CA

Scotland's Poetics of Space: An Experiment in Geospatial Semantics

Michael Gavin and Eric Gidal

11.16.17

This paper reports on a collaborative project that develops new applications of spatial text analysis. We offer a methodology to identify and evaluate correlations between semantic and geographic distance in a printed corpus.[1] Our work combines geographic information science (GIS) with corpus linguistics to study how places are described, how spatial categories coalesce and change, and how patterns in language correlate with patterns in geography.[2] At its base, we'll argue,

---

[1] Textual approaches to geographic questions have been pioneered in the humanities by scholars like Ian Gregory, Andrew Hardie, and Matthew Wilkens. See Ian N. Gregory and Andrew Hardie, "Visual GISting: Bringing Together Corpus Linguistics and Geographical Information Systems," *Literary & Linguistic Computing* 26, 3 (2011): 297-314; and Ian Gregory, David Cooper, Andrew Hardie, and Paul Rayson, "Spatializing and Analyzing Digital Texts: Corpora, GIS, and Places," in *Deep Maps and Spatial Narratives*, 150-78. See also Matthew Wilkens, "The Geographic Imagination of Civil War-Era American Fiction," *American Literary History*, 25, 4 (Winter 2013): 803:40.

[2] Werner Kuhn lays out basic principles of geospatial semantics in "Geospatial Semantics: Why, of What, and How?"*Journal on Data Semantics III* (2005): 1-24. Angela Schwering reviews various techniques, borrowed from corpus linguistics, for extrapolating geospatial concepts from texts in "Ap-

geospatial text analysis blends two key theoretical concepts. In GIS, the principle of "spatial autocorrelation" holds that nearby places tend to have similar characteristics at similar times. In computational semantics, the "distributional hypothesis" suggests that words with similar meanings tend to appear near each other in documents. Taken together, these hypotheses form a single idea: similar places at similar times tend to be described using similar words. To put this idea into action, we propose a data structure — the *word-place matrix* — that indexes a corpus to geographical features. Over a word-place matrix, measurements of word association are simultaneously measurements of spatial overlap. In just the same way that terms in a corpus can be gathered into conceptual groups, words in a word-place matrix can be associated with corresponding geographical regions. This very simple data format offers an easy-to-implement and theoretically well-grounded method for studying cultural topics with strong spatial components.

Our case study is nineteenth-century Scotland and the reception of the poems of Ossian. The Scottish poet and aspiring folklorist James Macpherson published a sequence of volumes in the 1760s and 1770s that adapted Gaelic ballads attributed to the mythical ancient bard Ossian.[3] His controversial "translations" incorporated names and themes from traditional bardic poetry concerning the battles of the warrior-king Fingal, but placed them within an entirely invented epic apparatus. Macpherson accompanied these epics with a host of philological and geographical annotations that presented Ossian as a poetic geographer whose elegiac verse could be mined for ethnographic and environmental information. Despite mounting evidence that Macpherson had cut many of the poems out of whole cloth, a mix of national pride and romantic fascination maintained a lively debate over their authenticity well into the nineteenth century, a debate

proaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey," *Transactions in GIS* 12, 1 (2008): 5-29. See also Krzysztof Janowicz, et al, "Geospatial Semantics and Linked Spatiotemporal Data – Past, Present, and Future," *Semantic Web* (2012): 321-32.

[3]Macpherson's story is told most completely in Fiona J. Stafford, *The Sublime Savage: A Study of James Macpherson and The Poems of Ossian* (Edinburgh: Edinburgh UP, 1988) and Paul deGategno, *James Macpherson* (Boston: Twayne, 1989). Dafydd Moore places the poetry within the context of eighteenth-century Scottish intellectual and literary history in *Enlightenment and Romance in Macpherson's "The Poems of Ossian"* (London: Ashgate, 2003) and Derick S. Thomson has offered the most extensive account of Macpherson's Gaelic sources in *The Gaelic Sources of Macpherson's "Ossian"* (Edinburgh: Oliver and Boyd, 1952). Howard Gaskill has edited the most comprehensive study of the poetry's reception in *The Reception of Ossian in Europe* (London: Thoemmes Continuum, 2004). Other significant collections of studies include Howard Gaskill, ed. *Ossian Revisited* (Edinburgh: Edinburgh UP, 1991); Fiona J. Stafford and Howard Gaskill, eds. *From Gaelic to Romantic: Ossianic Translations* (Amsterdam: Rodopi, 1998); Gerald Bär and Howard Gaskill, eds. *Ossian and National Epic* (Frankfurt: Peter Lang, 2012); and, most recently, Dafydd Moore, ed. *The International Companion to James Macpherson and the Poems of Ossian* (Glasgow: Scottish Literature International, 2017). See also the special issue of *Journal for Eighteenth Century Studies* 39, 2 (2016), edited by Sebastian Mitchell.

that increasingly turned to questions of place. As antiquarians strove to authenticate Macpherson's work, they gathered testimonies and manuscripts from local sources, finding echoes of Ossianic folklore throughout the Scottish Highlands.[4] References to Fingal and his heroes proliferated ever further as local enthusiasts across Scotland laid claim to pieces of the myth. For scholars of Scottish Romanticism, the intensity of this vogue suggests a deep-seated and widespread ambivalence about Scotland's place in the modern, British industrial order. Ossianic poetry was valorized in the Highlands as a remnant of a Gaelic heritage in opposition to Lowland Scots and English culture but celebrated by Scotland's literary and commercial classes as part of a collective British identity. This ambivalence goes back to Macpherson himself, who presented his books as acts of cultural salvage in response to the economic and political decimation of clan societies throughout the Highlands. He believed that Gaelic poetry flourished in regions largely isolated from trade and industry, but also that those same places were being invaded and contaminated by modernity. Ever since, scholars have debated how to characterize antiquarian poetry's participation in the period's larger shifts toward industrialization, urbanization, and British national identity.

We contribute to this line of inquiry by tracing the dispersion of Ossianic references across a corpus of nineteenth-century geographical writing. If the basic proposition of geospatial semantics holds, the above history suggests several verifiable hypotheses. Scottish places affiliated with Ossian should be recognizably different from others: in particular, they should have rural economies and rugged terrain typical of the Highlands; they should overlap significantly with other markers of Gaelic identity and with the history of clan politics; conversely, they should correlate less strongly with markers of urban infrastructure and agricultural improvement. Each of these expectations, we'll show, is borne out by the model. We also ask several general questions for which we could formulate no precise organizing expectation: How was industrialization imprinted in the language of geography? What kinds of words cluster spatially, and which distribute more evenly? Which kinds of agriculture and industry were most common in Ossianic space, and which were more common elsewhere? As topical markers, the terms "Ossian" and "Fingal" organize a wide array of geosemantic

---

[4]As Nigel Leask puts it, "the explanatory *mythe histoire* permeating Macpherson's prefaces and footnotes encouraged the reader to perform ideological acts of geographical as well as historical localisation, clues to which are sown like cryptic seeds in the verse itself, inspiring travellers to visit the Highlands in search of Fingalian locations and bardic fragments." "Fingalian Topographies: Ossian and the Highland Tour, 1760-1805" *Journal for Eighteenth Century Studies* 39, 2 (2016): 183-96, 186. Eric Gidal has studied how these quixotic yet innovative methods in environmental philology and statistical geography created a bibliographic record of Scottish industrialization. See Eric Gidal, *Ossianic Unconformities: Bardic Poetry in the Industrial Age* (Charlottesville: University of Virginia Press, 2015).

patterns that correlate with Scottish industrialization (most correlate negatively, as scholars would expect, but some correlate positively). Throughout, our primary aim remains methodological. Tracing the intersections among antiquarian and industrial space suggests new approaches for environmental history. We offer geospatial semantics as a new method for writing the histories of spaces and for revealing conceptual ecologies where literature, capital, technology, and environment entwine.

Historical background: Scottish industry & modernity

In 1700, Scotland was an agricultural economy: 90% of its people were dispersed among small villages where their primary trade - in cattle, timber, fish, and slate - was limited to their English and Dutch neighbors. By 1900, Scotland was an industrial powerhouse built upon textiles, heavy manufacturing, and international commerce. These changes have been described along many different axes: economic development,[5] technological diffusion,[6] social transformation,[7] working-class politics and culture,[8] and energy consumption.[9] T. C. Smout has charted how agricultural reform and the shift to fossil fuel consumption from the 1750s onward increased grain yields and mobilized labor for trade and industry, accelerating a demographic redistribution to urban centers, while also polluting the water, air, and soil, and reshaping Scotland's biogeography, as various plant

---

[5]On the economic history of Scottish industrialization, see R.H. Campbell, *Scotland Since 1707: The Rise of an Industrial Society* (Oxford UP, 1965) and *The Rise and Fall of Scottish Industry, 1707-1939* (Edinburgh UP, 1980). See also C.A. Whatley, *The Industrial Revolution in Scotland* (Cambridge UP, 1997) and T. M. Devine, C. H. Lee and G. C. Peden, *The Transformation of Scotland: The Economy Since 1700* (Edinburgh UP, 2005).

[6]On the technological history of Scottish industrialization, see John Butt, *The Industrial Archaeology of Scotland* (Newton Abbot Devon, 1967) and Geoffrey D. Hay and Geoffrey Stell, *Monuments of Industry: An Illustrated Historical Record* (Edinburgh: Royal Commission on the Ancient and Historical Monuments of Scotland, 1986).

[7]On Scottish social history, see T. M. Devine and R. Mitchison, eds. *People and Society in Scotland, Vol. 1, 1700-1830* (Edinburgh UP, 1988); W.H. Fraser and R. J. Morris, eds, *People and Society in Scotland, Vol. II, 1830-1914* (Edinburgh UP, 1990), and T. C. Smout, *A Century of the Scottish People, 1830-1950* (Yale UP, 1986).

[8]On Scottish working-class politics and culture, see W. W. Knox, *Industrial Nation: Work, Culture and Society in Scotland, 1800-Present* (Edinburgh UP, 1999). See also J.D. Young, *The Rousing of the Scottish Working Class* (Croom Helm, 1979).

[9]On the energy and environmental histories of Scottish industrialization, see T. C. Smout, *Nature Contested: Environmental History in Scotland and Northern England Since 1600* (Edinburgh UP, 2000) and *Exploring Environmental History: Selected Essays* (Edinburgh UP, 2009).

species rose, fell, and dispersed in new patterns.[10] C. H. Lee has traced how transatlantic trade and new credit and legal systems concentrated capital in Glasgow, where Scottish industrialists invested in new machinery and manufacturing processes.[11] R. J. Price has taken a longer view and mapped the geological factors contributing to industrialization: the concentrated coal reserves of the Central Belt and the successive glaciations that created watershed topographies perfectly suited to power early textile mills.[12] And Charles Withers, T. M. Devine, and W. W. Knox have focused on the cultural history of these changes. Gaelic-speaking people in the Highlands found their communal townships and clan relations reconfigured under legal and market pressures from the south and east, and industrial labor markets in Lowland urban centers upended class and social distinctions.[13]

These and many other measures of Scottish industrialization may be correlated with nineteenth-century practices of geographical inquiry made possible by an ever-expanding print culture. Of particular relevance to our study is the development of statistical geography. Sir John Sinclair's *Statistical Account of Scotland* (1791-99) offered a parish-by-parish census of natural, political, and cultural information. The *Account* was an attempt to construct, in Sinclair's words, "the Natural History and Political State of Scotland" through methods of distributed empirical observation and statistical collation he had adapted from natural philosophy and antiquarian survey.[14] Sinclair distributed questionnaires to every parish minister in Scotland, and their responses detailed "local and minute information" that could then be cross-referenced and compared. For Sinclair, "useful information" and the "promotion of improvement" were inextricable elements of a single project with many parts. The economic reorganization of Scottish society, Sinclair believed, would require the expansion of commerce and manufacturing as well as agricultural reform, developments that further demanded concomitant infrastructural investments in roads, canals, ports, and planned villages.[15] Improvements promoted by Sinclair and the Highland Societies aimed to

---

[10] T. C. Smout, "Land and Sea: The Environment" in T. M. Devine and Jenny Wormald, *The Oxford Handbook of Modern Scottish History* (Oxford UP, 2012), 19-38.

[11] C. H. Lee, "The Establishment of the Financial Network," in Devine, Lee, and Peden, *The Transformation of Scotland*, 100-127.

[12] See R. J. Price, *Scotland's Environment during the last 30,000 Years* (Edinburgh: Scottish Academic Press, 1983).

[13] See Charles W. J. Withers, *Urban Highlanders: Highland-Lowland Migration and Urban Gaelic Culture, 1700-1900* (East Linton: Tuckwell, 1988); T. M. Devine, *The Transformation of Rural Scotland: Social Change and the Agrarian Economy, 1660-1815* (Edinburgh UP, 1994); and Knox, *Industrial Nation.*

[14] John Sinclair, *Queries Drawn Up for the Purposes of Elucidating the Natural History and Political State of Scotland* (Edinburgh, 1790).

[15] John Sinclair, *Analysis of the Statistical Account of Scotland; with a General View of the History of*

increase the production of linen, wool, corn, and cattle, to centralize land holdings, and to continue the removal of traditional settlements. These ambitions place Sinclair's experiment in statistical information firmly within what Fredrik Albritton Jonsson has called a "Whig ecology for the Highlands," one that applied quantitative methodologies to a survey of the entire Scottish nation.[16] Sinclair's efforts were absorbed and expanded by a second *Statistical Account* published during the 1830s and 40s and, throughout the nineteenth century, by large-scale productions of topographical dictionaries, gazetteers, maps, and directories. This body of geographic information represents Scotland comprehensively as a history-bearing site of resource extraction, industrial manufacturing, and commercial distribution.

Sinclair was one of many defenders of Ossian and used information gathered in his *Statistical Account* to vindicate Macpherson's work. His fellow Highland Society member Henry Mackenzie used analogous methods for collating manuscripts and testimonials in his 1805 *Report on the Nature and Authenticity of the Poems of Ossian*, and one of Sinclair's parish sources, Alexander Stewart, drew upon both of these earlier works in composing his "Topography of some of the Principal Scenes of Fingal and His Warriors" published as an appendix to an 1807 Gaelic and Latin edition of the poems. Stewart and Sinclair also produced a richly rendered "Map of Ancient Selma, the residence of Fingal, with part of the Fingalian Territories in the Shire of Argyle" (See Figure 1). This map offers an early model of spatial literary and cultural investigation, but also a project of environmental history conducted during a period of accelerating change. Stewart defends its imaginative locations by collating Ossianic poetry with Gaelic and Latin toponymies, topographical details, and archaeological recoveries of Bronze Age structures, while at the same time noting later agricultural developments that had altered the landscapes and disturbed further potential evidence. The "Map of Ancient Selma" is in certain respects analogous to current geotagging techniques in digital literary studies. It visualizes a sequence of textual correlations between Sinclair's *Statistical Account* and the poems of Ossian whose truth claims are bolstered by their cartographical representation.

That Country, and Discussions on some Important Branches of Political Economy, 2 vols. (Edinburgh, 1825-26).

[16]Fredrik Albritton Jonsson, *Enlightenment's Frontier: The Scottish Highlands and the Origins of Environmentalism* (Yale UP, 2013). On Sinclair, see also Rosalind Mitchison, *Agricultural Sir John: The Life of Sir John Sinclair of Ulbster 1754-1835* (London: Geoffrey Bles, 1962); and Withers, "Scotland Accounted For: An Introduction to the 'Old' (1791-1799) and the New (1834-1845) Statistical Accounts of Scotland."
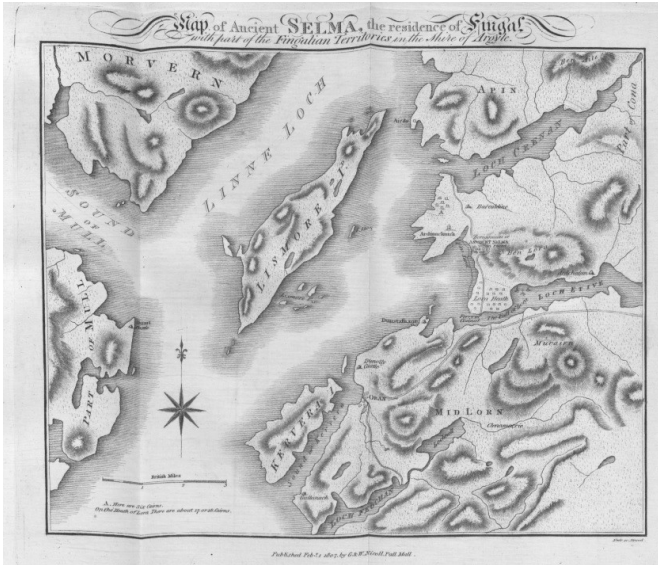
Figure 1. "Map of Ancient Selma, the Residence of Fingal, with Part of the Fingalian Territories in the Shire of Argyle," from The Poems of Ossian, in the Original Gaelic, with a Literal Translation into Latin (London, 1807).

Macpherson and his defenders cast bardic discourse in opposition to contemporaneous developments in industry and agriculture. Macpherson observed a "great change" suffered by "the genius of the highlanders": "The communication with the rest of the island is open, and the introduction of trade and manufactures has destroyed that leisure which was formerly dedicated to hearing and repeating the poems of ancient times."[17] In the appendix to his *Dissertation on the Poems of Ossian* (1765), Hugh Blair likewise suggested that modern economic and social developments had displaced the recitations of Ossianic poetry. "The manners of the inhabitants of the Highland counties have of late undergone a great change," he writes,

Agriculture, trades, and manufactures, begin to take place of hunting, and the shepherd's life. The introduction of the busy and laborious arts has considerably abated that poetical enthusiasm which is better suited to a vacant and indolent state. The fondness of reciting their old poems decays; the custom of teaching them to their children is fallen into desuetude; and few are now to be found except

---

[17]James Macpherson, *The Poems of Ossian and related works*, ed. Howard Gaskill (University of Edinburgh Press, 1996), 51.

old men, who can rehearse from memory any considerable parts of them.[18]

The "great change" observed by Macpherson and Blair overlays spatial and temporal divisions. On the one hand, modernity separates the Highlands from the rest of Scotland as a vestige of the past. On the other, the manners of modernity have intruded northward and have permeated the spaces of Ossianic myth. This elegiac defense of Ossianic poetry simultaneously distinguishes and absorbs the Highlands within a modernizing Scottish nation and British economic order, a rhetorical approach taken up by numerous agricultural and industrial reformers in the decades that followed.

Macpherson and Blair's observations suggest two related propositions: first, that Ossianic myth is rooted in places largely divided from the "agriculture, trades, and manufactures" of the Scottish Lowlands, and, second, that some aspects of those modernizing forces have nonetheless penetrated Ossianic space. Scholars from Peter Womack and Paul Baines to Nigel Leask and Sebastian Mitchell have underscored how Ossianic space proved amenable to both romantic obscurity and administrative specificity and how the tensions between the two were central to its geographical diffusion.[19] As Penny Fielding observes, "the regulated space of modernity fictionalises its 'traditional' past into romance space, and, in turn, these imaginary spaces act both as an alternative and as a prop to modernity."[20] Eric Gidal has demonstrated how this bifurcated experience of Ossianic geography provided a language for registering the accelerating spatial transformations of industrial modernity.[21] Applying geospatial semantics to this corpus of geographical writing allows us to test these ideas against a new set of interpretive parameters. When we turn to our corpus of later eighteenth- and nineteenth-century documents, we find clear evidence of this geographical hybridity.

Data & Methods

Modern British GIS and nineteenth-century Scottish geographical writing share a common vocabulary as well as, in many ways, a common sense of Scotland's physical and social space. In part because of the institutional continuity of the

---

[18] Macpherson, *The Poems of Ossian*, 404.

[19] Peter Womack, *Improvement and Romance: Constructing the Myth of the Highlands* (New York: Macmillan, 1989); Paul Baines, "Ossianic Geographies," *Scotlands* 4 (1997): 44-61; Leask, "Fingalian Topographies"; Sebastian Mitchell, "Landscape and the Sense of Place in The Poems of Ossian" in Moore, *International Companion*, 65-75.

[20] Fielding, *Scotland and the Fictions of Geography*, 81.

[21] Gidal, *Ossianic Unconformities*.

Ordnance Survey, place names and their locations within Britain's official geography have remained largely consistent, making it possible to match places mentioned in eighteenth- and nineteenth-century documents with corresponding features in the Survey's georeferenced datasets.[22] In the analysis that follows, we measure the distribution of geographic diction by tracing patterns of word use over a spatial model of Great Britain provided by the Ordnance Survey.[23]

The publications in our corpus are georeferentially rich, highly structured documents with entries that vary in length from a few sentences to several pages. Regularized typography and diction made it possible to use regular expressions to parse the entries from each document. In Nicholas Carlisle's *Topographical Dictionary of Scotland* (1813), for example, each entry begins with a new line followed by the toponym in all capital letters. (See Figure 2.) The precise typographical markers vary across the collection, so each title required a unique parsing algorithm. However, once the toponyms were identified, each was matched against the Ordnance Survey's 2015 gazetteer of named places. In the cases of a simple one-to-one match, the OS's identifier was assigned to the matching place name. When more than one possible candidate was found for a given place, the text of the description was searched for other toponyms, and the centroid was calculated for all possible places named in the description. The candidate place closest to that centroid was then selected as the match.[24] When no matches were found, a tightly constrained fuzzy search was used to find similarly named places. Over the entire collection, which totals approximately 7 million words, we captured 60,951 descriptions of 17,046 unique places. However, because of OCR errors and variations in spelling from the late eighteenth century to 2015, for some entries no suitable match could be found: 11,485 of the descriptions are "placeless" and so had to be excluded from many of the analyses performed below.[25]

---

[22] Rachel Hewitt traces the Ordnance Survey's origins back to the eighteenth century and to the aftermath of the Jacobite Rebellion, when Scottish rebels successfully evaded English forces by fleeing into the mountainous Highlands, inspiring military surveyors to create the first detailed maps of the region. See *Map of a Nation: A Biography of the Ordnance Survey* (London: Granta, 2010), especially 8-37.

[23] We used the 2015 version of the 1:50,000 Scale Gazetteer made available by the Ordnance Survey's OpenData initiative).

[24] Our process was modeled on the technique described by David A. Smith and Gregory Crane in "Disambiguating Geographic Names in a Historical Digital Library." Perseus Project/Tufts University, 2001. Using the centroid of shared references is only one possible approach. A survey of toponym resolution techniques can be found in Jochen Lothar Leidner, *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names* (University of Edinburgh, 2007), 81-109.

[25] Our toponym-matching algorithm was deliberately conservative, on the assumption that less, cleaner data would support more robust study than more, messier data. However, three types of error appear throughout the dataset: 1) the parsing algorithms sometimes failed to differentiate new entries, lumping descriptions together that should have been separated; 2) the toponym-matching
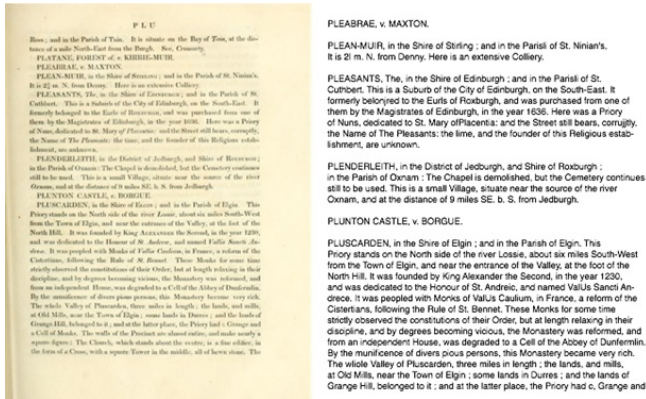
Figure 2. Nicholas Carlisle, A topographical dictionary of Scotland, and of the islands in the British seas (London: Printed for G. and W. Nichol, 1813). From The Internet Archive. https://archive.org/details/topographicaldic02carl.

Once the documents were parsed and their entries were tagged with GIS identifiers, the textual data was processed to support semantic analysis. In a typical application of computer-based semantics, a corpus is organized into a *term-document matrix*, with a row for each word and a column for each document.[26]

---

algorithm sometimes failed to choose the correct places when more than one town with the same name exists; and 3) the fuzzy search algorithm sometimes found false matches. Errors of the third type are fairly easy to spot, as when, for example, the town "Ide" was matched to section headings, "I.,"II.," etc. that the parsing algorithm misrecognized as description headings. We reviewed a random sample of 1,000 toponym matches and found 26 instances where a human researcher would have rejected the algorithm's match, for an estimated success rate of 97.4%. We have not attempted to estimate the other kinds of error. To avoid human bias, no hand correction was performed. Our goal in this essay is to demonstrate the validity of geospatial text analysis and to explore some of its interpretive possibilities for cultural and environmental history. However, many technical questions remain: How should error be minimized when resolving toponyms in historical corpora? How much "noise" is to be expected, in what kinds of collections analyzed over what kinds of spaces? How do different text-processing techniques affect the kinds of conclusions that can be drawn from geospatial text analysis? We set these important questions aside for now.

[26] The use of vector spaces to represent collections of documents was first pioneered by researchers in the field of information retrieval. The earliest attempt to imagine such a system can be found in Hans Peter Luhn, "A New Method for Recording and Searching Information," *American Documentation* 4, 1 (1953). The technique is most often associated with Gerard Salton: See G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," *Communications of the ACM* 18, 11 (1975): 613-20. For a survey of more recent applications, see Peter D. Turney and Patrick Pantel, "From Frequency to Meaning: Vector Space Models of Semantics," *Journal of Artificial Intelligence Research* 37 (2010) 141-188. For a gentle introduction to the mathematical principles that inform the method, see Dominic Widdows, *Geometry and Meaning* (Stanford: CSLI Publications, 2004). Drawing from Salton (1975), Turney and Pantel define the "bag of words hypothesis" as saying that "frequencies of words in a document tend to indicate the relevance of the document to a query" ("From

The value of each cell (w_{ij}) is the frequency with which each word (w_i) appears in each document (d_j). Our study modifies this basic format. We use a *word-place matrix* in which the rows represent words, just as above, but the columns represent places, rather than documents. The value of each cell (w_{ij}) is the sum of the frequency with which each word (w_i) appears in entries for each place (p_j) across all documents. Thus, all descriptions of Edinburgh are gathered into a single bag of words, which can then be compared to the bags of words that describe Glasgow, Aberdeen, and Dundee.[27] In the interest of computational efficiency, we restricted our vocabulary to the 12,000 most frequent words in the corpus, excluding stopwords. The result is a vector-space model with 12,000 rows across 17,046 dimensions, and each of those dimensions is associated with coordinates provided by the Ordnance Survey.[28]
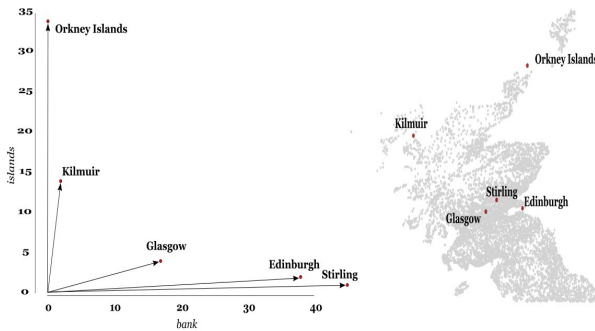

Figure 3. Places in semantic space. Left: Places distributed over two dimensions (bank, islands). Right: Locations among the British Ordnance Survey's named places gazetteer.

Because the word-place matrix is structured with variables that correspond to geographically defined points, measures of word usage can be correlated to mea-

---

Frequency to Meaning," 153). When applied to topoynms, the bag of words hypothesis suggests that frequencies of words in a set of descriptions about a place tend to indicate the relevance of the place to a similar keyword query.

[27] Drawing from Salton (1975), Turney and Pantel define the "bag of words hypothesis" as saying that "frequencies of words in a document tend to indicate the relevance of the document to a query" ("From Frequency to Meaning," 153). When applied to topoynms, the bag of words hypothesis suggests that frequencies of words in a set of descriptions about a place tend to indicate the relevance of the place to a similar keyword query.

[28] The Ordnance Survey does not use latitude and longitude coordinates. Instead, it uses an idiosyncratic spatial model, called the British National Grid, in which each place is located by its distance in meters east and north from an imaginary point just off the southwest corner of the British Isles.

sures of spatial distribution.[29] Our primary metric treats places as vectors in high-dimensional space and evaluates the cosine distances among them. (See Table 1.) When comparing places, the word counts operate as coordinates. For example, the word *islands* appears prominently in descriptions of Kilmuir and the Orkneys, while *bank* is more prevalent in accounts of Glasgow, Edinburgh, and Stirling. By treating these points as vectors from the origin (0, 0), it's possible to calculate the cosine of the angle that separates them. (See Figure 3.) The more acute the angle, the more similar the descriptive vocabulary. We define the semantic distance that separates any two places as the cosine distance that separates their respective column vectors:

$$SemDist\ (p_1, p_2)\ =\ 1\ -\ \frac{p_1 \cdot p_2}{\|p_1\| \, \|p_2\|}$$

The geographic distance that separates two places is estimated by taking the Euclidean distance between the easting and northing coordinates provided by the Ordnance Survey:

$$GeoDist(p_1, p_2)\ =\ \sqrt{(east_{p_1} - east_{p_2})^2 + (north_{p_1} - north_{p_2})^2}$$

These two primary measures outline the basic structure of our geosemantic model.[30] The first shows how similar places are. The second shows how far apart they are.

	Edinburgh	Glasgow	Kilmuir	Orkney Islands	Stirling
bank	38	17	2	0	45
city	345	117	1	0	5
coal	11	5	1	1	13
islands	2	4	14	34	1
royal	170	15	3	2	124
water	51	10	25	4	39
western	48	14	5	6	25

---

[29] A similar model, combining geographic information with a vector-space document model, is developed in Guoray Cai, "GeoVSM: An Integrated Retrieval Model for Geographic Information," in *GIScience 2002*, ed. Max J. Egenhofer and David M. Mark (Springer 2002), 65-79. Cai writes, "For geographic information, the document space can be divided into two kinds of subspaces: a geographic subspace and a thematic subspace. *Geographic subspace* is a two dimensional space. Documents may be considered similar or different based on the spatial relationships (inclusion, overlapping, adjacency, etc.) of their footprints in geographic space. *Thematic subspace* is a multidimensional space, where dimensions represent different thematic concerns of a document collection" (66).

[30] For a concise but welcoming explanation of these two calculations in the context of information retrieval, see Widdows, *Geometry and Meaning*, chapter 5.

Table 1. Sample word-place matrix. Word counts from the descriptions of each place are organized into a large matrix, making it possible to show what lexical features are most distinctive of each locale and to trace how words are distributed through the corpus.

Taken together, these measures allow us to test the interrelation between theories drawn from corpus linguistics and quantitative geography. In linguistics, the "distributional hypothesis" says that similar words tend to be used in similar contexts.[31] Within just about any corpus, words with similar meanings will tend to cluster together: legal documents use different words from medical documents, which use different words from love letters. These differences correlate to what we think of colloquially as the meanings of words.[32] This hypothesis is often stated as a dictum by J. R. Firth, "You shall know a word by the company it keeps."[33] In geography, the principle of "spatial autocorrelation" holds that nearby places will tend to have similar features: neighboring towns will, more often than not, sit at relatively even elevations, experience nearly identical annual rainfalls, and share common demographics and cultures.[34] Sometimes this principle is given as Tobler's First Law, which states that, in a geographic system, "everything is related to everything else, but near things are more related than distant things."[35] The distributional hypothesis says that words tend to cluster in documents; the principle of spatial autocorrelation says that geographic attributes tend to cluster in space. Geospatial text analysis combines these two ideas into a single proposition: Similar places tend to be described using similar words.[36]

---

[31] The distributional hypothesis was first advanced by Zellig Harris in "Distributional Structure," *Word* 10, 2-3 (1954). See also relevant discussion in Turney and Pantel (2010).

[32] The literature on computational approaches to semantic similarity is extensive. In addition to Turney and Pantel (2010), see also recent surveys in Katrin Erk, "Vector Space Models of Word Meaning and Phrase Meaning: A Survey," *Language and Linguistics Compass* (2012): 635-53; and Stephen Clark, "Vector Space Models of Lexical Meaning," in *The Handbook of Contemporary Semantic Theory, 2nd Edition*, ed. Shalom Lappin and Chris Fox (Wiley-Blackwell, 2015), 493-520. Most recently, discussion has tended to hinge on the use of machine-learning algorithms, like word2vec, that create dense, low-dimensional word-embedding models. See Mikolov, et al. "Efficient Estimation of Word Representations in Vector Space," *Computation and Language* (2013). For a detailed breakdown of the relationship between traditional distributional models and word-embedding models, see Omer Levy, Yoav Goldberg, and Ido Dagan, "Improving Distributional Similarity with Lessons Learned from Word Embeddings," *Transactions of the Association for Computational Linguistics* 3 (2015): 211-25.

[33] J. R. Firth, *Papers in Linguistics, 1934-1951*, (London: Oxford UP, 1957), 11.

[34] See Luc Anselin, *Spatial Econometrics: Methods and Models* (Kluwer Academic Publishers, 1988).

[35] Waldo Tobler, "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography* 46, 2 (1970): 234-40.

[36] Cai, who discusses this principle more narrowly in the context of information retrieval, explains: "When documents are put into the context of the geographic world, the potential spatial interactions between places (diffusion, movement of information through space) and the spatial patterns of docu-
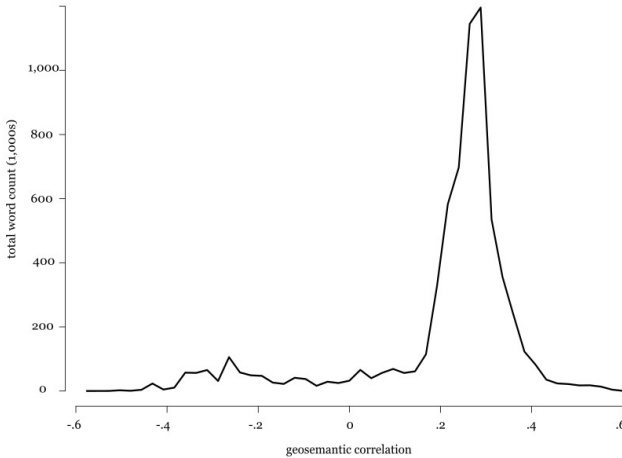
Figure 4. Geosemantic correlation by word count. Of the roughly 7 million words in the corpus, most (77%) are found in place descriptions that establish a connection between their descriptive vocabulary and their physical locations; that is, the semantic distances and the geographic distances that separate them from other places correlate strongly (between .2 and .4 Pearson coefficient).

This proposition suggests a quantitatively verifiable hypothesis, which we use to validate our geosemantic model. If the corpus actually describes real geographic variation, semantic distance should correlate positively with geographic distance. Our research began by demonstrating this correlation. For each place, we measured both kinds of distance against every other place in the corpus. We then calculated the correlation between these two series of distances (using the Pearson coefficient) to create a general metric showing how geo-correlative each place is. Over the entire corpus, the mean correlation between semantic and spatial distance is .2785, with most places (and most words) falling between .2 and .4. (See Figure 4) Small towns in idiosyncratically named regions are often highly geocorrelative because the names of the local parishes, rivers, lochs, and other features that appear in their descriptions are highly distinctive — this is often true across the corpus for both English and Scottish places. Places with longer, more substantive descriptions tend to correlate more strongly in the north, while southern English cities sit outside the main focus of our corpus. (Figure 5) For readers interested in such things, the p value for this correlation, when contrasted against a null hypothesis that assumes zero correlation, is less than 2.2e-16. We

ment distribution provide rich clues for judging relevance of a document in its associated geographic context." "GeoVSM," 73.

also tested the effect of error by measuring the geocorrelation of the corpus with its points randomly substituted: as the semantic data were randomly scattered over the model, its correlation to British space approached zero. [37]
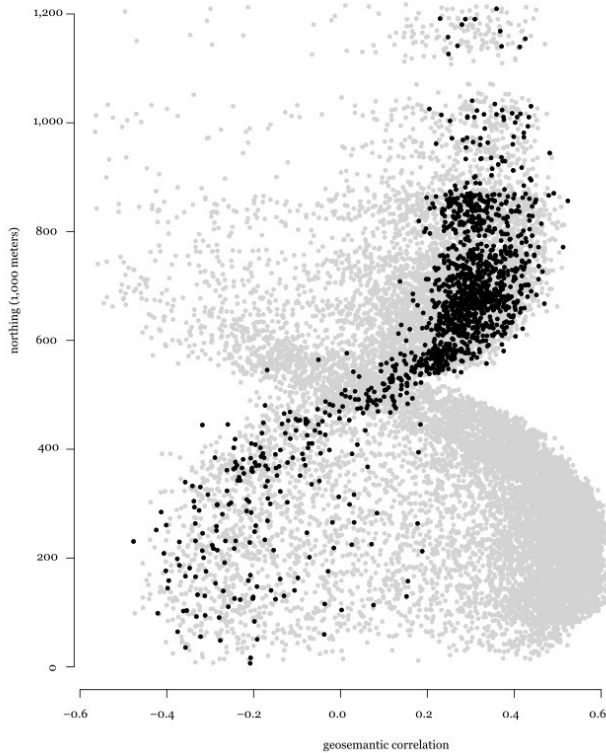

Figure 5.1. Geosemantic correlation by northing. Places in the British National Grid are located with "easting" and "northing" coordinates. Most of our textual data is based in Scotland, but a couple sources included English and Welsh towns. Of the places with more than 500 words of description (pictured in black), most

---

[37] To test the effect of error on the system, we measured geocorrelation while randomly substituting vectors one at a time, re-calculating the mean at each step. Before any substitutions had been performed, the mean Pearson correlation between semantic and geographic distance was 0.2785. After all 17,046 semantic vectors had been randomly re-assigned, the mean correlation was 0.0310. As expected, the declining correlation is neatly linear, with a perfectly randomized model approaching a minimum correlation of 0 and an ideal, error-free model approaching some as-yet-unspecified normal maximum.

are geocorrelative. Place descriptions with more than 500 words that fail to correlate spatially are almost entirely located in the south, outside the main focus of the corpus. This graph shows that our data is strong for Scotland but not for England or Wales.

Two general considerations are worth keeping in mind. First, these numbers suggest that our textual data overall is responsive to geography; that is to say, these figures validate our model by showing a clear correlation between semantic and spatial distribution. The geospatial semantic hypothesis offered above has been confirmed, at least for this corpus, so we have good reason to trust that our model is accurately capturing at least some spatial components to British geographical writing. Second, it must be noted that we have no way to evaluate whether this level of correlation is "high" or "low" when compared against other corpora. Our sense is that the correlation will prove to be strong, because the results of our analyses (described below) align closely with our qualitative understanding of Scottish cultural geography on every point for which we could formulate a precise expectation. However, the overall question of what counts as a "strong" or "weak" correlation remains open.

Another key feature of geospatial text analysis is that it allows researchers to trace the geographic distribution of descriptive terms. Which words were used where? When comparing words, we use the same basic calculation, but over the rows instead of the columns. We define the geosemantic similarity between any two words as the cosine similarity between their vectors in a word-place matrix:

$$Sim\,(w_1, w_2) \;=\; \frac{w_1 \,\cdot\, w_2}{\|w_1\|\,\|w_2\|}$$

Notice that when discussing places we usually refer to semantic distance and when discussing words we generally refer to semantic similarity. These are, of course, different ways of expressing the same basic concept. However, the terms "distance" and "similarity" have different connotations. When we talk about georeferenced places, we are concerned by how near or distant they are from each other, spatially and semantically. When we talk about words, we will be interested instead in how they spread over the Scottish landscape, and we'll look in particular for areas of overlap. We define the "semantic footprint" of a term as the signature of its distribution over a list of places. For example, the term *coal* has a semantic footprint of { 11, 5, 1, 1, 13 } over the places { Edinburgh, Glasgow, Kilmuir, Orkney Islands, and Stirling }. Like any other geographical attribute, a

semantic footprint can be analyzed spatially. In this sample, *coal* appears heavily in Edinburgh and Stirling, as do *bank* and *royal*, so the three terms are highly similar. (See Table 2.) The mean similarity score for words in relation to *coal* is .801: *coal*, *bank*, *royal*, *water*, and *western* are all above this average; *city* and *islands* are below. Z-score standardization measures how far each score deviates from the mean. We say that z-scores above 0 indicate a positive correlation and that z-scores less than 0 indicate a negative correlation. In this example, we say that *bank*, *royal*, *water*, and *western* all correlate positively with *coal* and that *city* and *islands* both correlate negatively. The absolute value of the standard score reflects the strength of the correlation.

	sim(w1,w2)	Z-score: (sim - mean) / deviation
coal	1	0.635
bank	0.998	0.629
royal	0.948	0.466
western	0.929	0.404
water	0.924	0.386
city	0.685	-0.394
islands	0.156	-2.121
Mean:	0.801	
Standard deviation:	0.306	

Table 2. Sample cosine similarity and standard scoring.

The semantic footprint of a term in a georeferenced corpus can also be plotted and mapped. Just as a weather map might show areas of cloud cover or rainfall, semantic maps can expose regions where terms predominate in place descriptions. When those terms are toponyms, they often cluster very coherently.[38] (See Figure 6.) For example, Argyllshire was a county in western Scotland, roughly corresponding to the modern county of Argyll and Bute. As one would expect, points where the term *argyllshire* is over-represented in the corpus tend to be places that were in or near Argyllshire itself. Similarly, the city of Edinburgh was a common reference point for towns in its vicinity, and so *edinburgh*'s semantic footprint extends to all suburbs and other nearby locations. Industrial, agricultural, and other terms of regional interest can be mapped as well. (See Figure 7.) The term

---

[38] Readers must keep in mind, however, that clustering algorithms tend to exaggerate that coherence by presenting as continuously distributed features that were, in reality, discretely placed. For example, the maps presented in this essay use a smoothing algorithm known as "hot spot" analysis to highlight main regions. See A. Getis and J. K. Ord, "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis* 24, 3 (1992) and J. K. Ord and A. Getis, "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application," *Geographical Analysis* 27, 4 (1995). For a summary discussion of the calculation, see A. Stewart Fotheringham, Chris Brunsdon, and Martin Charlton, *Quantitative Geography: Perspectives on Spatial Data Analysis* (London: SAGE Publications, 2000), 96-114. There exist different variations of the basic G statistic. We represent the value of each feature by taking the (non-inclusive) sum of the all features divided by their distance in meters.

coal clusters most tightly in Scotland's Central Belt, where coal production had significantly transformed the landscape by 1850, while descriptions of *fisheries* appear most distinctively along the coasts and harbors.
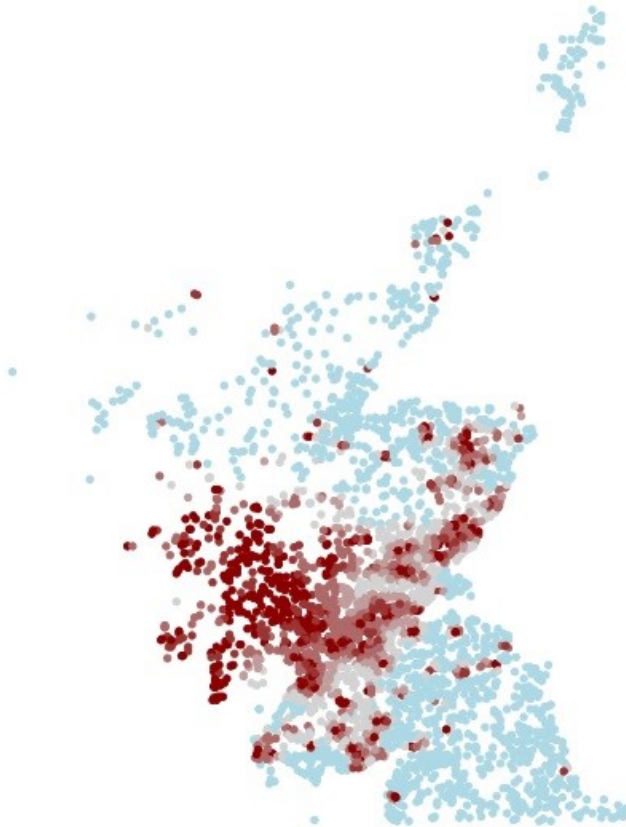

Figure 6.1. The semantic footprints of toponyms: argyllshire. Points in red indicate regional overrepresentation of a given term. Toponyms tend to cluster near their geographical locations.
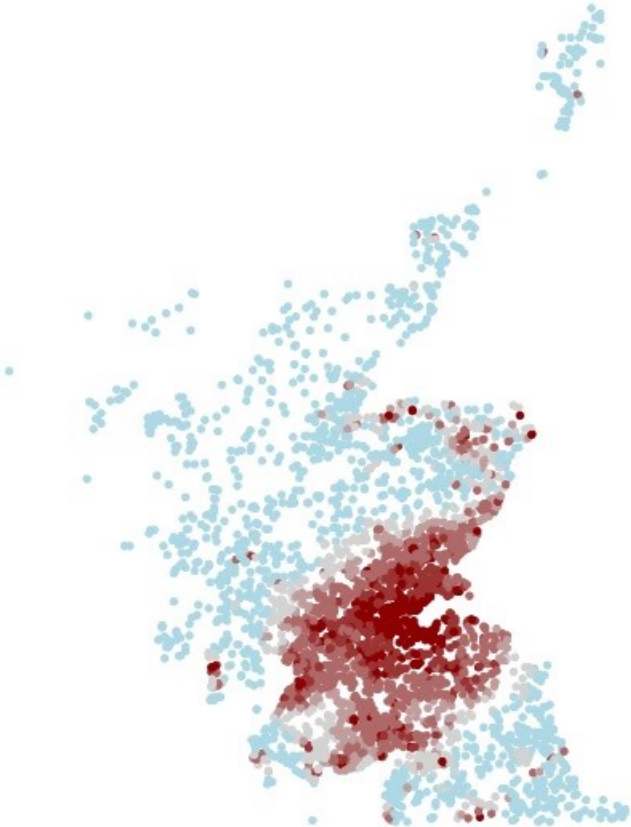
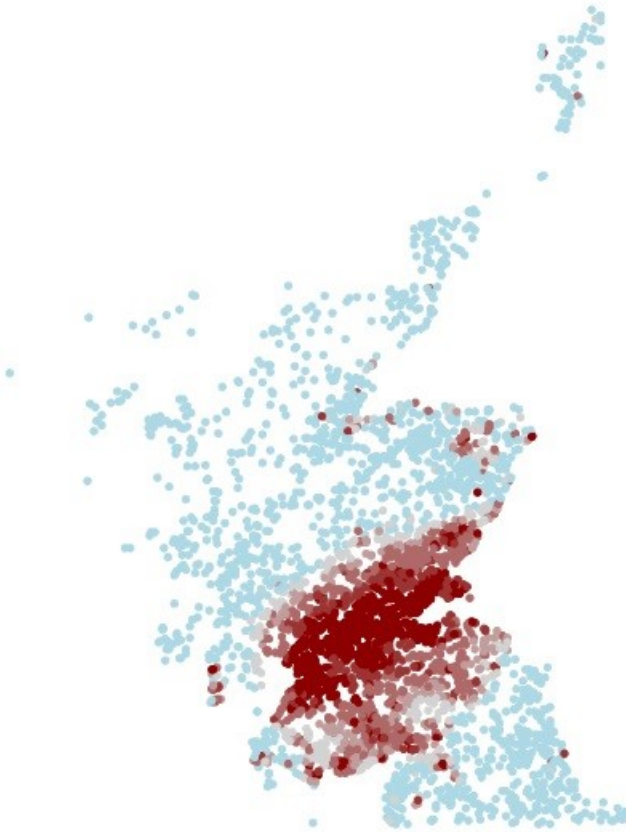Figure 6.2. The semantic footprints of toponyms: edinburgh.

Figure 7.1 The semantic footprint of coal. Points in red indicate regional overrep-resentation of a given term. Terms related to industry tend to cluster near their functional areas.
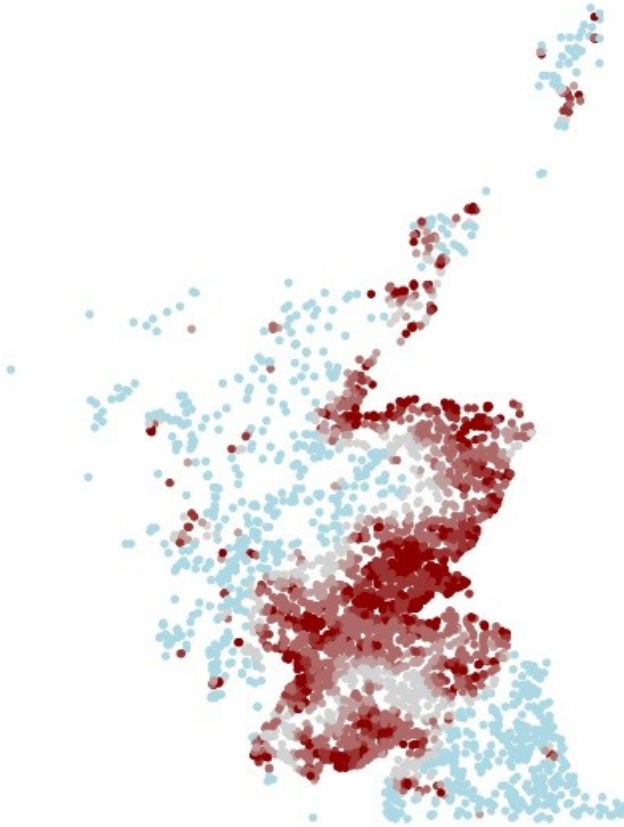
Figure 7.2. The semantic footprint of fisheries.

The above analytical methods exemplify a line of inquiry we call historical geospatial semantics, or "geosemantics" for short. Like geospatial text analysis more generally, geosemantics involves tagging a corpus and measuring the spatial distribution of references contained therein.[39] The specificity of geosemantics lies in the nature of the corpus and in the kinds of analysis it supports: text analysis becomes geosemantic when its meanings correlate with its places, such that a corpus can stand (provisionally) as a proxy for a real space. As in the examples of *coal* and *fisheries* illustrated in Figure 7, geosemantics simultaneously measures the distribution of concepts in a corpus of documents and the distribution of

---

[39]The project in the digital humanities most similar to ours is "Mapping the Lakes: Towards a Literary GIS," led by Ian Gregory. http://www.lancaster.ac.uk/mappingthelakes

features in geographic space. In what follows, we perform a spatial reading of nineteenth-century Scottish geography to explore how the distribution of references to Ossianic tradition compares with the distributions of terms that refer to Scotland's industrial development and changing energy systems. Our goal is to trace in outline the cultural, economic, and ecological contours of historical space, and to demonstrate how Ossian's poetry was folded into a larger story of Scotland's modernity.

Case Study (I): Fingal and the Energy Systems of Scotland

Literary scholars have long observed how Ossianic poetry romanticized Scotland's landscapes and its ancient past.[40] This geographical and conceptual association was first encouraged by Macpherson, who assured his readers of the poems' native origins and praised their "wild simplicity."[41] Evidence of this romanticization appears throughout our corpus. Folklorists following Macpherson gathered testimony from local residents, mapping a topography of Gaelic tradition across Scotland's rural counties.[42] Our analysis traces this topography by measuring the spatial distribution of the terms *fingal* and *ossian*, which our calculations will treat as a single composite vector.[43] References to Ossianic myth are spread fairly evenly over Scotland and appear with notable frequency in the mountainous regions of Argyllshire and Inverness. (See Figure 8.) The vocabulary used to describe Ossianic places reflects this distribution. Places where *fingal* or *ossian* appear are more likely than others to be described as *wild*, *rugged*, and *natural*. Other words reminiscent of Romantic landscape aesthetics that predominate in these areas include *mountains*, *gaelic*, *heroes*, *highland*, *language*, *hills*, *scenery*, and *lakes*. (See Table 3.) Terms most geosemantically similar to *fingal* and *ossian* identify many points of connection between Ossianic myth and a more general discourse of bardic nationalism: these include references to the Glencoe

---

[40] Elizabeth Bray, *The Discovery of the Hebrides: Voyagers to the Western Isles 1745-1883* (London: Collins, 1986); Malcolm Andrews, "The Highlands Tour and the Ossianic Sublime," in *The Search for the Picturesque: Landscape Aesthetics and Tourism in Britain, 1760-1800* (Aldershot: Scholar, 1989), 197-240; Womack, *Improvement and Romance*.

[41] Macpherson, *Poems of Ossian*, 474.

[42] Baines, "Ossianic Geographies"; Thomas A. McKean, "The Fieldwork Legacy of James Macpherson" *Journal of American Folklore* 114, 454 (2001): 447-63; Eric Gidal, *Ossianic Unconformities*; Nigel Leask, "Fingalian Topographies."

[43] To find all references to two words in a corpus, we simply add the two vectors together using basic addition. If we had wanted to find places where *fingal* and *ossian* were both mentioned together, we would use componentwise multiplication to find their points of intersection. For an explanation of vector composition, see Erk (2012).

Massacre (*glencoe, massacre*), discussions of clan politics (*clan, oath, allegiance, tribe, campbell, macdonald*), references to poetry and aesthetics (*poems, horror, bards, scenery, songs, terrific*), and Celtic identity and language (*gaelic, language, celtic, race*).
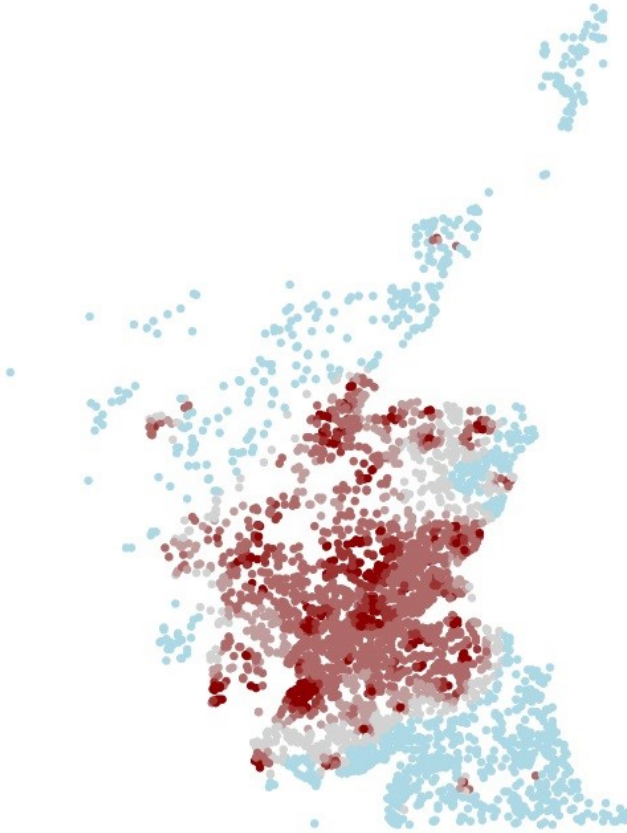
Figure 8. The semantic footprint of Ossianic myth. Map shows a hotspot analysis of the distribution of 345 uses of fingal or ossian.

Term	Frequency		
fingal	10.92	miles	4.08
wild	5.27	highlands	4.06
mountains	5.21	kilmorie	4.06
somerled	5.04	knowledge	4.05
glen	4.88	covered	4.05
mountain	4.86	northern	4.02
ossian	4.8	mac	4.02
heroes	4.75	distance	4
gaelic	4.72	climate	4

Term	Frequency		
highland	4.53	cascades	3.99
nan	4.52	spoken	3.98
language	4.52	places	3.97
districts	4.49	propagating	3.95
far	4.43	narrow	3.95
glens	4.28	eagle	3.95
district	4.26	forms	3.94
donald	4.24	natural	3.93
country	4.2	quite	3.91
hills	4.15	moun	3.9
among	4.15	goats	3.9
torrents	4.12	mountainous	3.9
lakes	4.1	whole	3.9
along	4.09	clan	3.89
scenery	4.09	several	3.89
rugged	4.09	remote	3.89

Table 3.1 Terms over-represented in places that mention Fingal or Ossian. Frequency displays the standardized difference (z-score) of each term's mean frequency in places that mention Fingal or Ossian, above those terms' average distributions over the corpus as a whole.

Term	Similarity		
ossian	12.67	highlands	3.65
fingal	11.68	tribe	3.64
glencoe	6.76	campbell	3.61
massacre	6.12	moun	3.52
instructions	5.94	cave	3.51
monzie	5.52	naked	3.49
glen	5.45	eagle	3.49
glenalmond	5.15	celtic	3.41
poems	5.09	argyleshire	3.4
wild	4.97	mercy	3.36
heroes	4.92	sudden	3.35
habitation	4.79	clan	3.35
oath	4.75	macpherson	3.35
highland	4.72	glens	3.34
gaelic	4.72	bards	3.32
bard	4.61	scenery	3.3
mountains	4.59	ledge	3.27
horror	4.42	macdonald	3.22
mountain	4.31	propagating	3.22
allegiance	4.28	accept	3.14
torrents	4.06	songs	3.13
plaid	4.02	donald	3.12
silence	4.01	far	3.11
language	4	race	3.09
cascades	3.74	terrific	3.09

Table 3.2. Terms geosemantically similar to *fingal + ossian*. Similarity displays the standardized cosine similarity (z-score) between each term and a vector that includes all uses of the terms *fingal* or *ossian*, distributed over the 17,046 places included in the word-place matrix. This table displays the 50 terms with the highest geosemantic similarity to that combined vector. In the tables that follow (Tables 4, 5, and 6), we use the same metric to examine the spatial relation between Ossianic myth and topical groups of other words.

The coherence of these results (their obviousness, even, to literary historians) suggests two aspects of Ossianic discourse that will be important to the analysis that follows. First, despite being widely and sparsely dispersed over the territories of Scotland, Ossianic myth was described using a remarkably consistent vocabulary. Second, references to Fingal and Ossian tended to appear in places that were themselves highly similar, often in mountainous rural areas near rivers. These characteristics of Ossianic topography suggest that its idealized landscapes were functionally distinct from the rest of Scotland, and, for this reason, they will correlate meaningfully (whether positively or negatively) to any other aspects of Scottish culture that had a strong spatial component. To tease apart these relationships we dig a bit deeper the vocabulary of the corpus, extending our analysis beyond the most similar terms to identify other topical groups in the data. We use a keyword-search method that selects terms based on their semantic similarity to an outside point of reference. Of everything that correlates spatially with *fingal* and *ossian* what's most similar to a farm or a factory? For the sake of expository clarity, we consider positive correlations separately from negative ones.

References to Fingal and Ossian predominate in the interior of Scotland, to the north and the west of highly populated areas where industry was most active, but those references were also widely distributed and therefore share significant overlap with other common descriptive terms. (See Table 4.) They correlate strongly with terms that describe agricultural activity, especially involving breeding livestock, such as *sheep* and *cattle*, but also with starchy vegetable produce like *potatoes*, *barley*, and *oats*, which were suitable for growing in the hilly terrain. Husbandry reform had already begun to transform the Highlands by the middle of the eighteenth century, and these developments intensified over the decades that followed: the spaces of Ossianic myth were also places where agricultural labor management (*tenants, maintain, hire, cottars, landlords*) and forms of improvement that consolidated ownership came to dominate farming activity on both sides of the Highland Line.[44]

livestock & produce				management				fishing	
sheep	2.72	root	1.39	tenants	2.63	indifferent	2.33	trout	2.15
cattle	2.44	potatoes	1.27	maintain	1.36	modes	1.97	bag	1.67
black	2.17	crops	1.2	tenantry	1.28	system	1.83	salmon	1.52
graziers	1.71	grass	1.08	hire	1.02	respects	1.83	abounding	1.29
faced	1.67	bread	0.89	exclusively	1.01	judicious	1.51	ascend	1.05
breed	1.51	barley	0.88	depend	0.96	greatly	1.47	pools	0.92
hardy	1.39	oats	0.82	occupants	0.89	trenching	1.36	char	0.84
pasturing	1.35	weeds	0.8	prefer	0.85	tenantry	1.28	fisheries	0.82

[44]On the Scottish agricultural revolution, see David Turnock, *Patterns of Highland Development* (London: Macmillan, 1970); E.S. Richards, *A History of the Highland Clearances: Agrarian Transformation and the Evictions 1746-1886* (London: Routledge, 1982); Devine, *The Transformation of Rural Scotland.*

livestock & produce				management				fishing	
flocks	1.32	vegetables	0.54	tacksmen	0.83	implements	1.21	poaching	0.76
goats	1.28	clover	0.46	cottars	0.81	altered	1.14	mackerel	0.73
pasturage	1.12	potato	0.43	landlords	0.78	improved	1.08	stored	0.73
devoted	1.04	seeds	0.4	encouraged	0.73	enclosures	0.9	caught	0.72
preferred	0.96	grasses	0.37	cottar	0.7	methods	0.83	rod	0.71
breeds	0.8	yielding	0.3	pendicles	0.65	encouraged	0.73	stake	0.61
reared	0.73	hay	0.15	lots	0.6	unproductive	0.71	flounder	0.51

Table 4. Select positive geocorrelations between Ossianic myth and terms related to agriculture and freshwater fishing. Topical keywords selected based on their semantic similarity to *sheep*, *cabbages*, *tenants*, *improved*, and *salmon*, respectively.

industrial production				coal and steam			
printfield	-1.28	calico	-1.39	splint	-1.38	foundries	-1.32
bleach	-1.08	silk	-1.3	hurlet	-1.37	engines	-1.15
printfields	-1	yarns	-1.29	cannel	-1.35	collieries	-1.09
bleachfields	-0.97	printing	-1.18	colliery	-1.32	propelled	-1.09
bleachfield	-0.94	worsted	-1.15	parrot	-1.23	engine	-0.93
threshing	-0.83	carpets	-1.14	seam	-1.15	wheel	-0.91
carding	-0.68	bleaching	-1.07	collieries	-1.09	packets	-0.89
snuff	-0.68	tannery	-1	workable	-1.04	factory	-0.86
breweries	-0.62	printfields	-1	metals	-1.01	furnaces	-0.81
thirlage	-0.62	bleachfields	-0.97	ell	-0.99	spindles	-0.78
multure	-0.6	bleachfield	-0.94	workings	-0.94	workers	-0.77
brewery	-0.57	factory	-0.86	pits	-0.93	machine	-0.71
flour	-0.55	spindles	-0.78	mining	-0.89	factories	-0.62
multures	-0.55	workers	-0.77	ironstone	-0.87	invention	-0.51
thrashing	-0.52	cotton	-0.76	seams	-0.78	thrashing	-0.51

Table 5. Select negative geocorrelations between Ossianic myth and terms related to industry. Topical keywords selected based on their semantic similarity to *mills*, *factories*, *coal*, and *engines*, respectively.

Key terms of Scottish technological development tend to correlate negatively with Ossianic myth. (See Table 5.) Scottish linens were produced in mills surrounded by large *bleachfields* and *printfields* where fabric was laid in the sun to wash its color. Grains were ground into *flour* and processed in *breweries*. Scottish factories became known over the nineteenth century for specializing in sophisticated fabrics like *silk*, *calico*, and *worsted*, while the *mining* industry stripped its landscape for *coal*. (The terms *splint*, *cannel*, and *parrot* all name kinds of coal, and *hurlet* was the name of a small, now-abandoned mining village near Glasgow.) In the second quarter of the nineteenth century, steam *engines* came to permeate Scottish manufacturing, but they did so largely in the main industrial regions, so most terms relating to the emergent fossil-fuel technologies correlate negatively with words like *fingal* and *ossian*.

This network of geospatial relations can be seen, too, by examining industrial terms that push against this trend by correlating positively with Ossianic references. (See Table 6.) When subsetted in this way, the semantic coherence of

industrial categories largely breaks down and what's left is the relatively narrow conceptual overlap that connects them to Ossianic space. Reading backwards from right to left, terms relating to *engines* lose the specificity of steam power, and the associations that remain have to do more generally with motion and kinetic energy. The discourse of coal mining was almost totally absent from Ossianic space, though mineralogical descriptions carefully catalogued local rock formations. Discussions of factories in towns where Ossian's poetry is mentioned tend to focus on problems with managing labor and maintaining profitability, but those discussions were relatively few in number and appear only obliquely through the data.

industrial production				coal and steam			
waterfalls	2.72	dwellings	1.62	vain	1.37	moved	1.21
turns	1.63	whisky	1.37	sought	1.24	constructing	1.17
driven	1.41	convenience	1.32	schistus	1.23	pressure	1.03
serviceable	1.34	moved	1.21	burning	1.19	power	0.94
moved	1.21	occupation	1.19	import	1.15	successfully	0.92
servitude	1.03	temporary	1.17	interruptions	1.11	sails	0.64
corn	0.73	exclusively	1.01	resembling	1.01	moving	0.62
saw	0.65	neighbour	0.81	slates	0.93	ply	0.57
moving	0.62	superseded	0.54	occurring	0.9	preparing	0.44
paper	0.49	department	0.44	worked	0.6	shoe	0.37
mills	0.49	blankets	0.44	basalt	0.6	conveying	0.32
mill	0.46	tile	0.35	attempts	0.59	horse	0.3
leather	0.26	stockings	0.33	limestone	0.53	touch	0.21
rolling	0.22	leather	0.26	discover	0.4	supplying	0.19
supplying	0.19	supplying	0.19	parting	0.39	driving	0.18

Table 6. Select positive geocorrelations between Ossianic myth and terms related to industry. Topical keywords selected based on their semantic similarity to *mills*, *factory*, *coal*, and *engines*, respectively.

Most interesting in this context are terms associated with *mill* and *mills* (far left column of Table 6), which, alone among references to industrialization, actually maintain semantic coherence in the context of Ossianic space. Of the 126 places with descriptions that mention either Fingal or Ossian, roughly half (62) also describe mills. Scotland's mills were crucial to its industrial development: they processed cotton and woolen textiles, paper, lumber, and grains (making flour or rolling oatmeal). Terms related to hydropowered mechanical production that correlate with *fingal* and *ossian* include *waterfalls*, *turns*, *driven*, *corn*, *saw*, *moving*, *paper*, *mills*, and *mill*. During industrialization's early phase in the later eighteenth and early nineteenth centuries, these mills depended upon hydropower and were therefore built alongside Scottish rivers, appearing in the counties that surround the main commercial center of Glasgow, such as Renfrewshire, Lanarkshire, and Ayrshire, but also appearing in regions further to the north, like Perth and Argyll, and even all the way to Uist in the Outer Hebrides.[45] It was in places

---

[45]John Shaw, *Water Power in Scotland 1550-1870* (Edinburgh: John Donald, 1984).

like these where parish ministers and local historians were likely to find echoes of Fingalian myth.

While Ossianic geographies tended to skew towards the Highlands away from the industries concentrated in Scotland's Central Belt, semantic patterns of both Gaelic tradition and early manufacturing permeated both regions. Later, over the course of the nineteenth century, the transition to coal-fired steam power centralized production, transport, and population in Scotland's cities.[46] But for much of the eighteenth and early nineteenth centuries, water-powered mills spread into villages, towns, and parishes, often in the same places where local antiquaries found traces of Scotland's mythic past. The bardic nationalism of Gaelic tradition coincided spatially and temporally with agricultural improvements, expanding infrastructure, hydropower, and manufacturing throughout the Scottish countryside. Thus, although we can say in general that Ossianic myth was associated with places perceived to be outside the sphere of industrial modernity, our analysis highlights a risk of anachronism in such generalizations. Whereas the idealized landscapes of antiquity tended to appear outside the cities and away from the factories and coal mines that epitomized the industrial age, those landscapes also draped in natural beauty the rivers and waterfalls where hydropowered mills processed corn, lumber, paper, and cotton. References to Fingal and Ossian actually correlate positively with the modes and technologies of industrial production that were prevalent during Macpherson's lifetime and during the height of antiquarian debate over the Ossianic tradition.

Case Study (Part II): Fingal and History

The above discussion raises an issue we have not yet addressed in detail: change over time. Scottish topographical dictionaries, gazetteers, and statistical accounts offer a wealth of historical details. The events surrounding local ruins and antiquities were carefully noted; battlefields were marked and often preserved; the political conflicts among the clans and Scottish nobles recounted. But chorographical writing also captures smaller events of local importance: the death of a particular parish minister, for example, or the date a neighborhood church was erected, or the year a town's post office first opened for business. Geographical description is a reminiscent genre through which places take shape as palimpsests of

---

[46] Andreas Malm has argued that this transition had less to do with any technological or financial advantage of steam power over hydropower, and more to do with economic factors that privileged a concentrated and regulated labor force. See Malm, *Fossil Capital.* See also G.N. von Tunzelmann, *Steam Power and British Industrialization to 1860* (Oxford: Clarendon, 1978).

local memory. As Mark Salber Phillips has observed in his study of Sinclair's *Statistical Account*, anticipations of future improvement are persistently measured not only against the present state of the parishes, but within localized accounts of social and economic change over time.[47] We hoped to capture some of this historical richness by mining the text for references to the past. Using regular expressions, we parsed the place descriptions by sentence and filtered them, keeping only those that refer to a specific year or century.[48] Of the 6.6 million words in the corpus, excluding stopwords, 1.2 million (19.1%) occur in such sentences. Overall we found more than 78,000 dated sentences, allowing us to measure the distribution of the collection's vocabulary over time. Whereas our primary analysis draws from the word-place matrix described above, our discussion in this section draws from what could be called a "word-decade matrix."

---

[47] Mark Salber Phillips, *On Historical Distance* (Yale UP, 2013), 97-114. Phillips argues that this difference in historical representation reflects the conflicting purposes of Sinclair and his many contributors: "From its editor's standpoint, the *Statistical Account* was designed as a grand inventory of the nation, and Sinclair looked forward to the day when the information he had compiled would be condensed into a series of tables and charts. But when we turn to the work of the nine hundred authors, we find a decidedly more historical vision. … [W]hat shines through the vast majority of the responses is that the ministers took Sinclair's invitation as an occasion to write the history of everyday life as it was reflected in the social and economic progress of their own parishes" (16).

[48] Our parsing algorithm began by dividing each entry whenever it discovered the pattern of a period followed by a space followed by a capital letter. Once parsed, we looked for sentences with four-character words that began with the number 1, removing any in which the year word was not preceded by a custom list of flag words that typically preceded years in references to specific dates throughout our corpus. We then searched over the parsed sentences again, adding any that included the words "century" or "centuries," and assigned them a date based on the word that preceded "century." When multiple years were mentioned in a sentence, we assigned that sentence with the mean of those years, and references to centuries were assigned to the middle of the century (i.e., references to "the seventeenth century" were assigned the value of 1650), sometimes causing spikes at the mid-century points in time-series graphs for words that were often included in sentences that refer to centuries. All references to specific years before 1000 are lost by this process, and in the time series graphs we use in this essay all references to centuries before 1000 are assigned the value of 1000. Through this process, 78,009 dated sentences were identified, a number small enough to allow human review, and we have found very few instances of sentences that were erroneously dated.
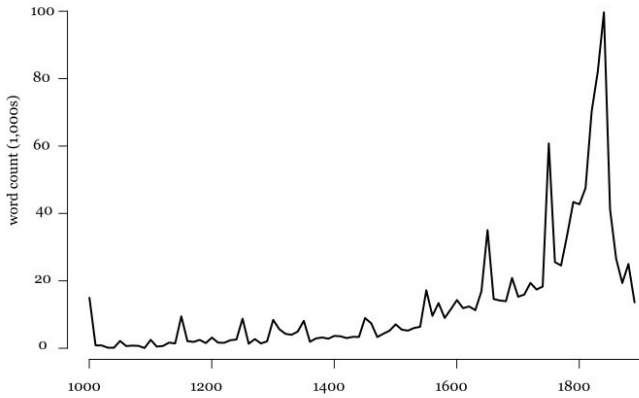
Figure 9.1. Word frequency by decade over the entire corpus. Note that all references to dates prior to 1,000 AD are gathered into a single point.
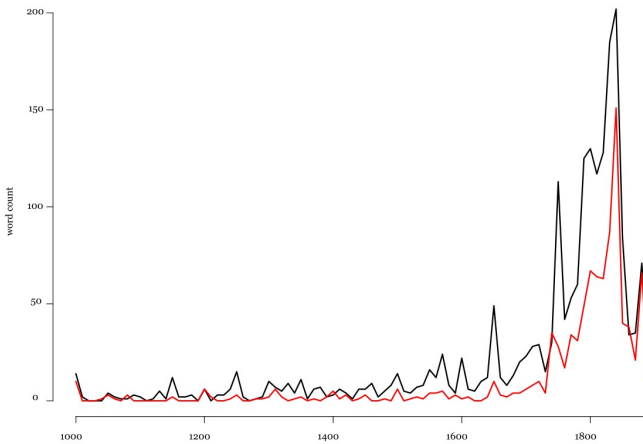

Figure 9.2. Frequency by decade for the words related to water-powered mills (black) and steam-powered engines (red). Values displayed represent the sum of decade totals for keywords highlighted in Table 6, concerning hydropower, and Table 5, concerning steam engines.

To understand how Scottish geographical writing represents its past, we examine this subset. Most dated sentences refer to events from the then-recent past, around when the Old and New Statistical Accounts of Scotland were first published. Nearly a third of the total dated word count refers to things that hap-

pened between 1790 and 1840. (See Figure 9.) Visualized as a time-series graph, the distribution of word frequencies looks like a pulsating wave that slowly gains in strength until it crashes into the mid-nineteenth century. Terms of industry follow a similar pattern and experience sharp spikes in the eighteenth and nineteenth centuries. Words related to hydropower rose during the mid-1700s and were followed a few decades later by references to steam power, drawing in ghostly silhouette the transition from hydropower to coal. Such patterns are common throughout the corpus, and indeed most high-frequency words, considered individually, also follow the same arc of modernity. However, some words display sharply divergent patterns, especially when they were associated with specific historical events. (See Figure 10.) For example, the words *bruce*, *mary*, and *jacobite*, in the context of dated sentences, refer almost exclusively to King Robert I (known as Robert the Bruce), Queen Mary I, and the Jacobite Rebellion, respectively. Sentences using words like these can be located very precisely within a fixed chronology, and matrix operations can be used to find conceptually coherent topics that organize historical events. These two semantic patterns — localized social and economic developments on the one hand and national chronicle on the other — may be graphed so as to visualize two orders of time informing these genres of national improvement.
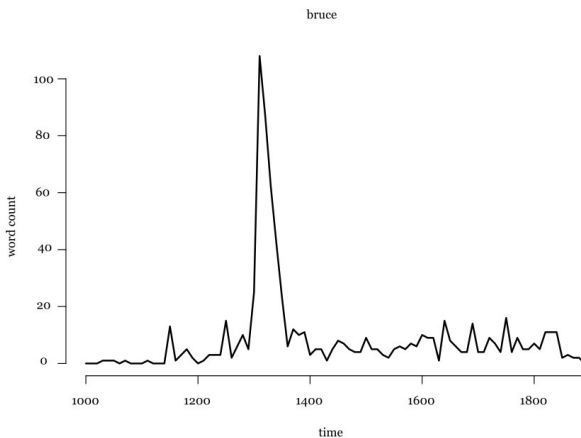

Figure 10.1. Frequency by decade for the word bruce. Historical figures like the Earl of Bruce and Queen Mary appear clearly in the data, as do events like the Jacobite Rebellion.
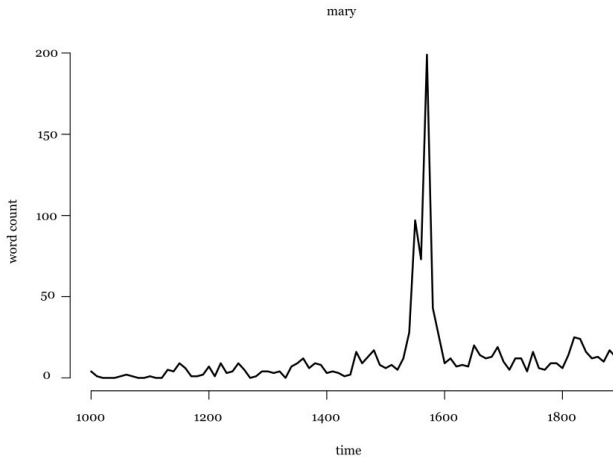
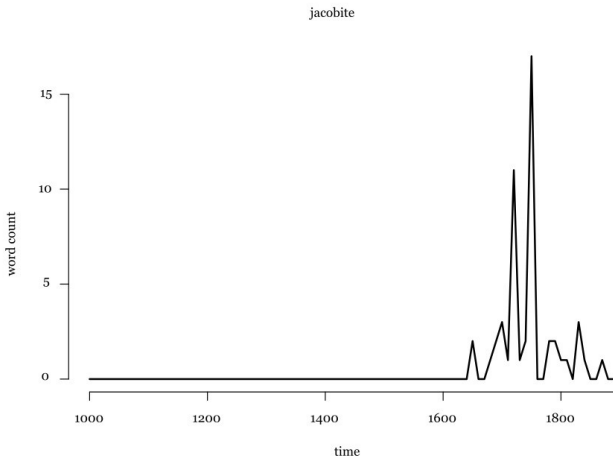Figure 10.2. Frequency by decade for the word mary.

Figure 10.3. Frequency by decade for the word jacobite.

Ossianic time, however, assumes a shape very different from those of modernity or historicity. Antiquarian speculations in these dictionaries and gazetteers repeatedly locate ruins, burial mounds, and monuments, but the actions of Macpherson's Ossianic poetry seem to exist outside of time and space. This tension between the abstractions of the poetry and the desires of both local

inhabitants and travelling tourists was a defining element of the Ossianic phenomenon and only heightened the poetry's simultaneous appeals to local pride and national identity. Cordula Lemke has recently observed that "the Scottish warriors' rootedness in a state of 'always already' provides a background to an ideal *tabula rasa* for nostalgic longing."[49] Ian Duncan has argued that such estrangement from historical time is central to the poetry's power for national myth. In Duncan's words, "historical extinction, sublimated into an aesthetic atmosphere, becomes essential to the poetic constitution of a modern national culture."[50] Accordingly, we find that the words *fingal* and *ossian* are less likely than average to be attached to the timeline of Scottish history. Of the 345 times either word is used, only 41 (12%) appear in sentences that refer to specific years or centuries, which is below average for the corpus. (See Figure 11.) Of these, some refer to the ancient past of Fingalian epic, referring to real events that were supposed to have happened during earlier centuries. Others follow the normal trajectory of modern description, because the sentences refer to the efforts of antiquarians who gathered Ossianic folklore, to the publication dates of books related to their work, or to other events that happened in places defined by Gaelic tradition. As a result, references to Fingal and Ossian exhibit a two-humped structure like that of Bactrian camel, with most activity appearing at either end of the chronology. References to the long periods we know as the Middle Ages and Renaissance were almost completely devoid of Ossianic vocabulary, and so in effect the discourse bridged modernity and antiquity while skipping over the political history of Scotland's intrigues and wars. Words that exhibit a similarly polymorphous temporality include terms like *irish* and *pictish* that were also invoked when chroniclers reflected upon the deep history of ancient peoples from whom the Scots as such had sprung. While Phillips is generally correct in his observation that "the narrative of Scotland's social and economic progress" is mapped in this literature by "the many tributary themes" of localized historical memory, Ossianic references are more than offhand antiquarian speculation. They can be seen to join the localized histories of regional topography with the national imperatives of political economy through an amorphous but productive pattern of mythic time.[51]

---

[49] Cordula Lemke, "Nostalgic Ossian and the Transcreation of the Scottish Nation," in Moore, *International Companion*, 52-64, 58.

[50] Ian Duncan, "Spawn of Ossian," in Evan Gottlieb, ed. *Global Romanticism: Origins, Orientations, and Engagements, 1760-1820* (Bucknell UP, 2015), 3-18, 8.

[51] Phillips, *On Historical Distance*, 106.

fingal & ossian

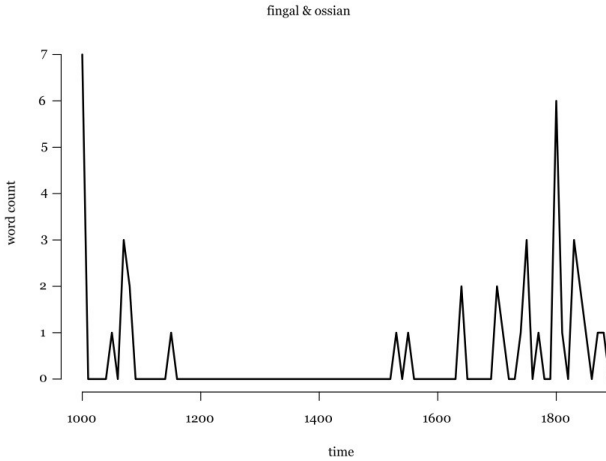

Figure 11.1. Frequency by decade for the words fingal and ossian (combined). Words that invoked the ancient past often leapt over the tumultuous events of Scotland's political history.
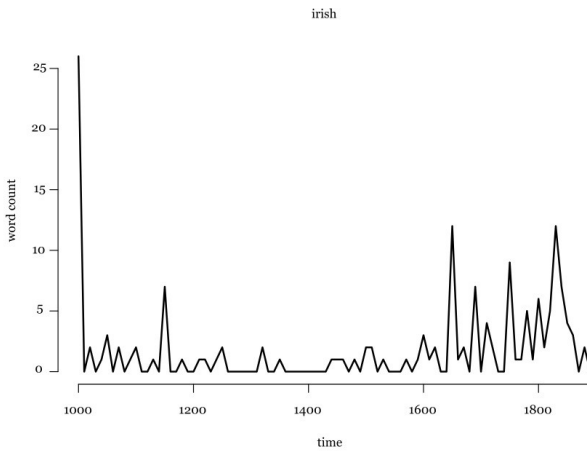
irish

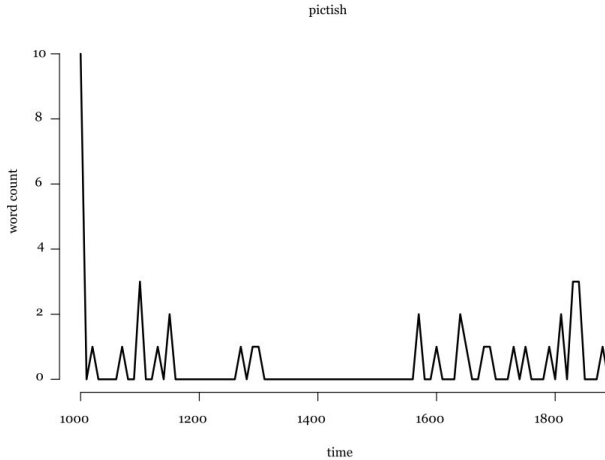Figure 11.2. Frequency by decade for the word irish.

pictish

Figure 11.3. Frequency by decade for the word pictish.

Conclusion

In this essay, we have demonstrated a set of techniques for geospatial text analysis. Our primary goal has been to demonstrate the validity of geospatial semantics and to offer a glimpse of its promise for historical research. However, we left many questions open. Some are narrow and specific while others are more general. How did choices we made in data processing affect the outcome? What if we had used different statistical measurements? To what extent does our analysis depend on the specificity of topographical description as a genre? Will other kinds of writing exhibit similar patterns? We noted above that semantic and spatial distances in our corpus correlate with an average Pearson coefficient of .2785. Is this a high number or a low number? How geocorrelative are newspapers, poems, and novels, and how do they change over time? What about fictional spaces? Would the methods described here successfully differentiate places in fantasy worlds like J. R. R. Tolkein's *Lord of the Rings* series, often taken as a descendant of Macpherson's bardic enterprise? What about time-based media like television and film? Finally, and perhaps most importantly, how do various genres respond to changes in physical environments? What is the imprint of place on human discourse, and vice versa?

Such questions highlight the need for better data collection and more sophisticated mathematical models. Our analysis in this essay has been kept deliberately simple, but further study might borrow more heavily from quantitative geography and might employ a wider array of spatial statistics. The same holds for our linguistic analysis. We chose to study Ossianic myth in part because the terms *fingal* and *ossian* are unusually strong topical markers: a more diffuse topic would require more sophisticated techniques for linguistic study, such as latent-semantic indexing or topic modeling. And because historical geospatial semantics will be of most potential interest to historians, the method will require a more robust vocabulary for describing change over time.

We have also suggested the unique applicability of these methods to the study of literature and environmental history. A cross-sectional and multi-scalar approach to language can provide useful insight into the cultural apprehension and technological reordering of ecologies and economies. These methods can help orient historical studies of language and literature by making explicit their underlying spatial components. Fredrik Albritton Jonsson has underscored how eighteenth-century methods of measuring demographic and material resources in the Scottish Highlands established a foundation for ecological inquiry and a mandate for sustainable environmental management. He has argued that "conservationist sentiment and aesthetic value were of less importance" for developing policies of proper stewardship "than the power to quantify the material basis of the economy."[52] The power to quantify the geosemantic structure of the printed record can provide analogous contributions to the cultural history of the environment. Our challenge moving forward will be to create data and models adequate to this task.

---

[52] Jonsson, *Enlightenment's Frontier*, 8.