University of Zurich[UZH]

Exploring the frequencies and thematic context of place names in the Swiss National Park in the past 110 years using computational text analysis

GEO 511 Master's Thesis

Author
Simona Di Vincenzo
16-711-996

Supervised by
Prof. Dr. Ross Purves
Dr. Stefanie Gubler (stefanie.gubler@scnat.ch)
Tamara Estermann (tamara.estermann@nationalpark.ch)

Faculty representative
Prof. Dr. Ross Purves

29.04.2024
Department of Geography, University of Zurich

GEO 511 Master's Thesis

Exploring the frequencies and thematic context of place names in the Swiss National Park in the past 110 years using computational text analysis

Student:
Simona Di Vincenzo

Supervisors:
Prof. Dr. Ross Purves
Dr. Stefanie Gubler (SCNAT)
Tamara Estermann (SNP)

Department of Geography, University of Zurich
Prof. Dr. Ross Purves
Zurich, April 29, 2024

Acknowledgements

First and foremost, I would like to express my heartfelt thanks to my supervisor, Prof. Dr. Ross Purves. His constant readiness to help has guided me through this lengthy process, always providing expert knowledge when needed. Special thanks go to my wonderful co-supervisors, Dr. Stefanie Gubler and Tamara Estermann, for their enriching conversations and enthusiasm for the work, which consistently inspired me.

To my family, thank you for their unwavering support without needing to ask. Especially to my sister Monica Di Vincenzo, who was always two steps ahead and took me along with her. Without her, I would not have come this far. To my housemates, Anna Ingwersen and Sophie Hauller, who not only provided me with a second home but also shared this significant phase of life with me, creating memories I cherish deeply.

To my dear friend Inga Birkhölzer, thank you for our friendship, your encouraging words when times were tough, and for proofreading this thesis. Thank you.

Furthermore, I would like to thank my good friend Anna Dieckmann for the countless hours we spent working together to finish this year, studying and supporting each other through our final challenges.

And finally, my gratitude goes to my boyfriend, Daniel Wiesehahn, who has supported and always believed in me.

Zurich, April 2024 Simona Di Vincenzo

Abstract

This Master's thesis examines the evolution of place names in the Swiss National Park over the past 110 years using computational text analysis. Analysing six diverse corpora, including atlas, Cratschla magazine, literary works, research publications and business reports, the study tracks the frequency and thematic contexts of 27 key place names to recognise political and environmental changes. The methodology combines frequency analysis, which observes the temporal dynamics of place names, with thematic analysis that delves into the contexts within which these names appear.

Key findings reveal significant consistency in mentioning certain locations such as Zernez, Il Fuorn, Trupchun, Stabelchod, Macun, Spöl, and Buffalora across all texts, underscoring their importance in the park's management and environmental narratives. The thematic analysis, supported by twelve defined fields from "Causes for Landscape Change", "Habitats", and "Animal", accurately depicted the locations' characteristics and habitats. The study also shows that variations in the depiction of place names across different document genres and publication years significantly affect their frequency, suggesting that a single metric like frequency is insufficient for understanding the full narrative scope of place names.

These insights highlight the need for more sophisticated analytical tools and methodologies. Challenges such as limitations in OCR technology and thematic categorisation were encountered, indicating a need for advanced preprocessing strategies and statistical methods such as TF-IDF, t-score, or mutual information. By refining these techniques, future research could provide deeper and more precise explorations into how place names document the environmental and historical narratives within the Swiss National Park.

Keywords: Toponyms, National Parks, Protected areas, Computational Text Analysis, Natural language processing, Optical Character Recognition, Pytesseract

Contents

List of Figures **ix**

List of Tables **xi**

1 Introduction **1**
1.1 Context . . . 1
1.2 Goals and Research Objectives . . . 3

2 Related Work **5**

3 Background **9**
3.1 The Swiss National Park . . . 9
3.1.1 Overview . . . 9
3.1.2 History . . . 11
3.2 History of Geographical Names in Switzerland . . . 12
3.2.1 History and Characteristics of the Places of Interest . . . 13

4 Methods **20**
4.1 Data sources, preparation and processing . . . 20
4.1.1 Document collections . . . 20
4.1.1.1 Overview . . . 20
4.1.1.2 Time coverage of the corpora . . . 24
4.1.1.3 Optical Character Recognition OCR . . . 25
4.1.1.4 Difficulties with OCR . . . 25
4.1.1.5 Post-processing of text-files . . . 27
4.1.2 List of place names . . . 28
4.1.2.1 Overview . . . 28
4.1.2.2 Potential problems with extracting place names . . . 28
4.2 Thematic fields . . . 31
4.2.1 Derivation of the thematic fields . . . 31
4.2.2 Noun extraction in the corpora . . . 31
4.2.3 Iterative and manual categorisation . . . 32
4.2.4 Description and characteristics of the thematic fields . . . 34
4.2.4.1 Primary Themes/Habitats & Animals . . . 34

4.2.4.2 Causes for landscape change 35
4.3 Development of Analysis Plots: Methodology and Application 37
4.3.1 Frequency Analysis: Steps to Creating Visualisations 37
4.3.2 Thematic Field Analysis: Procedure and Graphical Implementation 37
4.3.3 Evaluation: Precision, Recall, and $F1$ 38
4.3.4 Normalisation and its risks 41

5 Results and Interpretations **42**
5.1 Basic statistical overview 43
5.1.1 Token and locations counts in all corpora 43
5.1.2 Mentions of locations & thematic fields in paragraphs 44
5.2 Composition of thematic fields in the corpora 45
5.3 Most frequent places in the corpora 48
5.4 Atlas 52
5.4.1 Frequency of place name 52
5.4.2 Thematic context 54
5.5 Brunies (1914/48) 57
5.5.1 Frequency of place name 57
5.5.2 Thematic context 59
5.6 Business Reports 62
5.6.1 Frequency of place names 62
5.6.2 Thematic context 64
5.7 Relationship between primary themes and causes of landscape change 68
5.8 Relationship between primary themes and animals in the Business Reports and NPF publications 70

6 Discussion **72**
6.1 Summary of Key Findings 72
6.2 Methodological Reflections and Limitations 77

7 Conclusion **80**

Bibliography **82**

A Appendix **87**

B Appendix **95**
B.1 Cratschla 95
B.1.1 Frequency of place names 95
B.1.2 Thematic context 98
B.2 National Park Research publications (NPF) 99
B.2.1 Frequency of place names 99
B.2.2 Thematic context 101

8 Personal Declaration **103**

List of Figures

3.1 Location and of the Swiss National Park (Canton Graubünden) in the southeastern part of Switzerland (Source: (H. Haller, Eisenhut, and R. Haller, 2013, p. 15)) 10
3.2 Map Overview of Study Locations. 14

4.1 Overview of temporal coverage of the documents publications 24
4.2 Example of a satisfying result of OCR. 25
4.3 Example of a poor result of OCR. 26
4.4 Compilation for the categorisation in the thematic fields. 33
4.5 Process of annotation and evaluation. 39

5.1 Overview of the thematic field distribution in the 5 corpora. Normalisation: topic's occurrence count per corpus/total word count in said corpus. 46
5.2 Composition of thematic fields in the 6 corpora. 47
5.3 Overview of normalised frequencies of the locations in each of the five corpora. (Normalisation = count of specific place names in each corpus/total count of place names in said corpus) 50
5.4 Ranking the location from 1 to 27 for all 6 corpora. 51
5.5 The normalised frequency of field and place name in the Atlas (2013). The normalisation was conducted by dividing the count of specific place name by the total count of all place names. 52
5.6 Normalised co-occurrence bar chart of most relevant locations and thematic fields in the Atlas. 54
5.7 Absolute count of Co-occurrence of Causes for Landscape Change and Primary Themes/ Habitat with the Locations Ofenpass, Trupchun, and Spöl. 55
5.8 Normalised frequency of field and place names in Brunies (1914) and Brunies (1948). The normalisation was conducted by dividing the count of specific place name by the total count of all place names. 58
5.9 Difference between Brunies 1914 and 1948 in the appearances of field and place names. The normalisation was conducted by dividing the count of specific place name by the total count of all place names. 59
5.10 Example S-charl: Co-occurrence with the thematic fields in 1914/48. 60

5.11 This plot shows the sum of normalised appearances over the whole publication period of Business Reports, 1915-2021. Normalisation: Count of Specific Place Name in Publication/ total count of all place names. 62
5.12 Example: Buffalora, Spöl and Macun. 63
5.13 Examples Buffalora, Macun, and Spöl: Co-occurrence with the thematic fields. . 65
5.14 Absolute count of Co-occurrence of Causes for Landscape Change and Primary Themes/Habitat with the Locations Buffalora, Macun, and Spöl. 66
5.15 Co-occurrence in paragraphs of "primary themes" and "causes of landscape change". 68
5.16 Co-occurrence in paragraphs of "primary themes" and "animals". 70

B.1 This plot shows the sum of normalised appearances over the whole publication period, 1992-2023. The normalisation was conducted by dividing the count of specific place name by the total count of all place names. 95
B.2 Example 1: Macun, Spöl, Ova Spin and Trupchun. 96
B.3 Examples Macun, Spöl and Trupchun: Co-occurrence with the thematic fields. . 98
B.4 Sum of normalised frequency field/place names in the NPF. Normalisation: Count of Specific Place Name in Publication/ total count of all place names. 100
B.5 Example: Il Fuorn, Trupchun, and S-charl 101
B.6 Example Il Fuorn, S-charl, Spöl, and Trupchun: Co-occurrence with thematic fields. 102

List of Tables

4.1 Overview of the corpora and metadata. 23
4.2 Locations of interest. 28
4.3 Overview of alternative writings of the place names. 29
4.4 Thematic fields in their subcategories. The underlined thematic fields depict the human influence on landscape change. 31
4.5 Total noun count per paragraph and phrase per corpora. 32
4.6 Noun count per thematic field in the first, fourth and final iterative process. . . . 34
4.7 The nouns were categorised into the fields according to the following affiliations. 36
4.8 Example Atlas: Precision, Recall, and $F1$. 40
4.9 Precision, Recall, and $F1$ for all six corpora. 40

5.1 Overview of the total token count and the share of the 27 locations appearances in their respective 6 corpora. Table A shows the result with the processed text file (2) and Table B with processed text files (3) 43
5.2 Overview of the number of paragraphs, the percentage of locations, and the percentage share of the co-occurrence of locations and thematic fields in the paragraphs. 44
5.3 Thematic fields in their subcategories. The underlined thematic fields depict the human influence on landscape change. 45
5.4 Overview of normalised place name count in all corpora. The normalisation was done by the total count of place names in each corpus, e.g., the absolute count of the place name "Alp la Schera" in the Atlas/total counts of all place names in the Atlas. Each column sums to 1. The green cell with the number in brackets indicates the five highest frequencies. 49
5.5 Example: Most frequent nouns associated with Ofenpass, Trupchun, and Spöl. . 56
5.6 Example S-charl: Most frequent nouns associated with S-charl in 1914 and 1948. 61
5.7 Example: Most frequent nouns associated with Spöl, Buffalora, and Macun in single publications 67
5.8 Top 5 Nouns per Habitat that co-occur with the thematic field "Animals" in the Business Reports and NPF publications. 71

B.1 Example 1: Most frequent nouns associated with Spöl, Macun, and Trupchun in single publications. 99

1. Introduction

1.1 Context

Protected areas like national parks are crucial in preserving landscapes, ecosystems, and biodiversity. Besides ecological reasons — such as preserving biodiversity, native flora, and fauna - they also maintain cultural and historical characteristics of the landscape and provide invaluable recreational spaces (BAFU/WSL, 2022, p. 10). According to records in the UN database as of 2008, there are over 4'000 national parks worldwide, spanning an impressive expanse of approximately 5 million km², which accounts for about three percent of the Earth's terrestrial surface (Kupper, 2012, p. 9). The national park concept has its roots in 1872, with the founding of Yellowstone National Park, as documented by the World Database of Protected Areas, an initiative by the United Nations Environment Programme (UNEP) and the International Union for Conservation of Nature (IUCN) (Kupper, 2012, p. 11). This marked the beginning of a movement that has seen exponential growth over the past 150 years (Kupper, 2012). This was also the case in Switzerland with the founding of the Swiss National Park (SNP) on August 1, 1914, making it the first national park in the Alps and Central Europe (Bundesamt für Umwelt, 2024; Baur and Scheurer, 2014, p. 20). Unlike the other national parks of the time, the SNP not only stands as the first national park in the Alps but also as a living laboratory where the interplay between conservation efforts and natural dynamics can be observed (Baur and Scheurer, 2014, p. 20). With an area of 170.3 km², the park is the oldest in the Alps and central Europe, and according to the IUCN, it belongs to the highest protection class (category 1a). It is considered a strict nature reserve (Bundesamt für Umwelt, 2024). It offers a unique historical perspective on landscape development with as little human influence as possible due to the parks' longevity and due to the detailed (written) documentation by the SNP has accumulated. This thesis delves into the extensive documentation available to trace the evolution of the park's place names, revealing patterns of change and continuity. As we approach 2025, the SNP celebrates 111 years of dedication to protection, research, and education — the core tenets that form the foundation of the SNP. In anticipation of this significant milestone, the Swiss Academy of Natural Sciences (SCNAT) and the SNP are collaborating to curate a book and an exhibition centred around the "Landscape Change" theme. This thesis is intended to fit and contribute to this context. While research is a cornerstone of the Swiss National Park, with annual scientific publications focusing on habitat analyses of flora and fauna and landscape changes due to human intervention (Swiss National Park Data Center, 2024), the use of text-based sources to chronicle landscape change is

yet to be tapped into. My thesis aims to bridge this gap by leveraging textual sources to identify possible changes in the frequency and thematic context of place names within the park over time. This endeavour seeks to employ already established text analysis methodologies to recognise the thematic contexts of these place names. Computational text analysis will investigate the frequencies and thematic context in which 27 locations (located near/within the SNP) occur in text collections about the SNP. The texts consist of six corpora, each covering a different publication period. This analysis will use the **"Atlas des Schweizerischen Nationalparks. Die ersten 100 Jahre."** by Haller, H., A. Eisenhut, and R. Haller (2013), the first and fourth editions of Steivan Brunies' book **"Der Schweizerischer Nationalpark"** (1914/1948), the publication series **"Cratschla"** by the Federal National Park Commission (ENPK) and SCNAT Research Commission of the SNP (1992-2023), the **annual Business Reports of the ENPK** since 1915, and finally the publication of **"Nationalparkforschung in der Schweiz**" since 1918.

This work's overarching goals are to explore the frequencies and thematic context of the 27 place names in the Swiss National Park over the last 110 years using computer-assisted text analysis. In this analysis, two approaches were examined in depth. In the first part, called frequency analysis, the frequency of occurrences of place names in the last 110 years in the six corpora was examined and also illustrated over time. In addition to an overview of how often the place names occur, a few place names per corpus, which show interesting frequency patterns, were selected. Assumptions were made about possible reasons for these fluctuations and then checked for truth in the corpus. The thematic context in the corpora and place names was to be examined in the second part. For this purpose, twelve thematic fields, composed of six "Causes for Landscape Change" and five "Primary Themes/Habitats" fields and one thematic field containing the fauna of the park "Animal" field, were formed. The thematic fields were created by extracting all nouns from paragraphs and sentences containing a place name. The thematic analysis takes the most frequent and unique locations from the frequency analysis per corpus, a total of seven locations. It examines the thematic context of these places as a demonstration example. This means addressing questions about which thematic fields certain locations are most commonly associated with and if their thematic context sheds light on possible key events that shaped them. Here, findings and assumptions are also checked for truth in the corpora. Additionally, the relationships between the thematic fields are examined. With this approach, it should be possible to extract the thematic context of the 27 locations at any given point in time from the corpora and, if necessary, uncover the themes that have characterised these places over time. Great emphasis is also placed on the methodological approach of computer-assisted text analysis and the resulting visualisation of findings.

In fulfilling these aims, this thesis is structured as follows: In Chapter 2, (Related Work): This chapter outlines the existing literature relevant to this thesis, focusing on computational text analysis and its application in geographical and environmental studies. Chapter 3 (Background): The Swiss National Park is described in detail, covering its history, geography, and the evolution

of its conservation strategies. Locations of interest mentioned in the text are highlighted in *italics*. This chapter also provides the necessary background information about the place names used in the results chapter (3.2.1). Appendix A shows a complete list of descriptions. Chapter 4 (Methods): This chapter explains the data and methodology used to extract and analyse the place names and create thematic fields. Chapter 5 (Results): Findings and interpretations from the computational text analysis of the corpora Atlas, editions of Brunies, and Business Reports are presented. Due to the repetitive nature of the analysis across the six corpora, the scope of the results is extensive. Consequently, the results of the Cratschla and NPF publications have been moved to Appendix B. *Locations* mentioned in this chapter are italicised and linked to their descriptions, detailed in Subchapter 3.2.1 or Appendix A. Chapter 6 (Discussion): The results are discussed, the research questions (1.2) are answered, and the limitations of this study are presented. Chapter 7 (Conclusion): A summary of the key findings and exciting highlights are presented.

1.2 Goals and Research Objectives

This master's thesis explores the description of locations in the Swiss National Park over the last 110 years using computer-assisted text analysis. It provides a means or a guideline for making patterns recognisable with the given corpora and locations. Furthermore, it aims to identify suitable methods and tools within the Python work environment for efficiently processing scanned text collections into readable text files. This involves becoming acquainted with and utilising OCR (Optical Character Recognition) and other Python-compatible implementations.

The analysis has two objectives: a) to investigate changes in the frequency of appearances in the last 110 years of the locations in the corpora and b) to examine the thematic context of the 27 locations within the corpora. Specifically, it will assess how often, when, and in what context these place names occur in each corpus and the differences among the corpora. Considering the size and temporal span of the corpora, the findings from a) should reveal which place names are more or less frequently mentioned. Notable fluctuations in frequency will be scrutinised further. For example, a high frequency of a specific place name in a single publication year may indicate important historical or political events. Thus, a few examples of locations will be regarded in more detail for each corpus. Similarly, objective b) aims to analyse the thematic context of the most uniquely occurring place (based on the results of a)) through 12 thematic fields, which will be created by extracting all nouns from the corpora and categorising them. Similarly to a), only a few locations for each corpus will be analysed in depth as a role model example. In addition, I analyse how the 12 thematic fields relate and possibly influence each other.

The overarching aim of this thesis is:

Exploring the frequencies and thematic context of place names in the Swiss National Park in the past 110 years using computational text analysis.

Specifically, the following research questions will be addressed:

Research Questions:

i) How have the frequencies of place names in the Swiss National Park evolved over the past 110 years, and do they reveal political and environmental changes?

ii) How do thematic field co-occurrences reflect the known characteristics and any significant changes of the most frequent locations?

iii) How do variations in the depiction of place names across different types of documents reflect differing perceptions and priorities regarding the Swiss National Park?

iv) Which approaches and techniques are most suitable for extracting place names from historical sources such as journals, atlases, reports, and books in PDF?

2. Related Work

Computational text analysis uses computer science, linguistics, and statistics to examine text. This interdisciplinary approach uses the power of computers to process, understand, and generate insights from large volumes of text data that would be impossible for humans to analyse manually (Hirschberg and Manning, 2015). With the emergence of digital media, the amount of textual data available for analysis has exponentially increased, making computational methods useful and essential for extracting knowledge and understanding trends (Santana et al., 2023; Cambria and White, 2014). This applies to multiple research fields, not only computer science. To give a few examples: social and political sciences to study the possibilities, difficulties and dangers of big data and social behaviour (Alvarez, 2016; Atteveldt and Peng, 2018), medicine for improving healthcare services by studying thousands of patients surveys (Khanbhai et al., 2021; Ojo et al., 2024) and finally, in environmental sciences like geography (Purves, Koblet, and Adams, 2022), it helps understand spatial and environmental challenges, for instance, national parks face (Hausmann et al., 2020).

Understanding geography through text analysis combines the detailed world of geography with the insights of language studies and computer analysis. It allows us to study how places are described and understood in various texts, from old documents to online content (Purves, Koblet, and Adams, 2022). Computational text analysis has become a valuable tool in geography, offering new ways to explore spatial patterns, place names, environmental discourses, and the cultural significance of landscapes (Purves, Koblet, and Adams, 2022).

For example, Lund, Moncla, and Viehhauser (2022) explored how digital methods can detect what narratives were told about glaciers in different times and places using large text corpora and try to observe changes and shifts in the discourse about glaciers and how they were perceived. It can further be used for landscape character assessment (identifying and describing distinct landscape features for mapping) by analysing the perception of sound, smell and sight, as done by Chesnokova and Purves (2018). Using 7'000 online texts about the perception of the Lake District National Park in England, they successfully demonstrated the feasibility of constructing a broad, diverse collection of firsthand landscape experiences for analysis across multiple senses. Moreover, in the study by Luria and Campos (2022), keyword extraction tools like "YAKE!" (a set of statistical measures to identify and rank keywords within a single document) have been employed to understand evolving environmental narratives in post-industrial cities addressing gentrification.

The use of textual data to analyse the perceptions of surroundings is also prevalent in studies concerning protected areas like national parks (Heikinheimo et al., 2018; Hausmann et al., 2020; Teles da Mota and C. Pickering, 2021; Mangachena and C. M. Pickering, 2021; Sergiacomi et al., 2022). They play a crucial role in conserving biodiversity and supporting the population's well-being (Mangachena and C. M. Pickering, 2021; Hausmann et al., 2020). In his paper, Dilsaver (2009) promotes more integrative research approaches to understand better the complex roles that national parks play in nature conservation. The source of digital media provides fresh opportunities for geographic insights, particularly in resource-scarce fields like nature conservation (Heikinheimo et al., 2018). For instance, Hausmann et al. (2020) collected and analysed Instagram posts geolocated in South African national parks to improve socio-political understanding and support for nature conservation. These posts have been examined for sentiments and emotional content expressed by visitors, categorised as positive, negative, or neutral, and validated through manual annotation. Statistical tests such as ANOVA and Tukey post hoc tests compared sentiment across different parks to discern significant emotional differences, thus solidifying the importance of the results. This comprehensive approach offered robust insights into visitors' sentiments and emotions, aiding in better park management and conservation strategies (Hausmann et al., 2020).

A similar goal was pursued by Mangachena and C. M. Pickering (2021), who studied public discourse using tweets about seven South African national parks. The content and sentiment of the tweets were analysed based on word frequencies and categorised into topics such as biodiversity, tourism, natural attractions, and crime and safety. The NRC Word-Emotion Association Lexicon was used for the sentiment analysis to categorise the posts. The aim was to assess what topics were discussed for which parks, by whom, when, what emotions were expressed, and how such information could be useful for management. Studies of social media reviews discussing visitor experiences and sentiments about national parks are popular (Mangachena and C. M. Pickering, 2021).

Similarly, Sergiacomi et al. (2022) investigated TripAdvisor reviews about the Plitvice Lakes National Park in Croatia. They analysed visitors ' opinions using sentiment analysis and natural language processing approaches. The study aimed to understand the strengths and weaknesses of the park as perceived by the visitors based on sentiment scores and the frequency of specific terms. Specifically, in this study, the Rapid Automatic Keyword Extraction (RAKE) algorithm was used to isolate important words and phrases that frequently appear together, providing insights into the main topics discussed by the reviewers. The study reinforced the limitations that the dataset of social media users may not accurately reflect the broader population of park visitors and that sentiment analysis tools may struggle with detecting sarcasm and implicit expressions (Mangachena and C. M. Pickering, 2021; Hausmann et al., 2020). Natural Language Processing (NLP) tools can also have difficulties recognising colloquial languages, such as slang and emojis. It seems social media plays a crucial role in text analyses concentrating on national parks. (Mangachena and C. M. Pickering, 2021; Hausmann et al., 2020; Sergiacomi et al., 2022; Heikinheimo et al., 2018). This highlights that text is a rich yet underutilised source

of landscape information. It highlights the importance of creating analytical methods adapted to specific cultures, languages, and landscapes to tap into this potential fully.(Santana et al., 2023; Cambria and White, 2014).

Natural Language Processing (NLP) is fundamental at the core of computational text analysis methodology. NLP started in the middle of the 20th century with early experiments in translating languages and understanding human language by machines (Cambria and White, 2014, p. 49). It now includes complex methods to read, interpret, and even create human language, improving with new technological advances, especially in machine learning and deep learning (Cambria and White, 2014; Hirschberg and Manning, 2015; Dande and Pund#, 2023). There are many techniques and applications to create better algorithms to improve the understanding and analysis of human language to enable a more intuitive and efficient way of communicating between machines and humans (Dande and Pund#, 2023; Cambria and White, 2014; Khanbhai et al., 2021). Yet, even though the increasing digital content leads to NLP being a rapidly growing and resourceful field because of its high potential to gain insight into how people think and talk, the road to building such efficient and reliable systems is still long (Purves, Koblet, and Adams, 2022, p. 56). The reason is linguistic ambiguity and diversity, which means dealing with synonyms, idioms and context-dependent meanings (Dande and Pund#, 2023). A machine needs to "see beyond what is visible" to interpret the text as humans do (Cambria and White, 2014, p. 56). Moreover, machine and deep learning rely on complex and massive amounts of data, computational resources and energy (Purves, Koblet, and Adams, 2022, p. 56).

There are established methods and models that are simpler and still powerful for general text analysis. The most straightforward is the bag-of-words (BOW) model, which merely counts the frequency of (specific) words and indicates their importance in the document's content yet disregards grammar and word order. Nevertheless, it remains a useful model where improved applications continue to be explored and applied, including those that involve the semantic level (Lei Wu, Hoi, and Nenghai Yu, 2010; Zhang and Mayo, 2010). Taking a step further and integrating the approximate context is concordance and co-occurrence, which examine individual words within a certain distance to a node (term of interest, e.g. a place name) and look for meaningful combinations called collocates. Meaningful combinations imply that the pairs co-occur more than by random chance; this can be measured with statistical measures (Purves, Koblet, and Adams, 2022, pp. 57–61). For example, the term frequency-inverse document frequency (TF-IDF) measures the importance of a word within a document relative to a corpus, where "TF" represents the term frequency calculated by dividing the count of a word by the total number of words in the document. Conversely, "IDF" reflects how common or rare the word is across all documents, thus highlighting a word's uniqueness across the corpus. The product of TF and IDF scores balances a word's document frequency against its prevalence across all documents, assigning higher values to words that are uniquely significant to particular documents (Havrlant and Kreinovich, 2017; Purves, Koblet, and Adams, 2022, p. 59). Furthermore, the t-score compares the observed frequency of a word pair (or phrase) against the frequency expected by chance. It considers the variance of word frequencies and provides

a standardised way to assess the significance of word co-occurrences or mutual information (MI) that quantifies the strength of association between two words or terms (Brezina, 2018; Purves, Koblet, and Adams, 2022, p. 61). The core idea is that if two words appear together in the text more often than expected by chance, they might share a significant semantic or topical relationship. A fourth method is an unsupervised or supervised classification, where the text gets categorised in predefined classes, like, for example, the thematic fields used in this thesis. Supervised methods require annotated training data, whereas, on the other hand, unsupervised methods do not (Purves, Koblet, and Adams, 2022, p. 77). Yet, unsupervised typically come with subjectivity issues and a high investment of time and money and have lower accuracy than supervised methods (Papagiannopoulou and Tsoumakas, 2020).

Precision metrics like “Precision”, “Recall,” and “F1 score” are used to evaluate the performance of the classification process. This approach is useful for categorising large amounts of data. For example, the study of Jeon et al. (2024) successfully categorised complaints about building defects using unsupervised classification methods, proving that even though unsupervised methods normally perform lower, they can be a useful approach as they require less data preparation and computing resources. Another more complex method is keyword extraction for context, where a set of extracted representative phrases should briefly summarise the content of a document. Papagiannopoulou and Tsoumakas (2020) compiled and reviewed the most common unsupervised and supervised approaches for keyword extraction, as well as their advantages and challenges. Among the unsupervised methods, statistical methods work the best, especially with longer texts (Papagiannopoulou and Tsoumakas, 2020, p. 36). Similar to the limitations faced with digital data, printed texts can also suffer from sampling bias (Panzeri, Magri, and Carraro, 2008). The available texts might only represent specific regions, time periods, or demographic groups. Furthermore, analysing printed text presents challenges, including interpreting historical language, changes in language use over time, and variations in spelling and grammar (Asi et al., 2015). Optical Character Recognition (OCR) technology converts scanned printed text into digital form and can struggle with older fonts or degraded print (C. Patel, A. Patel, and D. Patel, 2012).

Despite recognising text as a potentially rich source of landscape information, its application remains notably under-explored in specific areas, such as within the Swiss National Park research. While research is a cornerstone of the Swiss National Park, with annual scientific publications focusing on habitat analyses of flora and fauna and landscape changes due to human intervention (Swiss National Park Data Center, 2024), the use of text-based sources to chronicle landscape change is yet to be tapped into. This thesis aims to bridge this gap by leveraging textual sources to identify possible changes in the frequency and thematic context of place names within the park over time. This endeavour seeks to employ already established text analysis methodologies to recognise the thematic contexts of these place names.

3. Background

3.1 The Swiss National Park

3.1.1 Overview

The Swiss National Park is located in the canton of Grisons and borders directly with Italy and Austria. The park is in the Alps and accessible from the Lower and Upper Engadin, Val Müstair, and Livigno (IT). The mountainous area, covering 170.3 km², is situated in the Romansh-speaking region of Switzerland, and in the local language, the park is correctly called Parc Naziunal Svizzer (Baur and Scheurer, 2014, pp. 14–15). The SNP lies within the territories of the municipalities of Zernez (68.6%), S-chanf (13.5%), Scuol (13.2%), and Val Müstair (4.7%) (Schweizerischer Nationalpark, 2024a). Hence, the municipality of Zernez, with two-thirds of the park's area, is by far the most important park municipality. According to the International Union for Conservation of Nature (IUCN), the park belongs to category 1a, meaning the reserve is declared in the highest protection class and wilderness area (Schweizerischer Nationalpark, 2024a).

The park consists of two sections - the central area (167 km²) near the lower Engadine Dolomites and the unexplored Macun area in the upper part of the SNP. The mass elevation is of interest to the Alpine region; the average altitude is 2323 m a.s.l., with absolute height differences ranging from 1400 m (Clemgia/Scuol) to 3174 m a.s.l. (Piz Pisoc) (H. Haller, Eisenhut, and R. Haller, 2013, p. 60). These differences are also reflected in the ground cover, consisting of forest (28%), alpine meadows/grasslands (21%), where most Alpine plants grow, and 51% raw soils, meaning scree and rock (Schweizerischer Nationalpark, 2024a). Particularly in higher elevations and extreme locations, vegetation-free zones can be found. Up to 2000 m a.s.l., forests dominate, and above this altitude, grasslands are widespread in elevations of 2100 to 2500 m a.s.l. Between 2000 and 2300 m a.s.l., the different vegetation types combine and intermix. Overall, the climate can be described as dry, rough and intense sun exposition (Schweizerischer Nationalpark, 2024a).

The SNP is not entirely untouched by human influence. Before the establishment of the Swiss National Park, the landscape was quite different from what visitors see today. The area was primarily used for agriculture and grazing. Human activities, including logging and livestock grazing, managed and altered the landscape. This utilisation significantly impacted the local flora and fauna, with some species being driven away or reduced in numbers due to habitat alteration (H. Haller, Eisenhut, and R. Haller, 2013, pp. 48, 50). Today, a 100 km long marked

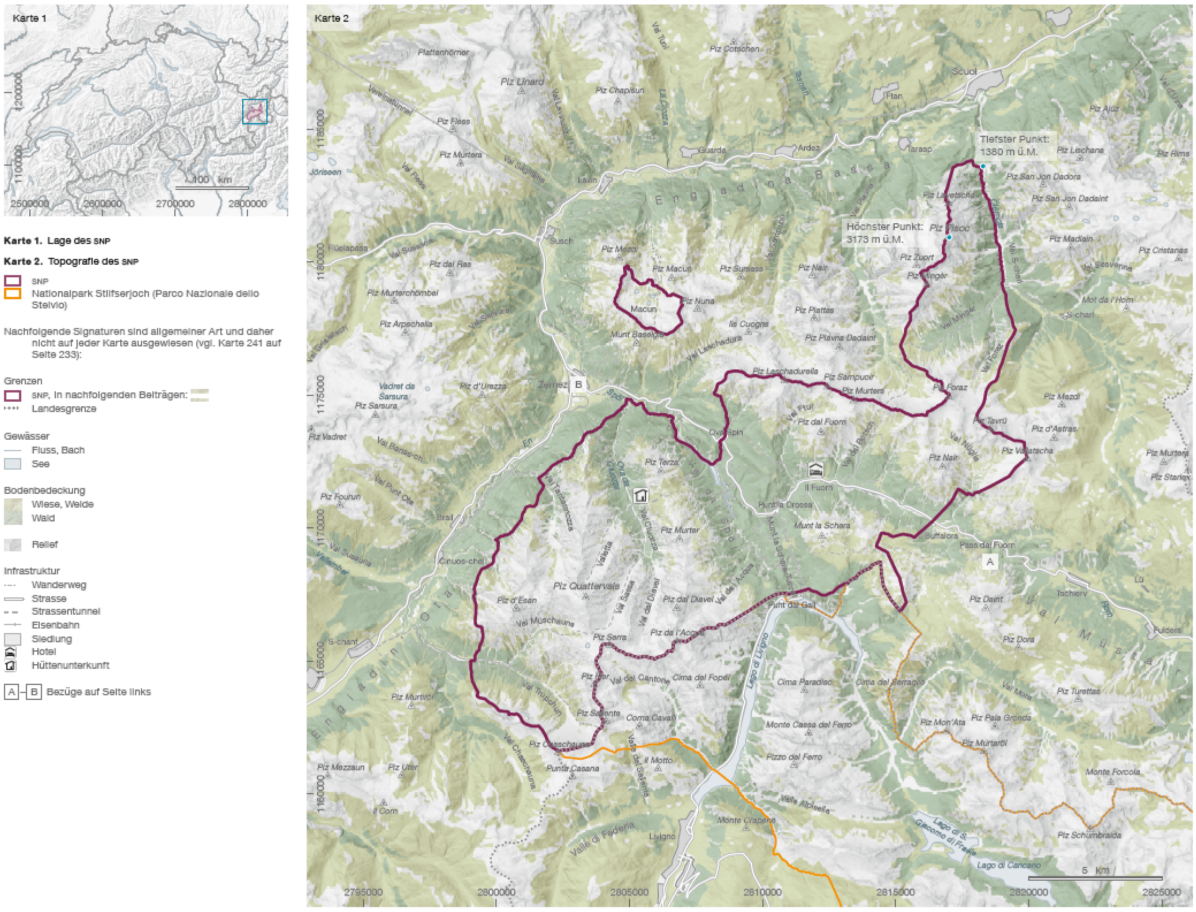


Figure 3.1. Location and of the Swiss National Park (Canton Graubünden) in the southeastern part of Switzerland (Source: (H. Haller, Eisenhut, and R. Haller, 2013, p. 15))

trail network and 15 buildings, including the only publicly accessible hut, Chamanna Cluozza in the Val Cluozza, and the laboratory house at Il Fuorn, human infrastructures can be found. The customs office Punt la Dross and the Hotel Parc Naziunal Il Fuorn are on the relatively busy Ofenpass road. Along the Ofenpass road, there are seven parking spots available. The SNP counts ca. 120’000 visitors annually, considering the park is closed to the public in winter (H. Haller, Eisenhut, and R. Haller, 2013; Schweizerischer Nationalpark, 2024a). Generally, there has been a linear increase in human presence throughout the park’s history, for instance, by incorporating one of the leading water inlets of the Spöl into the Engadin Power Stations (EKW) hydraulic system in the 1960s (H. Haller, Eisenhut, and R. Haller, 2013, p. 202). Also, the official opening of the Munt la Schera road tunnel in 1970 facilitated private traffic through the SNP (H. Haller, Eisenhut, and R. Haller, 2013; Schweizerischer Nationalpark, 2024a). Nonetheless, the SNP is considered a place of natural beauty (Engadin Tourismus AG, 2024; Tourismus Engadin Scuol Samnaun Val Müstair AG, 2024b; WegWandern.ch, 2019).

3.1.2 History

Note to readers: Throughout this chapter, places of interest relevant to this thesis are highlighted in *italics*.

At the beginning of the 20th century, the urge to protect and save native flora and fauna spread across the Western world (Kupper, 2012, p. 11). Industrialisation and technological progress only intensified the longing for pristine nature (Kupper, 2012, p. 40). As early as 1904, National Councillor Fritz Bühlmann proposed the idea of a large protected area (Schweizerischer Nationalpark, 2024a). The Swiss conservationists and later national park founders Fritz and Paul Sarasin, Carl Schröter, and Steivan Brunies (later the first park director of the SNP) chose the *Ofenpass* area for a good reason (H. Haller, Eisenhut, and R. Haller, 2013, p. 58). Its remoteness and landscape features met all the criteria for wild and original nature. The Val *S-charl* and the magnificent Swiss pine forest God *Tamangur* in that valley further reinforced their choice (H. Haller, Eisenhut, and R. Haller, 2013, p. 58). Brunies eventually suggested the *Val Cluozza*. The goal for the following decades was to connect these two areas. Although the Swiss pine forest *Tamangur* was designated a nature reserve in 2007, after its exclusion in 1935, *Tamangur* never became part of the SNP again, like *S-charl*. The first contracts were signed between the municipalities of *Zernez*, S-chanf, Scuol, and the Alpine cooperative Tavrü with the Conservation Commission of the Swiss Society of Natural Sciences (today's Swiss Academy of Sciences) in 1909 and 1910 (H. Haller, Eisenhut, and R. Haller, 2013, p. 58). These contracts established a leasing relationship for the next 25 years for the *Val Cluozza*. The project was financed with the help of the Swiss League for the Protection of Nature (today Pro Natura) (Schweizerischer Nationalpark, 2024a). In 1920, the Falcun area was allocated to the SNP by the municipality of *Zernez*, with the assurance to the municipality that potential hydraulic use of the *Spöl* would be permitted (H. Haller, Eisenhut, and R. Haller, 2013, p. 58). This decision was regretted in the years that followed. In 1957, the Swiss population finally approved the water use of the *Spöl* by the Engadine power plants within the national park area. The ecological system was sustainably changed, and human influence in the park increased. Another setback was the exclusion of the Val Tavrü in 1936 when the parties could not agree on a lease extension. However, the *Val Mingèr* and the Val Foraz were re-incorporated into the park. Not only did hydropower projects bring change to the SNP after World War II, but the park also became a popular destination, further increasing tensions between conservation, science, and tourism (H. Haller, Eisenhut, and R. Haller, 2013, p. 58). With the opening of the National Park Centre in Zernez in 1968, Zernez became the gateway to the national park and a tourist attraction. Further inaugurations for the National Park Centre in *Zernez* followed in 2008 and 2023, providing new momentum for public relations work (Schweizerischer Nationalpark, 2024d). The National Park Act of 1980 and the cantonal regulation defined the SNP's nature, purpose, ownership, tasks, and regulations, thus establishing a clear legal basis (Schweizerischer Nationalpark, 2024a). In the 1990s, the park authorities launched an expansion project. Of the 300 km² buffer zone, only the 3.6 km² *Macun* Lake plateau remained after the voters in *Zernez* rejected

it in 2000 (Schweizerischer Nationalpark, 2024a). Since the turn of the century, the Val Müstair and the SNP have been moving closer together, and in 2010, they jointly formed the UNESCO Biosphere Reserve (H. Haller, Eisenhut, and R. Haller, 2013, p. 58).

3.2 History of Geographical Names in Switzerland

A place name, or toponym, is a collective term for names of landscapes, settlements, bodies of water, mountains, and field names (in German: "Flurname") (Bundesamt für Landestopografie swisstopo, 2024). The significance of place names has been researched in Switzerland since the 19th century, as they carry historical information (Bicke, 2016). Toponyms witness the history of settlement, can demonstrate changes in vocabulary and language usage, and provide information about the composition and topographic characteristics of soils and locations in the case of field names (Bicke, 2016, p. 148). Place names can endure major historical upheavals and language shifts, as exemplified by "Aare" or "Rhine," which point to the pre-Roman era, or Zurich, which was known as "Turicum" in the same era (Bicke, 2016, p. 148). Landolt (2016) shed light on the rich history and challenges of spelling place names in Switzerland, particularly in accommodating local dialects versus standard German spelling. Over the years, various guidelines have been issued that influenced the spelling of place names. It was Albert Bachmann, a professor at the University of Zurich, who in 1916 suggested recording names not according to the local customary writing but according to the local customary speaking, for example, "Egg" instead of "Eck." This initiative was rejected by the federal Department of Justice in 1926 (Landolt, 2016, p. 140). Guntram Saladin tried, only a decade later, to push through the same project as his predecessor. It was not until 1948 that the Federal Department of Justice and Police issued clear directives for the recording and spelling of local names, which were broadly in line with Saladin's wishes, thus a "moderate" rendition of dialect pronunciation, including "Egg" instead of "Eck," although "Wäg" should be spelt as "Weg" (Landolt, 2016, p. 140). Field names were recorded in a standard Swiss German pronunciation that reflected the linguistic conditions of the Central Plateau in Switzerland (Landolt, 2016, p. 141). Although these regulations were seen as a compromise to preserve dialectal characteristics while standardising geographical place names for maps and official documents, the discourse on the correct spelling continues until this day (Landolt, 2016). Nowadays, cantonal authorities, together with the cantonal nomenclature commissions and the municipalities, determine the spelling (Bundesamt für Landestopografie swisstopo, 2024). One of the peculiarities of geographical names in Switzerland is multilingualism, as some objects have multiple names, such as "Ofenpass" and "Pass dal Fuorn". Also, some topographical objects are associated with a numerical indication, especially names concerning peaks, passes, or hills relating to altitude. The latter is the field names, which, especially for these names derived from forests, pastures, meadows, and fields, or carrying historical roots, often appear in the dialectical writing rather than in the standard language, such as "Charbunera" (= charcoal production) or "Murter" (= mortar, indicates alpine pastures) (Schorta, 1991). Precisely because field names reflect the circumstances of the place, a field name can be registered multiple times in Switzerland (Bundesamt für Landestopografie

swisstopo, 2024). For example for "Murter" alone there are 28 matches all in the canton Grisons (Schweizerisches Idiotikon, 2024). Local names often emerge when people seek to identify a place without relying on geographical coordinates (H. Haller, Eisenhut, and R. Haller, 2013, p. 56). This is also true for the SNP, where, over the years, park rangers have created field names that are now included on maps (H. Haller, Eisenhut, and R. Haller, 2013, p. 56).

3.2.1 History and Characteristics of the Places of Interest

The Swiss National Park (SNP) history has already given information about a few locations treated in this thesis. In this subsection, we will delve into a more detailed description and function of the most relevant locations, which will also be examined in the methods and analysis section. Figure 3.2 provides a clear overview of the 27 locations and where and how these places are situated and extend within the SNP. The spread of the locations already shows apparent differences. Are we dealing with a larger area or space, such as a valley, an alpine pasture, a body of water like *Spöl*, *Ova Spin*, or a peak like *Piz d'Esan*? The map clearly shows that place names can frequently represent different geographic locations. For example, the place name Tamangur can refer to six different geographical sites, such as Zondra da Tamangur (pine forest), Foppa da Tamangur (hollow near Tamangur), Murters da Tamangur (high mountain pasture), Tamangur Dadora (Outer Tamangur), Tamangur Dadaint (Inner Tamangur), or God Tamangur (forest or "The forest back there") (Tourismus Engadin Scuol Samnaun Val Müstair AG, 2024a). As mentioned, the same field name can be registered multiple times in Switzerland, as with Murtaröl, which is the diminutive form of Murter, meaning sheep pasture (Schorta, 1991). It appears three times in the SNP and Biosphere Reserve alone and 19 times across Switzerland (Schweizerisches Idiotikon, 2024). This should illustrate that when analysing place names, especially field names, it should be noted that they may not refer to a unique location. How precisely these places are written about in the corpora depends not only on the type of text source but is also influenced by their geographical location and surroundings. This is especially relevant for the thematic part, where the contextual embedding of the place names is analysed. For this reason, the locations' most important events and key dates and their surrounding vegetation, altitude, and the extent of the most relevant locations are described. At the end of each place description, hypotheses are presented on how the locations might co-occur with specific themes in the results. Assumptions are consistently made regarding the "Causes of Landscape Change", "Primary Themes/Habitat", and "Animals". Of the 27 location terms or 24 places (Ofenpass = Pass dal Fuorn, Stabelchod = Val da Stabelchod, Il Fuorn = Ofenberg), only the history and characteristics of seven places—Buffalora, Il Fuorn/Ofenberg, Macun, Ofenpass/Pass dal Fuorn, S-charl, Spöl, and Trupchun—are highlighted, as these are examined in greater detail in the individual corpora. Descriptions of the remaining places can be found in alphabetical order in Appendix A.

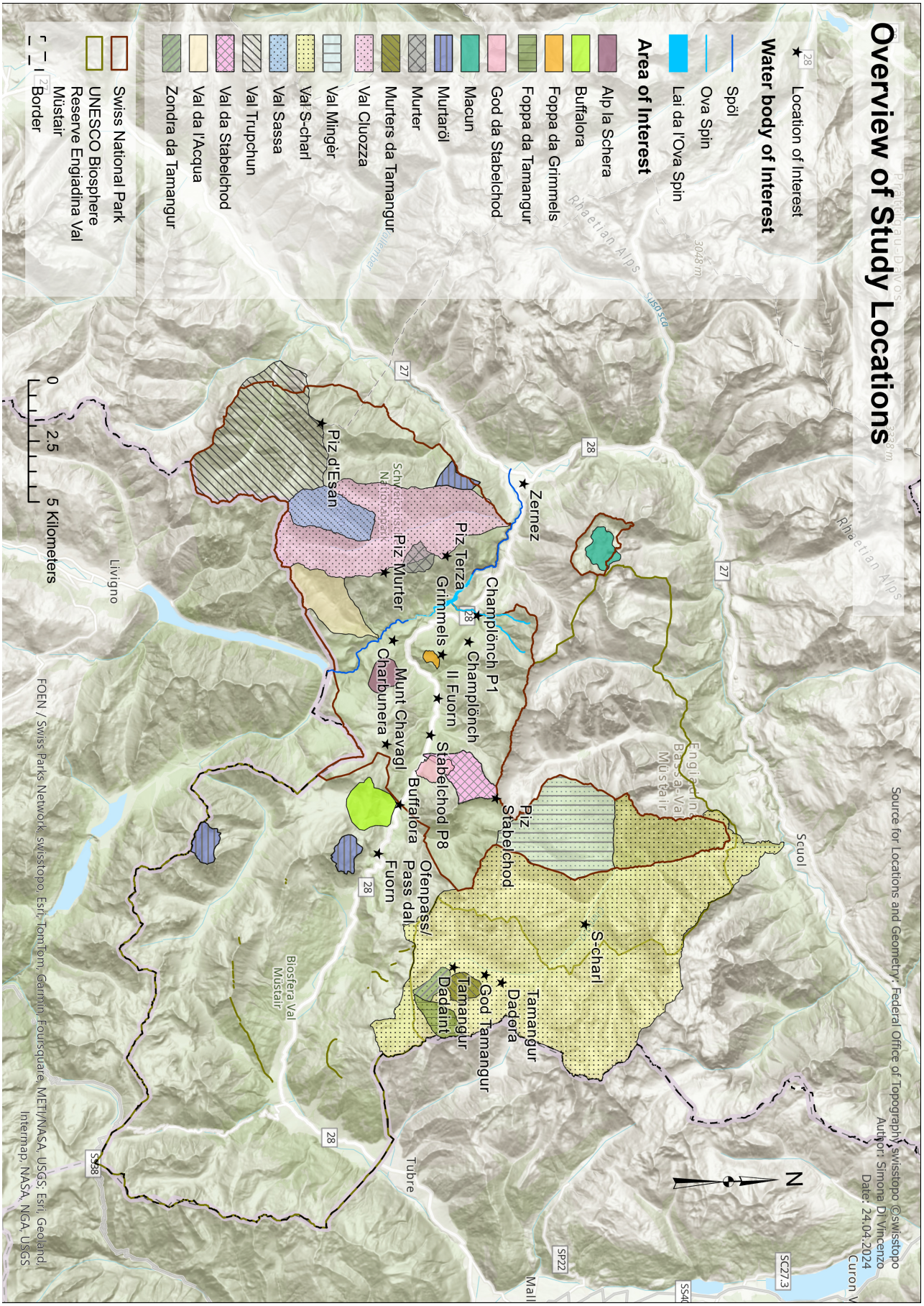


Figure 3.2. Map Overview of Study Locations.

Buffalora

Buffalora and Alp Buffalora are located just before the southeastern border of the National Park and, due to the good accessibility via the Ofenpass road, are popular destinations. An important contribution to monitoring is made by the meteorological station in Buffalora (1968 m a.s.l.), which has recorded an increase in the average temperature of 1.92°C from 1917 to 2022 (Schweizerischer Nationalpark, 2024e; H. Haller, Eisenhut, and R. Haller, 2013, p. 214). The largest increase is observed in spring and the smallest in autumn (Schweizerischer Nationalpark, 2024e). An increase in extreme weather events due to climate change can also be expected, as demonstrated by the debris flow event on July 25, 2022, at Piz Nair above Buffalora, when a heavy rain event triggered a debris flow. The age of the trees in the affected area suggests that such an event has not occurred in this location in the last 130 years (Schweizerischer Nationalpark, 2024f). The valley location and dryness generally characterise the climate in Buffalora. Alp Buffalora is part of the Val Müstair Parc Naziunal Biosphere Reserve (Baur and Scheurer, 2014, p. 89).
I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Anthropogenic Climate Change, Park Management and Research & Monitoring
- Primary Themes/Habitat: Infrastructure
- Animals: Low co-occurrence

Il Fuorn - Ofenberg

Il Fuorn, or Ofenberg in German, is located directly on the Ofenpass road (Pass dal Fuorn). The name Il Fuorn translates to "the furnace," more precisely a smelting furnace, a remnant from the 14th to 17th centuries when mining and iron smelting thrived in this area (H. Haller, Eisenhut, and R. Haller, 2013, p. 50). The historic building from 1489 is today the Hotel Parc Naziunal "Il Fuorn" and is one of the pieces of infrastructure within the park (Hotel Parc Naziunal Il Fuorn, 2024). Since 1947, laboratory space for researchers has been located in the lower area of the hotel, serving as a workspace and overnight accommodation (Kupper, 2012, p. 216). In 1951, a forest fire occurred near Il Fuorn during clearing work after an avalanche incident. Over the following decades, the area, once forested with mountain pines, was transformed into a grazing land by red deer and chamois. Today, predominantly pioneer plants and pasture grasses can be found (H. Haller, Eisenhut, and R. Haller, 2013, p. 198). Since 1991, bearded vultures have been released near Il Fuorn, and the site has become an important hotspot for visitors (Baur and Scheurer, 2014, p. 309). In the 1990s, red deer also began to resettle in the area, making it possible to observe the deer rut every autumn (H. Haller, Eisenhut, and R. Haller, 2013, p. 150). Il Fuorn serves as a tourist meeting point and a research location, and the mountain pine and pasture landscape allows for the admiration and study of bogs, plants, and insects.

I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring, Tourism & Guests
- Primary Themes/Habitat: Infrastructure, Grassland, Forest
- Animals: average co-occurrence

Macun

The 3.6 km² Macun Lake plateau is situated 2600 m a.s.l. and comprises 23 small and large mountain lakes. In Romansh, Macun means "ibex", which by chance can be observed alongside chamois (Schweiz Tourismus, 2024). As previously mentioned, Macun was integrated into the SNP territory only in 2000 (Baur and Scheurer, 2014, p. 270). Sparse vegetation, grass, lakes, and scree characterise the landscape. Its high-alpine lakes and ponds make it an important area for aquatic research (Schatz, Scheurer, and Steiner, 2012; H. Haller, Eisenhut, and R. Haller, 2013, pp. 204, 205). Over 200 species of diatoms have been found in the flowing waters and lakes (Baur and Scheurer, 2014, p. 277). Climate changes, such as ice melting and longer dry periods, have also been observed in Macun (Baur and Scheurer, 2014, p. 281). Between 1968 and 1991, waters originally devoid of fish were stocked with brown trout and non-native fish species by stocking measures and anglers (Baur and Scheurer, 2014, p. 282). These have managed to sustain themselves to this day; today, in addition to brown trout, Canadian lake char and minnows can be observed (Baur and Scheurer, 2014, p. 282). Since the inclusion of Macun, hydrological examinations are done annually (Baur and Scheurer, 2014, p. 381).
I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Anthropogenic Climate Change, Natural Processes, Park Management, Research & Monitoring
- Primary Themes/Habitat: Extreme Locations, Water Bodies, Infrastructure
- Animals: high co-occurrence (especially aquatic creatures)

Ofenpass - Pass dal Fuorn

The 2149 m a.s.l. high Ofenpass - Pass dal Fuorn in Romansh is one of the main areas of the SNP. The Ofenpass road runs through it, connecting Zernez with Livigno and the Val Müstair. The Ofenpass road was built in 1871 for alpine farming and iron ore transport, leading to large-scale logging (Baur and Scheurer, 2014, p. 75). The road was widened again at the end of the 1960s and is still the biggest source of noise in the SNP today, as there are seven rest areas, the Hotel Il Fuorn and the Punt La Drossa customs office along the road. Passenger car traffic has been possible on it since 1925 (H. Haller, Eisenhut, and R. Haller, 2013, p. 172). The road forms an aisle and is flanked by a tree-free embankment. Research has been conducted on these embankments since 1969 (H. Haller, Eisenhut, and R. Haller, 2013, p. 176). The snow heights allow water sources to reach the Ofenpass area. Although Pass dal Fuorn is the local name, the German spelling is often used in the corpora, except for cartographic representations, where

the Romansh name predominates (Baur and Scheurer, 2014; H. Haller, Eisenhut, and R. Haller, 2013; Kupper, 2012). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Tourism & Guests
- Primary Themes/Habitat: Infrastructure, Forest
- Animals: low co-occurrence

S-charl

The Val S-charl, with the breathtaking Swiss pine forest Tamangur, was, as mentioned, one of the areas explored for the founding of the SNP at the beginning of the 20th century alongside Val Cluozza and was intended to become the core of the SNP (H. Haller, Eisenhut, and R. Haller, 2013, p. 58). Like Il Fuorn, S-charl is one of the places where mining and iron smelting played a significant role from the 14th century onwards (Baur and Scheurer, 2014, p. 68). Among other events, the last bear in Switzerland was shot in S-charl in 1904 municipality (Baur and Scheurer, 2014, p. 79). During negotiations in 1910, the responsible municipality Scuol narrowly agreed to lease the Alp Tavrü (located in Val S-charl) to the SNP (Kupper, 2012, p. 122). The contract lasted until 1935 when no agreement was reached, and thus S-charl and Alp Tavrü were no longer part of the SNP (Kupper, 2012, pp. 122, 128). Today, the former administrative building of the lead and silver works in S-charl houses the Museum Schmelzra, which has been displaying the mining and bear exhibition since 1997 (Schweizerischer Nationalpark, 2024g).
I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Tourism & Guests
- Primary Themes/Habitat: Infrastructure, Forest
- Animals: average co-occurrence

Spöl

The Spöl is a 42 km long river, of which 14 km are on Swiss territory and the rest in Italy (schweizerfluss.ch, 2024). Its source is in the Val dell'Orsera at an elevation of 2640 m a.s.l., flowing through the Val da Spöl in the SNP before finally merging into the Inn at Zernez (schweizerfluss.ch, 2024). The Spöl has a turbulent history: in 1945, the construction of a hydroelectric power station on the Spöl was announced, followed by the construction of the power station in the 1950s, raising fundamental questions in nature conservation (Kupper, 2012, p. 216). Since 1970, the river has been used for hydroelectric power production by the Engadine Power Plants (EKW) (founded in 1954) (H. Haller, Eisenhut, and R. Haller, 2013, p. 60). The river underwent significant changes due to hydroelectric utilisation, accumulating algae and mosses in the riverbed (H. Haller, Eisenhut, and R. Haller, 2013, p. 118). The fish stock was subsequently intensified, even though this was not permitted within the SNP, yet the border of the SNP extended into the Spöl. By 1990, the biological character of the national park's waters had significantly changed (H. Haller, Eisenhut, and R. Haller, 2013, p. 118). The 1993 change in Swiss fish stocking policy, which prohibited the introduction of non-native species,

led to the designation of a conservation stretch in the lower section where fish could reproduce undisturbed (H. Haller, Eisenhut, and R. Haller, 2013, p. 118). Since 1990, attempts have been made to alter the flood level and stage smaller floods artificially, reintroducing the character of a mountain river. Among the inhabitants, such as the brown trout, natural selection resumed (H. Haller, Eisenhut, and R. Haller, 2013, p. 118). Unfortunately, the revitalisation initiated by the success of the dynamic residual water regime suffered a regrettable setback when an environmental accident occurred on the Spöl in March 2013 (Schweizerischer Nationalpark, 2024i). An operational incident caused by low water levels in the reservoir led to an interruption in the residual water supply and mudflows. Thousands of fish died as a result of this incident (Schweizerischer Nationalpark, 2024i).In recent years, there have been increasing discussions regarding the relevance of toxic polychlorinated biphenyls (PCBs) emissions in the Spöl area (Der Bundesrat - Das Portal der Schweizer Regierung, 2021). Since 2020, it is known that PCBs have entered the water through the hydropower plant, and remediation works are now being carried out (Der Bundesrat - Das Portal der Schweizer Regierung, 2021).
I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Rewilding, Park Management, Research & Monitoring
- Primary Themes/Habitat: Infrastructure, Water Bodies
- Animals: high co-occurrence

Trupchun
The area around Trupchun, attracting 30’000 visitors annually, is a true tourism magnet (Baur and Scheurer, 2014, p. 309). One reason for this is the high probability of wildlife sightings. Visitors can admire red deer, ibexes, chamois, marmots, bearded vultures, and golden eagles. Autumn is the ideal season to observe the red deer during their rutting period (H. Haller, Eisenhut, and R. Haller, 2013, p. 74). The area is located in the municipality of S-chanf. The valley’s upper end reaches the Fuorcla Trupchun into Italy. In the past, many research projects on animal and plant behaviour related to terrain forms were conducted here (H. Haller, Eisenhut, and R. Haller, 2013, pp. 44, 70, 74, 116). It was found, for example, that ibexes and chamois prefer slopes of 30-50°, while red deer favour flatter areas (H. Haller, Eisenhut, and R. Haller, 2013, p. 20). Val Trupchun is defined by its dry climate with high-temperature amplitudes and radiation intensities, which, along with the high density of ungulates, leads to challenging conditions for forest health (H. Haller, Eisenhut, and R. Haller, 2013, p. 90). Trampling creates vegetation-free areas, and the browsing of young shoots (browsing pressure) affects the growth rate and mortality of young trees (H. Haller, Eisenhut, and R. Haller, 2013, p. 90). This became known as the "red deer problem" in the 1970s and peaked in the 1990s. It was decided to implement population control to address these forest-wildlife conflicts (H. Haller, Eisenhut, and R. Haller, 2013, p. 74). This was achieved with the traditional Grisons high hunting season in September and a special hunt in late autumn. The special hunt is necessary because the red deer know the regular timing of the high hunting season and are smart enough not to cross this

protective boundary. Therefore, the special hunt was introduced in November and December to control the population (H. Haller, Eisenhut, and R. Haller, 2013, p. 74). Nonetheless, Val Trupchun boasts the highest biodiversity. This area may yield high hits in analysing the thematic fields of flora, fauna, park management, and extreme habitats. Since 1992, a climate station has been located in Trupchun (Baur and Scheurer, 2014, p. 381).
I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring, Tourism & Guests
- Primary Themes/Habitat: Infrastructure, Forest, Grassland, Extreme Locations
- Animals: high co-occurrence (especially ungulates)

4. Methods

Building on the methodologies discussed in the related work chapter, my study intricately combines elements of frequency analysis (similar to Bag of Words for counting occurrences), co-occurrence analysis (for examining relationships between place names and thematic fields), and a form of classification (in categorising nouns into thematic fields). The overarching methodology aligns most closely with co-occurrence analysis because it emphasises the relational aspect between specific place names and thematic fields, enhanced by frequency analysis techniques for quantification. The classification aspect is more about the organisation of my analysis rather than the predictive modelling typically seen in NLP classification tasks.

4.1 Data sources, preparation and processing

Six text corpora compiled and provided by Tamara Estermann (SNP) were carefully utilised for the computational text analysis conducted in this study. These corpora consist of several texts, ranging from books to a series of publications that thematically cover the area of the Swiss National Park. A list of locations (also referenced as place names) of interest was also compiled. It is precisely these 27 locations that are now used to identify and analyse their temporal and thematic context within the corpora. A detailed overview of the key data and metadata of the texts and locations will be explained in this chapter. This step is organised that first, the different document collections resp. corpora are described, followed by an explanation of the applied processing steps. The same approach is then applied to the place names. The methodology will delve into the computational extraction of the place names for the frequency analysis and visualisation and the creation of the thematic fields through classifying the nouns of the corpora. Precision and Recall are further used to evaluate the method.

4.1.1 Document collections

4.1.1.1 Overview

This thesis involves working with six document sources compiled and provided by the SNP. A contract has been established to govern the legal basis for using and spreading these documents. All documents can be released publicly, except for the contents of the annual Business Reports. The six document collections do not constitute a single corpus; each is treated separately. Therefore, in this study, we are working with six corpora, four of which consist of multiple

documents. The following collections resp. corpora will be used:

Haller, H., A. Eisenhut und R. Haller (Hrsg.) (2013): Atlas des Schweizerischen Nationalparks. Die ersten 100 Jahre. Nationalpark-Forschung. Schweiz 99/I. Bern: Haupt Verlag.
The Atlas of the Swiss National Park was published to celebrate its centennial anniversary and has a strong geographical connection to the region. This 247-page report is adorned with numerous high-resolution cartographic illustrations, addressing the changes in the park over the past years, encompassing aspects related to plants, animals, and human impacts, and further detailing the potentials and conflicts encountered. In eight chapters, researchers have gathered the fundamentals, history, flora and fauna, the impact of humans, and the challenges encountered in the last hundred years in the SNP into one comprehensive work. Central to this is the passion for discovery and research and the commitment to making the findings accessible to a broad audience. The Atlas includes numerous maps, illustrations, tables, and diagrams. This detail is also mirrored in the PDF file size of 4.2 GB, even though it comprises just a single document. The abundance of images, maps, and tables poses challenges in text processing with OCR (N. Islam, Z. Islam, and Noor, 2017; Mittal and Garg, 2020; Hamad and Kaya, 2016).

1st and 4th Edition (1914/1948) of "Der Schweizerische Nationalpark" by Steivan Brunies.
Steivan Brunies wrote both editions. The fourth edition also includes contributions from Dr. Prof. Hans Boesch and Dr. R. Billwiller. The Engadine native was born in the municipality of S-chanf and remained closely connected to the region into adulthood (Kupper, 2012, p. 110). He was not only one of the co-founders of the SNP but also served as the first director from 1914 to 1941, as well as the secretary of the newly founded Swiss League for the Protection of Nature (today known as Pro Natura) between 1909 and 1935 (Baur and Scheurer, 2014, p. 134). The editions follow a very similar structural framework. They cover the history of the park's establishment, spatial description, geological characteristics, climate, and flora and fauna. The editions differ only in their publication years, with the fourth edition likely providing a better representation of the park's changes. The book's length are 254 and 292 pages, respectively. The books are highly text-based, with occasional illustrations, tables, and black-and-white images included. OCR will likely require a subsequent manual revision since these are scanned PDFs, and the first edition is printed in Fraktur German. Issues may also arise from lengthy footnotes interrupting the text flow, page numbers, and image captions, which are common features in all corpora. This and the document format could also contribute to errors in OCR results.

Cratschla: Informationen aus dem Schweizerischen Nationalpark. Published by: Federal National Park Commission (ENPK) and SCNAT Research Commission of the SNP.
The Cratschla magazine covers fascinating processes currently underway in the Park. It provides insights into the results of ongoing and completed research projects and showcases the people working in park administration. It serves as a medium that facilitates transparency

among researchers, park management, and visitors. The new park director at that time, Klaus Robin, founded Cratschla (source: Tamara Estermann). Since 1992, the National Park Journal of Cratschla has been published twice a year, providing information and reports on developments within the National Park. Each issue often focuses on a specific theme, ranging from a particular plant or animal species to climatic changes. The older publications (until 1997) are more text-based and have a longer length of around 70 pages compared to the other publications, which consist of more images, bolder designs, and graphics. Both types can pose challenges for the OCR method, as the older versions involve scanned pages, and the newer ones may complicate text recognition due to their graphical and design features.

Business Reports of Federal National Park Commission (ENPK).
The annual Business Reports documenting the developments of the SNP have been published since 1915. Currently, Heidi Hanselmann, president of the ENPK, and Ruedi Haller, director of the ENPK, are responsible for these publications. The structure of the annual reports has changed over the decades. Initially, they were relatively short, spanning 5-10 pages. By the end of the 70s, the length of the reports gradually increased, with texts between 40-90 pages becoming common from the turn of the century. The largest report is from 2020, consisting of 104 pages. Naturally, the structure and layout evolved as well. The early decades typically feature text-heavy content, black-and-white images, and small illustrations. While the structure is mostly similar across reports—usually documenting the commission's activities, visits, the state of the park, research activities, and financial aspects—a clear chapter structure with tables of contents only appeared in the reports from the 1980s onwards. Fundamentally, they include personnel, operations, finances, and nature and research documentation. From 2004 onwards, the PDFs were available digitally, whereas before, they were only available in scanned form. All reports are generally text-heavy, but tables can influence text recognition and disrupt text flow.

Nationalpark-Forschung in der Schweiz (Volumes 1- 109). Earlier title until volume 84: Ergebnisse der wissenschaftlichen Untersuchungen im Schweizerischen Nationalpark.
This corpus is more differentiated and does not follow a clear structure compared to the other five. The common denominator here is research, which takes on a different focus depending on the publication. Furthermore, the PDFs are not divided by year but into chapters up until 2012, as they encompass individual works. This results in 150 documents rather than 102, equating to one document per year from 1918 to 2020. They showcase the diversity of research since 1918 and compile outstanding results regarding the park's development. They are highly scientific reports addressing the evolution of flora and fauna in the park, as well as the history of the SNP, some of which are dissertations and/or book publications, for example, volume 97 "Wildnis schaffen" by Patrick Kupper (2012). Volume 99 is the Atlas, examined as a standalone corpus in this work and, therefore, excluded from this corpus. The texts incorporate illustrations, maps, and tables depending on the publication type. Until the end of the 20th century, the texts

Sources		Atlas des Schweizerischen Nationalparks: Die ersten 100 Jahre	"Der Schweizerische Nationalpark" by Steivan Brunies (1st & 4th edition, 1914/1948)	Cratschla: Informationen aus dem Schweizerischen Nationalpark	Business reports of Federal National Park Commission (ENPK)	Nationalpark-Forschung in der Schweiz (NPF)
Properties	**genre**	informational	nature or environmental literature	popular science	technical reports	Research, nature, science, process documentation
	domain	scientific and cartographic: provides geographic geological, biological as well as historical information about the SNP.	literature	wide range of topics, including current research projects, seasonal updates, and offerings provided by the park.	science, research, park management, progress documentation	science, research, literature
	language	german	german	mostly german (small sections in english, italian, french, romansh)	german, french	german, french
	format	pdf (digitised, includes high resolution maps, pictures)	pdf (scan)	pdf (scan and digitised, includes pictures, graphs)	pdf (scan, digitised)	pdf (many figures and tables, scanned until year 1997)
	size	1 file (247 pp.) 4.20 GB	2 files: 1914 (1st edition): 77.2 MB (254 pp.) 1948 (4th edition): 77 MB (292 pp.)	64 files (1723 pp.) 496 MB	108 file (2610 pp.) 1.05 GB	150 files (about 12'844 pp.) 3.35 GB
Type/Publisher		book by Haller, H., A. Eisenhut & R. Haller (Hrsg.) (2013) 2. edition	book	journal Publisher: Federal National Park Commission (ENPK) and SCNAT Research Commission of the SNP	report Publisher: Federal National Park Commission (ENPK)	book, journals Publisher: SCNAT
Time frame		2013	1914 (1st edition) 1948 (4th edition)	1992 - ongoing (last issue 2023/02) published twice a year	1915 - 2021 (ongoing) published once a year	1918 - 2020 (ongoing) 109 volumes (few years are missing)
Spatial Coverage		approx. area of SNP	approx. area of SNP	approx. area of SNP	approx. area of SNP	approx. area of SNP
Rights / License		Usage with proper citation is permitted	Usage with proper citation is permitted	Usage with proper citation is permitted	private	Usage with proper citation is permitted

Table 4.1. Overview of the corpora and metadata.

were available in scans and not in a digitised form. The length of each document exceeds 100 pages and reaches, in individual cases, over 400 pages, making this corpus the most extensive collection in this study. The heterogeneity of the documents and the difficulties mentioned earlier can complicate the OCR process. Moving forward, this corpus will be referred to as the NPF publications.

Table 4.1 lists more detailed information about the collections and their characteristics, units, and copyright issues. All documents are available either scanned or in digital PDFs. I differentiate between scanned and digital PDFs because the scanned documents consist of images, and the OCR functionality is somewhat limited (e.g., the search function for locations), and digital PDFs, where OCR works very well. However, the pages' layout can disrupt the flow of reading the text (tables, graphics, images) (C. Patel, A. Patel, and D. Patel, 2012; N. Islam, Z. Islam, and Noor, 2017; Mittal and Garg, 2020). The use of the search function is also possible with scanned, newer PDFs, yet the OCR function in Adobe is already limited to complete text extraction. The Brunies' books were scanned with a special scanner because a conventional one could damage the old books. The collection consists of **325 files with 17'970 pages** that needed to be processed.

4.1.1.2 Time coverage of the corpora

A clear chronological overview of the periods covered by each corpus is provided in Figure 4.1. It illustrates the chronological timeline of publications and their overlaps. Both editions of Steivan Brunies books (1918 and 1948), and the Atlas (2013) represent individual book publications, but it can be assumed that the temporal scope addressed in these books is naturally broader. As the title of the Atlas suggests, "Atlas of the Swiss National Park. The First 100 Years", it describes events of the past century. In the fourth edition of the book "The Swiss National Park" by Steivan Brunies, it can also be assumed that the preceding years up to 1948 are mentioned. Business Reports and NPF publications cover the most extensive period and are therefore highly promising for the analysis. The arrows indicate the still ongoing publication of the corpora.

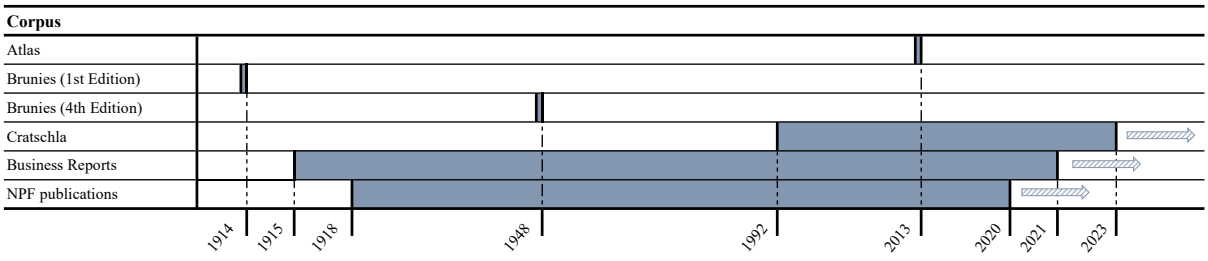


Figure 4.1. Overview of temporal coverage of the documents publications

4.1.1.3 Optical Character Recognition OCR

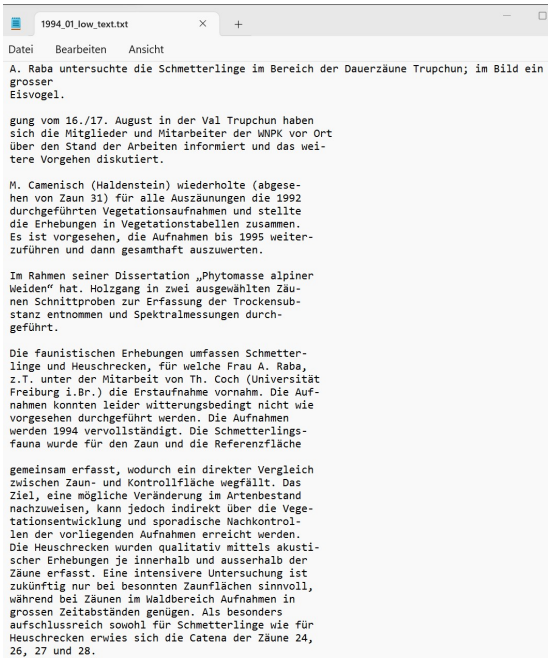
It is crucial to convert the texts into readable and processable formats for the analysis. The documents are available in PDF format, with newer publications (mostly from the 2000s onward) available in digitised form and earlier ones as simple scans from publications and books. The goal is to convert images into text files, meaning PDFs into .txt formats. For this, Python employed pytesseract, an open-source optical character recognition (OCR) tool. Pytesseract is a Tesseract software tool developed and mantained by Google. It provides over 130 languages and is freely accessible, proving an efficient choice for this step (Luria and Campos, 2022; C. Patel, A. Patel, and D. Patel, 2012). Individual PDF files were loaded, and the text was extracted and finally saved in text files. It recognised the German Fraktur language used in Steivan Brunies' book (1914). In the end, all documents were read and saved as text files, meaning there was a corresponding text file for each PDF document in each corpus. The conversion process and the output can be observed in Figure 4.2.

A. Raba untersuchte die Schmetterlinge im Bereich der Dauerzäune Trupchun; im Bild ein grosser Eisvogel.

gung vom 16./17. August in der Val Trupchun haben sich die Mitglieder und Mitarbeiter der WNPK vor Ort über den Stand der Arbeiten informiert und das weitere Vorgehen diskutiert.
M. Camenisch (Haldenstein) wiederholte (abgesehen von Zaun 31) für alle Auszäunungen die 1992 durchgeführten *Vegetationsaufnahmen* und stellte die Erhebungen in Vegetationstabellen zusammen. Es ist vorgesehen, die Aufnahmen bis 1995 weiterzuführen und dann gesamthaft auszuwerten.
Im Rahmen seiner Dissertation „Phytomasse alpiner Weiden" hat O. Holzgang in zwei ausgewählten Zäunen Schnittproben zur Erfassung der Trockensubstanz entnommen und Spektralmessungen durchgeführt.
Die *faunistischen Erhebungen* umfassen Schmetterlinge und Heuschrecken, für welche Frau A. Raba, z.T. unter der Mitarbeit von Th. Coch (Universität Freiburg i.Br.) die Erstaufnahme vornahm. Die Aufnahmen konnten leider witterungsbedingt nicht wie vorgesehen durchgeführt werden. Die Aufnahmen werden 1994 vervollständigt. Die Schmetterlingsfauna wurde für den Zaun und die Referenzfläche gemeinsam erfasst, wodurch ein direkter Vergleich zwischen Zaun- und Kontrollfläche wegfällt. Das Ziel, eine mögliche Veränderung im Artenbestand nachzuweisen, kann jedoch indirekt über die Vegetationsentwicklung und sporadische Nachkontrollen der vorliegenden Aufnahmen erreicht werden. Die Heuschrecken wurden qualitativ mittels akustischer Erhebungen je innerhalb und ausserhalb der Zäune erfasst. Eine intensivere Untersuchung ist zukünftig nur bei besonnten Zaunflächen sinnvoll, während bei Zäunen im Waldbereich Aufnahmen in grossen Zeitabständen genügen. Als besonders aufschlussreich sowohl für Schmetterlinge wie für Heuschrecken erwies sich die Catena der Zäune 24, 26, 27 und 28.
A. Raba photographierte zudem mit Ausnahme von Zaun 31 in allen Zäunen die *Vegetationsstrukturen*, welche nach Dichtekategorien ausgewertet werden. Auf der Alp Trupchun installierte M. Menz (Geogr. Inst. Uni Zürich) im Juli 1993 innerhalb des Zaunes 31 eine mobile, automatische Wetterstation, welche in halbstündigem Rhythmus Klimadaten über Wind, Boden- und Lufttemperaturen liefert.

15

(a) Screenshot of the text passage in the scan of Cratschla 1994/1, page 15.



1994_01_low_text.txt

Datei Bearbeiten Ansicht

```
A. Raba untersuchte die Schmetterlinge im Bereich der Dauerzäune Trupchun; im Bild ein
grosser
Eisvogel.

gung vom 16./17. August in der Val Trupchun haben
sich die Mitglieder und Mitarbeiter der WNPK vor Ort
über den Stand der Arbeiten informiert und das wei-
tere Vorgehen diskutiert.

M. Camenisch (Haldenstein) wiederholte (abgese-
hen von Zaun 31) für alle Auszäunungen die 1992
durchgeführten Vegetationsaufnahmen und stellte
die Erhebungen in Vegetationstabellen zusammen.
Es ist vorgesehen, die Aufnahmen bis 1995 weiter-
zuführen und dann gesamthaft auszuwerten.

Im Rahmen seiner Dissertation „Phytomasse alpiner
Weiden" hat. Holzgang in zwei ausgewählten Zäu-
nen Schnittproben zur Erfassung der Trockensub-
stanz entnommen und Spektralmessungen durch-
geführt.

Die faunistischen Erhebungen umfassen Schmetter-
linge und Heuschrecken, für welche Frau A. Raba,
z.T. unter der Mitarbeit von Th. Coch (Universität
Freiburg i.Br.) die Erstaufnahme vornahm. Die Auf-
nahmen konnten leider witterungsbedingt nicht wie
vorgesehen durchgeführt werden. Die Aufnahmen
werden 1994 vervollständigt. Die Schmetterlings-
fauna wurde für den Zaun und die Referenzfläche

gemeinsam erfasst, wodurch ein direkter Vergleich
zwischen Zaun- und Kontrollfläche wegfällt. Das
Ziel, eine mögliche Veränderung im Artenbestand
nachzuweisen, kann jedoch indirekt über die Vege-
tationsentwicklung und sporadische Nachkontrol-
len der vorliegenden Aufnahmen erreicht werden.
Die Heuschrecken wurden qualitativ mittels akusti-
scher Erhebungen je innerhalb und ausserhalb der
Zäune erfasst. Eine intensivere Untersuchung ist
zukünftig nur bei besonnten Zaunflächen sinnvoll,
während bei Zäunen im Waldbereich Aufnahmen in
grossen Zeitabständen genügen. Als besonders
aufschlussreich sowohl für Schmetterlinge wie für
Heuschrecken erwies sich die Catena der Zäune 24,
26, 27 und 28.
```

(b) Screenshot of the Output Textfile of the Cratschla 1994/1, page 15.

Figure 4.2. Example of a satisfying result of OCR.

4.1.1.4 Difficulties with OCR

It is recognised that certain obstacles exist in text recognition through OCR. Low resolution, low contrast, complex page layouts with tables, illustrations, maps, and page breaks can influence text recognition (C. Patel, A. Patel, and D. Patel, 2012; N. Islam, Z. Islam, and Noor, 2017; Mittal and Garg, 2020). This study has particular challenges with complex page layouts, colour designs, pages containing handwritten or italicised text, tables, and figures. A brief inspection reveals a relatively good outcome, with texts well reproduced, as shown in Figure 4.2. However,

there are other cases, as seen in Figure 4.3. This demonstrates that the cross-profile descriptions at the bottom left and the thematic overview at the top right in the Atlas disrupt the narrative flow. This issue is not unique to the Atlas; it is a common problem across all corpora. In the case of scanned PDFs, such as the books by Brunies, there are instances of incorrect word recognition where individual letters were inaccurately identified and substituted with special characters. The difference of "o" the letter and "0" the digit, is a known issue (C. Patel, A. Patel, and D. Patel, 2012, p. 50) While this issue is diminished in digitised versions, unique and elaborate, colourful page layouts in these cases have also impacted the flow of text.

(a) Screenshot of the text passage in the Atlas, p. 18.

```
grossen Aufwölbung. Dem südöstlichen Abtauchen
von diesem Gewölbe folgt eine immer mächtiger wer-
dende kontinentale Einheit (Chazforà-Formation/
Münstertaler Verrucano). Abb. 4 zeigt auch, dass der
Hauptbewegungssinn der kompressiven Phase der
Gebirgsbildung von SO nach NW gerichtet ist und
Überschiebungen und überrollte bzw. ausgewalzte

Minschuns so

3000 — — 3000
2000 —| — 2000
1000 — > | | | 4000
Engadiner Fenster (Penninikum) %, Sesvenna-Kristallin
G |
à.
© 0

Abb. 4. NW-SO-Querprofil durch die Engadiner Dolomiten, die als strukturelles Gewölbe über dem
verschuppten Sesvenna-Kristallin liegen und steil durch die Engadiner Linie begrenzt

sind (TRÜMPY et al. 1997; Legende siehe Seite 19).

18

atlas_v2.indb 18

ERDE
WASSER
LUFT
FLORA
FAUNA
MENSCH

GJ TEKTONIK

Falten entsprechend orientiert sind. Im deformierten
Schichtstapel fallt eine weitere Einheit auf: Die Raibl-
Gruppe. Hier handelt es sich um mergeliges und
gipsführendes, plastisch deformierbares Material,
das gewissermassen zum Schmiermittel der Decken-
bildung geworden ist.
```

(b) Screenshot of the Output Textfile of the Atlas, p. 18.

Figure 4.3. Example of a poor result of OCR.

Given the vast number of documents, manual revision is unrealistic, as it would take several months. Nonetheless, a manual check was performed for the Atlas and Cratschla series, fixing page breaks, text interruptions, and inaccurate text passages. This means the previously described interruptions in the Atlas have been manually removed. However, this meticulous process proved time-consuming and would significantly limit the analysis time frame. Since the aim of this study is the explorative investigation of the thematic context of place names in the corpora and not a method analysis, no revisions were made for the annual reports and NPF publications, allowing more time to analyse thematic fields. As a result, out of the six corpora, two are nearly error-free, while the other three, due to time constraints, are used in their flawed state

for analysis. Before proceeding with the analysis, the output text files undergo post-processing code to handle hyphens, capitalisation, and paragraphs, enabling removal. Further processing steps will be necessary, which will be discussed in more detail in the following chapter.

4.1.1.5 Post-processing of text-files

As evident in Figure 4.2 and Figure 4.3, numerous hyphens and paragraphs separate both sentences and words. Since the goal is to identify place names in the texts and analyse their context, important descriptions in this format may go unnoticed. Therefore, this post-processing phase created two additional versions for each output text file. In the first version, all hyphens, paragraphs, and capitalisation were removed, resulting in uninterrupted text. In the second version, the paragraph structure and capitalisation were retained, as detecting nouns in paragraphs will play a crucial role in creating thematic fields (more on this in subchapter 4.2). Furthermore, commas, colons, semicolons, and parentheses were removed here. In the co-occurrence analysis, which examined in which thematic fields the place names appear within the paragraphs, it was found that nouns located immediately next to one of these special characters were not counted for the co-occurrence. For each document in every corpus, there are now three text file versions, namely:

i) unprocessed text (direct output after applying OCR)

ii) processed text without paragraphs, hyphens, and only lowercase

iii) processed text with paragraphs and capitalised without commas, colons, semicolons, and parentheses.

As mentioned, the corpora Atlas and Cratschla have undergone manual verification, meaning their paragraph structures are more similar to the original text source than the other corpora. Another manual check was conducted on the old edition of Steivan Brunies' book, where Fraktur German failed to recognise some letters. Known errors include "<" or "{" instead of "c", "> " instead of "ck", and "s\ch" instead of "sch". Furthermore, the German sharp "ß" was transformed into the Swiss spelling "ss". As the NPF publications are published in English and French publications, the OCR was applied in the following languages: German, French, English, and Italian. Because the manual editing for Cratschla and the Atlas had been completed before their translation into English, Italian, and French, each of these corpora now contains two sets of text files: one in German and the other in multiple languages.
The document overview explains that NPF publications are divided into chapters or individual works. Such a structure poses difficulties for the temporal component because not every year has the same number of publications, or some years have none. A copy was created where publications from the same year were merged. For simplicity, a copy was also saved for all corpora, where all publications were combined into a single text document.

4.1.2 List of place names

4.1.2.1 Overview

As mentioned earlier, the text analysis will use a list of locations compiled by Tamara Estermann (SNP), Co-Supervisor of this thesis. The list of place names that were ultimately used in the analysis is visible in Table 4.2. Places with different language spellings, like Ofenpass and Pass dal Fuorn, Il Fuorn and Ofenberg, were treated as distinct entities, unlike Val Cluozza/Val Cluoza, where variations were merely in spelling.

List of place names	
Alp la Schera	Piz Terza
Buffalora	Spöl
Champlönch	Stabelchod
Charbunera	Tamangur
Grimmels / Grimels	Trupchun
Il Fuorn / Ofenberg	Val Cluozza / Val Cluoza
Macun	Val da l'Acqua / Val da l'Aqua
Munt Chavagl	Val da Stabelchod
Murtaröl	Val Mingèr
Murter	Val Sassa
Ofenpass / Pass Dal Fuorn	S-charl
Ova Spin	Zernez
Piz d'Esan	

Table 4.2. Locations of interest.

This approach resulted in the examination of 27 specific locations. Given the combination of 6 corpora containing over 300 documents, the potential for complexity is high. To simplify, an effort was made to merge all variant spellings of each location (see Table 4.3). In subchapters 3.2.1 and appendix A, each place's local descriptions and important historical events are included. Furthermore, the map 3.2 should also provide an overview of the geographical names involved, i.e., whether they are towns, settlements, valleys, mountains, or bodies of water.

4.1.2.2 Potential problems with extracting place names

It is crucial to compile different spellings, as a place name might have changed over the last 100 years. For this reason, the expert knowledge of Tamara Estermann was consulted, and the book of Andrea Schorta (1991) "Wie der Berg zu seinem Namen kam", and further literature was used to get more detailed information about the derivation of the place name and possible different spellings. The history and description of the key locations have already been provided in subchapter 3.2.1. Here, the focus is more on how the search for these locations was conducted within the texts and what problems were encountered. Initially, the question was whether to search only for standalone place names without including adjacent word parts. For example,

Alternative writings of the place names											
Alp la Schera	Alp la Schéra	alp la schera	alp la schéra								
Buffalora	Buffalóra	buffalóra	Buffaióra	buffaióra	Buffallóra	buffallóra	buffalora				
Champlönch	champlönch	Champ Löng	champ löng	tchamp-löntch	champlöng	Champlöng	Tchamp-löntch	Tchamp-löntch			
Charbunera	charbunera										
Grimmels	grimels	Grimmels	grimmels								
Il Fuorn	Il Suorn	il fuorn	JIFuorn								
Macun	macun										
Munt Chavagl	munt chavagl										
Murtaröl	murtaröl	murtarol	murtarôl								
Murter	murter	Murtèr	murtèr								
Ofenberg	ofenberg										
Pass dal Fuorn	pass dal fuorn										
Ofenpass	ofenpass	öfenpass									
Ova Spin	ovaspin	ova spin	Ova spin	ova Spin							
Piz d'Esan	Piz d'Ésan	piz d'esan	piz d'ésan	piz d'esen	Piz d'Ésen	Piz d'Esen	piz d'ésen				
Piz Terza	Piz Térza	piz terza	piz térza	piz tértza	Piz Tértza						
Spöl	spöl										
Stabelchod	Stabelchod	Stavelchod	stabelochod	stabelchod	stavelchod						
Tamangur	Tamangúr	Tamangür	tamangúr	tamangur	tamangür						
Trupchun	trupchun	Truphùm	trupchùm	Trupchum	trupchum	Trupchùm	truphùm	Trupchúm	trupchúm		
Val Cluozza	val cluozza	val Cluozza	Val cluozza	Val Cluoza	val cluoza	Val cluoza	val Cluoza	val cluótza	Val Cluótza		
Val da l'Acqua	Val dell'Aqua	Val da l'Acqua	val da l'aqua	val dell'aqua	val da l'acqua						
Val da Stabelchod	val Stabelchod	Val Stabelchód	V. Stabelchod	val da stabelchod	val stabelchod	val stabelchód	v. stabelchod	val stabelehod	val stabelochod		
Val Mingèr	val Minger	Val Mingér	val mingèr	val minger	val mingér						
Val Sassa	val sassa	Val sassa	val Sassa								
S-Charl	Scarl	Sharl	s-charl	s-carl	s-harl						
Zernez	Zernes	Zarnetz	Sarness	Cernetz	Zernetz	zernez	zernes	zarnetz	sarness	cernetz	zernetz

Table 4.3. Overview of alternative writings of the place names.

only "Ofenpass" would be extracted, not "Ofenpassstrasse". This could reduce the hit rate, so a rule that counts adjacent word parts was introduced, provided the place names are precisely written, including alternative spellings. However, this led to place names such as Pass dal Fuorn, Il Fuorn, or Val da Stabelchod and Stabelchod being counted twice. A context rule was then used to solidify these places and counteract this. For Il Fuorn, "Il" must precede it; for Pass dal Fuorn ", Pass dal", and the same goes for Val da Stabelchod with "Val da", "val", or "v.". After testing it on a small subset of the corpora, the context rule worked well. The locations were always correctly assigned. However, this does not guarantee that no errors will occur when applying the formula to all corpora. Another problem is field names like Murter or Murtaröl, which mean high "mountain pastures" or "sheep pastures" and are therefore often used as field names ("Flurnamen") in different regions of Grisons. The fact that the texts focus exclusively on the areas within the SNP is based on the assumption that the corpora are thematically related to this area. Ultimately, locations are counted, extracted, and processed only based on their spelling, and errors can occur, which may affect the final results.

4.2 Thematic fields

4.2.1 Derivation of the thematic fields

The idea here is to detect and visualise the context in which the locations appear in different corpora. Given that the thematic fields were self-created, the approaches to their development were numerous. The thematic fields should emerge from the texts, but determining the number and type of categories proved challenging. Since this work aims to contribute to the book publication on landscape change, scheduled for release in 2026 to mark the 111th anniversary of the SNP, it seemed appropriate to incorporate similar thematic fields to those addressed in the book. The book organises its content into six primary themes: Grasslands, Forests, Water, Extreme Locations, Infrastructure, and Humans. Additionally, it incorporates four categories that impact and contribute to and accelerate landscape transformation: Rewilding, Natural Processes, Anthropogenic Climate Change, and Human Influence. Further, for this thesis, we find the "Animals" in a stand-alone thematic field, as they can not be attributed to either of the descriptions. The initial categorisation process started with these 11 thematic fields. However, it became clear that differentiating between "Humans" and "Human Influence" posed a challenge, leading to their combination. This merger resulted in "Humans" becoming the most extensive thematic field, thus overshadowing the rest. To reduce this discrepancy and to increase transparency, the final step involved splitting the "Humans" category into three categories, namely "Park Management", "Research & Monitoring", and "Tourism & Guests". These changes were made during the categorisation process rather than being decided at the outset, leading to an iterative rather than a linear process. The details of this refinement process will be explored in the subsequent chapter, leading finally to the following **12 thematic fields**, differentiating between "Primary Themes/Habitats" and "Causes for Landscape Change":

Causes for landscape change	Primary themes/ Habitat	Additional field
Anthropogenic climate change	Extreme locations	Animals
Natural processes	Forest	
Rewilding	Grassland	
Park management	Infrastructure	
Research & Monitoring	Water bodies	
Tourism & Guests		

Table 4.4. Thematic fields in their subcategories. The underlined thematic fields depict the human influence on landscape change.

4.2.2 Noun extraction in the corpora

As previously mentioned, the thematic fields were to emerge from the texts themselves. The goal was to create a list of text files for each thematic field containing nouns relevant to the theme. This was achieved by extracting all nouns appearing in the same paragraph as a place

name and transferring them into a "total noun list" (saved as a .txt file). Initially, the work was conducted at the sentence level, but this approach proved too limiting as the context became more apparent at the paragraph level. Yet, both counts were added up, as duplicates were removed later. For this step, I worked with the processed text files (type 3) that still feature capitalised words and paragraph structure. Using a script for the calculation, Table 4.5 was created, showing the total count of nouns on both the paragraph and sentence levels for each corpus. The quantity of nouns in the individual corpora relatively accurately reflects the size of the text collections.

Extracted noun on sentence and paragraph level				
	Sentence		**Paragraph**	
	all	**unique**	**all**	**unique**
Atlas	3’163	1’656	9’681	4’146
Brunies (1914)	1’713	1’139	4’059	2’418
Brunies (1948)	2’542	1’599	5’263	2’938
Cratschla	33’423	10’298	76’894	18’974
Business Reports	56’383	12’227	67’010	14’820
NPF publications	88’211	21’613	135’013	29’590
Σ	185’435	48’532	297’920	72’886
Total of unique				**121’418**
Total without duplicate				**58’065**

Table 4.5. Total noun count per paragraph and phrase per corpora.

Brunies 1914 and 1948 have the lowest values, followed by the Atlas, Cratschla, Business Reports, and the NPF publications. The only surprise is the high number of nouns at the paragraph level in Cratschla, which exceeds that of the Business Reports. This could be attributed to the manual post-processing of the paragraph structure in the Cratschla series. Generally, it can be assumed that some nouns were not detected due to page breaks, tables, and graphics disrupting the paragraph structure (except in the case of the Atlas and Cratschla). 185’435 and 297’920 nouns were extracted across sentences and paragraphs, respectively. The substantial reduction in numbers observed in the "unique" columns reveals considerable repetition among the nouns within the texts. After duplicates were removed, the count of unique nouns notably decreased to only 58’065, which needed to be categorised.

4.2.3 Iterative and manual categorisation

The categorisation of nouns was conducted in an iterative semi-manual process. Throughout the categorisation process, exclusion criteria and modifications were subsequently implemented. Initially, a list of thematically relevant terms and word components for each thematic field was manually compiled, similar to a mind map. Since most of the text is in German, terms and word fragments were compiled mostly in that language. For instance, nouns containing words such

as "wasser", "bach", "fluss", and "see" were attributed to the thematic field "Water bodies" or "arve", "baum" and "rodung" to the thematic field "Forest". These components could appear at the word's beginning, middle, or end. This approach was employed across all thematic fields. An insight into how this was implemented in the code can be seen in Figure 4.4.

```
5
6   # Definieren der Themenfelder und deren Bedingungen
7   themenfelder = {
8       "Gewässer": ["flüsse", "hydro", "forelle", "Kaulquappen", "habitat", "fisch", "ufer", "Rückzug", "gewässer", "frost", "qu
9       "Grasland": ["Vegetation", "pflanze", "geis", "habitat", "biene", "rasen", "Rückzug", "Ochse", "Asung", "äsung", "orchide
10      "Wald": ["Arve", "Vegetation", "rodung", "wurzel", "hirsch", "habitat", "pilz","Reh", "Rückug", "Forst", "wald", "wälder"
11      "Extremstandorte": ["murgang", "rinnen", "grat", "gipfel", "kamm", "gebirge", "schotter", "beben", "ritz", "kurvatur", "h
12      "Mensch": ["Exkursion", "Tourist", "Reise", "Ferien", "info", "rodung", "Lärm", "Müll", "Konsum", "schmutz", "übernacht",
13      "Infrastruktur": ["übernacht", "stall", "Strassen", "Modellierung", "strom", "energie", "werk", "zaun", "Digitalisierung"
14      "Tiere": ["brut", "herde", "Prädatoren", "Gems", "Rind", "Kuh", "kühe", "population", "schmetterlinge", "kitz", "bär", "w
15      "Rewilding": ["revitalisierung", "Vegetation", "population", "Rekultivierung", "Begrünung", "habitat", "Prävention", "übe
16      "Einfluss des Menschen": ["Exkursion", "Schadstoff", "Tourist", "Reise", "Ferien", "info", "rodung", "Lärm", "Müll", "Kon
17      "Anthropogener Klimawandel": ["schmutz", "brand", "mess", "rekord", "energie", "kahlschlag", "Prävention", "emission", "ü
18      "Natürliche Prozesse": ["wetter", "Vegetation", "wolke", "beben", "frier", "witterung", "abfluss", "zyklus", "natürlich",
19  }
```

Figure 4.4. Compilation for the categorisation in the thematic fields.

Terms from the noun list that could not be categorised in the first round were then grouped in a separate list labelled "other". This resulted in the following classification rate visible in the left Table 4.6. The word lists attributed to the thematic fields were iteratively expanded to minimise the "other" list as much as possible. An active review of the "other" list was also undertaken to discover new word patterns. The "other" list was manually reviewed to find additional word structures. Yet, after the 4th iteration process, the size of the list nearly stayed the same and proved overwhelming and challenging to categorise. Therefore, the nouns were sorted by count frequency, indicating how often they appear in all texts. It was decided that nouns occurring fewer than five times would not be included. Initially, words were categorised without considering their frequency, as I wanted to use all possible nouns. To address this, the process was re-initiated with the entire noun list (n=58'065). By eliminating nouns that appeared less than five times across the texts, the list was narrowed down to 10'282 nouns. These remaining nouns were cross-referenced with those already placed into thematic fields during the first four categorisation steps. This exercise revealed a serious oversight in neglecting frequency up to that point; of the 10'282 nouns, only 2'414 had been previously identified, leaving 7'868 yet to be categorised. When manually categorising the nouns into thematic fields, it became evident that the lines between "Human" and "Human Influence" were too blurred to separate. Consequently, these two fields were combined. An interim review, as highlighted in the middle of Table 4.6, reveals that the "Human influence" category considerably outstrips others in size. The decrease in the noun count in the thematic fields from the 1st to the 4th iteration is due to the manual revision of the lists, in which many not-fitting nouns were removed. The challenge was not just the noticeable disparity but also the ambiguity surrounding the definition of "Human influence". To address these issues, and as previously mentioned in the chapter's introduction, the category was refined into three distinct themes: "Park management", "Research & Monitoring", and "Tourism & Guests". By deciding to categorise only nouns that appear five times or more, effectively 47'783 nouns were discarded at this stage. However, since nouns had initially been

Nouns per thematic field (1st Iteration)		Nouns per thematic field (4th Iteration)		Nouns per thematic field (final)	
Animal	3’554	Human influence	3’827	Park management	2’517
Infrastructure	3’220	Animals	1’985	Animals	2’270
Forest	3’130	Forest	1’394	Forest	1’452
Water bodies	2’796	Water bodies	1’357	Research/ Monitoring	1’444
Human	2’767	Infrastructure	1’145	Water bodies	1’402
Extreme locations	2’624	Grassland	918	Infrastructure	1’384
Human influence	2’432	Extreme locations	568	Tourism/Guests	1’074
Grassland	1’493	Anthropogenic climate change	479	Grassland	1’012
Anthropogenic climate change	1’252	Natural processes	390	Extreme locations	630
Rewilding	1’090	Rewilding	339	Natural processes	507
Natural processes	570			Anthropogenic climate change	494
Other	**33’137**			Rewilding	360
Total	**58’065**	**Total**	**12’402**	**Total**	**14’546**

Table 4.6. Noun count per thematic field in the first, fourth and final iterative process.

categorised without considering their frequencies, "only" 44’439 nouns were discarded, which amounts to 76.5% of all nouns. Furthermore, not only nouns below a frequency of 5 mentioning were discarded, but also nouns that do not hold a clear meaning without context, for example, "abends", "sommer", or "abschnitt", leading to a decreasing thematic field (for instance from the 1st to the 4th iteration). All lists were manually reviewed once more, a process that required an abundant investment of time. Nouns can also appear in multiple thematic fields. An Excel sheet was created to provide an overview, allowing for filtering by thematic fields and nouns. The final noun count per thematic field is presented in the right Table 4.6.

4.2.4 Description and characteristics of the thematic fields

This section aims to provide insights into the characteristics of the thematic fields and explain the rationale behind categorising nouns.

4.2.4.1 Primary Themes/Habitats & Animals

This category encompasses Grassland, Forest, Water Bodies, Extreme Habitats, and Infrastructure.

Grassland includes all plant species typically found in a meadow. Any nouns containing "-wiese" (meadow) or "-weide" (pasture) were categorised here. Shrubs, hedges, mosses, and peatlands were also classified under Grassland. Depending on the noun, peatlands and mosses could be categorised under Forest or Water Bodies. Nouns describing the grazing process were also classified as Grassland.

The **Forest** thematic field incorporates all types, such as tree species like larch, pine, fir, and arche. This thematic field includes wood, forest fires, roots, and fungi. Terms like vegetation, ecosystem, or biodiversity were assigned to thematic fields only if there was a clear context, like "forest vegetation".

The **Water Bodies** field includes lakes, rivers, and streams, along with their flora and fauna (fish, algae, etc.). Precipitation, groundwater, floodwater, fishing, and fish stocking are found here. Nouns related to fishing and fish stocking or residual water regulations also appear in the "Park Management" thematic field (category: causes of landscape change).

Extreme Habitats consist of nouns that describe high elevations, extreme weather conditions, and mountain structures. Specifically, terms like (block) glaciers, peaks, rockfalls, debris flows, scree, avalanches, permafrost, and talus slopes are included. This thematic field mostly overlaps with "Natural Processes" as avalanches, debris flows, and rockfalls are part of the process in the mountainous landscape.

The last thematic field, **Infrastructure**, which has the closest relation to humans, attempts to draw a clear distinction from the thematic fields of Park Management, Research/Monitoring, and Tourism/Guests, though overlaps naturally occur. For Infrastructure, the rule was to categorise it as an object/resource constructed or used by humans. This includes buildings, roads, facilities, power plants, iron, and ore indicative of mining. Overlaps with Park Management and Tourism were noted, for instance, with terms like "road traffic", which arguably belongs to all three fields, or bearded vulture observation huts, which are also assigned to the "Research" field.

Originally, **Animals** were also categorised into their habitats. However, this approach was eventually abandoned, as animals often inhabit multiple habitats. The only exception was the Water Bodies thematic field, where fishes were also categorised, as they are not found elsewhere. In the thematic field of animals, mammals, amphibians, reptiles, birds, insects and fish were classified. Also, grazing pressure (Äsungsdruck) and deer bite ("Verbiss") were included here, as animals cause them.

4.2.4.2 Causes for landscape change

Let's start with the smaller thematic fields of Rewilding, Natural Processes, and Anthropogenic Climate Change. In the **Rewilding** field, nouns that promote conservation are found, including terms like "renaturation", "revitalisation", "recolonisation", "ecologisation", "prevention of wildlife damage", and laws, regulations, projects that aim to support this concept. Since this is closely related to park management, there are overlaps. However, this category has few words since conservation is not necessarily equated with Rewilding, as this is already a part of the SNP.

In **Natural Processes**, the focus is on cycles experienced seasonally, annually, or cyclically, and lifelong by animals and nature. Examples include nouns like "soil formation process", "watershed", "erosion", "freezing processes", and "weather processes". As mentioned, events such as debris flows, avalanches and rock streams found in the Extreme Habitats field are also included here.

Classifying nouns under **Anthropogenic Climate Change** proved challenging, as there was a temptation to include terms like "melting", which cannot be easily categorised here without context. This field encompasses terms related to "temperature", wildlife populations, climate measurements, emissions, and species extinction.

Finally, the three thematic fields describing human influence are **Park Management**, **Research & Monitoring**, and **Tourism & Guests**. Since this categorisation was conducted in a later step to become more transparent, it was based on the following criteria visible in Table 4.7. The arrows indicate that, that an overlap occurred in those cases.

Research & Monitoring	Park Management	Tourism & Guests
• Site and animal observation	• Regulations, rules, measures	• Motorised traffic (private, public transport)
• Research Institutions & organisations	• Management: political and legal	• Guest houses
• Population counts (flora & fauna)	• Institutions (except research faculties)	• Leisure clubs
• Cartography/maps	• Land tenants/ agriculture/ Farmers	• Noise sources
• GIS & Modelling	• GIS (related to management work)	• Waste
• Monitoring & Measuring (temperature, water level, etc.)	• Monitoring & Measuring (in the sense of surveillance)	• Hikers, School trips, village habitants
	• construction, material transportation	
	• Fishing, Hunting and poaching issues	

Table 4.7. The nouns were categorised into the fields according to the following affiliations.

4.3 Development of Analysis Plots: Methodology and Application

4.3.1 Frequency Analysis: Steps to Creating Visualisations

For Tables 5.1 and 5.2, which aim to provide a brief statistical overview of the data, a script was created that counts the number of words, the total number of occurrences of place names, number of paragraphs, and number of paragraphs that reference at least one place name and one thematic field. Type 3 documents (see subchapter 4.1.1.5) were used for this calculation, as the paragraph structure was important. Paragraphs were defined as a new paragraph beginning/ending whenever there is an empty line in between. However, this also resulted in short sentences or titles counted as paragraphs, reflected in the many paragraphs in Table 5.2. The context rules for the place names were integrated into the code, ensuring as many locations as possible were found and correctly assigned. This was incorporated in all scripts whenever place names were searched for.

For the plots in the frequency analysis, a script was created for all corpora that counts the frequencies of place names, compiles them into a .csv file, and displays them in bar graphs (for individual publications like the Atlas and Brunies' books) or interactive line plots (for Cratschla, Business Reports, and NPF publications). For all corpora, there is a .csv file showing the absolute number of occurrences of locations and another .csv file displaying normalised frequencies. The frequencies were normalised by dividing each place name by the total occurrences of place names in the entire corpus. This normalisation was chosen to enable comparisons between the corpora. The processed Type 2 (lowercase, no line breaks) text files were used for the frequency analysis, as the paragraph structure is irrelevant in this particular case. In Table 5.4 and Plot 5.3, the sum of all normalised frequencies was compiled to simplify comparisons between the corpora. For the Atlas and Brunies' books, summation was unnecessary as they are individual publications. Heatmap 5.4 shows the sequence of places by frequency from 1 to 27. The position was derived from the corpora's normalised frequency Table 5.4. The table with normalised frequencies is the basis for the plots throughout the frequency analysis chapter. Bar graphs with frequencies in descending order were created in Excel for each corpus, as the .csv structure was already provided. The frequencies of Brunies' book editions were reviewed and compared in a subsection. For Cratschla, Business Reports, and NPF publications, in addition to bar plots, line plots were also created for a few interesting examples (max. 3-4 places per corpus), showing fluctuations across publications. The values for the example were extracted from the normalised .csv file of the requested corpus.

4.3.2 Thematic Field Analysis: Procedure and Graphical Implementation

Separate scripts were written for all these calculations and visualisations. For the graphics in chapter 5.2 "Composition of Thematic Fields in the Corpora", only the counts of nouns from the thematic fields in each corpus were recorded and saved in a .csv file. Type 2 text files were used, as it primarily involves assigning nouns in the corpora to thematic fields, and thus, the paragraph structure is irrelevant. For the Plot 5.1, the occurrences of the nouns were divided

by the total word count of the corpus, allowing for a comparison between the corpora. For Bar Plot 5.2, the ratios of the thematic fields are shown and compared again; here, the percentage shares of the fields per corpus are displayed (i.e., the sum of all percentage values of a corpus equals 1). To calculate co-occurrences between place names and thematic fields in paragraphs, all results were saved in a .csv table and visualised as bar plots. Normalisation was done as follows: hits of a specific place name with a category through all hits of the specific place name across all categories, meaning the sum of all values for a place name equals 1. The stacked bar plots thus allow for quick identification of the thematic context in which the place name was most frequently mentioned. The .csv files of co-occurrences also form the basis for visualising the thematic context in a geographical environment in the maps 5.7 and 5.14. For the bar charts in chapter 5.7, which calculated the relationship between the primary themes and the causes for landscape change, and for bar plot 5.16 in chapter 5.8, the same procedure as with the place names was applied, but no normalisation was performed. Lastly, nouns from the thematic fields most frequently occurring together were extracted for individual place names. This involved searching through all paragraphs for the selected place name in the desired corpus, counting and compiling the nouns, and saving them as a .csv file to make the top 10 most frequent nouns visible.

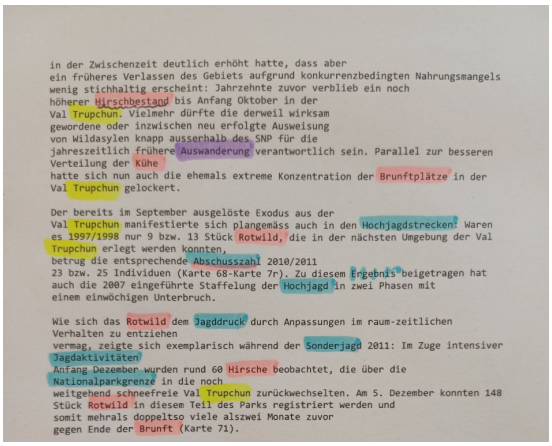
4.3.3 Evaluation: Precision, Recall, and $F1$

To assess the analytical method's respective code accuracy and reliability, which involves identifying and classifying nouns within thematic fields in the context of paragraphs containing place names, I used three key metrics: precision, recall, and the F1 score (Purves, Koblet, and Adams, 2022, p. 73). In this case, precision measures show that the code accurately identifies the nouns and accurately classifies them to a specific thematic field within paragraphs that include place names. It answers the following question: How many nouns were correctly classified into a thematic field? High precision in this context indicates that the method is reliable and that the nouns it classifies into a thematic field are likely correctly identified (4.1). Recall assesses the completeness of the method in capturing all relevant nouns associated with thematic fields within the textual corpora. It is calculated by dividing the number of correct positive results by the number of positives that should have been retrieved (the ground truth) (4.2). High recall indicates that our method effectively identifies and classifies the nouns. The F1 score provides a single metric that combines precision and recall, offering a balance between the two by calculating their harmonic mean (4.3).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.2}$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.3}$$


in der Zwischenzeit deutlich erhöht hatte, dass aber
ein früheres Verlassen des Gebiets aufgrund konkurrenzbedingten Nahrungsmangels
wenig stichhaltig erscheint: Jahrzehnte zuvor verblieb ein noch
höherer Hirschbestand bis Anfang Oktober in der
Val Trupchun. Vielmehr dürfte die derweil wirksam
gewordene oder inzwischen neu erfolgte Ausweisung
von Wildasylen knapp ausserhalb des SNP für die
jahreszeitlich frühere Auswanderung verantwortlich sein. Parallel zur besseren
Verteilung der Kühe
hatte sich nun auch die ehemals extreme Konzentration der Brunftplätze in der
Val Trupchun gelockert.

Der bereits im September ausgelöste Exodus aus der
Val Trupchun manifestierte sich plangemäss auch in den Hochjagdstrecken: Waren
es 1997/1998 nur 9 bzw. 13 Stück Rotwild, die in der nächsten Umgebung der Val
Trupchun erlegt werden konnten,
betrug die entsprechende Abschusszahl 2010/2011
23 bzw. 25 Individuen (Karte 68-Karte 7r). Zu diesem Ergebnis beigetragen hat
auch die 2007 eingeführte Staffelung der Hochjagd in zwei Phasen mit
einem einwöchigen Unterbruch.

Wie sich das Rotwild dem Jagddruck durch Anpassungen im raum-zeitlichen
Verhalten zu entziehen
vermag, zeigte sich exemplarisch während der Sonderjagd 2011: Im Zuge intensiver
Jagdaktivitäten
Anfang Dezember wurden rund 60 Hirsche beobachtet, die über die
Nationalparkgrenze in die noch
weitgehend schneefreie Val Trupchun zurückwechselten. Am 5. Dezember konnten 148
Stück Rotwild in diesem Teil des Parks registriert werden und
somit mehrals doppeltso viele alszwei Monate zuvor
gegen Ende der Brunft (Karte 71).

(a) Photo of the manual annotation of the "ground truth" data set. Text excerpt from p. 130 in the Atlas.

```
Place: Il Fuorn
  Category: Animals_final_low_cap
    schneehasen: 1 times
    kleinsäugern: 1 times
    schneemaus: 1 times
    paare: 1 times
    turmfalken: 1 times
    brutpaardichten: 1 times
    feldmaus: 1 times
    kleinvögel: 1 times
    grossinsekten: 1 times
    eidechsen: 1 times
    reptilien: 1 times
    reh: 1 times
  Category: Forest_final_low_cap
    waldgrenze: 1 times
  Category: Grassland_final_low_cap
    graslands: 1 times
  Category: Infrastructure_final_low_cap
    schloss: 1 times
  Category: Park_management_final_low_cap
    parkwächter: 1 times
  Category: Research_Monitoring_final_low_cap
    daten: 1 times
    nachweise: 1 times
    beobachtungen: 1 times
```

(b) Output of the model calculation

Figure 4.5. Process of annotation and evaluation.

This metric is particularly useful for our study as it addresses the potential trade-off between precision and recall. A high F1 score indicates that our method is precise and comprehensive. The precision evaluation steps are demonstrated here using a text excerpt from the Atlas. On approximately two pages, all place names (highlighted in yellow) and the respective nouns in the paragraphs were manually annotated, i.e., as shown in Figure 4.5a, the different colours indicate the affiliation of the nouns to the thematic fields. Subsequently, the frequencies of the nouns per place name were transferred to a table. This table was then compared with the output generated by the code (see Figure 4.5b). The output was produced using the same code that calculates the co-occurrence in the thematic analysis. Table 4.8 now lists the True Positives (TP), the nouns correctly identified and classified with the place names; False Positives (FP), which count the nouns that were misclassified, and False Negatives (FN), which show the number of nouns not detected by the code. Table 4.8 shows the place names with the respective TP, FP, and FN values. All FP values are zero, indicating that all nouns were correctly attributed to the appropriate thematic fields. A total of 11 nouns were not classified, as shown in the FN column. The review of the manual annotation revealed that the missing nouns are those that appear multiple times in a paragraph.

For example, the noun "red deer" appeared twice in the last paragraph (see Figure 4.5a), but the code calculated this as one occurrence, not two. This finding suggests that the size differences between thematic fields might be minimised within a paragraph. For instance, if three nouns from the "Animals" field, two from "Park Management", and one from "Natural Processes" are present in a paragraph, then the ratio should be 3:2:1 relative to the place name. However, this is likely only the case if the nouns from the same thematic field differ. If the three nouns in the "Animals" field were all "red deer", they would be counted as one, thus changing the ratio

within the paragraph itself. Additionally, one noun was not counted because the period directly behind the noun prevented its recognition. In Type 3 documents (with capitalised letters and paragraph structure), all special characters, including colons, semicolons, and parentheses, were removed for this reason. Still, periods were left since an analysis at the sentence level at this point could not be ruled out. Nevertheless, precision, recall, and the F1 score are in the high range, which denotes the code's performance as satisfactory.

Example Atlas: Calculating Precision, Recall, and F1			
	True Positive (TP)	**False Positive (FP)**	**False Negative (FN)**
Il Fuorn	19	0	1
Murter	19	0	1
Trupchun	38	0	6
Val Mingèr	19	0	1
S-charl	13	0	1
Zernez	13	0	1
Total	**121**	**0**	**11**
Precision:	1		
Recall:	0.917		
***F*1:**	0.957		

Table 4.8. Example Atlas: Precision, Recall, and *F*1.

This evaluation was conducted for all corpora. The values are visible in Table 4.9, and here, too, they are all in the high range. However, these values should be viewed cautiously, as multiple mentions were not counted correctly and would thus be significantly lower over the entire corpus length.

Precision, Recall, and F1 for all the 6 Corpora			
	Precision	**Recall**	***F*1**
Atlas	1	0.92	0.96
Brunies (1914)	1	0.77	0.87
Brunies (1948)	1	0.75	0.86
Cratschla	1	0.86	0.93
Business Reports	1	0.92	0.96
NPF publications	1	0.74	0.85

Table 4.9. Precision, Recall, and *F*1 for all six corpora.

4.3.4 Normalisation and its risks

The decision on how normalisation should be conducted led to frequent revisions. Initially, the frequencies of place names were normalised by the total occurrences of place names across the entire corpus and by following other approaches. For instance, frequencies were first divided by the total word count in the corpus. However, this method proved unsuitable due to considerable differences in length within a corpus and among different corpora. For example, earlier reports are only five pages long in the Business Reports, whereas newer ones can exceed 100 pages. This approach only amplified the discrepancy between the reports. To counteract this, the second attempt divided the occurrences of place names in a publication by the total word count of that given publication. This improved the variance within the corpus but led to distortions where short reports with frequent mentions of places disproportionately stood out. For these reasons, the normalisation method applied in this work was ultimately used, dividing the number of mentions of a specific place in a publication by the total occurrences of place names in the corpus. The issue of normalisation is an important aspect of this work because data can be misrepresented or even inaccurately presented, which should always be kept in mind.

5. Results and Interpretations

This chapter examines how often the 27 locations appear and in what context they are described in the six different corpora: Atlas, Brunies 1914 and 1948, Cratschla, Business Reports, and the NPF publications. It starts with a straightforward overview of token count, total place name count in the corpora, and the count of paragraphs with the percentage of place names and thematic field nouns. The following bar charts illustrate and compare the thematic composition in the corpora. Depending on the accuracy of the thematic fields themselves, the genres of the texts should be well reflected. Subsequently, an overview and ranking of the most to least frequent locations in all corpora is provided. The next section contains the most substantial part of the results. For each corpus, the frequencies of place names are calculated and visualised. In the case of Cratschla, the Business Reports, and the NPF publications, these frequencies are also plotted over time (i.e., throughout the publication series). A slightly different approach is followed for the individual publications (Atlas and Brunies books) and the publication series (Cratschla, Business Reports, and NPF publications). For the Atlas and editions of Brunies, assumptions are made based on frequency analysis as to why certain places appear more important than others, incorporating the background knowledge about the places (see subchapter 3.2.1 and Appendix A). The thematic context is then examined for a few important and interesting places, and the analysis focuses on which nouns are most frequently mentioned in connection with these places. This approach aims to determine if it is possible to infer potential historical events or the characteristic features of the selected places through the thematic fields and frequencies. This is then verified within the corpora themselves. The second approach, which affects Cratschla, the Business Reports and the NPF publications, involves conducting another frequency analysis, yet only a few intriguing places get plotted over time. When frequency anomalies are observed, assumptions are made about why this might be the case, taking the location's background information into account and subsequently examining the thematic context. The difference from the first approach is that only the co-occurrence of nouns of the thematic fields with the place names in periods showing increased frequencies or drops in frequencies is investigated, meaning that only certain publication years are considered. In the last chapter, the relationship between the habitats, causes of landscape and animals are analysed in the Business Reports and NPF publications without considering the locations. Because almost the same analytical procedure was conducted for the six corpora, the scope of the results is considerably long and, in places, repetitive. For this reason, the results of the Cratschla and the NPF publications have been moved to Appendix B. These two corpora were chosen because the NPF publications will

be focused on in the chapters 5.7 and 5.8, where the thematic fields are examined in relation to each other, and because the issues or publications of both corpora always revolve around a thematic focus, making it impossible to conclude timely thematic events. Nevertheless, these have also produced interesting insights, which can be read in the Appendix B.

5.1 Basic statistical overview

5.1.1 Token and locations counts in all corpora

A. Overview of total token count and place names in the 6 corpora (Type 2. processed text with lowercase)			
	Total tokens	**of which places names of interest**	**place name/total**
Atlas	99’689	829	0.83%
Brunies (1914)	52’721	369	0.70%
Brunies (1948)	70’925	593	0.84%
Cratschla	919’906	5’270	0.57%
Business Reports	1’349’228	12’248	0.91%
NPF publications	4’610’203	26’276	0.57%
Total	**7’102’672**	**45’585**	**0.64%**

B. Overview of total token count and place names in the 6 corpora (Type 3. processed text with capitalised words)			
	Total tokens	**of which place names of interest**	**place name/total**
Atlas	99’694	807	0.81%
Brunies (1914)	52’723	362	0.69%
Brunies (1948)	70’929	585	0.82%
Cratschla	834’029	5’159	0.62%
Business Reports	1’349’228	11’991	0.89%
NPF publications	4’695’162	25’913	0.55%
Total	**7’101’765**	**44’817**	**0.63%**
Difference between the Tables (A-B)	907	768	

Table 5.1. Overview of the total token count and the share of the 27 locations appearances in their respective 6 corpora. Table A shows the result with the processed text file (2) and Table B with processed text files (3)

Table 5.1 (A and B) provides an overview of the token count in the six corpora. The size differences are noticeable at first glance, with the NPF publications surpassing the others with over 4.6 million tokens. Naturally, it makes sense that the NPF publications and the Business Reports are the most extensive, as they cover the longest period (see Table 4.1). Since the processed text files are available in different formats (see subchapter 4.1.1.5), Table A shows the occurrences of the locations in the processed lowercase files (Type 2), and Table B respectively shows the occurrences in the processed text files in uppercase and with paragraph structure (Type 3). It was observed that Type 3 recorded a lower count of both words and place names,

specifically 768 fewer place mentions. The difference in word count is particularly noticeable in Cratschla and could be largely due to manual revisions of the text, where parts were elaborated, deleted or added. Most of the missing places mentioned in Type 3 are predominantly associated with the Cratschla corpora, the Business Reports, and the NPF publications. The extraction of place names was conducted similarly for Type 2 and Type 3. Furthermore, the function was first applied to a subset of all corpora, where the same number of places was identified. The issue arose when the function processed the entire corpus, complicating the troubleshooting process. A potential issue could be the paragraph structure, which might have caused individual places to be separated; this, however, is merely a hypothesis. It was also noted that place names are proportionally less common in Cratschla and the NPF publications than in the Atlas or the Business Reports. This is not surprising and relates to the nature of the corpora. The Atlas includes many maps and details about geographical, biological, and historical perspectives of places and undoubtedly references the investigated localities. Similarly, the Business Reports document the development of the SNP and its locations on a scientific and political level for over a century. The notable increase in mentions of place names from the first to the fourth edition of Brunies suggests potential local changes in the SNP in the first 40 years; more will be discussed in the frequency analysis chapter 5.3. Yet, overall, 0.64%, respectively, and 0.63% of the text contain the investigated locations. All corpora generally show a similar ratio of place names to total word count (<1%). For the frequency analysis, text documents of Type 2 and for the analysis of thematic fields, Type 3 were used.

5.1.2 Mentions of locations & thematic fields in paragraphs

Overview: Paragraph Count, Location and Thematic Field Mentions by Corpus (Type 3 - Text with Capitalised Words)

	Number of Paragraphs	Location Mentions in Paragraphs (%)	Location & Thematic Mentions in Paragraphs (%)
Atlas	2’258	384 (17.01%)	383 (16.96%)
Brunies (1914)	1’084	204 (18.82%)	198 (18.27%)
Brunies (1948)	1’828	335 (18.33%)	310 (16.96%)
Cratschla	11’718	2484 (21.2%)	2418 (20.63%)
Business Reports	55’997	6682 (11.93%)	6222 (11.11%)
NPF publications	199’039	14196 (7.13%)	12631 (6.35%)

Table 5.2. Overview of the number of paragraphs, the percentage of locations, and the percentage share of the co-occurrence of locations and thematic fields in the paragraphs.

Due to the importance of the paragraph structure in the thematic analysis, Table 5.2 shows the number of paragraphs in the corpora, the percentage of place names, and the percentage of places and thematic fields in the paragraphs, with the percentages in parentheses referring to the total number of paragraphs. Notably, at least one thematic field is present in almost all

sections where a place name occurs. This is unsurprising since all nouns were extracted at the paragraph level and assigned to thematic fields, though concessions were made for nouns appearing less than five times, which could be the reason for the slight decrease in percentage. Looking purely at the percentage, it is observed that most unique nouns (less than 5 mentions in all corpora) were probably attributed to the NPF publications, as the corpus has the lowest percentage (6.35%) of location and thematic mentions in a paragraph. In the thematic field analysis, Cratschla (20.63%) includes by far the most paragraphs proportionally, followed by the first edition of Brunies (18.27%), the Atlas and the second edition of Brunies (16.96%), the Business Reports (11.11%), and finally the NPF publications with 6.35%. This is particularly unfortunate since the Business Reports and the NPF publications cover the longest period. The high number of paragraphs is misleading since all words or text sections separated by at least one blank line were counted as a paragraph; this includes page numbers, table of contents, and title captions in the texts.

5.2 Composition of thematic fields in the corpora

In this chapter, we delve into the thematic composition of the corpora using the 12 self-created thematic fields (for a detailed description, see subchapter 4.2).
In Table 5.3, we see again the division into "Causes of Landscape Change", "Habitats", and "Animals". In a later chapter 5.7, this subdivision of the thematic fields will become again relevant, as the relationship between the "Causes of Landscape Change" and "Habitats" and the relationship between "Animals" and their "Habitats" will be analysed in the Business Reports and NPF publications. But before we get to that, this section will first illustrate the general distribution of the thematic fields in the six corpora. The highest sum of normalised occurrences belongs to the "Park Management" field across the corpora (visible in Plot 5.1). This is not surprising, given the field's overall size. It is especially pronounced in the Business Reports, which fit the corpus genre. It seems that the NPF references the fewest management-related issues. Also, the Atlas has quite a low percentage, indicating more science-oriented content in the books, similar to the NPF publications. The similarity between the Atlas and NPF

Causes for landscape change	Primary themes/ Habitat	Additional field
Anthropogenic climate change	Extreme locations	Animals
Natural processes	Forest	
Rewilding	Grassland	
Park management	Infrastructure	
Research & Monitoring	Water bodies	
Tourism & Guests		

Table 5.3. Thematic fields in their subcategories. The underlined thematic fields depict the human influence on landscape change.

publications is not unusual, as the Atlas is one of the NPF publications. Showing again as the "Research & Monitoring" thematic field is less common in the NPF and the Atlas than in Cratschla and the Business Reports. This could also be due to the nouns addressing monitoring attributes rather than the research part in the thematic field. The thematic field "Research and Monitoring" for Brunies editions is irrelevant, where park management, animals, forest, and infrastructure seem more important.

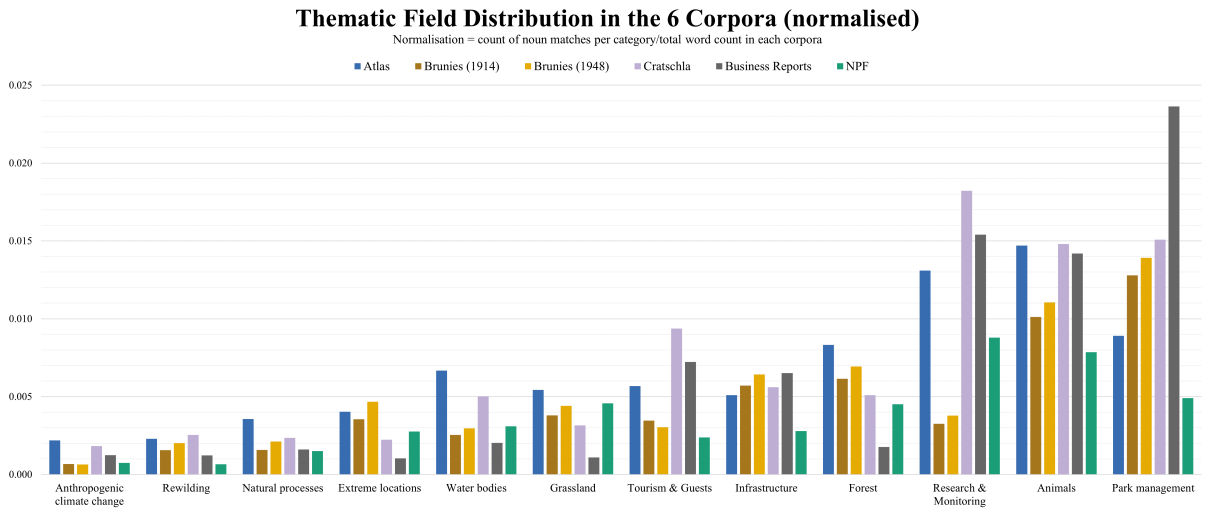


Figure 5.1. Overview of the thematic field distribution in the 5 corpora. Normalisation: topic's occurrence count per corpus/total word count in said corpus.

In the Brunies editions, discussions on animals and park management issues are at the forefront, followed closely by forest-related content. The dominance of topics such as park management, research and monitoring, tourism, and infrastructure in the Business Reports can be directly attributed to the standardised procedural documentation within these documents. While nature (flora and fauna) is a part of the reporting, it often emphasises the animal world. It is reasonable that the Atlas emphasises research on animals and water bodies, grasslands, and forests, which seem to have their rightful place in the said corpus. I would also like to draw attention to the subtler categories with a clear temporal dimension as they capture change, specifically "Anthropogenic Climate Change" and "Rewilding". For instance, "Rewilding" is especially pertinent in the Atlas and the Cratschla series. With the publication of the Atlas in 2013 and Cratschla spanning from 1992 to the present, both corpora cover a relatively recent period, making the prominence of the themes of "Anthropogenic Climate Change" and "Rewilding" in these collections all the more intriguing. One could infer that, particularly in the case of rewilding, it's a relatively new concept as the term emerged in the late 1980 (Lorimer et al., 2015, p. 41) and hence appears more frequently in the latest publications. The distinction between rewilding and nature conservation is not clear-cut, as both aim to protect and enhance the natural environment. Still, they differ significantly in their approaches and underlying philosophies. Rewilding seeks to restore natural processes and wilderness areas, often by reintroducing keystone species that have been extirpated or extinct in an area. On the other hand, nature conservation focuses on

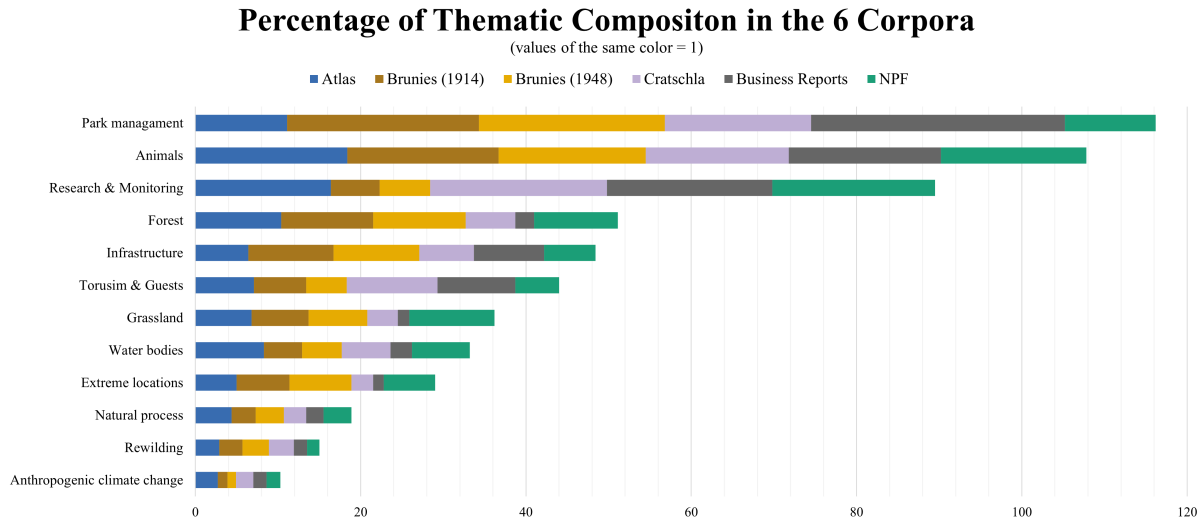


Figure 5.2. Composition of thematic fields in the 6 corpora.

preserving specific species, habitats, or landscapes to protect biodiversity or particular species from extinction. The difference between the two concepts is that rewilding aims to increase ecosystem resilience and self-sustainability. At the same time, conservation seeks to maintain the current state of an ecosystem or prevent further degradation (Lorimer et al., 2015). This unclear distinction between the approaches could also explain why rewilding appears in the Brunies editions and annual reports. Anthropogenic climate change is predominantly present in the Atlas and Cratschla and the Business Reports. Curiously, research in this area does not seem as relevant in the NPF publications as expected. The discrepancy between the thematic fields is relatively large due to the size of the thematic fields, ranging from 360 nouns (Rewilding) to 2'520 nouns (Park Management). Furthermore, one must remember normalisation, as the NPF publications have the highest word count, making the bars appear smaller. From the beginning of creating the thematic fields, it was pointed out that removing nouns with four or fewer appearances in the corpora could have affected the NPF publications in particular. There may be comprehensive work on climate change or rewilding in the NPF publications. Still, it might be somewhat overshadowed due to the normalisation process and missing nouns in the thematic fields. Normalising by the number of nouns per corpus would have mitigated this, but comparisons between the corpora would not have been possible. Figure 5.2 shows a different overview of each corpus. The stacked bar charts illustrate how well the self-created thematic fields reflect the genre of the book. In the Atlas, which compiles the most important findings and events of the last 100 years, animals are most frequently represented, followed by research. The reason animals occupy the highest spot is also due to them often being mentioned in connection with habitats.

For instance, insects, ungulates, and marmots are all also mentioned when discussing meadows. All habitats are well represented in the Atlas. In Brunies' books, park management constitutes the largest share in both editions, but animals are also prominently featured, especially in the fourth edition. Brunies' books (similar to the Business Reports) cover flora and fauna as well as

infrastructure, such as the construction of measuring stations. Cratschla aims to bring the park closer to the outside world and to visitors, which could explain the bigger portion of tourism and guests. The Business Reports discuss extreme habitats, grassland, and other habitats the least. Here, the most important themes are animals, tourism, and infrastructure. The regular distribution of thematic fields in the NPF publications reflects the diversity of the publications, with research and animals leading. The prominent role of park management can be explained by the often-mentioned history of the park's establishment in NPF publications. In any case, all habitats are researched and documented. The thematic composition of the corpora is important for the analysis of the thematic context of places. How places are written about depends not only on the places themselves but also on the scope of the texts. The hypotheses about which thematic fields most commonly co-occur with the places are based solely on the places themselves (subchapter 3.2.1 and Appendix A). In the following chapters, it will become clear whether the places' characteristics or the book's genre predominantly influence the places' descriptions.

5.3 Most frequent places in the corpora

Note to readers: The locations that are linked to the location descriptions in the subchapter 3.2.1 and Appendix A are written in *italics*.

This section illustrates the frequencies of the 27 locations across the publications. The mere number of locations in a corpus or across corpora provides little information about the individual places. For this reason, in Table 5.4, each location is listed along with its corresponding normalised occurrence in the six corpora. Normalised means dividing each specific place name count by the total place name count in the respective corpus. The green fields indicate the top five frequencies with the corresponding rank in the brackets for each corpus. Zernez appears to be the most relevant for all corpora except the Atlas, in which Ofenpass seems more relevant. Certain places generally show higher significance than others across all six corpora. Locations that are common for five corpora are Zernez (except in the Atlas), Spöl (except NPF), and Trupchun (except in the fourth edition of Brunies). For three corpora, it is Buffalora, Il Fuorn and S-charl. Buffalora and S-charl are especially relevant for the Brunies' books and the NPF publications; one might assume they were especially important in the first half of the 20th century, as that is where the timelines of Brunies and the NPF publications intersect. Ofenpass shows high relevance in the Atlas and Cratschla, suggesting a high relevance due to their thematic focus and for the SNP.
The same argument for Buffalora and S-charl applies to Stabelchod, often mentioned in Cratschla and the Business Reports. The table clearly shows the absence of the location Macun in the editions by Brunies. As known, *Macun* only became part of the SNP in the year 2000. Therefore, it can be assumed that until 1948, no negotiations or plans were underway for these future developments. *Murter* appears more relevant in Brunies' books than other corpora, indicating the location's early importance within the SNP. Not surprisingly, the unique relevance of Val

	Atlas	Brunies (1914)	Brunies (1948)	Cratschla	Business Reports	NPF
Alp la Schera	0.01327	0.00542	0.00842	0.01006	0.01388	0.02143
Buffalora	0.05187	(3) 0.09214	(3) 0.10606	0.02751	0.03927	(5) 0.07783
Champlönch	0.01689	0.01355	0.01178	0.01025	0.00914	0.00837
Charbunera	0	0	0	0.00038	0.00196	0.00042
Grimmels	0.01448	0.00813	0.01178	0.01195	0.02147	0.01842
Il Fuorn	(3) 0.08685	0.00813	0.03199	0.05693	(3) 0.10002	(2) 0.09682
Macun	(5) 0.07961	0	0	0.06091	0.04482	0.02401
Munt Chavagl	0.01689	0	0.00168	0.00797	0.01135	0.00845
Murtaröl	0.00724	0.00813	0.01515	0.00683	0.00425	0.01416
Murter	0.02292	(5) 0.04878	(5) 0.0505	0.01613	0.02400	0.04415
Ofenberg	0.00362	0.01355	0.00673	0.00266	0.01282	0.01225
Ofenpass	(1) 0.13269	0.00542	0.02189	(5) 0.06167	0.04809	0.04103
Ova Spin	0.05911	0.01626	0.02189	0.05085	0.01902	0.01945
Pass Dal Fuorn	0.01086	0	0.00842	0.00133	0.00049	0.01439
Piz d'Esan	0.00241	0.02168	0.02020	0.00190	0.00065	0.00506
Piz Terza	0.00603	0.00813	0.00842	0.00645	0.00351	0.00696
S-charl	0.01809	(4) 0.08943	(2) 0.11616	0.03074	0.02531	(3) 0.08289
Spöl	(4) 0.08203	(2) 0.11653	(4) 0.10438	(2) 0.12941	(3) 0.07422	0.06926
Stabelchod	0.07720	0.00813	0.03199	(4) 0.06376	(5) 0.06834	0.05046
Tamangur	0.00483	0.02168	0.01178	0.00645	0.00122	0.01431
Trupchun	(2) 0.1146	0.04878	0.03367	(3) 0.11139	(2) 0.17048	(4) 0.07787
Val Cluozza	0.02895	0.04878	0.01684	0.02486	0.03298	0.03775
Val da l'Acqua	0.01448	0	0.00673	0.00474	0.00425	0.00289
Val da Stabelchod	0.00724	0.00271	0.00337	0.01347	0.00882	0.00304
Val Mingèr	0.03739	0.02168	0.01178	0.02372	0.02621	0.02580
Val Sassa	0.01809	0.00271	0.01684	0.00968	0.01086	0.00891
Zernez	0.07238	(1) 0.39024	(1) 0.32155	(1) 0.24801	(1) 0.22257	(1) 0.21362

Table 5.4. Overview of normalised place name count in all corpora. The normalisation was done by the total count of place names in each corpus, e.g., the absolute count of the place name "Alp la Schera" in the Atlas/total counts of all place names in the Atlas. Each column sums to 1. The green cell with the number in brackets indicates the five highest frequencies.

Cluozza in Brunies (1914) stands out. Negotiations had been underway since 1908 to make *Val Cluozza* the core of the SNP, which should explain its high relevance in the first edition.
Figure 5.3 clearly shows the distribution of the frequencies of the locations once more. The sharp decline in frequencies after Zernez is even more evident here. However, this is not surprising since *Zernez* forms the political and touristic centre of the SNP. Zernez is also known for its advertising slogan, "Gateway to the National Park" (Schweizerischer Nationalpark, 2024b), and serves as the centre for the administration of the SNP. It also forms the transport hub in the Engadin and the junction towards the Ofenpass and the National Park. Unsurprisingly, these locations are frequently mentioned. Hydroelectric power and the ecological flooding of the *Spöl* are key issues and conflicts that have left a clear mark on the SNP. *Trupchun* plays a vital role for red deer and is a magnet for tourists. It is the premier tourism destination, with approximately 30'000 visitors annually. Since the SNP's early years, *Buffalora* has been an integral part of the park. By the 1920s, Buffalora, home to a meteorological station, became a crucial site for various monitoring projects, including climate, ungulates, rock glaciers, and vegetation growth. *S-charl*, considered a potential SNP site before its establishment, is now renowned for its Bear Museum and the historical significance of its silver and lead mines. This historical and ongoing relevance of S-charl to the SNP is evidenced by its frequency of mentions. We can identify five major drops in frequency. The largest, of course, occurs after Zernez, followed by a second distinctive drop after Trupchun and Spöl. After that, we find the places from Buffalora to Stabelchod ranking in 8th place, then from Macun to Val Mingèr in 13th place, Grimmels, beyond which the frequencies are consistently low to very low. This figure nicely shows the

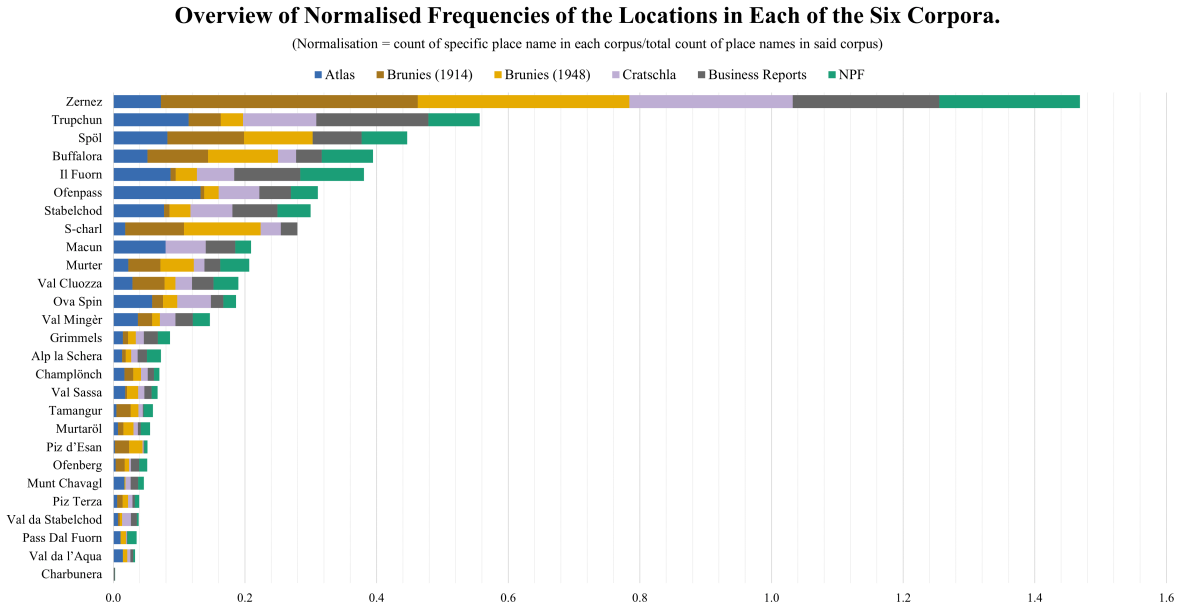


Figure 5.3. Overview of normalised frequencies of the locations in each of the five corpora. (Normalisation = count of specific place names in each corpus/total count of place names in said corpus)

locations that are less common in the corpora. For example, Charbunera is not mentioned at all in the Atlas and Brunies books and is barely mentioned in the other three. Also, Val da l'Acqua, Pass dal Fuorn, Piz Terza, Munt Chavagl, Ofenberg, and Piz d'Esan, to mention a few, rank relatively low in all corpora. As we already know, Pass dal Fuorn corresponds to the Ofenpass, indicating a general preference for using German spelling in the corpora. The opposite is true for Ofenberg, the German name for Il Fuorn. The reason *Il Fuorn* is more frequently used could likely be the Hotel Il Fuorn. *Val da l'Acqua* is a remote valley without hiking trails and is possibly mentioned primarily in the context of rock glacier occurrences. This could be analysed within the thematic fields. *Val da Stabelchod* is a floodplain experiencing debris flows, with the last significant event occurring in the summer of 2018 (Schweizerischer Nationalpark, 2024h). Here, it's important to incorporate the temporal component in Cratschla, Business Reports, and possibly the NPF. Charbunera is a so-called "field name" ("Flurname") that refers to the area's past charcoal production, but it seems to receive little attention in the texts, with only Cratschla, Business Reports and the NPF mentioning it. Regarding ranks, Figure 5.4 shows a heatmap displaying the 27 locations and their respective rank in each corpus. This visualisation helps with readability and comparison. This subchapter aimed to provide an overview of the frequencies across the six corpora and illustrate their distinct variances. The corpora's detailed frequency and thematic field analysis in the next chapter will focus on only a few locations. Those places were selected based on their relevance resp. high frequency in a corpus. High-relevant places frequently appeared across all the publications and received a high ranking in Figure 5.4. As Zernez is often the highest-ranked place in all corpora, I also focused on more uniquely important places in one corpus than the others. For example, "Ofenpass" for the Atlas or "S-charl" for Brunies books. Naturally, gaining insight into the thematic context of

all places in all corpora would have been exciting, but the sheer volume would have been too great to tackle in this thesis. Therefore, only a select few will be addressed.

Here are the top 5 places for each corpus. The **locations** written in bold will be the main focus of the detailed thematic analysis.

- Atlas: **Ofenpass, Trupchun,** Il Fuorn, **Spöl**, Macun
- Brunies 1914: Zernez, Spöl, Buffalora, **S-charl**, Murter/Val Cluozza/Trupchun
- Brunies 1948: Zernez, **S-charl**, Buffalora, Spöl, Murter
- Business Reports: Zernez, Trupchun, Il Fuorn, *Spöl*, Stabelchod. Selected based on the interesting pattern: **Buffalora, Macun.**

The following examples can be found in Appendix B:

- Cratschla: Zernez, **Spöl**, **Trupchun**, Stabelchod, Ofenpass. Selected based on interesting pattern: **Macun, Ova Spin**
- NPF publications: Zernez, **Il Fuorn, S-charl, Trupchun,** Buffalora. Selected based on the interesting pattern: **Spöl**.

Figure 5.4. Ranking the location from 1 to 27 for all 6 corpora.

5.4 Atlas

5.4.1 Frequency of place name

The Atlas overlooks the temporal component since it involves only one document, yet it nicely summarises the past 100 years of the SNP. The Atlas is a comprehensive edition that deals with history, flora and fauna, research, and the impact of humans. The frequency of mention of certain locations can reveal a lot about the context of the book. Figure 5.5 displays the most interesting places in the Atlas. By a considerable margin, Ofenpass and Trupchun are the most frequently occurring locations, with values of 0.133 and 0.115.

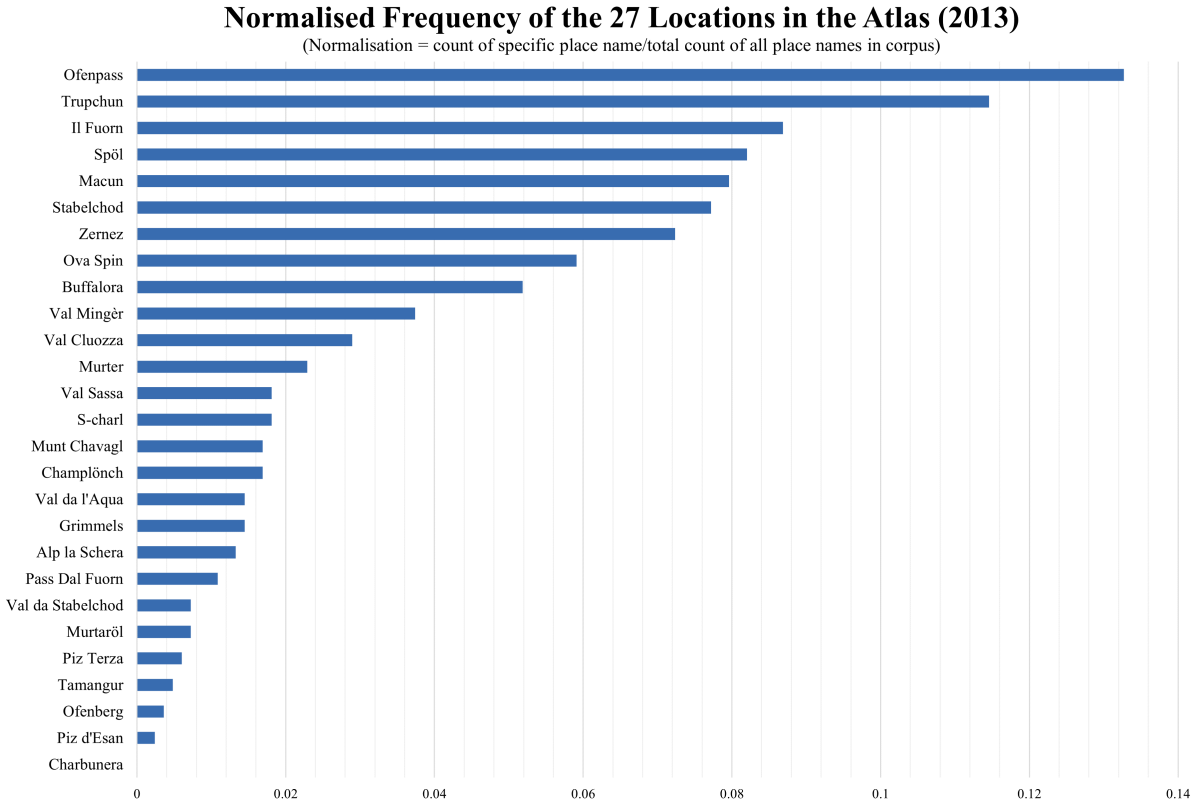


Figure 5.5. The normalised frequency of field and place name in the Atlas (2013). The normalisation was conducted by dividing the count of specific place name by the total count of all place names.

Following are the places Il Fuorn (0.087), Spöl (0.082) and Macun (0.08). Not at all relevant are Charbunera, Piz d'Esan (0.002), Ofenberg (0.004) and Tamangur (0.005). Pass dal Fuorn, the Romansh spelling of Ofenpass, is rather on the irrelevant side of the frequencies, indicating that in the Atlas, mostly the German spelling was used. A quick review of the corpus shows that most of the matching of Pass dal Fuorn was in cartographic illustrations. The same goes for Ofenberg and Il Fuorn, where the Romansh spelling is preferred. The top-rank places are not that surprising, as Ofenpass is one of the main and large areas in the SNP. Additionally, the *Ofenpass* area spans a considerable size, featuring high terrain, water bodies, forestation, and infrastructures like the heavily trafficked Ofenpass road, along which the Ova Spin hamlet, the Il

Fuorn Hotel, and the Buffalora meteorological station, as well as several rest areas, are located. The Atlas summarises developments and changes from a scientific perspective. The frequent mention of locations that are key research and monitoring sites is not surprising. For example, *Trupchun* is a significant location for wildlife behaviour research, where many management strategies regarding red deer issues have been developed. Il Fuorn, situated in the Ofenpass area, could often be mentioned due to the frequent references to the Ofenpass. This makes sense as the mining history heavily influences the pass's characteristics in Il Fuorn. *Spöl* and *Ova Spin* are the main locations where hydrological studies are conducted, both marked by a turbulent history due to the construction of the hydropower plant in the 1960s. The *Macun* lake plateau, only part of the National Park since the turn of the century, has since become the subject of hydrological studies. The *Buffalora* meteorological station is central to climate development studies. Operational since 1917, it has documented a clear increase in temperatures. Another important location is the Alp *Stabelchod*, which has been studying animal behaviour and plant growth through permanent plots for over 100 years. A pattern emerges that places with permanent plots and measurement stations are most frequently discussed. Locations like *Val Sassa* and *Val da l'Acqua*, both featuring rock glaciers, *Munt Chavagl* with the solifluction measurement station, or *Grimmels*, sharing similar conditions to Stabelchod, received noticeably less attention in the book.

5.4.2 Thematic context

For three studied locations, I want to examine whether the thematic context aligns with their descriptions and characteristics. For this purpose, Ofenpass, Trupchun, and Spöl have been selected. Although Il Fuorn was in third place, it was excluded as it lies in the Ofenpass region. In Chapter 3.2.1, assumptions about which thematic fields they most likely co-occur with were made for each location. It was assumed that park management and infrastructure are highly relevant for all three locations. Tourism and forest are expected to appear with *Ofenpass* and *Trupchun*, with the latter also co-occurring with the grassland habitat. Due to ungulate and hydrological studies, it was assumed that the research and monitoring field would often co-occur with Trupchun and Spöl. Compared to Ofenpass, a high occurrence of the animal thematic field is expected in both. *Spöl* is likely to have the strongest co-occurrence with the water bodies thematic field. Compared to the other two, the rewilding and natural processes thematic fields are considered relevant here due to fish stocking and restoration (artificial flooding). It is evident how the landscape features of the locations are reflected in the co-occurrence of thematic fields in Figure 5.6 and the map 5.7.

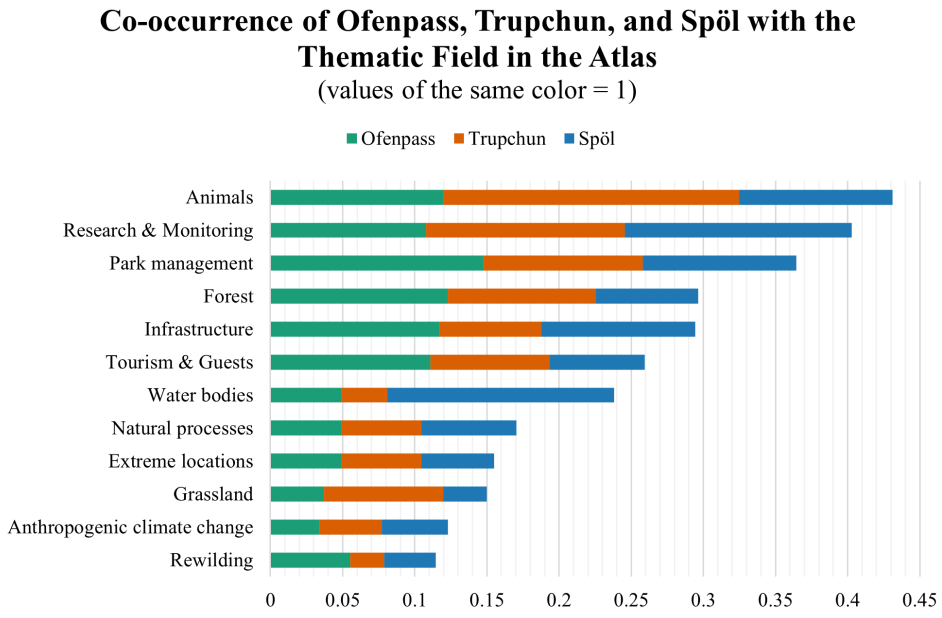


Figure 5.6. Normalised co-occurrence bar chart of most relevant locations and thematic fields in the Atlas.

Trupchun appears strongly with the animal's thematic field. It is a place of diverse biodiversity marked by hunting issues but also a popular excursion destination in the SNP and, therefore, matches the increased co-occurrence in these fields: animals, research/monitoring, park management, forest, grassland, and tourism. The Ofenpass has a high co-occurrence with the human thematic fields, park management, research and monitoring, tourism and guests, as well as the habitats, such as forests, infrastructure, and animals, especially visible in the map 5.7.

Co-occurrence of Ofenpass, Trupchun, and Spöl with the Causes of Landscape Change, Habitats, and Animals in the Atlas
[absolute count]

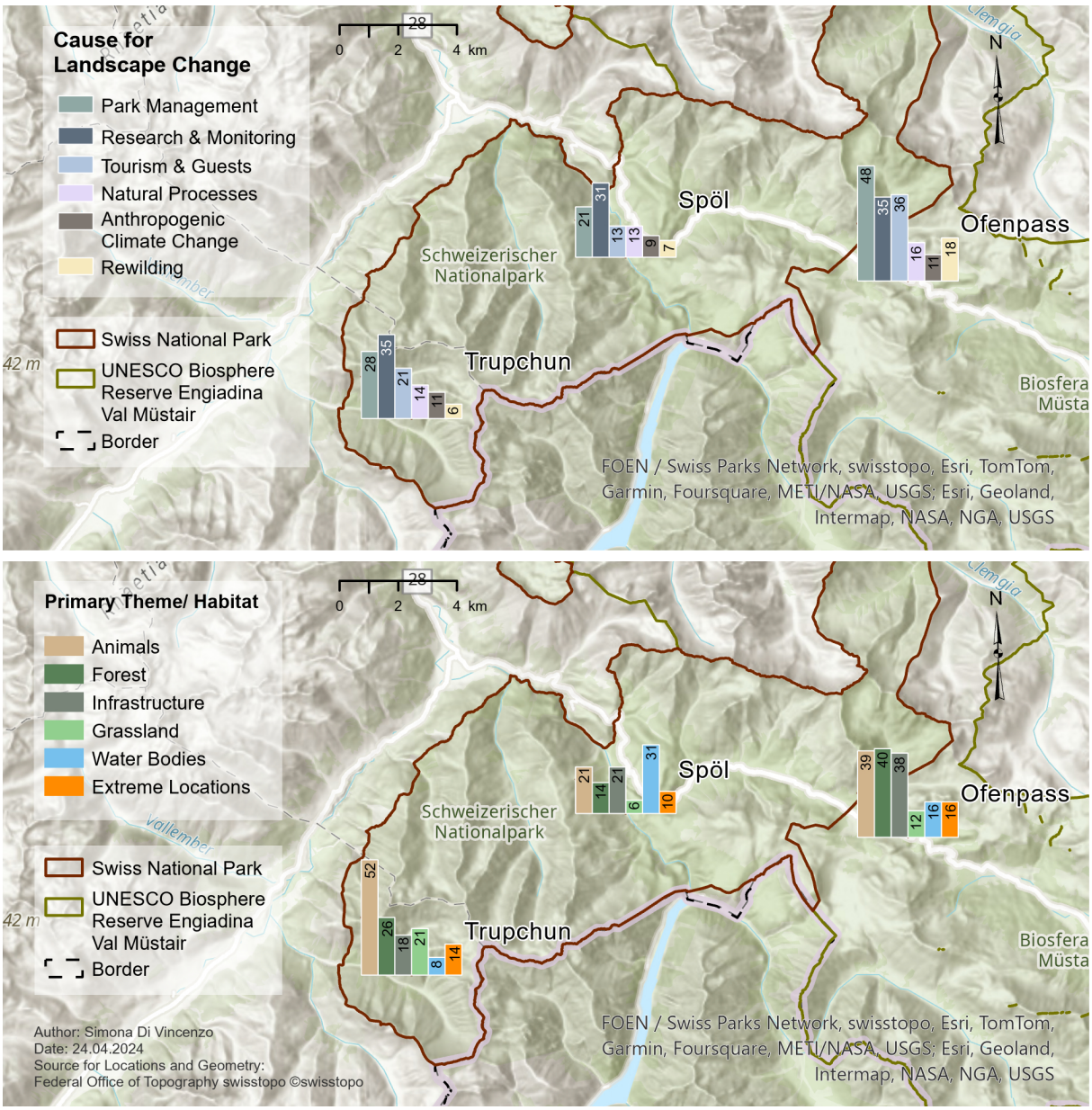


Figure 5.7. Absolute count of Co-occurrence of Causes for Landscape Change and Primary Themes/ Habitat with the Locations Ofenpass, Trupchun, and Spöl.

This contradicts the assumption that it is highly relevant to the animal field made at the beginning. The high human thematic fields and Infrastructure values might be due to the region's Ofenpass Road, bus stations, and Hotel Il Fuorn. Spöl had the most hits with "Water Bodies" but also with research, park management, and infrastructure, which is likely related to the hydroelectric power plant; animals also play an important role, which might indicate fish stocking. Yet, as in the categorisation process, the aquatic fauna was also included in the water bodies field, which could increase the water bodies percentage even more.

The smaller thematic fields of "Rewilding" and "Anthropogenic Climate Change" are not represented in the three locations. Yet, natural processes in Trupchun and Spöl show slightly elevated concordance, and rewilding occurs surprisingly in the Ofenpass area; the expectation that Spöl would exhibit elevated values here did not materialise. Generally, the map clearly shows that the Ofenpass strongly co-occurs with the "Causes of Landscape Change" themes and with forest and infrastructure habitats and animals. The "Animals" and "Research & Monitoring" themes are prominent in all three locations. The assumptions made here about the possible reasons why the place names co-occur so frequently with the thematic fields will now be verified by extracting the most common nouns. This will also provide insight into the content of the text and the immediate context of the locations (see Table 5.5).

Ofenpass	Trupchun	Spöl
verkehr (7)	tiere (15)	totholz (11)
jungvogel (6)	parkwächter (12)	hochwasser (8)
wasserfallfelsen (4)	beobachtungen (9)	inn (8)
vögel (4)	gämsen (7)	kraftwerke (6)
strasse (3)	rothirsch (7)	proben (5)
infrastruktur (3)	daten (7)	fische (5)
bartgeierpaar (3)	reh (6)	einzugsgebiet (5)
brut (3)	steinbock (6)	untersuchungen (4)
paare (3)	biomasse (6)	acqua (4)
schneedecke (3)	exposition (6)	ausgleichsbecken (4)
eis (3)	pflanzen (6)	
nationalparkbesucher (3)		
motorräder (3)		

Table 5.5. Example: Most frequent nouns associated with Ofenpass, Trupchun, and Spöl.

At first glance, it is apparent that Trupchun and Spöl are mentioned more frequently in connection with the nouns of thematic fields than Ofenpass, even though Ofenpass occurs more often in the text itself. The Atlas focuses on research, describing the SNP's habitat and history. Floods and power plants are terms that have shaped Spöl. The hydrological studies in the 1990s and the power plant construction in the 1960s are highlighted here with the nouns "hochwasser" and "kraftwerke". Interestingly, the noun "totholz" appears the most with Spöl, not expecting this connection. Yet, after a quick review of the Atlas, all the occurrences are concentrated in one chapter about dead wood, in which the valley of Spöl is mentioned as a hotspot for dead wood. As mentioned in the frequency analysis, these places often appear together, and this is also the case in the Atlas. The research is reflected in nouns such as "beobachtungen," "daten", "untersuchungen", and "proben" for Trupchun and Spöl. For the SNP, both places are actively involved in measurements and observations. Ofenpass has relatively low references to research; in the Atlas, the place is most often described in connection with "traffic." Road and infrastructure are also mentioned looking at "verkehr" being on the highest rank, "strasse",

and "motorräder". Yet, the higher percentage of the animal thematic field is reflected as well with "jungvogel", "bartgeierpaar", "brut", and "vögel", suggesting a focus on bird topics. This could reference the release of bearded vultures near *Il Fuorn* in 1991, located in the Ofenpass region. Similarly, it is visible that Trupchun is an important place for ungulates, as shown by the nouns "gämsen", "rothirsch", and "steinbock". The mentioning of "nationalparkbesucher", "motorräder", and "verkehr" also suggest the higher percentage of tourism & guests in the Ofenpass region. In summary, the topographic features and the habitat of the native animals at the three locations can be identified through the co-occurrence calculation and the extraction of nouns. However, the nouns alone are insufficient to understand the depth of the content about the location in the Atlas without background knowledge. Nonetheless, very typical and plausible descriptions of the places are found among the most common nouns.

5.5 Brunies (1914/48)

5.5.1 Frequency of place name

In the books by Steivan Brunies, there is no annual publication, so there is no detailed insight into how the significance of a location might have changed over the years or within these genres of books. However, with the two editions from 1914 and 1948, it is possible to delve a little into the events of the 34 years following the opening of the Swiss National Park. Figure 5.8 shows again the normalised frequency of the locations in both books. Some localities received more overall focus again, such as Zernez and Spöl. The high S-charl, Murter, and Piz d'Esan frequencies are interesting. The latter is not relevant for all other corpora. The high importance of *Zernez*, especially in the first edition, could reflect the municipality's importance in the SNP's origin as contract negotiations regarding the leasing of *Val Cluozza*, which is situated in the Zernez area. This would also explain the high importance of Val Cluozza from the first to the fourth edition. *S-charl* importance in the books compared to the other corpora mirrors the location's high relevance in the park's first decades, as *Tamangur* (located in Val S-charl) should have formed the heart of the SNP, which later declined.
Not mentioned is again Charbunera but also *Macun*, a territory which joined the SNP in the year 2000. Some localities can be found only in one edition, or there is a strong increase/decrease between the editions. For example, Munt Chavagl, Val da l'Acqua, and Pass dal Fuorn only appear in Brunies 1948. The lack of appearance in one or both editions could indicate a change in National Park history or an important occurrence like the reintroduction of animals to installing monitoring stations in certain locations. Figure 5.9 shows how the frequency has changed from 1914 to 1948. If values are higher than 0, resp. the bar extends to the right side; there was an increase after 1914. However, it's important to look at both plots 5.8 and 5.9, especially since the latter only shows the differences between the editions and does neglect the size relationships between the place names.
What is immediately noticeable is how Zernez and Val Cluozza's frequency decreased by approximately 0.07 and 0.03 since 1914, marking the most distinct decrease in the new edition.

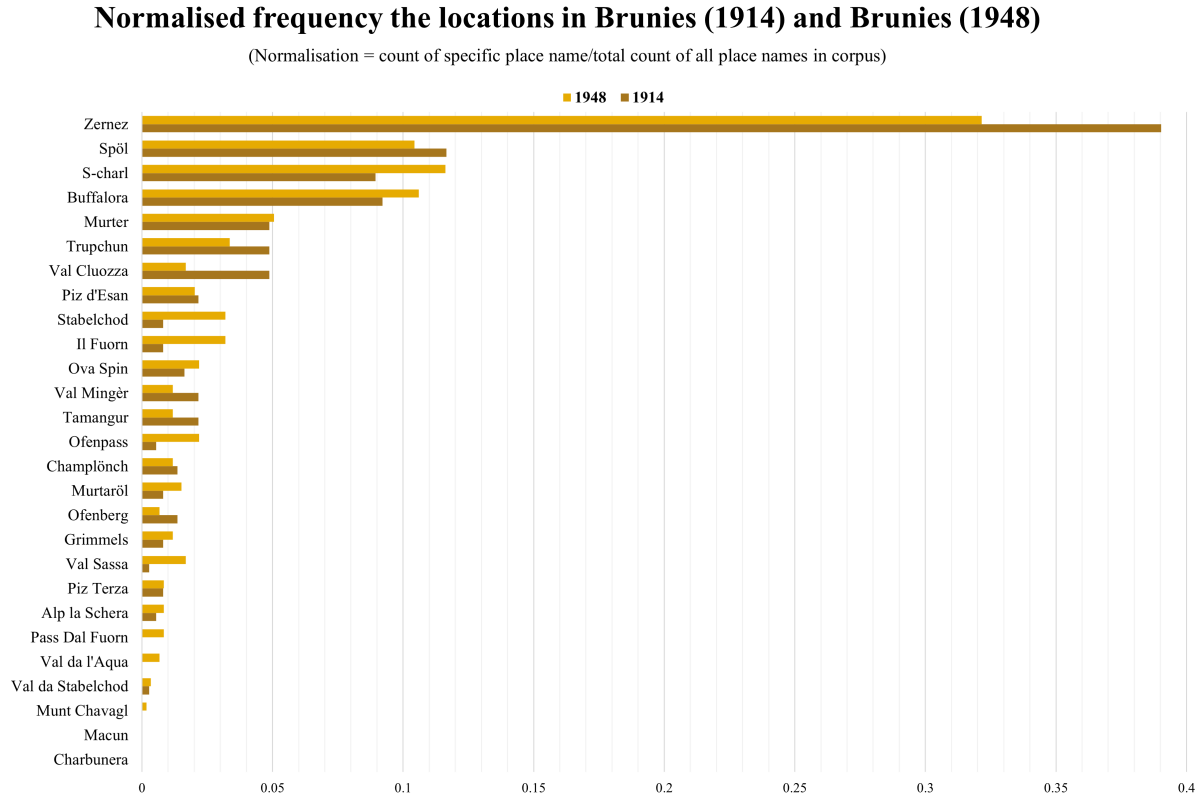


Figure 5.8. Normalised frequency of field and place names in Brunies (1914) and Brunies (1948). The normalisation was conducted by dividing the count of specific place name by the total count of all place names.

However, Zernez remains considerably present in both editions. Yet, why was Val Cluozza so much more important at the time of the park's foundation than in the middle of the century? Among the frequently occurring locations, there are notable declines for Trupchun (-0.02), Spöl (-0.01) and other places with lower frequency, such as Tamangur, Val Mingèr, and Ofenberg, which have also experienced a decline. On the other hand, S-charl has become even more important, but the notable highlights are the increases in Stabelchod, Il Fuorn, Ofenpass, and Val Sassa. In the cases of the one-sided mentions of Pass dal Fuorn, Val da l'Acqua, where a measurement station was installed in 1930, or Munt Chavagl, an interesting event could have been the reason. However, upon a brief examination of the corpus, no truly remarkable events are behind them. Munt Chavagl is only briefly mentioned on p. 100 in connection with the geological composition, which, however, suggests the later installed solifluction measurement station in 1968. Val da l'Acqua also discusses the rock glacier (p. 99), and Pass dal Fuorn is more often referred to as Fuornpass than Pass dal Fuorn, which reduced the number of hits. The decline of Tamangur and the increase in mentions of S-charl are also interesting. The negotiations to make *Tamangur* (Val S-charl) the heart of the National Park failed with the municipality responsible for Scuol in 1910, which could explain the decline of Tamangur, yet the mentions of S-charl increased. Why? A possible reason could be Alp Tavrü, also in Val S-charl, which belonged to the National Park until 1935 but was excluded after the contract expired. This issue could explain the further increase in the mention of S-charl. The next

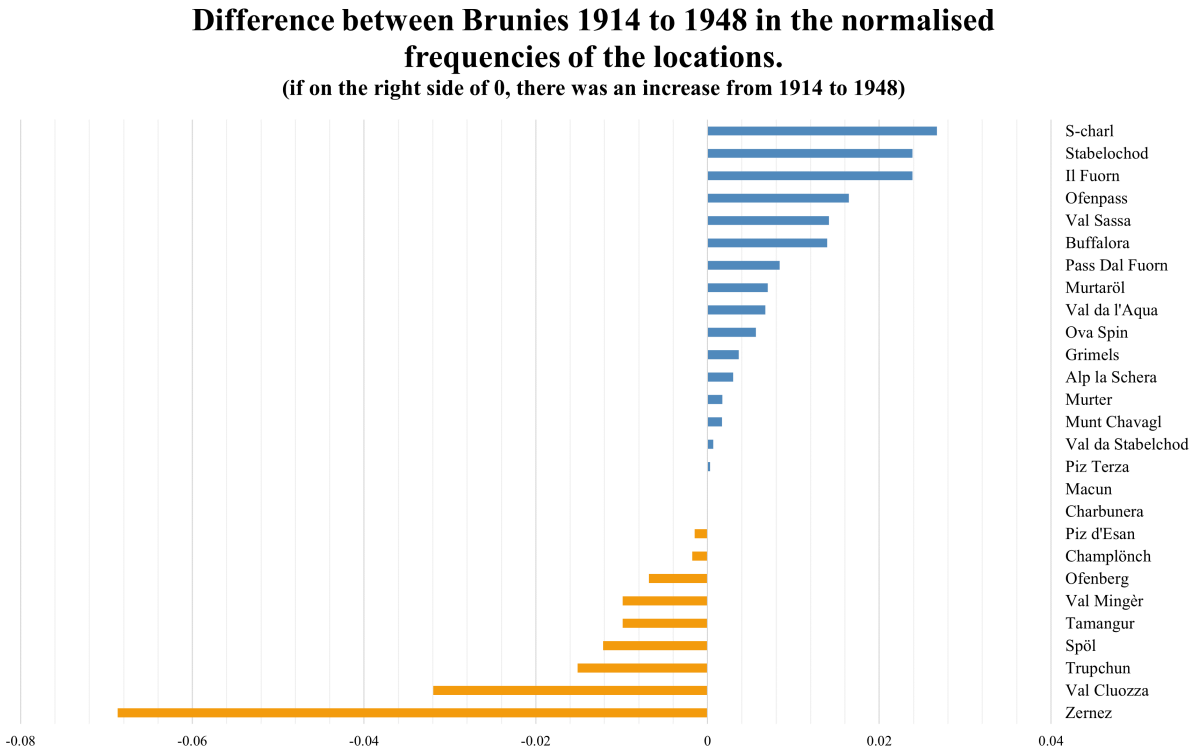


Figure 5.9. Difference between Brunies 1914 and 1948 in the appearances of field and place names. The normalisation was conducted by dividing the count of specific place name by the total count of all place names.

chapter will explore these questions concerning S-charl in depth. Generally, the frequency analysis of the Brunies editions provided little insight into significant historical events. The books' relatively small scope and genre certainly influenced the analysis, as, unlike the Atlas, no clear scientific findings over time are disclosed. Instead, it is more a description of the landscape and a collection of features that characterise these places.

5.5.2 Thematic context

The frequency analysis discovered that the fourth edition contained more extensive coverage of S-charl. In 1914, there was a noticeable presence of *S-charl* alongside park management issues; moreover, connections with animals, forests, and infrastructure were prominent (see Figure 5.10). In the description of the S-charl, it was assumed that there was a high co-occurrence with park management, tourism, infrastructure, and forests and an average co-occurrence with animals. For the most part, the assumptions align with reality, with a few exceptions, such as the high relevance of the animal thematic field in S-charl for both editions. However, Figure 5.2 demonstrates that Brunies' books generally contain a lot of writing about animals. This illustrates that the book's genre also significantly influences how places are written about, not just the location itself. Climate change topics are absent for the first edition and only minimal for the fourth edition. Looking at the fourth edition, a more balanced distribution emerges. Park management's emphasis diminishes, making way for a more frequent mention of infrastructure, possibly reflecting the narrative of construction history.

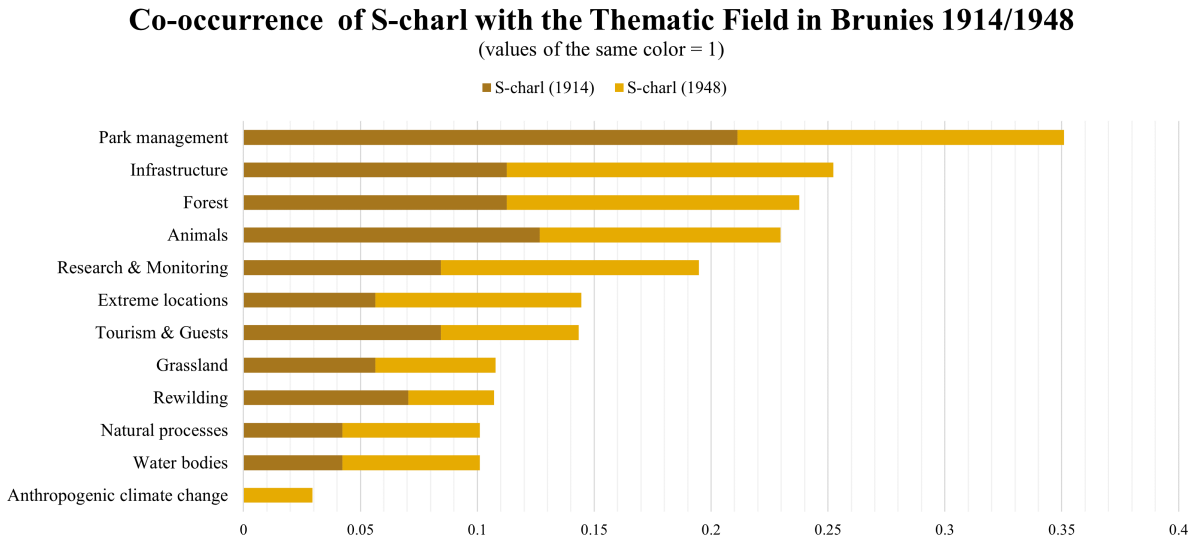


Figure 5.10. Example S-charl: Co-occurrence with the thematic fields in 1914/48.

The diminishing of the park management could indicate the exclusion of Alp Tavrü in 1935. Additionally, forest and extreme habitats have notably increased in presence. The thematic field of climate change in the fourth edition now includes S-charl, whereas rewilding sees a decline. Rewilding as a thematic field has nouns that could overlap with the definition of nature conservation (see subchapter 4.2.4. The higher co-occurrence in the first edition might be explained by the initial promotion of returning to the wilderness when the SNP was founded. Despite the smaller size of the thematic field, this aspect arguably reduces the impact and clarity of the "Rewilding" thematic field. It is debatable whether the concept of rewilding as understood today existed in 1914; rather, S-charl correlates more closely with conservation terms such as renaturation, now bearing a different connotation. Forest, next to park management and animals, is the most relevant thematic field in the corpora overall, mirrored in the S-charl description. It probably refers to the beautiful pine forest in Val S-charl. Examining the nouns that appear in the same paragraphs as S-charl, it is observed that the nouns 'gemeinde' and "bergwerke" are the most frequent nouns in the editions (see Table 5.6). Highlighting the substantial importance of the park management and infrastructure thematic fields.

The over-representation of park management terms in 1914, including municipality, mining, department, and hunters, is noteworthy. The significance of the Swiss stone pine forest, potentially related to discussions around making the Tamangur forest the core of the SNP—a goal that was not achieved—is consistent across editions. The fourth edition's emphasis on S-charl's mining history can be noted by the recurrence of terms like "bergwerke" and "bergbau." Surprisingly, "bär" appears seven times in association with *S-charl* in the first edition. This may indicate the last sighting and killing of a bear in 1904 in the mentioned valley, which might still be relevant in the first book. The frequent rewilding in the first edition is likely due to the "naturschutzkommission". As assumed, there are nouns in the thematic field that particularly pertain to nature conservation. The percentage is so high, even though this word only appears four times, due

to the generally low hit rates with nouns, which result from the infrequent occurrence of place names in both corpora. Upon review in the corpus, the exclusion of Alp Tavrü in 1935 received less attention than initially assumed. It is briefly mentioned on p.30 in the fourth edition. The remaining mentions deal with S-charl's mining history (e.g., p. 95), climatic conditions (e.g., p. 102), or its flora (p. 118). The high frequency of bears mentions in the first book is due to the detailed description of how the killing of the bear took place. The first book described the future park, while the fourth edition precisely documents the developments over 35 years, strictly following the research principle. Drawing context-related conclusions proves challenging with this corpus due to the brevity of the texts and the scarcity of place names, resulting in limited co-occurrences with thematic fields and little room for interpretation.

S-charl 1914		**S-charl 1948**
gemeinde (7)	jägern (2)	bergwerke (4)
bär (7)	beobachten (2)	stationen (4)
naturschusskommission (4)	bergwerke (2)	arvenwald (3)
weiler (4)	dorf (2)	arve (3)
arvenwald (2)	wiesels (2)	gemeinde (3)
wälder (2)	carex (2)	bergbau (3)
waldgrenze (2)	segge (2)	beobachtungen (3)
lärche (2)	gneis (2)	weiler (3)
eidgenossenschaft (2)	schneefall (2)	kirche (3)
bergbau (2)		blockhaus (3)
abteilung (2)		auerhahn (3)

Table 5.6. Example S-charl: Most frequent nouns associated with S-charl in 1914 and 1948.

5.6 Business Reports

5.6.1 Frequency of place names

In Figure 5.11, a similar frequency order of the locations in the Business Reports to Cratschla (see Appendix B). Most hits were made with Zernez (0.22), followed by Trupchun (0.17), seemingly an important location in the Business Reports. Further, Il Fuorn is positioned higher and Spöl lower in the reports than, for example, in Brunies' books (see Figure 5.4). Visible in the plot are dips varying in intensity—sometimes more pronounced, sometimes less—indicating that the frequencies of locations can be divided into five groups. The first group includes the most frequently occurring places by a considerable margin: Zernez and Trupchun. The second group, with frequencies between 0.07 and 0.1, comprises Il Fuorn, Spöl and Stabelchod. After Stabelchod, there's a decline, leading directly to the third group: Ofenpass, Macun, Buffalora, and Val Cluozza, with frequencies between 0.03 and 0.05.

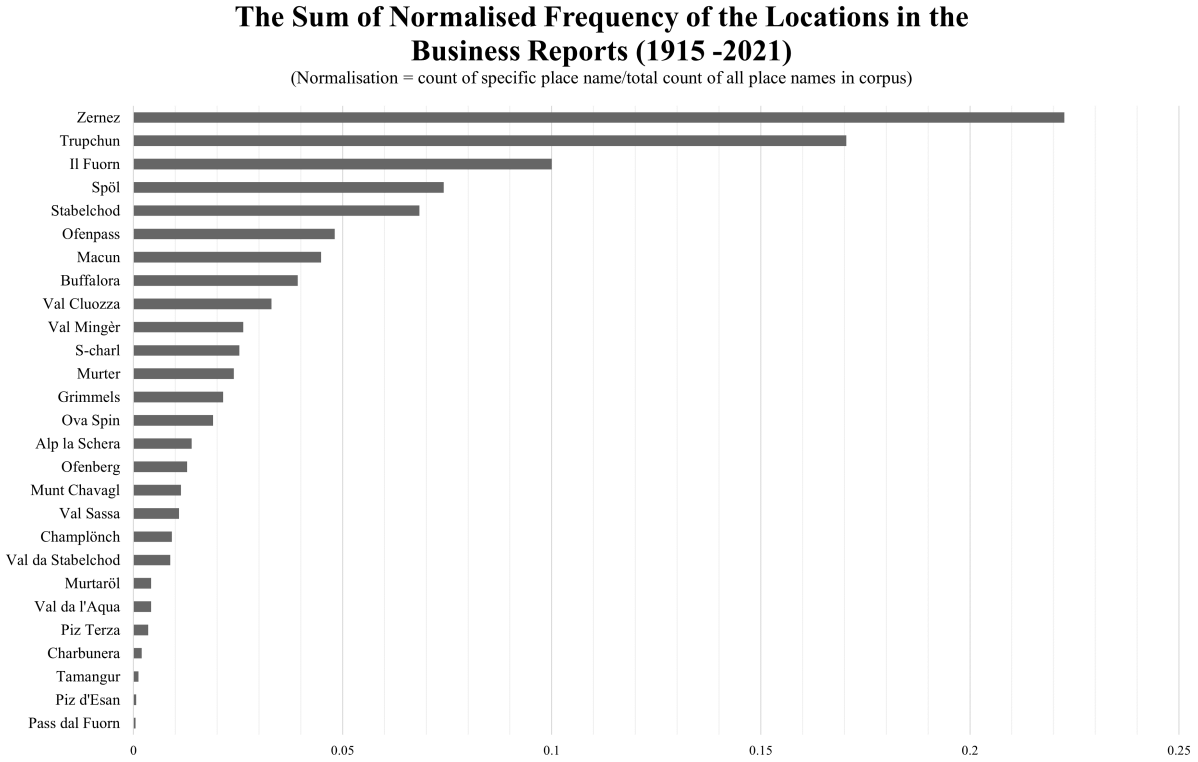


Figure 5.11. This plot shows the sum of normalised appearances over the whole publication period of Business Reports, 1915-2021. Normalisation: Count of Specific Place Name in Publication/ total count of all place names.

The high position of Macun is surprising as the Business Reports expanded over 100 years, and Macun only appeared in the SNP perimeter at the beginning of the century. Val Mingèr, S-charl, Murter, Grimmels and Ova Spin are in the fourth group. These locations seem to occupy all corpora more in the middle ranks. Notable exceptions are S-charl and Murter in the books of Brunies, indicating that those locations were more important in the first decades of the SNP. Ova Spin, though remaining in the middle, seems to rank lower compared to other corpora. In

the lower ranks, we find locations that are familiar to us. Interestingly, the "higher" mention of Charbunera, as Piz d'Esan and Pass dal Fuorn have even lower mentions. A pattern is slowly emerging: mentioning all place names have similar relevance and positions in the graphs across all corpora, except for the books by Brunies, with a few exceptions, such as S-charl. Ova Spin is also positioned lower in the Business Reports than in the other corpora. This is quite surprising, as *Spöl* and *Ova Spin* are often associated with each other. As mentioned, the reports have been available since 1915. In this sense, it spans over 100 years, meaning that finding interesting patterns and creating an overview for 27 field and place names can also become very daunting.

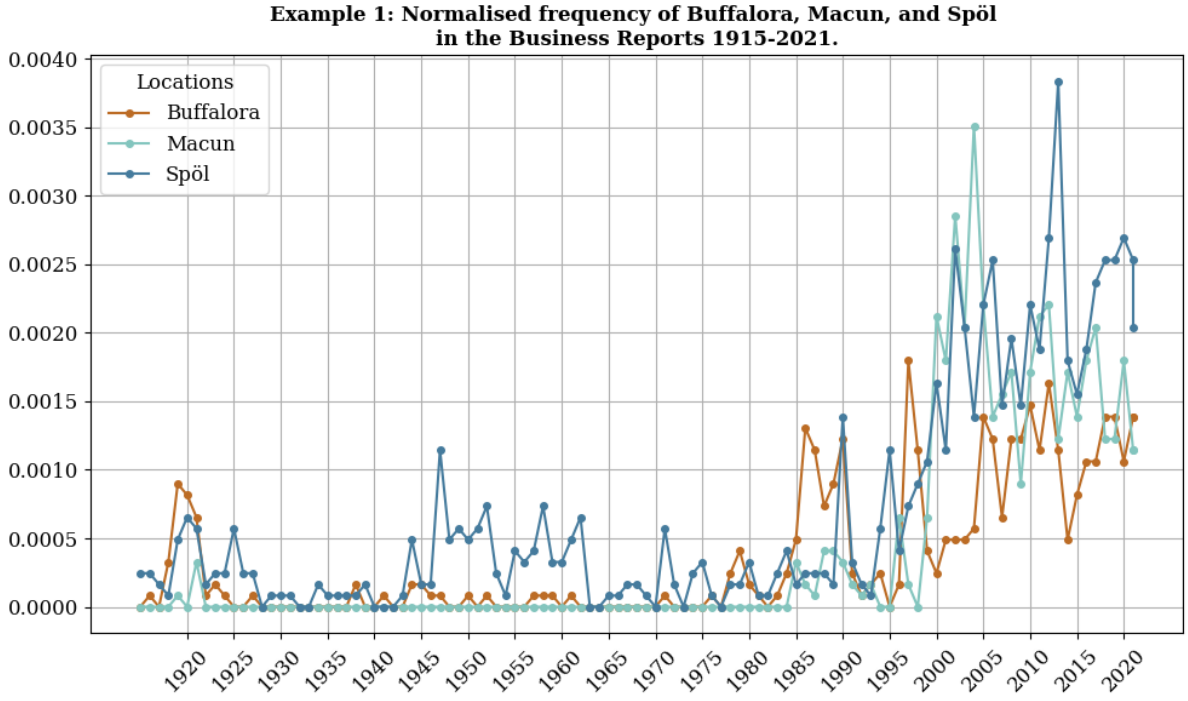


Figure 5.12. Example: Buffalora, Spöl and Macun.

As an example of a temporal progression, only three locations are depicted in Figure 5.12. The notable frequency increase in the 1990s also reflects the frequency behaviour of other locations. This is not surprising since the volume of the editions remarkably expanded and lengthened during the same period. This change in the Business Reports could be further attributed to the change in directorship. In 1990, Klaus Robin introduced park management to a more modern and digital era with Cratschla, which perhaps also contributed to the changes in the scope and content of the annual reports (source: Tamara Estermann). Again, this is a mere assumption. These locations were chosen as an example because strong manifestations were found for Buffalora and Spöl before the 1990s, and for Macun, the opposite occurred, with a peak around the 2000s. As mentioned, it is remarkable that Macun has reached seventh place in frequency within just 21 years, considering the reporting duration. In the case of *Spöl*, an increase is observed precisely in the 1950s, when the debate over a hydroelectric power plant intensified and eventually, in the 1960s, when it was built. Also, around 1920, there was a small

peak for Spöl, which could indicate negotiations with Zernez.
It involved a deal where SNP officials had already agreed to the hydropower project; however, they were allowed to release ibexes on *Piz Terza*. The substantial increase in mentions of Spöl from the 1990s onwards may not only be due to the length of the reports but also started in this period with hydrological experiments, where dynamic, artificial flooding was used to promote the wild, natural habitat of fish. The major peak in 2013 could be attributed to the environmental accident that represented a significant loss of flora and fauna. *Buffalora* hosts a meteorological station installed in the early years of the SNP around 1920. It is unclear why it became that relevant in the 1990s and 2000s. The reports may document the climate situation, which could explain why Buffalora has been frequently mentioned in recent years. However, this would not account for the low frequencies before the turn of the century. *Macun* became relevant in the 2000s due to its incorporation into the SNP. Furthermore, it is also an important location for hydrological research, which takes place there annually. These assumptions are confirmed upon review in the corpus, with Buffalora being associated with the meteorological station around 1920 and constantly referenced in more recent publications with the meteorological measurements. Spöl, all assumptions were fitting: the agreement in 1920, the construction of the hydropower plant in the 60s, and the two accidents in 2013 and 2016. Macun was not only mentioned due to hydrological studies but also regarding the development of the hiking network and the extent to which tourists visit the newly incorporated area. The fact that frequency analysis can shed light on events in the case of Business Reports is very clear. Here, the genre of the document plays an important role, too. Documenting key annual events is the core of the reports, allowing for a timely interpretation. This contrasts with Cratschla (Appendix B.1), where the thematic focus seems much more important, and retrospectives of a location are sometimes described. This characteristic could make the Business Reports the most important corpus of this work.

5.6.2 Thematic context

After reviewing the corpus, the frequency analysis of the Business Reports confirmed the assumptions about the events. Whether this is also visible with the thematic fields will be analysed. Figure 5.13 shows the percentage with which thematic fields the locations *Buffalora*, *Spöl,* and *Macun* generally occur most frequently. As always, these will be compared and discussed with the hypotheses made earlier in the place descriptions. Subsequently, it will be checked whether the reasons for the events in Spöl from 2013, in Macun from 2004 and in Buffalora from 1919-1921, which were already discussed in the frequency analysis in the previous chapter, also appear in the nouns of the thematic fields. This means that only the nouns from the reports of the affected years are specifically extracted to see if they accurately reflect the environmental accident in Spöl in 2013, the elevated frequencies in Macun in 2004 due to hydrological measurements, and the installation of the weather station between 1919 and 1921 with little background knowledge.

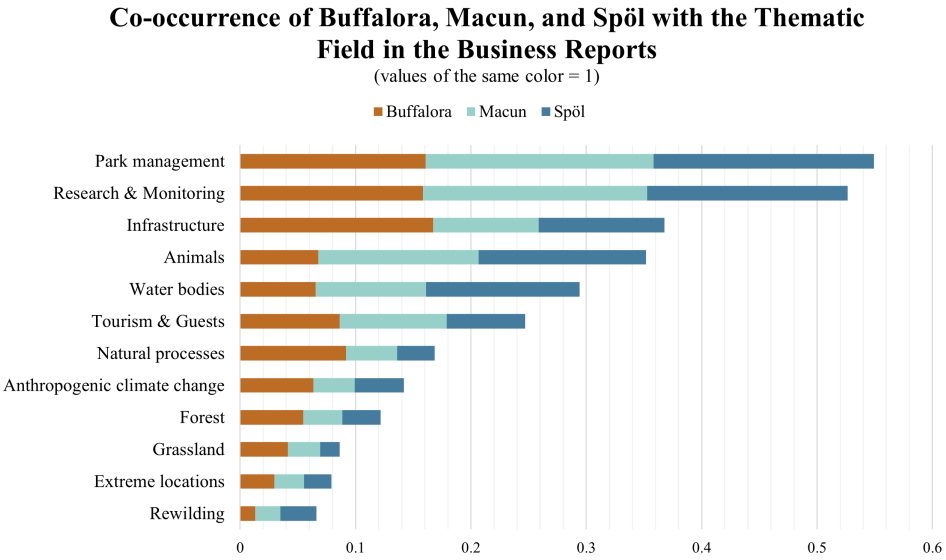


Figure 5.13. Examples Buffalora, Macun, and Spöl: Co-occurrence with the thematic fields.

In the hypothesis regarding *Buffalora*, it was assumed that it is most frequently associated with infrastructure, park management, and research and monitoring, likely due to the meteorological station there. This is greatly visible in the bar chart 5.13 and map 5.14, where the three are the highest-ranking thematic fields co-occurring with Buffalora. Compared to other thematic fields, Buffalora correlates further more with climate change and natural processes, where climate change is probably due to the temperature rise measured over decades, and natural processes might be attributed to heavy rainfall since the 2022 debris flow near Buffalora cannot be captured here as the reports are available until 2021. The increased correlation between tourism and guests could indicate Buffalora is an important visitor arrival location. Surprisingly, Spöl's highest co-occurrence is not with bodies of water but with park management and research & monitoring. This is possibly due to the management of the hydropower plant, its impact on the river, and the hydrological measurements carried out since the 1990s. A high co-occurrence in said thematic fields matched the hypothesis in the place description (3.2.1). Spöl has the highest frequency in the Business Reports compared to the other two locations, followed by the highest co-occurrence values, evident from the absolute numbers on map 5.14. That Spöl is a river is strongly indicated by the underrepresentation of forest, grassland, and extreme habitats and the overrepresentation of the water habitat. The influence of the hydropower plant may explain the high co-occurrence with infrastructure and park management. The research field is strongly represented in all three locations. At Spöl and Macun, there are greater discrepancies between the "Causes of Landscape Change" themes and between the habitats than at Buffalora. Except for the infrastructure, Buffalora appears to be relatively balanced in the fields of co-occurrences. Interestingly, grassland does not occur as frequently at Buffalora, probably because less is written about the Alp Buffalora than about the mountain inn directly on the Ofenpass road or the climate station. Animals are important for Spöl and Macun, likely referring more to aquatic life. Again, low occurrences are seen for causes of landscape change,

"climate change", "natural processes", and "rewilding". It was particularly assumed that natural processes should be more relevant for Macun and Spöl, but this is not evident here. Furthermore, habitats "extreme locations", "forest", and "grassland" have low co-occurrence since Spöl and Macun are water landscapes and Buffalora, beyond the climate station, is not so relevant for habitat research in SNP, as it is located outside the park boundaries. The high share of park management, research and monitoring can also be attributed to the genre of the books.

Co-occurrence of Buffalora, Macun, and Spöl with the Causes of Landscape Change, Habitats, and Animals in the Business Reports
[absolute count]

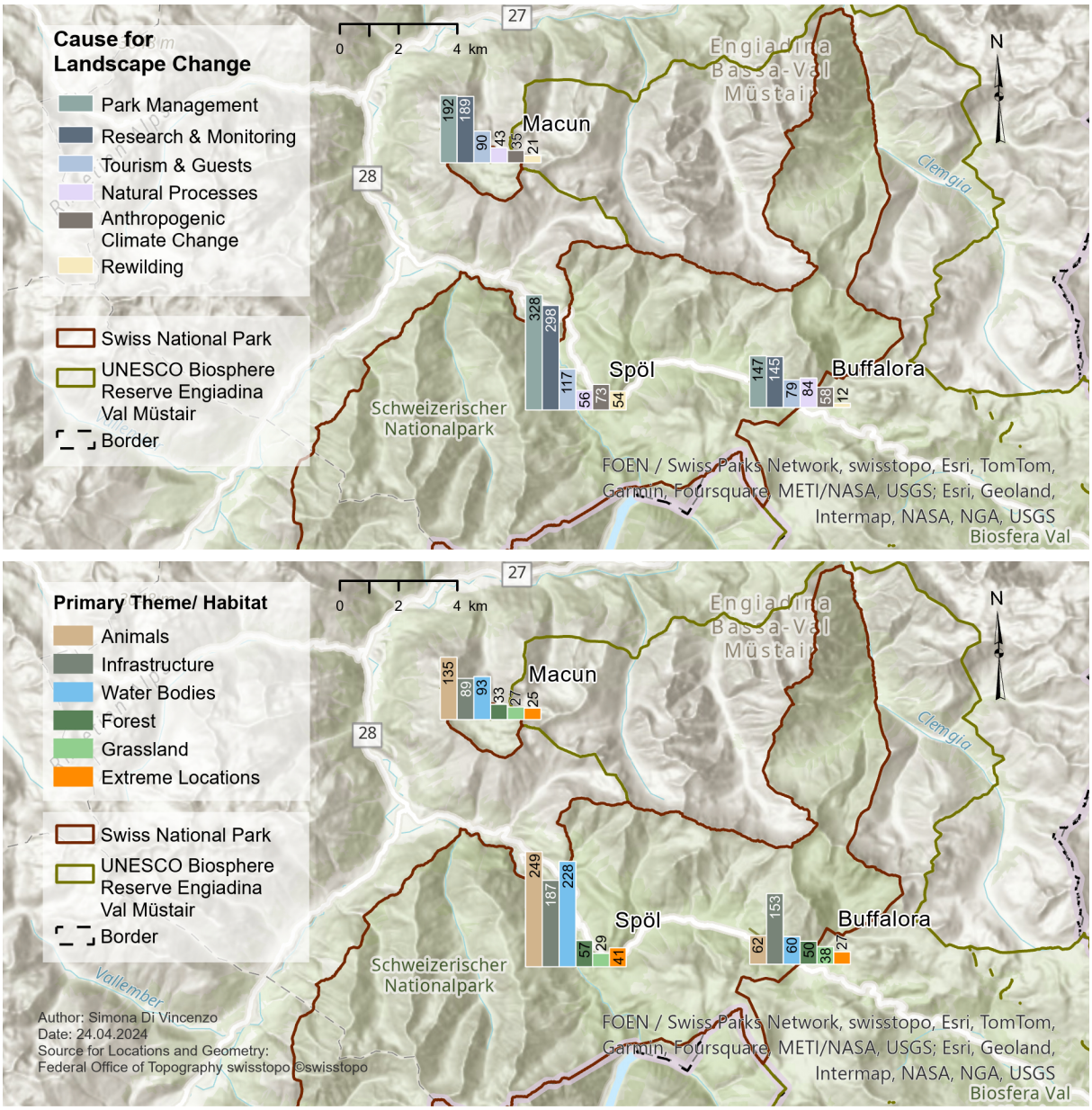


Figure 5.14. Absolute count of Co-occurrence of Causes for Landscape Change and Primary Themes/Habitat with the Locations Buffalora, Macun, and Spöl.

Table 5.7 shows the most common noun occurrences of Spöl, Buffalora, and Macun. The 2013 environmental accident is not reflected in this listing. Although nouns like "fische", "surveillance", "ekw", and "measures" could indicate that due to the accident at the EKW power plant, several fish have died, necessitating increased monitoring and measures. However, from the nouns alone, it is impossible to determine whether they refer to existing measures and monitoring or living fish, which means concluding from this context alone that the incident cannot be definitively determined. At Buffalora, the meteorological station is hinted at only to a limited extent. The associated nouns are rather arbitrary, like "meterologie", "temperature", and "lufttemperatur", which indicate a tendency to a meteorological station. Yet, the installation of one is not visible in the most frequent nouns. The description of Macun in the 2004 report, on the other hand, highlights the high relevance of animals, even more than just aquatic animals; ungulates are also unexpectedly mentioned. The nouns well capture the importance of Macun for scientific research and investigation "monitoring", "recherche", "projekt", and "forschenden".

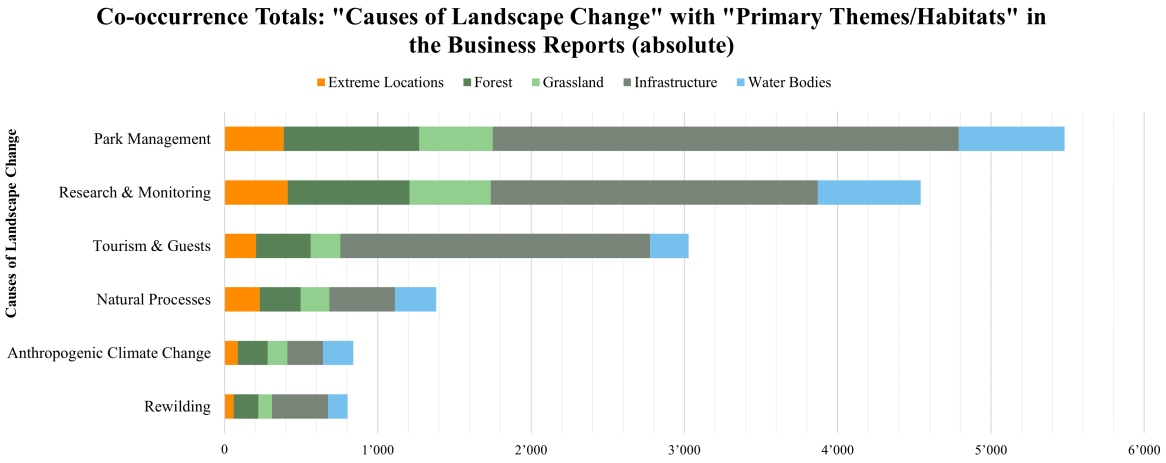
Buffalora 1919-1921	Macun 2004	Spöl 2013
parkwächter (8)	projekt (4)	fische (8)
wegerhaus (8)	monitoring (4)	brut (5)
beobachter (7)	gämsen (4)	surveillance (5)
meteorologie (7)	zähltotal (3)	chamanna (4)
prof (6)	exkursion (3)	hochwasser (4)
pilze (6)	veranstaltungen (3)	laichgruben (3)
betrieb (5)	vernissage (3)	ekw (3)
parkstationen (4)	fische (6)	station (3)
blockhaus (4)	recherche (3)	staumauer (2)
geographie (3)	forschenden (3)	mesures (2)
sonnenschein (3)	biodiversität (3)	revitalisierung (2)
temperatur (2)	geissen (3)	
lufttemperatur (2)	kitze (3)	

Table 5.7. Example: Most frequent nouns associated with Spöl, Buffalora, and Macun in single publications

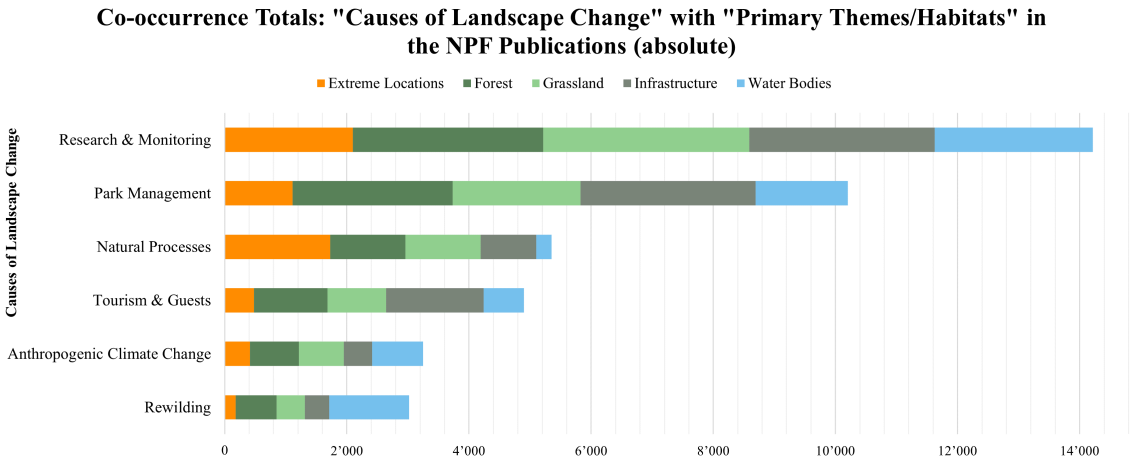
It is noticeable that in the case of the reports, one can conclude well about historical events (provided one has basic knowledge about the locations), and the general thematic field analysis reflects a good picture of the characteristics and habitats of the locations. However, on a micro level, the nouns are not only very vague but also show very low frequency, so this method can say little about the temporal context of the location but teaches more about the prevailing conditions and characteristics.

5.7 Relationship between primary themes and causes of landscape change

Using the Business Reports and NPF publications, this chapter examines the relationship between habitats, i.e., primary themes and causes of landscape change. Figure 5.15 shows both corpora's co-occurrence between the causes and habitats at the paragraph level. The thematic fields covering human influence on habitats are the most frequent in conjunction with habitats, especially in the Business Reports (Figure 5.15a).



(a) Co-occurrence in Business Reports.



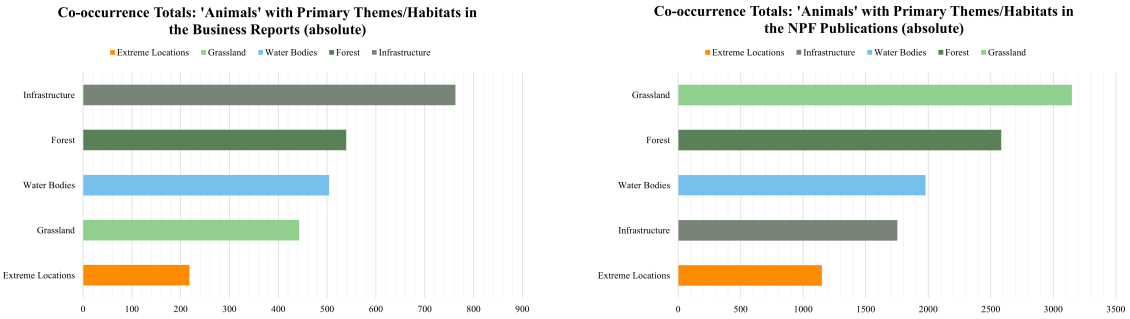
(b) Co-occurrence in NPF publications

Figure 5.15. Co-occurrence in paragraphs of "primary themes" and "causes of landscape change".

It is not surprising that the influence of these causes is particularly significant on infrastructures in the reports since they are man-made. The park management is primarily responsible for their maintenance, requiring installations for monitoring and research projects in the SNP, as well as transport connections, accommodation facilities, and attractions for SNP visitors. That these initiatives are documented in detail in the reports is not astonishing. In addition to infrastructures, human influences (park management and research & monitoring) also appear to impact the "Forest" habitat and bodies of water. Other causes, such as "Natural Processes",

"Climate Change", and "Rewilding", show significantly fewer co-occurrences with the habitats. Here, too, the impact on infrastructures appears to be the highest. Extreme locations strongly co-occur with natural processes. This can be explained by the fact that mudflows, rock streams, and avalanches are common in both thematic fields, potentially enhancing the co-occurrence. In the NPF publications (Figure 5.15b), the most matches with habitats were in the thematic field "Research & Monitoring," which makes sense since it concerns research publications. Nouns from research & monitoring are most often written in the context of grassland and with forest, infrastructure, and water bodies. Somewhat less with extreme habitats. The influence of natural processes is most common at extreme sites. Again, it can be deduced that avalanches, glacier streams, and scree slopes are all natural processes in extreme habitats. There is also much written about natural processes in water bodies and a bit less in grasslands and forests. Anthropogenic climate change and rewilding are fields that do not frequently occur in all habitats due to the thematic fields' size. Park management often co-occurs with infrastructure and forest, likely due to issues with grazing pressure, fire management, and deforestation. Despite the construction of hydroelectric power plants and a turbulent history, park management is not mentioned as frequently with water bodies compared to other fields, possibly due to the publications' focus on research. In Business Reports, as mentioned, the co-occurrence looks different and fits the genre of each corpus more. Overall, the strong relevance of research, park management, and natural processes in the NPF publications is evident. This again shows that the nature of the texts is essential.

5.8 Relationship between primary themes and animals in the Business Reports and NPF publications



(a) Co-occurrence in paragraphs of "primary themes" and "animals" in the Business Reports

(b) Co-occurrence in paragraphs of "primary themes" and "animals" in the NPF publications.

Figure 5.16. Co-occurrence in paragraphs of "primary themes" and "animals".

Figure 5.16 presents the total co-occurrence of the thematic field "Animals" with the habitats in the Business Reports (Figure 5.16a) and NPF publications (Figure 5.16b). Generally, the NPF publications exhibit more co-occurrences, which is also due to the generally higher number of paragraphs in the corpus overall. However, the primary interest lies in which corpus writes more or less about which habitats. It is immediately noticeable that in the NPF publications, animals associated with the habitats of forests and grasslands are an important research field. In contrast, the Business Reports write less about these habitats. The Business Reports seem to focus more on infrastructure and water bodies in connection with animals. As mentioned, how a place is written about depends greatly on the type of text. The NPF publications probably report on hydrological studies. At the same time, the Business Reports may deal more with the installation, documentation, and problematic events related to the infrastructure of the hydroelectric power station and the subsequent consequences for animal mortality. This could explain the frequent co-occurrence of infrastructure and water bodies with animals. Higher co-occurrences in grasslands and forests were anticipated since issues concerning ungulates and the installation and documentation of permanent observation sites also have high importance in the Business Reports. Nevertheless, it seems that all habitats are frequently researched in connection with animals in the NPF publications, the least being with extreme habitats. Still, even here, the thematic field is relatively small in terms of the number of terms compared to the others, which makes the high co-occurrence with animals in the Business Reports even more interesting. To look into more detail about their relationship, Table 5.8 shows what noun pair between animals and habitat are the most frequent terms concerning corpora. The table can be read as follows: On the left side of each word pair are the nouns describing animals, and on the right are their corresponding habitats. The left column shows the most common pairs in Habitats for Business Reports, and the right column shows those in the NPF publications.

Business Reports	NPF Publications
Extreme location	
adler/ firn (11)	bartgeier/ steinadler (28)
tiere/ lawinen (8)	gaier/ grat (11)
bouquetin/ exposition (8)	böcke/ hangneigung (11)
brut/ steinadler (6)	böcke/ schuttflächen (11)
fuchs/ steinadler (5)	geissen/ schuttflächen (11)
Grassland	
tierund/ pflanzenwelt (17)	tiere/ pflanzen (85)
hunde/ beeren (13)	rinde/ moos (37)
hunden/ blumen (12)	gaier-moos (31)
hirsche/ weiden (10)	gaier-curvuletum (31)
vogelwarte/dauerflächen (8)	geissen-weiden (27)
Water Bodies	
wildtierpopulationen/ fischerei (17)	tiere/ wasser (36)
fischbestandsaufnahmen/ fischerei (17)	tiere/ wassermilben (32)
wildtierpopulationen/ fischbestandsaufnahmen (16)	hydracarinen/ wassermilben (25)
fischbestandsaufnahmen/ laichfischfangaktionen (14)	phormidium/ cosmarium (25)
laichfischfangaktionen/ fischerei (14)	larven/ wasser (24)

Business Reports	NPF Publications
Forest	
wildtierpopulationen/ wald (16)	rinde/ holz (55)
vogelwarte/ god (15)	fischweiherquellen/ god (32)
fischbestandsaufnahmen/ wald (14)	hemipterenfauna/ heteropteren (28)
hunde/ wurzeln (13)	tiere/ waldgrenze (25)
tiergesundheit/wald (12)	huftiere-verjüngung (21)
Infrastructure	
hunden/ wege (43)	tiere/ falle (21)
wildtierpopulationen/ wanderwege (16)	rotwild/ markierung (19)
fischbestandsaufnahmen/ wanderwege (14)	weibchen/ stradin (18)
huftiere/ leitung (12)	tiere/ weg (17)
laichfischfangaktionen/ wanderwege (12)	männchen/ stradin (17)

Table 5.8. Top 5 Nouns per Habitat that co-occur with the thematic field "Animals" in the Business Reports and NPF publications.

At extreme locations, we often observe co-occurrences of species such as ibexes and chamois, eagles, and vultures, particularly associated with factors like exposure, slope inclination, or scree areas, appearing in both corpora. There are no significant differences in the contexts of how animals or specific species are discussed between the corpora. However, small differences are particularly noticeable in themes like water bodies, forests, and grasslands. In NPF publications, specialised terms for fauna or flora are often used, such as "Curvultum", "Hydracarina", "Phormidium", and "Hemipteran fauna". On the other hand, reports focus more on stocktaking and animal populations. Domesticated dogs are also relevant in reports, particularly in habitats like infrastructure, forests, and grasslands. One reason might be the use of dogs for game tracking or the prohibition of dogs in the SNP for visitors, which is often mentioned in the reports. In infrastructure contexts, distinctions between males and females are often made, and interestingly, in both cases, the Romansh term "stradin", meaning road, is used. Plus, "paths" and "markings" seem more important, which could indicate studies on ungulates. Also, "God" means forest in Romansh, and it is evident that both corpora use this spelling. Generally, the genre of the corpora can be identified by the choice of words, meaning that the scientific language in the NPF publications and the technical terms in the reports are evident. The animals are always appropriately assigned to their habitats in both corpora, except for dogs, which humans bring into the park.

6. Discussion

This chapter delves deeper into the findings from the computational analysis. Each research question is individually addressed, discussed, and answered. The overarching goal of this thesis was to explore the frequencies and thematic contexts of the place names. It is now determined how effectively the frequency and thematic field analyses have demonstrated political and environmental changes around the locations. Subsequently, the discussion reflects on the choice of methods, their strengths, and challenges, combined with the limitations of this work and the perspective of possible future research.

6.1 Summary of Key Findings

RQ1: How have the frequencies of place names in the Swiss National Park evolved over the past 110 years, and do they reveal political and environmental changes?

The frequency analysis clearly showed that certain places, such as Zernez, Il Fuorn, Trupchun, Stabelchod, Macun, Spöl, and Buffalora, are consistently mentioned more often than others across all corpora. They have maintained high frequencies across various document types, underscoring their central role in the park's geography and management. Conversely, places like Charbunera, Piz d'Esan, Tamangur, and Val da l'Acqua are less frequently mentioned in all corpora. The genre of the corpora influences the importance of certain locations to some extent. Moreover, publication year also plays a role to a certain point, as seen in the books by Brunies, where locations like S-charl received more focus, particularly in the first decades of the SNP. Similarly, Macun gained prominence in newer corpora such as the Atlas and Cratschla. Aggregated frequencies reveal the places most frequently mentioned in different corpora and the locations within the SNP that receive more extensive coverage. The development of these places over time and whether they can reveal political and environmental changes can only be examined by investigating corpora covering specific periods. This includes the Brunies editions for the first few decades, the Business Reports, Cratschla, and the NPF publications. The Business Reports, with examples like Buffalora, Spöl, and Macun, most accurately indicate environmental events, such as the environmental accident 2013 in Spöl. They also highlight political changes, such as Macun being included in the SNP around 2000. Reports are specifically valuable as they document events promptly when they occur. This timely documentation allows for the use of place name frequencies in these reports to indicate

significant events at specific locations reliably. Texts from different genres, such as Cratschla and the NPF publications, do not allow for interpretations of such events based on frequency alone. The editions of Cratschla, heavily influenced by thematic emphases, reveal more about the landscape features and significance for the flora and fauna of a place, covering immediate environmental events only minimally. For instance, following the environmental accident in 2013, a noticeable delay in the frequency of references to the location "Spöl" became apparent in the second issue of 2015, which also addressed the remediation efforts at the hydroelectric power plant. Consequently, events cannot be immediately discerned from the data as they can be from Business Reports. Each publication edition consistently focuses on a specific topic, compiling information about a place from the past to the present. This leads to issues providing a deep thematic focus on certain locations. Yet, it reduces the ability to interpret recent events and reveal a clear change over time in locations based solely on frequency analysis. The same applies to the NPF publications. It was impossible to refer to events as with the Business Reports through frequency analysis, but places particularly relevant for research or heavily researched places were identified. In summary, it can be said that variations in the frequencies of place names over the years have only changed significantly for a few places. Significant places like Zernez, Il Fuorn, Trupchun, and Spöl have been relevant since the park's founding or since their inclusion in the park and continue to be featured in the corpora because they generally show higher frequencies consistently. The same applies to places that appear with medium to low frequency. The fluctuations in frequencies are often limited to a certain range, which extends to the general importance of the place from the perspective of the SNP. Even low frequencies of popular places remain high relative to less frequently mentioned locations. Additionally, factors such as the type of corpora and differences in the length of publications significantly influence the pure frequency and their interpretation, especially when analysing frequency variations over time. For example, frequency increases in the case of the Business Reports successfully pointed to certain historical events. Still, the changes in report length and content at the end of the 1980s had an even greater impact on the frequency of writing about the places. A strong increase in almost all place frequencies was noted from then on. This significant interruption, however, comes from an internal decision by the park management and is not connected to any events at the mentioned locations. The same applies to the NPF publications, which vary greatly in scope, and this still shines through in the calculations and visualisations even after normalising the occurrences of places. To answer the research question precisely, the type and publication year of the corpora strongly influence how the frequencies vary. It can be said that places that were much written in the past are still extensively written about today. That timely documentation reports provide relatively good insight into events and changes occurring on site. Nevertheless, a frequency analysis is insufficient, as the characteristics of the corpora themselves make the extent of interpretation difficult.

RQ2: How do thematic field co-occurrences reflect the known characteristics and any significant changes of the most frequent locations?

The thematic field analysis aims to investigate whether the established thematic fields correlate with the characteristic properties of the studied locations and whether they reveal any changes in the discussion surrounding these places over time. Assumptions regarding potential co-occurrences of the locations with thematic fields such as "Causes of Landscape Change", "Primary Themes/Habitats", and "Animals" were outlined in the background chapter (see 3.2.1). Those assumptions were particularly met in the thematic analysis in the Atlas. The thematic contexts of the locations Ofenpass, Trupchun, and Spöl effectively highlighted these three sites' topographic features and animal habitats and aligned with the assumptions of the background chapter. For instance, it revealed that Trupchun is a key area for ungulate presence and research, Spöl is significant for energy production, and Ofenpass is a crucial route for tourism. This was further supported by the nouns extracted during the analysis. Similarly, the Business Reports for Spöl, Buffalora, and Macun confirm that thematic field analysis effectively captures these locations' physical characteristics and habitats. However, the ability of this method to detect changes in the discourse about these locations remains uncertain. For example, in 2013, the Spöl River experienced an environmental disaster that resulted in numerous fish deaths. Analysis of the nouns extracted from the relevant publications years identified terms like "fisch", "überwachung", "ekw", and "massnahmen". These terms likely relate to the incident at the EKW power plant, which required increased monitoring and intervention. However, it is challenging to definitively assert from the nouns alone whether they relate to the existing conditions or the incident itself, thus preventing a conclusive interpretation based solely on the context. It seems that nouns alone are insufficient to understand the contextual content of locations unless one has background knowledge about the history of these places and the park. The quality of the analysis is impacted not only by the ambiguity of the nouns but also by the type and length of the corpora and the frequency of occurrence of the locations within the texts. As mentioned in the results section of Brunies' books, drawing context-based conclusions solely based on the extraction of nouns is challenging due to the brevity of the text and the low number of mentions of locations and co-occurring nouns. One could argue that the Atlas has a similar scope to Brunies books, yet it focuses more on the SNP's geography, flora, and fauna. It is equipped with numerous maps, focusing more on the same place names, which have more frequency overall. The length and the overall frequency of place names in the Atlas and Brunies' books are quite similar (see Figure 5.2). However, the Atlas tends to concentrate on the same locations, while Steivan Brunies mentions a broader array of places. This broader scope, in turn, results in a lower overall frequency of place names across his works. This results in a less meaningful depiction of the locations' characteristics in the books' case. Moreover, the thematic fields' size and the locations' frequency play an important role, especially when determining whether the thematic fields can capture changes in the discourse about a location over time. For example, the Business Reports from the 1980s are much more extensive than earlier ones, meaning that more locations are mentioned in the newer reports, resulting in higher co-occurrence counts with the thematic fields. However, this makes it difficult to recognise contextual changes over

time, as shorter reports stand out less, resulting in lower matches with the thematic fields. This is further evident in the case of Brunies. Yet, the report's brevity alone does not exclude the co-occurrence of themes characteristic of the location; here, the nature of the text becomes even more relevant. A concise description from the Atlas likely provides a more comprehensive account of the geographical context of the location compared to a section from Brunies' books or a brief Business Report. In general, and to answer this research question, it can be said that the co-occurrence of the thematic fields with the studied place names provides a good depiction of the characteristics and habitats of the locations, especially with the Atlas, Cratschla, Business Reports, and NPF publications. However, the nouns are vague on a micro level and often show low frequency. Hence, this method reveals little about the temporal context but teaches more about the prevailing conditions and characteristics of the locations.

RQ3: How do variations in the depiction of place names across different types of documents reflect differing perceptions and priorities regarding the Swiss National Park?

Most of the examined corpora are characterised by a scientific focus. Notably, this includes the NPF publications, and the Atlas also delivers geographic, historical, and scientific insights into the SNP. Cratschla provides detailed and extensive updates and reports on developments within the National Park; each issue often centred around a specific theme. Although the Business Reports cover scientific findings and research developments, they primarily focus on reviewing the year's developments, goals, and performance, putting science in a secondary role. This is particularly evident in Chapter 5.8, where the co-occurrence between animals and habitats in Business Reports and NPF publications was examined. The use of language in NPF publications is more scientific, especially in the use of nouns, as many Latin terms are employed. Another distinct outlier is Brunies' books, classified as literary works. These books narrate the founding history of the SNP. Previous research questions suggested that the nature of the corpora significantly influences the frequency with which locations are discussed and the thematic fields co-occur. Frequency analysis has shown that all corpora periodically focus extensively on the same locations. Common sites across all corpora include Zernez, Spöl, Buffalora, Trupchun, Il Fuorn, and Stabelchod, noted for their management roles (Spöl, Zernez), scientific studies of flora and fauna (Buffalora, Trupchun, Stabelchod), and tourism hotspots (Zernez, Il Fuorn). Brunies' books uniquely highlight places like S-charl and Murter more frequently. This focus might be influenced by the era, as S-charl, especially the Val S-charl with the Tamangur pine forest, was intended to be the park's centrepiece. The frequent mention of the field name Murter may further reflect this period's influence. Field names are crucial to the SNP's history, often emerging when people identify and point to locations without using geographical coordinates, which was a typical practice in the SNP, where park rangers created and enforced the use of field names. Steivan Brunies, the park director then, often mentioned "Murter" in his writing, possibly demonstrating his connection to the place. This illustrates that the types of documents and the writers and editors affect the depiction of place names and the perception of places

within their contemporary context. Scientific reports and publications often highlight locations focusing on their ecological value or research importance, as seen with Trupchun, Stabelchod, and Spöl. Furthermore, Brunies' books demonstrate that the perception of place names can be a product of time and may reflect evolving customs practices.

RQ4: Which approaches and techniques are most suitable for extracting place names from historical sources such as journals, atlases, reports, and books in PDF?

Several approaches and techniques have proven most suitable when extracting place names from textual sources such as journals, atlases, reports, and books in PDF format. Working with OCR is crucial, particularly those tailored to handle different document formats. The use of Pytesseract has proven effective for this thesis. The open-source access, easy implementation, and wide language selection have reinforced the program choice. For instance, the results of the older documents that utilise Fraktur fonts have shown satisfying results. In this thesis, greater emphasis was placed on the post-processing of documents than on the preparation of the documents. The goal was to quickly convert the corpora into a readable format, resulting in the rather overlooked preparation work. Several points would have simplified working with this corpora. For example, divide the documents into language groups before reviewing them. Particularly, the NPF publications were originally PDF documents divided into chapters, which would have made separating French and German documents easier and could potentially have improved the results. While it was possible to specify multiple languages for processing, which was especially helpful for documents written in several languages, a prior separation may have minimised errors. Nevertheless, adaptive strategies are necessary for individual texts and corpora differences. Documents often present challenges such as complex layouts (for example, Cratschla) or deteriorated text quality (older documents in the corpora), which standard OCR technology may not adequately address. Combining OCR with manual verification processes ensures higher accuracy. This approach integrates the precision of human oversight to correct and refine the OCR output, which is necessary for texts where characters may fade. To find all the place names relevant to this work in the documents, post-processing steps for the corpora and crucial adjustment steps for the place names themselves are indispensable. Collecting and compiling all possible spellings of the place names is crucial. This includes collecting old and colloquial names for the places in the languages of the corpora. It is also beneficial to look up how the places are written in the corpora. Frequency analysis has shown that removing any special characters (such as «,», «;», «()», «""») and converting the texts and place names to lowercase minimises errors. This step is also relevant for thematic field analysis. Rules and expressions defining which places should be counted must be defined beforehand. For places like «Il Fuorn», «Pass dal Fuorn», «Stabelchod», and «Val da Stabelchod», context rules were introduced in the code to prevent double counting. For Il Fuorn, "Il" must precede it; for Pass dal Fuorn, "Pass dal" must precede it; and the same goes for Val da Stabelchod with "Val da", "val", or "v.". To ensure the effectiveness of the extraction techniques, it is important

to use evaluation metrics such as precision, recall, and the F1 score. These metrics assess the accuracy and efficiency of the extraction processes. Precision measures the correctness of the extracted place names, recall assesses the completeness of the extraction in capturing all relevant names, and the F1 score balances precision and recall. Regular evaluation using these metrics allows for ongoing refinement of the context rules and expressions, ensuring that the most effective methods are applied based on the type and condition of the documents. Overall, the method used in this work proved satisfactory, especially when dealing with many different documents. The extraction of place names (and thematic terms) was refined through evaluation steps, highlighting their importance. However, in retrospect, it would have been beneficial to invest more time in the pre-processing phase, particularly as the work involved dealing with texts that vary in style, length, and language.

6.2 Methodological Reflections and Limitations

The computational text analysis aimed to explore the thematic and temporal context of place names within the Swiss National Park across various corpora. This method enabled the detailed extraction and classification of nouns associated with the specific place names, providing a satisfying understanding of thematic fields. Despite the challenges encountered, OCR technology was instrumental in processing large documents, essential for managing and analysing extensive data sets. However, the OCR process faced significant challenges, particularly with historical texts and documents with complex layouts, occasionally resulting in inaccurate text recognition. C. Patel, A. Patel, and D. Patel (2012) noted that the text recognition of coloured images in Tesseract is more limited than grayscale. In hindsight, converting documents to grayscale might have generated better results. Images, tables, and page breaks undoubtedly affected the corpora's paragraph structure and text flow, influencing the contextual thematic analysis. Consequently, place names and nouns that typically occur together may have been separated and omitted from this analysis. The same goes for co-occurrences that were counted but would not go together. Different tools can convert PDFs to text, like pdftotext or PyPDF2 (Yildiz, Kaiser, and Miksch, 2005), which includes the option to handle page breaks. The use of Tesseract proved to be a good choice in this thesis. It can extract text from images and captions, yet, in certain cases, this ability adds to the disruption of the text flow. Another possibility is to write custom scripts and adapt them to the characteristics of each corpus, like page margins, headers, and footers, and remove them accordingly (Welbers, Atteveldt, and Benoit, 2017). The same goes for tables and images, yet most tools concentrate on improving the ability to extract text in this feature rather than excluding them (Yildiz, Kaiser, and Miksch, 2005; Fan and Kim, 2015; Li et al., 2020). With this in mind, it may have been beneficial to remove all the pages with cartographic images, for instance, in the Atlas, as the amount of work involved is still manageable. As mentioned, a longer preprocessing stage may improve the conversion from PDFs to text and possibly also lead to fewer text disruptions and better paragraph structure. The unsupervised, manual categorisation of nouns into thematic fields introduced a subjective element that could potentially influence classification outcomes. Additionally, excluding nouns appearing

four times or fewer may have led to overlooked co-occurrences. Complicating this issue, the method of extracting nouns as capitalised words did not account for linguistic variations, causing potential omissions in languages like French, Italian, and English, where nouns are not typically capitalised. Overall, the approach captured the relationship between thematic fields and place names, albeit roughly. However, the thematic fields differ greatly in content size, with the largest category containing 2'517 (Park management) and the smallest 360 (Rewilding) nouns, leading to less co-occurrence in a paragraph with the locations, affecting especially less mentioned locations. The fields such as "Anthropogenic Climate Change", "Natural Processes", and "Rewilding" furthermore are not well represented with only the noun as an entity. To capture the essence of these concepts, more context of the nouns is needed to contribute to interpreting their significance. For example, "Verschmutzung" could mean contamination or pollution, with the latter attributing more to climate change. However, as a stand-alone word, it lacks the necessary information to get categorised as such. This further leads to categorising nouns wrongly or discarding nouns completely that capture the concepts well. Additionally, nouns that were categorised across multiple fields have led to an increase in co-occurrence between these fields, especially impacting the results discussed in chapters 5.7 and 5.8, where the relationships among habitats, causes of landscape change, and animals are examined. Even though the thematic field analysis provides a good depiction of the characteristics and habitats of the locations, the extraction of the nouns applied in the examples of locations shows that without background knowledge about the places themselves, the meaning of the thematic context is vague and questionable. Moreover, nouns and place names that occur often together do not necessarily mean they are meaningful combinations. Precision and Recall proved valuable for optimising the code used in counting co-occurrences. However, the challenge of detecting the same noun appearing multiple times within a single paragraph indicates that differences in the size of thematic fields may be minimised within paragraphs, potentially influencing the outcomes of the thematic analysis. To improve the outcome of the analysis, incorporating the term frequency-inverse document frequency (TF-IDF) could have been a favourable addition. This method indicates how important a word is within a document in a corpus. The higher the score, the more important and unique the word is for a document (Havrlant and Kreinovich, 2017). Utilising this method or other statistical measurements would have positively influenced the results of this thesis and made them more reliable. Meaningful combinations could have been identified with the help of statistical measures, such as the t-score. It assesses whether the co-occurrence of two words is statistically significant and is primarily used to find meaningful collocations. These are pairs or groups of words that frequently appear together more often than would be expected by chance, or mutual information (MI), which measures the mutual dependencies between two words (Brezina, 2018). Using such advanced metrics, the study could have achieved a deeper and more nuanced understanding of textual relationships than using frequency alone. The methodological process highlights the complexity involved in quantifying, analysing, and understanding written natural language. It has particularly revealed that it is not a linear approach but is characterised by trial and error. For instance, the eval-

uation using precision and recall showed that the immediate presence of a special character next to a noun transforms it into a new word, which may not be immediately recognised and classified since a computer reads the sequence of letters, characters, and numbers. The subtle difference between the typographically correct apostrophe <’> and its replacement <‘> makes a significant impact when searching for names like Piz d’Esan using the replacement and with no matches found. These seemingly trivial aspects are often overlooked in method application because humans logically assume them. In general, developing and refining the methodology for this thesis provided valuable insights into the potential and limitations of computational text analysis. It underscored the importance of selecting appropriate tools and techniques according to the specific requirements and challenges of the research data.

7. Conclusion

This thesis employs computational text analysis techniques to thoroughly explore the evolution of place names within the Swiss National Park over the past 110 years across six corpora: Atlas, Steivan Brunies' 1st and 4th editions, Business Reports, and NPF publications. The frequency of 27 place names was studied and depicted over time (specifically for Cratschla, Business Reports, and NPF publications). Furthermore, the thematic context of the most frequently mentioned or notably interesting locations per corpus was analysed in greater depth, focusing on seven locations. To examine the thematic context, twelve thematic fields were established by extracting all nouns from paragraphs and sentences containing a place name, which resulted in six fields composing the "Causes for Landscape Change", five the "Primary Themes/Habitat", and one dedicated to the park's fauna, named "Animal". These fields facilitated the analysis of the thematic context for each corpus's most frequently mentioned or unique locations. The study aimed to explore how the frequencies of these place names have changed over time and whether their fluctuations suggest political and environmental changes. It also examined how thematic fields co-occur with frequently mentioned locations, their relationship to the characteristics of these locations, and their reflection of potential political and environmental events. Key findings indicate that places such as Zernez, Il Fuorn, Trupchun, Stabelchod, Macun, Spöl, and Buffalora consistently appeared frequently across all corpora, underscoring their important role in the park's geography and management. In contrast, places like Charbunera, Piz d'Esan, Tamangur, and Val da l'Acqua were mentioned less frequently. The Business Reports provided the most accurate reflections of environmental events. They highlighted political changes, while texts with a strong thematic focus, such as Cratschla and NPF publications, were less effective at pinpointing specific events. The analysis also revealed that the genre and publication year of the corpora significantly influenced the frequency variations, suggesting that frequency alone is not a sufficient metric to capture the evolving narratives of place names. Regarding thematic analysis, the calculated co-occurrences accurately depicted the locations' characteristics and habitats. Nonetheless, the ability of thematic analysis to detect discourse changes over time remains uncertain, as indicated by the difficulty in interpreting whether terms extracted from the publications relate to ongoing conditions or specific incidents. By comparing results across all corpora, variations in the depiction of place names show that scientific publications and reports focus on the ecological values or research importance of locations. At the same time, literary works like those by Brunies might emphasise historical aspects and personal connections to places. The insights derived from this study highlight the connection between geography and language and

underscore the possibilities of examining place names to capture and reflect political and environmental changes. Regarding methodological approaches, OCR, specifically Pytesseract with open-source access, easy implementation, and wide language selection, has proven effective for converting PDFs to readable documents. Post-processing steps for the corpora texts, including removing special characters and converting them to lowercase, further minimised errors. Implementing context rules and expressions (e.g., Il Fuorn vs. Pass dal Fuorn) to extract place names is essential to prevent double-counting. To ensure the effectiveness of the extraction techniques, it is important to use evaluation metrics such as precision, recall, and the F1 score. These metrics assess the accuracy and efficiency of the extraction processes. However, the methodology used, and therefore the results, are affected by challenges and limitations. The OCR process encountered challenges, especially with historical texts and documents with complex layouts, occasionally resulting in inaccurate text recognition. Furthermore, images, tables, and page breaks undoubtedly affected the corpora's paragraph structure and text flow, influencing the contextual thematic analysis. Pre-processing steps like dividing the documents into language groups before reviewing them, transforming them into greyscale for better text recognition, or removing all the pages with cartographic images would have reduced text disruption and improved paragraph structure. Furthermore, the manual categorisation of nouns into thematic fields introduced a subjective element that could affect the classification outcomes. Moreover, nouns and place names often occurring together do not necessarily mean they are meaningful combinations, demonstrating that frequency alone is limited as a measure. To improve the depth of the analysis, incorporating the term frequency-inverse document frequency (TF-IDF), t-score, or mutual information (MI) could have been a beneficial addition. In conclusion, while this thesis has uncovered valuable insights into how place names in the Swiss National Park have evolved and are written about, there were some challenges, particularly with OCR and how the thematical fields were categorised. To refine this research, better pre-processing strategies should be included. This might mean creating tailored OCR solutions that can handle the quirks of older texts or using more detailed layout analyses for documents with complex formats. Integrating more sophisticated statistical methods like TF-IDF, t-score, or mutual information would also be beneficial. These approaches and tools would improve the quality of data extraction and allow for a more comprehensive and precise exploration of how place names and their associated narratives evolve in the researched corpora. Looking ahead, these improvements promise endless possibilities to broaden and deepen the exploration, offering more detailed insights into their stories within the studied corpora. This thesis has merely scratched the surface of exploring place names within the Swiss National Park. Continuing to work with this data, including the suggested improvements, could unlock an even deeper understanding of the significance the place names hold for the environment and the history of the Swiss National Park.

Bibliography

Alvarez, R. Michael, ed. (2016). *Computational social science: discovery and prediction*. Analytical methods for social research. Cambridge, United Kingdom New York, NY, USA Port Melboure, Australia Delhi, India Singapore: Cambridge University Press. 327 pp. ISBN: 978-1-107-10788-5 978-1-107-51841-4.

Asi, Abedelkadir et al. (2015). "Simplifying the reading of historical manuscripts". In: *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 826–830.

Atteveldt, Wouter van and Tai-Quan Peng (2018). "When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science". In: *Communication Methods and Measures* 12.2, pp. 81–92. ISSN: 1931-2458. DOI: 10.1080/19312458.2018.1458084.

BAFU/WSL (2022). *Landschaft im Wandel. Ergebnisse aus dem Monitoringprogramm Landschaftsbeobachtung Schweiz (LABES)*. Umwelt-Zustand 2219. Bundesamt für Umwelt (BAFU); Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft (WSL), p. 53.

Baur, Bruno and Thomas Scheurer, eds. (2014). *Wissen schaffen: 100 Jahre Forschung im Schweizerischen Nationalpark*. 1. Aufl. Nationalparkforschung in der Schweiz Bd. 100,1. Bern: Haupt Verl. 391 pp. ISBN: 978-3-258-07862-5.

Bicke, Hans (2016). "ortsnamen.ch – das Internetportal der Schweizer Ortsnamenforschung". In: *Sprachspiegel (SchweizerischenVerein für die deutsche Sprache SVDS)* 5, pp. 147–153.

Brezina, Vaclav (2018). "Semantics and Discourse: Collocations, Keywords and Reliability of Manual Coding". In: *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press, pp. 66–101. ISBN: 978-1-107-12570-4. DOI: 10.1017/9781316410899.003.

Brunies, Steivan (1914). *Der Schweizerische Nationalpark*. 1st ed. Basel: Frobenius AG.

— (1948). *Der Schweizerischer Nationalpark*. 4th ed. Basel: Benno Schwabe & Co. Verlag.

Bundesamt für Kultur BAK (2024). *Zernez*. ISOS-Bundesinventar der schützenswerten Ortsbilder der Schweiz von nationaler Bedeutung. URL: https://www.gisos.bak.admin.ch/sites/2313 (visited on 04/23/2024).

Bundesamt für Landestopografie swisstopo (Jan. 8, 2024). *Geografische Namen der Schweiz*. Bundesamt für Landestopografie swisstopo. URL: https://www.swisstopo.admin.ch/de/geografische-namen-der-schweiz (visited on 04/13/2024).

Bundesamt für Umwelt (Mar. 25, 2024). *Schweizerischer Nationalpark*. URL: https://www.bafu.admin.ch/bafu/de/home/themen/thema-landschaft/landschaft--fachinformationen/landschaften-nationaler-bedeutung/paerke-von-nationaler-bedeutung/schweizerischer-nationalpark.html (visited on 03/25/2024).

Cambria, Erik and Bebo White (2014). "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]". In: *IEEE Computational Intelligence Magazine* 9.2, pp. 48–57. ISSN: 1556-603X. DOI: 10.1109/MCI.2014.2307227.

Chesnokova, Olga and Ross S. Purves (2018). "From image descriptions to perceived sounds and sources in landscape: Analyzing aural experience through text". In: *Applied Geography* 93, pp. 103–111. ISSN: 0143-6228. DOI: 10.1016/j.apgeog.2018.02.014.

Cratschla: Informationen aus dem Schweizerischen Nationalpark (1992). URL: https://nationalpark.ch/about/cratschla/.

Dande, Abhay A. and Dr. M. A. Pund# (2023). "A Review Study on Applications of Natural Language Processing". In: *International Journal of Scientific Research in Science, Engineering and Technology*, pp. 122–126. ISSN: 2394-4099, 2395-1990. DOI: 10.32628/IJSRSET2310214.

Der Bundesrat - Das Portal der Schweizer Regierung (Oct. 19, 2021). *Giftquelle im Idyll*. URL: https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-85501.html (visited on 04/13/2024).

Dilsaver, Lary M. (2009). "Research Perspectives on National Parks". In: *Geographical Review* 99.2. Publisher: Routledge _eprint: https://doi.org/10.1111/j.1931-0846.2009.tb00430.x, pp. 268–278. ISSN: 0016-7428. DOI: 10.1111/j.1931-0846.2009.tb00430.x.

Engadin Tourismus AG (2024). *Auf Expedition im «Wilden Osten»*. Engadin. URL: https://www.engadin.ch/de/natur/nationalpark/ (visited on 04/21/2024).

ENPK/SCNAT (1995a). "75 Jahre Steinbock im Nationalpark". In: *Cratschla: Informationen aus dem Schweizerischen Nationalpark* 3.1. Ed. by Federal National Park Commission (ENPK) and SCNAT Research Commission of the SNP. URL: https://nationalpark.ch/about/cratschla/ (visited on 04/13/2024).

— (1995b). "Huftiere - und botanische Vielfalt im Nationalpark". In: *Cratschla: Informationen aus dem Schweizerischen Nationalpark* 3.2. Ed. by Federal National Park Commission (ENPK) and Research Commission of the SNP (SCNAT).

ENPK/Stiftung Schweizerischer Nationalpark Zernez (1915). *Schweizerischer Nationalpark Geschäftsberichte*.

Fan, Miao and Doo Soon Kim (2015). "Open-domain Table Detection Using Large-scale PDF Files without Annotation". In: URL: https://api.semanticscholar.org/CorpusID:11967133.

Haller, Heinrich, Antonia Eisenhut, and Rudolf Haller (2013). *Atlas des Schweizerischen Nationalparks: die ersten 100 Jahre*. 2. Auflage. Nationalpark-Forschung in der Schweiz. Series Number: 99/1 Series: Nationalpark-Forschung in der Schweiz. Bern: Haupt Verlag. 247 pp. ISBN: 978-3-258-07801-4 978-3-258-07902-8.

Hamad, Karez and Mehmet Kaya (2016). "A Detailed Analysis of Optical Character Recognition Technology". In: *International Journal of Applied Mathematics, Electronics and Computers* 4 (Special Issue-1), pp. 244–244. ISSN: 2147-8228. DOI: 10.18100/ijamec.270374.

Hausmann, Anna et al. (2020). "Understanding sentiment of national park visitors from social media data". In: *People and Nature* 2.3. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pan3.10130, pp. 750–760. ISSN: 2575-8314. DOI: 10.1002/pan3.10130.

Havrlant, Lukás and Vladik Kreinovich (2017). "A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation)". In: *International Journal of General Systems* 46, pp. 27–36. URL: https://api.semanticscholar.org/CorpusID:298072.

Heikinheimo, Vuokko et al. (2018). "Digital Imaginations of National Parks in Different Social Media: A Data Exploration". In: Publisher: [object Object]. DOI: 10.5281/ZENODO.1472745.

Hirschberg, Julia and Christopher D Manning (2015). "Advances in natural language processing". In: *ARTIFICIAL INTELLIGENCE* 349.6245, pp. 261–266.

Hotel Parc Naziunal Il Fuorn (2024). *Il Fuorn • Hotel im Schweizer Nationalpark*. Hotel Parc Naziunal Il Fuorn. URL: https://www.ilfuorn.ch/ (visited on 03/26/2024).

Islam, Noman, Zeeshan Islam, and Nazia Noor (2017). "A Survey on Optical Character Recognition System". In: *ArXiv* 10.2. URL: https://api.semanticscholar.org/CorpusID:1040908.

Jeon, Kahyun et al. (2024). "Dynamic building defect categorization through enhanced unsupervised text classification with domain-specific corpus embedding methods". In: *Automation in Construction* 157, p. 105182. ISSN: 0926-5805. DOI: 10.1016/j.autcon.2023.105182.

Khanbhai, Mustafa et al. (2021). "Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review". In: *BMJ Health & Care Informatics* 28. Publisher: BMJ Publishing Group Ltd Section: Original research, e100262. ISSN: 2632-1009. DOI: 10.1136/bmjhci-2020-100262.

Kupper, Patrick (2012). *Wildnis schaffen: eine transnationale Geschichte des Schweizerischen Nationalparks*. Nationalpark-Forschung in der Schweiz Band 97. Bern Stuttgart Wien: Haupt Verlag. 371 pp. ISBN: 978-3-258-07719-2.

Landolt, Christoph (2016). "Fast so vielfältig wie die Flurnamen: Die Schreibweise auf den Landeskarten. Eidgenössische Regeln und kantonale Ausnahmen." In: *Sprachspiegel (Schweizerischen Verein für die deutsche Sprache SVDS)* 5.

Lei Wu, Steven C H Hoi, and Nenghai Yu (2010). "Semantics-Preserving Bag-of-Words Models and Applications". In: *IEEE Transactions on Image Processing* 19.7, pp. 1908–1920. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2010.2045169.

Li, Minghao et al. (2020). "TableBank: Table Benchmark for Image-based Table Detection and Recognition". English. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Ed. by Nicoletta Calzolari et al. Marseille, France: European Language Resources Association, pp. 1918–1925. ISBN: 979-10-95546-34-4.

Lorimer, Jamie et al. (2015). "Rewilding: Science, Practice, and Politics". In: *Annual Review of Environment and Resources* 40.1, pp. 39–62. ISSN: 1543-5938, 1545-2050. DOI: 10.1146/annurev-environ-102014-021406.

Lund, Katrín Anna, Ludovic Moncla, and Gabriel Viehhauser (2022). "Glacial Narratives: How Can They Be Captured?" 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. 93–108. ISBN: 978-1-911529-56-9. DOI: 10.5334/bcs.e.

Luria, Sarah and Ricardo Campos (2022). "Greening a Post-Industrial City: Applying Keyword Extractor Methods to Monitor a Fast-Changing Environmental Narrative". In: *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis*. Ed. by Ross S. Purves, Olga Koblet, and Benjamin Adams. Ubiquity Press, pp. 109–132. ISBN: 978-1-911529-56-9. DOI: 10.5334/bcs.f.

Mangachena, Joy Rumbidzai and Catherine Marina Pickering (2021). "Implications of social media discourse for managing national parks in South Africa". In: *Journal of Environmental Management* 285, p. 112159. ISSN: 0301-4797. DOI: 10.1016/j.jenvman.2021.112159.

Mittal, Rishabh and Anchal Garg (2020). "Text extraction using OCR: A Systematic Review". In: *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 357–362. DOI: 10.1109/ICIRCA48905.2020.9183326.

Ojo, Adegboyega et al. (2024). "Prioritising national healthcare service issues from free text feedback – A computational text analysis & predictive modelling approach". In: *Decision Support Systems* 181, p. 114215. ISSN: 0167-9236. DOI: 10.1016/j.dss.2024.114215.

Panzeri, Stefano, Cesare Magri, and Ludovico Carraro (2008). "Sampling bias". In: *Scholarpedia* 3, p. 4258. URL: https://api.semanticscholar.org/CorpusID:39834664.

Papagiannopoulou, Eirini and Grigorios Tsoumakas (2020). "A review of keyphrase extraction". In: *WIREs Data Mining and Knowledge Discovery* 10.2, e1339. ISSN: 1942-4787, 1942-4795. DOI: 10.1002/widm.1339.

Patel, Chirag, Atul Patel, and Dharmendra Patel (2012). "Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study". In: *International Journal of Computer Applications* 55.10, pp. 50–56. ISSN: 09758887. DOI: 10.5120/8794-2784.

Publikationsreihe Nationalpark-Forschung in der Schweiz (1918). Red. by Akademie der Naturwissenschaften Schweiz (SCNAT). 1-109. URL: https://scnat.ch/de/id/2ksbC.

Purves, Ross S., Olga Koblet, and Benjamin Adams (2022). *Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis.* Publication Title: Ubiquity Press. London: Ubiquity Press. ISBN: 978-1-911529-58-3 978-1-911529-59-0 978-1-911529-56-9 978-1-911529-57-6. DOI: 10.5334/bcs.

Santana, Brenda et al. (2023). "A survey on narrative extraction from textual data". In: *Artificial Intelligence Review* 56.8, pp. 8393–8435. ISSN: 1573-7462. DOI: 10.1007/s10462-022-10338-7.

Schatz, Ferdinand, Thomas Scheurer, and Benno Steiner (2012). *Ergebnisse 70 Jahre Gewässerforschung im Schweizerischen Nationapark*. Vol. 98. Publikation aus der Reihe Nationalpark-Forschung in der Schweiz. 120 pp.

Schorta, Andrea (1991). *Wie der Berg zu seinem Namen kam - Kleines Rätisches Namenbuch mit zweieinhalbtausend geografischen Namen Graubündens*. 3rd ed. Chur: Terra Grischuna Verlag. 160 pp.

Schweiz Tourismus (2024). *Seenplatte Macun*. Schweiz Tourismus. URL: https://www.myswitzerland.com/de-ch/reiseziele/seenplatte-macun/ (visited on 03/26/2024).

schweizerfluss.ch (2024). *Spöl*. schweizerfluss.ch. URL: https://schweizerfluss.ch/spoel/ (visited on 03/26/2024).

Schweizerischer Nationalpark (Mar. 25, 2024a). *Grundlagen & Impressionen*. Parc National Suisse. URL: https://nationalpark.ch/about/nationalpark/ (visited on 03/25/2024).

— (2024b). *Besuchen*. parc naziunal svizzer. URL: https://nationalpark.ch/besuchen/ (visited on 04/16/2024).

— (2024c). *Erlebniswege - Kinderpfad Champlönch*. parc naziunal svizzer. URL: https://nationalpark.ch/besuchen/erlebniswege/ (visited on 03/27/2024).

— (2024d). *Eröffnung des komplett neu inszenierten Nationalparkzentrums in Zernez*. parc naziunal svizzer. URL: https://nationalpark.ch/medianews/eroeffnung-nationalparkzentrum/ (visited on 04/21/2024).

— (2024e). *Klimawandel –Auswirkungen im Nationalpark*. parc naziunal svizzer. URL: https://nationalpark.ch/wissenschaft/klimawandel/ (visited on 03/26/2024).

— (2024f). *Murgang Buffalora - Von tabula rasa zu neuem Leben*. parc naziunal svizzer. URL: https://nationalpark.ch/fr/forschung/murgang-buffalora/ (visited on 03/26/2024).

— (2024g). *Schmelzra*. parc naziunal svizzer. URL: https://nationalpark.ch/besuchen/schmelzra/ (visited on 03/26/2024).

Schweizerischer Nationalpark (2024h). *Val da Stabelchod auf neuen Wegen*. parc naziunal svizzer. URL: https://nationalpark.ch/medianews/val-da-stabelchod-auf-neuen-wegen/ (visited on 04/23/2024).

— (2024i). *Zustand der Fischpopulation im Spöl*. parc naziunal svizzer. URL: https://nationalpark.ch/medianews/zustand-der-fischpopulation-im-spoel/ (visited on 03/26/2024).

Schweizerisches Idiotikon (2024). *ortsnamen.ch*. ortsnamen.ch. URL: https://search.ortsnamen.ch/de/record/803298244/ (visited on 03/27/2024).

Sergiacomi, Carlotta et al. (2022). "Exploring National Park Visitors' Judgements from Social Media: The Case Study of Plitvice Lakes National Park". In: *Forests* 13.5. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 717. ISSN: 1999-4907. DOI: 10.3390/f13050717.

Swiss National Park Data Center (2024). *Publications*. parc naziunal svizzer. URL: https://www.parcs.ch/snp/publications.php (visited on 04/10/2024).

Teles da Mota, Vanessa and Catherine Pickering (2021). "Geography of Discourse about a European Natural Park: Insights from a Multilingual Analysis of Tweets". In: *Society & Natural Resources* 34.11. Publisher: Routledge, pp. 1492–1509. ISSN: 0894-1920. DOI: 10.1080/08941920.2021.1971809.

Tourismus Engadin Scuol Samnaun Val Müstair AG (2024a). *Bergunterkunft Ova Spin*. Engadin Scuol Zernez. URL: https://www.engadin.com/de/bergunterkunft-ova-spin-0 (visited on 03/27/2024).

— (2024b). *Nationalpark: Val Spöl*. Engadin Scuol Zernez. URL: https://www.engadin.com/de/touren/nationalpark-val-spoel (visited on 04/13/2024).

— (2024c). *Ova Spin - Il Fuorn (Nationalpark)*. Engadin Scuol Zernez. URL: https://www.engadin.com/de/touren/ova-spin-il-fuorn-nationalpark (visited on 04/13/2024).

WegWandern.ch (Mar. 16, 2019). *Nationalpark*. URL: https://wegwandern.ch/graubuenden/nationalpark-wandern/ (visited on 04/21/2024).

Welbers, Kasper, Wouter van Atteveldt, and Kenneth Benoit (2017). "Text Analysis in R". In: *Communication Methods and Measures* 11, pp. 245–265. DOI: https://doi.org/10.1080/19312458.2017.1387238.

Wikipedia (Mar. 28, 2024). *Zernez*. In: *Wikipedia*. Page Version ID: 243528780. URL: https://de.wikipedia.org/w/index.php?title=Zernez&oldid=243528780 (visited on 04/13/2024).

Yildiz, Burcu, Katharina Kaiser, and Silvia Miksch (2005). "pdf2table: A Method to Extract Table Information from PDF Files." In: pp. 1773–1785.

Zhang, Edmond and Michael Mayo (2010). "Improving Bag-of-Words model with spatial information". In: *International Conference of Image and Vision Computing New Zealand*. International Conference of Image and Vision Computing New Zealand (IVCNZ). Queenstown, New Zealand, pp. 1–8. ISBN: 978-1-4244-9631-0 978-1-4244-9629-7. DOI: 10.1109/IVCNZ.2010.6148795.

A. Appendix

Alp la Schera
The Alp la Schera, located at an elevation of 1755 m a.s.l., attracts over 10'000 visitors annually, making it a highly popular tourism hotspot (Baur and Scheurer, 2014, p. 309). Since the park's establishment, two avalanches occurred nearby, in 1917 and 1978, destroying large parts of the standing forest (Baur and Scheurer, 2014, p. 331). An increase in the Alpine marmot population was noted on the high mountain pasture in the 1990s (H. Haller, Eisenhut, and R. Haller, 2013, p. 140). The vegetation is similar to that found in Stabelchod. I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Park Management, Tourism & Guests
- Primary Themes/Habitat: Forest, Grassland
- Animals: high co-occurrence (especially ungulates)

Champlönch
Champlönch translates to "long field," formed by meltwater channels at the edge of the glacier (Baur and Scheurer, 2014, p. 126),(H. Haller, Eisenhut, and R. Haller, 2013, p. 65). From 1985 to 1988, long-term ornithological observations were conducted in the mountain pine forest of Muottas Champlönch. Champlönch has 5 permanent areas rich in mountain pines (Baur and Scheurer, 2014, pp. 195, 222). At Muottas Champlönch, the capercaillie population at the end of the 20th century was surprisingly high despite unfavourable forecasts. This is because the hill will become increasingly overgrown, thus no longer serving as a refuge for the species (H. Haller, Eisenhut, and R. Haller, 2013, p. 132). An exciting tourist attraction in Champlönch is the children's educational trail, launched in 2009. With an app from the Swiss National Park, users receive an auditory signal and information about animals and features of the location at 10 different spots. This approach eliminates the need for installing physical educational signs (H. Haller, Eisenhut, and R. Haller, 2013, p. 76)(Schweizerischer Nationalpark, 2024c). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Park Management, Research & Monitoring, Tourism & Guests
- Primary Themes/Habitat: Infrastructure, Forest, Grassland
- Animals: high co-occurrence (especially ungulates)

Charbunera

The place name resp. field name "Charbunera" indicates that charcoal was produced here before the establishment of the National Park (Tourismus Engadin Scuol Samnaun Val Müstair AG, 2024b). Additionally, the name is a remnant from the past, as the location of Charbunera is often explained in relation to other place names. The forest section at Charbunera shows signs of damage caused by ungulates (ENPK/SCNAT, 1995a). Black woodpeckers have also been spotted in this area (ENPK/Stiftung Schweizerischer Nationalpark Zernez, 1915). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes
- Primary Themes/Habitat: Forest
- Animals: high co-occurrence (especially ungulates)

Grimmels

The place name "Grimmels" appears in various forms, close to each other and within the SNP. There are Foppa da Grimmels (a ravine), Muottas da Grimmels (peak/elevation), Murteras da Grimmels (alpine pasture), and Margun Grimmels (Schweizerisches Idiotikon, 2024). In the search results, not only the specific name Grimmels was considered but also word parts occurring before or after it. However, the texts often refer to it simply as Grimmels. The name Grimmels could derive from "grumus" (Latin for mound, hill) (H. Haller, Eisenhut, and R. Haller, 2013, p. 65). Like the Alp Trupchun and Stabelchod, the alpine pasture has been shaped by decades of livestock farming (H. Haller, Eisenhut, and R. Haller, 2013, p. 92). In the 1930s, permanent plots were fenced to study deer grazing (Kupper, 2012, p. 219). Phenological studies are conducted, as well as research on insects and animal populations (Baur and Scheurer, 2014; H. Haller, Eisenhut, and R. Haller, 2013, p. 359). It is a moderately visited area, from which one can enjoy a wide view of the Ofenpass from Alp Grimmels (Baur and Scheurer, 2014, pp. 296, 309). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Park Management, Research & Monitoring
- Primary Themes/Habitat: Grassland, Infrastructure
- Animals: high co-occurrence

Munt Chavagl
The summit of Munt Chavagl is particularly important for the national park due to its climate station, which has been measuring solifluction since 1968. This represents the longest continuous data series recording earth flows. The climate station continuously measures air, surface, and soil temperatures, snow depth, wind speed, and radiation. The station was upgraded in 1995 and expanded from 40 to 100 movement markers. (Baur and Scheurer, 2014, pp. 131, 133). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Park Management, Research & Monitoring
- Primary Themes/Habitat: Extreme Locations, Infrastructure
- Animals: low co-occurrence

Murtaröl
Murtaröl is the diminutive form of Murter, meaning sheep pasture (Schorta, 1991). This, however, leads to a similar situation where the place name is frequently found in Grisons. In the Val Müstair alone, there are nine reference points for Fuorcla Murtaröl, Alp Murtaröl, Vadret da Murtaröl, and Aua da Murtaröl (Schweizerisches Idiotikon, 2024). A measurement station named Murtaröl is located at the southeastern entrance of the SNP near Buffalora (H. Haller, Eisenhut, and R. Haller, 2013, p. 28). Historical avalanche barriers are found at Murtaröl (H. Haller, Eisenhut, and R. Haller, 2013, p. 76). In 1961, Murtaröl was incorporated into the SNP (Kupper, 2012, p. 118). Due to its proximity to Zernez, Murtaröl is an important location for visitors wishing to quickly detour into the SNP. The view of the Val Cluozza is also rewarding. I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring
- Primary Themes/Habitat: Extreme Locations, Infrastructure
- Animals: low co-occurrence

Murter
Murter, in Romansh, means "mortar" (Mörser) and refers to high mountain pastures or sheep pastures (H. Haller, Eisenhut, and R. Haller, 2013, p. 56). This place name is widespread in Switzerland for many locations fitting this description. There is Murter in the municipality of Schanf (west outside the SNP), Murter ridge near Surses, Murter in the Val Müstair, Murter (with the Alp and Fuorcla) belonging to Zernez located in the SNP, and finally Murter Dadora and Murter Dadaint in the municipality of Scuol (Schweizerisches Idiotikon, 2024). This illustrates how place names are commonly used in Switzerland but can also lead to misunderstandings if one is not physically present at the location. Since the corpora thematically deal with the SNP, we assume the references pertain to Murter Dadora/Dadaint in Scuol and the Murters associated with Zernez. The alpine pasture Murter is basin-shaped at an elevation of 2463 meters above sea level and is also home to Alpine marmots and sometimes even ptarmigans (H. Haller, Eisenhut,

and R. Haller, 2013, pp. 138, 140). Some deadwood can also be found on the pasture, and many fossil shells and snails have been discovered due to the alternating layers of clay, marl, and limestone (H. Haller, Eisenhut, and R. Haller, 2013, pp. 44, 86).
I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Research & Monitoring
- Primary Themes/Habitat: Extreme Locations, Infrastructure
- Animals: high co-occurrence

Ova Spin
The Lai Dad Ova Spin is a reservoir located in the municipality of Zernez and is operated by the Engadine Power Plants. The 73-meter-high dam was constructed between 1965 and 1968 and is just outside the SNP perimeter. However, the compensation basin extends up to 1 km into the SNP (H. Haller, Eisenhut, and R. Haller, 2013, p. 14). The catchment area of the basin covers 1116 km², including the catchment area of the Spöl, which flows into the Lai Dad Ova Spin. In addition to many tributaries, two stream branches, Ova Spin Dadoura and Ova Spin Dadaint, share the same name as the lake. A hamlet above the basin on the Ofenpass road with a bus stop also bears the same name (Tourismus Engadin Scuol Samnaun Val Müstair AG, 2024c). The Ova Spin fortification, restored and supplemented in 1933, is considered a national military historical monument and an interesting feature for tourists (Baur and Scheurer, 2014, p. 64). During the planning in 1961, adjustments were made to protect the SNP. The compensation basin was significantly reduced from 28 million cubic meters to 6.5 million cubic meters of usable capacity (H. Haller, Eisenhut, and R. Haller, 2013; Kupper, 2012, pp. 60, 242). Another fire event, similar to that at Il Fuorn, occurred in 1964 at Lai Dad Ova Spin, highlighting the lack of a fire management concept in the SNP area (H. Haller, Eisenhut, and R. Haller, 2013, p. 178). The compensation basin is also relevant for research concerning mercury content and sediment deposits (H. Haller, Eisenhut, and R. Haller, 2013, pp. 180, 206). The situation and measures for fish populations and stocking are similar to those at the Spöl. I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring, Tourism & Guests
- Primary Themes/Habitat: Infrastructure, Water Bodies
- Animals: high co-occurrence (aquatic creatures)

Piz d'Esan

Piz d'Esan is a mountain peaking at 3127 m a.s.l. (H. Haller, Eisenhut, and R. Haller, 2013, p. 28). At the time of the park's founding, there were glacier areas between Piz d'Esan and Val da l'Acqua, which have disappeared for decades (ENPK/Stiftung Schweizerischer Nationalpark Zernez, 1915). After the first ibexes were released into Val Trupchun in 1934, they have been spotted over the years at Piz d'Esan as well (ENPK/SCNAT, 1995b). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural processes
- Primary Themes/Habitat: Extreme Locations
- Animals: average co-occurrence

Piz Terza

The peak of the mountain is at 2684 m a.s.l. The mountain is primarily associated with sightings of ibexes, chamois, and golden eagles, particularly due to its steep slopes. On June 20, 1920, four male and three female ibexes were released into the wild (H. Haller, Eisenhut, and R. Haller, 2013, p. 158). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural processes
- Primary Themes/Habitat: Extreme Locations
- Animals: high co-occurrence

Stabelchod & Val da Stabelchod

Stabelchod primarily refers to the alp at an elevation of 2291 meters above sea level and the valley behind it, *Val da Stabelchod*. Stabelchod also has a history of ore smelting, similar to Il Fuorn and Alp Trupchun (H. Haller, Eisenhut, and R. Haller, 2013, p. 52). Like Grimmels and Alp Trupchun, the meadows were created by human activities such as grazing cows and sheep. The transition from livestock to deer grazing nearly doubled the variety of plant species (H. Haller, Eisenhut, and R. Haller, 2013, p. 92). This was studied using a permanent observation plot of 127 m², established in 1919. In Stabelchod, regular surveys of breeding birds, marmots, plants, and insects are conducted and analysed (H. Haller, Eisenhut, and R. Haller, 2013, pp. 122, 140). However, it is assumed that it will take several centuries before subalpine pastures like Stabelchod and Alp la Schera will reforest (Baur and Scheurer, 2014, p. 150). Until then, it serves as a research area to examine how meadow areas with livestock farming before the park's establishment change compared to deer utilization afterwards (Baur and Scheurer, 2014, p. 158). In the alluvial plain and scree landscape of *Val da Stabelchod*, 26 young bearded vultures were released between 1991 and 2007 (H. Haller, Eisenhut, and R. Haller, 2013, p. 126). Deer are also increasingly found in the forest in the Stabelchod valley. Stabelchod is an important location for research and, following Trupchun, an important tourist meeting point. The hut in Stabelchod, with its seating areas and rest stop, invites visitors to linger (H. Haller, Eisenhut, and R. Haller, 2013, p. 166). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring, Tourism & Guests
- Primary Themes/Habitat: Grassland, Infrastructure
- Animals: high co-occurrence

Tamangur

Tamangur belongs to the municipality of Scuol, located in the upper Val S-charl and means "hut of the miners." Several place names are describing the area, such as Tamangur Dadora (Outer-Tamangur), Tamangur Dadaint (Inner-Tamangur), God Tamangur (forest or "The forest back there"), Murters da Tamangur (high mountain pasture), Zondra da Tamangur (Tamangur- pine forest) (Tourismus Engadin Scuol Samnaun Val Müstair AG, 2024a), and Foppa da Tamangur (Tamangur valley basin). Tamangur is a moor and Swiss pine forest landscape. Particularly, the area of God da Tamangur is characterised by an extraordinary density and stands up to the forest line. The Swiss pines can be dated back up to 800 years. For these reasons, it has been a natural forest reserve since 2007. The Alp Dadaint is still being farmed today, whereas Dadora is no longer used. As mentioned, the Swiss pine forest was of great interest to the SNP at its foundation but was never incorporated (H. Haller, Eisenhut, and R. Haller, 2013, p. 58). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management
- Primary Themes/Habitat: Forest
- Animals: average co-occurrence

Val Cluozza

In 1908, it was determined that Val Cluozza should be transformed into a nature reserve (Kupper, 2012, p. 71). The 21.74 km² valley officially became part of the SNP in 1910, making it the first area to be included. A lease agreement lasting 25 years was signed with the municipality of Zernez and was extended by 100 years upon its expiration (Kupper, 2012, pp. 72, 118). The Cluozza hut, built in 1910, accommodated park rangers and visitors (H. Haller, Eisenhut, and R. Haller, 2013, p. 218). It underwent a complete renovation in 1993 and was finally placed under cantonal monument protection in 2021. Like Il Fuorn, this area was also used for mining and clear-cutting (Baur and Scheurer, 2014, p. 76). Research continues in the region, particularly phenological observations, such as the measurement of dust deposition on plant growth in 1920 or observations of the game (Jagdwild) in 1924, specifically focusing on ibex (Kupper, 2012; Baur and Scheurer, 2014, pp. 8, 96, 202, 239). Val Cluozza provides habitat for game and species such as the adder or the black grouse (H. Haller, Eisenhut, and R. Haller, 2013, p. 132). The suspicion of a bear hibernating in the Val Cluozza-Falcun section during the winter of 2007/2008 was a significant event (H. Haller, Eisenhut, and R. Haller, 2013, p. 146). Generally, it is a rather barren area with mountain pine stands and deadwood (H. Haller, Eisenhut, and R. Haller, 2013, pp. 82, 86). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring, Tourism & Guests
- Primary Themes/Habitat: Forest, Infrastructure, Grassland
- Animals: high co-occurrence

Val da l'Acqua

Two valleys away from Val Sassa, we find Val da l'Acqua, home to another rock glacier. The two valleys share much in their measurement history; however, unlike Val Sassa, an annual measurement series in Val da l'Acqua only started in 1930 instead of 1917 (Baur and Scheurer, 2014, p. 381). Another fascinating phenomenon is the spring's origin. Since 1999, the Ova Val da l'Acqua has been an example of an undisturbed alpine watercourse in water monitoring. The Ova da l'Acqua is fed by the local rock glacier (Baur and Scheurer, 2014, p. 283). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Park Management, Research & Monitoring
- Primary Themes/Habitat: Extreme Locations, Infrastructure, Water Bodies
- Animals: low co-occurrence

Val Mingèr

Val Mingèr has been part of the SNP since 1911. After the contract expired 25 years later in 1935, it was renewed in 1937 (Kupper, 2012, p. 118). Historically, the valley was important for livestock summer grazing and mining, which is also evident in the origin of its name. In Romansh, "minier" means miner, indicating the historical significance of the valley in mining (Baur and Scheurer, 2014; Schorta, 1991, p. 71). Over 300 years of mining have left their mark; the barren landscape shows that the forest area was likely cleared multiple times (Baur and Scheurer, 2014, p. 73). The landscape change in Val Mingèr can be seen and measured using aerial photographs from 1946 and 2000. Changes in canopy cover were detected in 55% of the forested areas (H. Haller, Eisenhut, and R. Haller, 2013, p. 188). Phenological observations were conducted in Val Mingèr from 1994 to 2008 but were discontinued as park rangers visited the valley less frequently. There are still some permanent vegetation plots (194 Baur and Scheurer, 2014; H. Haller, Eisenhut, and R. Haller, 2013, p. 184). In 1920, the barren mountain landscape became an iconic visual representation of the national park. It remains a popular hiking area for visitors, though not as relevant as Trupchun (76 Baur and Scheurer, 2014; H. Haller, Eisenhut, and R. Haller, 2013, p. 164). In addition to the flora, the H. Haller, Eisenhut, and R. Haller (2013) also shows the presence of butterflies, various insect species, snow hares, and Alpine marmots, which can be seen in Val Mingèr. I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring
- Primary Themes/Habitat: Extreme Locations, Infrastructure, Forest, Grassland

- Animals: high co-occurrence

Val Sassa

In 1917, the research commission assigned the task of measuring the movements of rock glaciers in Val Sassa and Val da l'Acqua (H. Haller, Eisenhut, and R. Haller, 2013, p. 117). For the first time in 1923, detailed descriptions of the rock glaciers in Val Sassa, Val da l'Acqua, and Val Tantermozza were published (H. Haller, Eisenhut, and R. Haller, 2013, p. 210). However, the first systematic study was not conducted until 1951 to 1973, when aerial photographs were used to calculate a movement rate of half a meter per year (Baur and Scheurer, 2014, p. 128). Geophysical studies followed in the 1960s, with further research conducted in 1972 and 1978, describing how the rock glacier had moved constantly since 1920 (H. Haller, Eisenhut, and R. Haller, 2013, p. 210). The rock glaciers have been measured for over a hundred years, making them the oldest time series (H. Haller, Eisenhut, and R. Haller, 2013, p. 129). Although the rock glaciers are a natural wonder, Val Sassa is seldom visited, possibly due to the hike's difficulty level (Baur and Scheurer, 2014, p. 334). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Natural Processes, Park Management, Research & Monitoring
- Primary Themes/Habitat: Extreme Locations, Infrastructure, Water Bodies
- Animals: low co-occurrence

Zernez

The municipality of Zernez, although located spatially outside the SNP perimeter, forms the gateway to the National Park, with its administration centre in the National Park House (founded in 1968, housed in the Baroque Wildenberg Castle since 2007) and the museum (moved to a new building in 2008), making it an important point of contact for park visitors (Bundesamt für Kultur BAK, 2024). In 1909, Zernez, one of the first municipalities, signed the lease agreement, transferring the Val Cluozza and, later, in 1920, the Falcun area to the SNP (Kupper, 2012, p. 70). In 2015, the municipality merged with the neighbouring municipalities of Lavin and Susch but continued under Zernez. Zernez is one of the most forest-rich municipalities in Switzerland and, at the same time, a crucial transport hub: The Bever–Scuol line of the Rhaetian Railway has been passing through Zernez since 1913, and the Ofenpass road channels traffic through the SNP (Wikipedia, 2024). I assume a relevant co-occurrence with the following thematic areas:

- Causes of Landscape Change: Park Management, Research & Monitoring, Tourism & Guests
- Primary Themes/Habitat: Infrastructure
- Animals: low co-occurrence

B. Appendix

B.1 Cratschla

B.1.1 Frequency of place names

Cratschla covers 1992 to 2023 and has two issues published annually (*Cratschla: Informationen aus dem Schweizerischen Nationalpark* 1992). Since the magazines always have a thematic focus, potential frequency peaks can also be attributed to the specific content focus of an issue. Nevertheless, as always, we first look at the sum of the frequency of occurrences over the entire period to get an initial size ratio between the places (see Figure B.1).

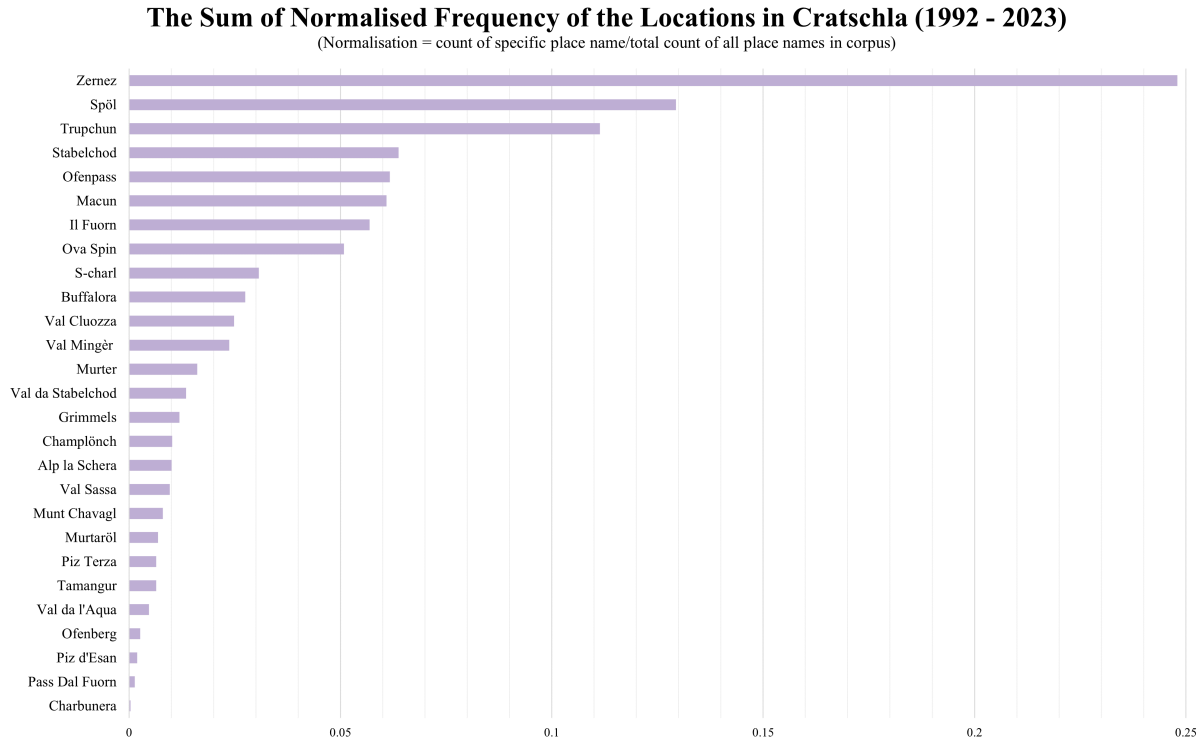


Figure B.1. This plot shows the sum of normalised appearances over the whole publication period, 1992-2023. The normalisation was conducted by dividing the count of specific place name by the total count of all place names.

Once again, it is no surprise that Zernez and Spöl secure the leading positions, though Zernez distinctly dominates with a frequency of 0.25, whereas Spöl has only 0.13. Zernez stands out as a clear outlier. Its top ranking is expected, considering Zernez is the central hub for the

park's administration and the National Park Centre. Cratschla publications frequently update on current events, such as excursions, lectures, and cinema and theatre events, alongside personnel updates, including staff changes. These factors contribute to Zernez's pronounced presence in Cratschla. Besides Spöl, other key locations include Trupchun, Stabelchod, Ofenpass, Macun, Il Fuorn and Ova Spin. The high frequencies of Macun are interesting due to its inclusion only being in the year 2000. Compared to Brunies, S-charl and Buffalora also seem less important (Brunies, 1914; Brunies, 1948). Charbunera is last, though we see an appearance of Pass dal Fuorn, Piz d'Esan, and Ofenberg. Ofenpass again has a higher frequency than Pass dal Fuorn. Similar locations appear as important or less important than in the Atlas. Interestingly, more has been written about Tamangur than about Val da l'Acqua. The rock glacier phenomenon in Val Sassa seems to receive more attention, possibly due to the longer series of measurements or because more measurements are generally conducted there. The aggregation provides a good initial analysis of which localities have been written about the most resp. the least in Cratschla over the last 30 years. However, it does not consider fluctuations in the frequency of place names. Due to the number of places, a line graph showing the changes over time can quickly become cluttered and confusing. Although Cratschla covers the last 30 years, this represents less than a third of the SNP's lifespan, limiting the interpretation of possible events. Nevertheless, four locations were plotted over time (across the publication series) to identify exciting events and patterns and to show the frequencies at Cratschla (see Figure B.2).

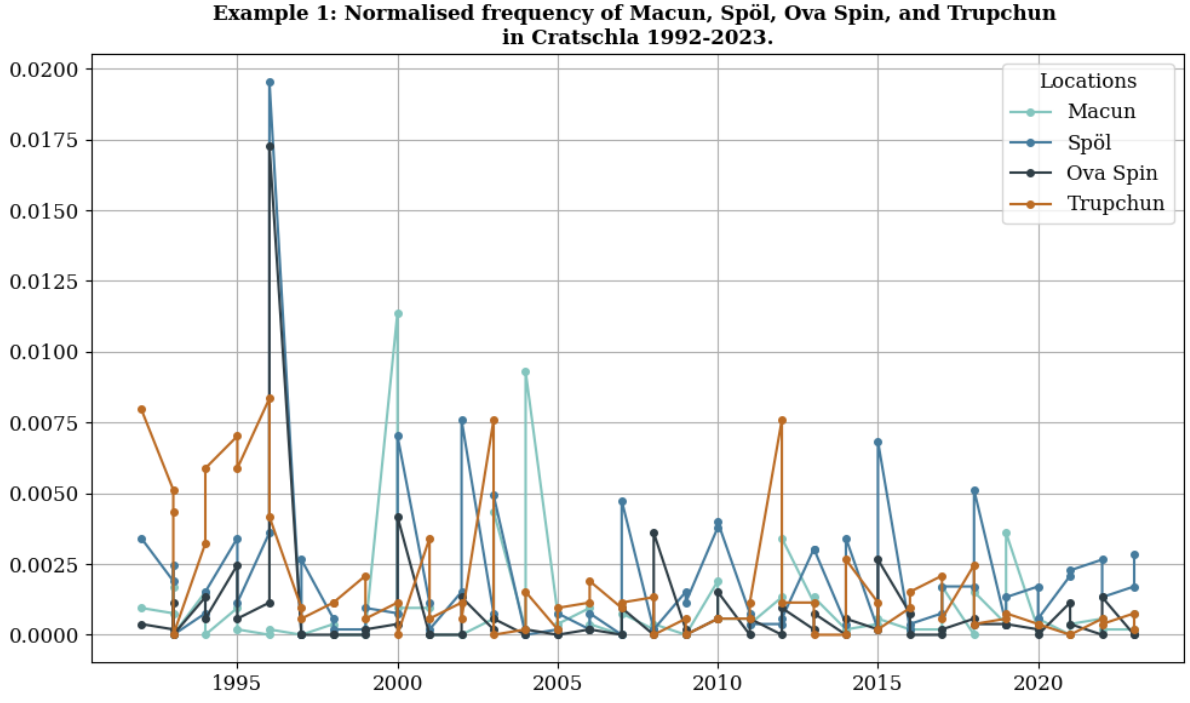


Figure B.2. Example 1: Macun, Spöl, Ova Spin and Trupchun.

I selected Macun, Spöl, Ova Spin, and Trupchun, especially since Spöl and Ova Spin are physically connected, yet they frequently do not appear together in the publications. Macun was

chosen to demonstrate the spike at the beginning of the 2000s. Additionally, they are all related to the theme of water. To represent a different habitat, the frequently mentioned Trupchun was selected. An observable example is that Spöl and Ova Spin, both bodies of water, have the highest number of mentions in the 2nd edition of 1996 (see Figure B.2). It makes sense, as the thematic focus of that issue was the rinsing of the reservoir Ova Spin, which is connected to the Spöl. Trupchun seemed to receive particularly high attention before the turn of the century. When Macun became part of the SNP in 2000, an entire edition was dedicated to the lake plateau under the titles "Park Expansion Discussion" and "Mysterious Macun," which naturally explains the distinct peak in mentions. The spike in the second edition of 1996 is, as mentioned, thanks to the main theme of the issue, namely the draining of the Ova Spin balancing reservoir in 1995. The peak in 2000, particularly for Macun but also for Spöl and then Ova Spin, is due to the thematic focus on park expansion in the first issue and the thematic emphasis on dynamic flooding in Spöl in the second issue. In 2004, Macun was relevant again, not for the lake landscape but for the crystalline subsoil, which was the main topic of the second issue that year. Ova Spin and Spöl do not always appear together, such as in 2002, 2003, 2015, and 2018, where Spöl was more important. In these issues, Spöl is either the focus of the issue (2015) or discussed in terms of revitalisation (2003) and Spöl as a habitat and energy supplier (2002, 2018).

The second issue is unsurprising in all these cases since habitat focuses often alternate. The first issue often addresses the terrestrial habitat, and the second addresses the aquatic world. Ova Spin is only more important than Spöl in the second issue, 2008, where Ova Spin is associated with the Engadine power plants (EKW) and the power grid. It makes sense that Ova Spin is mentioned less often than Spöl, as it is mostly located outside the SNP boundaries. This example is also very representative of the other locations. The thematic alternating focus is also evident in Trupchun. Trupchun was strongly themed until 1996 and received more attention in 2003 and 2012. The first publications (both issues) focused heavily on ungulates and ibex themes, where Trupchun provides habitat. After 1996, the stronger thematic focuses of the issue came to the forefront. Although the frequencies of Trupchun generally decrease, it is mostly the case that Trupchun is mentioned more often in the first issue than in the second. Also, the peaks in 2003 and 2012 are found in the first issue, which again thematises ungulates. The variation of thematic focuses after 1996 might not be coincidental. As is well known, a new park director, Heinrich Haller, took office that year, succeeding his predecessor, Klaus Robin (founder of Cratschla). These changes always influenced park management, and just as Robin introduced Cratschla, the change might have altered the magazine's structure (Source: Tamara Estermann). This example illustrates well that the corpus genre and who/how it is written strongly influence what it is written about. Moreover, the thematic focus of the editions greatly influences the frequencies.

B.1.2 Thematic context

The analysis of thematic field co-occurrence within Cratschla revealed intriguing patterns associated with the entities Macun, Spöl, and Trupchun (see Figure B.3). For Macun, a pronounced relationship was evident with "Water Bodies" and "Research and Monitoring", suggesting a substantial focus on aquatic ecosystems and the scientific aspects within its discourse. Spöl's highest co-occurrence was with "Water Bodies". Trupchun showed a dual peak in co-occurrence, with both "Park Management" and "Research and Monitoring" strongly featured. This highlights a strong commitment to managing natural resources and thorough scientific investigation. All three locations co-occur with the human thematic fields, though Spöl appears less relevant for "Tourism and Guests". The connection of the hydropower plant with Spöl seems more evident with the high value in the infrastructure field. Overall, the three locations do not often co-occur with Rewilding, Grassland, climate change and natural processes, though the latter, in the case of Spöl and Macun, have slightly higher co-occurrence. In general, the distribution of the co-occurrence with the thematic fields represents the landscape of the location and its characteristics well. These themes reflect the corpus's emphasis on the ecological and hydrological significance of Spöl, the natural beauty and scientific interest of Macun, and the rich wildlife and conservation efforts in Trupchun.

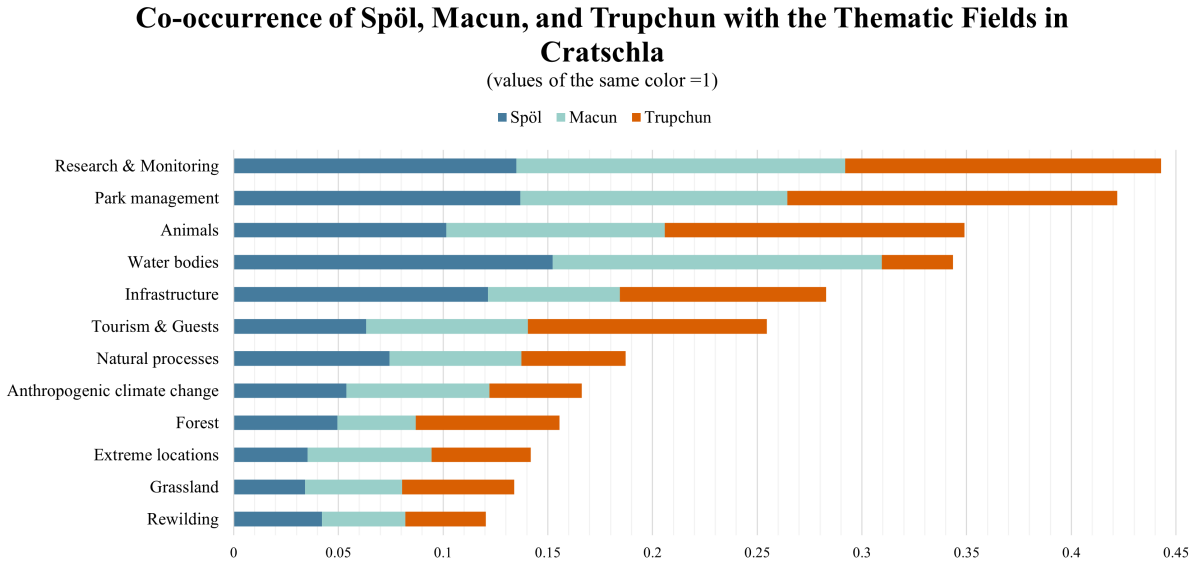


Figure B.3. Examples Macun, Spöl and Trupchun: Co-occurrence with the thematic fields.

To delve deeper into the peaks of specific locations, Table B.1 displays the most common nouns: 2015 for Spöl, 2004 for Macun, and 2012 for Trupchun. As previously mentioned, these editions feature discussions where Spöl is highlighted as both a habitat and an energy source, Macun for its crystalline subsoil, and Trupchun for its emphasis on ungulates. The 2015 discussion on Spöl vividly brings up the themes of hydropower and energy with nouns such as "ekw," "stauanlage," "wasserkraft," and "energie," alongside mentions of "hochwasser", "regeneration", and "massnahmen". This edition effectively portrays Spöl in terms of its habitat

and role as an energy source through these nouns. Similarly, Trupchun's discussion on ungulates, research, and their popularity among tourists can be inferred from nouns like "steinbock," "huftiere," as well as "forschung," "wanderung." In the case of Macun, though, the focus shifts more towards the lakes and their fauna rather than the crystalline subsoil, indicating that while the thematic focus of the corpus and the habitats of these locations are aptly captured through thematic fields when filtering out the most frequent nouns across the corpus, pinpointing specific events using a publication like Cratschla proves challenging.

Spöl 2015/02		**Macun 2004/02**	**Trupchun 2012/01**
ekw (10)	hochwasser (3)	seen (29)	steinböcke (9)
stauanlage (6)	restwassersanierung (3)	daphnien (16)	tiere (7)
sanierung (5)	bach (3)	temperaturen (14)	wanderung (7)
wasserkraft (5)	restwasser (3)	gewässer (11)	wanderer (7)
energie (4)	quellflur (3)	fische (11)	weg (6)
arbeitsgruppe (4)	arbeiten (3)	fische (11)	huftiere (5)
see (4)	regeneration (3)	wasser (9)	gemeinde (4)
kraftwerke (3)		weiher (9)	projekt (4)
staumauer (3)		blockgletscher (9)	gämsen (3)
massnahmen (3)		untersuchungen (8)	lawinen (3)
arbeiten (3)		bäche (7)	route (3)
inn (3)		fischen (7)	forschung (3)

Table B.1. Example 1: Most frequent nouns associated with Spöl, Macun, and Trupchun in single publications.

B.2 National Park Research publications (NPF)

B.2.1 Frequency of place names

As Figure B.4 shows, the high frequency for Zernez (0.2) is not so surprising, followed by a remarkable drop for Il Fuorn (0.1). All locations are mentioned at least once in the NPF publication series. However, it can be said that Charbunera (0.0004) practically has a negligible prevalence. On the other hand, it is interesting to see why and in what context it has been mentioned. Localities in the higher ranks in all corpora, such as Trupchun, Buffalora, Spöl, Stabelchod and Ofenpass, are similarly important in the NPF publications. Yet, S-charl and Murter appear to be more common here. Pass dal Fuorn gets mentioned considerably more than in the Business Reports, even though both corpora cover similar periods.
In the lower ranks, next to Charbunera, we have Val da Stabelchod, Val da l'Acqua, Piz d'Esan, and Piz Terza. Comparing the plot with the Business Reports, since both corpora cover a similar period, many similarities and exciting differences can be seen regarding the locations mentioned. For example, the Business Reports discuss Trupchun more frequently than the NPF publications. Tamangur, on the other hand, is more relevant in the NPF publications than in the Business Reports. Val Cluozza, Spöl, Ova Spin, Val Mingèr, and Grimmels are given comparable levels of attention. Locations such as Piz d'Esan, Val da l'Acqua and Piz Terza are deemed less important in both corpora. This variance and overlaps could provide valuable clues about the thematic focus of each corpus.

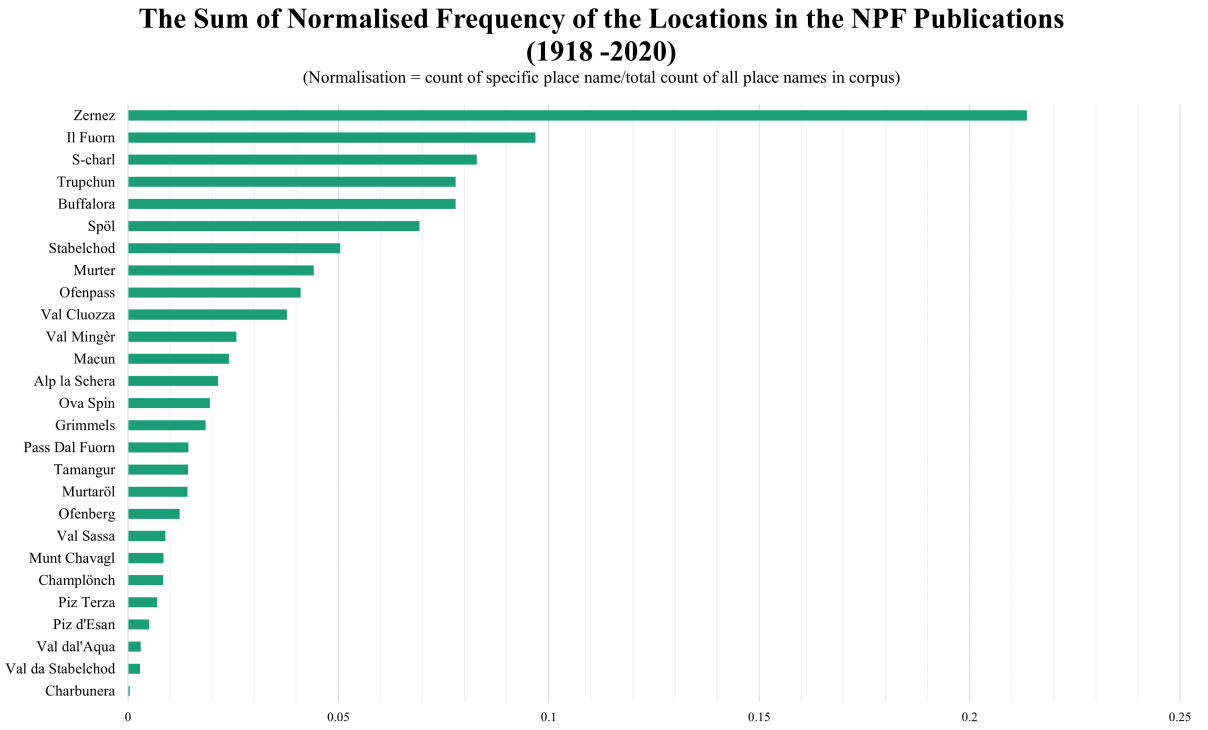


Figure B.4. Sum of normalised frequency field/place names in the NPF. Normalisation: Count of Specific Place Name in Publication/ total count of all place names.

The line graph B.5 shows how often Il Fuorn, S-charl, Spöl, and Trupchun have been mentioned in the NPF publications. The NPF publications were not published annually or at a constant interval and vary in length, as they can be scientific publications or books (*Publikationsreihe Nationalpark-Forschung in der Schweiz* 1918). The intervals, including the breaks, are visible here. All three locations show high frequencies for 1960 and 1963-65, particularly S-charl in 1960 and 1964. Also, 2006 seems to be a relevant publication year for Trupchun and Il Fuorn. Il Fuorn was first mentioned in 1928, and S-charl not until 1946. Generally, the frequency of mentions for Il Fuorn, S-charl, and Trupchun increased in the 1940s and dropped after 1965, with exceptions in 2006 and 2011-13. Upon review, the editions from 1943 (high values for Il Fuorn), 1960 were half in French and German. The 2006 publications focused on ungulate research, especially in the Il Fuorn and Trupchun areas, with the latter being known for red deer issues. The overlap of these two locations is no coincidence, as red deer also resettled in Il Fuorn in the 1990s. The German publication about mining in the SNP in 1960 explains the high rate of S-charl, which, like Il Fuorn, has a mining history. However, the high frequency of Trupchun is due to the French publication about higher fungi populations in the subalpine zone. As with the other corpora, Spöl shows increased repetition around the 1960s, when the hydropower construction took place, and again in 2012 and 2015, where the environmental accidents (2013 and 2016) can only partially explain the spikes. Upon closer inspection, these outliers seem to have little to do with the events. Still, Spöl is associated with water research (2012) and studies on habitats and being an energy supplier (2015). The individual high frequency in 1918 and

1924 is more related to research on the flora and fauna in the water than to historic events. This example shows that frequency analysis alone does not indicate events but points to the locations' landscape features. These studies can examine these in detail and are not solely dedicated to a specific time, such as the mining publication dealing with a time before the SNP. Conversely, the Business Reports are more current and document annual events. In this sense, the NPF publications are similar to the Cratschla series: the frequency suggests more about the thematic significance of a location than about a temporal event.

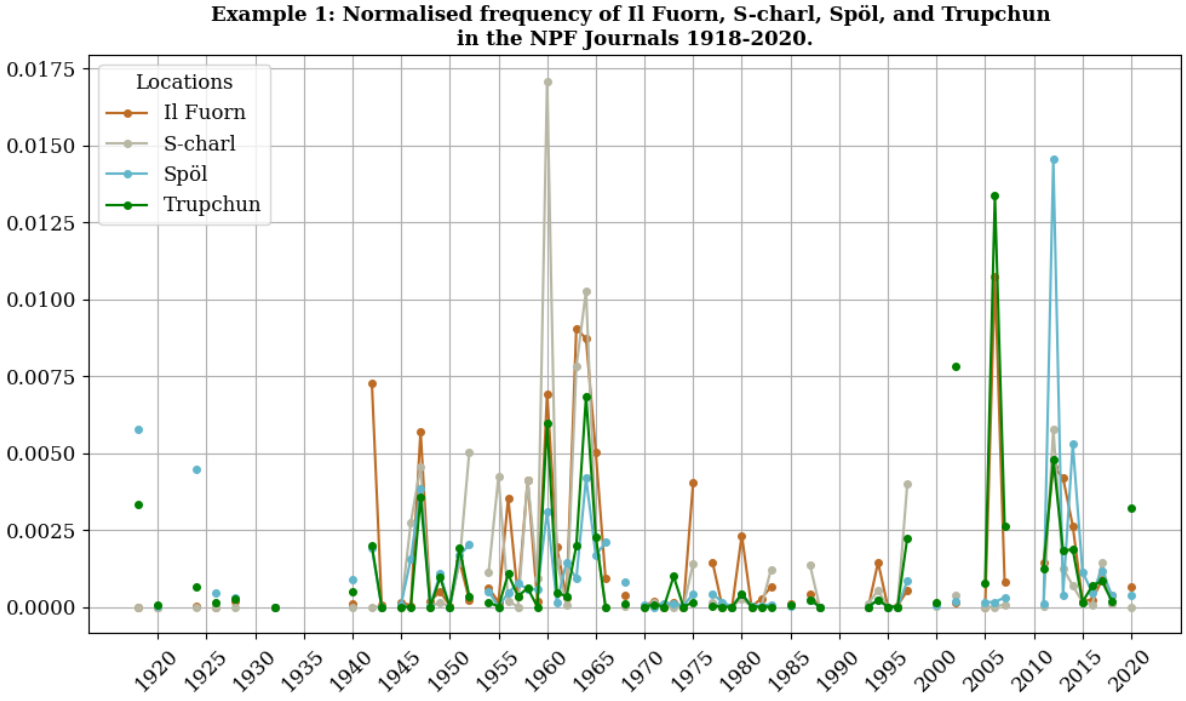


Figure B.5. Example: Il Fuorn, Trupchun, and S-charl

B.2.2 Thematic context

Four locations, Il Fuorn, S-charl, Spöl, and Trupchun, were analysed across 12 different thematic fields. The thematic fields of "Forest," "Infrastructure," "Research and Monitoring," and "Grassland" are most frequently discussed. The strong co-occurrence of S-charl with "Infrastructure" and "Forest", which could be attributed to the presence of the Tamangur Forest in Val S-charl and/or its history of wood clearance. Trupchun is associated with all habitats except for water bodies and extreme locations about equally often. This could also be related to the size of the thematic field of extreme locations. The infrequent association of water bodies with Trupchun may be because this element is not prevalent in the area, and the matches might have been more associated with precipitation terms. The highest focus with Trupchun is on the animals, indicating it may be where animals are more common or extensively studied. Interestingly, compared to the Atlas, where animals played a significant role overall, habitats receive more focus in the research publications. S-charl and Spöl also emphasize animals,

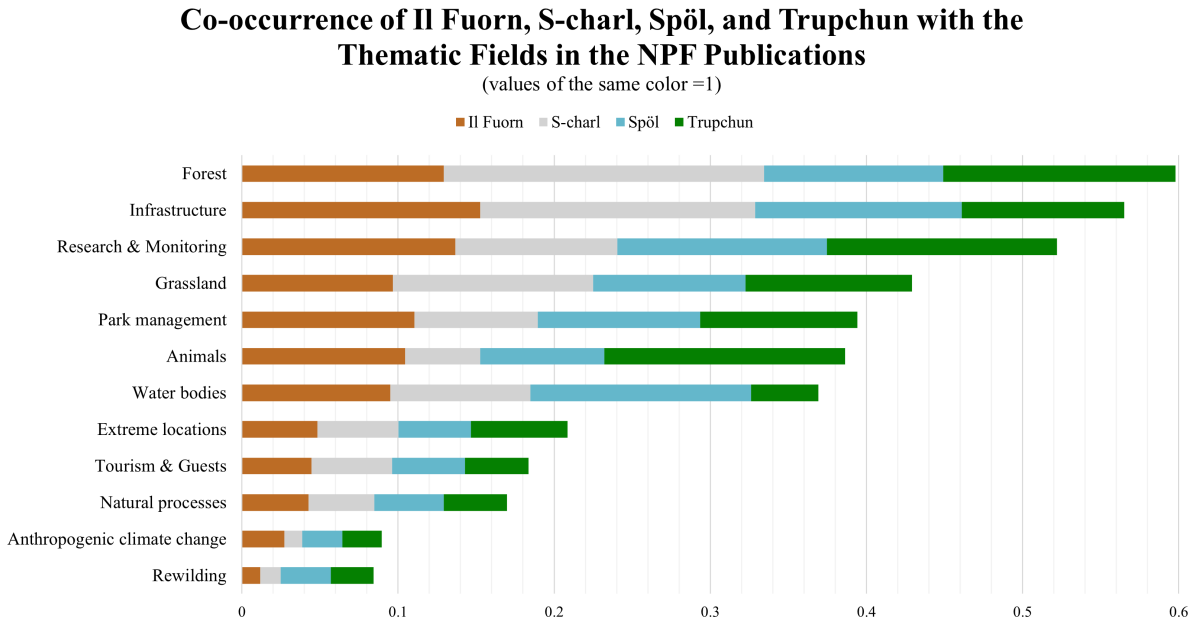
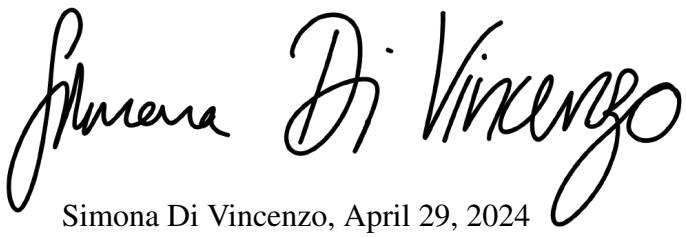


Figure B.6. Example Il Fuorn, S-charl, Spöl, and Trupchun: Co-occurrence with thematic fields.

though to a lesser extent, suggesting they still consider animal life an important aspect of their ecosystems. The analysis of forests and grasslands reveals the types of environments in these areas. All areas display a good balance in primary themes/habitats, highlighting the park's varied research into ecosystems. Il Fuorn and S-charl feature the most human-made structures, possibly referencing research facilities and the hotel. Regarding water ecosystems, Spöl records the highest co-occurrence with the thematic field of "water bodies." The topics of "Climate Change Effects," "Natural Processes," and "Rewilding" are less frequently mentioned in the NPF publications. Given the relatively narrow scope of these fields and because each publication typically focuses on a specific ecosystem and/or the park's history, these themes may be easily overlooked. Park management and scientific research are of high importance across all locations. Surprisingly, tourism is most prominently noted in S-charl, suggesting it is a popular destination for visitors. In summary, the high co-occurrence in the primary fields and research reflects the corpus genre well. However, the low concordance in themes related to causes of landscape change (except those attributed to human activity) is not as representative. This is likely due to their considerably small size compared to the other fields.

8. Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis. Furthermore, in this thesis, AI applications such as ChatGPT 4, Grammarly, and DeepL Translate were used for language translation and to improve the readability and quality of the text.



Simona Di Vincenzo, April 29, 2024