

**Exploring Qualitative Geographies in Large Volumes of Digital Text: Placing
Tourists, Travellers and Inhabitants in the English Lake District**

Ian N. Gregory, Department of History, Lancaster University, UK

Robert Smail, Independent Researcher, UK

Joanna E. Taylor, Department of English, American Studies, and Creative Writing,
University of Manchester, UK

James O. Butler, Department of Archaeology, Classics and Egyptology, University of
Liverpool, UK

**This is an original manuscript of an article currently (May 2024) in press with Taylor & Francis
for publication in *Annals of the Association of American Geographers*, to be available at:**

<https://www.tandfonline.com/toc/raag20/current>

© The authors 2024. This work is openly licensed via CC-BY 4.0

Abstract:

Hitherto, approaches to analysing geographies in large collections of digital texts have used coordinate-based locations derived from toponyms (place-names). While this approach provides a useful starting point, it ignores both non-toponym references to place and textual structure, while stressing Euclidean space. This paper addresses these issues. Focussing on a corpus of writing about the English Lake District before 1900, it explores the different geographies associated with tourists, travellers and inhabitants. Alongside toponyms, we identify the concept of a geo-noun to associate place with geographical features. Sense of place is then identified using the adjectives, nouns and verbs that writers associate with locations and locales. This methodology allows us to identify and define places described in large volumes of digital text both holistically and textually.

Keywords: place; travel writing; geographical text analysis; digital texts, spatial networks
English Lake District

Introduction

In his *Guide through the District of the Lakes* (1835), William Wordsworth divides people in the Lake District into three main groups: tourists, travellers and inhabitants. Wordsworth's *Guide* worried that 'tourists', speeding from one site to another, would neglect to connect with the landscape. More discerning 'travellers' could be expected to learn how to appreciate the place's complexities and nuance. The people Wordsworth defined as 'inhabitants' could fully read the landscape, discerning hidden messages in its soil and rocks and vegetation. These different responses continue to have social and ecological implications (Hess 2012), but they also impact on a place's cultural identities.

This paper explores how these labels – tourist, traveller, and inhabitant – affect our understanding of the pre-twentieth-century Lake District’s cultural geographies. Using Agnew’s (2011) work as a touchstone, we explore how spatio-textual analyses can uncover relationships between tripartite aspects of place: location (where an object or activity is located), locale (the setting in which activities take place, or the ‘material things’ that constitute somewhere [Cresswell 2004, 7]), and sense of place (the intangible elements that make a place unique). Specifically, we explore how Lake District writers described places for different types of people.

Previous work has used geographical information systems (GIS) to investigate differences between tourists, travellers and inhabitants (Taylor and Gregory 2022). However, the intrinsic nature of GIS, with its reliance on tabular attribute data linked to precisely defined coordinate-based spatial data (Bolstad 2019; Longley et al. 2015), means that using GIS for qualitative research based on textual sources is complex. As Sally Bushell observes, geographical information in sources like literature is ‘slippery’; even where it seems to denote real-world places, textual spaces are often blended with something unmappable in a Cartesian sense (Bushell 2012). There have been many calls to make GIS more relevant to a broader range of sources, specifically those qualitative datasets for which precise spatial locations may be either unavailable or unsatisfactory. To date, progress in advancing alternative approaches has been limited (see, for example, Bergmann and Lally 2021; Cope and Elwood 2009; Gao 2022; Goodchild 2011; Purves, Winter, and Kuhn 2019). This paper responds to the call for a ‘mixed model that combines GIS mapping with other forms of mapping to permit greater complexity’ (Bushell et al. 2022, 13). We aim to create a representation of place as an aggregate concept that can be made up of locations taken from toponyms (place-names) and locales taken from geographical features, combined with the authors’ descriptions to give a more complex sense of place.

[Figure 1 (a, b and c) – tourist, traveller and inhabitant PNCs near here]

To do so, we draw on and develop geographical text analysis (GTA), which explores how toponyms relate to other terms in the text. GTA uses methods from geographical information science (GISc), natural language processing (NLP), and corpus linguistics to identify and analyse the geographies described in large textual corpora. GTA starts by geoparsing the corpus to identify toponyms and allocate them to a coordinate (Gregory et al. 2015). Next, it uses place-name co-occurrences (PNCs) to identify search-terms within a set number of words of each toponym. For example, Figure 1 charts the toponyms that co-occur within ten words of ‘tourist’, ‘traveller’ or ‘inhabitant’. The maps are summarised using density smoothing to highlight clusters of PNCs. The maps show the differences in the geographies associated with these terms. Tourist PNCs cluster around major settlements and transport routes. The pattern for travellers also has clear clusters, but these are more dispersed. The PNCs for inhabitants, by contrast, do not cluster in any meaningful way (Taylor and Gregory 2022). GTA can provide a meaningful overview of which locations are being associated with particular themes or peoples. It has been used to explore Lake District cultural geographies (Donaldson, Gregory, and Taylor 2017), geographies of poverty in modern British newspapers (Paterson and Gregory 2019), nineteenth-century disease (Murrieta-Flores et al. 2015; Porter, Atkinson, and Gregory 2015), and geographies of Early Colonial Mexico (Jiménez-Badillo et al. 2021).

While GTA begins to enable textual geographies to be explored and analysed, as currently conceived it has several limitations. Most seriously, its definition of place is limited to toponyms for which coordinates can be found. It usually renders all place types as points, which is unhelpful for extensive features such as administrative zones or lakes. GTA also removes the narrative structure from the texts, showing where toponyms are near each other on the map but not whether they are close textually. Finally, GTA makes extensive use of

spatial analysis techniques that use Euclidean space (Fotheringham, Brunsdon, and Charlton 2000; Lloyd 2010), which may not be the way writers perceive places and the spaces between them.

Here, we present an extension of GTA that moves away from place as toponym represented by coordinates in Euclidean space to using networked space to represent the relative geographies writers describe in their texts. Place becomes an aggregate concept based on clusters of geographical features that co-occur in the texts, alongside the sense of place described by the writers. Using the Corpus of Lake District Writing (CLDW) we explore the various places that writers associate with Wordsworth's three groups – tourists, travellers, and inhabitants – to ask how these terms affected historical understandings of the Lake District.

b. Background and Methods

i. Locating the Lake District and Corpus of Lake District Writing

[Figure 2: Background map near here]

The Lake District, summarised in Figure 2, is a mountainous region in North-West England. In his *Guide*, Wordsworth exhorted his readers to imagine 'the main outlines of the country' as if they were positioned on a cloud between Scafell and Great Gable in the central Lakes. From there, the reader would see spread below them 'a number of valleys, not fewer than eight' appearing 'like spokes from the nave of a wheel.' On the eastern side, the hills running south from Ullswater to Windermere complete the wheel. This area became a national park in 1951, and a UNESCO World Heritage Site in 2017. It is relatively small, measuring around 40 by 30 miles. There are over 200 fells, although only four (Scafell, Scafell Pike, Helvellyn, and Skiddaw) are over 900 metres. The region also contains sixteen major lakes and numerous smaller ones, including the largest (Windermere) and deepest (Wast Water) lakes

in England. The largest towns (Penrith, Kendal, Lancaster, and Carlisle) are all beyond the National Park's boundaries, and the major settlements within it all lie along the main north-south transport route. This road, today's A591, passes from Bowness-on-Windermere, through Ambleside and Grasmere, over the pass of Dunmail Raise, and on to Keswick.

The Lakes have a rich literary heritage; since the late seventeenth century, works by writers including Celia Fiennes (1688), Daniel Defoe (1727), Thomas Gray (1775) and Thomas West (1778) have popularised them as a visitor destination. These early advocates of Lakeland travel were followed by the Lake Poets, notably William and Dorothy Wordsworth, Samuel Taylor Coleridge, and Robert Southey, writing in the late eighteenth and early nineteenth centuries. While authors like these celebrated Lakeland life, they also contributed to its transformation. During the nineteenth century, local industries such as farming, mining, quarrying, charcoal burning, and textiles were overshadowed by the expansion of tourist infrastructure (Horn 1980). The railway's arrival in the mid nineteenth century confirmed the Lake District as a popular tourist destination (Yoshikawa 2020).

The Corpus of Lake District Writing (CLDW) comprises 80 texts dating from 1622-1900, spanning this period of growth in the region's tourist industry (Taylor and Gregory 2022, 7-10, Rayson et al. 2017).¹ The CLDW is predominantly made up of tourist guides and travel narratives, combining well-known authors like William Wordsworth and Harriet Martineau with lesser-known figures, including Edwin Waugh and Priscilla Wakefield, alongside anonymous publications, such as *Black's Shilling Guide to the English Lakes* (Anon 1853; Baddeley 1900). Together, these works reveal how the Lake District developed as a cultural landscape. Its descriptions of place are rich and multi-layered, and it is important that an analysis of senses of place is similarly complex. In traditional literary studies, sense of place

¹ See: <https://github.com/UCREL/LakeDistrictCorpus>

would be inferred by close reading individual texts and comparing examples. Developing an understanding of senses of place from a large collection of texts, however, requires a different kind of layering, combining concepts of place to incorporate the personal and emotional alongside the geographical.

The CLDW's 1.5 million words were transcribed with minimal markup, before toponyms were tagged using the Edinburgh Geoparser. This automated process uses Part-Of-Speech (POS) tagging to identify proper nouns that are candidates to be toponyms. Candidate toponyms are compared against a place-name gazetteer which, in its simplest form, is a list of place-names with their coordinate-based locations (Southall, Mostern, and Berman 2011; Berman, Mostern, and Southall 2016). Where there is a match between the candidate toponym and the gazetteer, we assume that the candidate is a toponym and it can be allocated the appropriate coordinate. Additional contextual information may be used to improve on this (Grover et al. 2010).

With a corpus as rich in geographical description as the CLDW, geoparsing is an error-prone operation. To improve accuracy, two further stages were added. The first used 'concordance geoparsing,' whereby subsets of the corpus were manually checked and errors corrected and recorded. This was repeated for additional subsets, with existing errors automatically corrected and additional errors being corrected and recorded. Once the majority of errors had been identified, the corrections could be applied to the entire corpus (Rupp et al. 2015).

Manual checking and editing was also performed on the entire corpus (Butler et al. 2017).

Despite this extensive process, problems remain, particularly with ambiguity. For example, it may not be clear whether 'Windermere' or 'Coniston' refer to lakes or settlements.

Additionally, there are issues with how to represent features like counties, rivers and major lakes that cannot be well represented using points. Alongside toponyms, elements like adjectives, nouns (excluding toponyms and other proper nouns) and verbs were also

identified using POS-tagging (Hardie 2007; Voutilainen 2003) implemented through the Spacy Natural Language Processing toolkit (Spacy 2022).

These tasks resulted in a collection of texts that are useful for toponym-based analysis. Once spellings had been standardised, 3,965 unique toponyms were identified, of which 3,243 could be geoparsed. These gave a total of 43,254 toponym instances, with an average of 0.73 toponyms per sentence. Of the geolocated toponyms, 1,656 (51.1% of unique toponyms) referred to locations in or near the Lake District, providing a total of 35,490 instances (82.1% of total instances). Frequently occurring toponyms tend to be found in and around the Lake District, while the many named locations outside of the Lakes tend to have low frequencies. While these toponyms provide a rich field for analysis, writers in the CLDW do not only define place simply through toponym-based locations. To understand how the Lake District also functions as a series of locales, we need also to work with more abstract notions of place.

ii. Identifying locales with geographic feature nouns

In Agnew's understanding, location is complemented by locale which encompass the 'settings where everyday life activities take place', alongside the 'where of social life and environmental transformation' (Agnew 2011). Rather than being represented by toponyms, locales are concerned with types of features that affect the communities who live at or visit somewhere. To identify these, we developed the concept of 'geo-nouns:' nouns that refer to physical or human geographical features. These features were defined using a list of 'seed' terms (e.g. 'lake', 'mountain', 'river', 'and 'town'). Collocation analysis (Hunston 2002; McEnery and Hardie 2011) was then used to identify words POS-tagged as common nouns found significantly more frequently than expected within five words of each seed term. The list was manually examined to identify nouns referring to geographical features. Once a provisional list of geo-nouns had been identified, the process was repeated with the additional

terms until no new geo-nouns were added. This process revealed a compact list of geo-nouns used to describe Lakeland locales. In total 153 unique geo-nouns (including lexemes) were identified, occurring a total of 75,928 times in the corpus, averaging 1.35 times per sentence.

Unlike geolocated toponyms, we cannot georeference geo-nouns. Thus, to move beyond toponym-based analysis we need to introduce new, non-Euclidean, forms of visualisation. Additionally, as observed above, Euclidean mapping of toponyms imposes a space-based representation of geography that does not express the more qualitative understanding of place experienced and described by writers. In this paper, we turn away from conventional mapping – which stresses spatial proximity – towards a network analysis approach that highlights the places that writers associate with each other.

iii. Visualising senses of place

A network consists of nodes, represented using points, and edges, the lines that join associated nodes together (Barthélemy 2011; Newman 2018). Social network analysis is the best-established form of network analysis, wherein nodes typically represent people and edges indicate connectedness, such as being related or having communicated in some identifiable way. Examples include Ahnert and Ahnert's (2015) use of a corpus of letters to conduct a social network analysis of Protestants in Early Modern England based on whether the individuals have written to each other, and Wilson, Buchnea, and Tilba's (2018) investigation into business relationships based on whether directors sat on the boards of the same firms. The use of network analysis to investigate geographical data has, to date, been relatively limited. Golitko et al. (2012) use it to analyse Mayan obsidian supply; Bernardini and Peebles (2015) explore Puebloan villages in the south-western United States; Gidal and Gavin (2019) use it to explore Scotland's historical postal network; and Bushell et al. (2022) use a form of network analysis to investigate geographical elements in literary texts.

Writers rarely describe a place using a single toponym or geo-noun. Instead, place descriptions are built up through clusters of co-occurring words. Here, for instance, Samuel Taylor Coleridge (1802) describes the hamlet of Ulpha:

I found myself coming down upon Ulpha again, about a mile above the House & Kirk which I had just quitted / however I was not sorry, to have another view of that lovely Place, and it brought me in full sight of a fine water fall on the opposite Hill on the other side of the Duddon, seemingly a short mile above the Kirk / I saw Houses to my right, & an [old] Man with his Daughter, a sweet Girl, burning Bracken[.]

The hamlet is more than only the houses and the church, or even the hill and the Duddon. To these geographical features Coleridge adds his opinions ('lovely', 'fine') and describes how his understanding of the place is inflected by the people who live there. Defining aggregate places requires us to identify toponyms or geo-nouns that are frequently associated with each other, or co-occur (McEnery and Hardy 2011), alongside words or phrases indicating perceptions of that place. For the purposes of this paper, we define all words that occur within the same sentence as co-occurring. Nodes represent toponyms or geo-nouns; an edge forms where two of these co-occur. For example, focussing on toponyms, the sentence 'I began therefore to calculate whether it was practicable to proceed from Coniston Water to the head of Great Langdale, go up the mountain from that place, and afterwards ride to Ambleside in the evening.' (Baines 1829) gives three nodes: 'Coniston Water,' 'Great Langdale,' and 'Ambleside', and three edges linking Coniston Water to Great Langdale, Coniston Water to Ambleside, and Great Langdale to Ambleside. The only geo-noun in this sentence is 'mountain' so this is recorded as an instance of a node without any edges.

As with the relationship between mapping and spatial analysis, network analysis frequently requires statistical approaches to formally identify patterns. A cluster in a network can be

thought of as a community in which nodes are allocated to the group to which they are best connected. In this paper we used the modularity method proposed by Lambiotte, Delvenne, and Barahona (2009) to identify clusters of toponyms or geo-nouns that frequently co-occur. Clusters defined in this way are felt to be distinct places as defined by the writers themselves regardless of whether they are in close spatial proximity. Clusters of toponyms define location-based places, while clusters of geo-nouns define locale-based places.

To complete Agnew's tripartite definition of place we also require sense of place. Sense of place differs from the other two elements because, as we see with the example from Coleridge's tour above, it cannot exist independently but must instead be linked to one or both of location or locale. Sense of place can be derived from the POS tags that co-occur with toponyms or geo-nouns. Three types of words are particularly helpful in defining sense of place: verbs (excluding auxiliary verbs) provide information about what occurs at a place; nouns (other than proper nouns and geo-nouns) tell us what other features are associated with the place; and adjectives tell us much about how the place is perceived.

We are interested in adjectives, nouns and verbs that co-occur significantly more than expected with our groups of Lakeland people and with places defined using clusters of toponyms or geo-nouns. This requires a measure similar to place-name co-occurrence keywords (Paterson & Gregory 2019). For example, if we are interested identifying the sense of place words associated with tourists for a particular set of toponyms, we count the instances of the sense of place words that co-occur with both 'tourist[s]' and one of our toponyms. Log-likelihood is then used to compare this with the number of times each of these sense of place word co-occurs with 'tourist[s]' and any other toponym. In this way the statistically significant sense of place words can be identified.

c. Analysis: Tourists, travellers, and inhabitants

i. Overview

[Table 1 near here]

When Wordsworth distinguished between tourists, travellers, and inhabitants in his *Guide*, his aim was to encourage the tourist away from superficial encounters and towards the discernment associated with the traveller. The most sensitive and attuned travellers could, in turn, access the deep connection with the region that its inhabitants displayed. While other terms could be used to describe this last group, our use of ‘inhabitant(s)’ echoes the most common language from across the CLDW. In Wordsworth’s *Guide*, ‘inhabitant(s)’ is the most frequently used term to describe people who live in the Lake District, and throughout the corpus it is the most frequently used term to describe the local population; it occurs 406 times whereas resident, for example, occurs 49 times. ‘Inhabitant’ can, on occasion, refer to non-human residents too; Wordsworth refers to ‘native Scotch firs’ as ‘these old inhabitants’ for example, but these instances are rare and, where they do occur, indicate an intimate link between human and non-human residents that – as we will shortly see – is pertinent to this group’s perceived sense of place. Wordsworth’s ambition notwithstanding, in the CLDW these three groups remain distinct, and their differences can be modelled as a geographic phenomenon. Table 1 summarises the sentence-level co-occurrences between tourists, travellers, or inhabitants² and toponyms or geo-nouns.

We can see that the corpus refers to travellers more frequently than either of the other two groups: 596 sentences refer to travellers, compared with 400 for tourists and 410 for inhabitants. Superficially, tourists and travellers are both comparably associated with place; there are, on average, 3.48 instances of place-reference co-occurrences per sentence with tourists, and 3.63 with travellers. The difference lies in the type of place with which they are

² Note that for simplicity the words ‘tourist,’ ‘traveller’ and ‘inhabitant’ will always be assumed to include their plural form unless otherwise stated.

associated. Tourists are most closely associated with locations: 1.62 toponyms occur per tourist sentence, compared with 1.27 for travellers. Travellers, meanwhile, are more closely associated with locale; references to travellers co-occur with geo-nouns 2.36 times per sentence, whilst tourists and geo-noun instances co-occur in 1.86 sentences. The implication is that the CLDW directs tourists more carefully than it does travellers; in line with Buzard's (1993) interpretation of the two terms, the various route plans and itineraries direct tourists to particular locations, while travellers are expected to follow their own routes off the beaten track.

Inhabitants, by contrast, are much less linked to place with 2.77 instances per sentence. This is still higher than the background frequency across the corpus as a whole, but it is telling that this is noticeably lower than either tourists or travellers. Again, there is a clear difference in the type of reference to place. Inhabitants are associated with locales, co-occurring with geo-nouns 1.94 times per sentence (more than tourists, but less than travellers), but are rarely associated with location, with only 0.82 toponym instances per sentence. This low proportion indicates inhabitants' marginalisation from central Lakeland locations.

[Figure 3: Top 10 toponyms graph near here]

[Figure 4: Top 10 geo-nouns graph near here]

Figures 3 and 4 break this down further by summarising the rankings of the individual toponyms and geo-nouns most associated with the three groups. Despite the overall differences, the places associated with tourists and travellers are broadly similar with the same top five geo-nouns: 'road', 'lake', 'mountain', 'pass' and 'inn' (albeit in a slightly different order), and similar popular toponyms, primarily those along the most common tourist routes through the region such as Keswick, Ambleside and Windermere. Thus, although Table 1 showed that tourists are more associated with named locations and

travellers more associated with locales, Figures 3 and 4 show that within this the types of locations and locales associated with both groups are similar. The places associated with inhabitants are very different to the other two groups. They are largely associated with human locales: towns and houses, as well as (less frequently) villages and parishes. Perhaps even more telling is how the CLDW relates inhabitants to location: the popular toponyms associated with inhabitants tend to be associated with the major market towns around the Lake District's edges (Kendal and Penrith), administrative zones (Cumberland and Westmoreland), or locations outside the region's main centres (Ulpha). Thus, as well as being less associated with place than the other two groups, the places constructed for inhabitants, both as locations and locales, tend to be different. . For a better understanding of how each group practised place throughout the Lake District, though, we need to use our co-occurrences with network analysis to attend to the individual places that make up these patterns.

ii. Tourist places

[Figure 5: Tourist toponyms near here]

Figure 5 uses a network graph to identify the major tourist locations. The figure shows that Keswick and Ambleside are the most used toponyms for tourists and are also the most connected in the sense that they co-occur with each other in tourist sentences more than any other two toponyms. They, along with Grasmere, are also literally connected, by the main road through the Lakes. More generally, tourist locations are more heavily concentrated in the north, particularly areas within an easy day-trip distance from Keswick (including Borrowdale and Buttermere). There is another cluster around Ambleside (including Grasmere and Great Langdale). The larger towns of Penrith and Carlisle, main access points for the Lakes, also link to Keswick. It is also noticeable that the textual clusters created by

modularity also tend to cluster spatially; in other words, toponyms close to each other in the text also tend to be near each other on the map. This clustering provides further support for the idea that tourist guides are written for people who will travel from one location to the next along relatively accessible routes between them.

[Table 2 near here]

Table 2 summarises the main clusters from Figure 5. It provides the most common toponyms that define the cluster alongside a selection of the cluster's statistically significant geo-nouns and sense of place terms. Two types of place emerge: those with a practical function, and those with more aesthetic appeal. The Keswick-Penrith-Skiddaw and Windermere-Bowness-Kendal clusters both focus on the functional details of journey planning. The geo-nouns for both focus on modes of transport, including 'train', 'road', 'boat', 'ferry', and 'railway'. The 'l.', 'n.' and 'w.' come from the abbreviation for the London & North Western Railway Company that served Keswick. The sense of place words focus on facilities, such as accommodation and guides, as well as on directing the tourist around the area ('mile', 'right', 'proceed', 'turn', 'east'). While some of the other geo-nouns, such as 'stream' and 'hamlet', might suggest aesthetically pleasing features, these are not reinforced by appropriate sense of place words; 'interesting' is the only term of judgement in these clusters. In fact, close reading of the individual passages that form these clusters suggests that these features tend to be used as landmarks on journeys rather than features to be appreciated. Samuel Leigh's (1830) directions for the tourist exploring the tarns near to Wastwater is indicative; he describes how Sprinkling Tarn is 'connected with Sty Head Tarn by a stream about a mile long'. The stream is not of interest in its own right, but as a guide between the two features. Clusters representing popular day-trip routes, taking in Buttermere, Wast Water, Ennerdale, and Crummock Water in the north, the Langdales and Dungeon Ghyll Force in the central

Lakes, or Coniston in the south, show more interest in the aesthetic appeal of these places – though even here, much of the detail is in how to travel to the location. The tourist might take a carriage to a ‘fine’ prospect in the Langdales or – for the more adventurous excursionist – follow a ‘track’ into ‘difficult’ and ‘unfamiliar’ scenery in the north. An ‘inn’ at Coniston might give the tourist a base to discover sights both ‘agreeable’ and ‘harsh’, a balance between the beautiful and sublime aesthetics for which the Lakes were renowned (Donaldson, Gregory and Taylor 2017).

[Figure 6: Tourist GFNs near here]

[Table 3 near here]

Figure 6 and Table 3 show that the picture is similar when we examine tourist locales. ‘Lake’ is by far the most common tourist geo-noun with 99 co-occurrences, followed by ‘road’ with 62, and ‘pass’ and ‘mountain,’ with 43 each. These four geo-nouns between them account for a third of the total tourist locale co-occurrences, showing that, for tourists, locales have a much more limited and focussed vocabulary than toponyms. The most common edge is between ‘road’ and ‘pass’, while the remainder of the top five link ‘lake’ with ‘boat’, ‘mountain’, ‘pass’ and ‘shore’. The strong links between these four geo-nouns recall the CLDW’s tourist itineraries, which focus on visits to lakes including Windermere, Derwentwater, and Buttermere with views of conspicuous landmarks like Skiddaw or the Langdale Pikes. That clusters associated with the most remote parts of the Lake District (pass-inn-bridge, linked to Scafell and Sty Head) are also associated with words linked to tourist infrastructure (‘accommodation’, ‘refreshment’, ‘enquire’) or directions (nouns such as ‘head’ as in the head of the valley, ‘north’ and ‘mile’; verbs such as ‘descend’, ‘turn’, and ‘cross’; adjectives such as ‘east’, ‘adjacent’, and ‘high’), suggests the extent to which tourists were more focused on amenities than aesthetics.

Certain locales invite more reflection on their aesthetic properties. The mountain-valley cluster – with associated locations at the popular tourist attraction the Bowder Stone, as well as Bow Fell, Castle Crag, and Rosthwaite – seems to invite a sense of freedom unusual in the usually prescriptive tourist itineraries with terms such as ‘escape’, ‘rejoice’, and ‘run’.

However, these first two, as well as the prominence of the Nag’s Head Inn, are attributable to one author, Samuel Barber (1892), who describes his observations of a group of vacationing University of London students:

Down the grassy slope by the ghyll at the foot of the mountain three gray objects rapidly descend in Indian file towards the high road that runs along its base; these are collegian tourists that have walked over from Patter-dale, rejoicing in their escape from the anxieties of the Senate House, and now making straight for the Nag's Head.

Here, the students’ joy appears to come from a twofold source: their day of walking across the fells, and their imminent arrival at the pub. Occurrences of ‘run’ refer, variously, to the aqueduct that was to take water from Thirlmere to Manchester (Barber, 1892), a tourist train that ran from London to Windermere and Keswick (Anon, 1853), and a road ‘running by the side of a narrow stream’ in St. John’s Vale (Anon, 1852). The tourist, it seems, enjoyed freedom vicariously through the objects in the locale. At other spots, the guidebooks aimed to encourage the tourist to absorb some of the qualities of their surroundings. In the waterfall cluster, the tourist is invited to ‘admire’ the ‘picturesque’ scenery in ‘tranquil’ settings – although the bleakness and sterility of the fells threatens to overwhelm the pleasant associations. The tourist interested in gentler countryside – consisting of villages, hills, and wood – such as that found on the southern route into the Lakes via Lancaster and Ulverston might enjoy an evening walk in ‘rich’, ‘fine’, ‘pleasing’ and even magnificent surroundings. Connecting these locales together is the ubiquitous ‘road’, the most common geo-noun in the second largest cluster. The importance of roads again indicates the importance of practicality

to locale-based tourist places. Although travel along these thoroughfares might be an ‘adventure’, there is little in the sense of place words to suggest that, for tourists, travel on roads was much more than a practical activity with little of aesthetic pleasure. The experience was very different for travellers.

iii. Traveller places

[Figure 7: Traveller toponyms near here]

The locations and locales formed by toponyms and geo-nouns for travellers share much with those for tourists, but the relationships between them are distinctive. Figure 7 shows the clusters of co-occurring toponyms that make up the major traveller locations. While Keswick and Ambleside again have the highest number of toponym instances, the more dispersed pattern means that they have 6.4 and 4.0 instances per hundred traveller sentences compared to 12.0 and 7.5 for tourists. Unlike with tourists, no single edge dominates the graph, and, as well as Keswick and Ambleside, clusters are formed around places including Penrith, Buttermere, Ullswater, and Patterdale. The points around the south-western region – Ulverston and the Furness Peninsula – are largely attributable to a single traveller, Edwin Waugh (1860 and 1861).

[Table 4 near here]

The more dispersed pattern for traveller locations is further illustrated by Table 4. As with tourists, traveller toponyms that co-occur in the text tend also to be close together spatially. Like with tourists, too, a northern cluster centred on Keswick is dominant, as is a southern one focused on Ambleside and Bowness-on-Windermere. The sense of place terms indicate that both of these clusters form part of descriptions of travellers’ arrivals into the Lake District: these clusters contain most in the way of practical terms (‘tour’, ‘m[iles]’, ‘follow’, ‘proceed’), alongside the traveller’s delight at the ‘beautiful’ or ‘sublime’ views in and

around the towns. Despite the Keswick cluster spreading into remoter areas, the geo-nouns for this cluster are largely associated with human features such as ‘parish’, ‘village’ and ‘town.’

The second cluster centres on Patterdale and Grasmere, intersected by Helvellyn; the geo-nouns and sense of place words (‘scree’, ‘path’, ‘stream’, ‘high’, ‘bleak’, ‘ascend’) indicate a more intimate knowledge of the mountain tracks than we saw with tourists. This and the Ambleside cluster emphasise the traveller’s own movement; while the geo-nouns associated with toponyms for tourists refer to carriages, boats, and trains, the traveller relies on their feet (‘foot’, ‘pedestrian’) or horseback to explore in more intimate detail. One result is that the traveller is more inducted into the sublime, rather than only the beautiful; the bleakness and darkness ascribed to this cluster is indicative of a successful search after this more challenging aesthetic experience. By contrast, the third cluster is a strange mix of the names of the three Lake District counties – Cumberland, Westmoreland and Lancashire – and places around Coniston and the south-western Lakes. It is characterised by a ‘leisurely’ enjoyment of ‘beautiful’ views: ‘eye’, ‘look’ and ‘see’ are prominent among the sense of place words, and examples from this cluster emphasise Coniston a location from which to enjoy a prospect across all three counties.

[Figure 8: Traveller GFNs near here]

[Table 5 near here]

Similarly, travellers share the same four main geo-nouns with tourist: ‘road’, ‘lake’, ‘mountain’ and ‘pass’, although for travellers ‘road’ has the largest number of instances compared to ‘lake’ for tourist. For travellers, these four examples make up over a quarter of geo-noun instances but, as Figure 8 and Table 5 show, the relationships between them are complex. While for tourists the major geo-nouns all connected back to ‘lake’, the edges

across the travellers' geo-nouns graph are much more evenly distributed. In Figure 8, 'lake' and 'road' lie at the centre of one main cluster; the most common edge is between 'lake' and 'mountain', while the next three are all between 'road' and respectively 'lake', 'pass' and 'mountain,' suggesting a more well-rounded appreciation of the Lake District's landscape beyond the lakeside viewing stations favoured by tourists. That travellers' understanding of locale was more complex is indicated by the plethora of unconnected geo-nouns surrounding the central cluster.

In general, as we can infer from Table 5, traveller locales form more coherent sets of features than those for tourists. The largest traveller locale contains both 'road' and 'pass' with the majority of other geo-nouns referring to human features such as 'inn', 'house' and 'village.' The associated toponyms, alongside sense of place terms that focus on the quality of the thoroughfares, accommodation and, importantly, ale suggest that the moments at which travellers were most interested in infrastructure was at the entry and exit points to the district and not, as with tourists, as a primary means of traversing the region. The 'lake' cluster is similarly cohesive, including terms such as 'point', 'water', 'boat' and 'island'. As with tourists, travellers to the Lakes were primarily interested in views, but where tourists focus on 'picturesque' scenes, travellers concentrate on what is 'beautiful' or 'sublime'.

The 'mountain' cluster is perhaps the site of the greatest difference between tourists and travellers; while tourists admire 'grassy' and 'attractive' features, such as at Bow Fell or Castle Crag, travellers 'toil' up 'difficult', 'steep', and 'tremendous' slopes. Getting to the 'top' is important for giving the traveller a sense of control over the landscape; Ann Radcliffe (1795) was indicative when she memorably wrote that reaching the summit of Skiddaw allowed her to '[command] the whole dome of the sky'. Those kinds of extensive views depend on the weather, though ('cloud', 'sun', 'shadow'); Harriet Martineau (1855), for instance, wishes her reader better luck in climbing High Street than she had on her previous

visit, when ‘the wind laid the whole party flat on the summit of the pass’. There is a focus on difficulty and reward (‘perseverance’, ‘pleasure’, ‘consolation’); Edward Baines (1829), for example, advises that reaching Stickle Tarn and the summit of Harrison Stickle are rewards for ‘great exertion, or rather great patience and perseverance’. The contrast between mountains for tourist and travellers is revealing: for tourists, mountains are to be things to be looked at, while for travellers they are to be climbed. The toponyms in this cluster – including Coniston Water and the comparatively gentle Hampsfell – bely this effort, and are representative of a problem throughout these and the tourists’ clusters: it would be difficult to guess to which cluster the toponyms refer. It is also noticeable that, for travellers, locale-based places are much more numerous and dispersed than the very concentrated location-based places.

A key difference between tourists and travellers is the way they respond to and describe the places they visit. While features that make up traveller locales are similar to those for tourists, the total number of sense of place words show that much richer senses of place are developed for traveller locales. Apart from the roads cluster, the senses of place terms are much more focused on aesthetic experience; even where practical suggestions for routes are invoked, the terms used are indicative of the sense of adventure the traveller experienced. They might ‘discover’ something at a ‘majestic’ river, ‘conduct’ themselves on an ascent of a ‘delightful’ (or ‘horrid’) valley, or even ‘penetrate’, or ‘shrink’ from, the apparently impenetrable rocks of Borrowdale. In these terms, we encounter a Lake District full of mystery, and difficulty, waiting for the traveller to discover and conquer it. When we turn to inhabitants, we find a very different relationship to place.

iv. Inhabitant places

Most writers in the CLDW were not residents of the Lake District. Even the few with Lakeland connections – including William Gilpin, Thomas West, William Wordsworth, John Ruskin, Harriet Martineau, and William Collingwood – were not part of the communities they describe as ‘inhabitants’. ‘Inhabitants’ tends to refer to the Lake District’s working poor and, while there were several notable working-class writers in the area during this period, they do not feature in the CLDW. An analysis of inhabitants’ place in the CLDW, then, is a reflection of how poorer communities were perceived and represented by the leisured classes.

[Figure 9: Inhabitant toponyms near here]

That inhabitants are far less associated with toponyms that tourists and travellers is not necessarily a reflection of their experience of the Lakes, but rather a symptom of the ways in which writers in the CLDW – who were more usually not from the area – represented inhabitants in a manner that is disenfranchised from their homeland. Figure 9 reinforces this perceived sense of placelessness; the lack of edges indicates that inhabitant locations are usually described using only a single toponym, rather than the multiple toponyms far more commonly used for tourists and travellers. Seathwaite and Ulpha, farming communities in the north and south-west Lakes respectively, have the highest number of inhabitant co-occurrences at six. The map of inhabitant geographies derived from the CLDW thus looks very different to the tourist or traveller counterparts; while Keswick remains prominent, Ambleside and other tourist locations in the central Lakes are much reduced in importance. Instead, communities around the western Lakes, and peripheries to the south and north-east, are more significant.

[Table 6 near here]

This sparsity of toponyms for inhabitants is reflected in table 6, where we have reduced the threshold of the clusters so that over 5% of total inhabitant instances are shown rather than

10% used in Tables 4 and 5. Even with this reduced threshold, there are only three major clusters and these are highly dispersed across the region. Indeed, the largest cluster refers to most of the region, having two counties (Cumberland and Westmoreland) among its most frequent toponyms. The other two clusters are more typical of tourist or traveller clusters in that there is a northern one centred on Keswick that includes nearby Derwent Water and Skiddaw, and a southern one centred on Kendal that includes Windermere and Ambleside. Even these are distinguished from tourists and travellers by focusing on the market towns rather than surrounding areas, as the geo-nouns in the first and third clusters make clear. Rather than assessing aesthetics or infrastructure, associated sense of place terms highlight historic ‘attachment’ to places from the Romans onward, and the focus is a qualitative assessment of the various ways that the place has been ‘cultivated’ and ‘occupied’.

The Keswick cluster, surprisingly, focuses on rural geo-nouns: ‘vale’, ‘slope’, ‘grove’, ‘wood’, ‘ridge’. Nevertheless, the sense of place words still focus on human working relationships with the landscape. ‘Poor’ and ‘small’ in the Keswick cluster are used by several writers to refer to the agricultural conditions, specifically the fields, in this area.

‘Steal’ is drawn from a single anonymous guide to *Keswick and its Neighbourhood: A handbook for the Use of Visitors* (Anon, 1852) where the author describes the town as impoverished and inaccessible prior to the arrival of tourism, quoting a 1751 magazine article which claimed that the poorer inhabitants subsisted by stealing black lead. ‘Aversion’ and ‘terror’ are found in this cluster because both Thomas Gray (1775) and Thomas West (1778) state that the inhabitants refer to the name of Skiddaw Fell ‘with a sort of terror and aversion’, a reflection of eighteenth-century fears about mountains (Nicholson 1955). The phrase is re-produced by later writers such as Charles Mackay (1846). ‘Sun’ has a similar origin in that both Gray and West refer to a proverb repeated by Keswick’s inhabitants, which says that the sun always shines on the hills by Keswick.

Although not large enough to appear in table 6, the cluster associated with Ulpha and Seathwaite has the fifth and eight most common toponyms for inhabitants. They also co-occur more than any other inhabitant toponyms. Many of the sense of place terms around Ulpha, a small and relatively inaccessible agricultural community in the south-west Lakes, are negative, with words such as ‘disagreeable’, ‘apprehensive’, ‘discontent’ and ‘objection’. The area seems to be disadvantaged by the influence of certain prominent figures, notably the ‘bishop’, ‘curate’ and ‘schoolmaster.’ However, these instances come from a single story that originally appeared in Wordsworth’s *River Duddon* sonnet series (1820), in which he tells of the Bishop of Chester’s attempt to merge the curacies of Seathwaite and Ulpha and how this was opposed by Ulpha’s inhabitants. This story is re-told in later texts such as Gibson (1849) and Waugh (1861). Like earlier Gray and West references around Keswick, these repetitions suggest the importance of a small number of writers in defining the sense of place ascribed to the Lake District’s inhabitants.

[Figure 10: Inhabitant GFNs near here]

[Table 7 near here]

That inhabitants are more likely to be associated with the more urban aspects of the Lake District is reiterated by Figure 10 and Table 7, which show the information for inhabitant locales. As we saw with Table 1, the richer set of geo-nouns suggests that inhabitants are more associated with locale than location. Inhabitant locales are similar to those for tourists and travellers in so much as they are made up of multiple geo-nouns rather than the isolated toponyms that tends to define locations. The largest cluster is dominated by instances of ‘town’, the geo-noun most associated with inhabitants. Towns are associated with a diverse range of other geo-nouns including ‘market’, ‘country’, ‘river’ and ‘coast’. The town toponyms that co-occur with this cluster are not the obvious major tourist towns, but instead

peripheral market towns such as ‘Ulverston’, ‘Whitehaven’, ‘Milnthorpe’ and ‘Cockermouth’. This literal decentralising of inhabitants speaks to the extent to which they are marginalised in the CLDW; writers tend to view inhabitants as part of the scenery. That inhabitants are portrayed as objects to look at is reinforced by their lack of association with roads. While for both tourists and travellers, ‘road’ is one of the most common geo-nouns and central to one of the largest clusters, for inhabitants they are far less important, having only fifteen instances and few edges. This difference reflects the way that inhabitants are represented as being more static, with roads and other forms of transport being largely irrelevant to the ways they are represented. Indeed, inhabitants are often described, in an almost semi-mythic way, as being contained by the mountains and so protected from the vices of the outside world. The great pedestrian Joseph Budworth (1792) is far from the only writer in the CLDW to celebrate the way the mountains seem to ‘serve as barriers to keep out many follies and vices’, and so ‘preserve’ the region and the ‘harmless lives [of] the inhabitants of the lakes’. The region’s topography seems, in accounts like this, to be a defining feature of inhabitants’ identities.

As with travellers, inhabitant locales have a richer sense of place than their locations, though these tend to focus on quality-of-life assessments. The rural areas take on the idealised qualities that tourists (and, to a lesser extent, travellers) privileged in their Lakeland narratives. The mountain cluster celebrates the place’s fertility; the country and village clusters idealise the ‘calm’ and ‘good’, yet ‘ignorant’, ‘simplicity’ of farming families focused on rearing sheep, with an implication that ‘pastoral’ beauties influence the characters of those who live there. Together, these clusters reinforce a myth of what Thomas Gray called the ‘happy’ poor. The more urban clusters, however, provide a more realistic reflection of everyday life. The town cluster comments that Lakeland towns are ‘small’ and ‘neat’, although historically they had the propensity for outbreaks of contagious disease (‘infected’, ‘garment’); the ‘house’ cluster finds

that rents and finance are a major issue ('yearly', 'schoolmaster', 'wage'), largely thanks again to Wordsworth's explanation of the Ulpha/Seathwaite debate; and the street cluster notes town planning (the 'square' towers and 'narrow' streets), as well as some of the functions of the castles and other historic buildings ('dungeon', 'vault', 'hang'). Taking these senses of place together, the implication is of a simple community focused on everyday worries, living in a 'beautiful' place that might be either 'lonely' or a 'refuge'. The harsh realities of subsisting on small farms in the northern uplands are overlooked in this outsiders' understanding of inhabitants' senses of place.

d. Conclusion

This paper has demonstrated how a tripartite understanding of place can be uncovered in historical texts by extracting toponyms, geo-nouns, and sense of place descriptors in the form of statistically significant adjectives, nouns and verbs that co-occur with relevant toponyms and geo-nouns. By adapted digital approaches such as GIS and network analysis we have moved towards Bushell's (2020) call for digital models for literary geography that represent relative rather than absolute space. This allows us to demonstrate the similarities and differences in the way that place is constructed and described for different groups of people using large volumes of digital text.

While our approach opens up new ways of understanding place in textual sources, it has a number of limitations that could be addressed in further work. Here, place is defined primarily either by toponym or by locale. An alternative would be to combine these to bring place closer to Agnew's integrated definition, with locations and locales acting together to define place. Relatedly, methods that create summaries of sense of place based on the words attached to them could be developed further, along with methods to compare the senses of place for different places (Wartmann and Purves 2018). Our analysis is primarily concerned with nodes,

the places themselves, rather than with the edges, or what connects places. An additional level of sophistication would be to add more meaning to, or nuance across, edges, such as by allocating edges to classes (e.g. landscape appreciation or function). This approach could be enhanced through the use of subject-predicate-object triples (Weikum et al. 2021) whereby a phrase such as ‘Skiddaw provides a good view of Derwent Water’ would be split into two nodes: ‘Skiddaw’, the subject; ‘Derwent Water’, the object; linked by the edge (‘provides a good view of’), a predicate which is both functional (Derwent Water is visible from Skiddaw) and aesthetic (the view is good). This method could then be used to provide an analysis that stresses what writers think connects places. Nevertheless, our approach has enabled us to demonstrate the similarities and differences in the way that place is constructed, described and experienced for different groups of people using large volumes of digital text. When we explore and map place in this way, as a multi-dimensional experience that combines external factors with subjective perception, we open up new avenues for exploring human relationships with the world.

Note: As yet no single software environment provides the tools to explore texts in both geographical and network space. This required us to use a combination of different software tools: ArcGIS for the GIS analysis, Gephi for network analysis, and AntConc to assist with the corpus linguistics. Custom-written Python scripts were used to allow data to be incorporated into Gephi and ArcGIS in the appropriate formats as neither of these packages is written to accommodate natural language texts.

Acknowledgements:

We acknowledge support from the “Understanding imprecise space and time in narratives through qualitative representations, reasoning, and visualisation” grant funded by the UK's Economic and Social Research Council (ES/W003473/1) and US National Science Foundation (NSF#2225179). The research was also supported by “Revealing long-term change in vegetation landscapes: The English Lake District and Beyond” funded by the UK's Arts and Humanities Research Council (AH/T006153/1). We would like to acknowledge the support of all of the team members on these two projects for their ideas and support.

References:

- Agnew, J. 2011. Space and place. In *Handbook of Geographical Knowledge*, ed. J. Agnew and D. Livingstone, 316-330. London: Sage.
- Ahnert, R. and S.E. Ahnert. 2015. Protestant letter networks in the reign of Mary I: A quantitative approach. *English Literary History*, 82: 1-33.
- Anon. 1852. *Keswick and its Neighbourhood: A Hand-book for the Use of Visitors*. Windermere: J. Garnett.
- Anon. 1853. *Black's Shilling Guide to the English Lakes*. London: A. and C. Black
- Baddeley, M.J.B 1900. *Black's Shilling Guide to the English Lakes*. London: A. and C. Black.
- Baines, E. 1829. *A Companion to the Lakes of Cumberland, Westmoreland, and Lancashire*. London: Simpkin and Marshall.
- Barber, S. 1892. *Beneath Helvellyn's Shade. Notes and Sketches in the Valley of Wythburn*. London: E. Stock.
- Barthélemy, M. 2011. Spatial networks. *Physics Reports* 499, 1-101.
- Bergmann L. and N. Lally. 2021 For geographical imagination systems. *Annals of the Association of American Geographers* 111: 26-35.
- Berman, L., R. Mostern, and H. Southall. 2016 *Placing Names: Enriching and integrating gazetteers*. Bloomington IN: Indiana University Press.
- Bernardini, W. and M.A. Peeples. 2015. Sight communities: The social significance of shared visual landmarks. *American Antiquity* 80: 215-235.

- Bolstad, P. 2019, *GIS Fundamentals: A first text on Geographic Information Systems*. 6th edition. Ann Arbor, MI: XanEdu.
- Budworth, J. 1792. *A Fortnight's Ramble to the Lakes in Wesmorland, Lancashire, and Cumberland*. London: Hookham & Carpenter
- Bushell, S. 2012. The Slipperiness of Literary Maps: Critical Cartography and Literary Cartography. *Cartographica* 47: 149-160.
- Bushell, S. 2020. *Reading and Mapping Fiction. Spatialising the Literary Text*. London: Routledge.
- Bushell, S., J. O. Butler, D. Hay, and R. Hutcheon. 2022. Digital Literary Mapping: I. Visualizing and Reading Graph Topologies as Maps for Literature. *Cartographica* 57: 11-36.
- Butler, J.O., C.E. Donaldson, J.E. Taylor and I.N. Gregory. 2017. Alts, Abbreviations, and AKAs: Historical onomastic variation and automated named entity recognition. *Journal of Map and Geography Libraries* 13: 58-81.
- Buzard, J. 1993. *The Beaten Track: European Tourism, Literature, and the Ways to Culture, 1800-1918*. Oxford: Oxford University Press.
- Cope, M. and S. Elwood. 2009 *Qualitative GIS: A mixed methods approach*. London: Sage.
- Cresswell, T. 2004. *Place: A short introduction*. Oxford: Blackwell.
- Defoe, D. 1727. *A Tour thro' the Whole Island of Great Britain*. London: G. Strahan.
- Donaldson, C., I.N. Gregory and J.E. Taylor. 2017. Locating the beautiful, picturesque, sublime and majestic: Spatially analysing the application of aesthetic terminology in descriptions of the English Lake District. *Journal of Historical Geography* 56: 43-60.

- Fiennes, C. 1888. *Through England on a Side-Saddle in the Time of William and Mary*. London: Field & Tuer.
- Fotheringham, A.S., C. Brunsdon, and M. Charlton. 2000. *Quantitative Geography: Perspectives on spatial data analysis*. London: Sage.
- Gao, S. 2022. Towards place-based GIS. In *New Thinking in GIScience*, ed. B. Li, X. Shi, A.-X. Zhu, C. Wang and H. Lin, 51-58. Singapore: Springer.
- Gibson, A.C. 1849. *The Old Man; or Ravings and Ramblings round Conistone*. London: Whittaker.
- Gidal, E. and M. Gavin. 2019. Infrastructural semantics: postal networks and statistical accounts in Scotland, 1790-1845. *International Journal of Geographical Information Science* 33: 2523-44.
- Golitko, M., J. Meierhoff, G.M. Feinman, and P.R. Williams. 2012. Complexities of collapse: The evidence of Maya obsidian as revealed by social network graphical analysis. *Antiquity* 86: 507-523.
- Goodchild, M.F. 2011. Formalizing place in geographical information systems. In *Communities, Neighborhoods and Health*, ed. L. Burton, S. Kemp, M.-C. Leung, S. Matthews, and D. Takeuchi, 21-33. Singapore: Springer.
- Gray, T. 1775. *Journal of A Visit to the Lake District in 1769*. London: J. Dodsley.
- Gregory I.N., C. Donaldson, P. Murrieta-Flores, P. and Rayson. 2015. Geoparsing, GIS and textual analysis: Current developments in Spatial Humanities research. *International Journal of Humanities and Arts Computing* 9: 1-14.

Grover, C., R. Tobin, K. Byrne, M. Woollard, J. Reid, S. Dunn, and J. Ball, J. 2010. Use of the Edinburgh geoparser for georeferencing digitized historical collections. *Philosophical Transactions of the Royal Society A* 368: 3875-89.

Hardie, A. 2007. Part-of-speech ratios in English corpora. *International Journal of Corpus Linguistics* 12: 55-81

Hess, S. 2012. *William Wordsworth and the Ecology of Authorship: The Roots of Environmentalism in Nineteenth-Century Culture*. Charlottesville: University of Virginia Press.

Horn, P. 1980. *The Rural World, 1780-1850: Social Change in the English Countryside*. London: Routledge.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Jiménez-Badillo D., P. Murrieta-Flores, B. Martins, I. Gregory I., M. Favila Vazquez, and R. Licerias-Garrido. 2021. Developing geographically oriented NLP approaches to sixteenth century historical documents: Digging into Early Colonial Mexico. *Digital Humanities Quarterly*: 14(4). <http://digitalhumanities.org/dhq/vol/14/4/000490/000490.html>

Lambiotte R., J.-C. Delvenne, and M. Barahona. 2009. Laplacian Dynamics and Multiscale Modular Structure in Networks *arXiv*. <http://arxiv.org/abs/0812.1770v3>

Leigh S. 1830. *Leigh's Guide to the Lakes and Mountains of Cumberland, etc*. London: Samuel Leigh

Lloyd, C.D. 2010. *Local models for Spatial Analysis*. 2nd ed. Boca Raton FL: CRC Press.

Longley P.A., M.F. Goodchild, D.J. Maguire and D.W. Rhind. 2015 *Geographical Information Science and Systems*. 4th ed. Hoboken NJ: John Wiley & Sons.

MacKay, C. 1846. *The Scenery and Poetry of the English Lakes. A Summer Ramble*. London: Longman.

Martineau, H. 1855. *A Complete Guide to the English Lakes*. Windermere: J. Garnett.

McEnery, A.M. and A. Hardie. 2011. *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Murrieta-Flores P., A. Baron, I. Gregory, A. Hardie and P. Rayson P. 2015. Automatically analysing large texts in a GIS environment: The Registrar General's reports and cholera in the nineteenth century. *Transactions in GIS* 19: 296-320.

Newman, M. 2018. *Networks*. 2nd ed. Oxford: Oxford University Press.

Nicholson, N. 1955. *The Lakers: The adventures of the first tourists*. London: Robert Hale.

Paterson L.L. and I.N. Gregory. 2019 *Representations of Poverty and Place: Using geographical text analysis to understand discourse*. Cham, Switzerland: Palgrave.

Porter C., P. Atkinson, and I. Gregory. 2015. Geographical Text Analysis: A new key to nineteenth-century mortality. *Health and Place* 36: 25-34.

Purves R.S., S. Winter and W. Kuhn. 2019. Places in information science. *Journal of the Association for Information Science and Technology* 70: 420-37.

Radcliffe, A. 1795. *A Journey made in the Summer of 1794, through Holland and the Western Frontier of Germany: With a return down the Rhine: To which are added Observations during a tour to the lakes of Lancashire, Westmoreland, and Cumberland*. London: G. G. & J. Robinson

Rayson, P., A. Reinhold, J. Butler, C.E. Donaldson, I.N. Gregory, and J.E. Taylor. 2017. A deeply annotated testbed for geographical text analysis: The Corpus of Lake District Writing.

GeoHumanities'17: 1st ACM SIGSPATIAL Workshop on Geospatial Humanities: 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.

Association for Computing Machinery ACM. doi: 10.1145/3149858.3149865

Rupp, C.J., P. Rayson, I. Gregory, A. Hardie, A. Joulain, and D. Hartmann. 2014. Dealing with heterogeneous big data when geoparsing historical corpora, *Proceedings of the 2014 IEEE Conference on Big Data* 80-83.

Southall H.R., R. Mostern and M. Berman. 2011. On historical gazetteers. *International Journal of Humanities and Arts Computing* 19: 127-45.

Spacy. 2022. *Industrial-Strength Natural Language Processing in Python*. <https://spacy.io/>

Taylor J.E and I.N. Gregory. 2022. *Deep Mapping the Literary Lake District: A Geographical Text Analysis*. Lewisburg: Bucknell University Press.

Voutilainen, A. 2003. Part-of-speech tagging. In *The Oxford Handbook of Computational Linguistics*, ed. R. Mitkov, 219-232. Oxford: Oxford University Press.

Wartmann F and R.S. Purves. 2018. Investigating sense of place as a cultural ecosystem service in different landscapes through the lens of language. *Landscape and Urban Planning* 175: 169-83.

Waugh, E. 1860. *Over Sands to the Lakes*. Manchester: A. Ireland.

Waugh, E. 1861. *Rambles in the Lake Country and its Borders*. Manchester: John Heywood.

Weikum, G., X.L. Dong, S. Razniewski and F. Suchanek. 2021. Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases. *Foundations and Trends in Databases* 10: 108–490.

West, T. 1778. *A Guide to the Lakes: Dedicated to the lovers of landscape studies, and to all who have visited, or intend to visit, the lakes in Cumberland, Westmorland, and Lancashire.*

London: Richardson & Urquhart.

Wilson, J.F., E. Buchnea, and A. Tilba. 2018. The British corporate network, 1904-1976: Revisiting the finance-industry relationship. *Business History* 60: 779-806.

Wordsworth, W. 1820. *The River Duddon: A Series of Sonnets.* London: Longman.

Wordsworth, W. 1835. *Guide through the District of the Lakes.* 5th ed. London: Longman.

Yoshikawa, S. 2020. *William Wordsworth and Modern Travel: Railways, Motorcars and the Lake District, 1830-1940.* Liverpool: Liverpool University Press.

Author biographies:

IAN N. GREGORY is a Distinguished Professor of Digital Humanities in the Department of History at Lancaster University, Lancaster, LA1 4YT, UK. Email:

I.Gregory@lancaster.ac.uk. His research interests focus on the use of geospatial methods across humanities disciplines.

ROBERT SMAIL is an independent researcher. His core research interests span the history of science and environmental history. Email: Robert.Angus.Smail@gmail.com

JOANNA E. TAYLOR is Lecturer in Nineteenth-Century Literature and Digital Humanities in the Department of English, American Studies, and Creative Writing at School of Arts, Languages and Cultures, University of Manchester, M13 9PL. E-mail:

joanna.taylor@manchester.ac.uk. Her research sits at the intersection between literary studies, digital humanities, and environmental history.

JAMES O. BUTLER is a Postdoc Digital Research Associate in the Department of Archaeology, Classics and Egyptology at The University of Liverpool, L69 7WZ, UK. E-mail: james.butler@liverpool.ac.uk. His core research interests include the semantics of names and association, proximal spatial description, keyness-led corpus linguistic analysis, and digital library design.

Figures:

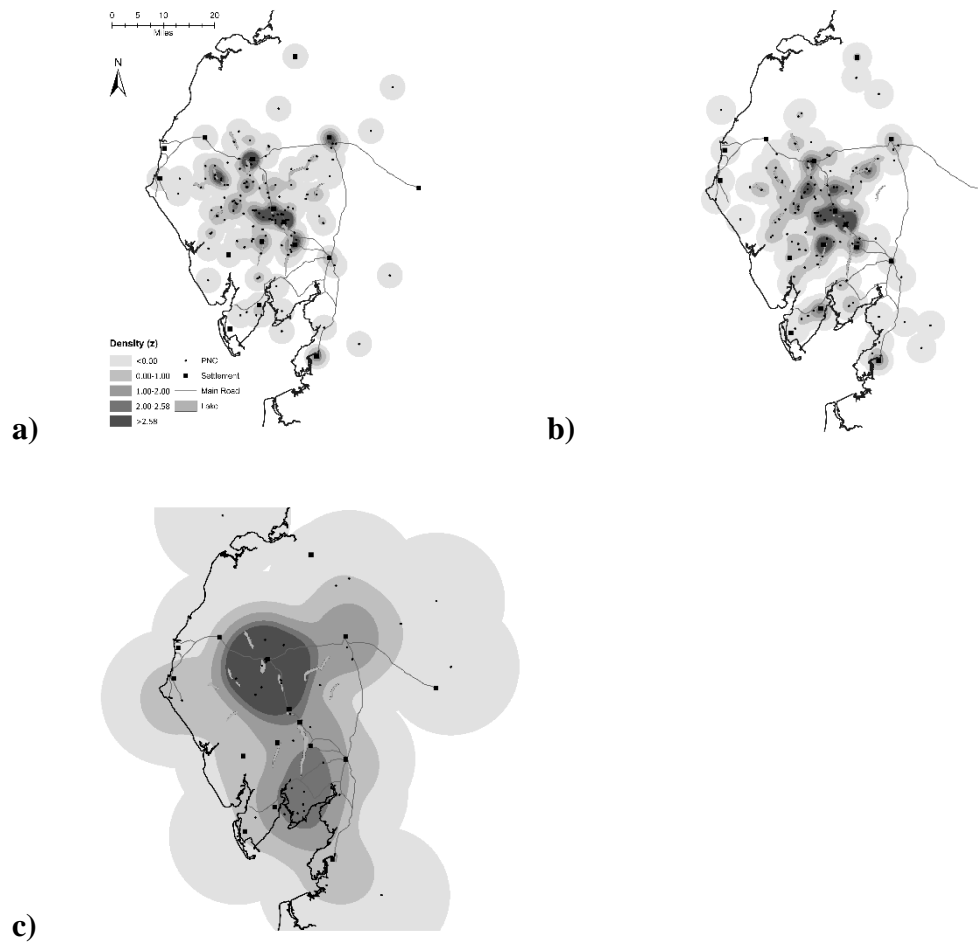


Figure 1: Density smoothed maps of co-occurrences between toponyms and the words:
a) ‘tourist[s]’, b) ‘traveller[s]’, and c) ‘inhabitant[s]’. PNCs refer to the locations where place-names co-occur within 10 words of the word. Modified from Taylor and Gregory (2022).

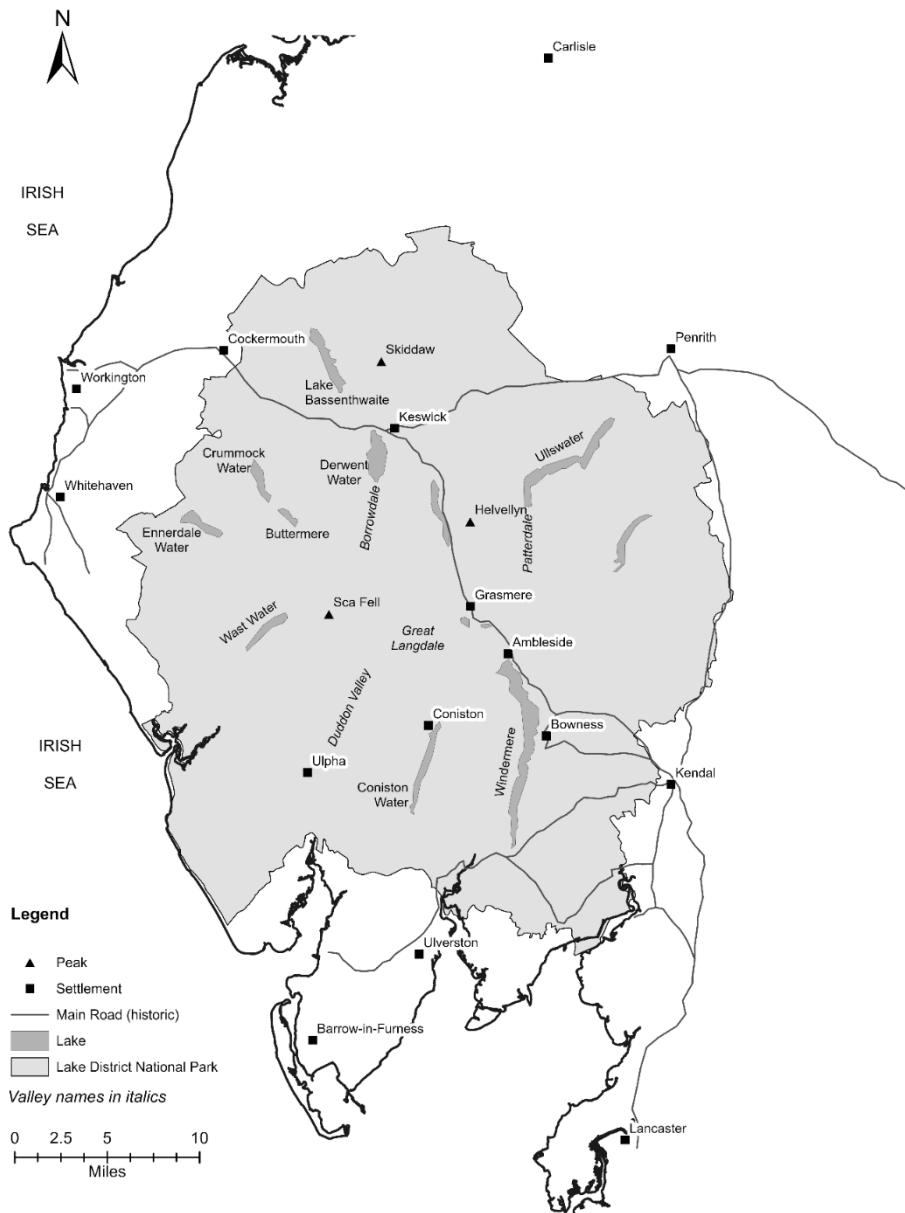


Figure 2: The major features of the Lake District

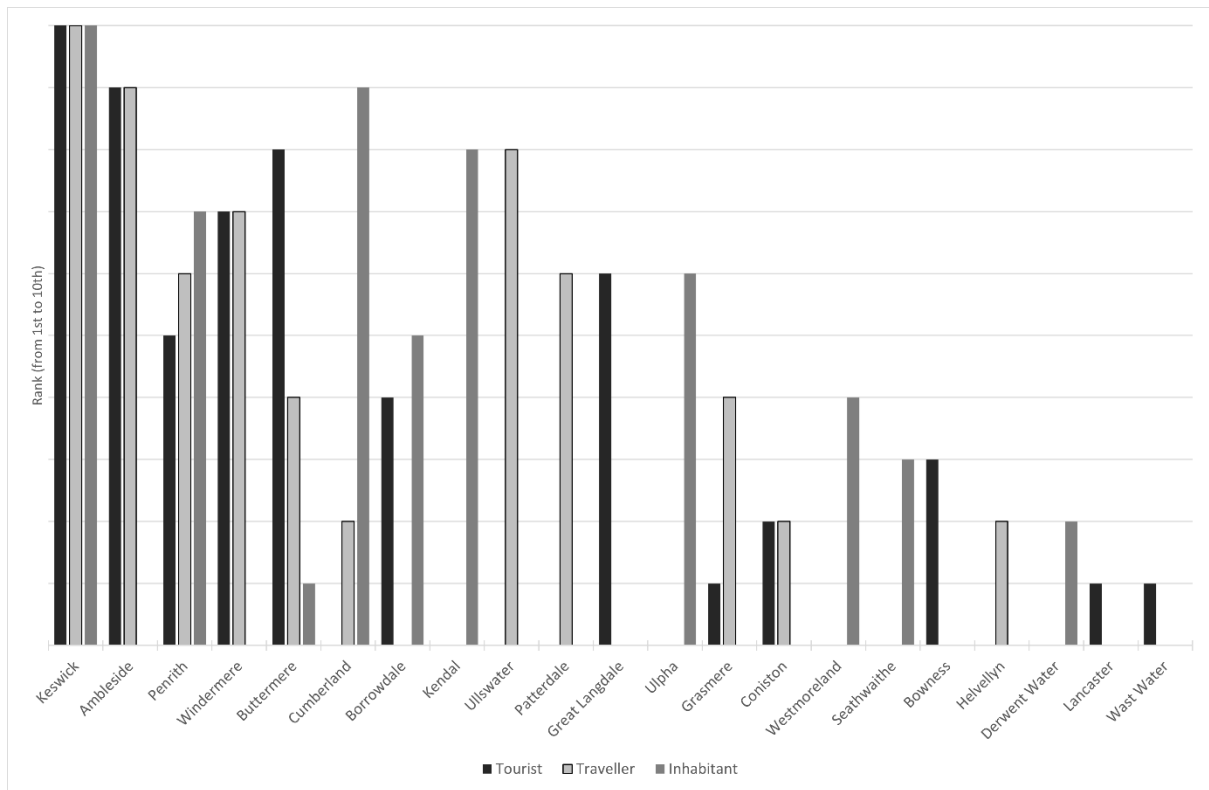


Figure 3: The top 10 toponyms in the CLDW co-occurring with tourists, travellers and inhabitants. Bars are proportional to position with top ranks being the highest and those outside the top 10 not having a bar. Co-occurring means in the same sentence.

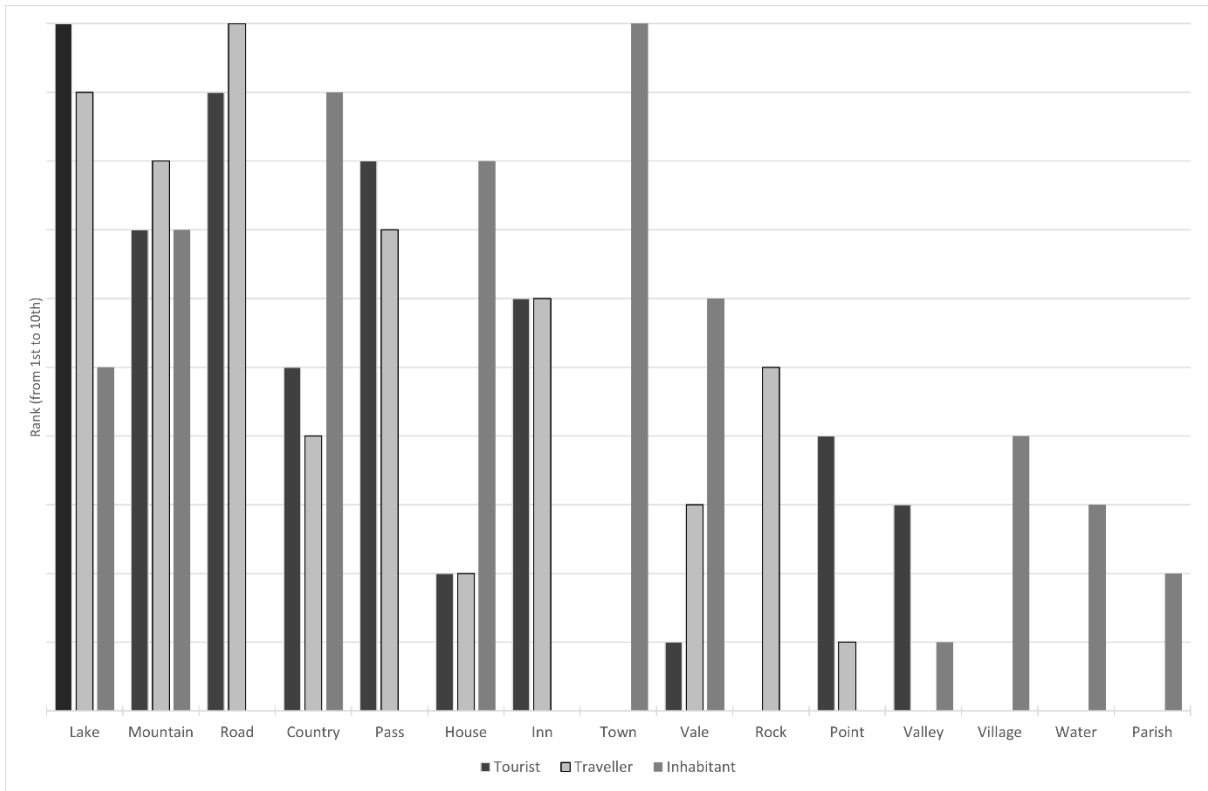


Figure 4: The top 10 geo-nouns co-occurring with tourists, travellers and inhabitants.

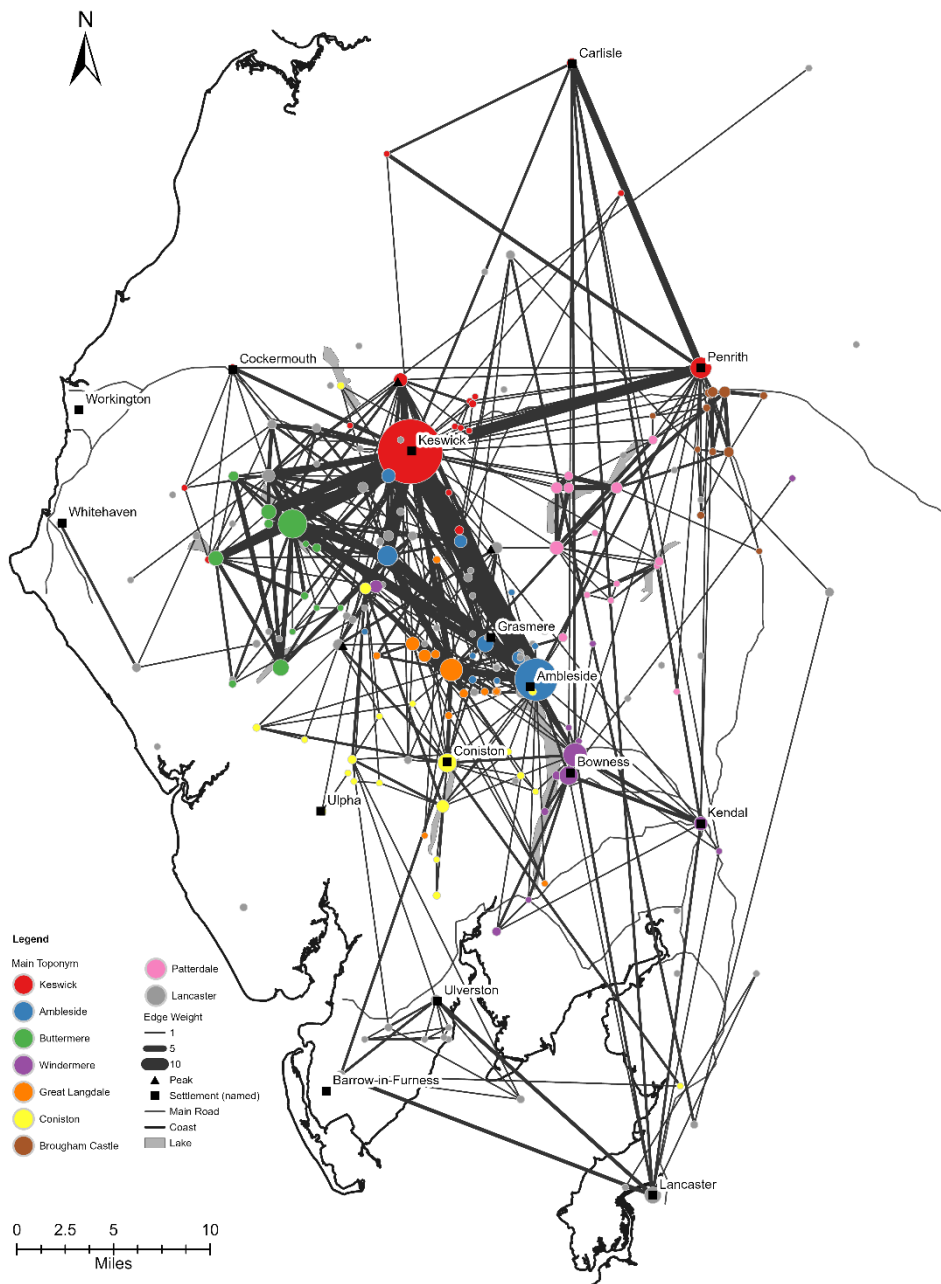


Figure 5: Network of co-occurring toponyms for tourists. Only toponyms from the Lake District area are included. Node sizes are proportional to the number of instances of the toponym. Clusters are defined by modularity with the most frequent toponym shown on the legend. The network is geo-located.

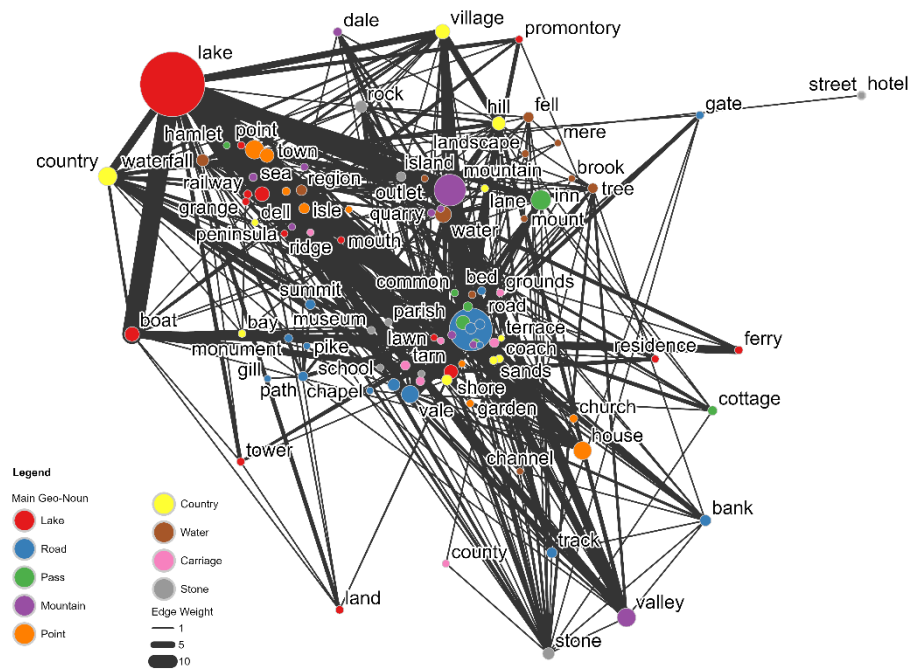


Figure 6: Network of co-occurring geo-nouns for tourists. Node sizes are proportional to the number of instances of the geo-noun. Clusters are defined by modularity.

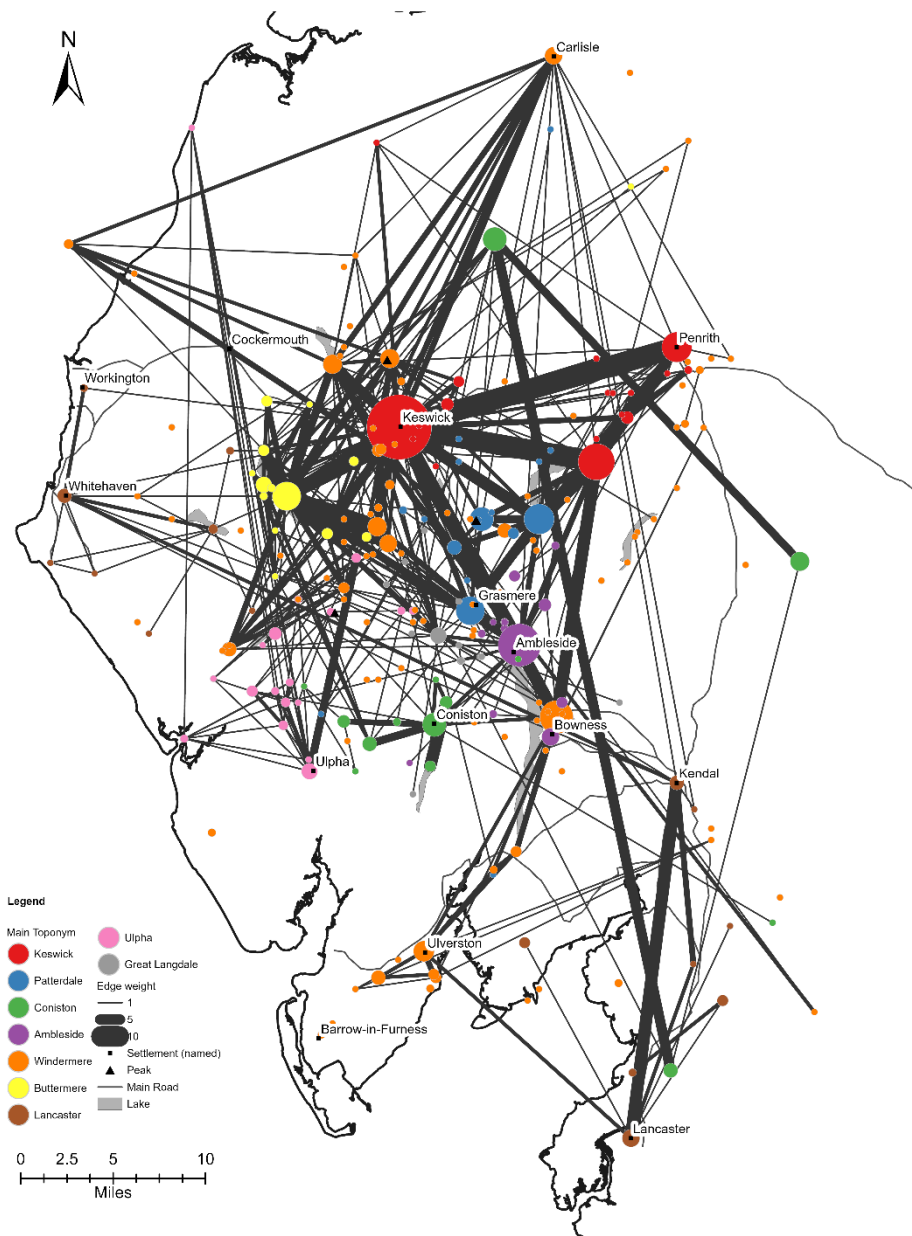


Figure 7: Network of co-occurring toponyms for travellers. Details as for figure 5.

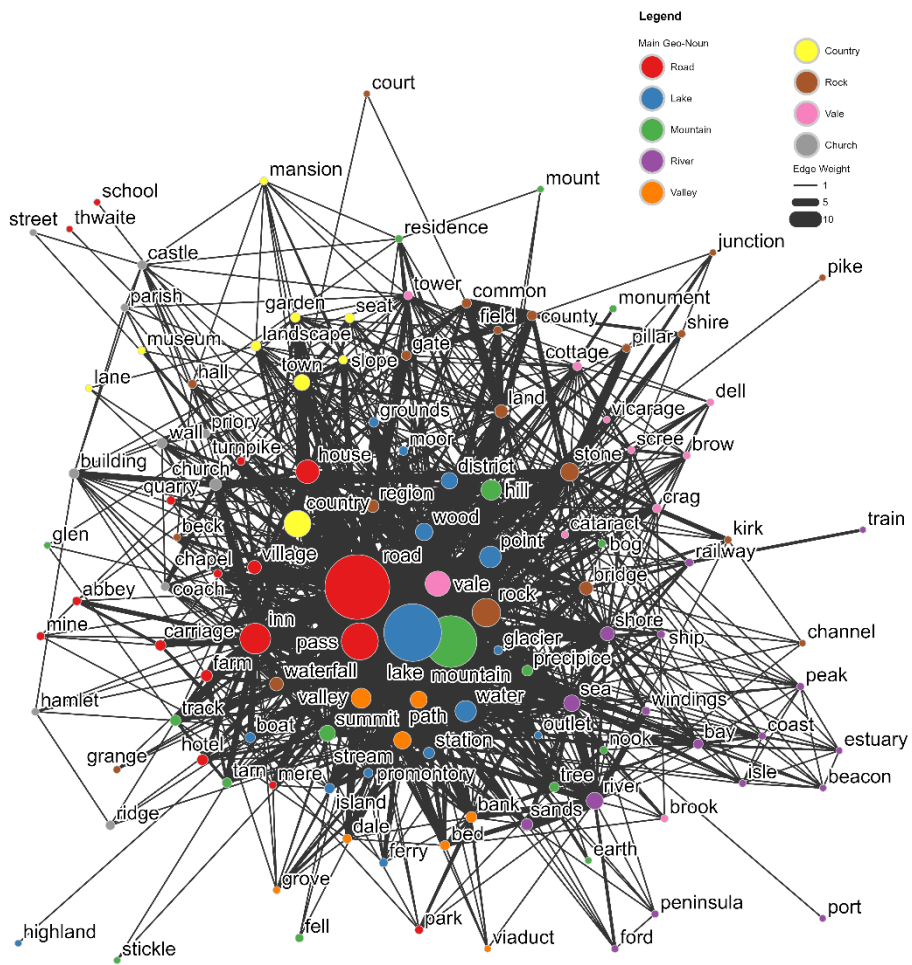


Figure 8: Network of co-occurring geo-nouns for travellers. Details as for figure 6.

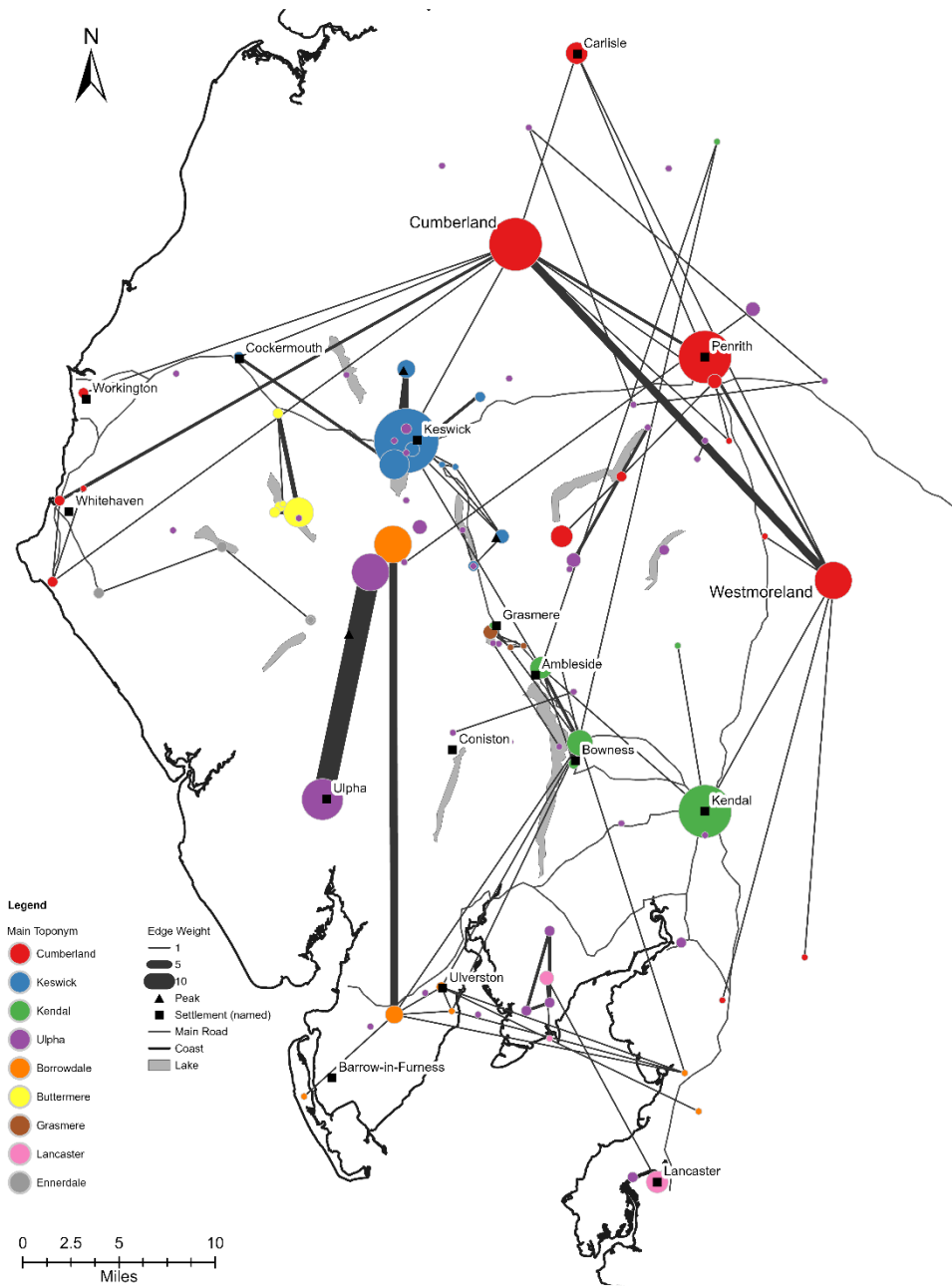


Figure 9: Network of co-occurring toponyms for inhabitants. Details as for figure 5.

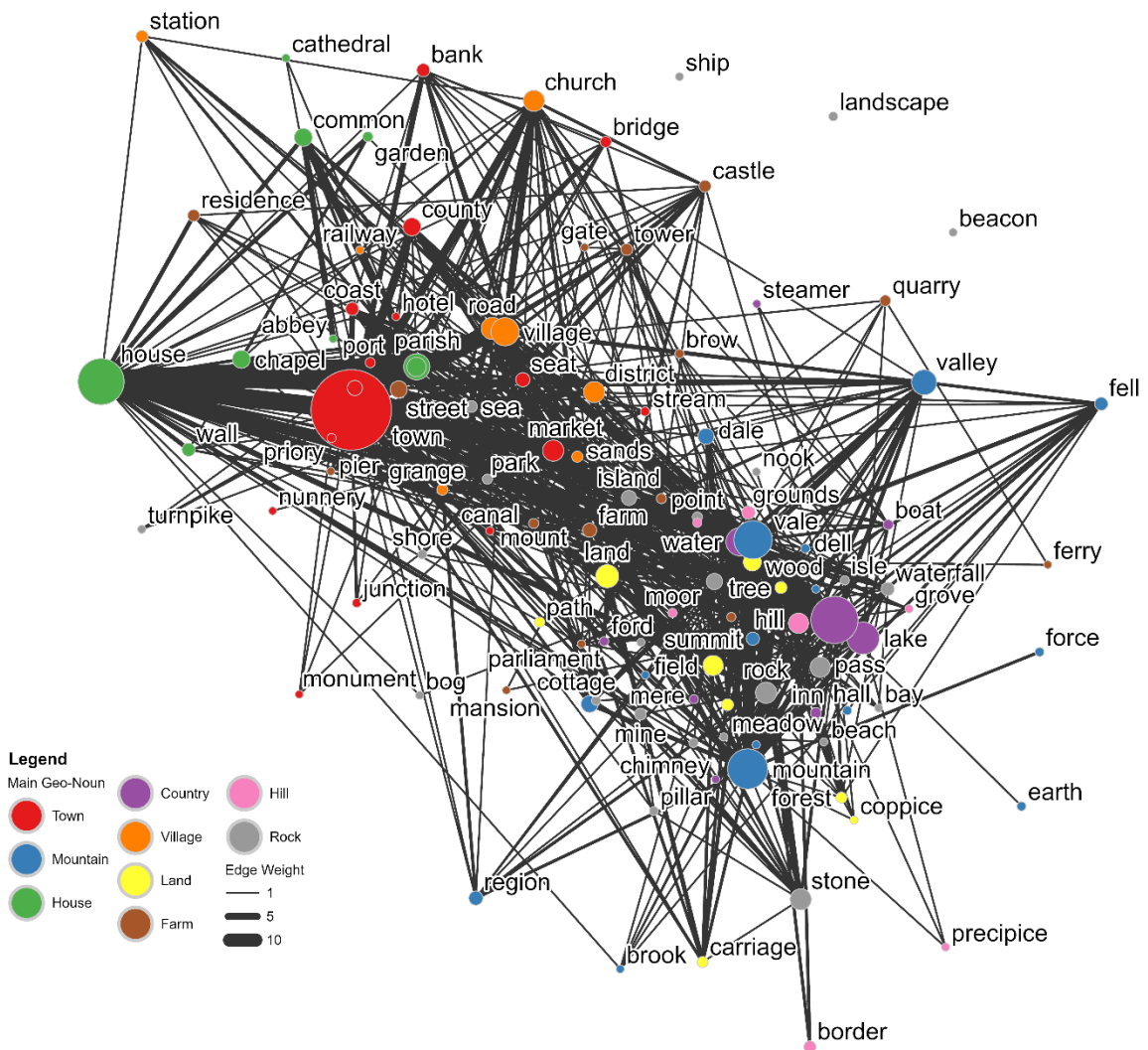


Figure 10: Network of co-occurring geo-nouns for inhabitants. Details as for figure 6.

	Tourists	Travellers	Inhabitants	Total in CLDW
Sentences	400	596	410	56,248
Total place reference co-occurrences (per sentence)	3.48	3.63	2.77	2.12
Toponym co-occurrences (n)	646	755	338	43,254
Toponym co-occurrences (per sentence)	1.62	1.27	0.82	0.73
Geo-noun co-occurrences (n)	745	1,409	798	75,928
Geo-noun co-occurrences (per sentence)	1.86	2.36	1.94	1.35

Table 1: The relative importance of toponyms and geo-nouns in describing the places associated with tourists, travellers and inhabitants. ‘Sentences’ refers to the number of sentences containing the word ‘tourist[s]’, ‘traveller[s]’ or ‘inhabitant[s]’. A co-occurrence occurs every time a toponym or geo-noun is found in one of these sentences.

N	Common Nodes (n>5)	Significant geo-nouns		Significant Sense of Place words	
		n	Indicative words	n	Indicative words
91	Keswick (48), Penrith (13), Skiddaw (7)	10	<i>Stream; hamlet; train; road; town</i>	29	<i>mile; right; north; l.; n.; w.; proceed; turn; enquire</i>
77	Ambleside (30), Borrowdale (12), Grasmere (9), Derwent Water (7), Thirlmere (6), Rydal (6)	4	<i>Dale; cottage; coach; water; bridge</i>	4	<i>lovely; little; many; return</i>
67	Buttermere (19), Wast Water (9), Ennerdale (8), Crummock Water (8)	9	<i>Mount; track; mansion; valley; region</i>	21	<i>difficult; unfamiliar; tempting; day; end; excursion; see; descend; get</i>
55	Windermere (15), Bowness-On-Windermere (11), Kendal (7), Lake District (6)	8	<i>Boat; ferry; shore; lake; railway</i>	31	<i>east; interesting; guide; quarter; abode; accommodation; find; hire; supply</i>

49	Great Langdale (14), Langdale Pikes (7), Dungeon Ghyll Force (6)	9	<i>Force; carriage;</i> <i>country; lane; tarn</i>	31	<i>conspicuous; rich;</i> <i>fine; car; finer; foot;</i> <i>ascend; visit; go</i>
48	Coniston (11), Coniston Water (6)	6	<i>Rock; river; inn;</i> <i>crag; point</i>	16	<i>agreeable; broad;</i> <i>harsh; enterprising;</i> <i>inclination; name;</i> <i>strike; trace; bear</i>

Table 2: Major clusters of locations associated with tourists with their associated geo-nouns and sense of place words. Only clusters with more than 10% (40) of the total tourist instances are included. ‘N’ refers to the total instances of all nodes in the cluster, numbers of instances of each common node is given in brackets. Only nodes with more than 5 instances are shown. ‘n’ refers to the total number of unique geo-nouns or sense of place words significant at the $p < .05$ level with those significant at $p < .01$ shown in italics.

N	Common Nodes (n>5)	Significant toponyms		Significant Sense of Place words	
		n	Indicative words	n	Indicative words
163	Lake (99), boat (14), district (14), shore (13)	17	<i>Bowness-On- Windermere; Belle Isle; Ferry House; Pooley Bridge; Lylulphs Tower</i>	63	<i>east; adjacent; beautiful; head; guide; beauty; north; hire; supply; consult</i>
145	Road (62), Vale (19), River (10), Bank (8), Stream (8), Track (7), Mine (6), Path (6), Summit (6)	22	<i>Striding Edge; Greenside; Kirkstone Beck; Saddleback; River Greta</i>	45	<i>high; right; agreeable; right; lead; adventure; descend; turn; lead</i>
96	Pass (43), Inn (23), Bridge (13)	8	<i>Sca Fell; Nether Beck; Crook; Over Beck; Sty Head</i>	18	<i>little; mile; way; accommodation; refreshment; cross; call; enquire</i>
82	Mountain (43), Valley (21)	6	<i>Bowder Stone; Nags Head; Bow Fell; Castle Crag; Rosthwaite</i>	26	<i>high; attractive; grassy; feature; end; escape; surround; rejoice; run;</i>
71	Point (21), House (20), Town (13), Station (7)	3	<i>Grasmere; Helvellyn; Rydal Mount</i>	25	<i>neat; next; western; poet; gentleman;</i>

					<i>admission; visit; tell; speak;</i>
65	Country (21), Village (14), Hill (12), Wood (7)	6	<i>Morecambe; River Lune; Lancaster; Wast Water; Ulverstone</i>	35	<i>rich; fine; pleasing; evening; gaze; magnificence; walk; exhibit; dine</i>
56	Water (17), Waterfall (10), Fell (7), Region (7), Tree (6)	5	<i>Esthwaite; Buttermere; Ullswater; Ambleside; Crummock Water</i>	77	<i>comfortable; bleak; picturesque; marvellous; sterile; minute; appearance; tranquil; come; admire</i>

Table 3: Major clusters of locales associated with tourists with their associated toponyms and sense of place words. Details as for table 2.

N	Common Nodes (n>5)	Significant geo-nouns		Significant Sense of Place words	
		n	Indicative words	n	Indicative words
101	Keswick (38), Ullswater (20), Penrith (16),	11	<i>parish; stone; village; grounds; tower</i>	46	<i>beautiful; sublime; excellent; tour; skirt, m.; colouring; leave; come; follow</i>
66	Patterdale (16), Grasmere (15), Helvellyn (12), Wyburn (6)	10	<i>scree; path; stream; cottage; dell</i>	72	<i>high; bleak; charming; dark; pastoral; sterile; sweet; foot; horse; pedestrian; ascend; cross; lead</i>
66	Coniston (12), Cumberland (12), Westmoreland (9), Lancashire (6),	7	<i>tarn; track; farm; moor; summit; inn</i>	32	<i>beautiful; long; circular; leisurely; head; source; eye; pleasure; flow; look; present; see;</i>
61	Ambleside (24), Bowness-On- Windermere (8)	0	-	21	<i>high; western; favourable; horseback; foot; beauty; proceed; look; come</i>

Table 4: Major clusters of locations associated with travellers with their associated geo-nouns and sense of place words. Details as for table 2.

N	Common Nodes (n>5)	Significant toponyms		Significant Sense of Place words	
		n	Indicative words	n	Indicative words
349	Road (122), Pass (64), Inn (51), House (36), Village (15), Hotel (10), Carriage (9)	21	<i>Kendal; Threlkeld; Furness; Santon Bridge; Burton-in-Kendal</i>	115	<i>excellent; tolerable; unpleasant; mile; foot; accommodation; ale; rest; return; rejoin; sing</i>
272	Lake (107), Point (32), Water (31), Wood (23), District (21), Station (10), Boat (8), Island (8), Grounds (7), Moor (6), Promontory (6)	14	<i>Waterhead; Mellbreak; High Stile; Red Pike; Ullswater</i>	172	<i>fine; beautiful; sublime; view; time; scene; see; take; appear</i>
195	Mountain (95), Hill (29), Summit (20), Track (9), Precipice (9), Tarn (7), Tree (7)	13	<i>High Yewdale; Coniston Water; Hampsfell; Harrison Stickle; Strands</i>	106	<i>difficult; steep; tremendous; top; pleasure; ascent; tell; distinguish; toil</i>
101	River (22), Sea (21), Shore (16), Sands (11), Bay (7)	13	<i>Ulverston; Lancaster; River Keer; River Leven; Old Man of Coniston</i>	75	<i>clear; majestic; pleasant; guide; place; day; appoint; discover; meet</i>

99	Valley (28), Stream (23), Path (22), Bank (10), Bed (7), Dale (6)	10	<i>Grisdale Beck; Patterdale; Deep Dale; Helvellyn; Grasmere</i>	76	<i>ancient; sublime; delightful; horrid; foot; head; horse; lead; conduct; ascend</i>
97	Country (43), Town (20), Landscape (8), Garden (7), Seat (7)	2	Carlisle; Lorton	70	<i>mountainous; rich; dull; nature; eye; left; adjoin; belong; censure</i>
83	Rock (46), Waterfall (15), Region (13),	5	Borrowdale; Hartsop; Greta Bridge; Dove Crag; Kirkstone Beck	57	<i>Immense; chaotic; dreadful; fear; difficulty; rain; penetrate; dare; shrink</i>
65	Vale (39)	18	<i>Green Crag; Fornside; Armboth Fell; Newlands; Buttermere; Derwent Water</i>	61	<i>sweet; bleak; classic; way; head; south; go; pour; descend</i>

Table 5: Major locales associated with travellers and their associated toponyms and sense of place words. Details as for table 2.

N	Common Nodes (n>5)	Significant geo-nouns		Significant Sense of Place words	
		n	Indicative words	n	Indicative words
61	Cumberland (13), Penrith (13), Westmoreland (9)	8	<i>town; county; force;</i> <i>junction; coast</i>	25	<i>fine; rural; upper;</i> <i>attachment; cliff;</i> <i>creek seaport;</i> <i>contain; carry; view</i>
41	Keswick (16), Derwent Water (7), Skiddaw (4)	11	<i>vale; slope; grove;</i> <i>wood; ridge</i>	35	<i>little; small; poor;</i> <i>large; excellent; sun;</i> <i>distance; terror; call;</i> <i>shine; steal; cultivate</i>
30	Kendal (16), Windermere (6), Ambleside (5)	9	<i>town; market; station;</i> <i>lake; tree</i>	15	<i>great; roman;</i> <i>industrious; head;</i> <i>side; ground; build;</i> <i>situate; occupy;</i>

Table 6: Major clusters of locations associated with inhabitants with their associated geo-nouns and sense of place words. Clusters with more than 5% (21) of the total inhabitant instances are included, other details as for table 2.

N	Common Nodes (n>5)	Significant toponyms		Significant Sense of Place words	
		n	Indicative words	n	Indicative words
142	Town (73), market (14), county (11), river (8), seat (7), bank (6), coast (6)	16	<i>River Greta;</i> <i>Ulverston;</i> <i>Whitehaven;</i> <i>Conishead; Carnforth</i>	66	<i>small; neat; ancient;</i> <i>mile; side; distance;</i> <i>contain; situate;</i> <i>derive</i>
132	Mountain (33), Vale (31), Valley (18), Cottage (10), Dale (9), Region (7), Fell (6), Summit (6)	9	<i>Keswick; Yewdale;</i> <i>Troutbeck; Souther Fell; Brothers Water</i>	102	<i>mountainous;</i> <i>delightful; fertile;</i> <i>life; winter; refuge;</i> <i>consider; enter; place</i>
103	House (39), Parish (19), School (12), Chapel (11), Common (11), Wall (6)	10	<i>Great Crosthwaite;</i> <i>Borrowdale; Penny Burn; Ulpha;</i> <i>Seathwaite</i>	174	<i>yearly; industrious;</i> <i>lawful; schoolmaster;</i> <i>money; wage;</i> <i>support; maintain</i>
99	Country (40), Lake (25), Water (21)	4	<i>Grasmere Water;</i> <i>Lake Windermere;</i> <i>Rydal Water; Rydal</i>	66	<i>requisite; bottom;</i> <i>ignorant; kind; good;</i> <i>simplicity; sheep; ply;</i> <i>serve; improve</i>
78	Village (21), Road (15), Church (14), District (14)	6	<i>High Cark;</i> <i>Allithwaite;</i>	57	<i>pastoral; calm;</i> <i>lonely; time; farmer;</i>

			<i>Flookburgh; Lorton; Portinscale</i>		<i>clergyman; take; decline; fish</i>
62	Land (16), Field (13), Wood (11)	1	Keswick	76	<i>little; spongy; beautiful; soil; plough; crop; furnish; convert; grow</i>
51	Street (10), Farm (7), Castle (5), Residence (5), Tower (5)	6	<i>Egremont; Lancaster; Derwent Water; New Galloway; Fylde</i>	47	<i>square; narrow; cheerful; person; canopy; dungeon; stand; hang; vault</i>

Table 7: Major clusters of locales associated with inhabitants with their associated toponyms and sense of place words. Details as for table 2.