

CHAPTER 12

Surveying the Terrain and Looking Forward

Ross S. Purves

Department of Geography; URPP Language and Space, University of Zurich,
Switzerland

Olga Koblet

Department of Geography, University of Zurich, Switzerland

Benjamin Adams

Department of Computer Science and Software Engineering, University of
Canterbury, Christchurch, New Zealand

In our introduction we described the three pillars of the workshop which gave rise to this book. The first was concerned with the identification of themes related to the environment, and the nature of the multidisciplinary questions which might be explored through text. The second concentrated on the resources and methods available to us which might enable addressing these questions, and the third focused on the development of individual, illustrative case studies. These pillars give a useful framework for some concluding remarks, identifying areas of common ground and potential for future work. These remarks are structured around four elements. Firstly, we discuss the nature of

How to cite this book chapter:

Purves, Ross S., Olga Koblet, and Benjamin Adams (2022). "Surveying the Terrain and Looking Forward." In: *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. Ed. by Ross S. Purves, Olga Koblet, and Benjamin Adams. London: Ubiquity Press, pp. 233–244. DOI: <https://doi.org/10.5334/bcs.l>.
License: CC-BY 4.0

the collections analysed, before moving to the methods used to explore these data. We then step back, and explore not only the nature of the questions posed in the book, but discuss the potential and limitations of some of the results presented in the book. Finally, we set out a research agenda for future work, picking out some potential themes for research. We do not make claims for exhaustivity in any of these elements – rather, our aim is to illustrate the, largely to date unrealised, potential for the analysis of unstructured text in addressing pressing scientific, societal and policy-related environmental research questions.

Our case studies embraced a very broad range of resources for their analysis. The nature of these sources reflected not only the types of questions being explored, but also the research interests and backgrounds of those exploring them. Thus, for example, Tobias Zuerrer's starting point given his background in English literature, was a classic novel of the 19th century, Mary Shelley's *Frankenstein*. As an out-of-copyright classic novel, where the protagonists travel in a range of landscapes, it provided an accessible and appropriate starting point for questions concerning ways in which the industrial revolution was reflected in descriptions of urban and natural scenes.

Sarah Luria and Ricardo Campos were concerned with similar questions, but chose a particular location, the historical Canal District of Worcester, Massachusetts and diverse narratives about the deindustrialisation of the city over time. By definition, their study required a collection of texts from different times and different authors. Since one author, Sarah, was familiar with the story of the Canal District, she hand-built a small but very diverse corpus of texts capturing these different voices in quite different genres, ranging from poetry through to news reporting. The selection of texts was neither objective nor exhaustive, but (in common with many approaches from the humanities) no such claims are made. Karen Jones, Diana Maynard and Flurina Wartmann took a somewhat similar approach to building a corpus of historical documents about Loch Lomond in Scotland. However, their approach was different in that they started from a small set of historical documents identified by searching online archives, before comparing their documents with lists compiled about the region. Their aim was therefore to create a more systematic corpus of all travel writing about Loch Lomond in the 18th and 19th centuries, and they acknowledge that doing so would likely require digitisation of sources currently only available in analogue form. Unlike Sarah and Ricardo, Karen and her colleagues focused on a particular genre of writing, that of travel guides, implying that these sources are likely to throw light on views of the region from a particular perspective, which may be quite different to that experienced by its inhabitants. Nonetheless, both of these studies used geographically constrained collections to specifically ask questions about particular locations.

Two further studies also chose to focus their research on a particular location, though in these cases the choice of place, Rannoch Moor, was an artefact of a set of texts used in discussions at the workshop from which this book

stemmed. Joanna Taylor and Ben Adams used writing from W. H. Murray and Robert Macfarlane about Rannoch Moor as a starting point for a discussion about the influence of gender on writing about wild places, adding travel writing and poems to build an initial corpus. However, the location is not central to their hypotheses, and they extended their work to include more nature writing, in the form of the Guardian's Country Diary column. Here, Joanna and Ben deliberately choose a collection of texts of the same genre and from the same source available online through an application programming interface (API). Their second collection was much larger than the small collections used by our authors so far, consisting of more than 6000 articles. Simon Scheider, Ludovic Moncla and Gabriel Viehhauser started by analysing the same two pieces from Murray and Macfarlane, but with a very different aim. They wished to explore how frames of reference are used in writing, in order to improve methods which extract locations from such texts. They used these texts to develop a model, which they then tested on a different genre of text, the historical novel *Kidnapped* by Robert Louis Stevenson. This choice again gave the team access to an out-of-copyright and digitised source, which much like *Frankenstein* included rich spatial texts describing the main protagonists' journeys through Scotland.

The final two case studies were less directly concerned with explicit spatial locations, and more so with writing about specific subjects. Joanna Taylor and her colleagues investigated how landscape is assigned value in a very technical genre of writing, so-called Landscape Character Assessments, which are widely used in landscape policy and management in the UK. This technical form of writing meant that the authors could look for specific elements of documents in their corpus, but they were hindered by the structure of the documents available online. Unlike most of the other texts analysed in the book, these were often Portable Document Formats (PDFs), meaning that before applying computational methods to raw text this text first had to be extracted from structured documents, and elements such as figure captions, text contained in tables, information boxes had to be filtered from the main body of the text. This case study illustrates clearly that even where a specific genre is digitally available, these texts may require considerable preprocessing before substantive analysis can commence. It also hints at the related issue of context – how much, for example, of the writing in the historical guidebooks analysed by Karen Jones and colleagues, rely on context given by accompanying maps and sketches, and what is lost when our analysis ignores this? Katrín Lund, Ludovic Moncla and Gabriel Viehhauser took a different starting point, a specific location in Iceland, for their diachronic study of how glaciers are captured in narrative. In contrast to the other works we report on, they started with much larger initial collections in multiple languages and of quite different genres. These included British parliamentary proceedings, a well-known German news magazine, *Der Spiegel*, and German versions of the Swiss Alpine Club's yearbook

from the corpus Text+Berg. They extracted potentially relevant articles using keyword searches for 'glacier' and its synonyms. However, as they acknowledge, such an approach to corpus building does not deal with word sense ambiguity, for example with respect to metaphorical uses of glacier in text. Since such usage can be both a function of genre and time, it is an important consideration in not only the building of corpora but also the filtering of relevant texts for further analysis.

To scholars from the humanities, this discussion of the nature of sources and their origins is perhaps obvious. However, for those more accustomed to working with other forms of digital data – for example, in the form of terrain models or land cover data – it is important to emphasise the challenges in understanding how the collections selected can, and do, influence our analysis. Equally important is a recognition that large collections are not necessarily more effective ways of studying specific questions computationally, and that the importance of domain knowledge with respect to the theme under investigation cannot be overestimated. Our case studies illustrate the breadth of sources amenable to computational analysis, and leave us ready to discuss the methods applied by our teams. However, before we do so it is important at this point to make some caveats. In particular, most of the texts we worked with were in English, and their settings were European and North American. Thinking about ways of including both questions and sources from the Global South is an important challenge, and one that we do not address here.

This book is predicated on the potential of computational analysis of environmental narratives to extract information not otherwise accessible. Each of our teams used a combination of techniques to analyse text, and importantly for the reader of this book, the focus was very much on the use of existing approaches, rather than development of new methods. Many of these methods were similar to those we introduced in Chapter 3, and together they give a good overview of starting points for future work.

Our first case study, from Katrín Lund and colleagues sought to explore the narratives existing in three distinctive collections over time. Computationally, standard methods were applied to characterise the three corpora over time, using simple frequency-based approaches to suggest potential themes in these corpora. Such methods are essentially language-independent, and the authors chose to remove only stop words before analysing the remaining tokens in each corpus. By using collocation, the team zoomed in on some of the typical themes discussed with respect to glaciers, though arguably those identified are relatively unsurprising given the nature of the corpora. Through microreading of individual texts the importance of the use of 'glacier' metaphorically, particularly in the parliamentary corpus, became apparent. This importance of metaphor, and the potential of its prevalence as a function of individual collections is an important methodological consideration, since its detection requires examination of the source material in detail.

The computational approach taken by Sarah Luria and Ricardo Campos was in many ways similar to that of Katrín, Ludovic and Gabriel. Just like the previous study, the aim was to identify salient terms used in documents, however rather than characterising an entire corpus, Sarah and Ricardo summarised individual documents from a hand-picked corpus. Because some of the documents in this corpus were not in a suitable digital form, an initial pre-processing step using optical character recognition (OCR) was required and, as is often the case with historical texts, some post-processing was also necessary to deal with errors in the OCR process. Rather than simply extracting high-frequency terms or significant collocates, Sarah and Ricardo applied a bespoke piece of software, Yake!. Yake! uses an unsupervised approach to extract keywords and is language-independent. This means it can be applied directly to individual documents or a complete corpus. In practice, simple features such as frequency and collocates are used to identify important keywords, which may also take the form of *n*-grams. Because Yake! was used on a small corpus, and because Sarah had detailed knowledge about both the process and the sources being explored, it was possible to perform a much more detailed, but qualitative, evaluation of the terms extracted and represented as word clouds. Interpreting these word clouds was only possible given Sarah's underlying knowledge of the corpus, and the conclusions drawn are thus inherently dependent on both the macroreading performed by Yake! and Sarah's microreading of the individual texts.

Karen Jones, Diana Maynard and Flurina Wartmann investigated historical travel writing about Loch Lomond in Scotland in their case study. Similar to Sarah and Ricardo, after selecting an initial collection of documents, pre-processing was necessary, though in this case with the aim of removing extraneous material such as indexes and dividing books into small enough sections for processing. Like the previous two case studies, Karen and colleagues looked at individual words and their use in their collections, however their starting point is more semantically constrained. Using the text analysis toolkit GATE, which Diana has played a key role in developing, the team identified parts of speech and extracted landscape elements and place names. Doing so required the creation of curated lists of relevant terms. Unlike the first two case studies, Karen and her colleagues explicitly link narrative to space by, for example, mapping the order in which locations were discussed in the texts, and compared these computationally extracted lists with microreadings of the text.

They also explored co-occurrences, but did so for their lists of landscape terms and investigated their relationship with a list of more abstract terms relating to landscape. Like Sarah and Ricardo, the importance of interpreting and discussing the computational results from a particular perspective, in this case that of an environmental historian is once again a key element of the case study.

Tobias Zuerrer also investigated landscape perception in text, setting out to explore a potential dichotomy (that of urban and natural landscapes) using a curated set of seed terms containing both generic landscape terms and

toponyms. Tobias used an off-the-shelf tool, AntConc, to explore the frequency of seed terms, and used concordances to perform a simple microreading and annotate positive and negative connotations. He very effectively demonstrates how an existing tool can be used to analyse texts with no need for programming skills, and provides an excellent example of what is achievable through carefully considered questions and existing tools with respect to the computational analysis of text.

The first four case studies all took essentially exploratory approaches, using simple methods such as term frequency, collocation and order to select and discuss particular environmental narratives, visualising these through word clouds, tables and simple maps. Joanna Taylor and Ben Adams took a similar initial starting point to exploring how gender influenced the use of pronouns in writing about Rannoch Moor, using concordance plots and collocates. Having demonstrated that authors, in an initially small corpus, appeared to describe the landscape differently according to gender, Joanna and Ben tested their hypothesis by developing a supervised classifier capable of assigning gender to a text based on the language used. Doing so required training and test data with gender annotations, and they used an existing tool to automatically assign annotation based on forenames. For the classification itself, they used a well-known classifier, Naive Bayes, which treats a document as a bag of words and assigns it a probability of belonging to a particular class. In a back and forth that lies at the core of the methodological approach taken, Joanna and Ben then identified collocates used with pronouns in their corpus and used these as a basis for further microreading.

Joanna and Ben relied on a supervised classification in their study of diversity of voices about wild places. Such a classification implied in turn that classes exist. In their investigation of value in the widely used instrument Landscape Character Assessment, Joanna Taylor, Meladel Mistica, Graham Fairclough and Timothy Baldwin took a different approach and used an unsupervised approach, topic modelling. Similar to the work of Karen and colleagues, an important methodological pre-processing step was the extraction of relevant text. Here, the challenge was not noisy Optical Character Recognition (OCR) on historical texts, but rather extracting meaningful structure from rich PDF documents which use text boxes, figures, and tables to ease reading and salience for humans. After preparing the texts, topic modeling, and specifically latent Dirichlet allocation (LDA), assigned individual words the probabilities of being associated with particular topics. Since topics also consist of statistically related words, Joanna and her colleagues used LDA as a way of exploring the extent to which value was explicitly and implicitly described in LCA. Interestingly, using computational methods allowed the authors to reframe their understanding of value in terms of LCA through microreading. Here, somewhat in contrast to the earlier studies, the computational analysis of the texts explicitly suggested new ways of interpreting the content, through a change in the way in which

LCAs were read paying attention to properties of the text revealed through topic modelling.

The final case study, from Simon Scheider, Ludovic Moncla and Gabriel Viehhauser, explored how space was referenced in environmental narratives. Like several other studies, the authors annotated texts, in this case though as a first step in defining differing ways in which frames of references were constructed. Using this annotation, it was possible to propose a set of rules which were then implemented in the *Perdido Geoparser*. Thus, like all of the other studies, Simon and colleagues applied well-tested existing methods to their problem. They emphasised the challenges in annotating frames of reference consistently, and the need for an iterative process to capture a concept about which humans do not always agree. The results of applying their approach to a new text illustrated that the challenge is not simply encoding rules correctly (as captured by precision), but having a training data set with sufficient examples to cover possible cases – as reflected by the low recall of their approach. Crucially, the methods developed by Simon, Ludovic and Gabriel allow us to start to extract complex spatial frames of reference, such as are commonly found in the texts which form the subject matter of this book.

In listing the methodological approaches taken in this book, a few points stand out. Firstly, and most importantly, all of our authors adopted what Joanna Taylor called a ‘multiscalar analysis’ – that is to say used different approaches to collecting, analysing and interpreting texts, moving fluidly back and forth between macro and microreadings. This multiscalar approach was enabled in all but one single-authored piece by inherently multidisciplinary teams, who worked together to bring a range of approaches to the table. This is perhaps best illustrated by Sarah and Ricardo’s piece, which by documenting some initial misunderstandings helps uncover the need for a constant dialogue in such work.

Secondly, all of our teams made use of relatively long established methods, rather than state of the art machine learning approaches which are currently being applied to a wide range of tasks in natural language processing and are gaining popularity in the digital humanities. This does not, we think, mean that these methods do not have potential in the analysis of environmental narratives. Rather, however, where simple off-the-shelf methods allow first exploratory insights, these may be an effective way of starting discussions between disciplines such as those exemplified in our case studies.

As we argued in the introduction, starting these discussions also requires that meaningful questions are identified. A strength, we would argue, of the case studies here is the interdisciplinary inputs to both the research questions and the methodological approaches taken. Arguably, the results are often modest, and many are either inconclusive or suggest starting points for further work rather than delivering deep insights. We have already shown great variation in the nature of the collections analysed which contrasted strongly with the relatively

consistent use of standard off-the-shelf methods in much of the analysis. What though of the questions asked in our studies – do these reflect the breadth we found in the ways that workshop attendees asked questions when given a single short text to reflect on, or does the relative homogeneity of the methods applied constrain the ways in which these texts are approached computationally?

In the introduction we used a simple framework – the 5Ws & H (what, why, when, where, who and how) as a tool to categorise ways in which questions were asked of our texts. What happens when we do the same with our case studies?

Katrín and colleagues set out to explore the influence of different voices in narratives about glaciers at multiple locations and times. Computationally, they used corpora in different languages (English and German), of different genres (parliamentary records, news reporting and mountaineering yearbooks) and with historical depth as proxies for the questions where, who and when? In practice, they could start to ‘track down the traces of the multitude of voices ... hidden in large text corpora’, but also struggled to reconcile the qualitative nature of the questions suggested by microreading of environmental narratives with the broad conclusions that could be drawn through a macroreading limited by the need to deal more effectively with metaphor.

These limitations are interesting, since in many ways the approach taken by Sarah and Ricardo was very similar. They too wanted to explore narratives about a particular location – the Canal District (where), changes in these narratives over time (when), and explore both the voices (who) and forms (how) of these narratives. Their focus on a particular district, their use of a curated corpus, whose constituent parts one of the authors was very familiar with, and their hermeneutic back and forth led them to be very positive about the possibilities of understanding the process of revitalisation and gentrification in the Canal District through text analysis. We believe this points to an important dichotomy between the needs of scholars concerned with environmental narratives and those interested in computational methods. Despite the allure of running methods over very large corpora, these methods are essentially limited by the need for microreading to understand context, allowing an iterative back and forth with the material being researched.

This in turn leads us to an argument for what Joanna Taylor and colleagues argued for as multiscale approaches. In their approach to exploring a very specific genre of document (landscape character assessments), the team focused on understanding a specific aspect – value (what?) and the ways it was implicitly and explicitly described in these documents (how?). By again building in a back and forth between computational analysis and microreadings of their texts, Joanna and team were able to point to some discrepancies in the ways LCA sets out to describe character without ascribing value, and the reality of the close link between character and distinctiveness and value. Furthermore, by identifying a potential link in who wrote these LCAs, and in particular their

geographic origins further potential questions are hinted at (where, and by whom is landscape value ascribed?).

Perhaps unsurprisingly Joanna, this time with Ben Adams, once again took a multiscalar approach to exploring differences in voices (who) describing wild(er)ness (what). Here, the starting point was an analysis of individual documents to explore the influence of gender on descriptions of wild places, in particular through the use of pronouns (a classic how question). The locations of these descriptions (for the initial analysis around Rannoch Moor, and then more broadly concentrated in the UK) are not explicitly considered, but, just like the language of analysis (English here), are important caveats, since these empirical results are specific to the languages, locales, and corpora analysed.

Nonetheless, historical writing and literature influences how environments are perceived today. Karen Jones with Diana Maynard and Flurina Wartmann looked at historical travel writing to identify patterns in how a particular location, Loch Lomond (where) was described (what). Karen and her group took a similar hybrid approach to that argued for by Joanna, emphasising their exploratory approach by naming it 'forensic fishing.' Interestingly, this study is the only one which puts locations from text into their geographic context on topographic maps. Furthermore, the study hints at an underlying process in the writing of these texts – the process of place-making, in this case through the genre of travel guides – and the diversity of voices ignored in this writing.

Tobias Zuerrer takes a different approach to this forensic fishing in his analysis of *Frankenstein*, and starts from a guiding question: whether there are differences in how urban and natural landscapes are conceptualised in the novel. His question can be seen as an exploration of what and where, guided by a specific and well-formulated research question.

Our final study, and the only one to specifically develop new methods, is all about location, and as such poses the question as to where specific passages can be located – an important task if we are to georeference using more complex approaches than the simple toponym-lookup applied by Karen Jones and her team. Simon Scheider and colleagues did so by developing and implementing a model which captures different ways (a how question) spatial information is conveyed through frames of references.

Analysing our case studies, we see much more diversity in the questions posed than the methods used. We find ample examples of studies addressing five of the six questions (what, when, where, how and who) and it appears that the relative homogeneity of computational methods is overcome by on the one hand the diversity of the sources used by the teams, and on the other by the disciplinary backgrounds of those working on a particular problem. Each of these questions can, at least at a superficial level, be answered by, for example, counting and extracting words appropriately. Thus, for example, we could explore ways in which climate is described over time by extracting and comparing

adjectives for some given corpus, stratified by time. Of course, as our case studies show, by combining micro and macroreadings of material, it is possible to interpret in much more depth. However answering (as opposed to posing) why questions computationally requires, we suggest, building upon the foundations laid in this book.

Since this book is about environmental narratives, it is also worth reflecting on the lenses through which our teams explored the environment. Change was an important theme – both as anthropologically driven climate change in the context of changing attitudes to glaciers, and with respect to attitudes as the urban environment of the Canal District was ‘revitalized’. The influence of the past on current ways of exploring Loch Lomond and its environs can be seen as one way of exploring what and how a particular landscape is valued, contrasting effectively with the analysis of a particular contemporary management genre, Landscape Character Assessment. Potential differences between the ways in which particular sorts of environments are described in a single text, the urban and natural of Mary Shelley’s *Frankenstein* suggests another way of exploring how the environment is implicitly and explicitly valued in text. Turning all of these ideas on their head, and looking not at the observed, but the observer, and studying the influence of gender on ways in which environments are written about reminds us that narrative should be studied in context, and that power (who writes, which languages are digitised, for which languages are tools available, who researches) has important consequences for any interpretations. Equally, understanding the diversity of ways that locations can be described, and recognising that this is not language-independent, emphasises the need to develop methods applicable not across cultures, but rather to cultures. Although we almost exclusively worked with English, it is also worth pointing out that the methods applied may struggle with texts not written in modern English amenable to use of basic text analysis.

In a book of this kind, it is traditional to close with a research agenda. Such agendas though often become prescriptive, limiting the diversity of research in a field and constraining imagination. With this in mind, we have chosen to present not a research agenda, but rather some starting points for future work, which we believe might be fruitful in developing the potential of computational analysis of environmental narratives. Our ideas are seeded by the potential demonstrated by the case studies in this book.

Perhaps one of the most surprising findings in hindsight was the power of computational analysis in individual volumes or small, curated corpora, rather than the big data analysis so often trumpeted as the way forward in contemporary research. Our teams and their interdisciplinary compositions were much better suited to a productive mixture of macro and micro-analysis. This micro-analysis, was most productive where the thematic specialists were involved in the selection of texts and their qualitative exploration. At the scale of our case studies, which can be seen as pilots for future research, an important limiting

factor was the volume of text which could be read, rather than computationally analysed. We suggest that future work on environmental narrative takes heed of this productive combination of well-chosen research questions and small, thematically focused collections as a starting point for research.

A second key finding of the work carried out in this book was the utility and effectiveness of existing, well-known methods for text analysis. Although the promise of machine learning and machine understanding of text is one which has gained much attention in natural language processing, we believe that effective research on environmental narratives should take advantage of existing, well-established and crucially, well-understood methods. For example, simple approaches to classification can deliver more than adequate performance for many of the questions of interest to our teams. Research on the computational analysis of environmental narratives should of course take advantage of general trends in text analysis. However, since the focus of work on environmental narratives is on explanation and understanding, future research should consider carefully how to combine existing, well understood methods with the most potential to generate insightful results.

This finding does not however preclude methodological development of particular relevance to environmental narratives. The study by Simon Scheider and colleagues on frames of references is an excellent example of such a study, since the ways in which locations are described in narratives of this nature are very rich, and unlikely to be identified in more general corpora such as news or social media. Working on a particular genre of texts was informative and led to much more productive research. This leads us to our third suggestion: that the development of methods can be productive when driven by use cases and collections directly relevant to understanding the environment. This emphasis on understanding, as opposed to methods in isolation, chimes with current debates in the digital humanities about the use of computational methods (Robertson and Mullen, 2021), and we believe it is an important result of the truly interdisciplinary process of writing this book.

Development of new methods presupposes that a variety of resources exist. These include collections of text about the environment, annotations of texts with respect to, for example, landscape preference or sentiment and resources such as environmentally specific gazetteers and lexicons. In writing this book all of these elements were hard to find, and we encourage future researchers to give much more consideration to not only reproducibility and replicability at the level of individual publications, but also considering shared tasks and resources more intensively to further research. Examples with great potential for more effective use include the Corpus of Lake District Writing (Rayson et al., 2017), the Text+Berg corpus (Volk et al., 2010) and resources available through APIs such as the Guardian's Country Diary, all of which were used within the case studies in this book.

As we wrote the text accompanying our case studies, one major challenge was exploring results, especially from larger datasets, with static visualisations, often limited to displaying the most highly ranked terms. More effective visualisations, capable of linking different views and moving beyond simply placing documents or texts on a map remains an important challenge for future work. Computational analysis of environmental narratives could thus be a very productive area for interdisciplinary work on effective and efficient visualisation and visual analytics. This need for exploration is important, as the approaches developed here will provide the bridge between macroanalysis and micro-reading. Providing more integrated ways to move from overviews of datasets, through zooming and filtering to details on demand, as proposed by Schneiderman in his influential information-seeking mantra (Schneiderman, 2003) would be an important contribution if we are to more effectively interpret material in context.

Productive interdisciplinary work lay at the heart of the case studies around which this book is based, and also lies at the core of our last recommendation. Future work should start not from the identification of methods or datasets, but with productive questions posed by experts with underlying thematic expertise. The potential of environmental narrative as source for computational analysis mediated by humans is, we believe enormous, and hope this book can stimulate future work in the field.

References

- Rayson, Paul, Alex 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: *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, pp. 9–15. DOI: 10.1145/3149858.3149865.
- Robertson, Stephen and Lincoln Mullen (July 2021). "Arguing with digital history: Patterns of historical interpretation". In: *Journal of Social History* 54.4, pp. 1005–1022. ISSN: 1527-1897. DOI: 10.1093/jsh/shab015.
- Shneiderman, Ben (2003). "The eyes have it: A task by data type taxonomy for information visualizations". In: *The craft of information visualization*. Amsterdam: Elsevier, pp. 364–371. DOI: 10.1016/B978-155860915-0/50046-9.
- Volk, Martin, Noah Bubenhofer, Adrian Althaus, Maya Bangerter, Lenz Furrer, and Beni Ruef (2010). "Challenges in building a multilingual alpine heritage corpus". In: *Seventh International Conference on Language Resources and Evaluation (LREC)*.