Understanding the Spatial Complexity in Landscape Narratives Through Qualitative Representation of Space

Erum Haris [1]
School of Computing, University of Leeds, UK

Anthony G. Cohn
School of Computing, University of Leeds, UK
The Alan Turing Institute, London, UK

John G. Stell
School of Computing, University of Leeds, UK

Abstract

Narratives are the richest source of information about the human experience of place. They represent events and movement, both physical and conceptual, within time and space. Existing techniques in geographical text analysis usually incorporate named places with coordinate information. This is a serious limitation because many textual references to geography are ambiguous, non-specific, or relative. It is imperative but hard for a geographic information system to capture a text's sense of place, an imprecise concept. This work aims to utilize qualitative spatial representation and natural language processing to allow representations of all three characteristics of place (location, locale, sense of place) as found in textual sources.

2012 ACM Subject Classification Computing methodologies → Knowledge representation and reasoning

Keywords and phrases Narratives, Qualitative spatial representation, Natural language processing

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.37

Category Short Paper

Supplementary Material *Dataset*: `https://github.com/UCREL/LakeDistrictCorpus`

Funding The support of the Economic and Social Research Council (ESRC) under grant ES/W003473/1 is gratefully acknowledged.

Acknowledgements We also thank the entire team of the Spatial Narratives project (`https://spacetimenarratives.github.io/`) for their discussions on the CLDW and this work.

1 Introduction

Narratives are a fundamental way of organizing experiences and giving them meaning. Narrative theory privileges time by emphasizing the sequence of events, yet all narratives also imply a material, spatial world. Narratives represent events and movement, both physical and conceptual, within time and space. Most work to date looking at geographies within digitized texts has focused on extracting and mapping well-recognized toponyms i.e. place names with geographic coordinates. However, in practice, people conveniently log and share their narrative experiences in imprecise natural language. They likely recall locations qualitatively [6]. For instance, they would share information like "Lake Gardens is a 10-minute walk east

[1] Corresponding author. **Contributions**: EH wrote the original submission based on work done by AGC; AGC and JGS acquired the funding and commented on the revised version.

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 37; pp. 37:1–37:6
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of the monorail station" instead of "From latitude, longitude 3.177383, 101.7076 walk to 3.1733, 101.6959" . In such cases, qualitative and approximate spatial statements are more useful than exact location coordinates [5].

The term "landscape narratives" describes the interaction and mutual relationship between story and place. Place has multiple characteristics including: the location of an object or event; the natural and built physical environment that makes up a place, termed its locale; and sense of place, the accumulated events, actions, and memories attributed to a location [1]. In this work, we aim to preserve the narrative structure of text data, and move much further by incorporating geographical features that cannot be easily mapped ("a town", "the hills"), relative spatialities ("near to", "20 minutes from"), and senses of place ("picturesque","enclosed"). This will be achieved by combining natural language processing (NLP) and qualitative spatial and temporal reasoning (QSTR) to extract, locate, and contextualise imprecise information about places.

This work particularly explores the Corpus of Lake District Writing (CLDW) consists of travel writing and tourist literature from 1622 to 1900 describing the English Lake District. It contains 80 geoparsed texts by famous writers, such as Wordsworth and Coleridge, and works from lesser-known writers and travel guides. The corpus portrays leisure journeys where the aim is to describe the landscape and evoke an emotional response. It offers us the opportunity to assess how human experiences of space are represented in writing. It has already been extensively analysed in geographical information systems (GIS) using quantitative places [3] [13]. Applying our methods to these texts will allow us to develop an enhanced understanding of what semantic, grammatical, and geographical tropes can be discovered in individual texts and entire corpora, allowing us to better understand the senses of place recorded by individual writers and their aggregate grouping. The scope of the details provided in this paper is limited to spatial information part of the larger work.

2 Qualitative Spatial Representation for Textual Data: Background and related work

Spatial representation usually locates objects in a quantitative frame. Natural language, on the other hand, offers an imprecise and vague setting. The emergence of qualitative spatial representation (QSR) provided a systematic description of space in this domain. It defines locations as regions, but without the need for geometric information. It uses a common-sense level of abstraction to represent spatial knowledge in terms of connections or spatial relationships between one region and another region. Hence, spatial relations are one of the fundamental aspects for formal qualitative representation of space. Examples include topological, direction and distance associations [5]. The difficulty lies in analysing textual data to identify connections between regions, so one can perceive spatial representations that go beyond toponyms [12].

Additionally, QSR makes use of the logical properties of relationships between entities, enabling data consisting of entities with qualitative relationships between them to be handled as a network of nodes and labelled links. This provides a computational representation of qualitative data even in cases, such as geographical relationships, where specific details are unknown [10]. In this way, it can push spatial study beyond the limitations imposed by quantitative geographical information. Some notable and relevant studies include the work on the analysis of the 16th century Mexican maps [10] to model complex and diverse spatial information, including social and symbolic conceptions depicted in the maps. The study explores the implications of qualitative spatial reasoning for historical and archaeological research. Another interesting study [9] adopted this technique along with corpus linguistics and NLP in humanitarian forensic research to analyze social and media releases from official sources to gain an understanding of the death of migrants at the Texas-Mexico border. This illustrates the utility of qualitative spatial representation in Humanitarian GIS. Semantic role labelling [7] draws on natural language semantics for the extraction of qualitative descriptions from text is yet another application of qualitative spatial reasoning. A pertinent study [8] presents mapping of natural language to formal spatial representation using a two-level approach where the first level deals with spatial role labelling and the second level maps these arguments to formal spatial calculi.

3 Proposed Methodology

This work is part of a larger project focusing on the extraction, qualitative representation, analysis and visualization of the CLDW [11] . Each phase of the project forms an independent module. In this context, the proposed work will utilize QSTR to demonstrate how one can extract a network of geographical and temporal entities and relationships combining location, locale, and sense of place from the CLDW. The first stage is the development of an appropriate reasoning mechanism using existing and extended qualitative spatial and temporal calculi which refer to the sets of relationships encoding spatial and temporal semantics with associated inference mechanisms. This stage will primarily use the annotations from the first module which focuses on the corpus linguistics and NLP techniques for named-entity recognition (NER) and related tasks [4]. Spatial relations can be understood from multiple, sometimes conflicting viewpoints. Existing research [2] identifies a "user level" corresponding in our case to the writer's intended meaning. Two other levels in [2] are called "geometrical" and "computational" which provide respectively an abstract mathematical denotation and an implementation. An alternative abstract level to [2] is to specify meanings of relations by logical formulas. As part of the first stage, we aim to develop an ontology to define various categories of spatial and temporal entities and relations exist at the user level. The identified entities and relations will be used to construct semantic triples. An analysis of the meanings of spatio-temporal relationships in the corpus requires the transformation of these extracted user-level triples to a suitable abstract level, and then to the computational level to connect with geographical visualizations. The overlap between these relationships and existing spatio-temporal calculi will be identified. Calculi will be designed that extend current AI-focused work in QSTR to narratives. This will not only allow the expression of qualitative relationships, but also those which are vague and imprecise.

In the second stage, semantic representations of narratives will be developed as networks with locations, temporal entities and events as network nodes and spatio-temporal relationships as edges, or links between nodes. This will be followed by a network analysis step since patterns within networks will represent more complex relationships. Hence, the overall task is to explore the extent to which existing QSTR reasoning methods are adequate to allow deeper meanings to be extracted from these representations and design new inference methods to allow hidden meanings and consequences to be made explicit.

3.1 Representing spatial information using QSR

This section provides details on representing spatial information using QSR. It presents an interpretation of the different types of qualitative relations found in the CLDW as narrative writing. The purpose here is to illustrate the complexity of describing space, which in turn requires a range of inference mechanisms to appropriately represent respective relations.

Hence, it is imperative to provide a non-mathematical description of the mapping from object-entity relation to possible geometric space representation using QSR inference rules. From the given examples, it can be realized that some of the relationships are straightforward, while others being abstract require development of inference mechanisms. A few relations are highly domain-specific and require background knowledge before modelling. Consider a snippet from the CLDW related to one of the landmarks named Pooley Bridge with NER tagging represented in figure 1 (The NER and semantic tagging system for extracting spatial entities has been developed by our colleagues at Lancaster as part of the first module [4]). Here, an interpretation of possible relationships and ambiguous terms is presented for a few sentences:

▶ **Example 1.** *From Penrith two roads lead to Pooley Bridge, about six miles distant, which spans the Eamont just at its issue from Ulleswater. Either road may be taken, be we recommend that which follows the Shap road to Eamont Bridge. Carleton Hall is near to it on the left. Cross the bridge, and take the first road to the right. At this point, on the left, are the druidical remains called King Arthur's Round Table and Mayborough. Immediately after crossing Pooley Bridge, the road runs along the western shore of Ulleswater to Patterdale, a distance of ten miles; but, before proceeding along it, the tourist would do well to take a walk of a few miles along the eastern shore, in the direction of Martindale, from several points on which he will obtain a good view of the lake.*

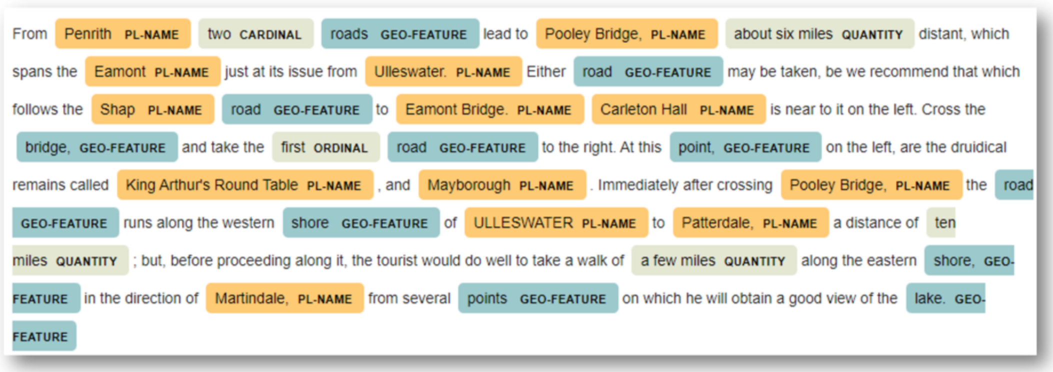

Figure 1 NER and semantic tagging system developed by our project colleagues I.Ezeani and P.Rayson [4].

Sentence 1: From Penrith two roads lead to Pooley Bridge, about six miles distant, which spans the Eamont just at its issue from Ulleswater.

QSR-based relations and interpretation:

- **place(penrith)**
- **place(pb)** : Pooley Bridge, pb, is both a bridge and a town, one needs to consider both, and have two different logical names
- **distance(penrith, pb, about(6), miles)** : approximate measurements with about(), a reasoning mechanism could be developed for "aboutness"
- **road(road1)** : "roads" are really "routes" since there are various roads and road segment which constitute them
- **road(road2)** : one could explicitly say that road1 and road2 are 6 miles long or possibly have a rule which says that if r is a road from x to y and the distance between x and y is z, then r must be at least z long

- **end(road1,penrith)** : a separate start(,) relation would be required if road is to be made oriented
- **end(road1,pb)**
- **end(road2,penrith)**
- **end(road2,pb)** : a further statement could be added to note the implicit fact that there are *only* 2 roads from penrith to pb
- **bridge(pb)**
- **spans(pb,eamont)** : some rules could be added about spanning and bridges that lets one infer that one can get from one side to the other via the bridge, and also one can only cross a river via a bridge or a tunnel or a ford. Moreover, all bridges span something and have two ends
- **river(eamont)** : not explicit in the text, requires background knowledge
- **source(eamont,ullswater)**
- **lake(ullswater)** : not explicit in the text, requires background knowledge

Sentence 2: Carleton Hall is near to it on the left.
QSR-based relations and interpretation:

- **near(carleton Hall,pb)** : "near" is a vague term and one has to separately consider what axioms/rules might apply to it in this kind of geographical context. Vague spatial terms have received a lot of attention in the literature but without any definitive treatment. Note that "near" is not transitive, i.e. from near(a,b) and near(b,c) we cannot conclude near(a,c).
- **direction(carleton Hall, left eamont bridge,shap)** : Directions provide different frames of references. Here, a direction predicate is defined with four arguments: the "figure" (i.e. the thing being pointed out), the direction (here left), the "ground" (i.e. the place from where the direction is being pointed out from), and the direction the person pointing is facing. Note that if roads have a start and an end then fourth argument is not required. Almost certainly a number of different direction predicates for the different situations are needed.

4 Conclusion

This paper presents a small part of a larger research project which aims to uncover spatial and temporal dynamics of narratives. The research will create a step-change in the way we explore the geographies described in large textual collections by exploring how GIS and related tools can identify, analyse and visualise qualitative senses of place alongside the quantitative spatial information more typically used in geographical information science (GISc).This will allow us to redefine the way that computer technologies represent place and transform the abilities of social scientists and humanists to understand and interpret narrative accounts about place.

References

1 John Agnew and David N. Livingstone. *The SAGE Handbook of Geographical Knowledge*. SAGE Publications Ltd, London, 2011. doi:10.4135/9781446201091.

2 Eliseo Clementini. A conceptual framework for modelling spatial relations. *Information Technology and Control*, 48:5–17, 2019. doi:10.5755/j01.itc.48.1.22246.

3 Christopher Elliott Donaldson, Ian N. Gregory, and Joanna E. Taylor. Implementing corpus analysis and gis to examine historical accounts of the english lake district. in: Historical atlas. In Peter Bol, editor, *Historical atlas: its concepts and methodologies*, pages 152–172. Northeast

Asian History Foundation, Seoul, 2017. URL: https://eprints.lancs.ac.uk/id/eprint/81425/.

4 Ignatius Ezeani, Paul Rayson, and Ian N. Gregory. Extracting imprecise geographical and temporal feferences from journey narratives. In *Proceedings of Text2Story — Sixth Workshop on Narrative Extraction From Texts*, 2023. URL: https://ceur-ws.org/.

5 Erum Haris, Keng Hoon Gan, and Tien-Ping Tan. Spatial information extraction from travel narratives: Analysing the notion of co-occurrence indicating closeness of tourist places. *Journal of Information Science*, 46(5):581–599, 2020. doi:10.1177/0165551519837188.

6 Gregor Josseź, Klaus Arthur Schmid, Andreas Züfle, Georgios Skoumas, Matthias Schubert, Matthias Renz, Dieter Pfoser, and Mario A. Nascimento. Knowledge extraction from crowd-sourced data for the enrichment of road networks. *Geoinformatica*, 21(4):763–795, 2017. doi:10.1007/s10707-017-0306-1.

7 Parisa Kordjamshidi, Martijn Van Otterlo , and Marie-Francine Moens. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Trans. Speech Lang. Process.*, 8(3), 2011. doi:10.1145/2050104.2050105.

8 Parisa Kordjamshidi, Martijn Otterlo, and Marie-Francine Moens. From language towards formal spatial calculi. In *Proceedings of the Workshop on Computational Models of Spatial Language Interpretation at Spatial Cognition*, volume 620, pages 17–24, 2010. URL: http://ceur-ws.org/Vol-620.

9 Molly Miranker and Alberto Giordano. Text mining and semantic triples: Spatial analyses of text in applied humanitarian forensic research. *Digital Geography and Society*, 1:100005, 2020. doi:10.1016/j.diggeo.2020.100005.

10 Patricia Murrieta-Flores, Mariana Favila-Vázquez, and Aban Flores-Morán. Spatial humanities 3.0: Qualitative spatial representation and semantic triples as new means of exploration of complex indigenous spatial representations in sixteenth century early colonial mexican maps. *International Journal of Humanities and Arts Computing*, 13(1-2):53–68, 2019. doi:10.3366/ijhac.2019.0231.

11 Space Time Narratives Project. Understanding imprecise space and time in narratives through qualitative representations, reasoning, and visualisation. https://spacetimenarratives.github.io/, 2023.

12 Robert Smail, Ian N.Gregory, and Joanna E.Taylor. Qualitative geographies in digital texts: Representing historical spatial identities in the lake district. *International Journal of Humanities and Arts Computing*, 13(1-2):28–38, 2019. doi:10.3366/ijhac.2019.0229.

13 Joanna E. Taylor and Ian N. Gregory. *Deep Mapping the Literary Lake District*. Bucknell University Press, 2022. doi:10.2307/j.ctv2v55bsf.