Towards an Extensible Framework for Understanding Spatial Narratives

Ignatius Ezeani, Paul Rayson, Ian Gregory
Lancaster University
United Kingdom

Erum Haris, Anthony Cohn, John Stell
University of Leeds
United Kingdom

Tim Cole
University of Bristol
United Kingdom

Joanna Taylor
The University of Manchester
United Kingdom

David Bodenhamer, Neil Devadasan
Indiana University Purdue University Indianapolis
United States

Erik Steiner, Zephyr Frank, Jackie Olson
Stanford University
United States

ABSTRACT

Spatial narratives help us to organize experiences and give them meaning. Previous approaches to understanding geographies in textual sources focus on geoparsing to automatically identify place names and allocate them to coordinates. Those are highly quantitative, and are limited to named places with coordinates, and have little concept of time. Narratives of journeys indicate that human experiences of geography are often subjective and more suited to qualitative representation. Geography is not limited to named places but incorporates the vague, imprecise, and ambiguous, e.g "the camp", or "the hills in the distance", and relative locations such as "near to", "on the left", "north of" or "a few hours' journey from". Places are organized worlds of meaning, characterized by experience, emotion, and memory as well as by geography. In this paper, we discuss our approach to gaining more insight from textual data beyond the toponyms and introduce an extensible framework for extracting, analyzing, and visualizing spatial elements that define the *'locale'* as well as the *'sense of place'* referenced in text using two test corpora –the Corpus of the Lake District Writing and Holocaust Survivors' Testimonies.

CCS CONCEPTS

• Information systems → Geographic information systems; **Presentation of retrieval results**; **• General and reference → Experimentation**; **• Human-centered computing → Collaborative and social computing systems and tools**.

KEYWORDS

spatial narratives, toponyms, place names, geographical feature nouns, location, locale, sense of place, spatio-textual regions, corpus annotation, datasets, named entity recognition, qualitative spatial representation, ontology

ACM Reference Format:
Ignatius Ezeani, Paul Rayson, Ian Gregory, Erum Haris, Anthony Cohn, John Stell, Tim Cole, Joanna Taylor, David Bodenhamer, Neil Devadasan, and Erik Steiner, Zephyr Frank, Jackie Olson. 2023. Towards an Extensible Framework for Understanding Spatial Narratives. In *Proceedings of 7th ACM SIGSPATIAL Workshop on Geospatial Humanities (31st ACM SIGSPATIAL GeoHumanities Workshop '23).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3615887.3627761

31st ACM SIGSPATIAL GeoHumanities Workshop '23, Nov 13–16, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/3615887.3627761

1 INTRODUCTION

Scholars who want to better understand the geographies in texts initially embraced geographic information systems (GIS) to manage spatial information, under rubrics such as historical GIS and literary GIS. These tools and methods were essential but not sufficient. GIS and its related technologies allow users to determine the geometry of space; fuzzy data, conceptual space, and relative time too often pose insurmountable problems for these tools. It will be necessary to replace this more limited quantitative representation of space with a view that emphasizes the intangible and socially constructed world captured in texts and not simply the world that can be measured.

Our project[1] aims to understand space and time in narratives through qualitative representations, reasoning, and visualization. We start from the perspective that many digital textual sources contain multiple representations of place and geography. These representations can be explicit such as a placename (e.g. 'Keswick') and much progress has been made in mapping and analyzing these using geoparsing and geographical text analysis. However, there are implicit representations of place e.g. '*the town*', '*a mountain*', or relational such as in the sentence '*Shortly after leaving Keswick we crossed a stream where we turned off the road and began to climb the hill*'. We, therefore, consider it necessary to start by defining our understanding of the underlying concepts of 'location', 'locale', and 'sense of place', originally formulated by Agnew and Duncan [2], in this work:

- **Location** refers to the 'geometric' dimensions of place [8] that are easily mapped in GIS using latitude and longitude. Texts often use toponyms for locations which are mostly reduced to coordinates for mapping.

[1] https://spacetimenarratives.github.io/

- **Locale** describes the material, social, cultural, political, and economic dimensions of space. Cresswell [8] describes this as 'the material setting for social relations', signaling a complex set of qualities that are not easily reducible to a single point on a map. In practice, we can identify nouns for geographical features such as 'house', 'lake', 'mountain' etc. which are often vague or ambiguous and cannot be directly mapped in Euclidean space.
- **Sense of place** is more complex still and includes all of *'the subjective and emotional attachment people have to place'* [8]. Scholars often distinguish between two paths of enquiry related to the sense of place: place as a locus of attachment, which privileges the bond with a geographic place; and place as a center of meaning, which focuses on how experience, emotional responses, and relationships shape the perception of a place [32]

This paper presents our preliminary approaches to developing computational methods to define, identify, and extract entities and relationships that reflect these types of references from the text for further analysis and visualization. We used a combination of techniques from natural language processing (NLP), geographical information science (GISc), qualitative spatio-temporal representation and reasoning (QSTR), and visual analytics. The contributions of this work include:

1. expanding the discourse on spatio-temporal elements in historical texts beyond geo-coding to include the locale and sense of place by defining additional spatial elements while preserving the narrative structure.
2. annotating the text corpus to create datasets that reflect occurrences of spatial elements of interest - toponyms, geographical features, events, times, or sentiments - and their interactions and relationships.
3. developing computational methods and tools to automatically extract, analyse and visualise these spatio-temporal references to gain relevant insight from written text.

2 PROBLEM DEFINITION

Using the collection of texts described in section 4, we want to be able to explore them individually and collectively with the aim of understanding and building a narrative around their location, locale, and sense of place. Specifically, we would like to know, among other things:

1. how writers as a group shape the sense of place
2. whether the sense of place is temporally or geographically continuous or otherwise
3. how and when tropes relate to a sense-of-place.
4. whether there are spatial and linear trajectories embedded in texts.
5. whether these trajectories are grouped by experience, genre, or subject.

To address these questions, we recognise the essence of not only effectively identifying and extracting toponyms and other features from text, but also to analyse them for better understanding. Figure 1 presents an overview of our framework for addressing these questions. At its basic level, the framework receives text input, extracts the relevant spatial entities, and then analyses and visualises it. For this work, these entities are discussed in section 5.1.

3 RELATED WORK

Extracting geospatial information from textual data is considered essential by researchers. This is because, unlike other datasets on place, textual data contains valuable human experience information e.g. peoples' *feelings* toward a place [17, 18]. Such information is important in creating effective computational models of places [12, 23], and also for geospatial data that are only available in unstructured historical texts e.g. old newspaper archives, historical archives [27]. In these cases, harvesting geospatial data from texts, otherwise referred to as 'geoparsing', is necessary for enabling advanced spatial analysis using geoparsers [4, 19]

In general, nouns are important in text processing [39]. Identifying, extracting and analysing geographical feature nouns is an essential aspect of geospatial analysis and ontology design [9, 22, 36]. Understanding the emotional expressions around a place is a key to fully appreciating its sense of place. Technique for extracting and analysing place emotions from texts are commonly explored by researchers [3].

Qualitative spatial representation and reasoning (QSRR or QSR, and sometimes extended to QSTR to encompass both temporal and spatial aspects), provides computational solutions for handling qualitative spatial relations, such as 'next to', 'in front of', 'to the left of' and more. Such relationships are commonly encountered in spatial humanities data, presenting a challenge to the traditional GIS. In his review, Stell [38] delves into the origins of QSR and advocates for exploring how these methods can complement GIS as a computational tool within the field of humanities.

In geographic space description, spatial semantics does not offer a mechanism for drawing deductions. Hence, techniques from QSR are inevitable to obtain a systematic elucidation of space by relating one statement to another, using a common-sense level of abstraction and making inferences. Smail et al. [35] acknowledge the utility of QSR in defining places qualitatively, such as "Keswick" or "the road", both can be used as regions in QSR. Such an analysis would facilitate spatial representation beyond toponyms and coordinate-based geography. At the application level, QSR has been used to analyze 16th-century Mexican maps [25] to model intricate and diverse spatial information, encompassing both social and symbolic aspects portrayed in the maps. Another work [24] employs a similar approach in combination with corpus linguistics and NLP for humanitarian forensic research to analyze social and media reports from official sources and gain insights into the migrants' deaths. A notable study by Kordjamshidi *et al.* [20] introduces a method for mapping natural language to formal spatial representation. It follows a two-tiered approach; the first level is dedicated to spatial role labeling, and the second level maps these roles to formal spatial calculi.

4 DATA SOURCE AND ANALYSIS

Our methods –which will be explained in section 5.2 –focused on two distinct corpora – Corpus of the Lake District Writing and Holocaust Survivors' Testimonies. Table 1 shows the comparative

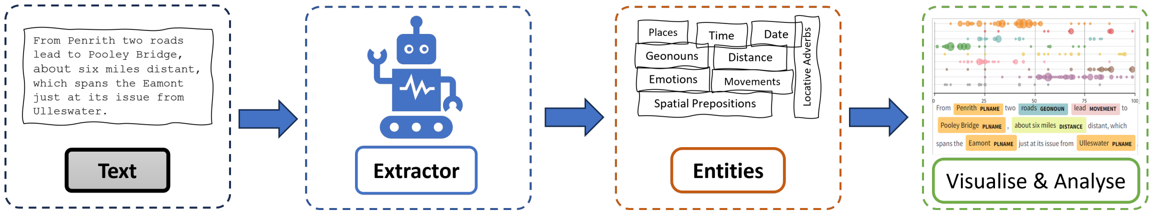

Figure 1: This is an overview of the frameworks for extracting and visualising the spatial entities in text corpora

analysis of the two corpora at the document, sentence, and word (token) levels.

4.1 Corpus of Lake District Writing

The Corpus of Lake District Writing (CLDW)[2] comprises 80 texts and around 1.5 million words that describe the Lake District [41]. The earliest texts are from the seventeenth century and run through to the early twentieth century containing travel literature, fiction, histories, letters, and diaries. It includes works by well-known Lake Poets such as William Wordsworth and Samuel Taylor Coleridge. There are also accounts of visits to the Lake District by prominent writers such as Daniel Defoe, Celia Fiennes, and other less well-known writers. There are also a number of tourist guides stretching from Thomas West's (1778) "A Guide to the Lakes" to Black's (1900) "Shilling Guide to the English Lakes" [14]. While drawn from a variety of styles and genres, the majority of the corpus comprises tourist guides and travel narratives.

4.2 Holocaust Survivors' Testimonies

The Corpus of Holocaust Survivors' Testimonies (HST) comprises a random selection of transcripts of one thousand oral history interviews undertaken by the USC Shoah Foundation Visual History Archive[3] in the 1990s. This corpus of around 21 million words represents only a fraction of the more than a thousand interviews in the archive. Each of the thousand transcripts follows a broadly similar format. They include a series of questions posed by the interviewer and the corresponding answers from the survivor being interviewed. The focus of each interview is mainly on the individual's experiences during the Holocaust which are explored in a broadly chronological order. Each interview – generally of around two hours duration – devotes approximately 20 percent of the time to pre-war life, 60 percent to wartime experiences focused on the events of the Holocaust, and 20 percent to post-war life [34]. In short, these are not full life histories, but more focused interviews asking about wartime experiences across a series of sites of incarceration or hiding. These sites serve as anchors in the narratives that describe survivors' wartime trajectories.

[2]CLDW and the gold standard dataset [33] are available here: https://github.com/UCREL/LakeDistrictCorpus

[3]Information about the USC Shoah Foundation Visual History Archive can be found https://sfi.usc.edu/what-we-do/collections

5 EXPERIMENTAL METHOD

We start by defining the key features or entities of interest in the text as presented in section 5.1. These entities are identified, categorized, and marked up in the text as described in section 5.2 using a combination of automatic and manual methods.

5.1 Spatial Entities

PLNAME:
This refers to any occurrence of a place name or toponym in the text. The initial approach applied a combination of methods including searching for gazetteer entries and using a named entity recognition tool.

GEONOUN:
These are geographical feature nouns appearing in the text. We started with a manually created list of 139 nouns (e.g. ***river***, ***road***, ***waterfall***, etc) as well as their inflections (*rivers*, *roads*, *waterfalls*).

EVENT:
This captures the descriptions of events or activities such as a '*ride*', '*walk*', or '*excursion*'. We bootstrapped the process via semantic tagging with PyMUSAS (movement) followed by manual corrections

EMOTION:
These are expressions of sentiments and emotions captured in the text. For example e.g. *'**delightful** excursions'*, *'**pleasant** walk'*, or *'**horrible** ravines'*. We applied sentiment analysis based on a standard sentiment lexicon [21] to extract the sentiment words as well as compute a sentiment score for each textual unit.

DATE or TIME:
References to dates (e.g. '*March 1803*', '*the beginning of the century*') and time (e.g. '*10 o'clock*', '*late in the evening*') references were extracted in the text

PERSON: Names and references to people Spacy NER and manual annotation

5.2 Spatial Entity Extraction

Spatial entity extraction is the process that identifies and extracts relevant spatial elements in text using the **Extractor** upon which the Spatial Entity Extraction Demo tool being developed by our team is based. The workflow presented in Figure 2 highlights two key components – **Dataset Creation** and **Model Training** – involved in the creation of the extraction model based on standard NLP entity extraction libraries.

Table 1: Comparative analysis of the size and contents of the Corpus of the Lake District Writing and Holocaust Survivor's Testimonies

	Corpus of the Lake District Writings	Holocaust Survivor's Testimonies
File count	80	1000
Sentence count	57,451	816,800
Words (tokens) count	1,524,718	21,516,122
File size range (words)	1,072 – 96,029	4852 – 84,051
Averages file size (words)	19,059	21,516

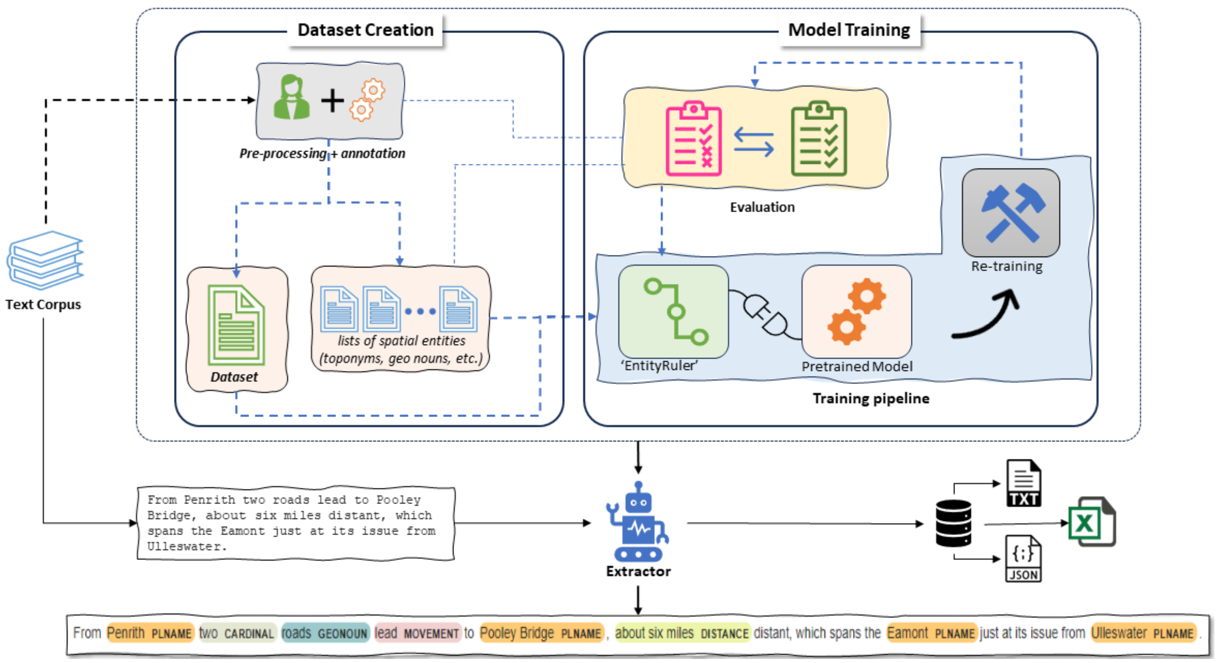

Figure 2: This is an overview of the dataset creation and the spatial entity extraction workflow. There are three key connected components – pre-processing and annotation, model training and *model evaluation* – and other components that produced the Spatial Entity Extractor

5.3 Dataset Creation

This mostly happens within the *'pre-processing and annotation'* stage in the workflow presented in Figure 2. Here, the raw corpus (i.e. the CLDW) is pre-processed using a semi-automatic method to correct and remove the errors (typos, misspellings, etc.) and prepare the data instances in the required structure i.e. paragraphs in this case. The choice of the CLDW (instead of the HST) corpus was informed by the fact that there is an existing gold-standard version deeply annotated with placenames (PLNAME) and other metadata [33]. Additional annotations were introduced to include other categories as shown in Table 2. This workflow is intentionally iterative as we constantly review the categories and approaches to extracting them.

We initially applied the methods described by Ezeani *et al.* in [10] involving a combination of regular expression (regex)[4], *spaCy*'s[5] named entity recognizer and semantic tagging with PyMUSAS[6]. Human annotators reviewed and corrected the outputs from these systems to improve the quality of the annotations. The main output from this section is the dataset from CLDW as well as the lists of all possible tokens and spans belonging to each of the categories mentioned above. Table 2 shows an overview of these entity categories and their examples in our corpora.

5.4 Model Training

This section focuses on the training of the **Extractor** – a bespoke and more generalizable model for extracting the relevant spatial

[4]A regular expression [13] is a sequence of characters that specifies a search pattern in the text. See https://en.wikipedia.org/wiki/Regular_expression
[5]spaCy is a free and open-source python library for general NLP [16].
[6]PyMUSAS is an open-source Python implementation of the semantic tagger for English and other languages: https://pypi.org/project/pymusas/

Table 2: Entity categories and examples: This table lists and describes the spatial entity categories as well as the examples and sources from the two corpora CLDW and HST.

Category	Description	Example	Source
PLNAME	Place names	`From **Penrith** two roads lead to **Pooley Bridge**`	CLDW
GEONOUN	Geo feature nouns	`Cross the **bridge**, and take the first **road** to the right.`	CLDW
TIME	Temporal references	`If I survive **tonight**, I will remember the 9th of April`	HST
DATE	Date mentions	`Americans will probably be here **the following day**.`	HST
EVENT	Events or activities	`You lost your brother and father in the **Holocaust**?`	HST
MOVEMENT	Movements	`**Cross** the bridge, and **take** the first road to the right.`	CLDW
SPPREP	Spatial preposition	`Carleton Hall is **near** to it on the left.`	CLDW
LOCADV	Locative adverb	`I wasn't **here**, I was sleepwalking`	HST
DISTANCE	Distance measure	`Lowther Castle is **about four miles** from this bridge.`	CLDW
EMOTION	Sentiments	`The march is a **nightmare**, completely **obliterated**.`	HST

entities from text as a named entity recognition (NER) task. NER methods are commonly applied to spatial entity extraction [6, 28, 40]. Existing NER tools can extract toponyms without gazetteers as well as spatio-temporal references in text.

Using the annotated dataset as well as lists of spans (tokens or exact phrase matches) in relevant entity categories (e.g. gazetteers, list of geonouns, spatial prepositions, locative adverbs, etc.), we included additional rules to the already existing named entity recognition pipeline in the `spaCy` thereby tweaking a standard English NLP model in spaCy e.g. `en_core_web_sm`. This was achieved with the `EntityRuler` feature. The resultant extraction model is evaluated with the gold standard subset of the CLDW which contains 28 texts (about one-sixth of the entire corpus), carefully selected to be representative of the entire corpus. The customized tag `<cdplace>` was used to mark up placenames e.g. a variety of different locations (regional, national, and international), landmarks, and geographical formations. Empirical evaluation can only be done on the placenames because we are still working on the gold standard annotations for other categories. However, the performance of the model is generally improved by manually and iteratively inspecting the error points, reviewing the entire process, and retraining the model which we refer to as the **Extractor**.

5.5 Evaluation

We have limited gold standard datasets as work is still ongoing on the annotation and review processes of our corpora. Therefore, we are only able to evaluate the **Extractor** model on a few entity categories –place names (PLNAME), geographic feature nouns (GEONOUN), locative adverbs (LOCADV) and distance measures (DISTANCE). The baseline model uses regex rules with the entity category lists to extract spatial entities. We also applied a standard named entity recognition feature from an off-the-shelf open-source NLP library spaCy. The performances of these two were compared with that of our **Extractor** model which leveraged the `spaCy` architecture with our existing domain-specific list of entities.

It is important to mention that although the standard `spaCy`'s entity recogniser does not explicitly contain the entity tags we used, it was possible to assume that its entity labels LOC, GPE, and FAC refer to place names and therefore can be converted to our PLNAME tag. Also, our DISTANCE tag is mostly equivalent to the `spaCy`'s QUANTITY tag. However, there were no such equivalent tags for GEONOUN and LOCADV.

The fully trained **Extractor** model is then used to parse an input text to create a standard spaCy Doc object which provides us additional linguistic annotations in addition to the required spatial entities. Our pipeline provides a visualisation module that lets us highlight and annotate these entities in the text - see examples in Figure 3. Also, as shown in Figure 2, the output from the model can be post-processed to produce other formats (e.g. `.txt`, `.xlsx` and `.json`) which are useful in other downstream analysis tasks or complex visualizations. Sections 6 and 7 describe two of the recent applications of the outputs from the extraction process presented in this section.

6 SPATIO-TEXTUAL REGIONS

Steiner *et al.* opined that *"raw spatial clustering ignores the structure or sequence of the narrative source, and thus omits critical information about the unfolding of a platial experience depicted by the author"* [37]. So they defined an STR as a *"clustered set of toponyms and a contiguous section of text describing those toponyms"*.

6.1 Toponyms in STRs

By applying spatial clustering to toponyms extracted from one of the Lake District texts – *The English Lakes* [Nelson and Sons (1857)] - they identified 8 major clusters (Grasmere, Cockermouth, Coniston, Pooley Bridge, Skiddaw, Windermere, Buttermere, and Keswick) as well as the "Outside region" that indicates places outside the Lake District. Figure 4 shows the number of places in each cluster by paragraphs which fairly indicates regional shifts as the narrative progresses.

6.2 Sense of Place in STRs

Unlike "location" (the place coordinates) and "locale" (the physical attributes and activities surrounding these coordinates) [1], "sense of place" is not easily defined despite its centrality to our identity as social beings. Feld & Basso defined it as the *'the experiential and expressive ways places are known, imagined, yearned for, held, remembered, voiced, lived, contested and struggled over.'* [11]. In Figure 5, Steiner *et al.* showed that analyzing sentiment scores (obtained from the extractor model) across a narrative sequence

Table 3: Performance scores: Table shows the F1 performance scores of the baseline regular expression `Regex` method, the `spaCy` off-the-shelf NLP model and our `Extractor (Ext)` model.

Entity Category	CLDW			HST		
	Regex	SpaCy	**Ext**	Regex	SpaCy	**Ext**
PLNAME	85.76	87.33	**90.20**	52.34	71.61	**72.86**
GEONOUN	82.44	–	**85.78**	78.51	–	**68.33**
LOCADV	61.55	–	**63.85**	55.94	–	**57.07**
DISTANCE	67.73	73.25	**79.29**	60.75	63.60	**68.08**

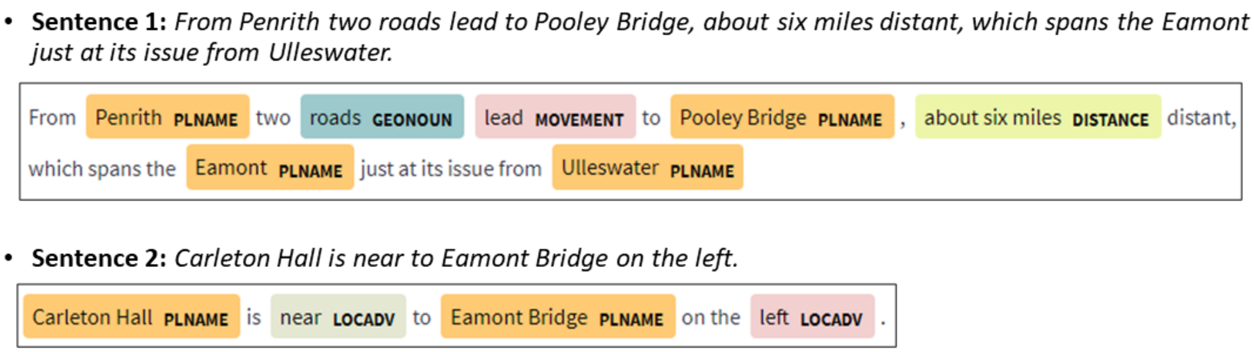

Figure 3: A visualization of the output from the extraction model with two example texts

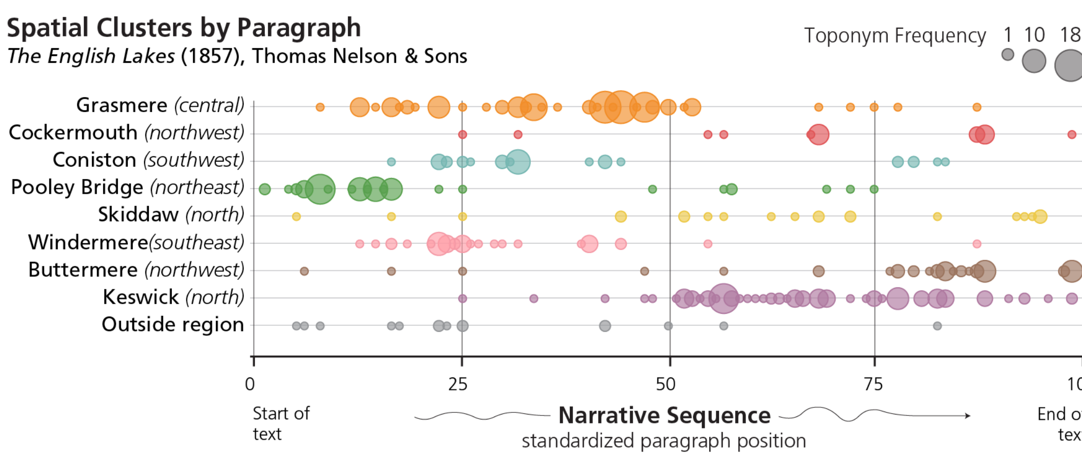

Figure 4: Number of toponyms in each spatial cluster by paragraph. The x-axis shows standardized paragraph position (0-100%) along the narrative sequence with the first paragraph on the left and last paragraph on the right. Source: Steiner *et al.* [37]

gives some insight into the sense of place of the STRs as portrayed in text.

7 QUALITATIVE SPATIAL REPRESENTATION

Understanding spatial narratives requires the ability to identify not just the spatial elements but also if and how they interact. We are also applying qualitative spatial and temporal reasoning to make more sense of the output from the extraction process. For example, Haris *et al.* [15] advocate systematically applying qualitative spatial representation (QSR) to identify and interpreting possible relationships using the outputs from the extraction process discussed in section 5.2. Table 4 shows a list of possible QSR-based relations and their interpretations from the example sentences in Figure 3.

In Table 4, the analysis of **Sentence 1** shows that `Penrith`, `Pooley Bridge`, `Eamont`, and `Ulleswater` are place names. However, analysis establishes that `Pooley Bridge` can be ambiguous (a 'town' or a 'bridge'). The instance of 'town' is defined by the 'bridge' class i.e. `bridge(pb)`. Also, `Eamont` and `Ulleswater` are tagged places but background knowledge indicates that they both

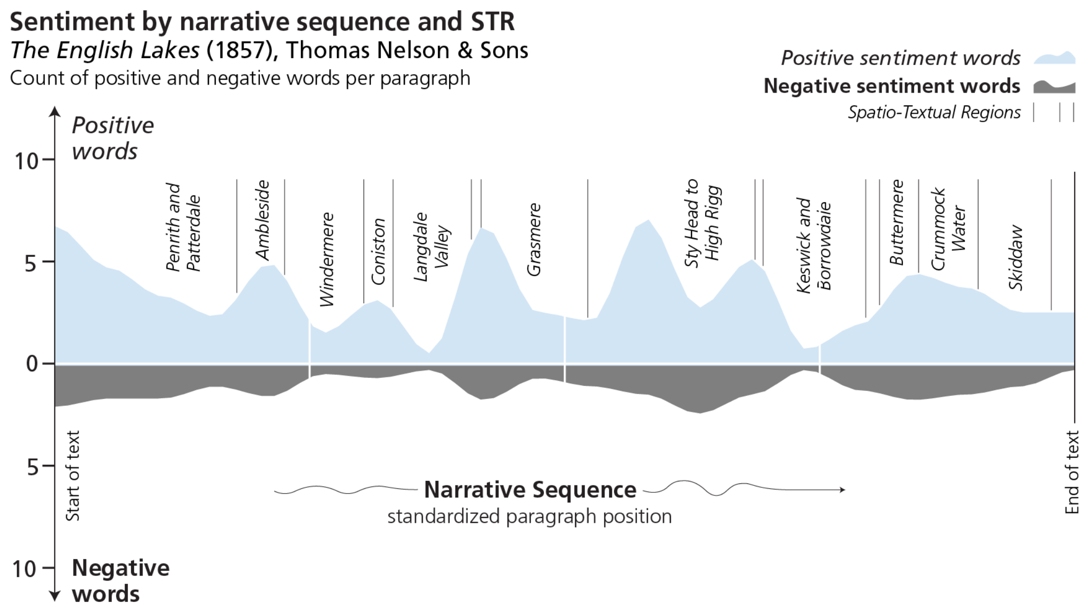

Figure 5: Sentiment by narrative sequence and Spatio-textual Regions in Nelson’s 1857 Text. Regions differ in their positive, negative, and total sentiment, with the peak emotions appearing in the central portion of the text. Source: Steiner *et al.* [37]

belong to the ‘river’ and ‘lake’ classes respectively. This example also exposes the need for a reasoning mechanism for approximate measurements (e.g. “aboutness”) for the ‘distance’ relationship between two places. Also, the distance between two places on a road can indicate the minimum length of the road. The end (or the beginning) of each road can be defined by the ‘start’ or ‘end’ predicate with each instance of the ‘road’ and ‘place’. Some rules could be added about spanning and bridges that let 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.

Sentence 2 is an example of an important and frequent representational challenge. The term ‘near’ is vague and requires specific rules in this kind of geographical context. Despite the attention received by vague spatial terms in the literature, there has been no definitive treatment. A key point to bear in mind is that ‘near’ is not transitive, i.e. from `near(a,b)` and `near(b,c)`, we cannot conclude `near(a,c)`. Directions provide different frames of reference. The ‘direction’ predicate can be defined with these 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.

Analyzing the spatio-temporal relationships within the corpus involves a multi-step process. Initially, we need to extract and convert user-level triples into an abstract representation, which consists of logical formulas describing the relationships. From there, we transition to a computational level, where we integrate these abstractions into geographical visualizations. In these visualizations, semantic representations of narratives are developed, which transform them into networks, with locations, temporal entities, and events serving as network nodes and spatio-temporal relationships forming the connections between these nodes. Subsequently, we perform network analysis to uncover intricate patterns that capture more nuanced and complex relationships.

8 ONTOLOGY DESIGN AND DEVELOPMENT

Ontologies are a powerful tool for the knowledge representation of a particular domain [30]. An ontology provides a collection of classes into which individual entities can be classified as well as the ability to represent spatial and other relationships between entities. In addition, some properties of classes and relationships can be stated as part of the ontology. Geo-ontologies usually base their foundation on the primary schema of `GeoSPARQL` ontology [7] which contains “spatial object” as the main class with two subclasses namely “feature” and “geometry”. A spatial object is defined as anything spatial (being or having a shape, a position, or an extent). A feature is a spatial object and has a geometry. For example, a river is a feature and hence can be linked to a point object (subclass geometry) which can represent a geocoordinate [7]. An important design feature of the `GeoSPARQL` and almost all the other related ontologies [7, 28, 31, 42] describing geographical aspects is that they define qualitative spatial relations as object properties e.g. {**Entity**: Place1, **ObjectProperty**: west, **Entity**: Place2}.

The ontology platform `Protégé` [26] is being used to construct an ontology in which, for example, the fact that `Pooley Bridge` belongs to the class `Bridge` can be recorded. Relationships between entities, such as the fact that `Pooley Bridge` spans the `River Eamont`, as also recorded and in `Protégé` these relationships are called ‘properties’. Going beyond particular facts about individual

Table 4: A list of possible QSR-based relations and interpretations from the example sentences in Figure 3.

QSR-based relations	Interpretations
From Penrith two roads lead to Pooley Bridge, about six miles distant, which spans the Eamont just at its issue from Ulleswater	
place(penrith)	Penrith is a place.
place(pb)	Pooley Bridge, pb, can be a place (as here) or a bridge. For convenience, here we use the same name, pb, for both the place and the bridge, but in practice it is necessary to have separate names in the formalism, as we discuss below in section 8.
distance(penrith, pb, about(6), miles)	distance in miles between places
road(road1); road(road2)	road1 and road2 belong to the class road.
end(road1, penrith); end(road1, pb)	road1 ends in Penrith and Pooley Bridge.
end(road2, penrith); end(road2, pb)	Also road2 as above.
bridge(pb); spans(pb, eamont)	Pooley Bridge is also a bridge and 'spans' Eamont.
river(eamont); source(eamont,ullswater)	Eamont is a river and issues from Ullswater.
lake(ullswater)	Ullswater is a lake.
Carleton Hall is near to Eamont Bridge on the left.	
near(carleton hall,pb)	Carleton Hall is near Pooley Bridge.
direction(carleton Hall, left, eb, d1)	Carleton Hall is left of Eamont Bridge. Here, the final argument, d1, is needed since in order to interpret "left of", one needs to know the direction the traveller is heading. That is not given in this sentence but needs to be inferred from the surrounding text.

entities, the description logic supported by Protégé allows general statements about classes and properties to be stored as axioms. This is possible with the examples mentioned above: *"bridges have two ends"* and *"every bridge spans something"*, and this allows some of the reasoning tasks to be carried in one of the reasoners associated with Protégé. The process of representing information in the corpus within an ontology has itself helped to clarify some of the distinctions it is proving necessary to make. For example, in the texts, the name "Pooley Bridge" is used both for a place and for a particular bridge. While this causes no problem for human readers, in the ontology we cannot use the same name to identify both entities. Thus there are separate identifiers for Pooley Bridge qua bridge and for Pooley Bridge qua place, both being related to the name "Pooley Bridge" via a "hasName" datatype property in Protégé.

Representing static entities, their classes, and relationships is thus possible in the ontology. However, the main interest is not this basic setting but more complex entities such as journeys within the setting. A particular journey described in a narrative will have a sequence of spatial entities with which it interacts via spatial relations and actions. For example: crossing the river Eamont, passing through Pooley Bridge, returning to Penrith. In the course of such a journey the experiences, such as seeing a certain lake through trees, or being caught in a rainstorm, can be identified within the ontology as classes of event having instances in the journey and ordered in sequence. We anticipate that representing journeys, or paths through landscapes described in narratives, in this way in the ontology will provide a means of exploring similarities and differences between journeys as sensed rather than as traced on a map.

Based on the outcomes of the spatial entity annotation and extraction process, we define a general taxonomy for the narrative corpora which includes a listing of spatial and temporal entities and relationships along with their associated features shown in Figure 6. The features could also be associated with actions in space and time. The elaboration of a feature class and its relationship with subclasses will let us relate sense of place with particular entities. The ontology is being designed based on this basic taxonomy of the narrative domain; the ontology schema incorporates standard place vocabularies and also follows the conventions used in the standard geo-ontologies earlier referred to. To delineate the variety of place names appearing in the CLDW corpus, the geospatial categories proposed by Rayson et al. [33] have been utilized since their annotation scheme provides a comprehensive list of geospatial categories most suited for the CLDW corpus, which include: -Country. - Region. - Settlement. - Height. - Lake. - Waterway. - Waterfall. - Vale. - Woodland. - Island. - Pass. - Specific. - Feature. - House. - Farm. - Inn. -Street. - Battlesite. - Poetic. Although in [33], definitionally specific categories have been merged into a general category, we intend to introduce them as subclasses using standard place ontologies. Ballatore and Adams [3] curated an exhaustive vocabulary of nouns used to describe places, both natural and man-made place types, using several sources including GeoNames Ontology, the DBpedia Ontology, and Wordnet. Similarly, standard vocabularies for spatial and temporal relations, including Ordinance Survey's spatial relations[7], GeoSPARQL, OWL-Time[8] and other related ontologies will be used.

[7] https://www.ordnancesurvey.co.uk/linked-data/ontology/spatialrelations.owl
[8] https://www.w3.org/TR/owl-time/

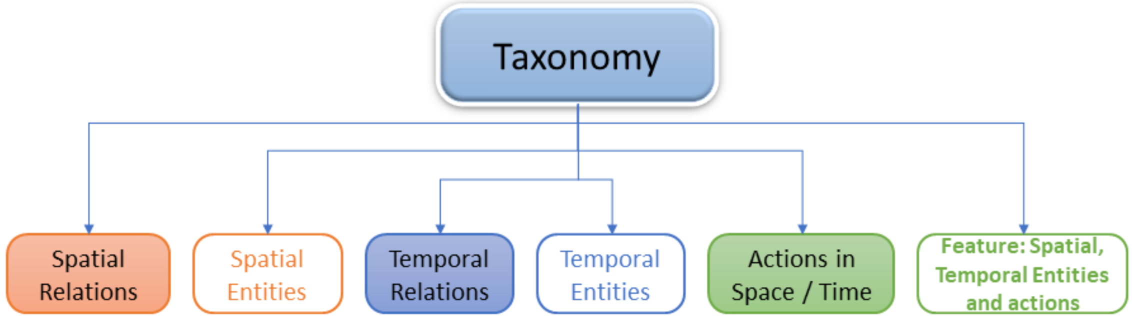

Figure 6: Taxonomy of the narrative domain: A listing of spatial and temporal entities and relationships along with their associated features.

9 CONCLUSION AND FUTURE WORK

In this work, we set out to expand the discourse on the methods for automatically extracting spatial elements that enable us to analyze and understand imprecise references to space and time in narratives. By leveraging two distinct corpora –the corpus of Lake District Writing and the Holocaust Survivor's Testimonies –we have presented our preliminary work on an extensible framework that not only sheds light on the intricate interplay between space and time in narratives but also offers a versatile tool (Spatial Entity Extraction Demo) for researchers.

It is our view that this framework will support many downstream tasks in the processing of spatial narratives. A key contribution of this paper is the design of the workflow for spatial element definition, corpus annotation, and the design and development of the spatial element `Extractor` model as well as the demonstration of the use cases for the outputs from the extraction model, including analysis, visualization of spatio-textual regions, qualitative spatial representation, and ontology design and development.

It is evident that this framework, drawing from the unique contexts of the Lake District and Holocaust survivors, will serve as an invaluable compass for exploring the rich tapestry of spatial narratives in written text. By bridging the gap between traditional narrative theory and computational systems, our work opens up possibilities for a better understanding of textual narratives in spatio-temporal dimensions. It promises a more comprehensive understanding of how we perceive, interact with, and ultimately experience the world around us beyond geo-coding.

Our future work will focus on deeply annotating our two experimental corpora –CLDW and HST. We hope to achieve this through a thorough review of our entity and relationship annotation schema based on the qualitative spatial representation and ontology designs that are currently ongoing. We have presented a top-level view of the ontology but there are many spatial relations and individual data instances that are yet to be developed and included. Similarly, temporal ontological concepts will be incorporated and the relationship between relevant classes will be defined using standard time ontologies [5, 29]. This will hopefully include additional relationship information as well as improve the performance of our entity extraction model and the demo tool.

10 ACKNOWLEDGMENTS

The authors 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). Cohn is also partly supported by the Alan Turing Institute.

REFERENCES

[1] J.A. Agnew. 1987. *Place and Politics: The Geographical Mediation of State and Society.* Allen & Unwin, Boston. https://www.worldcat.org/oclc/15016665

[2] John Agnew and James Duncan. 1989. *The Power of Place: Bringing Together Geographical and Sociological Imaginations.* Boston: Unwin Hyman.

[3] Andrea Ballatore and Benjamin Adams. 2015. Extracting place emotions from travel blogs. In *Proceedings of AGILE*, Vol. 2015. AGILE Conference, Lisbon, Portugal, 1–5.

[4] Adrien Barbaresi. 2017. Towards a toolbox to map historical text collections. In *Proceedings of the 11th Workshop on Geographic Information Retrieval.* ACM, New York, USA, 1–2.

[5] Sotiris Batsakis, Euripides GM Petrakis, Ilias Tachmazidis, and Grigoris Antoniou. 2017. Temporal representation and reasoning in OWL 2. *Semantic Web* 8, 6 (2017), 981–1000.

[6] Lucie Cadorel, Alicia Blanchi, and Andrea GB Tettamanzi. 2021. Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text. In *Proceedings of the 11th on Knowledge Capture Conference.* ACM Digital Library, New York, NY, United States, 41–48.

[7] Nicholas J. Car and Timo Homburg. 2022. GeoSPARQL 1.1: Motivations, Details and Applications of the Decadal Update to the Most Important Geospatial LOD Standard. *ISPRS International Journal of Geo-Information* 11, 2 (2022), 1–30. https://doi.org/10.3390/ijgi11020117

[8] Tim Cresswell. 2014. *Place: an introduction.* John Wiley & Sons, Toronto, Canada.

[9] Curdin Derungs and Ross S Purves. 2014. From text to landscape: locating, identifying and mapping the use of landscape features in a Swiss Alpine corpus. *International Journal of Geographical Information Science* 28, 6 (2014), 1272–1293.

[10] Ignatius Ezeani, Paul Rayson, and Ian N Gregory. 2023. Extracting Imprecise Geographical and Temporal References from Journey Narratives.. In *Text2Story@ ECIR.* CEUR Workshop Proceedings, Republic of Ireland, 113–118.

[11] S. Feld and K.H. Basso. 1996. *Senses of Place.* School of American Research Press, Santa Fe, New Mexico. 11 pages. https://www.worldcat.org/oclc/35043056

[12] Michael F Goodchild. 2010. Formalizing place in geographic information systems. In *Communities, neighborhoods, and health: Expanding the boundaries of place.* Springer, Switzerland, 21–33.

[13] Jan Goyvaerts. 2016. Regular Expressions Tutorial. https://web.archive.org/web/20161101212501/http://www.regular-expressions.info/tutorial.html. Accessed: 2022-10-10.

[14] Ian Gregory, Christopher Donaldson, Andrew Hardie, and Paul Rayson. 2018. Modeling space in historical texts. In *The Shape of Data in the Digital Humanities.* Routledge, University of Birmingham, 133–149.

[15] Erum Haris, Anthony G. Cohn, and John G. Stell. 2023. Understanding the Spatial Complexity in Landscape Narratives Through Qualitative Representation of Space. In *12th International Conference on Geographic Information Science (GIScience 2023) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 277),*

Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 37:1–37:6. https://drops.dagstuhl.de/opus/volltexte/2023/18932

[16] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017).

[17] Yingjie Hu and Benjamin Adams. 2020. Harvesting big geospatial data from natural language texts. In *Handbook of Big Geospatial Data*. Springer, Cham, 487–507.

[18] Yingjie Hu, Xinyue Ye, and Shih-Lung Shaw. 2017. Extracting and analyzing semantic relatedness between cities using news articles. *International Journal of Geographical Information Science* 31, 12 (2017), 2427–2451.

[19] Morteza Karimzadeh, Wenyi Huang, Siddhartha Banerjee, Jan Oliver Wallgrün, Frank Hardisty, Scott Pezanowski, Prasenjit Mitra, and Alan M MacEachren. 2013. GeoTxt: a web API to leverage place references in text. In *Proceedings of the 7th workshop on geographic information retrieval*. ACM, NY, United States, Orlando, Florida, 72–73.

[20] Parisa Kordjamshidi, Martijn Van Otterlo, Marie-Francine Moens, Robert Ross, Joana Hois, and John Kelleher. 2010. From language towards formal spatial calculi. In *Proc. of 1st Workshop COSLI’10*, Vol. 620. CEUR-WS. org; http://ceur-ws.org/Vol-620, Oregon, USA, 17–24.

[21] Bing Liu et al. 2010. Sentiment analysis and subjectivity. *Handbook of natural language processing* 2, 2010 (2010), 627–666.

[22] BA Meiring. 1993. The syntax and semantics of geographical names. *Training course in Toponymy for Southern Africa* 1, 1 (1993), 1–19.

[23] Helena Merschdorf and Thomas Blaschke. 2018. Revisiting the Role of Place in Geographic Information Science. *ISPRS International Journal of Geo-Information* 7, 9 (2018), 364.

[24] Molly Miranker and Alberto Giordano. 2020. Text mining and semantic triples: Spatial analyses of text in applied humanitarian forensic research. *Digital Geography and Society* 1 (2020), 100005.

[25] Patricia Murrieta-Flores, Mariana Favila-Vázquez, and Aban Flores-Morán. 2019. 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 (2019), 53–68.

[26] Mark A. Musen. 2015. The Protégé Project: A Look Back and a Look Forward. *AI Matters* 1, 4 (jun 2015), 4–12. https://doi.org/10.1145/2757001.2757003

[27] Paolo Nesi, Gianni Pantaleo, and Marco Tenti. 2016. Geographical localization of web domains and organization addresses recognition by employing natural language processing, Pattern Matching and clustering. *Engineering Applications of Artificial Intelligence* 51 (2016), 202–211.

[28] Vatsala Nundloll, Robert Smail, Carly Stevens, and Gordon Blair. 2022. Automating the extraction of information from a historical text and building a linked data model for the domain of ecology and conservation science. *Heliyon* 8 (10 2022), e10710. https://doi.org/10.1016/j.heliyon.2022.e10710

[29] Feng Pan and Jerry R Hobbs. 2006. Time Ontology in OWL. *W3C working draft, W3C* 1, 1 (2006), 1.

[30] Cecilia Reyes Peña and Mireya Tovar. 2019. Ontology: Components and Evaluation, a Review. *Res. Comput. Sci.* 148 (2019), 257–265. https://api.semanticscholar.org/CorpusID:214487268

[31] Babak Ranjgar, Abolghasem Sadeghi-Niaraki, Maryam Shakeri, and Soo-Mi Choi. 2022. An ontological data model for points of interest (POI) in a cultural heritage site. *Heritage Science* 10 (03 2022), 13. https://doi.org/10.1186/s40494-021-00635-9

[32] Christopher M Raymond, Marketta Kyttä, and Richard Stedman. 2017. Sense of place, fast and slow: The potential contributions of affordance theory to sense of place. *Frontiers in psychology* 8 (2017), 1674.

[33] Paul Rayson, Alexander Reinhold, James Butler, Chris Donaldson, Ian Gregory, and Joanna Taylor. 2017. A deeply annotated testbed for geographical text analysis: The Corpus of Lake District Writing. In *GeoHumanities’17 Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*. Association for Computing Machinery (ACM), Redondo Beach CA USA, 9–15. https://doi.org/10.1145/3149858.3149865

[34] Noah Shenker. 2015. *Reframing Holocaust Testimony*. Indiana University Press, Indiana University. http://www.jstor.org/stable/j.ctt16gz8z7

[35] Robert Smail, Ian Gregory, and Joanna Taylor. 2019. Qualitative geographies in digital texts: Representing historical spatial identities in the Lake District. *Int’ernational Journal of Humanities and Arts Computing* 13, 1-2 (2019), 28–38.

[36] Barry Smith and David M Mark. 2001. Geographical categories: an ontological investigation. *International journal of geographical information science* 15, 7 (2001), 591–612.

[37] Erik Steiner, Zephyr Frank, Ian Gregory, David Bodenhamer, and Ignatius Ezeani. 2023. Spatio-textual Regions: Extracting Sense of Place from Spatial Narratives. In *Proceedings of the Fourth International Symposium on Platial Information Science (PLATIAL’23)*, R Westerholt and F. B. Mocnik (Eds.). PLATIAL’X, Dortmund, Germany;, 1–8.

[38] John G Stell. 2019. Qualitative spatial representation for the humanities. *International Journal of Humanities and Arts Computing* 13, 1-2 (2019), 2–27.

[39] Lauren M Stuart, Julia M Taylor, and Victor Raskin. 2013. The importance of nouns in text processing. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. California Digital Library, University of California, USA, 1390–1395.

[40] M. A. Syed, E. Arsevska, M. Roche, and M. Teisseire. 2022. GeoXTag: Relative Spatial Information Extraction and Tagging of Unstructured Text. *AGILE: GIScience Series* 3 (2022), 16. https://doi.org/10.5194/agile-giss-3-16-2022

[41] Joanna E Taylor and Ian N Gregory. 2022. *Deep Mapping the Literary Lake District: A Geographical Text Analysis*. Rutgers University Press, New Brunswick.

[42] Barney Warf and Daniel Sui. 2010. From GIS to neogeography: ontological implications and theories of truth. *Annals of GIS* 16, 4 (2010), 197–209.

Received 15 September 2023; revised 04 October 2023; accepted 06 October 2023