example, Gregory, 2003; Gregory & Ell, 2007; Heyward et al 2011). What is more, as spatially organised databases, GIS also allow one to combine and analyse datasets that contain spatial information about the same regions or locations. As long as these regions or locations are defined using a real-world coordinate system (such as latitudes and longitudes) or a projection like Universal Transverse Mercator or British National Grid, any GIS database can be integrated with data from another.

The process of adding coordinate information to a corpus is called georeferencing. This process can either be performed manually or automatically. Manual georeferencing involves reading through a corpus and recording all of the place-name entities it contains. This procedure is commonly known as geoparsing. Once the corpus has been geoparsed, the recorded place-name entities can then be arranged in a database table and paired with metadata including geographical coordinates from a gazetteer. Once this spatial information has been added, the table can be read into any GIS application. Automated georeferencing involves using automated geo-parsing tools such as the Edinburgh Geoparser, which has been developed by a research team at the University of Edinburgh (see Grover et al, 2010). The Edinburgh Geoparser is a web-based georeferencing tool that consists of two interlinked components. The first of these components is a geo-tagger, which uses a named entity recognition (NER) sub-component to identify the place-name entities in an HTML or plain-text input file. The tagger automatically inserts <enamex> elements around these place-names, and inputs the file to the geo-resolver (the second component), which resolves the tagged place-names by cross-referencing them with a preselected gazetteer. We have already discussed and demonstrated this process elsewhere (Murrieta-Flores et al, 2015). What matters here is that the output of the automated geoparser system is a georeferenced file, containing tagged place-names annotated with geographical coordinate information.

Traditionally, GIS have been the domain of disciplines that use quantitative data, particularly in the environmental and social sciences. Within the humanities GIS have also been widely used by both historians and archaeologists. In history, they have proven especially valuable in subfields such as historical demography, which works with large amounts of quantitative data from censuses and similar tabular sources (see, for example, Gregory, 2008; Fotheringham et al, 2013). In the case of Archaeology, GIS have long been

used as management tools in excavations and field surveys. They have also been used in the research environment in tandem with advanced spatial and statistical analysis (see, for example, Wheatley and Gillings, 2002; Conolly and Lake, 2006). More recently, researchers have started to apply GIS to sources more directly applicable to humanities disciplines concerned with literature and the arts (Bodenhamer et al, 2010). One of the more important developments here has been the use of GIS in the analysis of literary and artistic representations of specific places and landscapes (see, for example, Cooper and Gregory, 2011; Cooper, Donaldson and Murrieta-Flores, 2016). In the Spatial Humanities project, for example, we are using GIS both to study the history of travel and tourism in the Lake District (an internationally renowned cultural landscape in the county of Cumbria in the northwest of England) and to chart how popular perceptions of the region have changed over time.

The Lake District is an ideal focus for a study that uses GIS because it is a region that has long been portrayed in poems, guidebooks and paintings. One thinks of William Wordsworth and J. M. W. Turner, of course; but one can also think of an array of writers and artists – from Celia Fiennes and Daniel Defoe to Alfred Wainwright and the Abraham Brothers – who have shaped our perception of this region and its landscapes. The Lakeland’s ties with twentieth-century art and literature are particularly notable. These include both the authors of famous children’s stories, like Arthur Ransome and Beatrix Potter, as well as important Modernist writers and artists, such as Kurt Schwitters, Michael Roberts, Kathleen Raine and, of course, the region’s own twentieth-century laureate: the poet, playwright and novelist Norman Nicholson.

2 Mapping Nicholson’s Poetry and Letters

In anticipation of the centenary of Nicholson’s birth, in January 2014, and in order to share in the task of bringing his works to a larger audience, we undertook a proof-of-concept GIS analysis of his five major verse collections as well as his network of correspondents, based on letters now held in the John Rylands University Library in Manchester.

2.1 Mapping the Poems

As a lifelong resident of Millom, on the west coast of Cumbria, Nicholson stands alongside other late-Modernist poets like R. S. Thomas and Patrick Kavanagh as a writer who shaped his craft through the close study of the landscape and the community in which he lived. Like Thomas and Kavanagh, moreover, Nicholson’s writings bespeak a conviction in the adequacy of local places, people and events to serve as the subjects of his art. We were particularly struck by how Nicholson’s poetry engages with the geography of Cumbria, and especially the landscapes around Millom and along the west Cumbrian coast. It occurred to us that GIS technology and other visualisation techniques could help us explore the nature of this engagement by facilitating an analysis of Nicholson’s five major verse collections: *Five Rivers* (1944), *Rock Face* (1948), *The Pot Geranium* (1952), *A Local Habitation* (1972) and *Sea to the West* (1981).

Although, as noted above, the Spatial Humanities project has made advances in the automated georeferencing of digitised text corpora (Murrieta et al 2015), in this study, we opted to extract this data manually due to the lack of a ready-to-use digitised version of these texts. Accordingly, creating a GIS of Nicholson's five collections involved a four-step process. In the first place, we had to read through each collection to identify every place-name reference it contained. These place-names were then entered into a table as individual rows, to which were added relevant attribute information, such as the title of the poem and the collection in which it appears. The place-names were then assigned geographical coordinates using Geonames, an open-access gazetteer. Once georeferenced in this manner, the full list of place-names was uploaded into ArcGIS to create a series of interactive maps displaying the places named in each book (Figure 1). As we wanted to have a more precise sense of the specific places with which Nicholson dealt in his poetry, we decided to exclude references to bodies of water, and large administrative regions – both of which are too spatially extensive to be adequately represented by point data. In total, then, the georeferencing process revealed that Nicholson's five collections contain 201 references to 169 unique place-names.

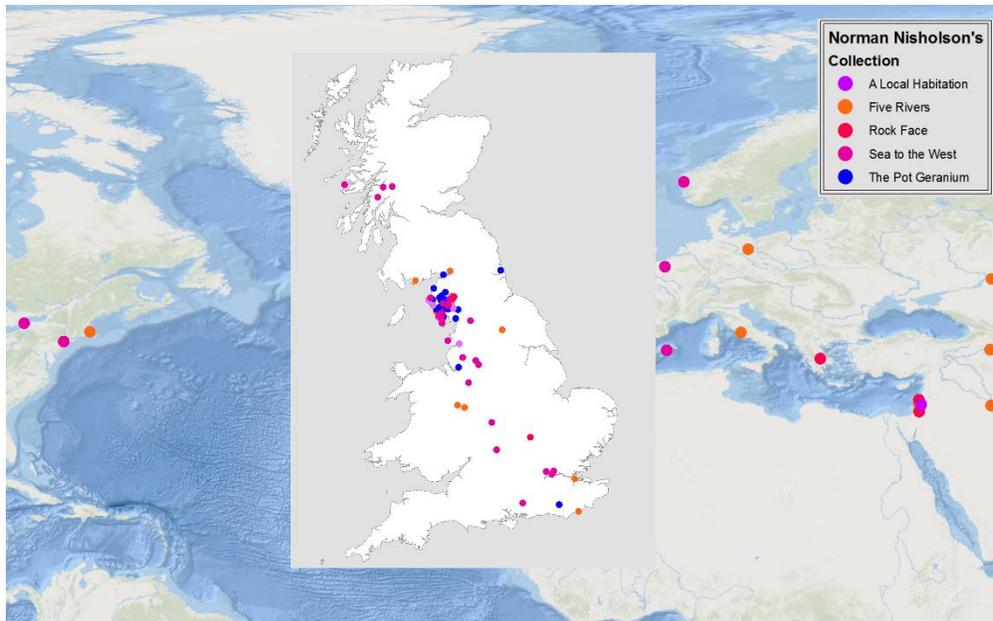


Figure 1: A GIS of Nicholson's five major poetry collections

Now, naming places is obviously not the only technique that Nicholson uses to create a sense of place in his poetry. Many of his finest poems evoke distinctive places and scenes, but contain no place-names whatsoever. Consider the opening stanza of his poem 'Coastal Journey':

A wet wind blows the waves across the sunset;
There is no more sea nor sky.

And the train halts where the railway line
Twists among the misty shifting sand,
Neither land nor estuary,
Neither wet nor dry. (Nicholson, 1944: p. 61).

When read in the context of Nicholson's collection, the imagery of these lines leads one to infer that the poem describes the scenery around the viaduct over the Duddon estuary, near Millom: a remote place, and one quite literally suspended between the elements of water, earth, and sky. But such speculations are really of only secondary importance, as the poem does not give us a named location. What matters here is that the intertidal landscape Nicholson invokes, with its 'misty shifting sand,' provides a perfect counterpart to the flux and flow of the neither-nors that shape the speaker's meditations. Here image and thought coalesce to give a concrete sense of setting.

Such examples must be kept in mind. Yet, many of Nicholson's poems *do* contain references to locatable places and landmarks, and many have place-names in their titles: examples include 'Egremont', 'Cleator Moor', 'Whitehaven', and 'St Bees'. Accordingly, although it does not provide an utterly comprehensive geography of Nicholson's poetry, focusing on the places mentioned in his five collections does call attention to how his relationship with place and geography evolved throughout his career.

Permit a few statistics to carry the point (Figure 2). Nicholson's first collection, *Five Rivers*, contains references to 57 places; 46 of those 57 places are in the UK, and 19 of those 46 are along the west Cumberland coast. In his second collection, *Rock Face*, one finds only 13 places named, 8 of which are located in the UK. *The Pot Geranium* contains the names of 41 places, 10 of which are located in (or near) Millom and 24 of which are found elsewhere in Cumbria. *A Local Habitation* is, as one might expect, even more emphatically local. For although it contains references to only 27 places in all, 17 of these are places in or around Millom. Finally, *Sea to the West* contains 63 place-names (the most of any of Nicholson's volumes), 57 of which are in the UK. 12 of these 57 places are in or near Millom, and a further 25 are places in Cumbria.

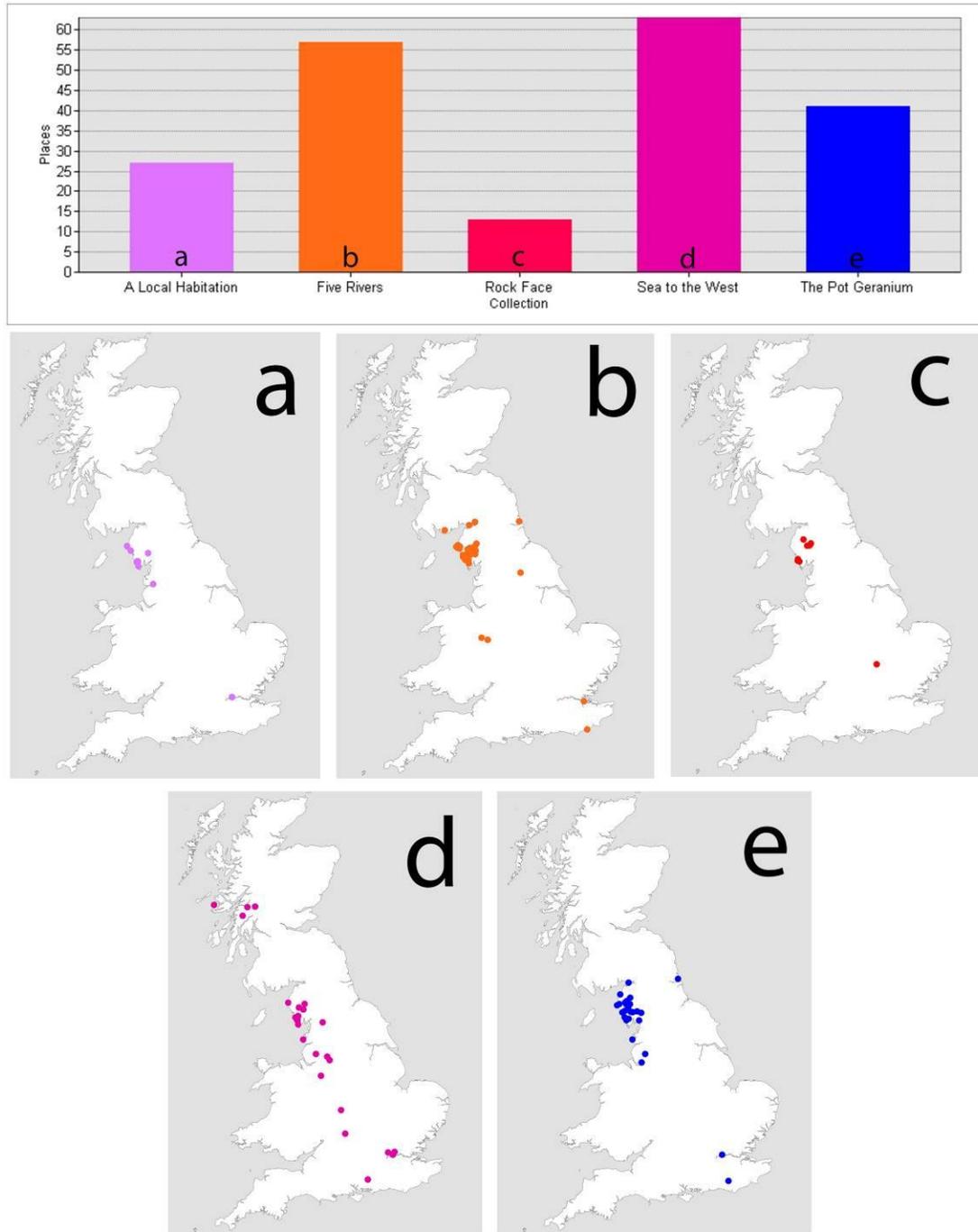


Figure 2: Locations and graphic count of the collections

Plotting and counting place-names might seem like a peculiar way of appreciating poetry; nevertheless, it does point up an important fact about the way Nicholson's identity as a poet developed. Specifically, it alerts us to the fact that Millom, with its landscape and its inhabitants, came to figure more prominently in the poems Nicholson included in his last

three collections. Significantly, this pattern corresponds to one that has been noted in previous assessments of Nicholson's poetry. As Neil Curry (the editor of Nicholson's *Collected Poems*) has observed, the course of Nicholson's career as a poet is one gradually tending towards both a championing of his own identity as a provincial writer and, concomitantly, an acceptance of his 'own "one small radius of rock" as a microcosm of the whole world' (Curry, 1994: p. xx). Indeed, although it is true that *Five Rivers* is distinguished by its stirring invocation of the towns, villages and waterways of the Cumbrian coast, and although it is equally true that *Rock Face* is distinguished by, amongst other things, its subtle appeal to local scenery, neither book addresses Millom with the fervour and frankness of Nicholson's later volumes.

Significantly, a GIS mapping of Nicholson's verse also calls attention to spatial patterns that have previously received less attention from his readers. Indeed, as the maps displayed here indicate, the geography of his poetry extends well beyond Northwest England. Although most of the places named in his five volumes are located either within Cumbria or elsewhere in the UK, those volumes do nevertheless contain a small but substantial number of references to places in Europe, the Mediterranean and the Middle East (Figure 1). Intriguingly, on returning to Nicholson's poems, one discovers that many of these distant places are invoked either (as in 'The Pot Geranium') as points within a geographical constellation of relationship linking Millom with the wider world or (as in 'The Bow in the Cloud') as part of symbolic landscapes, such as the 'mythical land of Palestine': that place 'of cloud and mist and dream', which being neither wholly England nor the Holy Land is found not in the Ordnance Survey but written on the map of the mind and the soul (Nicholson, 1953: p. 222).

Our findings thus shed an illuminating light on the geography of Nicholson's poetry. But, as Angharad Saunders, amongst others, reminds us, the geography of a writer's work is not limited to the places with which his or her works engage. Indeed, it also includes the geography of the social and relational networks in which that writer participates. In this way, as Saunders explains, the geography of a writer's works includes not only the locations mentioned in his or her publications, but also the geographies in which his or her 'letters, journals, memoirs, [unpublished] writings and diaries' circulate (see Saunders, 2013: 4). With this in mind, we decided to augment our analysis of the geography of Nicholson's poetry with a consideration of the geography of his circle of correspondents. Suitably, as we discovered, the geography we encountered when mapping Nicholson's poems – with its various local and global connections – complements the one formed by his network of correspondents.

2.2 Mapping the Letters

Using Frances Baker's catalogue of the Norman Nicholson Archive at the John Rylands University Library, we next set out to explore the geographical distribution of the people who wrote letters to Nicholson between 1940 and 1987. In total, this collection comprises

565 letters from some 275 different correspondents. Although this, of course, represents only a portion of the letters Nicholson received during his lifetime, the collection nonetheless affords us terrific insights into the variety of different people with whom he corresponded.



Figure 3: Radial Map depicting the Distribution of Nicholson's international correspondents

The visualisation techniques we employed to examine the geographical distribution of Nicholson's correspondents included GIS-based radial and density maps as well as word clouds. This combination of cartographic and schematic visualisations is significant, because it helps to bring different aspects of the collection into focus. For example, whereas the radial map (Figure 3) points up the location of Nicholson's correspondents around the globe, the density map (Figure 4) alerts us both to the location of his correspondents within the UK and to the proportion of letters sent from each of those locations. Similarly, word-clouds (Figures 5 & 6) provide us with an abstract way of modelling how frequently each of the correspondents and locations appear in the collection.

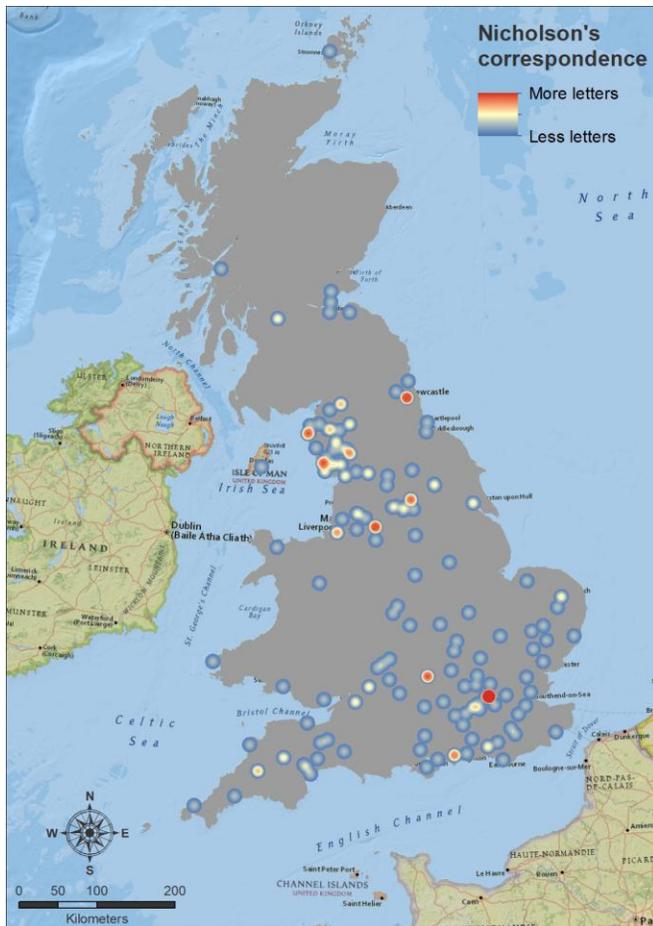


Figure 4: Density Map depicting the distribution of Nicholson's UK-based correspondents and the relative amount of letters received from each location

Creating these visualizations drew our attention to a number of interesting patterns – perhaps most significantly, the surprisingly global distribution of Nicholson's correspondents. Of the 167 different places represented in the corpus of letters, 140 were found to be located in the UK. The other 27 were located in places around the world, including Canada, Australia, Denmark and the United States. This only goes to show that Nicholson's community, though rooted in Millom and in England, extended much farther around the world. We were especially surprised to discover among the non-British correspondents a high proportion of Italians and Scandinavians, many of whom seem to have written to Nicholson about translations and critical studies of his works.

London, UK



Figure 5: Word cloud displaying the frequency on the locations of Nicholson's correspondents

We were also intrigued to see that London (the metropolitan centre against which Nicholson is famous for having defined his identity as a provincial writer) is the place-name that appears most frequently in the collection. Although this might seem counterintuitive, it makes sense when one remembers the role London played, and continues to play, as a centre for literary and artistic life. Notably, the offices of Faber and Faber, who published Nicholson's five major collections are located there. Noticing these patterns has sharpened our awareness of the complexity underlying Nicholson's status as a writer who, though proudly provincial, nonetheless lived and worked in an increasingly global and cosmopolitan age.



Figure 6: Word cloud ranking the names of Nicholson's correspondents in terms of the number of letters received from each

3 Conclusions

In this particular case study our aim was to experiment with a relatively small, manually built dataset using geographic visualisations and common corpus linguistics approaches such as frequency analysis. There are many other aspects of this dataset that remain to be explored, not least the intriguing relationship between the maps we have created and the development of Nicholson's use of place-names in his poetry. There are also different methodological approaches that future research might exploit. For the present, it suffices to say that, as this case study indicates, GIS technology can provide researchers with a powerful tool for exploring a variety of different textual sources, including both works of literature, such as poems, and other forms of writing, such as letters. This article has analysed a modest collection of such sources looking at the major works and correspondence of a single writer. The combined potential of modern computing power and the ability of GIS to organise and summarise vast amounts of information could allow researchers to go much further, especially if one were to take advantage of the increasing availability of large digital corpora.

The on-going and future work of the Spatial Humanities project moves in the latter direction. Currently, we are working to analyse corpora as diverse as a 1.5-million word corpus of Lake District writing, a 13-million word corpus of material that accompanies British census and public health reports, and a corpus of newspapers that runs to many billions of words. There is, however, a tension at the core of this work. On the one hand there is the desire to use distant reading approaches, which enable us to summarise and integrate large amounts of material and information. On the other hand, there is the need to remain true to more traditional humanities paradigms, which have proven that close reading and interpretation is necessary for understanding the kinds of subtle and nuanced information that textual sources contain. The success of using GIS and other forms of digital geospatial technology as a resource for revealing new understandings and insights from digital texts depends on achieving the right balance between these two approaches.

References

- Bodenhamer, David J., Corrigan, John and Harris, Trevor M. (2010, eds) *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington & Indianapolis: Indiana University Press.
- Cooper, David and Gregory, Ian N. (2011) Mapping the English Lake District: A literary GIS. *Transaction of the Institute of British Geographers*, 36, 89-108.
- Cooper, David, Donaldson, Christopher and Murrieta-Flores, P. (2016, eds) *Literary Mapping in the Digital Age*. Farnham: Ashgate.
- Conolly, James and Lake, Mark (2006) *Geographical information systems in archaeology*. Cambridge: Cambridge University Press.
- Curry, Neil (1994, ed.) *Norman Nicholson: Collected Poems*. London: Faber and Faber.
- Fotheringham, A. Stewart, Kelly, Mary H. and Charlton, Martin (2013) The demographic impacts of the Irish famine: Towards a greater geographical understanding, *Transactions of the Institute of British Geographers*, 38, 221-237.
- Gregory, Ian, (2003) *A Place in History: A Guide to Using GIS in Historical Research*. Oxford: Oxbow Books.
- Gregory, Ian N. (2008) Different places, different stories: Infant mortality decline in England & Wales, 1851-1911, *Annals of the Association of American Geographers*, 98, 773-794.
- Gregory, Ian N. and Ell, Paul S., (2007) *Historical GIS: Technologies, Methodologies and Scholarship*. Cambridge: Cambridge University Press.
- Grover, Claire, Tobin, Richard, Byrne, Kate, Woollard, Michael, Reid, James, Dunn, Stuart and Ball, Julian (2010) Use of the Edinburgh geoparser for georeferencing digitized historical collections, *Philosophical Transactions of the Royal Society A*, 368, 3875–3889.
- Heyward, Ian, Cornelius, Sarah and Carver, Steve, (2011) *An Introduction to Geographical Information Systems*, 4th edn. Harlow: Pearson Prentice Hall.
- Murrieta-Flores, Patricia, Baron, Alistair, Gregory, Ian, Hardie, Andrew, and Rayson, Paul (2015) Automatically analysing large texts in a GIS environment: The Registrar General's reports and cholera in the nineteenth century. *Transactions in GIS*. doi: 10.1111/tgis.12106.
- Norman Nicholson (1944) *Five Rivers*. London: Faber and Faber.
- Nicholson, Norman (1953) The Comic Prophet, *The Listener* (6 August), 222.

Saunders, A (2013) The Spatial Event of Writing: John Gallsworthy and the creation of *Fraternity*, *Cultural Geographies*, 20.3, 285-98.

Wheatley, David and Gillings, Mark (2002) *Spatial technology and archaeology: the archaeological applications of GIS*. London: Taylor and Francis.

Biography of the authors

Chris Donaldson is Lecturer in Romanticism at the University of Birmingham

Patricia Murrieta-Flores is Senior Researcher in Archaeology at the University of Chester

Ian N. Gregory is Professor of Digital Humanities at Lancaster University

Acknowledgements

We would like to thank Frances Baker for her catalogue of the Nicholson Archive at the John Rylands University Library. Thanks are also due to the European Research Council, who funded this research under the EU's Seventh Framework Programme (FP7/2007-2013) / ERC grant 'Spatial Humanities: Texts, GIS, places' (agreement number 283850).