



Extracting positive descriptions and exploring landscape value using text analysis in the Cairngorms National Park

Daniela Mariño, Ludovic Moncla, Ross Purves

► To cite this version:

Daniela Mariño, Ludovic Moncla, Ross Purves. Extracting positive descriptions and exploring landscape value using text analysis in the Cairngorms National Park. 7th ACM SIGSPATIAL International Workshop on Geospatial Humanities, Nov 2023, Hamburg, Germany. pp.44-47, <10.1145/3615887.3627758>. <hal-04274439>

HAL Id: hal-04274439

<https://hal.science/hal-04274439v1>

Submitted on 6 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Extracting positive descriptions and exploring landscape value using text analysis in the Cairngorms National Park

Daniela Mariño
Department of Geography
University of Zurich
Switzerland
daniela.Mariño@uzh.ch

Ludovic Moncla
LIRIS, UMR 5205
Univ. Lyon, INSA Lyon, CNRS
France
ludovic.moncla@insa-lyon.fr

Ross S. Purves
Department of Geography
University of Zurich
Switzerland
ross.purves@geo.uzh.ch

ABSTRACT

Being in natural places is known to influence well-being. Therefore, increasing attention is being paid to the diverse ways in which our emotions can be influenced by places, and this paper investigates this notion in a protected area – the Cairngorms National Park in Scotland. We assume that positive sentiment correlates with pleasant (hedonic) experiences, and extract positive emotions and objects associated with them, from textual descriptions written about places in the park.

To do so, we first apply sentiment analysis and select descriptions with positive sentiment. Second, we filter the initial dataset to reduce the impact of prolific users and ambiguously positive words. From an initial set of 33790 descriptions we are left with 3031 positive descriptions. Third, we apply transducers to find semantic sequences for patterns detection. Finally, these sequences are annotated and assigned to one of five classes – activities, aesthetics, features, places and times. Features are by far the most common class, making up more than 50% of all unique nouns associated with positive emotions.

CCS CONCEPTS

• **Information systems** → *Information systems applications* → *Spatial-temporal systems* → Geographic information systems

KEYWORDS

Landscape perception, values of nature, natural language processing, sentiment analysis, transducers.

ACM Reference format:

Daniela Mariño, Ludovic Moncla & Ross S. Purves. 2023. Extracting positive descriptions and exploring landscape value using text analysis in the Cairngorms National Park. In *GeoHumanities '23*, November 13, 2023, Hamburg, Germany © 2023 Copyright is held by the owner/author(s). ACM ISBN 979-8-4007-0349-2/23/11. <https://doi.org/10.1145/3615887.3627758>



This work is licensed under a Creative Commons Attribution International 4.0 License.

GeoHumanities '23, November 13, 2023, Hamburg, Germany
© 2023 Copyright is held by the owner/author(s).
ACM ISBN 979-8-4007-0349-2/23/11.
<https://doi.org/10.1145/3615887.3627758>

1 Introduction

For many of us, going to green, quiet places provides us with a respite from everyday life. There is a wealth of literature showing that natural landscapes at very different scales can provide concrete benefits to our health, and such positively contribute to our lives [1]. But many natural landscapes are under threat [2], and understanding what and how such landscapes are valued is the subject of this paper.

Our starting points are twofold. First, the recent IPBES report on the diverse values and valuation of nature emphasized the need to understand different ways in which landscapes are valued. Second, it set out a values assessment typology emphasizing, amongst others, the need for descriptive indicators reflecting nature's contribution to people [3].

Among the ways in which values can be assessed, the IPBES report highlights two groups of methods. In statement-based valuation, individuals are directly asked about their experiences of a location, typically through methods such as Public Participation GIS and surveys [4]. By contrast, behavior-based valuation observes what people do in nature, linking their activities to value. Here, the links between value and behavior are implicit – for example, we assume that people take more photographs of landscapes that they prefer [5]. However, both sets of methods have limitations. Statement-based valuation is time consuming and difficult to scale up across many sites or over long time periods [6], while behavior-based valuation often lacks specific information about what or why a particular location in a landscape is being valued [7].

To address these issues, we propose using natural language processing methods to extract and classify positive experiences in landscapes. Similar to previous work [8], we wish to identify individual experiences. Thus, our approach has parallels with statement-based valuation, since we aim to find specific descriptions of positive experiences in the landscape. These positive experiences can be manifested in different ways, but are typically grouped in the literature as hedonic – that is to say pleasurable ways in which we can experience landscapes [9]. These might include therefore tranquil calm places, activities we enjoy that are exciting, or arousing and aesthetically appealing locations that we choose to share with our colleagues [10]. Key though is that

those having these experiences are sufficiently moved that they report them. This reporting, through writing accounts of experiences at given locations, brings us to the second parallel with behavior-based valuation, but by using natural language, we aim to move beyond simply locating these experiences, and better understand the ways in which a particular landscape experience is considered positive.

In this paper, we introduce a pilot study, investigating the use of text processing methods to extract positive experiences in a single protected area. Our approach is iterative and combines off-the-shelf natural language processing tools with lexicons and rulesets to extract positively connotated descriptions and the objects referred to therein.

2 Study site and data

In our pilot study, we chose a single protected area, the Cairngorms National Park, as our site of interest. The park is found in the north-east of Scotland. It was designated national park in 2003 and extended in 2013, becoming the biggest in the UK with 4528 km² (6% of Scotland's total area) [11]. Large areas of the park are considered wild, and within the national park many areas have been designated as protected natural heritage. However, the park area also includes settlements, and more than 18000 people live within its boundaries. The landscapes found in the park thus include mountainous regions, with large areas over 900m, broad valleys (known locally as *straths*), in which wetlands and settlements are found, and regions of moor and forest. A long history of settlement has resulted in a rich cultural heritage, complementing the natural beauty of the park [11]–[13].

We used data from the Geograph Britain and Ireland project [14] to search for positive experiences in the park. Geograph is a project where individuals can contribute pictures and descriptions of Great Britain and Ireland. At the time of writing this article, more than 13000 Geograph contributors submitted over 7 million images and associated texts. Within the Cairngorms National Park, we identified 33790 descriptions, contributed by 3623 users. These texts are the starting point for our analytical workflow.

3 Analytical workflow

Our workflow's aim was to understand the sorts of positive experiences reported in the national park. To operationalize the workflow, we carried out the following steps:

1. Identified candidate positive descriptions.
2. Reduced the effects of participation inequality in the corpus.
3. Associated and disambiguated words expressing emotions in candidate positive descriptions.
4. Extracted objects (aspects) associated with positive emotions in candidate descriptions.
5. Bottom-up classification of objects associated with positive emotions.

We used an off-the-shelf sentiment analysis tool to filter descriptions which may be positive from our texts. Using NLTK, and the VADER lexicon, the descriptions were given aggregate

scores between -1 and +1 (classifying descriptions as negative, neutral or positive) [15]. Within the national park, we found that the bulk of descriptions (66%) were classified as neutral, 21% were positive and the remaining 13% were negative. The 21% positive corresponded to 7199 descriptions, written by 424 users.

However, like all user generated content, these data are biased towards prolific contributors [16]. In our case, the most active user wrote 1572 descriptions or 22% of all positive descriptions in the park. Since we wished to explore diverse values associated with the landscape, we needed to reduce the impact of such very prolific users on our dataset.

We used two parameters to understand the effects of participation inequality: Ripley's K (a measure of spatial clustering) and the gradient *g* of a log-log plot of term rank against term frequency in the corpus. This gradient indicates the extent to which the corpus follows Zipf's law, and thus the semantic content of our corpus. To test the influence of prolific contributors on the spatial and semantic properties of our corpus, we progressively removed descriptions written by prolific contributors and recalculated Ripley's K and *g*. We found similar values of both Ripley's K and *g* to the corpus as a whole when we only considered contributors who provided 200 or less descriptions. Doing so, we reduced the corpus size to 4441 descriptions and 419 users – in other words the 5 most prolific contributors were excluded from further analysis.

The remaining 4441 descriptions had an average length of 25 tokens, and contained 9458 unique tokens. In this set, we then searched for matches within a list of words associated with emotions – the NRC Word-Emotion Association Lexicon [17]. For 3069 descriptions (69%), we found at least one match. We used WordNet synsets to deal with word sense ambiguity, removing 38 more descriptions.

For the remaining 3031 descriptions, we found 4453 tokens that matched the Word-Emotion Association Lexicon, of which 54% were identified through part of speech tagging as adjectives, 13% nouns, 12% verbs and 12% adverbs.

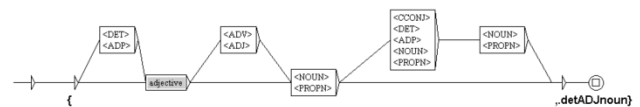


Figure 1: Finite transducer to detect object associated with an emotion root as an adjective

Then we iteratively developed a set of four finite-state sentence transducers to extract annotations. Figure 1 illustrates the finite transducer structure, and Table 1 lists the four implemented cases. For each extracted annotation, we next identified the objects with which these positive emotions were associated [19, 20] – a task known in sentiment analysis as aspect detection [18].

To carry out this aspect detection, we made a simplifying assumption that these objects are dependent nouns in the extracted annotations. This approach has been shown in previous work to be robust, and well suited to applications in specialized corpora with clearly defined tasks [21].

Case A: Adjective* + Noun is a positive quality of an entity in the place. This is a case rooted in the list of positive adjectives, followed by a noun:

Example: ‘attractive birds’, ‘lovely bridge’, ‘attractive small glen’

Case B: Verb + Adjective* + Noun is a positive quality of an activity related to the place. This variant is rooted in the list of adjectives, but requires a verb before the adjective:

Example: 'is a lovely sandy beach', 'provides excellent salmon'

Case C: Verb* + Noun is a positive action experienced in a place. The root is the list of positive verbs with a related noun:

Example: ‘enjoy the peace and tranquility’, ‘enjoying a dip’

Case D: Article + Noun* + Noun is an entity evoking a positive place. The root is based on the list of positive nouns, which is combined with a noun:

Example: ‘calm superb dark pool, ‘the beauty of a walk’

Table 2 shows the total number of sequences (524) identified by the four finite transducers. The majority (449 or 86%) were associated with cases A and B, both of which use adjectives as roots. Indeed, all cases where adverbs were used, the annotations were already subsumed into these examples, and thus we did not implement separate transducers for adverbs. Having extracted sequences, such as the examples shown in Table 1, we annotated the nouns related to each positive experience. Using an iterative coding process, we identified five broad classes of nouns: activities (e.g., *picnic*, *walking*, *swimming*); aesthetics (e.g., *clarity*, *view*, *visibility*); features (e.g., *bridge*, *birch*, *loch*, *path*), places (e.g., *area*, *location*, *spot*), and times (e.g., *afternoon*, *day*, *September*).

Table 2: Cases of sequences for positive detection

Annotations		Associations				
		Activities	Aesthetics	Features	Places	Times
Case A	331	9	38	157	61	66
Case B	118	8	33	46	26	5
Case C	33	6	11	11	3	2
Case D	42	2	6	23	8	3
Total	524	25	88	237	98	76
Unique terms		19	22	146	38	25

Each of these classes captures different ways in which locations are valued. For example, activities are often used in reporting and understanding recreational uses of landscape [22]. Descriptions of landscapes in terms of their aesthetics have a long tradition in the British Isles, and the importance of landscape beauty as an inherent characteristic is enshrined in policy documents [23]. Previous research on landscapes and text has found that anthropogenic and biophysical objects are commonly described in texts related to landscape [24, 25], and here we see that these are reported in more positive descriptions than any other category. This result is particularly important, since these objects capture the properties of the environment itself which are valued. The places category captures generic locations, whose properties are not specified, but appear nonetheless to be holistically important. Finally, times are a

common way of describing positive experiences, and these may relate to ephemeral conditions – *a beautiful morning or a pleasant winter's day*.

Figure 2 shows the individual sentiment aspects extracted for each class as word clouds. Note the total number of unique aspects (250 terms) was unevenly distributed across the classes. For example, for features, we find more than 50% of all unique aspects (146) associated with positive experiences. These lists reveal great diversity in the ways in which the Cairngorms National Park is valued – and thus bring us one step nearer to our goal of combining statement and behavior-based valuation.

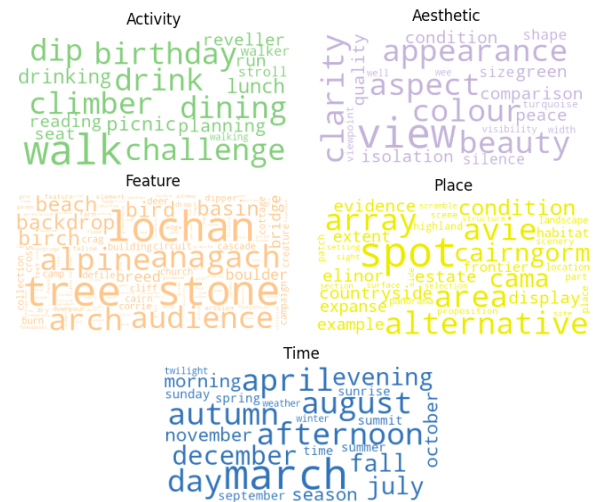


Figure 2: Nouns frequency by category

Our final result, Figure 3, maps the locations of the annotated positive experiences in the park as a whole. Here, we note that most of these experiences happen relatively near to infrastructure, typically in the form of roads running along the valley to the north and south of the region, and that the concentration of positive experiences appears to be less dense in the wilder central areas. However, care is needed here, since this may simply reflect the overall distribution of contributions to *Geograph*, and further work would be needed to test whether positive experiences were influenced directly by accessibility.

The Cairngorms National Park Authority highlight scenery, wildlife and cultural heritage as important attractions. All of these occur frequently in our feature class, and we find a predominance of natural objects. Examples include water bodies (e.g., *loch* and *beach*), living objects (e.g., *tree*, *bird*, *birch* and *deer*) and other elements of the landscape (e.g., *basin*, *stone*, and *boulder*). Less natural elements of the landscape, reflecting its cultural heritage, are present but less common (e.g., *bridge*, *building*). In the aesthetic class, very important in British notions of landscape, there is less diversity in usage of terms, with a few prominent generic adjectives and nouns (e.g., *beauty*, *clarity*, *colour*, *silence*), most of them relating to visual and aural experiences. Activities include not only outdoor activities (e.g., *climbing* and *walking*), but also celebrations and more sedentary activities (e.g., *birthday*, *reading*, *drinking*, *dining*). Finally, times encapsulate not only seasons (e.g., spring and autumn), but also times of day (e.g. *morning* and *afternoon*) and specific months of the year (e.g., *march*, *april* and *august*).

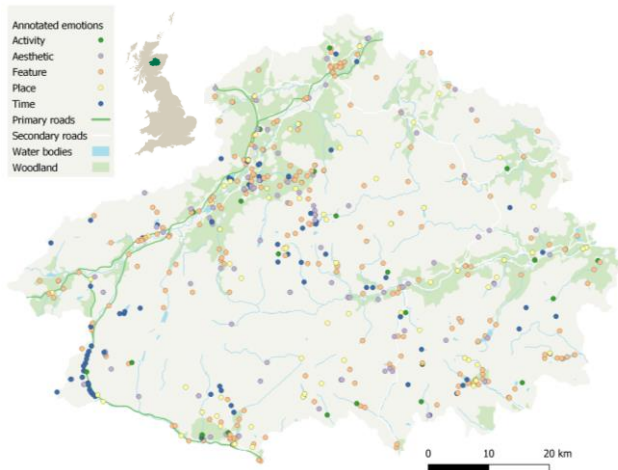


Figure 3: Locations of annotated positive experiences in the Cairngorms National Park (Contains OS data © Crown copyright and database right2022)

5 Conclusions

Our pilot study successfully extracted positive descriptions in the Cairngorms National Park, and the aspects associated with these appear plausible. For example, lochans are small lakes, and these locations are often considered scenic by visitors – indeed the relationship between water and scenicness has been observed in many previous studies of landscape values [10].

Nonetheless, our approach has limitations – perhaps most importantly, building the pipeline was time consuming, and involved many decisions, all of which could change the results. In future work, we will explore different approaches to extracting both positive experiences and their aspects, for example by investigating the use of transfer learning based on the currently annotated set of data. Equally, we note that a lack of a value or aspect, in our results, does not mean that a location does not play an important role in the make-up of a landscape – our approach is inherently biased towards accessible locations. However, these results are encouraging, and we intend to expand our study to compare all national parks in the Great Britain, exploring variations in ways in which these important landscapes are, and potentially, have been valued.

ACKNOWLEDGMENTS

We gratefully acknowledge all the contributors to Geograph British Isles (Creative Commons Attribution-ShareAlike 2.5 License) whose descriptions we analyzed in this paper. We also gratefully acknowledge funding from the Swiss National Science Foundation project EVA-VGI 2 [186389].

REFERENCES

- [1] N. Fagerholm, S. Eilola, and V. Arki, “Outdoor recreation and nature’s contribution to well-being in a pandemic situation - Case Turku, Finland,” *Urban for Urban Green*, vol. 64, Sep. 2021, DOI: 10.1016/J.UFUG.2021.127257.
- [2] R. Scolozzi, U. Schirpke, C. Detassis, S. Abdullah, and A. Gretter, “Mapping Alpine Landscape Values and Related Threats as Perceived by Tourists,” <http://dx.doi.org/10.1080/01426397.2014.902921>, vol. 40, no. 4, pp. 451–465, May 2014, DOI: 10.1080/01426397.2014.902921.
- [3] IPBES, “Summary for policymakers of the methodological assessment of the diverse values and valuation of nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES),” Jul. 2022, DOI: 10.5281/ZENODO.7410287.
- [4] N. Fagerholm et al., “Perceived contributions of multifunctional landscapes to human well-being: Evidence from 13 European sites,” *People and Nature*, vol. 2, no. 1, pp. 217–234, Mar. 2020, DOI: 10.1002/pan3.10067.
- [5] K. F. Tieskens, B. T. Van Zanten, C. J. E. Schulp, and P. H. Verburg, “Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape,” *Landscape Urban Planning*, 177, pp. 128–137, Sep. 2018, DOI: 10.1016/J.LANDURBPLAN.2018.05.002.
- [6] F. M. Wartmann, E. Acheson, and R. S. Purves, “Describing and comparing landscapes using tags, texts, and free lists: an interdisciplinary approach,” *International Journal of Geographical Information Science*, vol. 32, no. 8, pp. 1572–1592, Aug. 2018, DOI: 10.1080/13658816.2018.1445257.
- [7] D. R. Montello, M. F. Goodchild, J. Gottsegen, and P. Fohl, “Where’s downtown? Behavioral methods for determining referents of vague spatial queries,” *Spatial Cognition & Computation*, vol. 3, no. 2–3, pp. 185–204, 2003, DOI: 10.1080/13875868.2003.9683761.
- [8] O. Koblet and R. S. Purves, “From online texts to Landscape Character Assessment: Collecting and analyzing first-person landscape perception computationally,” *Landscape Urban Planning*, vol. 197, May 2020, DOI: 10.1016/J.LANDURBPLAN.2020.103757.
- [9] C. D. Ryff, J. M. Boylan, and J. A. Kirsch, “Eudaimonic and Hedonic Well-Being,” in *Measuring Well-Being*, Oxford University Press, 2021, pp. 92–135, DOI: 10.1093/oso/9780197512531.003.0005.
- [10] F. M. Wartmann, O. Koblet, and R. S. Purves, “Assessing experienced tranquility through natural language processing and landscape ecology measures,” *Landscape Ecology*, vol. 36, no. 8, pp. 2347–2365, Aug. 2021, DOI: 10.1007/S10980-020-01181-8.
- [11] NatureScot, “Landscape Character Assessment Review: Cairngorms landscape evolution and influences,” 2019. Accessed: Jul. 05, 2023. [Online]. Available: www.historicenvironment.scot/
- [12] Cairngorms National Park, “Cairngorms National Park Partnership Plan 2022–27,” 2022. Accessed: Jul. 05, 2023. [Online]. Available: <https://cairngorms.co.uk/working-together/partnershipplan/>
- [13] Protected Planet, “Explore the World’s Protected Areas,” 2023. Accessed: Jul. 05, 2023. [Online]. Available: <https://www.protectedplanet.net/country/GBR>
- [14] Geograph® Britain and Ireland Project, “Geograph Project Limited,” Accessed: Jul. 05, 2023. [Online]. Available: <https://www.geograph.org.uk/>
- [15] S. Elbagir and J. Yang, “Twitter sentiment analysis using natural language toolkit and Vader sentiment,” *Lecture Notes in Engineering and Computer Science*, vol. 2239, pp. 12–16, 2019.
- [16] D. S. Shah, H. A. Schwartz, and D. Hovy, “Predictive Biases in Natural Language Processing Models: A Conceptual Framework and Overview,” in the *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 5248–5264, DOI: 10.18653/v1/2020.acl-main.468.
- [17] S. M. Mohammad and P. D. Turney, “Crowdsourcing a Word-Emotion Association Lexicon,” Aug. 2013, [Online]. Available: <http://arxiv.org/abs/1308.6297>
- [18] B. Liu, “Sentiment Analysis”. Cambridge University Press, 2015, DOI: 10.1017/CBO9781139084789.
- [19] E. Egorova, L. Moncla, M. Gaio, C. Claramunt & Ross, and S. S. Purves, “Fictive motion extraction and classification,” *International Journal of Geographical Information Science*, vol. 32, no. 11, pp. 2247–2271, 2018, DOI: 10.1080/13658816.2018.1498503.
- [20] M. Gaio and L. Moncla, “Extended Named Entity Recognition Using Finite-State Transducers: An Application to Place Names,” 2017, Accessed: Jul. 14, 2023. [Online]. Available: <https://hal.science/hal-01492994>
- [21] L. Moncla, M. Gaio, J. Noguera-Iso, and S. Mustière, “Reconstruction of itineraries from annotated text with an informed spanning tree algorithm,” <http://dx.doi.org/10.1080/13658816.2015.1108422>, vol. 30, no. 6, pp. 1137–1160, Jun. 2015, DOI: 10.1080/13658816.2015.1108422.
- [22] F. M. Wartmann, M. F. Baer, K. T. Hegetschweiler, C. Fischer, M. Hunziker, and R. S. Purves, “Assessing the potential of social media for estimating recreational use of urban and peri-urban forests,” *Urban for Urban Green*, vol. 64, Sep. 2021, DOI: 10.1016/J.UFUG.2021.127261.
- [23] J. Glover, “Landscapes Review,” 2019. Accessed: Sep. 07, 2023. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833726/landscapes-review-final-report.pdf
- [24] C. Derungs and R. S. Purves, “From text to landscape: locating, identifying and mapping the use of landscape features in a Swiss Alpine corpus,” *International Journal of Geographical Information Science*, vol. 28, no. 6, pp. 1272–1293, Jun. 2014, DOI: 10.1080/13658816.2013.772184.
- [25] F. M. Wartmann and R. S. Purves, “Investigating sense of place as a cultural ecosystem service in different landscapes through the lens of language,” *Landscape Urban Planning*, vol. 175, pp. 169–183, Jul. 2018, DOI: 10.1016/j.landurbplan.2018.03.021.