



Alts, Abbreviations, and AKAs: Historical Onomastic Variation and Automated Named Entity Recognition

DOI:

[10.1080/15420353.2017.1307304](https://doi.org/10.1080/15420353.2017.1307304)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Butler, J. O., Donaldson, C. E., Taylor, J. E., & Gregory, I. N. (2017). Alts, Abbreviations, and AKAs: Historical Onomastic Variation and Automated Named Entity Recognition. *Journal of Map and Geography Libraries*, 13(1), 58-81. Advance online publication. <https://doi.org/10.1080/15420353.2017.1307304>

Published in:

Journal of Map and Geography Libraries

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Alts, Abbreviations, and AKAs: Historical onomastic variation and automated named entity recognition.

James O. Butler, Christopher E. Donaldson, Joanna E. Taylor, and Ian N. Gregory.

ABSTRACT:

The accurate automated identification of named places is a major concern for scholars in the digital humanities, and especially for those engaged in research that depends upon the gazetteer-led recognition of specific aspects. The field of onomastics examines the linguistic roots and historical development of names, which have for the most part only standardised into single officially recognised forms since the late nineteenth century. Even slight spelling variations can introduce errors in geotagging techniques, and these differences in place-name spellings are thus vital considerations when seeking high rates of correct geospatial identification in historical texts. This article offers an overview of typical name-based variation that can cause issues in the accurate geotagging of any historical resource. The article argues that the careful study and documentation of these variations can assist in the development of more complete onymic records, which in turn may inform geotaggers through a cycle of variational recognition. It demonstrates how patterns in regional naming variation and development, across both specific and generic name elements, can be identified through the historical records of each known location. The article uses examples taken from a digitised corpus of writing about the English Lake District, a collection of 80 texts that date from between 1622 and 1900. Four of the more complex spelling-based problems encountered during the creation of a manual gazetteer for this corpus are examined. Specifically, the article demonstrates how and why such variation must be expected, particularly in the years preceding the standardisation of place-name spellings. It suggests how procedural developments may be undertaken to account for such georeferential issues in the Named Entity Recognition strategies employed by future projects. Similarly, the benefits of such multi-genre corpora to assist in completing onomastic records is also shown through examples of new name forms discovered for prominent sites in the Lake District. This focus is accompanied by a discussion of the influence of literary works on place-name standardisation – an aspect not typically accounted for in traditional onomastic study – to illustrate the extent to which authorial interests in regional toponymic histories can influence linguistic development.

BODY:

Introduction

This article explores the procedural shortcomings of automated named entity recognition for corpora that contain a range of textual source material published prior to 1900. In what follows, we illustrate a range of place-name related issues that should be anticipated within any historical corpus. Since even slight spelling variations can introduce errors in geotagging techniques, this is a vital consideration to ensure higher rates of correct geospatial identification in historical texts. The article argues that the careful study and documentation of these variations can assist in the development of more complete onymic records, which in turn may inform geotaggers through a cycle of variational recognition. This article is not a critical assessment of Named Entity Recognition (NER) procedures as they currently exist, but an analysis of the specific types of name-based

issues found within a particular resource: the Corpus of Lake District Writing (CLDW). The article uses this corpus as a case study to demonstrate our approach to, and solutions for working with, occurrences of onomastic variation that create difficulties for geotagging techniques. In particular, the article seeks to encourage the integration and adoption of onomastic records in the historical and geographical analysis of textual source materials, including works of literature.

The Field of Onomastic Studies

Place-names provide a richly detailed subject for linguistic research, with the names of most settlements possessing histories that span hundreds of years in one form or another. The academic interest in place-names results from the role place-names play as descriptive markers of topography that can furnish insights into the geographical, social, and cultural history of specific points within a visible environment. Shifts and variations in the spelling of place-names are the basis of much of the interpretative discourse within traditional onomastic research. These changes are traced through a documented lineage in order to determine both the oldest forms of place-names and the features or locations to which those place-names most likely pertain based on regional and temporal norms. Gelling (1988) labels names ‘signposts to the past’, as the range of elements (both topographically derived *generics* and unique markers of *specifics* for a particular point) and their linguistic progression yield significant data for historical inquiries. Every region comprises a distinctive onomastic base landscape: a stock of common linguistic and cultural elements that feature across many of the individual names within that region. Identifying the developmental and variational patterns in these elements assists in the automatic identification of potential non-standard name forms.

As unique markers of reference and narrative framing, place-names provide a highly desirable category for dedicated mark-up and analysis within digital corpora. Digitising place-names allows for a wide range of mapping, analytic, and visualisation processes to be adopted. Gazetteers provide unified resources that geotag the lexical category, through pre-formed lists of place-names typically spanning multiple tiers of embedded categorisation ranging from the macro to the micro: for example, The United Kingdom – England – Cumbria – Furness – Ulverston – Hoard Hill – Sir John Barrow Monument (Hoard Lighthouse, The Hoard). Tagging software automatically provides geo-referenced data for every entity recognised in this manner, no matter in what context the reference appears. The gazetteer lookup process identifies instances of a name form as it appears within a gazetteer, highlights it, and inserts the necessary referential data allowed by the software, which can then be used for dedicated work or study (outlined in Donaldson, Gregory, and Taylor, 2016). There are three major issues with this approach that can be identified from this overview of the process: first, the inclusion of necessary spatial information at a local (minor names) level; second, the extent to which this automated process is able to compensate for forms that deviate from the modern, officially recognised spelling of a place; and third, the ability to distinguish between unique points that otherwise share common elements of a name. The importance of spatial description and reference, alongside a publication period spanning 250 years, renders the CLDW a strong resource for investigating the extent to which onymic variation influences domain (topographic) recognition.

Place-name Standardisation and the Corpus of Lake District Writing

Studies dedicated to the name of a single place can encompass several linguistic possibilities borne from the range of spelt forms (e.g. Whaley, 2010). Research into shared elements of names (e.g. Hough, 2010) demonstrates the expansive range of current and historical forms that may be derived from an original onymic root. Responding to such findings, scholars such as Whaley have advised that ‘ideally, a whole run of spellings beginning soon after the supposed date of coining’ should be ascertained in order to see past the ‘often unintelligible or deceptive’ modern forms (2016: xvi). In short, higher numbers of variant forms are desirable, so that more can be learnt about the particular socio-historic development of each unique name. Although Baron and Rayson (2009) discuss the benefits of automatic linguistic standardisation for digital corpora, the practices they advocate run counter to the aims of traditional onomastic (or indeed any historical-linguistic) research. Most place-names now have an officially standardised spelling, but this is a relatively recent development. In countries such as the UK, the mass standardisation of place-name spellings has largely occurred over the past two centuries through a staggered, gradual, and – often – politically contentious process (see Hewitt, 2011). This is a phenomenon to which our case-study corpus, the CLDW, attests. Before elaborating on this issue, however, it will be helpful to provide a brief overview of the corpus itself.

The CLDW is a heterogeneous sample of eighty texts (comprising more than 1.5 million word tokens) published between the years 1622 and 1903. In the main, these texts were selected from the lists of works catalogued by Bicknell (1990) and Norgate and Norgate (2014), who provide useful bibliographies of historical Lake District travelogues and guidebooks. Accordingly, for the most part, the texts in the CLDW are works of topographical literature that offer descriptive accounts of journeys to and through specific places in the Lakeland region. Significantly, though, the CLDW also contains a selection of fictional writing about the Lake District, including works in verse and prose that are concerned with the myths and legends that pertain to particular named locations. The broad date range of the CLDW, and the loco-specific nature of the works it contains, distinguish it as an ideal resource for illustrating and modeling responses to each of the three major issues we identified in the previous section.

The eighty texts included in the current version of the CLDW were manually transcribed as part of a set of interlinked research projects at Lancaster University, undertaken between 2007 and 2012 (see Rupp et al. 2013). Once transcribed, these eighty texts were parsed using data obtained from the Unlock Text Webservice and the Digital Exposure of English Place-Names project in conjunction with the Edinburgh Geoparser: an open access Natural Language Processing tool that, firstly, identifies place-names in the texts using an NER sub-component and, secondly, disambiguates them and assigns each place-name with a set of mappable co-ordinates (see Grover et al., 2010). As Rupp et al. (2013) explain, the output of the Edinburgh Geoparser was submitted to a thorough error analysis in order to eliminate any immediately apparent false-positive inclusions. The Geoparser, for an initial sample of just twenty-right texts, returned data for 5,048 place-names, which was reduced to 4,583 after these corrective measures were taken (falling in line with the general statistical outlier figure suggested by van Dalen-Oskam and Zuggert, 2003); however this initial sweep also recorded 1,276 that had not been recognised. As many of the texts in the corpus featured references to geographical features and locations outside of the Lake District, the precision and recall of the output data of the Edinburgh Geoparser was less than optimal. This data was, accordingly, of

limited value for conducting the kinds of historical analyses it was intended to facilitate.¹ However, the scale of the corpus permitted a dedicated manual cataloguing of named spatial references across the complete dataset to be conducted within a three-month period. The manual gazetteer found 3,682 unique places referenced throughout the corpus of eighty texts.

This cataloguing was tailored to the spatio-referential focus of the CLDW. The manual collection of all related name forms under a single entry for each referent allows this article to explore two aspects of GIS in relation to historical corpora that have not yet been subject to much critical attention: first, how onomastic research may both inform and benefit from spatial humanities projects; and second, how dedicated gazetteer tailoring can strengthen the geospatial profiles of such resources.

The standardised forms of the names of larger settlements were usually established at earlier dates, since these settlements were generally of higher relative importance within the economic development and administration of the region. The names of these settlements were, therefore, more frequently used in printed materials, and this resulted in a higher level of consistency. The anticipated variance in form is much smaller for such places. The smaller villages, hamlets, and other toponymies that were discussed less frequently in official capacities (if at all), by contrast, yield the largest range of variations. These are also the sites that typically do not appear in standard gazetteer; as these features are central to much of the CLDW material accurately tracing all possible forms for each unique site is a central concern for this corpus. Accounting for variant forms is necessary to ensure utmost accuracy for any resulting linguistic, literary, or geographic research to be conducted against these data points. Donaldson, Gregory, and Taylor (2016) briefly touch upon name-based issues, using the settlement *Bonmess* (alongside its variant forms *Bonus* and *Boulness*) as an example of gazetteer constraints, but the extent to which non-standard historical name forms may affect corpora should be examined in closer detail, so that methods to counter each type of issue may be more viably recommended.

Onomastic research typically focuses on the use of non-fictional records of localities in order to source historical forms. As a consequence, the role of literary works in shaping a namescape is frequently overlooked. The English Place-Name Society records for the two sample locations (Lodore and Coniston) do not contain the majority of the variant forms that feature within the CLDW and discussed in the following section, nor does it include any account of the Lake District's (significant) literary history. Studying the corpus for such influence confirms the claims of scholars such as Fulford (2013) and Yoshikawa (2014) with respect to the influence of the Romantic poets (particularly William Wordsworth, Samuel Taylor Coleridge, and Robert Southey, as will be discussed later in the paper) on the development of the Lake District's regional identity during the nineteenth century, and their influence on the eventual standardisation of name forms.

A typical entry for onomastic work and place-name dictionaries (reprinted from Whaley, 2006) takes the following form:

¹ Lakeland Geotext Explorer, *Spatial Humanities: Texts, GIS, Places*
<<http://www.lancaster.ac.uk/fass/projects/spatialhum/geotext/>> [accessed 07/01/2017].

AUTERSTONE NY4521 Barton.

Otherstone 1761, *Otterstone* 1773, *Alter-* or *Oder Stone* 1839, *Auterstone* 1859 (PNWestm II, 212).

◆ ‘A great limestone crag and scar, possibly so called from a supposed resemblance to an altar’ (PNWestm). The 1st el., that is, may be *altar* from OE *altar*, and spellings in *au-* are common 13th-16th cent. See also **stone**.

This entry indicates the earliest appearances of four different spellings, within a 100-year period. Preserving – as well as actually capturing – all variant spellings within a corpus is of significant importance for place-name analysis, as well as wider historical research. The specificity of any gazetteer is entirely dependent on the resources underpinning its creation, and although the Lancashire Place-Names Survey group will release a dedicated regional resource in the near future (planned for 2017), this will be derived primarily from a single source family: OS maps of the county, published 1900. The resource will offer a focused snapshot of the officially recognised names of major places and locations across the Lake District and facilitate dedicated county-based recognition, but it will suffer similar limitations to its wider regional counterparts when applied to material outside the immediate boundaries of the period, and will not contain the alternate forms, such as those shown in the record above.

A number of the authors in the CLDW draw attention to the lack of standardised name forms, typically apologising to the reader and explaining that they are dependent on local pronunciation to guide their writing. For many writers, the lack of standardisation was a sign of the Lake District’s backwardness. Arthur Young, for instance, complains that: ‘I am sensible throughout this tour of mis-spelt names; but many of the places I mention are not to be found in maps, I am obliged, therefore, to write from the ear’ (Young, 1770: 175). Similarly, Joseph Budworth warns that his spellings of place-names are phonetic. The opening page of his *Fortnight’s Ramble to the Lakes* offers the reader a vocabulary of ‘provincial Words’, which bears the caveat that ‘As the Author has no Vocabulary to go by, he judges [spelling] from the ear’ (Budworth, 1792: i). Such declarations, coupled with attempted transcriptions from local guides, allow for an assessment of the Cumbrian accent as it was comprehended by visitors. The historic market town of Ulverston, for instance, is written on one occasion as ‘Owston’ and on another as ‘Ooston’, and the Bowder Stone, a free-standing rock in Borrowdale Valley, appears as Boodther. Alexander Craig Gibson is representative of the touristic fascination with the Cumbrian conflation of words when he carefully records being directed to ‘t’Ho’ penny Yall’ us’ – or, the Halfpenny Ale House (Gibson, 1849: 118). This factor is the primary reason for the presence of variant forms in the CLDW, as non-native documenters depended upon local accounts of the region. Such a range is typical for any non-major (city) place-name during this period.

Samuel Taylor Coleridge, for his part, was particularly fascinated by the meanings of regional words. In a notebook entry written on a walking tour in August 1802, Coleridge records with delight the reasons behind the Cumbrian names that sounded so strange to him:

I found that the common Names, Towers, & Tozers are the same — /
er signifies ‘upon’ — as Mite-er-dale the Dale upon the River Mite /
Donnerdale — a contraction of Duddon-er-dale, the Dale upon the
River Duddon — So Towers, pronounced in the Vale *Te-ars* — &
Tozers is [are] those who live on *the Toes* — i.e. upon the *Knobbly* feet of

the Mountain / (Letter: STC to Sara Hutchinson, 6 August 1802, *CL II*: 845)

Coleridge was delighted that his habit of discovering human forms in the landscape appeared to be one that was deeply embedded in native Cumbrian culture. In another place, he played with the term ‘ness’ (meaning ‘ridge’), and explores the word’s phonetic and the mountain’s visual resemblance to ‘nose’ (Letter: STC to Sara Hutchinson, 1-5 August 1802, *CL II*: 839). Often, though, Coleridge failed to hear correctly when locals told him the names of places. He wrote these onomastic musings at what he described as an ‘Estate House called Toes’, at the foot of Scafell and the head of Wastwater. He thought that that it was ‘not possible that any name can be more descriptive of the Head of Eskdale’. In fact, Coleridge misunderstood his hosts’ accents: he was actually staying at Taws farmhouse, rendering his onomastic readings of the names he has collected on his journey problematic at best. Coleridge’s mishearing – and subsequent etymological misunderstandings – of these place-names is indicative of the kinds of processes that lay behind the onomastic variety on display in the CLDW. Coleridge hears a place-name in conversation with a local farmer; he records it in his notebook; later, he reflects upon what he supposes to be the etymological history of the region. Coleridge’s text was not, in the event, published, but it is nevertheless clear how the kinds of variation apparent in works that were made publicly available could very quickly confuse the Lake District’s toponymic heritage.

By demonstrating the extent to which this sort of historical onomastic variation appears throughout the CLDW, we hope to encourage discussion for the development of procedures to detect the issues typical of names which should be anticipated in any resources covering publication time frames prior to the twentieth-century. The use of corpora to assist in the creation of more complete historic namespaces adds this specialised branch of linguistics to the fields discussed by Gregory, Donaldson, Murietta-Flores, and Rayson (2015) as potentially benefiting from digital-spatial humanistic developments. Name records are typically built from a limited range of source (parish or map-based) material, so the provision of easily-searchable historical discourse that extends beyond this remit has the potential to expand significantly the capacities of these developmental inventories. These same records may, in turn, enhance the identificatory profiles for NER systems. This connection highlights the integral role onomastic consideration should have in tagging historical corpora, and the extent to which the two fields can mutually benefit from one another.

Place-name Variation across the CLDW

Two examples of complete naming inventories are provided from the gazetteer, containing all the spellings that appear in the texts that the CLDW comprises. These two locations – the waterfall at Lodore and lake at Coniston – offer the highest number of variant forms. They both suffer potential referential ambiguity, a consideration that impacts both onomastic development and spatial analysis.

LODORE (group: waterfall): Loder, Loadoar, Loadoar, Lowdore, Lowdore water-fall, Fall of Lodore, ‘the Niagara of Derwent Water’, Low Door, Gulph of Lodore, Fall of Lowdore, Cataract of Lodore, Falls of Low Dore, Fall of Low Dore, Falls of Lodore, Lodoor water-fall, Lowdoor, Lod. Falls, Lawdore,

Lawdoor, Lowdore Waterfall, Borrowcascade, Barrowcascade, [Borrowcascade], Cascade of Lodore.

CONISTON (group: lake): Thurstan, Coniston Lake, Coniston Water, Thurston, Conistone, Conistone Lake, Cunnistone Lake, Thurston Lake, Coniston Mere, Lake of Coniston, Conis- ton, Conyngs Tun, Conyngeston, Thorstane's watter, Turstinus.

As specifically noted within the title of this article, there are three distinct strands of name spelling variation identifiable across these two exemplar locations:

1. Alternate Forms

As might be expected, the most common issue is the spelling not conforming to the modern, standardised form. Lodore presents nineteen iterations, and Coniston nine, with these totals including additional terrain descriptors as part of their onymic identity. Every name was taken from the corpus as it appeared. The reasons for these slightly different forms correspond with the traditional onomastic issues discussed above, but for this amount to be found across the material comprising the corpus is demonstrative of the need for NER to be able to adapt to provide the necessary accuracy for geospatial analysis of such material. The alternates provided above highlight the subtle spelling shifts (e.g. *Lodoor*, *Lawdoor*, *Lawdore*) that may be anticipated across historical texts. Even individual sources are not consistent in their reference. These are the forms that provide critical material for onomastic research, but Whaley's (2006) entry for *Lodore*, for example, contains only five of the nineteen noted variants. Such a level of omission leads to potentially incomplete analysis as to how or why particular spellings were adopted as the standard form for a name. Even singular occurrences of minor variation might offer valuable insight as to a formational path.

2. Abbreviations

The process of the deliberate shortening of forms, typically those mentioned earlier within the text to facilitate easier internal reference, has the unintentional effect of obscuring the form for wider external recognition. The reader can identify *Lod. Falls* as one and the same with *Lodore*, but even though it is only missing three letters from its standard form, this would not be recognised as the same entity by any gazetteer software.

3. AKAs

Not every name naturally passes down a single path of linguistic development, with many locations changing their name completely at some point in their history. Such is the case with both exemplar locations, which contains names that do not hold any similarity with their standard form. The roots of Lodore have been traced to 1209: *Lagbedure*, a notable example of a rare hybrid between the Old Norse *lagr* ('low') and Old English *duru* ('door'). Whaley (2006), however, makes no reference to the forms derived from *Borrowcascade* found within the corpus. These forms appear later in the corpus, after the Lake District had become a popular tourist destination, and the shift was likely a result of their serving as one of the exemplar falls for the region. Conversely, the lake of Coniston undergoes a significant shift from its original form (the earliest record being *turstiniwatra* from 1157-63), likely derived from the name of a Nordic clan leader, with the adjoining settlement noted as *Coningeston* ('the King's estate or village') within the same document. The lake came to be referred to as *Coniston* on account of its proximity to the settlement,

and its earliest forms are chiefly referenced as a note of historical interest in the texts. There are, however, instances where these ancient forms are used romantically as actual locational references.

In addition to these linguistically-rooted variations, there is a fourth category of potential error that can affect the appearance of any name within a textual corpus, whether standard or one of the variant forms: typographical errors. Gregory, Cooper, Hardie, and Rayson (2015) note the potential for formatting to obscure the identification of any text that has been digitised. These errors have been introduced at some point in the history of a text, whether by the original printer, or during the later digitisation process, and can also lead to variations in the spelling of place-names. This category derives from two separate sub-issues: mid-word line-breaks, and misspellings. The latter is typically the consequence of two letters being switched around (e.g. *Borrrwocascade*), where the nature of variation results from accidental transposition or mistranslation, rather than an intentional linguistic shift. The appearance and format of mid-word breaks varies significantly between historical texts, and whilst digital transcription can convert these to a regular form (e.g. an en-dash followed by the `</ lb>` TEI markup, which may allow the two parts to be stitched together for later analysis), this depends entirely on the development protocols of individual corpora, as it is not automatically accounted for by NER systems. Consistency in the final format is key, but within the CLDW the style of these formatting marks varied significantly. Any words caught within these physical shifts are erroneously broken up into separate elements, and treated as such by gazetteer scans.

The customised gazetteer built for the CLDW was designed to allow for a wide range of geospatial appraisal. Alongside cataloguing every occurrence of variant forms for each unique place, it contained two supplementary fields of externally-defined information: Topographic Category, and In/External Placement. These aspects were included to extend the types of geospatial analysis that could be conducted on the corpus material. The latter was a binary option, intended to facilitate the ability to toggle the inclusion of places outside the Lake District alongside those within it, to increase the versatility of comparative assessments. Topographic assignment was the labelling of every named entity according to one of the nineteen classes deemed most relevant to the particular interests of the travel-writing material that comprise the corpus. These classes are broad enough to cover any new additions that might be made. They were chosen for their functional distinction (in terms of disambiguating referential elements), and their topographic relevance for the discourse particular to the region. Figure 1 shows the full range of categories chosen for this custom markup scheme.

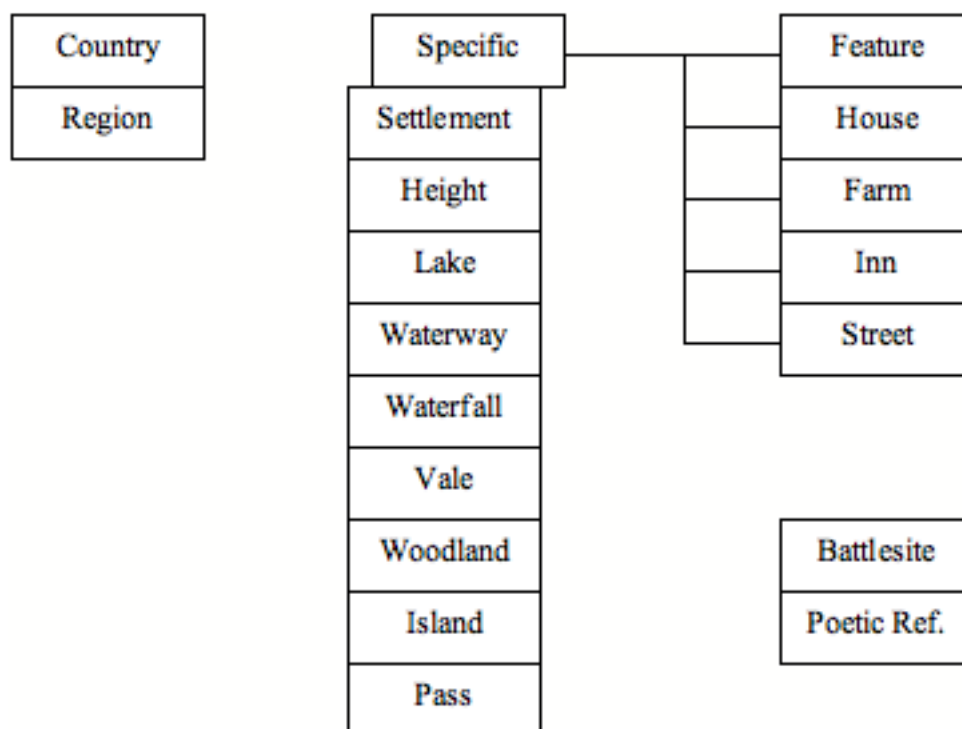


Figure 1: Categories for the CLDW’s topographic-led gazetteer.

This topographic attribution of every named site that appears in the corpus will enable detailed analysis to be undertaken on the loco-specific trends of how different authors guide their readers through the area. The gazetteer’s intent was to catalogue every valid place-name for each unique location referenced throughout the corpus; as the topographic categories indicate, this was not restricted solely to major names.

As noted in the opening paragraphs, non-commercial gazetteers are limited in the scale as to which they operate; GeoNames, for instance, does not recognise many of the entities listed under the heights category within the CLDW, nor does it contain several smaller but significant features, such as the local landmark *Hoad Hill*. 3,718 distinct locations were found across the corpus, with 1,277 of these entries featuring at least one additional alternative form that geotaggers would struggle to identify (assuming every name, major and minor, could be covered by a gazetteer). The major lakes of the region consistently saw high rates of variation in spelling issues, as shown in Table 1.

Ennerdale	8
Grassmere	8
Derwentwater	9
Ullswater	12
Thirlmere	12
Bassenthwaite	13

Table 1: Hydronymic variation for major Lake District features.

Given the role of these places across the dominant genres of the source material in the CLDW (travelogues and guidebooks), missing any non-standard forms of the name would significantly impact geospatial analysis. These examples alone provide sufficient evidence to promote corpus analysis as an important approach for finding previously-

unaccounted historical forms for onymic studies and records, but as the material that comprise such resources can span a wide range of genres, so too can evidence for non-academic interest in naming be uncovered.

Minor Names and Local Features

To return to Coleridge's documented interest in naming patterns, it is interesting to note that the class of hyper-specific minor names that Coleridge records are absent from standard gazetteers because they identify details that do not appear on OS maps: significant rocks, parts of fells, individual farmhouses. Other Lakeland texts adopt a similarly detailed approach to the naming of Lake District places, and working with a region-specific corpus like the CLDW emphasizes the quantity of place-names missing from a conventional gazetteer. Visitors to the region throughout the eighteenth and nineteenth centuries were astounded at the number of toponyms in local use that transformed an apparently wild space into a series of highly specific locales. In her travelogue *A Journey Made in the Summer of 1794*, Ann Radcliffe writes with palpable astonishment of the degree to which the Lakeland fells had been linguistically mapped in a way not recorded on any official document:

Not only every fell of this wild region has a name, but almost every crag of every fell, so that shepherds sitting at the fire-side can direct each other to the exact spot among the mountains, where a stray sheep has been seen. (Radcliffe, 1795: 422)

To the unwitting visitor, the fells themselves appear to be discreet entities. What the colloquial names ascribed to 'almost every crag of every fell' reveals, however, is the extraordinarily detailed knowledge of the local terrain that is derived from personal experience working with – and living near – the landscape. The toponymic specificity of this region is indicative of the central role that the topography played in the Lake District's cultural and linguistic heritage. Only a relatively small amount (134, of which 112 are internal) of these minor features that denote a specific point on an otherwise named aspect of the landscape are referenced in the corpus. The unappended 'Specific' label was applied to similar points of marked tourist value that are not natural formations, such as bridges, ruins, cairns, etc. These minor names hold a significant presence (totaling 570 entries, 402 internal) within the gazetteer.

The names of inns, public houses, and other commercial establishments are especially transient, and typically comprise generic forms that are not linguistically or identifiably distinct without an expanded context for their placement. There are several possible inns named *King's Arms*, *Queens Head*, *Black Bull*, or *White Lion* across the Lake District, and the forms that these non-permanent minor names take cannot be predicted. They are valid names of distinct and identifiable places, but their forms require a higher level of contextual framing than geoparsing techniques can currently offer. This set of minor names is also prevalent throughout the corpus (with 176 unique entries in the Lake District, and 64 external), given their important role to the tourist industry. Only the entities that bore a settlement as part of their name (such as *The Buttermere Inn*, *The Furness Abbey Hotel*, *The Ullswater*) were recognised by the geotagger, but – as might be expected – only partially, with the incorrect referent. Commonly-held names of little permanence represent the trickiest class of names to automatically identify and connect with accurate geospatial information. The use of location-specific concordance is essential for the true

identification of all sites that bear minor-names of these types. Wordsworth's home at Dove Cottage in Town End, a hamlet adjacent to Grasmere, is indicative of the transitory nature of building names. Prior to the Wordsworths' tenancy there (1799–1808), the cottage had been a pub known as *The Dove and Olive Bough*. In fact, it was not named Dove Cottage until around 1890, eighty-two years after the Wordsworths left the dwelling (Atkin, 2013).

Two additional classes of referential oddities found within the CLDW are the names of poetic place and the sites of battles. The former class contains only 28 entries, which are not able to be assigned any mappable coordinates, but still hold semantic value through their referential use. The upland nature of the region's topography exposed visitors to greater extremes – of height, temperature and rain – than they tended to be used to. The terrain encouraged detailed descriptive prose throughout much of the source material (the appearance and development of three specific adjectives is discussed in greater detail by Donaldson, Gregory, and Taylor, 2016). Another strand of poetic detailing brings metaphoric comparison to fantastic places into frequent use by authors. Comparisons to such mythical places as *El Dorado*, *Mount Olympus*, *Arcadia*, and the *Garden of Eden* appear throughout the corpus. However, this class of onymic-derived allusions straddle fiction and historical reality; places of classical referential significance (such as *Babylon* and *Bandusia*) reference symbolic ideals, rather than literal places. Identifying such place-names – those that fall far from the subject under investigation, both literally and figuratively – as a unique set is an important consideration for texts that feature any prosaic amelioration.

The reverse process of referential assignation creates similar issue, where historical battle naming conventions typically denote a local significant feature or settlement which serve as the marker of the event. Acceptable nomenclature was deemed an important protocol of conflict in order to ensure chronicler consistency. Smith and Crane (2001: 130-131) discuss the problem of referential leaps of a similar ilk, where straight geotagging would be technically correct in identifying the place-name, but semantically false as the referent is not the place but rather the event. This issue of entity metonymy has been studied within NER systems (Brun, Ehrmann, and Jacquet, 2007, provide an overview of the challenges of extracting deep syntactic relationships for this purpose), but requires both an awareness of its presence and application of additional identification routines to filter references to events from those of the place itself – both of which may feature together, dependent on the context of the discourse. Marking these entities as the sites of battles allows the end-user to toggle their inclusion as it may suit individual research requirements. Although there are only 14 entities of this type in the CLDW, every occurrence references a battle that occurred outside the Lake District (e.g. *Agincourt*, *Bannockburn*, *Flodden Field*, *Waterloo*, four of these references refer to the site as a place at other parts of the corpus.

Despite the manual onomasticon (complete name-data set) checks to remove obvious false-positives, a number of tagged entities that superficially appear as valid illustrate additional issues in place-name use and automated identification. For example, there is a close relationship between family names and those of associated places (whether the former supplied the latter, or this attribution is reversed), such as Sir Christopher Lowther of *Lowther* (recognised twice), the knightly family of Harrington, or Mr Crosthwaite, Matthews, or Jenkin. Likewise, some generic topographic elements have come to reference a specific place in an unmodified state: for example, *Wike*, *Waterfall*, *Neb*, *Gates*, *Syke*, *Seat*, *Thwaite*. Any discussion of these elements is mistakenly tagged as

belonging to those referents. There is also the adoption of other lexis into modern place-names that are intended as such within the dataset. Examples of this class of erroneous identification include: *Priest*, *Louder*, *Cherry Tree*, and *Norseman*. The transference of major names to reference local features is a common aspect of minor names serving to evoke a sense of distance or some association with the general characteristics of the referent (Field, 1989: xvii). As these are recognisable places, borrowed and repurposed for local use, they are still tagged as the primary place, despite the fact that they are located far outside the boundaries of the Lake District. An example of this transpositional confusion is provided by Williamson (1884: 79), who notes that: ‘by an authority on the Pillar this false rock has been christened *Pisgab*, and the gap, *Jordan*’. These names are figurative attributions from the local perspective.

Problems also arise from identical names shared by multiple valid references, whether these are more prominent occurrences that are not the intended referent, or shared names of equal potential within a local area. Otley (1823) specifically comments on the Wallow Crag (of Hawes Water and Keswick – the latter is now officially distinguished as ‘Walla Crag’), Whitbarrows (of Witherslack and Penrith), and Eagle Crag (of Patterdale, Ennerdale, and Buttermere), to name three topographic examples of equal referential likelihood. Donaldson, Gregory, and Taylor (2016) also note the issue of identical names referring to multiple locations, using an example of *Bonmess* as a settlement along the Solway Firth as well as the probable (within the remit of the CLDW’s focus) primary candidate on the shores of Windermere. The difficulties in weighing geotaggers to prefer particular occurrences around a specified region is not a novel concern for context-free automated geo-tagging (see Kim, Vasadarni, and Winter, 2016), but the custom gazetteer drew attention to a related issue that has not seen critical attention: that is, similarly-named but topographically-distinct entities. These entries (taken solely from the corpus) share a common full-name element, but refer to a range of distinct environmental features, with the example locations holding close parallels with the following:

Lodore (waterfall): *High Lodore* (settlement), *Lodore House* (specific – house), *Lodore Hotel* (specific – inn), *Lodore* (waterway),

Coniston (lake): *Coniston fells* (height), *Coniston Old Man* (height), *Monk Coniston* (settlement), *Church Coniston* (settlement), *Coniston* (settlement), *Coniston Waterhead* (settlement), *Coniston Bank* (specific – house), *Coniston Hall* (specific – house), *vale of Coniston* (vale), *Coniston Beck* (waterway).

The original geoparsing process for the corpus struggled to distinguish between these locations as a result of the shared name element, with priority given to the most prominent feature that bore it. The latter set poses particular problems for gazetteers derived from present day resources, as *Monk Coniston*, *Church Coniston*, and *Coniston Waterhead* have shifted in reference, amalgamated, and reappeared in new contexts, since the corpus material was published (the British History Online account of this area [Farrer and Brownbill, 1914] discusses these shifts in greater detail, noting that 1894 saw major shifts in the township boundaries). The category labels above reflect the true intended referent within the corpus material. The category of ‘Vale’ is particularly challenging in this regard, as such areas require a primary focal referent to establish placement, but refer to a wider spatial concept that is distinct from that entity. For example, the *Vale of Coniston* refers to all the settlements just discussed, as well as numerous topographic features, taken together as a whole. Entries of this type require specialist knowledge of the region. Given the linguistic developments seen across other entities, it would be easy

to assume that the three settlements above were analogous to the town of Coniston itself, especially with the shared semantic link between ‘monk’ and ‘church’. They are derived from the same linguistic and topographic root, but are still distinct entities.

Onomastic research frequently overlooks the role of contextual literary works in shaping a namescape. The Lake District, being so remote from the easily navigable spine of England, presents a valuable opportunity to explore such relationships through digital humanities projects that span a range of written genres beyond the material that typically provides onymic record data. One of the example locations chosen for this case study also provides an opportunity to demonstrate such a connection. Among Lakeland waterfalls, the falls of Lodore have long held pride of place—not least because of the frequency with which they have figured in the works of poets whose writings have influenced modern conceptions of the Lake District’s scenic beauty. Notably, the region’s most famous poet William Wordsworth chose to pay special tribute to Lodore at the beginning of one of his earliest printed poems (see Wordsworth 1793: 1). Nearly a decade later, in 1802, Coleridge thought the waterfall ‘beyond all rivalry the first & best Thing of the whole Lake Country’ (Letter: STC to Sara Hutchinson, 25 August 1802, CL II: 54). Intriguingly, though, poetical accounts of Lodore can be seen to have done as much to increase the waterfall’s fame as they did to establish the standardised spelling of its name. Long before the first instance of Lodore recorded by Whaley — ‘Lodore Cascade 1901 (OS)’ (2006: 216) – the Revd. Dr John Dalton employed the spelling Lodore in his *Descriptive Poem* of 1755 (Dalton 1755: 19-20; ll. 244, 258), which celebrates the way the waterfalls’ tremendous force enchants the eye, overwhelms the ear, and thrills the mind. The declaration with which Dalton first invokes the falls is indicative: ‘Let other streams rejoice to roar | Down the rough rocks of dread Lodore’ (ll. 243–44). Here assonance, alliteration, and rhyme conspire to create a sense of auditory exhilaration, with the sound of ‘rejoice’, ‘roar’, ‘rough rocks’, ‘down’, and ‘dread’ creating a euphonic pattern with ‘Lodore’. These rhymes helped to establish a perceived pronunciation, and the poem was an important moment in the development of a standardized name for this site.

Dalton’s *Descriptive Poem* has long been recognised as one of the earliest and most influential accounts of the Lake District. In his seminal bibliography of Lake District guidebooks and travelogues, for example, Peter Bicknell distinguishes Dalton as one of ‘the first of a long line of dons and divines who were to record the Lakes’, and he describes Dalton’s poem as a foundational document in the history of Lake District tourism (Bicknell 1990: 1, 22). Far less attention, however, has been given to the way that Dalton’s poem, which was widely quoted and reprinted, helped to introduce spellings that have trickled down through the canon of writing about the Lakes region and into modern orthography. One should, of course, be cautious of attributing absolute orthographical influence to a single text, especially given that the spelling used in eighteenth-century printing was largely left to the whim of the typesetters (see Flint 2011). But it is nonetheless noticeable that whereas significant early accounts of the Lake District such as Thomas Gray’s epistolary journal (1769), Thomas West’s *Guide to the Lakes* (1778), and William Gilpin’s *Observations, Relative Chiefly to Picturesque Beauty* (1786) favour different spellings (*Landoor*, *Lowdore*, and *Lodoar* respectively), Dalton’s ‘Lodore’ passed into several subsequent works that either quoted his poem or imitated its evocation of Lodore’s ‘roar’. Hence, one finds in Adam Walker’s *Remarks Made in a Tour From London to the Lakes* (1792) descriptions of both ‘the tremendous Cascade of Lodore [...] dashing from rock to rock with a hideous roar’ and ‘Lodore thunder[ing] a ground Bass with its roaring cascade!’ (Walker 1792: 92, 102). Even the relevant lines from

Wordsworth's aforementioned poem, *An Evening Walk*, attest to Dalton's influence. Here, we read of the poet listening 'to hear the roar | That stuns the tremulous cliffs of high Lodore' (Wordsworth 1793: 1; ll. 5-6). Notice here again the euphonic pairing of roar and Lodore.

In the long-term, however, the echo of Dalton's poem one finds in Wordsworth verse is less important than that discernible in the writings of Wordsworth's fellow 'Lake poet' Robert Southey. In the Lakeland section of his oft-reprinted *Letters From England: By Don Manuel Alvarez Espriella* (1807), Southey's fictitious persona, the enterprising Don Espriella, dubs Lodore 'the most famous [waterfall] in all this country' and he admires the way the celebrated 'cataract [...] thunders down' the mountainside 'through a chasm in the rocks, which are wooded with birch and ash trees' (1807: 149, 152). The cascade, Espriella exuberantly concludes, 'is at once the finest work and instrument of rock and water that I have ever seen' (1807: 152). Here, once again, it is as much the site of Lodore as its sound that compels the writer and informs his account. Moreover the spelling preferred is Lodore as opposed to the various variant spellings that one finds in other texts from this period. Priscilla Wakefield's *A Family Tour Through the British Empire* (1808: 74), for example, praises 'the famous fall of Lowdore'.

Even more significant than the celebration of Lodore found in Southey's *Letters from England*, however, is that which appears in his poem 'The Cataract of Lodore, Described in Rhymes for the Nursery'. Written in 1820 for the amusement of Southey's children, the poem is an exuberant experiment in verbal and metrical variation. 'The Cataract of Lodore' exceeds even Dalton's *Descriptive Poem* in the extent to which it uses alternating rhymes, rhythms, and schemes to capture the visual and auditory sensation of standing before the waterfall, as reproduced in the Appendix for this paper. Latter-day critics have tended to regard 'The Cataract of Lodore' as an amusing oddity. Southey's most recent biographer, William Speck, describes the poem as 'an impressive exercise in onomatopoeia' (2006: 189). More recent criticism has, however, approached the poem with even greater sensitivity. Notably, Tim Fulford has argued that 'The Cataract of Lodore' is both 'a visual as well as a sound poem': it is a poem that uses lineation to mirror the drop of the cascade, which widens towards its base; equally, though, it is a poem whose playful use of metre and diction at times approaches the quality of 'pure sound' (Fulford 2013: 54).

Southey later offered it to Joanna Baillie for publication in her *Collection of Poems, Chiefly Manuscript, and from Living Authors* (1823). The poem proved a great success, and rapidly rose to prominence among visitors to the Lake District. Following its publication, Baillie wrote to Southey to say how the poem had 'pleased and amused her', and she predicted that 'the younger part of [her] readers [will be] running about with portions of it in their mouths and shaking their heads to the measure, for these six months to come' (Baillie, 2010: 27). Frequently quoted, referenced, reprinted, and anthologized, 'The Cataract of Lodore' became a signature piece—and one deeply imprinted in the popular cultural consciousness by the countless tourists and travellers who invoked it in their accounts of visiting the famous waterfall. But it is important to recognise that as much as Southey's poem succeeded Dalton's in shaping popular perceptions of the falls of Lodore, these poems also collectively exercised an influence on the gradual standardisation of the spelling of the waterfall's name, particularly with regards to its rhyming with 'uproar'. None of this literary influence is noted in linguistic records, but it serves as a primary example of the additional context that may be ascertained from consulting the wider range of material offered through digital corpora.

Conclusion

Historical corpora that comprise multiple genres offer a wide range of datasets to discover unofficial linguistic shifts and non-conventional components that might otherwise not be accounted for in traditional onomastic analysis. This article has highlighted the importance of retaining original name forms, and it has explored the reasons behind typical variations in those forms, in order that an understanding of toponymic development may assist in the development of geotagging modifications and refinements that can increase the accuracy of NER procedures. Patterns in regional naming variation and development, across both specific and generic name elements, can be identified through the historical records of each known location, which offers significant potential for identifying new spelling forms, allowing for a more complete entry record for the place, leading to improved accuracy for spatial tagging. The provision of large amounts of untraditional (literary) source material for finding new name forms holds great benefit for the field of onomastics, which in turn is an essential consideration for the geospatial identification of non-standard modern name forms that this article has demonstrated. The three primary problems that contributed to a significantly impacted entity recognition rate were: spelling variation, geospatial disambiguation, and lack of coverage for non-major sites. Resolutions for these issues, as we have suggested, may be developed through the further integration of historical onomastic records, as well as an awareness of typical spelling-based variations encountered in name elements, into NER procedures.

This case study has demonstrated the limitations of word-list gazetteers that provide only a single official standardised form (or small amount of potential variants). Such a format cannot accurately capture the geospatial identity of historical documents. Techniques that can adapt these standard forms to account for the four types of issues identified within this article are essential for improving geospatial recognition, and this in turn may offer significant benefit to onomastic research. The potential for digital corpora to allow more complete referential histories has been discussed, as has the importance of exploring the role of published literature, beyond maps and regional records, as important contextual factors in directing onymic development and standardisation.

ACKNOWLEDGEMENTS

The research leading to these findings has received funding from the Leverhulme Trust, as part of the Geospatial Innovation in the Digital Humanities research project (Ref: RPG-2015-230), and from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007–2013): ERC grant Spatial Humanities: Texts, GIS, Places [agreement number 283850]. We are grateful to Dr Claire Grover and the Unlock Text support team in Edinburgh, and to Jayne Carroll and the Digital Exposure of English Place-names (DEEP) project for granting up access to the their data. We are also grateful for the data and resources made available online by Martin and Jean Norgate, creators of the Old Cumbria Gazetteer, *Guides to the Lakes* <<http://www.geog.port.ac.uk/webmap/thelakes/html/lgaz/lgazfram.htm>>.

APPENDIX: Southey's Literary Influence of the Naming of Lodore.

THE CATARACT OF LODORE

DESCRIBED IN RHYMES FOR THE NURSERY, BY ONE OF
THE LAKE POETS

How does the water come down at Lodore?

Here it comes sparkling,
And there it lies darkling;
Here smoking and frothing,
Its tumult and wrath in,
It hastens along, conflicting strong;
Now striking and raging,
As if a war waging,
Its cavern and rocks among.

Rising and leaping,
Sinking and creeping,
Swelling and flinging,
Showering and springing,
Eddying and whisking,
Spouting and frisking,
Turning and twisting,
Around and around
Collecting, disjecting
With endless rebound:
Smiting and fighting,
A sight to delight in;
Confounding, astounding,
Dizzying and deafening the ear with its sound.

Receding and speeding,
And shocking and rocking,
And darting and parting,
And threading and spreading,
And whizzing and hissing,
And dripping and skipping,
And whitening and brightening,
And quivering and shivering,
And hitting and splitting,
And shining and twining,
And rattling and battling,
And shaking and quaking,
And pouring and roaring,
And waving and raving,
And tossing and crossing,
And flowing and going,
And running and stunning,
And hurrying and scurrying,
And glittering and frittering,
And gathering and feathering,

And dinning and spinning,
 And foaming and roaming,
 And dropping and hopping,
 And working and jerking,
 And guggling and struggling,
 And heaving and cleaving,
 And thundering and floundering;
 And falling and brawling and sprawling,
 And driving and riving and striving,
 And sprinkling and twinkling and wrinkling,
 And sounding and bounding and rounding,
 And bubbling and troubling and doubling,
 Dividing and gliding and sliding,
 And grumbling and rumbling and tumbling,
 And clattering and battering and shattering,
 And gleaming and streaming and steaming and beaming,
 And rushing and flushing and brushing and gushing,
 And flapping and rapping and clapping and slapping,
 And curling and whirling and purling and twirling,
 Retreating and beating and meeting and sheeting,
 Delaying and straying and playing and spraying,
 Advancing and prancing and glancing and dancing,
 Recoiling, turmoiling and toiling and boiling,
 And thumping and flumping and bumping and jumping,
 And dashing and flashing and splashing and clashing,
 And so never ending, but always descending,
 Sounds and motions for ever and ever are blending,
 All at once and all o'er, with a mighty uproar,
 And this way the water comes down at Lodore.

(Baillie, Ed. 1823: 280–83)

BIBLIOGRAPHY

Baillie, J. [ed.] (1823) *Collection of Poems, Chiefly Manuscript, and from Living Authors*. London: Longman & Co.: 280–83.

Baillie, J. (2010) *Further Letters of Joanna Baillie*, ed. by T. McLean. Madison and Teaneck: Fairleigh Dickinson University Press.

Baron, A. & Rayson, P. (2009) 'Automatic standardisation of texts containing spelling variation: How much training data do you need', In: M. Mahlberg, V. González-Díaz, & C. Smith (Eds), *Proceedings of the Corpus Linguistics Conference: CL2009*. Lancaster: Lancaster University.

Bicknell, P. (1900) *The Picturesque Scenery of the Lake District, 1752–1855: A Bibliography*. Winchester: St Paul's Bibliographies.

British History Online – Townships: Hawkshead and Monks Coniston. Digitised from: Farrer, W. and Brownbill, J. (1914) *A History of the County of Lancaster: Volume 8*. London: Victoria County History, 376-380. <<http://www.british-history.ac.uk/vch/lancs/vol8/pp376-380>> [accessed: 07/12/2017].

- Brun, C., Ehrmann, M., & Jacquet, G. (2007) A Hybrid System for Named Entity Metonymy Resolution. In: Z. Vetulani & H. Uszkoreit (Eds) *Human Language Technology: Challenges of the Information Society (Third Language and Technology Conference, LTC 2007, Poznan, Poland, October 5-7, 2007, Revised Selected Papers)*, 118–130.
- Budworth, J. (1792) *A Fortnight's Ramble To The Lakes In Westmoreland, Lancashire, And Cumberland*. London: Hookham and Carpenter.
- Coleridge, S.T. (1966) *The Collected Letters of Samuel Taylor Coleridge: Volume 2.*, 6 vols., ed. by E.L. Griggs. Oxford: Clarendon Press.
- van Dalen-Oskam, K and Zundert, J.J. (2003) 'Finding Names with the Autonom Parser'. *DigiCULT.info* 6, pp. 47-49
- Dalton, J. (1755) *A Descriptive Poem, Addressed to Two Ladies*. London: J. & J. Rivington.
- Donaldson, C.E., Gregory, I.N., & Taylor, J.E. (2016). Implementing corpus analysis and GIS to examine historical accounts of the English Lake District. In: *The Making of a Historical Atlas*. Northeast Asian History Foundation.
- Donaldson, C.E., Bushell, S.C., Gregory, I.N., Rayson, P.E., & Taylor, J.E. (2016). 'Digital literary geography and the difficulties of locating 'Redgauntlet Country''. *Studies in Scottish Literature* 42.2, 174-83.
- Field, J. (1989) *English Field Names: A Dictionary*. Gloucester: Alan Sutton.
- Flint, C. (2011) *The Appearance of Print in Eighteenth-Century Fiction*. Cambridge: Cambridge University Press.
- Fulford, T. (2013) *The Late Poetry of the Lake Poets: Romanticism Revised*. Cambridge: Cambridge University Press.
- Gelling, M. (1988) *Signposts to the Past: Place-Names and the History of England*. Chichester: Phillimore. [3rd edition, 1998].
- Gregory, I.N., Donaldson, C.E., Murieta-Flores, P. and Rayson, P. (2015) 'Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research'. *International Journal of Humanities and Arts Computing* 9, 1–14.
- Gregory, I.N., Cooper, D., Hardie, A., & Rayson, P. (2015). 'Spatializing and analyzing digital texts: corpora, GIS and places'. In: D. Bodenhamer, J. Corrigan, & T. Harris (Eds.), *Deep maps and spatial narratives*. (Spatial Humanities). Bloomington, Ind.: Indiana University Press.
- Grover, C., Tobin, R., Byrne, K., Woollard, M., Reid, J., Dunn, S., and Ball, J. (2010) 'Use of the Edinburgh Geoparser for Georeferencing Digitized Historical Collections'. *Philosophical Transactions of the Royal Society A*, 368, 3875–89

- Hewitt, R. (2011) *Map of a Nation: A Biography of the Ordnance Survey*. London: Granta.
- Hough, C. (2010) 'The name-type Maid(en)well'. *Nomina* 33, 27–44.
- Kim, J., Vasardani, M. and Winter, S. (2016) 'Similarity matching for integrating spatial information extracted from place descriptions'. *International Journal of Geographical Information Science*, 1–25.
- Norgate, M. and Norgate, J. (2014) *Guides to the Lakes*.
 <<http://www.geog.port.ac.uk/webmap/thelakes/>> [accessed 7/12/2017]
- Otley, J (1823) *A Concise Description of the English Lakes, and Adjacent Mountains: with general directions to tourists, notices of the botany, mineralogy, and geology of the district, observations on meteorology, the floating island in Derwent Lake, and the black-lead mine in Borrowdale*. London: John Richarson.
- Radcliff, A. (1795) *A Journey made in the Summer of 1794*. London: G.G. & J. Robinson.
- Rupp, C.J., Rayson, P., Baron, A., Donaldson, C.E., Gregory, I.N., Hardie, A., and Murrieta-Flores, P. (2013) 'Customising geoparsing and georeferencing for historical texts'. *Proceedings of the 2013 IEEE International Conference on Big Data*, 59-62.
- Smith, D.A. and Crane, G. (2001) 'Disambiguating Geographic Names in a Historical Digital Library'. *Research and Advanced Technology for Digital Libraries (Volume 2163, Lecture Notes in Computer Science)*, 127–136.
- Southey, R. (1807) *Letters From England: By Don Manuel Alvarez Espriella, Vol. 2*. London: Longman & Co.
- Speck, W.A. (2006) *Robert Southey: Entire Man of Letters*. New Haven, CT: Yale University Press.
- Wakefield, P. (1808) *A Family Tour Through the British Empire, 2nd edn*. London: Darton & Harvey.
- Walker, A. (1792) *Remarks Made in a Tour From London to the Lakes*. London: G. Nicol.
- Whaley, D. (2006) *A Dictionary of Lake District Place-Names*. Nottingham: English Place-Name Society.
- Whaley, D. (2010) 'Scarborough Revisited'. *Nomina* 33, 87–100.
- Williamson, C. N. (1884) 'The Climbs of the English Lake District'. *All the Year Round* 35, 77-82, 110-14.
- Wordsworth, W. (1793) *An Evening Walk*. London: J. Johnson.