University of Zurich[UZH]

Comparing Semantic Geographic Information of Toponyms through Landscape

GEO 511 Master's Thesis

Author
Songlin Wang
21-739-545

Supervised by
Prof. Dr. Ross Purves
Dr. Sara Racca

Faculty representative
Prof. Dr. Ross Purves

30.04.2024
Department of Geography, University of Zurich

Abstract

Toponyms are names assigned to particular geographic objects. They could contain a greater degree of semantic meanings, particularly those about landscape, compared to names of other objects. Various research has been conducted to investigate the relationship between toponyms and landscape, seeking the way that people perceived landscape characters to name geographic objects, in various regions and countries including Switzerland. However, only small subsets of toponyms in Switzerland were studied, via somewhat limited methods. In this research, I aimed to focus on more generic parts, which usually serve as the spatial classifiers in toponyms, of Swiss toponyms in German and Italian, to study and compare the landscape they denote. I first extracted nature-related generic parts (but not related to water areas or very tiny landforms) from gazetteer, together with studying their meanings via multiple sources. Then, I selected the generic parts that were found to be about convex landforms, and those about open areas as case studies for comparisons. Landscape, including topography and land cover, of related objects' peripheral areas were investigated by first calculating histograms regarding elevations, slopes, and land cover types for each of them, then constructing bag of words using these histograms for each generic part. For each case study, all the bags of words were first aggregated, then partitioned using K-Means clustering, and patterns of each generic part were found according to their histograms' distribution based on the clustering result. Cosine similarities were also calculated, and the clustering of each individual bag of words was also conducted to compare generic parts with their patterns together. The research found that some of the generic parts can represent different kinds of landscapes. What's more, even though some generic parts were told to denote similar kinds of landscapes, they show different patterns in this research. Besides, cross-language comparisons were also conducted, via which I drew the conclusion that generic parts in different languages are not translatable. In all, this research uncovered the meanings of a larger set of generic parts in Swiss toponyms, which could potentially pave the way for further investigations about the processes by which people assigned names to geographic objects.

Keywords: Toponyms, Landscape, Generic Parts, Extraction, Bag-of-words, Clustering, Comparison

Acknowledgements

First of all, I express my sincere gratitude to my supervisors Prof. Dr. Ross Purves and Dr. Sara Racca, for their patient guidance and their invaluable advice, which played an important role in advancing my thesis forward. They always take the trouble to listen to my ideas and answer my sometimes tedious questions, with which I made a lot of progresses while working on my thesis.

Then, I would like to thank my parents for their supports in my studying during these years. I would also thank my meemaw, who enlightened me in learning English when I was a little kid, which inspired me to explore the whole world. Her rigorous altitude in doing research always inspires me a lot as well.

Finally, I would also like to express my special thanks to Prof. Dr. Krzysztof Janowicz at Uni-Wien, who is going to be one of the supervisors during my next academic journey. His useful advises during my interview contributed to further advancements in my thesis.

Contents

List of Figures

List of Tables

Acronyms

BoW	Bag of Words
CORINE	Coordination of Information on the Environment
DHM	Digital Height Model
DTM	Digital Terrain Model
ISCED	International Standard Classification of Education
NDVI	Normalized Difference Vegetation Index
SOM	Self-Organizing Map
TIFF	Tagged Image File Format
TLM	Topographic Landscape Model
UNESCO	United Nations Educational, Scientific and Cultural Organization

Chapter 1 Introduction

1.1 Background and Motivation

Various ‘objects’ in our world have names, which usually seem to convey semantic information. Let’s start with people’s names. Apparently, the names of most of us were not given arbitrarily. They could, for example, embrace the expectations of our family members. Take a common given name ‘Albert’ as an example. Derived from Old High German, this name has its semantic meaning ‘noble, bright, and famous’ (Wayback Machine, 2018). However, it couldn’t be true that all people with their first name ‘Albert’ are indeed ‘noble, bright, and famous’. In other words, we could not infer people’s characteristics through their names. This was summarized by Coates (2006) that names are ‘senseless’, since they are ‘less intuitively obvious’. However, when it comes to the names of places, further discussions are needed regarding this conclusion. Is it still valid for such names? Are they ‘senseless’ as well, or could they, on contrary, reflect some characteristics of the corresponding places? Hollis and Valentine (2001) discussed these and have verified that, compared to people's names, toponyms, which are names given to places, could usually ‘carry a greater degree of semantic meanings’, and are ‘less arbitrary’. Perono Cacciafoco and Cavallaro (2023) also claimed that the acts of naming of places are caused by people’s intention of, for example, safeguarding their memory and commemorating important events and people etc. These objectives for naming places could often be related to landscape, leading to toponyms that frequently reflect geographical characteristics (Burenhult & Levinson, 2008; Helleland, 2002), as well as having a close relationship with landscape terms (Levinson, 2008). Considering such characteristics of toponyms, Burenhult and Levinson (2008) proposed the following questions:

> ‘How are landscape features selected as nameable objects (‘river’, ‘mountain’, ‘cliff’)? Are there universal categories? What is the relation between landscape terms (common nouns) and place names (proper nouns)? How translatable are landscape terms across languages, and what ontological categories do they commit to?’ (Burenhult & Levinson, 2008)

These questions have been discussed in various ways by many scholars recently, and I hope to find their answers as well. I retrieved for some toponyms in Zürich first through Ortsnamen.ch, which is the result of Swiss toponyms research containing more than 740,000 geo-referenced toponyms (Schweizerisches Idiotikon, 2023), and explored the landscape-related information most of them contain. Let us have a look at two among them: ‘Milchbuck’ and ‘Strickhof’ here, which are quite close to Campus Irchel, therefore we must be familiar with them. ‘Milchbuck’ contains ‘Milch’ (in English means ‘milk’) which refers to a place where people store milk, or where juicy grass grows which promotes milk yield, and ‘buck’ refers to a hill. ‘Strickhof’ was once called ‘Strick’, which refers to a path with a property, while in the 18th century ‘hof’ was added to it as the courtyard there was expanded, marking the formation of a farm, based on that property (Schweizerisches Idiotikon, 2023). Toponyms like these could usually reveal people’s collective memory, and could be seen as part of the cultural heritage, as Perono Cacciafoco and Cavallaro (2023) highlighted, while studying the semantic geographic information embedded in them that denote geographic objects could provide us with more insights into how people, especially ancient people, 'perceived, conceptualized, classified, and utilized the environment' (Thornton, 1997). In other words, such 'archaeology of language' allows us to explore the relationships between human beings and the landscape and people's feelings about them, hence researchers nowadays are managing to document and recover the origins and the stories of such toponyms, return them to the new generations, and safeguard the 'feeling of the places' (Perono Cacciafoco & Cavallaro, 2023). These purposes also resonate with my interests, therefore, I conducted my research on Swiss toponyms, adopting an interdisciplinary approach to explore more about the landscape they could represent, in order that the 'auroral sensation' of ancient people living there while naming places, the way they perceived landscape, could be revealed.

1.2 Research Objectives

The concrete information including semantic meanings, histories, as well as changes in history of plenty of Swiss toponyms could be found through multiple resources. As shown above, the information of toponyms 'Milchbuck' and 'Strickhof' could be found through 'Ortsnamen.ch' (Schweizerisches Idiotikon, 2023), where 'Buck' was found to refer to a hill. Nonetheless, from an ontological perspective, such information is still not enough. For this example, with merely the word 'hill' as an explanation for 'Buck' in toponyms, the 'big picture' of all the geographic objects with such names is still ambiguous. Neither we could know whether all of them denote similar objects or not, nor could we tell the differences between 'Buck' with other convex-landscape-related terms, like 'Bool', 'Büel' etc. denoting hills, or 'Berg', 'Horn' etc. denoting mountains (Schweizerisches Idiotikon, 2023). Similar problems also exist when talking about other kinds of landforms, like concave ones, saddle ones etc. (Mark & Smith, 2004). To solve these problems, quantitative methods could be more helpful, via which and with massive data of toponyms, the features of objects with exact names could be shown in a better way. Various questions including 'Does 'Buck' always denote an exact kind of landscape?', 'Does it denote a similar landscape as 'Büel?', etc., could be answered. Investigating this in the study, I was able to get closer to the answers to the first two questions from Burenhult and Levinson (2008). Besides, since Swiss people speak different languages (Swiss Federal Council, 2024), cross-language analysis could be interesting as well, i.e., to compare toponyms in different language regions. I also intended to conduct this in the research, via which the fourth question from Burenhult and Levinson (2008) could also be discussed.

1.3 Structure of the Thesis

This thesis is organized into 5 chapters, including this current chapter. The contents of the next four chapters are summarized as follows.

In Chapter 2, I will first introduce some definitions, concepts, and further essential information about toponyms as well as landscape, which are necessary to be clarified before talking about the main implementations of data pre-processing and data analysis. Related works about the association between them will also be discussed, followed by research gaps found based on them and the research questions I raised to be answered in this study.

Chapter 3 begins with the selection of study areas, followed by the introduction of data sources and the implementations of data pre-processing. Then, detailed methods for the analysis will be shown. Divided into two parts, the extraction of meaningful terms from toponyms, and the analysis to compare them using landscape, will be introduced respectively.

In Chapter 4, I will show the results of this research, which consists of two different parts: extraction of meaningful terms and analysis of the association and comparison as well. Multiple visualizations generated will also be shown to make the results clear.

In Chapter 5, I will further discuss my definitions, approaches and findings, and the limitations of this research as well, by coming back to the related works and research questions.

Finally, in Chapter 6, I will wrap up this research, and talk about some 'new research gaps' left for possible future works.

Chapter 2 State of the Art

This chapter is divided into four main sections. In the first section, I will start by introducing ‘names’, or ‘proper nouns’, then turn to the more specific one ‘toponyms’. Types of approaches to toponymy, which means the study of toponyms, will be briefly discussed afterward. I will talk about the typology of toponyms thereafter, from where the semantic information contained in them could be revealed. Then the structures of them from a morphological view would be introduced. In the second section, the other main element of this research ‘landscape’, will be discussed. This section begins with various definitions of it, followed by a best one in the scope of this research and the aspects of it under such definition. Then the ways people usually use to measure (numerical aspects) or classify (categorical aspects) landscapes will be summarized. In the third section, multiple studies discussing or analyzing the association of toponyms with landscape will be shown, both those focusing on Swiss toponyms and those investigating the toponyms from other countries or regions. Finally, I will talk about research gaps, then research questions I wanted to answer to fill them.

2.1 Toponyms

2.1.1 From Names to Toponyms

According to Perono Cacciafoco and Cavallaro (2023), Bredart et al. (2002), and Redmonds (2004), ‘names’, or more specifically ‘proper nouns’, are linguistic signs, or ‘special words’, made up of single or multiple words that can denote a range of entities, including people, animals and other living beings, places or landmarks like rivers, buildings, settlements, ‘things’ like furniture, temporal names like months, dates, etc. while nouns other than names are usually called ‘common nouns’. Names are given consciously and intentionally, rather than arbitrarily, with transparent etymologies, and could usually contain certain meanings (though people might forget them over time), carry plenty of social information, and emphasize specific connotations (Radding & Western, 2010). Hence, just like Coates (2006) said, they are not ‘meaningless’, though sometimes they could be ‘senseless’ (which means sometimes they are ‘less intuitively obvious’). Among all kinds of names, human names, like the first name ‘Albert’ I talked about in 1.1, are also called ‘anthroponyms’, and research of them are called ‘anthroponymy’ (Perono Cacciafoco & Cavallaro, 2023). ‘Toponyms’, or ‘place names’, on the other hand, are another kind of names, which was focused on in this study. Given to particular places, they indicate natural features like mountains, rivers, forests, and human-made features as well, and ‘tie to the way humans conceptualize and organize spaces’ (Perono Cacciafoco & Cavallaro, 2023). Compared to other kinds of proper names like anthroponyms, they usually ‘carry a greater degree of semantic meanings’ (Hollis & Valentine, 2001), and ‘exist in relation to geographical objects’ (Helleland, 2002). Moreover, scientific study of all kinds of names is called onomastics, under which the study of human names, like ‘Albert’, is called anthroponymy, while the study of place names, like ‘Milchbuck’, is called toponymy (Perono Cacciafoco & Cavallaro, 2023).

Not only toponyms' origins, etymologies, developments and changes over time, but also their cultural and sociological characteristics, utilities, and values in society, are what toponymy refers to (Perono Cacciafoco & Cavallaro, 2023). In the next section, I am going to present different types of approaches to toponymy in detail and discuss the differences they could make for the understanding of human culture.

2.1.2 Approaches to Toponymy

There are two different approaches to toponymy, including intensive approaches and extensive approaches, according to Tent (2015). The intensive one focuses on 'biography' of toponyms, in other words, 'etymology', and usually covers meanings and origins of them (Tent, 2015). In details, intensive research seeks the answers to the following questions, as Tent (2015, p. 68) summed up:

- 'Who named the place?'
- 'When was the place named?'
- 'Why was it given this particular name?'
- 'What does the name mean?' and 'What kind of feature is it?'
- 'Where does the name come from?' and 'Where is the place located?'

On the contrary, extensive toponymic research looks for patterns in a collection of place names based on datasets or corpora of toponyms, gazetteers, maps, atlases, and so on, according to Tent (2015). As he claimed, toponyms serve as independent variables in this case, which can be tested against dependent variables like regions, toponyms type, or features type. The analysis of the distribution of geographical features through toponymic datasets is an example of this kind of research (Perono Cacciafoco & Cavallaro, 2023; Tent, 2015). As for my study, since I used massive data of toponyms to analyze their patterns in a global view, it should be seen as an extensive one. Nevertheless, it was conducted with the help of multiple intensive approaches (via which adequate information about the meanings of most toponyms could be found). Just as Tent (2015) highlighted, both intensive and extensive paradigms contribute to each other, in addition to their contributions to the discipline of toponymy as a whole.

2.1.3 Typologies of Toponyms

Being the labels for places indicating different kinds of geographic objects (Perono Cacciafoco & Cavallaro, 2023), toponyms could be classified in different ways. Four of them will be summarized in the next paragraphs, as they are relative to this research, serving as the basis for me to decide whether to keep or to delete exact kinds of toponyms for conducting the analysis.

The first way to divide toponyms is through the types of objects they denote. According to Perono Cacciafoco and Cavallaro (2023), toponyms could indicate natural features (mountains, rivers, forests, etc.) or human-made ones (buildings, streets, cities, etc.). They stated that those indicating natural features include but are not limited to:

- Hydronyms: 'the names of all kinds of water bodies, including rivers, streams, brooks, lakes, and seas'. They could further be classified as oceanonyms (the names of oceans), pelagonyms (the names of seas), potamonyms (names of rivers), limnonyms (names of lakes), and micro-hydronyms (the names of smaller and more localized water bodies, like springs and wells).

- Oronyms: names of mountains, hills, and hillocks.
- Speleonyms: names of caves, chasms, grottoes, mines, and entire underground systems.
- Dryonyms: names of forests.

Those that denote human-made features include but are not limited to:

- Odonyms, which are the names of 'streets, avenues, boulevards, drives, lanes, and other denominations relating to inhabited areas'.
- Urbanonyms: the names of 'urban elements, like streets, blocks, parks, avenues, drives, churches, buildings, etc.'.

The naming of human-made features in urban areas can be influenced a lot by people's names, politics and economy etc., but less by landscape (Light & Young, 2015), hence problems might be raised for this study, and were thus not considered. Imagine the people's name 'Zuckerberg'. Usually in toponyms 'Berg' means 'mountain' (Schweizerisches Idiotikon, 2023), however, for those objects named after this 'Zuckerberg', they might have nothing to do with mountain at all. Such names could lead to biases for this study. Note that, this does not mean that naming of natural features is totally not influenced by these factors. But as it's rarely influenced, according to Light and Young (2015), I supposed that they would not cause severe problems.

The second way is to divide them through their scales. In this way, they could be divided into macro-toponyms, which are 'toponyms of larger or major geographical sites', like 'countries, cities, regions, and even major streets', and are usually well known, as well as micro-toponyms, which are 'toponyms of smaller geographic features, whose names are usually only known by a smaller set of people', or 'unofficial names that hardly anybody knows of' (Clark, 2009; Miccoli, 2019; Perono Cacciafoco & Cavallaro, 2023). The boundary between these two can be blurred, depending on the perspective of researchers, according to Perono Cacciafoco and Cavallaro (2023). For example, 'Bahnhofstrasse' in the city center of Zürich could be a macro-toponym for people living in or near Zürich, since it's a major street where people there usually go shopping, and 'Matterhorn' could also be a macro-toponym for people in Switzerland since it's famous among them, however, for many 'outsiders' like tourists that are totally not familiar with Switzerland, these names can be micro-toponyms, as they are hardly well-known among them. 'Zürich', 'Switzerland', 'Alps' etc. on the opposite, are definitely macro-toponyms for all people. In this research, the toponyms used for analysis were micro-toponyms (at least from the points of view of 'tourists'). As the naming of such 'smaller geographic features' was found to have a strong connection with the landscape (Penko Seidl, 2008), their names were helpful in making this analysis more accurate.

The third way is to divide them based on their origins into endonyms and exonyms. According to the Glossary of terms for the standardization of geographical names (Kadmon, 2002), an endonym means 'the name of a geographical feature in one of the languages occurring in that area where the feature is situated', while an exonym is 'the name used in a specific language for a geographical feature situated outside the area where that language has official status, and differing in its form from the name used in the official language or languages of the area where the geographical feature is situated'. These mean that, an exonym is not the original name of a place. For example, 'Genève' is an endonym of the city in the southwest of Switzerland, while it also has an exonym called 'Genf'. In this research, I only considered endonyms, the detailed reason of which would be explained in 3.2.1.2, together with the implementations.

The fourth way is to divide them through the motivation of naming. Blair and Tent (2021, p. 41-43) classified toponyms into the following types based on this:

- Descriptive toponyms: names that 'indicate an inherent characteristic of features', like those denoting the topography of features, the relationships between features and other features nearby, the locations or orientations of features, and the function of features (like 'Park').
- Associative toponyms: names that 'denote something associated with features or their contexts', including something in the local natural environment like animals perceived with or associated with the features, occupations, habitual activities, or related artefacts associated with the features, and manufactured structures associated with the features.
- Evaluative toponyms: names 'reflecting the emotive reactions of the namers, or strong connotations associated with the feature'. Positive responds and negative responds to features are both included.
- Occurrent toponyms: names that 'record events, incidents, occasions or times and dates when the features were named'.
- Copied toponyms: names 'copied from another place or language, including the use of features from other places and the use of name-form which the features have in other languages'.
- Eponymous toponyms: names from people, including 'names of the namers, of notable people, of colleagues, of family members or friends, and of other associated people', or names from other named entities, including names about animals, and names about non-animate entities like 'notable occasions, concrete entities, expedition vessels, literary, biblical, or mythical entities etc.'.
- Innovative toponyms: those names containing 'people's playing with humorous intent with languages', and those names that were 'specifically made with pleasing sound, positive connotation, or appropriate meaning'.

Similarly, Rennick (2005, p. 298-301) divided toponyms into:

- Local or descriptive names, considering:
    - 'Location, direction, position, or distance compared to other places or features.'
    - 'Shape, size, odor, color.'
    - Characteristic of 'the natural environments', including 'terrain, topography and landscape, soil and minerals, water bodies, animals, and plant life'.
    - Approbation and disapprobation.
- Historic events.
- Names taken from other places or features (earlier residences, nearby features etc.).
- Personal names.

- Subjective names that reveal more about the namers (aspirations and ideals of namers, nicknames of namers, humorous names etc.).
- Mistake names (that were misinterpreted by people).
- Names derived from multiple sources. (They could be classified into multiple classes mentioned above.)
- Underived names including those for which there is 'not even a guess much less a folk etymology'.

Through these classifications we can see that different toponyms contain different kinds of semantic information. Some are about landscape (like 'Milchbuck' remarked above, which is a descriptive name denoting the local features based on these classification methods), while some are not (for instance, about people, animals, events, people's emotions while perceiving landscape, from names of other places etc.). Though under the scope of this study, only the first groups of toponyms were extracted for analysis, it was not my hypothesis that, other kinds of semantic information of toponyms are meaningless. Interesting patterns might also be revealed by studying those toponyms, which can be the scope of other studies.

In this sub-section, I talked about the specific kinds of toponyms needed for my study. However, such exact toponyms were not 'prepared' by other people. To get them, data pre-processing and extraction were needed to massive toponyms data, contained in gazetteers, the implementations of which would be introduced in 3.2.1.2 as well as 3.3.

2.1.4 Structures of Toponyms

It is common for a European toponym to be compound in structure, with one element called the 'generic part', having classifying function and defining the type of the place, and the other element called the 'qualifier part', indicating the characteristic of the place, according to Hough (2006) and Gammeltoft (2005). Hough (2006) stated that in Germanic languages, the generic part of a toponym usually comes second, following the qualifier part in the first place. Let's still take 'Milchbuck' as an example here. Apparently, it's compound in structure, and it's composed of 'generic' part 'Buck' and 'qualifier' part 'Milch'. 'Buck' here has the 'classifying function', meaning the type of the place is a hill, while 'Milch' here has the 'descriptive function', indicating that people stored milk there or juicy grass grows there (Schweizerisches Idiotikon, 2023). However, not all European toponyms have such forms. For example, Ursini (2016) argued that in Italian, things are though quite different. According to him, in an Italian toponym, the pre-nominal part of it usually serves as the 'spatial classifier', which is of the same meaning with 'generic part', and the 'qualifier part' usually comes second. For example, in many toponyms 'Monte' appears first, with the classifying function, meaning that the related places are mountains, followed by the various kinds of qualifier parts (Olivieri, 1931). Note that, a toponym might also contain a single generic part only, for example, in German-speaking areas in Europe, some places are only named 'Buck', while in Italian-speaking areas in Europe, some geographic features are only named 'Monte' (Olivieri, 1931; Schweizerisches Idiotikon, 2023).

2.2 Landscape

2.2.1 What is Landscape: Different Aspects

The word 'landscape' originated in Western Europe during the thirteenth century, denoting 'both a tract of land organized by people and its visual appearance', according to Antrop and Van Eetvelde (2019). They argued that, it's a complex, multi-scale and dynamical system, and is continuously transforming. Mücher et al. (2010) also stated that landscapes themselves 'have resulted from long-term interactions of natural abiotic, biotic, and anthropogenic processes'. Hence, 'landscape' has 'etymologically multiple meanings' essentially, and is a holistic concept, containing elements from the natural, physical world and from the cultural world and human society (Antrop & Van Eetvelde, 2019).

Because of the characteristics of landscape, there exist different aspects to define it. As Bastian (2008) stated, 'landscape' can be explained from either positivistic or constructivistic positions, or both. People explaining it in a positivistic way assume that it is 'a really existing, defined part of the Earth's surface, where all the components (geological structure, relief, soils, climate, waters, plants, animals, and humans incl. settlements, roads, land use, etc.) are existing'. While the constructivistic view is held especially by social scientists, assuming that 'landscape is a construct in our mind'. For example, Hard (1970) said that landscape 'is a primarily aesthetic phenomenon, closer to the eyes than to the mind, more related to the heart, the soul, the moods than to the intellect'. Recently, there are growing trends of defining it in a comprehensive way to balance between those two aspects (Bastian, 2008), as it involves both human perception and physical reality (Brabyn, 2009). For example, it was defined as the appearance of the land by Brabyn (2009), where the 'appearance' embeds people's perception of landscape. According to Mücher et al. (2010), landscape was regarded as forming recognizable, although usually heterogeneous, parts of the earth's surface, and as showing a characteristic ordering of elements. Simensen et al. (2018) defined it as a geographical area, characterized by its content of observable, natural and human-induced, landscape elements. The Council of Europe Landscape Convention (2000) provided another definition to landscape:

> 'Landscape' means an area, as perceived by people, whose character is the result of the action and interaction of natural and/or human factors.' (Council of Europe Landscape Convention, 2000, p. 2)

Clearly, they also defined landscape 'comprehensively', from the points of view of both the entities, and people's perceptions to them.

2.2.2 Landscape for This Research

In this research, the definition of landscape was adopted from the Council of Europe Landscape Convention (2000). Specifically, two main components were focused on, including topography and land cover, as they are broad enough to cover most of the 'perceptible' natural features. In the next paragraphs, I am going to interpret what 'topography' and 'land cover' means in detail.

According to Mark and Smith (2004), topography is the shape of the Earth's surface. It is ‘a fundamental dimension of the environment, shaping or mediating many other environmental flows or functions’. They stated that ‘there exists a major divergence in the way that topography is conceptualized in different domains’. As they summarized, people including cartographers and geomorphologists usually conceptualize topographic variability in a quantitative way, as a continuous field of elevations or as some discrete approximation to such a field, while people like anthropologists, ecologists etc. usually conceptualize this same variability as ‘(special sorts of) objects, with locations, shapes, and names of their own’. It was pointed out that, in general landforms could be classified into five categories: convex, concave, horizontal planar, vertical planar, and saddle points, however, with those two ways to conceptualize it, the answers to some essential questions like: ‘What exactly are mountains?’, were hard to be given (Mark & Smith, 2004). They claimed that some landforms like mountains, might have crisp boundaries, therefore it could be hard to separate them sharply from their surroundings. What’s more, the boundaries of different landforms in the same categories as mentioned above, might be blurred, like ‘mountains’ and ‘hills’ (Mark & Smith, 2004). This was also discussed by Villette (2021), who stated that sometimes people define mountains as ‘an elevation standing high above the surrounding area, with steep slopes and local relief greater than 300m’, however, such objects might not be considered as ‘significant’ in the chain of Alps. Therefore, the integration of the qualitative ontology of landform features as well as categories, with the quantitative field-based ontology, was essential to describe the terrain surface, which could be fulfilled by, for example, DTM (Digital Terrain Model) (Mark & Smith, 2004). In the next sub-section, this will be talked about in detail.

Besides, land cover is ‘the physical material at the surface of the earth’, according to Fisher et al. (2005), and is ‘the material which we see, and which directly interacts with electromagnetic radiation and causes the level of reflected energy which we observe as the tone or the digital number at a location in an aerial photograph or satellite image’. They noted that, land cover is different from land use, which in contrast, is a description of how people use the land (Fisher et al., 2005). Since human-made objects in urban areas like roads, buildings, towers etc. were not in the scope of my study, land cover data could be more helpful than land use data here.

In the next part, I will lead into the details about how people usually do to measure topography, and to classify land cover, based on the definition of them mentioned above.

2.2.3 Measurements and Classifications

As discussed in the last sub-section, DTM is a good way to conceptualize the terrain surface. The following definition to it was given by Miller and Laflamme (1958):

> ‘The digital terrain model (DTM) is simply a statistical representation of the continuous surface of the ground by a large number of selected points with known X, Y, Z coordinates in an arbitrary coordinate field’ (Miller & Laflamme, 1958).

As Mark and Smith (2004) pointed out, with DTM, the terrain surface could be conceptualized as a single-valued function of position in a 2-dimensional geographic spatial domain. In the book *Digital Terrain Modeling: Principles and Methodology* by Li et al. (2005), it was stated that data for DTM could be from field surveying based on the terrain surface, photogrammetry by stereo pairs of aerials (or space) images and photogrammetric instruments, and cartographic digitization using existing topographic maps and digitizers. With these data, the DTM surface could be reconstructed, then after validation, DTM product could be produced. He highlighted that, compared to traditional analog representation, DTM has plenty of advantages. It has a variety of representation forms such as topographic maps, vertical and cross sections, as well as 3-D animation, greater feasibility of automation and real-time processing, and easier multi-scale representation, while it doesn't have accuracy loss of data over time owing to the use of digital medium (Li et al., 2005). It was summarized that, information of landforms (elevation and slope etc.), terrain features (settlements, boundaries, transportation networks, hydrographic features etc.), natural resources and environments (soil, geology, vegetation, climate, etc.); as well as socioeconomic data (population distribution, industry and agriculture, capital income, etc.), could all be represented by DTM data (Li et al., 2005).

Meanwhile, the author also summarized some terrain descriptors. It was claimed that, a terrain surface could be described by the concepts of roughness and irregularity and characterized by different numerical parameters,, in details, frequency spectrum (which can transform the terrain surface from the space domain to the frequency domain and characterize it by the frequency spectrum), curvature, covariance and auto-correlation etc. were told to be able to describe the terrain, and slope, relief, and wavelength were also told to be able to measure the roughness of the surface (Li et al., 2005).

As for land cover, there exist many different sources of data as well. As Anderson et al. (1976) pointed out, traditionally the data of it was generated by ground surveys involving enumeration and observation. Started in 1930s, remote sensing techniques including the conventional aerial photography has been used for classifying land cover, and nowadays it's used widely. NDVI (Normalized Difference Vegetation Index) derived from satellite data have been used for land cover classification frequently (Defries & Townshend, 1994). Nowadays, with the development of machine learning, multiple methods of it have been involved in the classification of land cover, for example, support vector machines (Huang et al., 2002), decision tree (Friedl & Brodley, 1997), random forests (Gislason et al., 2006), and joint deep learning models (Zhang et al., 2019) etc. Based on different research objectives, researchers usually classify land cover into different kinds of categories. For example, Anderson et al. (1976) determined 9 categories, which include 'urban or built-up land, agricultural land, rangeland, forest land, water, wetland, barren land, tundra and places with perennial snow or ice', in their classification system for classifying land cover types of America,. While Defries & Townshend (1994) and Friedl & Brodley (1997) classified land cover of the entire world into 11 classes in their research, including 'broadleaf evergreen forest, coniferous evergreen forest and woodland, high latitude deciduous forest and woodland, tundra, mixed deciduous and evergreen forest and woodland, broadleaf deciduous forest and woodland, wooded grassland, grassland, shrubs and bare ground, bared ground and cultivated ground', which mostly based on NDVI and UNESCO definitions. Clearly, this classification method considered different climate types in different parts of the world, due to their broad research area. Besides, Huang et al. (2002) classified land cover in a relatively smaller area: eastern Maryland, into only 6 types, including 'closed forest, open forest, woodland, non-forest land, land-water mix and water'. In this research, I also managed to make the categories of land cover suitable for the context of it. The detailed information can be found in 3.2.3.2.

2.3 The Association of Toponyms with Landscape

2.3.1 Overview

As mentioned in Chapter 1, toponyms could contain semantic information and could denote geographical objects, according to Helleland (2002). He argued that a toponym normally exists in relation to a geographical object, real or imagined. Jordan (2012) also said that 'geographical names reflect spatial characteristics, most frequently natural characteristics, but sometimes also characteristics of settlement history, land use and economy, former feudal relations, historical events, etc.'. When they reflect natural ones, according to Jordan (2012), they refer mostly to location, morphology, exposition, waters, vegetation, soil conditions and mineral resources. Thus, geographers are quite interested in answering the research questions of Burenhult & Levinson (2008) shown in 1.2, as the answers of which can 'provide evidence of environmental settlement and social conditions at the time a name was coined, and determine the origins and processes of places, identify spatial locations, and operationalize meanings in space and in human-nature relationships' (Savage, 2020). In the following sections, various research that focus on such questions would be summarized, both those investigating Swiss toponyms, and those discussing toponyms of other regions. Besides, it was also mentioned that cross-language analysis could be interesting. As Mark and Turk (2017) proposed, 'people from different places and cultures use various conceptual categories for landscape features, as reflected in toponyms'. For example, an object is called a 'mountain' by people living at place A, however, a similar object of it at place B might only be considered as a hill by people living there. Such differences could be reflected by toponyms in these places. This hypothesis is called 'Ethnophysiography Hypothesis' (Mark & Turk, 2017). Some research conducting cross-language analysis and discussing the last research question of Burenhult & Levinson (2008), as well as this hypothesis, will be summarized in the next sections as well.

2.3.2 Previous Research

2.3.2.1 Research on Swiss Toponyms

To explore the meanings of generic parts of Swiss toponyms, Derungs et al. (2013) conducted a study aiming at finding their relations with landscape. Using topographic characteristics (elevation and slope), they calculated values including relief (the maximum difference between the elevation of two raster cells within the buffer zone), standard deviation in elevation (which is related to surface roughness), mean slope and standard deviation of slope of each toponym's related location's three buffer zones (200m, 400m and 2000m), as a 12-direction descriptor of topography. Similarity between toponyms were computed both quantitatively and qualitatively, using cosine similarity and 'Self-Organizing Map', or SOM. Here SOM is a method 'created through an unsupervised artificial neural network procedure allowing us to see the arrangement of geographic kinds in a space based on conceptual similarity' (Mark et al., 2001). Through the research, Derungs et al. (2013) found that 'the SOMs of related areas of 'Spitze' and 'Horn' were autocorrelated, and almost congruent, while for 'Berg' it was quite different and showed a diffuse pattern', though all of these three denote 'convex landforms' according to sources like 'Ortsnamen.ch' (Schweizerisches Idiotikon, 2023). They also found that terms denoting passes through mountains in different languages in Switzerland including 'Pass', 'fuorcla' and 'col', represented similar topographies, with some overlap with Horn and Spitze, while for terms denoting cultivated fields: 'champs', 'Acker', 'Acher', the related topographic characteristics occupied a different region of the SOM, overlapping slightly with the region occupied by Berg. Through cosine similarity, they also found that generic parts that describe similar features have high cosine values, even for those in different languages (Derungs et al., 2013). Therefore, the 'Ethnophysiography Hypothesis' should be rejected. The SOM representation of topographic vectors in their research is shown in Figure 2.1, while Figure 2.2 shows their cosine similarity analysis results.

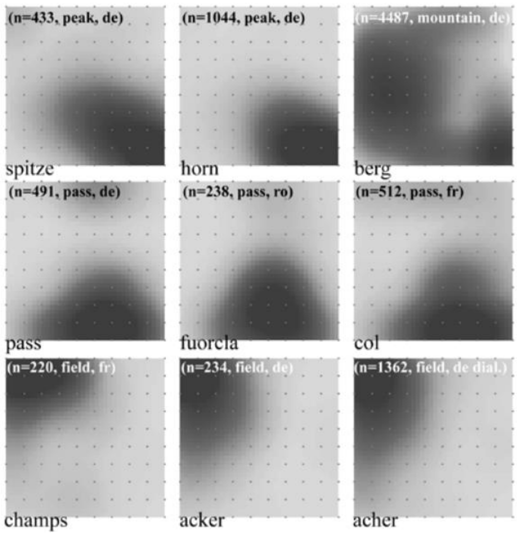


Figure 2.1: SOM representation of topographic vectors from Derungs et al. (2013).

Villette (2021) on the other hand, investigated landscape through toponyms in Switzerland to study the landscape terms in micro-toponyms and landscape conceptualizations, and to explore 'the relationship between the natural and the cultural aspect of landscape'. She chose two pairs of meaningful elements of micro-toponyms from Canton of St. Gallen: 'Wald', 'Holz', and 'Riet', 'Moos' , which refer to 'forest' and 'wetland' respectively and explored the relationship between them and the geographical aspects. She selected elevation and area as general physical properties, land cover or land use data from CORINE dataset, as well as topographical wetness index for the analysis, and found that: "micro-toponyms containing 'Wald' are at high altitudes with steep slopes, associated mainly with coniferous forests, followed by mixed forests and a small proportion of broad-leaved forests, while those containing 'Holz' are at lower altitudes with gentle slopes, associated with mixed and broad-leaved forests and have a negative value for coniferous forests', and "'Riet' and 'Moos' micro-toponyms are no longer associated with wetland, with 'Riet' associated with environments less specific than 'Moos' and can be linked to places at higher elevations", through the calculations of cosine similarities as well (Villette, 2021). She highlighted that the associations may be justified by these terms' semantic meanings.

To sum up, through these studies, groups of generic parts in Swiss toponyms were compared using various landscape characteristics, and their similarities and differences were discussed. As Villette (2021) said in her thesis, this kind of research could 'enable the linguistic process of the act of naming to be revealed and provide information about which features were salient enough to be used as a name for a toponym' (Villette, 2021).

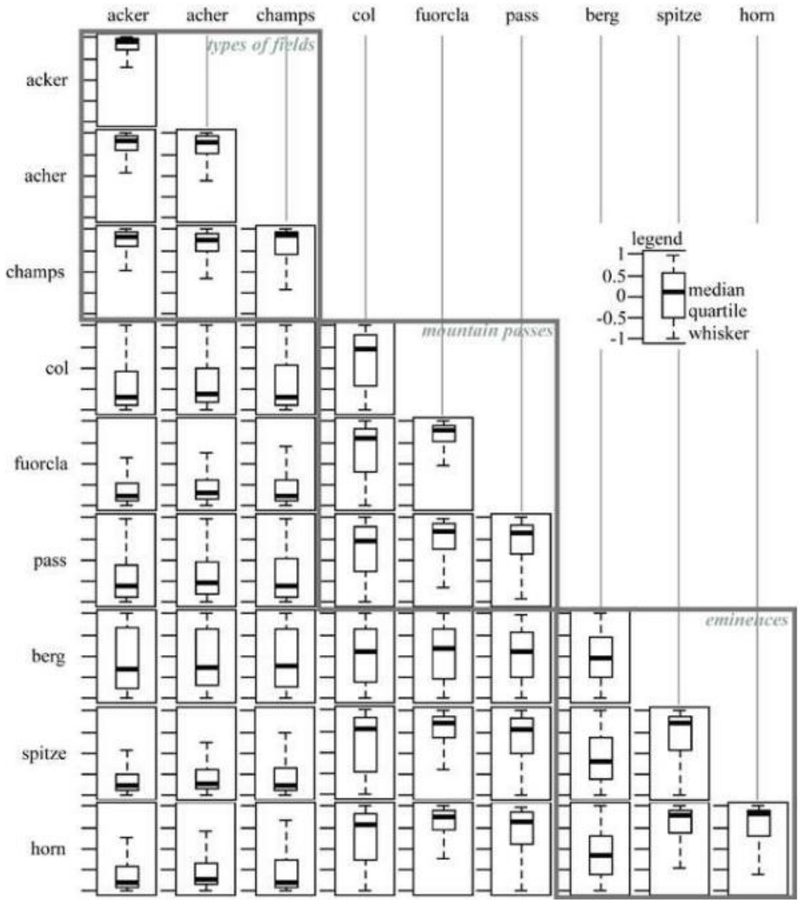


Figure 2.2: Cosine similarity results of topographic vectors from Derungs et al. (2013).

Conedera et al. (2007) investigated historical land use in southern Switzerland using toponyms referring to 'burnt', by calculating topographic parameters such as elevation, slope and aspect, potential forest vegetation unit, and the geo-referenced fire events. In their research, a probable use of 'eco-clouding' (action of burning a field to fertilize it or to stop the natural evolution of this field into a bush and a forest) were supposed to be the collective memory of ancient people. They found most of the toponyms associated with pastureland and probably also a large portion of the generic toponyms are related to the habit of using fire to clear brushwood to improve pastureland or to eliminate trees, and summarized that their research made it possible to reveal forgotten landscape practices that are still present in the toponymic landscape and highlighted the power of the information contained in toponyms (Conedera et al., 2007).

2.3.2.2 Research on Toponyms of Other Regions

Various research investigating the relationship of toponyms (proper nouns), landscape terms (common nouns), and landscape has been conducted in many other regions as well. Senft (2008) investigated toponyms of Kaile'una Island, and found that most of them are not 'semantically transparent', which means they rarely 'include' landscape terms. But it was proved that they could 'map to' plenty of landscape terms, including 'various hydrology terms, corals, islands, types of soil, gardens, villages and village parts, landmarks, plants, stones etc.' (Senft, 2008). The third question of Burenhult & Levinson (2008) was therefore answered regarding those toponyms. Cablitz (2008) and Jett (1997) studied toponyms in the Marquesas Islands and the Navajo toponyms of the Canyon de Chelly system in Arizona, respectively. Both of them found a number of toponyms containing descriptive labels of landscape features (or geographic entities). These lead to the representation of the interaction between environment and the local residents. Jett (1997) therefore summarized that toponyms may 'provide insights into cultures, linguistics, histories, habitats, and spatial and environmental perceptions'. Atik and Swaffield (2017) also verified that 'toponyms could reflect humans' connection to landscape, explaining how it was valued, changed, used or interpreted' via research focusing on toponyms in Otago Region in New Zealand.

Some researchers used toponyms to explore the landscape. For example, Atik et al. (2022) used them to explore rural landscape characteristics of the Serik district in the Turkish Mediterranean. By categorizing them into four groups: natural features (including vegetation and topography), land use, cultural features, and visual features (including color, form and position), they showed that characteristics of rural landscape there were dominated by two components: biophysical features and the position of a locality (Atik et al., 2022). They highlighted that toponyms can help identify and explain 'how the landscape looks visually', 'how it has evolved over time' etc. (Atik et al., 2022). Calvo-Iglesias et al. (2012) on the other hand, used them to identify the historical landscape of the agras field system in Northwest Spain, as they supposed that they often show a strong inertia in time, therefore they indicate the former landscape, though it might has disappeared recently (Calvo-Iglesias et al., 2012)

The spatial patterns of toponyms were also focused on by some researchers. Qian et al. (2016) revealed the historical settlement patterns of people with different ethnic groups in Guangdong, China via this. Using kernel density and Pearson's correlation, they found that, 'Zhuang and Hoklo people tended to settle in plains with adequate water resources, near paddy fields or near irrigated fields, and near wharfs, the Cantonese and Hakka people, on the opposite, tended to settle in places with higher terrains, near towns and far from wharfs' (Qian et al., 2016). Atik et al. (2022) on the other hand, used toponyms for geolocating. They found that 'the density of oronyms (toponyms associated with topography) increase where topography elevates swiftly and becomes steeper, and the density of hydronyms (toponyms associated with water) also rise correspondingly where water features are most abundant', and argued that it's possible to detect the pattern of resources in the landscape by spatial overlaying of place name densities on the actual biophysical data (Atik et al., 2022).

Cross-language research of toponyms and landscape also exist. Feng and Mark (2017) explored how gazetteers could be used to examine 'Ethnophysiography Hypothesis' mentioned above. Toponyms of mountain and hill in Malaysia and Indonesia (with generic parts: 'gunung' and 'bukit') were extracted and analyzed, revealing that they are used differently in two countries: "They are used interchangeably for eminence features in Indonesia, regardless of size (area extent), but use of ‘gunung’ is stricter for Malaysia, referring to larger eminence features” (Feng & Mark, 2017).

2.4 Research Gaps and Research Questions

2.4.1 Research Gaps

In Switzerland, though some ‘extensive’ approaches could be found studying several specific common generic parts of toponyms, their semantic geographic information, and the association of them with landscape using quantitative methods for calculation, there are still far more such generic parts waiting to be studied. As introduced in 2.3.2.1, only four generic parts including ‘Wald’, ‘Holz’ and ‘Ried’, ‘Moos’ were studied and compared by Villette (2021), and only nine other generic parts ‘Spitze’, ‘Horn’ and ‘Berg’; ‘Pass’, ‘Fourcla’ and ‘Col’; ‘Champs’, ‘Acker’ and ‘Acher’, were analyzed by Derungs et al. (2013). Still take ‘convex landforms’ as example. Among these investigated ones, only ‘Spitze’, ‘Horn’ and ‘Berg’ denotes this kind of landscape. However, upon searching on Ortsnamen.ch (Schweizerisches Idiotikon, 2023), far more terms denoting convex landforms were found emerged, like ‘Buck’, ‘Bool’, etc., both of which denote hills. More comparisons of these terms could also yield valuable and interesting information about toponyms, which has not been covered by other research.

Besides, data and variables are quite limited in those studies. Derungs et al. (2013) used DTM data only, calculating relief, standard deviation of elevation, mean slope and standard deviation of slope for the description of landscape. However, topography should not be the only element of ‘landscape’, especially while comparing generic parts that mainly denote land cover types of corresponding positions, where data of land cover should be more necessary. This would make studies quite in line with the definition of ‘landscape’ as entities ‘perceived by people’ as well (Council of Europe Landscape Convention, 2000). As for variables, using merely several values about, for example, elevations and slopes for a zonal area might be not so representative, hence a more global way for the calculation of descriptors is needed, which could also make the methods more persuasive.

Last, cross-language analysis answering questions like ‘how translatable those toponyms are’ (Burenhult & Levinson, 2008) could rarely be found in Switzerland. Though few generic parts were indeed compared by Derungs et al. (2013), there also exists far more ones to be further discussed. Through the comparison of such terms, ‘how language or administrative boundaries might impact the act of naming’ could be revealed, as Villette (2021) highlighted. Therefore, it was also worth conducting such an analysis.

2.4.2 Research Questions

My main goal of this research was to investigate and compare the exact meanings of those generic parts of toponyms conveying similar geographic semantic meanings (like ‘convex landforms) using landscape, to ‘gain more insights into the way people perceive, conceptualize, classify, and utilize the environment’, and ‘to deepen our understanding of people’s connection to the landscape’ (Thornton, 1997). More specifically, the following questions were proposed:

RQ1: Does a specific generic part of Swiss toponyms in German always relate to the same kind of landscape (including topography and land cover)? Considering other generic parts that are identified to denote same categories of geographical objects with it, do the sets of objects they denote show similar patterns regarding landscape?

RQ2: How translatable are some of the generic parts of Swiss toponyms in German compared to those in Italian, still from the perspective of landscape?

Note that, in this thesis I sometimes use 'the pattern of a generic part' to denote 'the patterns of those objects whose names contain this generic part, regarding landscape', to express in a more concise way.

Chapter 3 Data and Methods

3.1 Study Areas

German, French, Italian and Romansh are official national languages of Switzerland. Among these four languages, Swiss German, which is a collection of distinct Alemannic dialects, is the most widely spoken one. The dialects and subdialects of it are usually quite different from the formal one that is used for written communication. Besides, in the western part of Switzerland, French is widely spoken, with various dialects as well, and in the Canton of Ticino as well as the southern valleys of Graubünden, Italian is the official language. Romansh is only spoken in certain parts of Graubünden by several 10,000 people (Swiss Federal Council, 2024). Figure 3.1 shows the spatial distribution of these four linguistic areas in Switzerland (Swiss Federal Statistical Office, 2022).

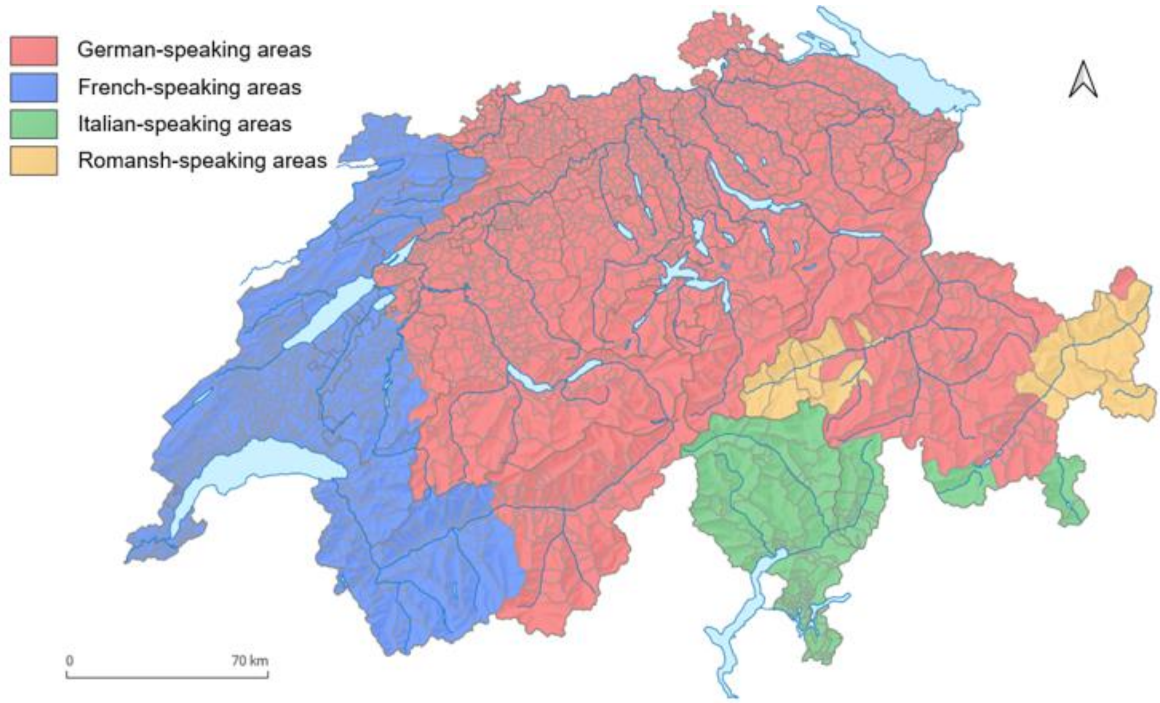


Figure 3.1: Spatial distribution of four main linguistic areas by municipality in Switzerland (Swiss Federal Statistical Office, 2022).

Due to the existence of different linguistic areas, Swiss toponyms are in different languages as well. In this research, German toponyms in Switzerland were essential for RQ1. Obviously, their related objects are mostly in German-speaking areas, with few exceptions called exonyms, whose related objects are in other linguistic areas. These exonyms are different from the official toponyms in local languages, as mentioned in 2.1.3 (Kadmon, 2002), therefore the way people named such objects initially could not be revealed, which is however essential for this research. Consequently, they were deleted during data pre-processing, which is shown in 3.2.1.2. In this case, German-speaking areas in Switzerland served as the only study areas for answering RQ1. Additionally, to answer RQ2, Italian toponyms in Switzerland were chosen for the comparisons, as resources regarding Italian toponyms (Olivieri, 1931) and their generic parts (Ursini, 2016) were found, which could be the theoretical basis for the study. Similarly, Italian-speaking areas also served as the study areas for this study.

3.2 Data and Pre-processing

To solve the research questions, toponyms data and landscape data (consisting of topography data and land cover data based on the definition of landscape for this research) were necessary. In this section I will introduce the sources and details of these datasets, as well as the pre-processing processes for getting the data needed for the analysis.

3.2.1 Toponyms Data

3.2.1.1 Data Source and Overview

A gazetteer is a geographical dictionary that provides structured information on named features, linking their names with locations and types (here 'features' means distinct physical elements or objects in the landscape, such as mountains, rivers, and buildings etc.) (Goodchild & Hill, 2008). SwissNAMES3D, as the most comprehensive gazetteer in Switzerland with a 6-year updating cycle, contains information and precise locations of over 400,000 landscape features in the Swiss Topographic Landscape Model (TLM) that have names (Swiss Federal Office of Topography Swisstopo, 2023a). As the data volume of this gazetteer is adequate for the analysis, and the attributes are helpful for filtering out proper types of toponyms for the analysis, it was considered as the data source for toponyms. The Geodatabase files from SwissNAMES3D comprises three thematic geo-referenced layers: points, polylines, and polygons, respectively. Among them, points layer includes topographic features like peaks and passes, features with field and local names (in German 'Flur- und Lokalnamen'), public transport stops, as well as buildings etc. Polylines layer includes features like transport infrastructures, sport facilities and watercourses. Polygons layer includes features like lakes, regions, terrains, and sites etc. Layer 'TLM_NAMEN_ALLE' saves toponyms and some supplementary information of each of them, connected with those three thematic layers via three relationship layers based on the UUIDs (Swiss Federal Office of Topography Swisstopo, 2023a). Note that, one single toponym could represent multiple features, as distinctive features might share one exact name, like 'Pfäffikon' exists in both Schwyz and Zürich. Conversely, a single feature could have multiple names as well, for example, 'Genève' as an endonym and 'Genf' as an exonym. Figure 3.2 from Swiss Federal Office of Topography Swisstopo (2020) shows the logical database model of the layers.

For achieving the research goals, the points layer was selected as it is the largest (with more than 300,000 records) among three thematic layers, with many records denoting landscape-related objects. The other two layers are, on the contrary, relatively smaller, and most records are human- or water related. As there aren't various generic parts in hydronyms in Switzerland for conducting the comparison (Schweizerisches Idiotikon, 2023), and human-related objects' names could lead to biases as discussed in 2.1.3, these two layers were excluded. Table 3.1

summarizes the attributes of the points layer and the toponyms layer, with explanations of their attributes according to Swiss Federal Office of Topography Swisstopo (2020).

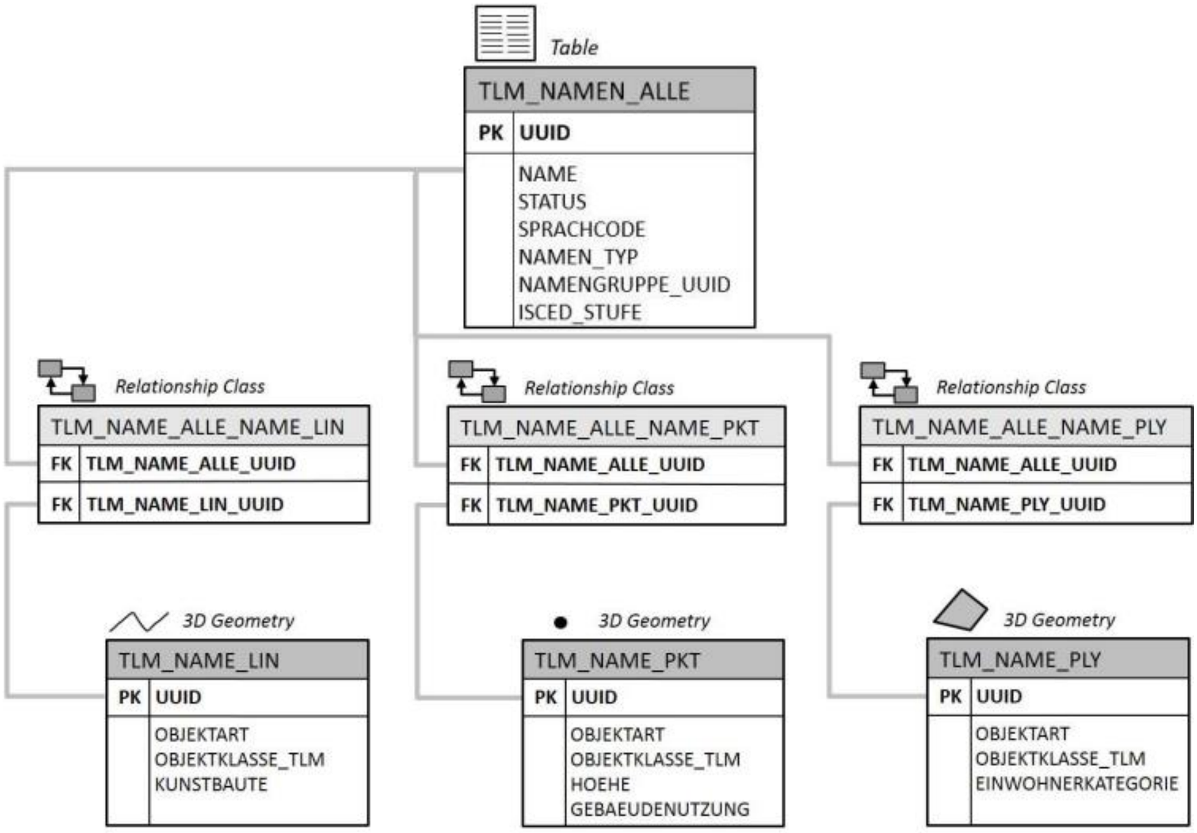


Figure 3.2: The logical Geodatabase model of SwissNAMES3D. The top one is the toponyms layer, those three at the bottom are objects layers including points, polylines, and polygons, while the others contain their relationships (Swiss Federal Office of Topography Swisstopo, 2020).

3.2.1.2 Data Pre-processing

To ensure the analysis to get proper, reasonable and valid results, I cleaned the data after joining the points layer and the toponyms layer together according to the attributes. The concrete filtering processes and reasons are as follows:

- As claimed in 2.1.3, odonyms and urbanonyms were not considered in my study. Thus, only objects with ‘OBJEKTARTKLASSE_TLM’ codes to be ‘700’, ‘1100’ or ‘1400’, whose ‘GEBAEUDENUTZUNG’ codes are ‘999998’ were kept (see Table 3.1 for what these codes denote). Among them, objects with ‘OBJEKTART’ codes to be ‘1409’ were also deleted, since it represents ‘street passes’ whose names are also odonyms (Swiss Federal Office of Topography Swisstopo, 2020). Regarding ‘ISCED_STUFE’ code from the toponyms layer, meaning ‘the level of International Standard Classification of Education’ (Swiss Federal Office of Topography Swisstopo, 2020), toponyms with this attribute not being ‘999998’ indicate that the related objects are education-related, therefore their names might be urbanonyms, and in theory should be deleted. However, after the previous processing, for all the remained records, this value was ‘999998’. Hence, nothing was needed to do with it anymore.

- As mentioned in 2.1.3, problems might be raised due to exonyms while conducting the analysis, as they sometimes could not reflect how people perceived landscape initially, even if the generic part of it might have been translated in some way, for example, from ‘Berg’ in German to ‘mont’ in French. The reason is that, though here ‘Berg’ and ‘mont’ both denote mountains (Schweizerisches Idiotikon, 2023), the types of mountains they denote could still be distinct, which is exactly my hypothesis. Simply assuming they are the same equal to simply denying the hypothesis without evidence. Therefore, those objects with their names’ attribute ‘NAMEN_TYP’ to be ‘2’ were deleted. After this, no record with ‘STATUS’ code to be ‘3’ was remained anymore (see Table 3.1 for what these codes denote). Therefore, nothing was needed to do for it.

Though the attribute ‘HOEHE’ contained in this dataset saves the elevations of related features, it is still far from being able to represent ‘topography’. To realize this in a comprehensive way, zonal operations and functions were employed in this research. With such implementation, the elevation values of both the object’s position and its neighbors were considered for representing topography (DeMers, 2001; Tomlin, 1990). In this case, more details about, for example, elevation and slope, of the surrounding area of this object could be told. The related algorithms will be introduced in a more concrete way in 3.4. Hence, additional data containing global topographic information of Switzerland was emergently required. In the next sub-section, such data will be talked about.

3.2.2 Topography Data

3.2.2.1 Data Source

SwissALTI3D is an extremely precise high-resolution digital terrain model (DTM) describing the surface of Switzerland without vegetation and development. It is delivered as a grid with an aperture width of 0.5m or 2m, and the size of each cell is 1km * 1 km. Compared with another product called ‘DHM25’, which saves the elevation of the surface of Switzerland with an aperture width of 25 meters, this product could obviously reflect more detailed information about the earth surface (Swiss Federal Office of Topography Swisstopo, 2004, 2024). The coordinate system of this product is also LV95 LN02, and it’s also updated every 6 years (Swiss Federal Office of Topography Swisstopo, 2022, 2023b). The TIFF files from Swisstopo are spatial indexed based on the grid, in other words, each cell of the grid is saved as one single TIFF file. The indices are saved as a part of the file names. For example, the file named 'swissalti3d_2019_2501-1120_2_2056_5728' indicates that, the related cell’s lower left corner is at (2501000, 1120000) and upper right corner is at (2502000, 1121000). In all, there are 43,628 TIFF files, which implies that there are 43,628 1km * 1km cells, in other words, the data covers an area of 43,628 km^2 in all. That is a bit larger than the area of Switzerland, as the data of Liechtenstein and of Lago (Lake) Maggiore are also contained. Meanwhile, 25 meters were extended beyond the borders of Switzerland and Liechtenstein (Swiss Federal Office of Topography Swisstopo, 2023b). This is helpful for the research, as investigating the topography of the zonal area of a position that stands too close to the borders became more convenient.

Table 3.1: Overview of the points layer as well as the toponyms layer in the Geodatabase of SwissNAMES3D (Swiss Federal Office of Topography Swisstopo, 2020)

Layer	Attribute	Introduction
Points	OBJEKTARTKLASSE_TLM	Classifies objects into eight main classes with the following codes as labels: objects with label '400' are buildings, with label '700' are small morphological objects, with '800' are single objects, with '1100' are objects with local or field names, with '1400' are passes and peaks, while with '1700', '2000' and '2200' are for transportation.
	OBJEKTART	Further classifies objects into 36 sub-classes based on 'OBJEKTARTKLASSE_TLM'.
	GEBAEUDENUTZUNG	Represents the building types of objects. For objects that are not buildings, it is '999998'.
	HOEHE	Saves altitudes of objects.
	geometry	The geometry of an object is in (X, Y, Z) format with LV95 LN02 as the coordinate system.
Toponyms	NAME	Saves toponyms.
	STATUS	A toponym labelled '1' means it is the main or the official toponym, labelled '2' means it is another toponym used by people, labelled '3' means it is an exonym.
	SPRACHCODE	Represents the language of a toponym.
	NAMEN_TYP	A toponym labelled '100' means it is a unique toponym, labelled '0' means it is an endonym, labelled '1' means it is an exonym, labelled '2' means it is a pair of different toponyms.
	ISCED_STUFE	Classifies the toponyms based on International Classification of Education (ISCED). For those toponyms not about education, it is '999998'.

3.2.2.2 Data Pre-processing

For this research, I chose the DTM data whose grid has an aperture width of 2m, as this data is meticulous enough for representing the whole map of an area's topography while using zonal operations. Note that some smaller geographic features, such as mounds or holes created by marmots, cannot be detected via this data due to their size, but they are not considered part of 'landscape' in this context. As for the 'grid-based spatial indexing', though it could expedite spatial queries (Rigaux et al., 2022), it's actually not helpful for this research as I did not need to do any spatial query at all, but only matrix slices, while generating the zonal area as the research area of a position, which will be introduced in 3.4. The analysis could even take more time with the existence of it, as data from multiple cells, in other words, multiple files, might be required while generating zonal areas if the position is close to the boundary of a cell, or the buffer area is too large, or both. Accessing more files implies taking longer time. Thus, a large bounding box of Switzerland was created, containing all these cells, with the DTM data. Figure 3.3 a) is a visualization of how this DTM data looks like after merging. For those places lying outside of Switzerland and Liechtenstein but being included in the bounding box, elevation data was filled with -9999.0 to keep the regular boundary of the bounding box, which is parallel to what Swisstopo did for all the cells.

3.2.3 Land Cover Data

3.2.3.1 Data Source

Swiss land use statistics provides detailed and accurate information on land use and land cover in Switzerland determined via aerial photographs, together with the help of swissALTI3D and DHM25 etc., containing geo-referenced data of 4.1 million points, which distributes at 100m intervals regularly (Swiss Federal Statistical Office, 2021). Each point is assigned an X and Y coordinate for georeferencing, along with a 'land use' category and a 'land cover' category. Land use here refers to the socio-economic use of the land, while land cover here refers to the physical cover of the land, according to 2.2.2. With these two categories, numerous ways for classifying land cover are introduced. First, the combination of these two forms the standard nomenclature with 72 categories. They are further aggregated into 4 main domains and 17 and 27 classes for better statistical reliability and for the use on a small scale. Additionally, the land use nomenclature NOLU04 has 46 core categories, and are further aggregated into 10 classes and 6 domains. Last, the land cover nomenclature NOLC04 has 27 core categories, and are further aggregated into 6 domains. These 9 classification methods, delivered 4-5 times during different time periods with a 6-year updating cycle, result in 36-45 indices in all for each point (Swiss Federal Statistical Office, 2021). Besides, the coordinate system of this product is also LV95 LN02 (Swiss Federal Statistical Office, 2021).

3.2.3.2 Data Pre-processing

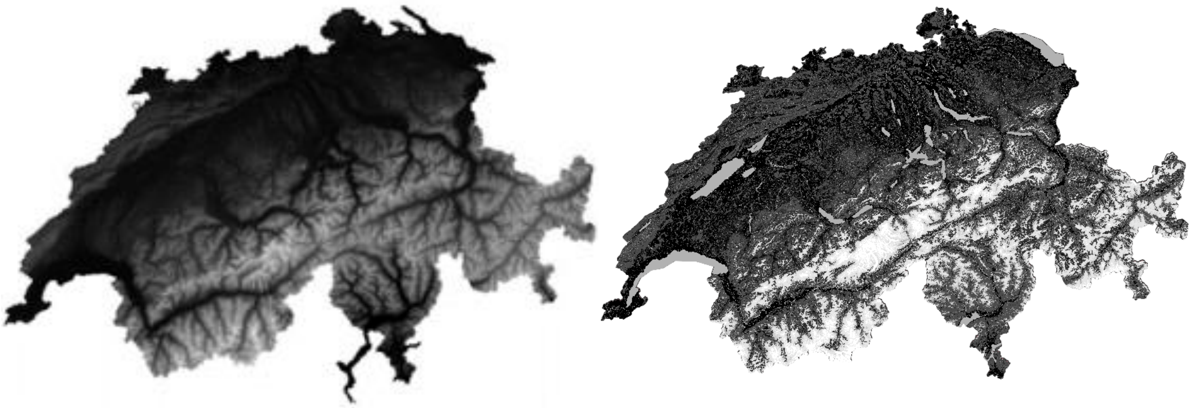
In order that this research could get valid results, which is also in line with the research goals, the following decisions were made to find the best way to classify land cover.

- During the most recent updating period from 2020, only data of cantons of Vaud and Geneva have been delivered. The latest fully-available data was gradually updated from

2013 to 2018 (Swiss Federal Statistical Office, 2021). In this research I decided to use the data from the latter one only. Of course, another option is to use the newly updated indices for points in cantons of Vaud and Geneva and keep the data from the previous delivery for everywhere else. But as these two cantons totally lay in the French-speaking area of Switzerland as shown in Figure 3.1, which is not the study area of this research, such effort is obviously unnecessary. Therefore, I used the classification results updated from 2013 to 2018 only.

- As for classification methods, the one combining 'land cover' and 'land use' defined by Swisstopo together and classifying those points into 27 different land cover or land use types fulfills my requirements the best, as it classifies land cover into enough but not overly detailed classes. However, as 'land use' was not what I cared about in my study, according to 2.2.2, further aggregation was needed to make the land cover classification more fitted. These 27 classes were eventually aggregated into 11 new classes, which enabled me to get more proper results. The ways of this aggregation and the reasons why some of the original classes were aggregated to a single new class, or why some original classes were retained, are concluded in Table 3.2.

Following these pre-processing processes, the '.csv' file was converted into vector data, using the coordinates contained in the file, then into raster data with the same bounding box as topography data to maintain methodological consistency. Likewise, for those places that lay outside of Switzerland and Liechtenstein within this bounding box, land cover data was filled with -9999.0 as well. Figure 3.3 b) is a visualization of the land cover data after merging these categories. Note that, in Figure 3, for both topography and land cover, the brighter a place is, the larger the value it has. In the case of topography, the 'value' represents altitude, whereas for land cover, it corresponds to the code of land cover category (see Table 3.2 for the type of land cover a code denotes).



(a) (b)

Figure 3.3: The visualizations of topography data (a) as well as land cover data (b) after data pre-processing.

Table 3.2: Reclassification of land cover

New Class	Original Class(es)	Reasons
Human-built or -changed objects (new code 1)	Industrial and/or commercial areas, residential areas, public buildings, agricultural buildings, unspecified buildings, roads, railways, airports, special settlements, recreational and green areas, fruit growing areas, vineyard areas, horticultural areas, farmlands (codes 1-14).	As pointed out above, land use was not concerned in the study, therefore they are aggregated. I only used this class to detect whether an object had been changed by people a lot in the analysis. Since for farmlands a lot of exploits were made as well, it's also aggregated into here.
Grasslands (new code 2)	Natural grasslands, Alps grasslands (codes 15, 17).	Differences exist in elevation, which DTM could represent. Pastures were not aggregated to class '1' as it's a land cover type and is not changed a lot since ancient times, hence it's worth being kept unchanged.
Pastures (new code 3)	Home pastures and Alps pastures or Jura pastures (codes 16, 18).	
Forests (new code 4)	Forests (code 19).	The densities or types of vegetations in these classes are different (Swiss Federal Statistical Office, 2021). In this case, they were kept unchanged.
Dissolved forests (new code 5)	Dissolved forests (code 20).	
Shrublands (new code 6)	Shrublands (code 21).	
Trees (new code 7)	Trees (code 22).	
Water areas (new code 8)	Lakes and rivers (code 23, 24).	Hydronyms were not investigated.
Unproductive vegetation (new code 9)	Unproductive Vegetation (code 25).	They represent different land cover categories under the definition of landscape.
No vegetation (new code 10)	No vegetation (code 26)	
Glaciers (new code 11)	Glaciers (code 27)	

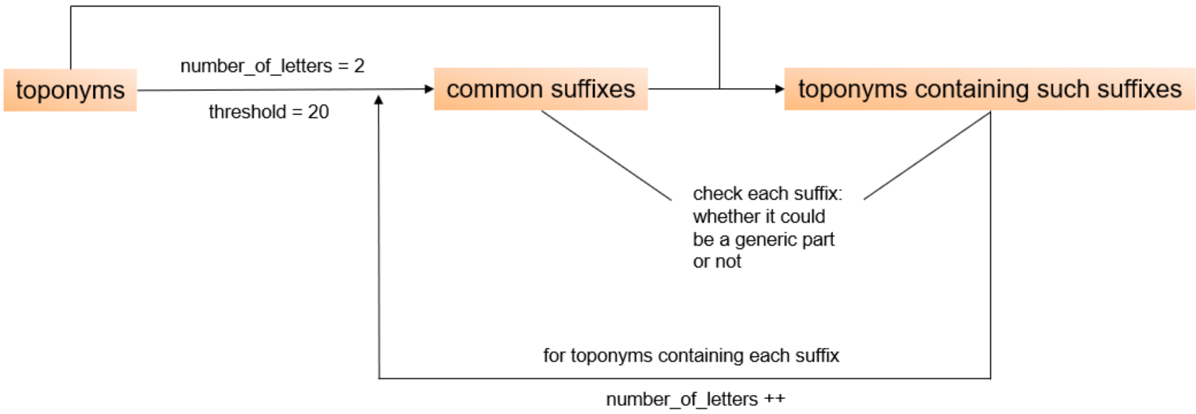
These pre-processed data were used to address the research questions through two phases of analysis, which will be introduced in the next two sections respectively. The first of which involved the exploration of the generic parts contained in German and Italian toponyms in Switzerland, and the extraction of toponyms that contain them, which would be introduced in section 3.3. While in the second phase, I investigated the landscape (including topography and land cover) of objects whose names contain each generic part using 'bags of words', and compared the patterns of those generic parts that were identified to denote similar kinds of objects, which would be talked about in 3.4.

3.3 Extraction of Generic Parts

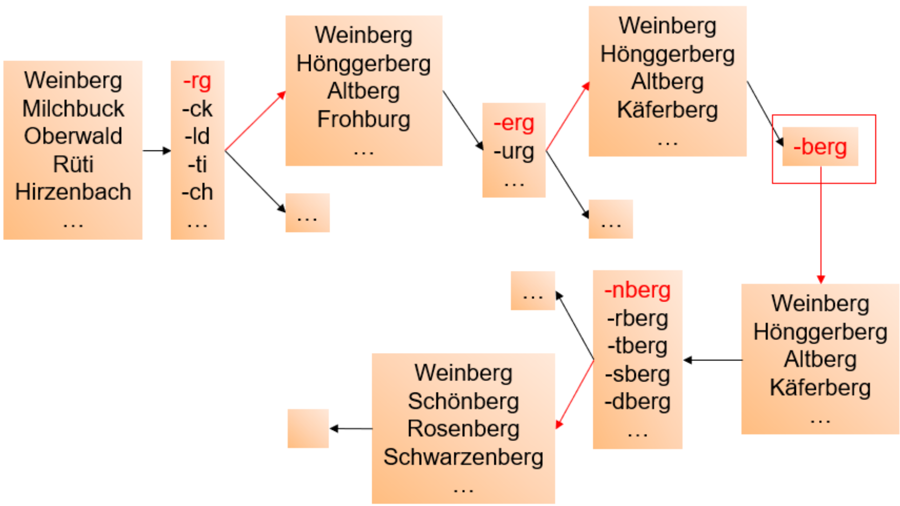
Traditionally, the process of geographic knowledge discovery is the formation of a conceptual view of the system first, followed by the development of hypotheses to be tested, and finally the groundwork to verify or falsify these hypotheses, but with the development of Big Data, data-driven science started to play a crucial role in geo-knowledge discovery, making data-exploration to be an important part for the formation of conceptual view and for hypothesis generation (H. J. Miller & Goodchild, 2015). In the context of this topic, a traditional approach might involve collecting generic parts of toponyms based on theories, then investigating the landscape of objects whose names contain such generic parts. Nevertheless, toponyms dataset I get allowed for a data-driven approach, as it provides massive toponyms data in Switzerland for various kinds of objects, therefore it's helpful for the discovery of generic parts. Compared to the traditional way, this approach is more objective and could reflect reality better, since it could bring out those generic parts that have not been studied yet. In the following paragraphs, such methods would be shown. As German toponyms and Italian toponyms are quite different in their forms, they were processed in diverse ways, which would be discussed separately as well. The attribute 'SPRACHCODE' as shown in Table 3.1 could be helpful to distinguish between German toponyms and Italian toponyms in Switzerland.

3.3.1 German Toponyms in Switzerland

Germanic toponyms are often compound in structure, with qualifier parts coming first followed by generic parts (Hough, 2006). Therefore, extracting generic parts could be transformed into extracting suffixes. The process started by extracting the last 2 letters of all German toponyms in Switzerland, as there isn't any single letter that could serve as the generic part itself. Grouped by these 2-letter terms themselves, the number of occurrences for each of them in toponyms was calculated. Since it was suggested that, in order that clustering method, which I intended to use in the second phase, could get valid and representative results, the sample size should better be at least 10 times the number of clustering variables (Qiu & Joe, 2009; Sarstedt & Mooi, 2019), I decided that only when a generic part's occurrences in toponyms exceeds 20 could it be kept for the next step, as I expected to have at least two variables: topography and landscape. Based on this, a number of 2-letter suffixes were found, and for each of them, I determined whether it could serve as a generic part or not, referring to sources about Swiss toponyms (Huber, 1885; Schweizerisches Idiotikon, 2023). Then, the process went on to the next iteration. For those toponyms containing each of these 2-letter suffixes, I proceeded to find 3-letter suffixes, still with the threshold as 20, and checked their meanings as well to find whether they could indeed be generic parts. The process went so on and so forth like this, by adding number of letters until nothing could be found. Throughout the process, once a suffix was identified to be a generic part in toponyms (but not in toponyms about human-built objects in urban areas, about water areas or about very tiny objects), it was noted down. Note that, this iteration must not be stopped after this, as there might be another generic part, containing this generic part as its suffix, and they should be distinguished from each other (like ‘Schorn’ was found to be a generic part in toponyms containing ‘Horn’, which is also a generic part). Figure 3.4 (a) shows a schematic diagram of this process.



(a)



(b)

Figure 3.4: (a) A diagram showing the process of German toponyms’ generic parts extraction. (b) An example of German toponyms’ generic parts extraction.

Figure 3.4 (b) illustrates an example of this generic parts extraction process. It starts with a list of toponyms from the gazetteer, which is the upper left one. By extracting suffixes with 2 letters, it gets the result shown in the next box. Then, for each suffix, it generates a new list of toponyms. For instance, the third box containing ‘Weinberg’ etc. is the toponyms with suffix ‘rg’ (which appears over 5,000 times in toponyms), then a new list containing 3-letter suffixes is generated. This process continues so on and so forth like this, with longer suffixes, until no suffixes could be found. During this process, for each suffix I found, resources are used to investigate their meanings, with which 'Berg' is found to serve as a generic part in toponyms, means 'mountain' which is related to landscape (Schweizerisches Idiotikon, 2023). With the extraction process for all those 3-letter suffixes ending with 'rg' finished, the study turns to 'ck'. Note that, the toponyms and suffixes shown in this figure are only for illustrative purposes, they could of course not represent the whole gazetteer, which contains more than 300,000 toponyms.

During this process, some variants of the generic parts could be found as well. Besides of plural forms, it was also not a surprise that some generic parts of toponyms had evolved into various written forms in different parts of German-speaking areas in Switzerland, as Swiss German has plenty of dialects. For example, 'bärg' was found to be another form of 'Berg' and 'Stein' was found to be written as 'stei' based on some dialects (Schweizerisches Idiotikon, 2023). Some of them could be found during extraction, as they appear a lot, while some of other rarely used terms (even with less than 5 appearances), could also be mined with the help of rich information from Schweizerisches Idiotikon (Huber, 1885; Schweizerisches Idiotikon, 2023). For example, while searching for information about 'Berg', I found that it has another form 'bärg'. Moreover, it was also discovered that, the diminutive forms of some generic parts also exist. For example, 'Bergli' was discovered to be a diminutive form of 'Berg' (Schweizerisches Idiotikon, 2023). Comparing the use of these two could also yield interesting insights, through which one might be able to answer: 'Does a diminutive really denote smaller objects, compared to 'normal' forms?'. Thus, they were also kept for the analysis in this research.

During suffixes extraction, constructions with the form "preposition + (article +) toponym" were discovered, for example, 'Hinder Rötelberg' and 'Bi da Platta', etc. These could probably lead to biased results, as 'Bi da' means 'near the', and 'Hinder' means 'behind' (Huber, 1885). The landscape of an object 'behind a mountain', 'near the flat', for example, could be different from it of exactly the mountain, or the flat. Therefore, they should not be mixed up. There is of course no doubt that forms like 'Hinder *berg', 'Bi da *platta' can be regarded as other generic parts, however, as constructions with form "preposition + (article +) toponym" were only rarely found, they were not studied in the research, and such toponyms were deleted from the dataset to avoid biases. Note that, words that sometimes serve as prepositions could serve as qualifier parts of toponyms as well, like 'Hinderberg' and 'Oberplatta'. Here 'Hinder' and 'Ober' serve as 'descriptions', rather than 'prepositions', hence they were not deleted.

Via these steps, German toponyms with different generic parts (including different variants due to plural forms and dialects, and diminutive forms) were filtered out. Meanwhile, toponyms composed of one single term, like 'Berg', were also included. They could be regarded as being composed of barely generic parts, without any qualifier parts providing descriptions about the place, as stated in 2.1.4. A new dataset was built for each of the generic parts thereafter.

3.3.2 Italian Toponyms in Switzerland

Italian toponyms have a different form from German toponyms. As mentioned in 2.1.4, the pre-nominal parts of them serve as spatial classifiers, namely generic parts (Ursini, 2016). Thus, the extraction of generic parts of Italian toponyms could be transformed into the extraction of the first words of toponyms, which is simpler than extracting generic parts in German toponyms as iteration is not needed here. Figure 3.5 shows the process for extraction.

Similarly, the threshold was set as 20 here as well, in order that the clustering analysis could get valid results. Some variants of generic parts could also be found for Italian toponyms, as various dialects also exist in Lombard language, which is the vernacular language spoken in Italian-speaking areas in Switzerland (Jones, 2015). For example, 'Mott' was discovered to be another form of 'Motti', which means 'hill' (Olivieri, 1931). Such various forms were collected as well. Additionally, via the extraction of the first words of toponyms, some prepositions and articles could also be found, as they also appear at the first place (Ursini, 2016). Like German toponyms, these terms could lead to biases, therefore toponyms containing them were deleted.

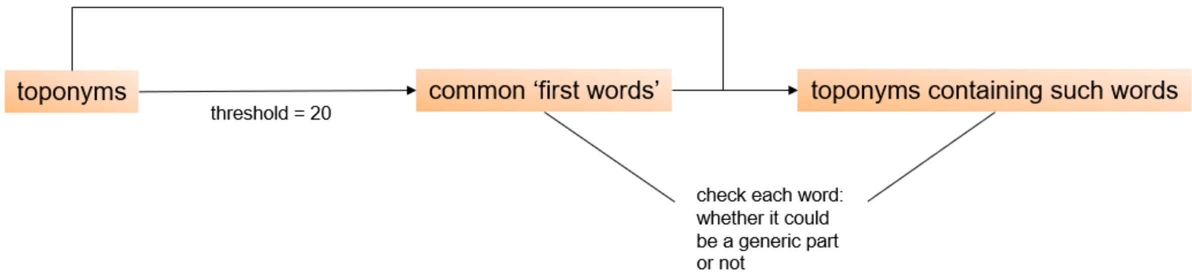


Figure 3.5: A diagram explaining the process of Italian toponyms' generic parts extraction.

After these steps, toponyms with different generic parts in Italian (as well as variants in plural forms or with dialects) were filtered out as well, and a new dataset was built for each of them.

The results obtained in this phase were saved in a database. In detail, this database includes multiple layers, each of which saves the toponyms with the same exact generic part (sometimes with variants), as well as the corresponding positions. German toponyms with generic parts of diminutive forms were put into distinct layers from those with 'normal' generic parts. All these generic parts were further classified into several classes, regarding their 'semantic geographic information', in other words, the typologies of landscape they denote. This will be concluded in 4.1. This new database paved the way for phase 2 of my research.

3.4 Comparison of Generic Parts using Landscape

To address the research questions, I conducted comparisons among generic parts classified into the same category in the first phase (for example, generic parts conveying semantic meanings about 'convex landforms', which will be concluded in 4.1) using landscape. In this section, I'm going to present the approach adopted for the comparisons.

3.4.1 Construction of Research Areas

As claimed in 3.2.1, I employed zonal operations in this study to describe a position's landscape. Specifically, this involved considering not only the elevation value as well as the land cover code of this position itself, but those in its neighboring areas as well. A global view of this position’s features in topography and landscape could therefore be shown. Obviously, before proceeding with the analysis, the construction of zonal areas for all the positions to be studied was needed, to achieve this, rectangle buffer areas were applied, and proper buffer distances were determined as follows:

For topography, I initially tested 500m, 1000m and 2000m as distances to construct buffer areas. An area might lack representativeness while using buffer distance of less than 500m, as it might fail to capture the topography of the entire peripheral area of a position, while too large an area with a buffer distance exceeding 2000m (means this area is larger than 4000m * 4000m in area) could easily incorporate unrelated features, such as a distant peak near the position under study, which is actually the position of an object in the flat. Similarly, for land cover, I selected buffer distances of 200m, 500m, and 1000m, as too large an area can include nearby villages, leading to potential biases of the results. After conducting tests using data from various layers in the database, I found that, for topography, buffer distances of both 500m and 1000m worked well, while with 2000m as the buffer distance, some distant geographical features could already be included. Likewise, for land cover, a buffer distance of 200m yielded optimal results, while of 500m and 1000m could lead to the inclusion of many villages nearby. Consequently, I decided on a buffer distance of 1000m for topographic analysis, and of 200m for land cover analysis.

To generate the buffer areas of positions in Switzerland, one must handle the coordinate system carefully. The coordinate system of all the datasets is LV95 LN02, where points on the lower border of an area should have the smallest Y coordinate. However, in Python matrices, points on that border have the largest X coordinate. Additionally, the lower left point of the DTM and the land cover data has coordinates (2480000, 1070000), which is the origin of the entire area. Yet, the origins of matrices in Python are always the upper left point. Furthermore, since the topography data has an aperture of 2m and the land cover data has a resolution of 100m, buffer areas must be divided by these values respectively to obtain the correct research areas needed. Given these considerations, coordinate transformations should be done carefully. To illustrate this process, let's take the generation of the research area for investigating land cover, for the point located at (2684912, 1250492) as an example:

- First, find the nearest point with valid land cover data. Both the X and the Y coordinate of this point could be divided by 100, resulting in (2684900, 1250500).
- Then, apply the buffer distance of 200m to determine the borders of the research area, using the LV95 LN02 coordinate system. The lower left point should be calculated as (2684700, 1250300), while the upper right point is (2685100, 1250700).
- Divide the coordinates by the aperture value (100), then set the origin’s coordinate from (2480000, 1070000) to (0, 0) by subtracting 24800, 10700 from X and Y values of all points in the area respectively. After this transformation, the coordinate of the lower left point of this area is (2047, 1803), while the upper right point is (2051, 1807).
- Change the origin point from the lower left to the upper left, and exchange X and Y coordinates. Since the entire area has 2300 rows, the X value of the upper left point is adjusted to ‘2300 minus the maximum Y value’. With this ‘Y_max’ being 1807, this X value becomes 493. Similarly, the lower right point’s X value becomes 497. Their Y values correspond to the X values got in the last step respectively.

- Eventually, the research area's upper left point would be at (493, 2047) while the lower right point would be at (497, 2051), with the coordinate system of a 2D NumPy array.

Finally, if a research area contains -9999.0 in its values, this area should be excluded from the analysis, as -9999.0 represents 'no data'. This situation typically occurs when the area extends beyond Swiss borders, which is not common as not so many objects stand near the border. Thus, this exclusion should not cause significant data reduction issues.

3.4.2 Calculation of Local Descriptors

In this step, I described each research area with a local descriptor of landscape using histogram. The distribution of values about topography and land cover of this area served as the values of it. For topography, elevation and slope were both included in constructing the histogram. The reason for using slope is that it could describe the undulation of the terrain, which could not be well-represented by the distribution of elevation. For land cover, simply using the classification codes for generating the histogram could be fine. The way to calculate slope at a position is:

$$I_x(i,j) = \frac{I(i,j+1) - I(i,j-1)}{4}, I_y(i,j) = \frac{I(i+1,j) - I(i-1,j)}{4}$$

$$magnitude = \sqrt{{I_x}^2 + {I_y}^2}, direction = \arctan\left(\frac{I_y}{I_x}\right)$$

where 'I(x, y)' means the elevation of position (x, y), while 'I_x(x, y)' and 'I_y(x, y)' mean the derivative of the terrain with respect to x and y direction, respectively. Here I concerned only about the value of slope (which is the magnitude of the gradient), while the direction was not calculated. The reason why the denominators for calculating the gradient were 4 rather than 2, which is normally used is, the aperture value of DTM data is 2, rather than 1.

With these values, a 31-bin histogram combining the distribution of elevations (occupy 10 bins), the distribution of the magnitudes of slopes (occupy 10 bins), and the distribution of land cover types (occupy 11 bins), were generated for each research area. For elevation, cut-off values were decided to be 400, 800, 1200, 1600, 2000, 2400, 2800, 3200, 3600, which allowed most positions in Swiss plateau and Alpine valleys in German-speaking areas fall within the second bin, and valleys in Ticino fall within the first and the second bin. The last 5 bins represent areas in the major region of the Alps, and the others could denote 'Voralpen' (Pre-Alps) areas (Swiss Federal Council, 2023; Swiss National Center for Climate Services NCCS, 2023). Cut-off values for slope were determined to be 0.15, 0.3, 0.45, 0.6, 0.75, 0.9, 1.05, 1.2, 1.35, since these values were found to represent different terrain undulation levels the best. These correspond to the slopes of 6, 12, 18, 24, 30, 36, 42, 48, 54 degrees respectively, given the assumption that diagonal directions always maximize gradients. Besides, for land cover, things were simpler, with each bin representing a distinct land cover type mentioned in 3.2.3.2. Overall, the sum of all the values of a histogram should be 3, with 1 to be the sum of frequencies for elevation distribution, 1 to be the sum of them for slope distribution, and 1 to be the sum of them for land cover distribution.

3.4.3 Construction of ‘Bag-of-Words’

Bag of words (BoW) is a feature extraction technique commonly employed in machine learning and information retrieval algorithms. It is a simplified representation of all of the words in a text according to frequency while ignoring grammar, word order and context (Mitkov, 2014). Also, it’s used frequently in solving problems in Computer Vision (Sivic & Zisserman, 2009). In this study, this technique was employed for representing landscape. Through the last step, one histogram was generated for each position, and here, histograms for all the positions in one exact layer (where the toponyms contain a same generic part and its variants) were put together, forming a bag of ‘words’. In this way, each generic part was associated with a ‘bag of word’, representing landscape characteristics. Figure 3.6 shows the process of this construction, with the upper one to be the 31-bin histogram, or the ‘local descriptor’ of the landscape at one exact position, and the lower one to be the two-dimensional ‘bag of word’, containing that ‘local descriptor’ as one vector. In the next step, this ‘bag of word’ was used for further analysis, which will be discussed in the next sub-section.

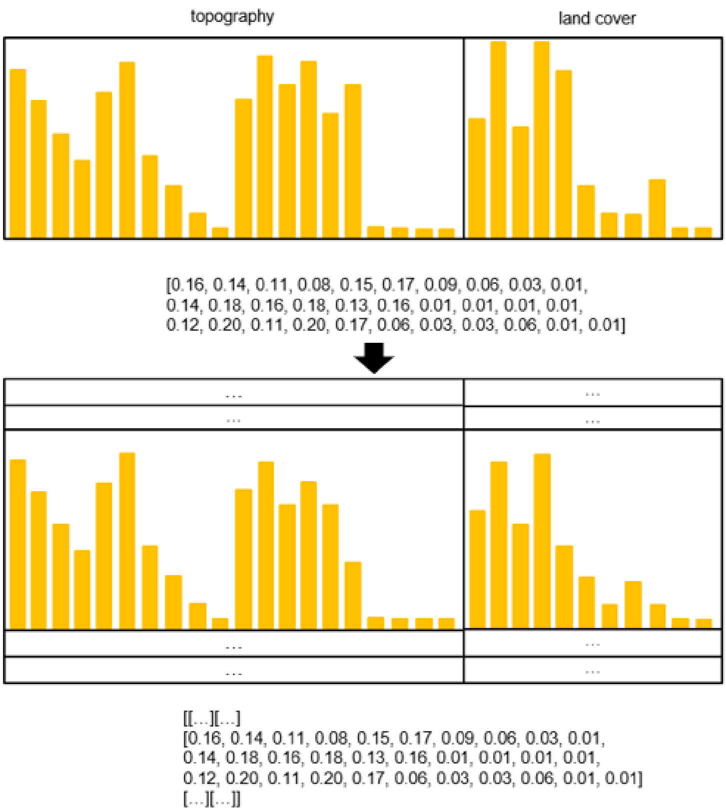


Figure 3.6: An illustration of ‘bag of word’ construction using ‘local descriptors’. Note that, the values shown here are only for illustrative purposes and are not from any ground truth.

3.4.4 Construction of ‘Landscape Dictionaries’

In the previous step, a bag of word was built for each layer, while to compare the generic parts conveying semantic meanings within the same category, all of these ‘bags’ were aggregated for each category, which led to the creation of several more comprehensive bags of words, with labels being the related generic parts. Following this, for getting a smaller amount of ‘landscape words', K-Means clustering was applied to partition all the histograms (or local descriptors) within each comprehensive bag of words into k clusters. Thereby, k cluster centers for each of the categories were determined, forming a 'landscape dictionary'. 'Elbow method' was used for defining a proper k value. In detail, via trying different values of k for a K-Means clustering, the plot of Within-Class Sum of Squares versus k could be generated, as shown in Figure 3.7, and the 'elbow point' could be determined via the plot. The results of this step were employed for the comparisons among generic parts, which will be introduced in the next subsection.

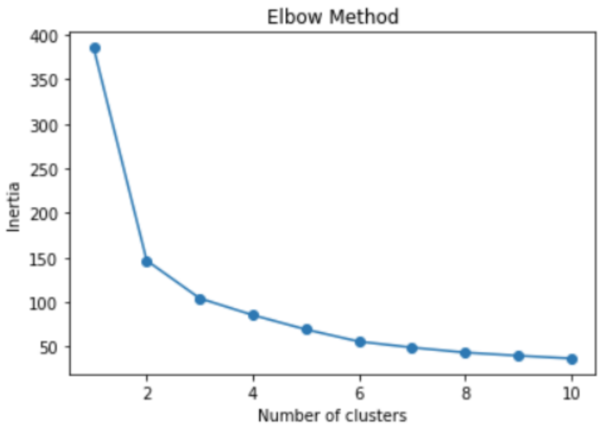


Figure 3.7: An example of the plot for 'Within-Class Sum of Squares versus k'. Here the elbow point should better be at the position with k=2, as the WCSS could be significantly decreased compared with it when k=1, but with further increasing of k, no significant change of WCSS could be witnessed. Note that, the plot shown here is only for illustrative purposes and is not from any ground truth as well.

3.4.5 Comparison of Generic Parts

In this step, the distribution of the local descriptors within the bag of words for each generic part was calculated by assigning each labeled descriptor to a cluster according to the K-Means clustering result of the related category as the one shown in the upper part of Figure 3.8 (which is the 'comprehensive bag of words' for convex landforms), while the lower part of this figure shows an example of the final clustering result. Such results could provide a more global view of the landscape each of the layers represents, allowing comparisons among generic parts based on their patterns. Furthermore, for pairs or sets of generic parts seeming to have similar patterns, cosine similarities were calculated, to investigate whether they indeed represent similar objects or not. The following formula was used for the calculation of cosine similarity of two vectors, while 'a' is a vector (i.e., a descriptor) in bag of words 'A', and 'b' is a vector in bag of words 'B' (Manning et al., 2009):

$$\cos(a, b) = \frac{a \cdot b}{||a||\ ||b||}$$

Using this method, with m descriptors in 'bag' A and n descriptors in 'bag' B, m*n results could be calculated, where each one represents the cosine similarity of a vector in A and it of a vector in B, without repetition. The results were then used for analyzing the similarities of A and B. If the cosine similarities for most pairs of vectors from two bags respectively are close to 1, the corresponding generic parts could be considered as denoting similar landscape. While if they are mostly close to 0, the corresponding generic parts could be seen as denoting totally different landscapes. No cosine similarity result here could be negative, as all the values in all histograms were greater or equal to 0.

For each 'bag of words' denoting each of the generic parts, K-Means clustering was also used to get auxiliary results about whether this generic part could denote totally different kinds of landscape or not, together with their patterns generated via the above-mentioned processes.

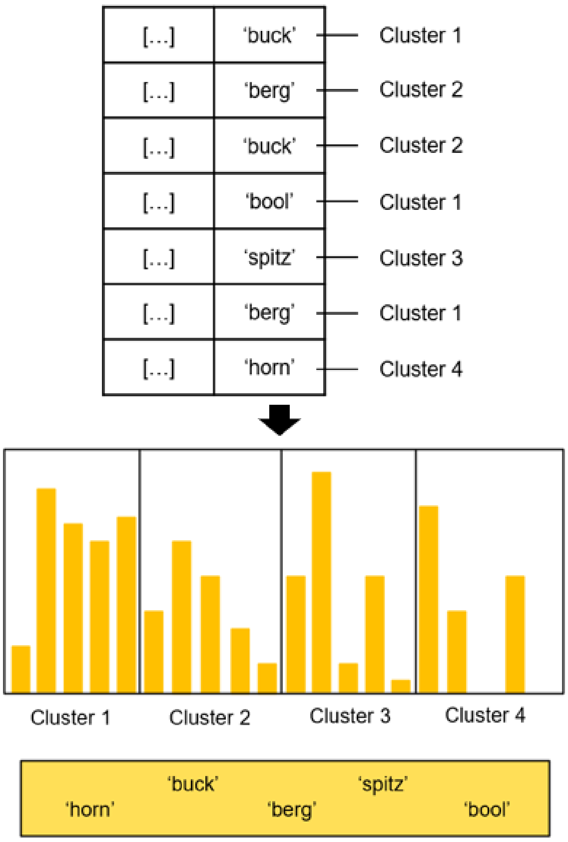


Figure 3.8: An illustration of K-Means clustering result. Note that the figures are only used for illustration, but do not represent anything about the results or my hypotheses.

Chapter 4 Results

In this chapter, I'm going to present the results that I got in the two phases of this research. In details, in 4.1 I will show the results of the extraction of generic parts, which served as the basis for the analysis in the second phase, which is the comparisons of generic parts that were told to convey similar semantic information using landscape (including topography and land cover), whose results would be presented in 4.2.

4.1 Extraction of Generic Parts

During the extraction process, a number of generic parts conveying semantic meanings related to landscape (excluding water areas and 'tiny' landforms) were identified, with the assistance of information from Ortsnamen.ch (Schweizerisches Idiotikon, 2023). Initially, as claimed in 3.3, I determined to collect only generic parts appearing in toponyms for at least 20 times for the analysis, following the suggestion of Qiu and Joe (2009) for clustering methods. However, considering that I ultimately determined the histograms to consist of 31 bins rather than 2 in section 3.4, theoretically I should have modified that threshold to 310 based on their suggests. Nevertheless, since the nature of this study was using the clustering of comprehensive bags of words to find the patterns of each generic part, while clustering for each bag of words only played an auxiliary role, and since 10 among these 31 variables referred to the distribution of elevations, 10 to the distribution of slopes, and 11 to the distribution of land cover types, I made a 'concession' by treating all of them as 3 variables: elevation, slope and land cover. Eventually, I retained all the generic parts that appear in toponyms for more than 30 times, as summarized in Appendix A. Additionally, their variants in plural forms or various dialects, diminutive forms, and their meanings according to Ortsnamen.ch (Schweizerisches Idiotikon, 2023) are shown there as well.

To sum up, 52 generic parts of German toponyms were found, with 21 of them have diminutive forms, and 9 generic parts of Italian toponyms were discovered. They could be classified into 2 main classes, denoting topography and land cover respectively. Moreover, they were further classified into 10 more detailed categories based on the semantic meanings, as outlined in Table 4.1. The number of generic parts found in German or Italian of each class, together with some examples of generic parts, are listed as well. Note that, 3 generic parts in German toponyms were found to convey semantic meanings related to both topography and land cover: 'Matt' was found to mean 'Flat grassy area', 'Wang' means 'Unforested, grassy, steeply sloping slopes', and 'Tobel' means 'Forested valley' (Schweizerisches Idiotikon, 2023), thus, the overall amount of generic parts in this table equals to 64, which is 3 more than the amount of generic parts I found.

Table 4.1 indicates that, within the main class 'topography', the amount of generic parts found to denote convex landforms was the largest, both in German and in Italian. Similarly, for those related to land cover, generic parts represent 'open areas' and 'forests' were the most abundant.

Thus, I selected categories ‘convex landforms’ and ‘open areas’ as case studies for the second phase of my study, in which I compared generic parts in those two categories respectively. The comparisons between ‘normal’ forms and diminutive forms, and cross-language comparisons, were also included here. Note that though ridges could be considered as ‘convex landforms’, it’s clear that they are distinct from mountains and hills in shape, hence generic parts denoting ridges were listed separately, and were not compared together with generic parts about other convex landforms.

Table 4.1: Categories of generic parts regarding their semantic meanings. Numbers in brackets are the amounts of diminutive forms of generic parts found in each category.

Category	In German	In Italian	Example(s)
Topography	42 (15)	7	-
• Convex landforms (Mountains/Hills)	16 (6)	3	‘Berg’ in German, ‘Monte’ in Italian
• Concave landforms	9 (3)	0	‘Tal’ in German
• Flats	6 (2)	1	‘Bode’ in German, ‘Piano’ in Italian
• Slopes	6 (2)	2	‘Halde’ in German, ‘Costa’ in Italian
• Saddles	2 (1)	1	‘Furgge’ in German, ‘Passo’ in Italian
• Ridges	2 (1)	0	‘First’ in German
• ‘Hills between other hills’	1 (0)	0	‘Chanzel’ in German
Land cover	13 (8)	2	-
• Forests	5 (4)	1	‘Wald’ in German, ‘Bosco’ in Italian
• Open areas (Grasslands/Meadows)	6 (3)	0	‘Riet’ in German
• Bushes	2 (1)	1	‘Loo’ in German, ‘Bosch’ in Italian

4.2 Comparison of Generic Parts using Landscape

Based on the results of the first phase, a total of 25 distinct bags of words were created for the generic parts denoting convex landforms, among which 16 bags were for 'normal' generic parts in German, 3 for generic parts in Italian (both include variants), and 6 for diminutive forms (in German). As described in 3.4, they were aggregated together, followed by the implementation of K-Means clustering. Subsequently, the distributions of each generic-part's bag of words were studied based on the clustering results. The clustering outcomes as well as the patterns of each generic part are presented in 4.2.1. Additionally, a total of 6 distinct bags of words were created for the generic parts denoting open areas, which are all in German. For this category, I did not conduct cross-language comparison, as generic part in Italian was not found, nor the comparisons of normal forms with diminutive forms, as it's hard to compare the 'size' of land cover via my implementations in this study. The clustering outcomes as well as the patterns of each generic part denoting land cover are presented in 4.2.2.

4.2.1 Comparison of Generic Parts Denoting Convex Landforms

Table 4.2 concludes all the generic parts (normal forms) denoting convex landforms with the amounts of toponyms containing each of them used for the analysis. As for their variants, their diminutive forms' variants, and their concrete meanings from 'Ortsnamen.ch' (Schweizerisches Idiotikon, 2023), please refer to Appendix A.

After implementing K-means clustering to the aggregated bag of words for convex landforms, it was partitioned into the following 6 clusters:

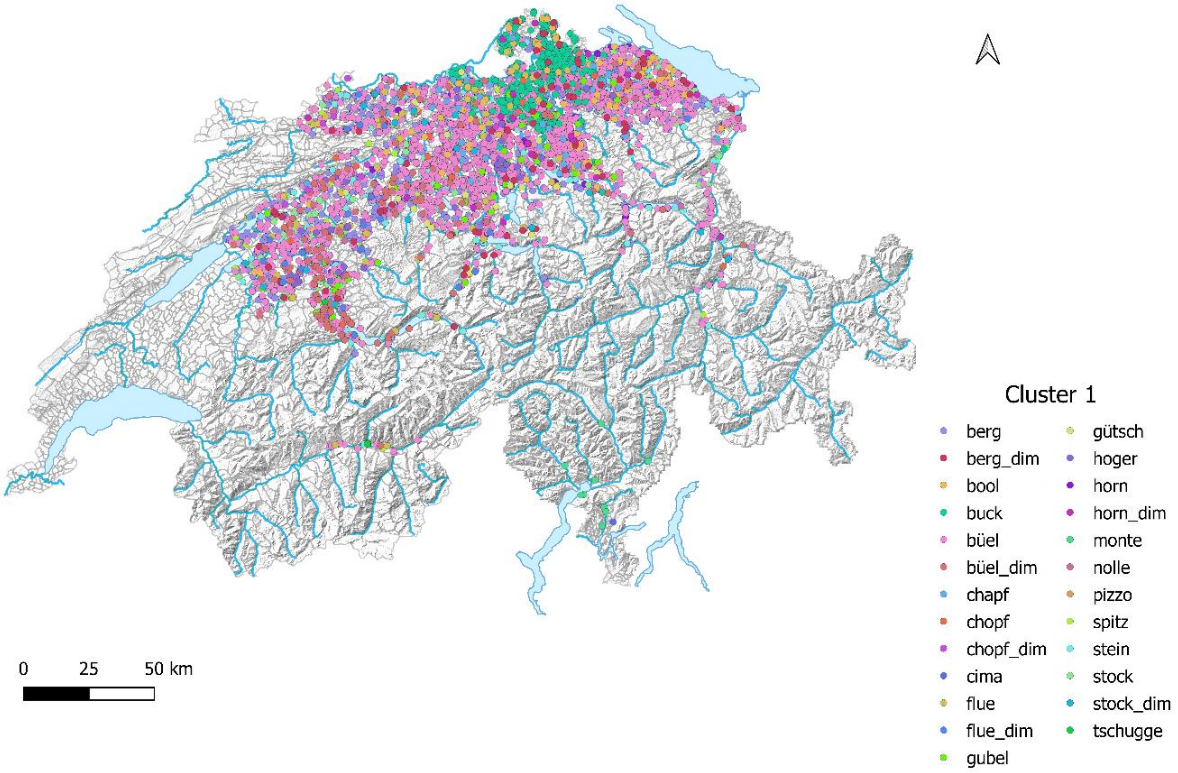
- Objects in cluster 1 are mostly located in Swiss Plateau and occasionally also in valleys, with elevations in their peripheral areas typically below 800m, and are characterized by gentle slopes. Additionally, on average, over 60% of their surrounding areas have been changed by people. Figure 4.1 shows the facts of this cluster, including the distributions of objects and the cluster center. Overall, this cluster contains 4911 records.
- Objects in cluster 2 are mostly located in Swiss Plateau and occasionally in valleys as well, with elevations in their peripheral areas typically below 800m, and characterized by more or less gentle slopes. The peripheral areas of them have not been changed a lot by people, with around 60% to be forests on average. Figure 4.2 shows the conditions of this cluster, including the distributions of objects and the cluster center. Overall, this cluster contains 3557 records.
- Objects in cluster 3 are mostly located in 'Voralpen' ('Pre-Alps') (including 'Luganer Voralpen' or 'Lugano Prealps') or in Jura areas, with elevations in the peripheral areas typically between 800m and 1200m, and with slopes that are not so gentle. Grasslands and forests are on average the main land cover types in the surrounding areas, followed by pastures and human-related features. Figure 4.3 shows the conditions of this cluster, including the distributions of objects and the cluster center as well. Overall, this cluster contains 2255 records.
- Objects in cluster 4 are mostly located in 'Voralpen' ('Pre-Alps') (including 'Luganer Voralpen' or 'Lugano Prealps') areas, with elevations in the peripheral areas typically

between 1200m and 1600m, sometimes between 800m and 1200m or between 1600m and 2000m, and with slopes that are a bit steep. The most prevalent land cover type on average in the peripheral areas is forests, followed by pastures, grasslands and dissolved forests. Figure 4.4 shows the facts of it, including the distributions of objects and the cluster center. Overall, this cluster contains 2109 records.

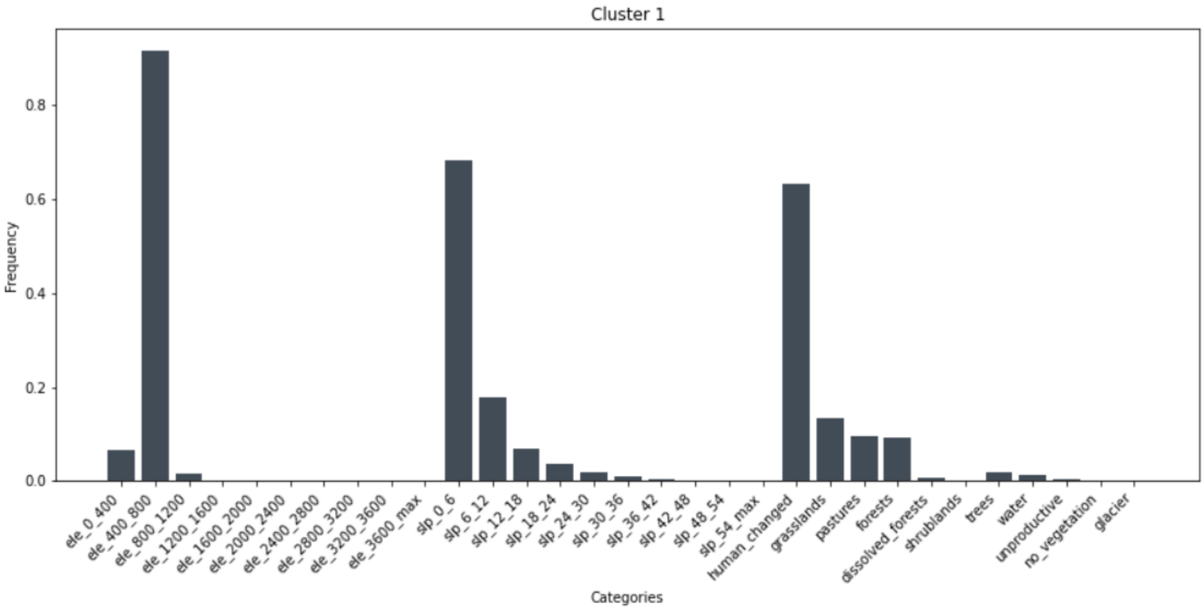
- Objects in cluster 5 are mostly located in the major region of Alps, but usually near to the 'Voralpen' areas, with elevations in the peripheral areas typically between 1600m and 2000m, sometimes also between 2000m and 2400m or between 1200m and 1600m, and with steep slopes. The most prevalent land cover type on average in the peripheral areas is pastures, then unproductive vegetations, and occasionally no vegetation exists there. Figure 4.5 shows the conditions of it, including the distributions of objects and the cluster center. Overall, this cluster contains 2156 records.

Table 4.2: Generic parts denoting convex landforms. 'GER' is the abbreviation of 'German' here, while 'ITA' is the abbreviation of 'Italian' here.

Generic Part	Amount	Language	Generic Part	Amount	Language
'Berg'	4160	GER	'Hoger'	84	GER
'Berg' (dim)	379	GER	'Horn'	916	GER
'Bool'	124	GER	'Horn' (dim)	103	GER
'Buck'	549	GER	'Nolle'	92	GER
'Büel'	4341	GER	'Spitz'	387	GER
'Büel' (dim)	274	GER	'Stein'	895	GER
'Chapf'	129	GER	'Stock'	898	GER
'Chopf'	524	GER	'Stock' (dim)	141	GER
'Chopf' (dim)	66	GER	'Tschugge'	118	GER
'Flue'	1081	GER	'Cima'	126	ITA
'Flue' (dim)	103	GER	'Monte'	375	ITA
'Gubel'	108	GER	'Pizzo'	254	ITA
'Gütsch'	115	GER	-	-	-



a)

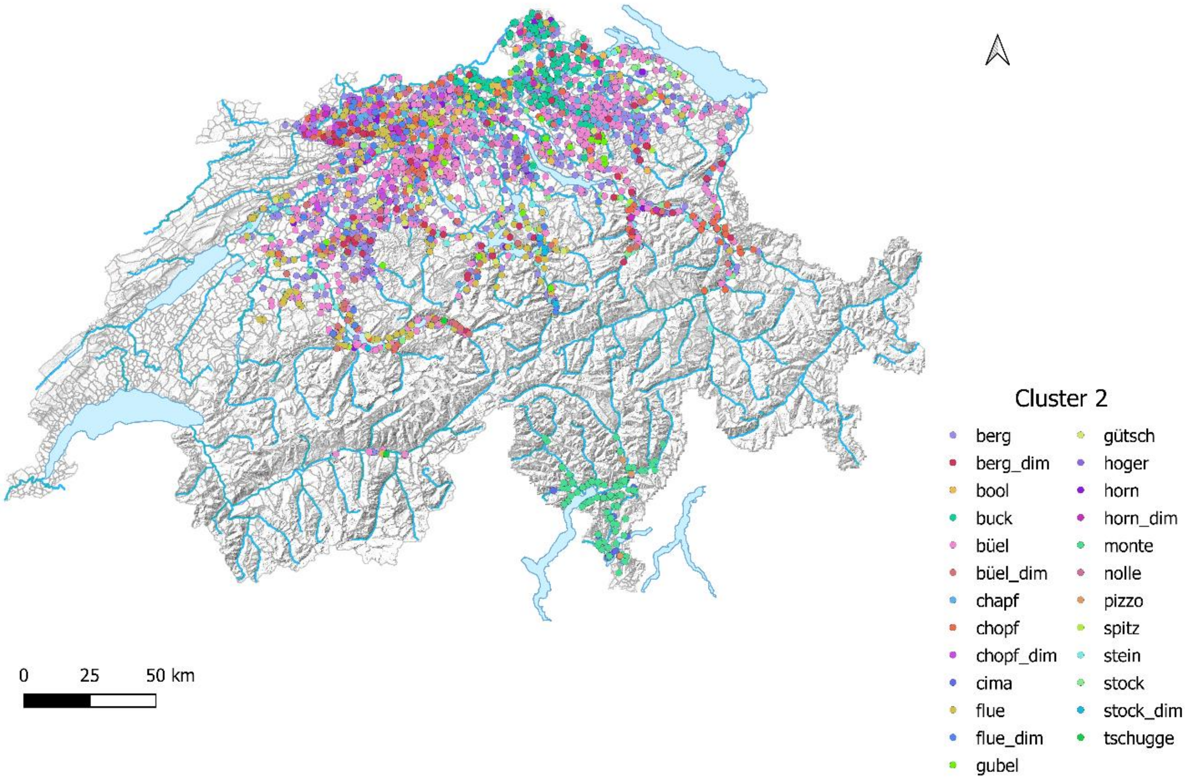


b)

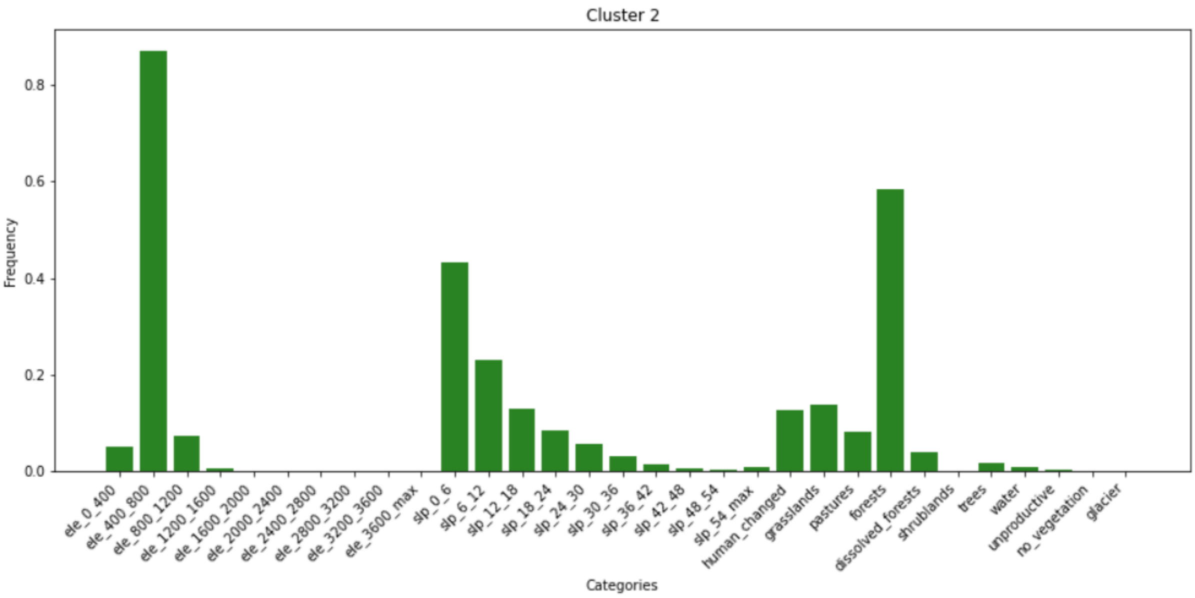
Figure 4.1: a) Distribution of the objects whose histograms were partitioned into cluster 1 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 1.

- Objects in cluster 6 are mostly located in the major region of Alps, around the main ridge, with elevations in the peripheral areas usually greater than 2000m, and with very steep slopes. Mostly the peripheral areas of them are not covered by any vegetation.

Figure 4.6 shows the facts of it, including the distributions of objects and the cluster center. Overall, this cluster contains 1354 records.

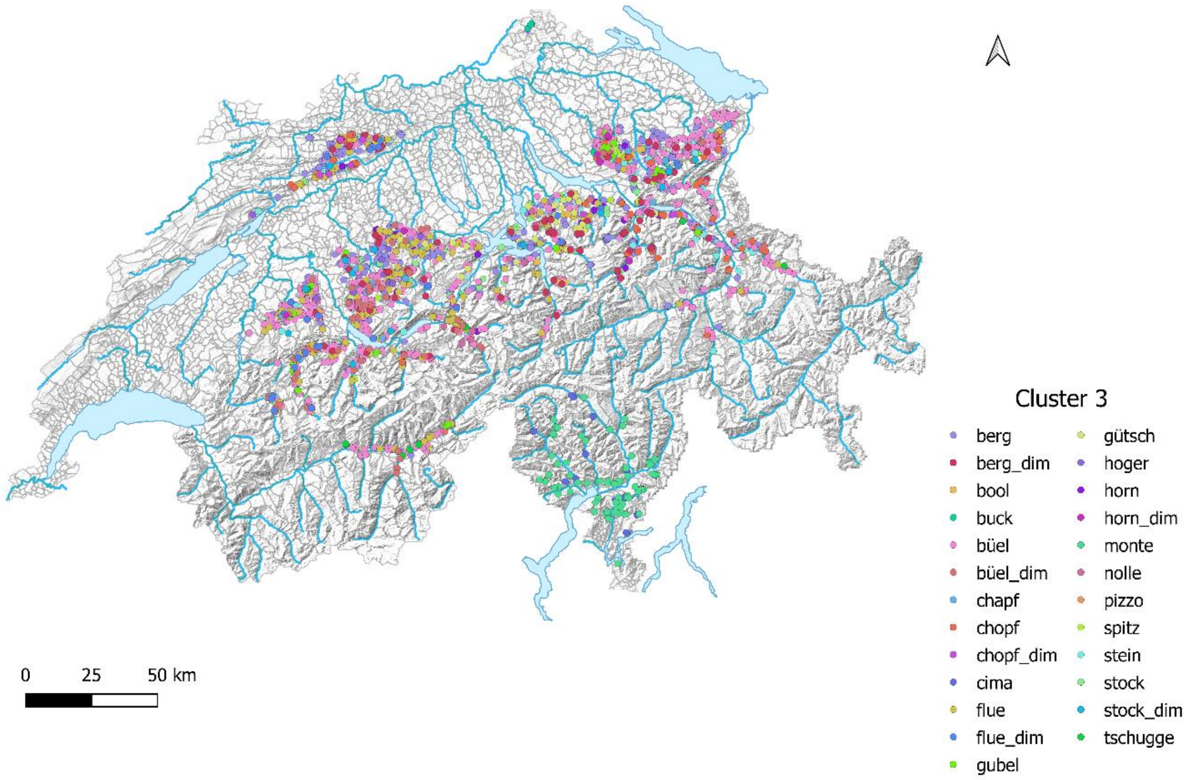


a)

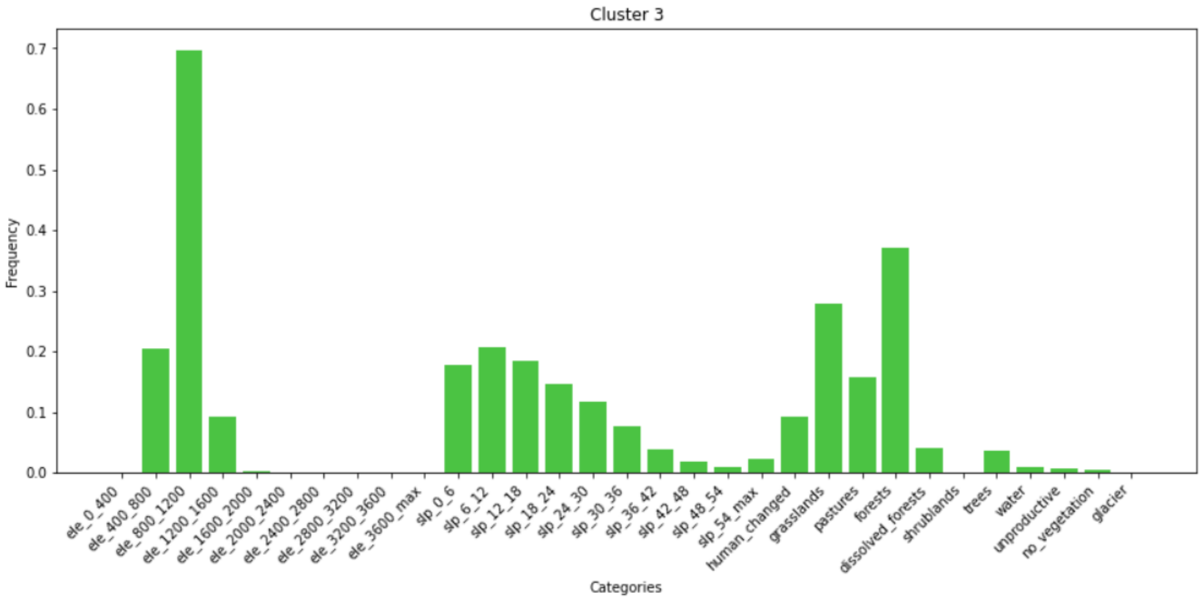


b)

Figure 4.2: a) Distribution of the objects whose histograms were partitioned into cluster 2 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 2.



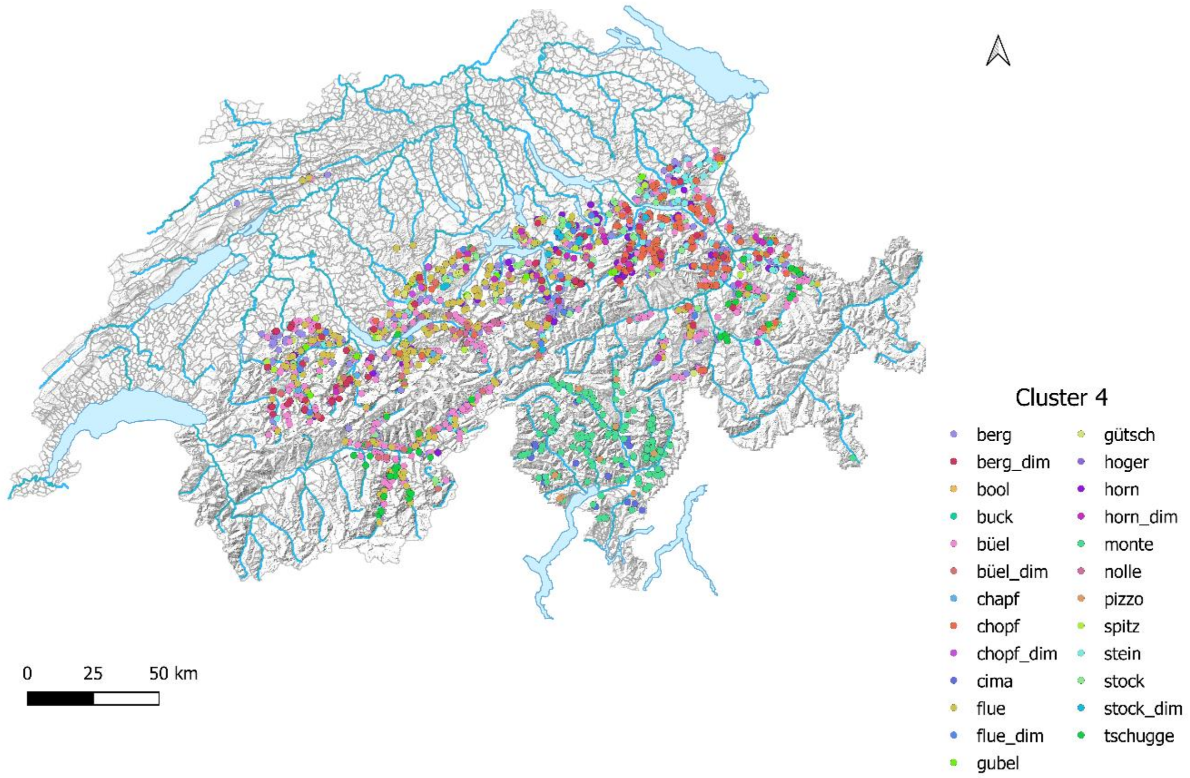
a)



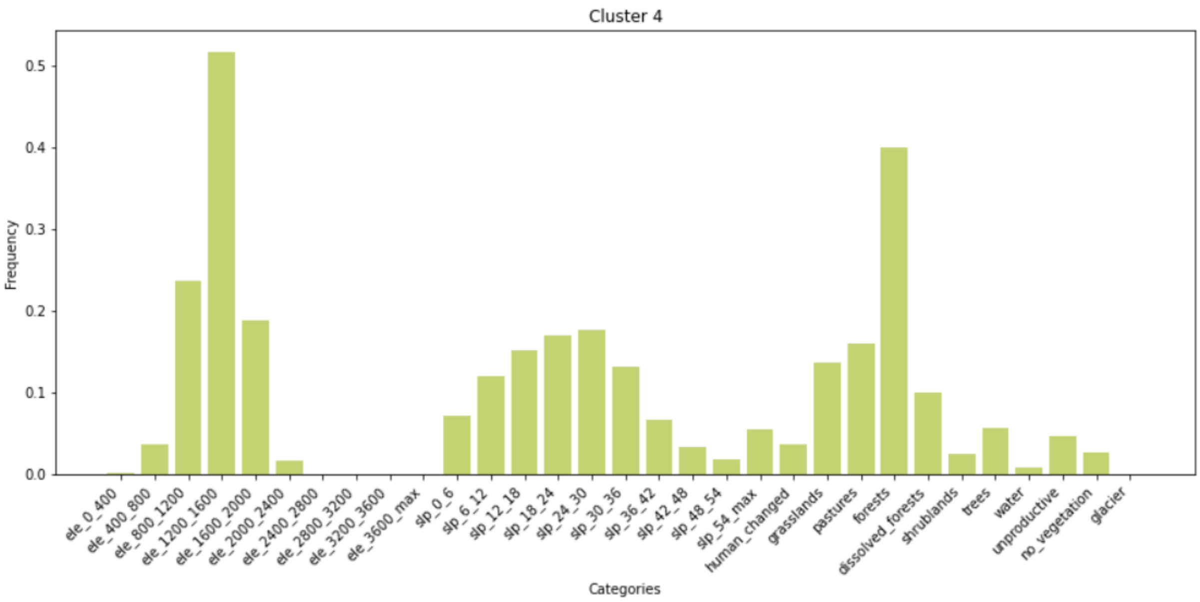
b)

Figure 4.3: a) Distribution of the objects whose histograms were partitioned into cluster 3 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 3.

With these clustering results, I investigated the patterns of those generic parts denoting convex landforms, by finding the distributions of the local descriptors within each of those 25 bags of words. In the following paragraphs, I will discuss the patterns of these generic parts. Meanwhile, for pairs or sets of generic parts that show similar patterns, I also conducted cosine similarity analysis, which are also included here. Note that, in the visualizations illustrating patterns, these clusters are marked by 'cluster 0' to 'cluster 5', where the 'cluster 0' there corresponds to cluster 1 mentioned above, 'cluster 1' means cluster 2 mentioned above, and so on and so forth.

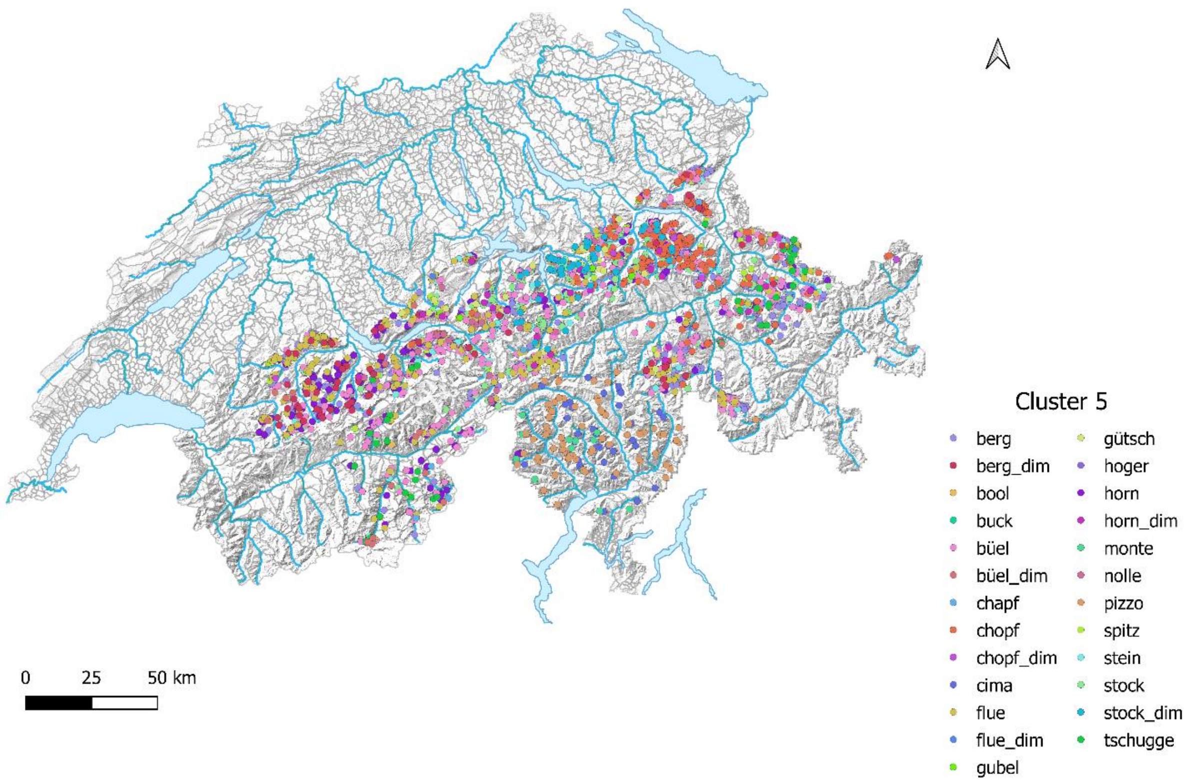


a)

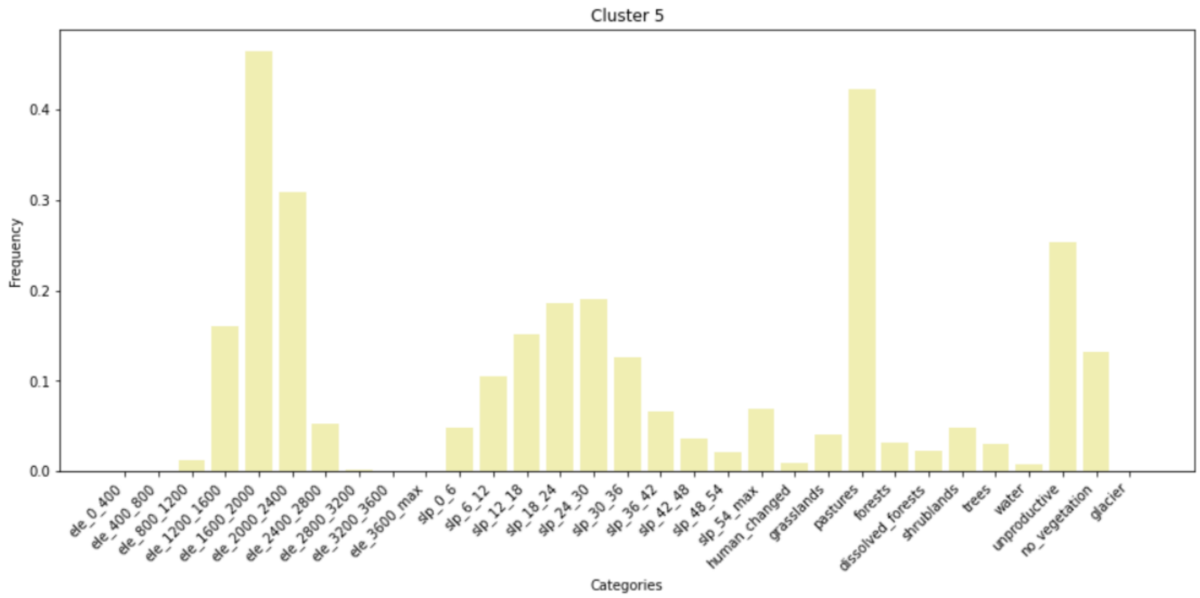


b)

Figure 4.4: a) Distribution of the objects whose histograms were partitioned into cluster 4 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 4.



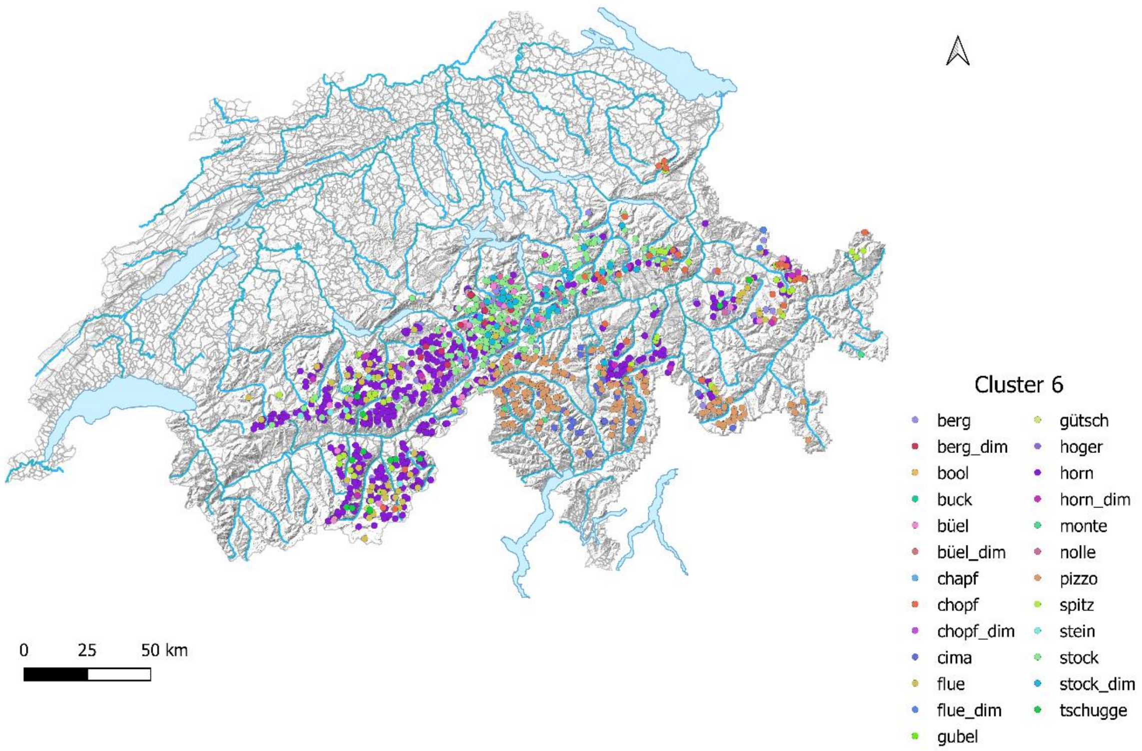
a)



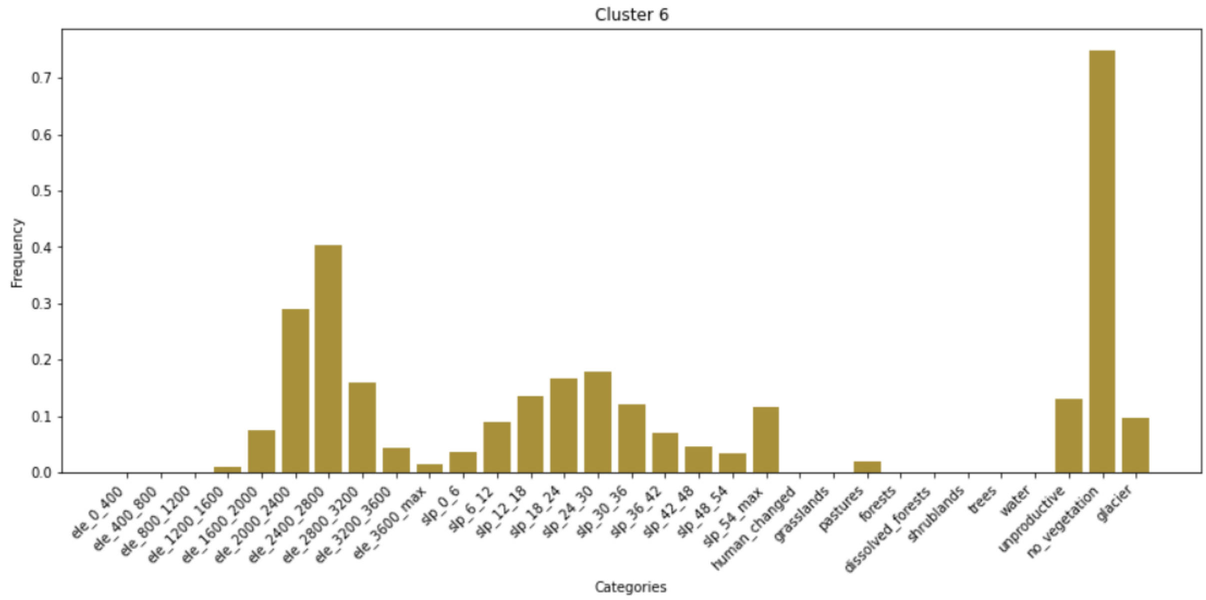
b)

Figure 4.5: a) Distribution of the objects whose histograms were partitioned into cluster 5 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 5.

First of all, I compared the ‘normal’ generic parts, both in German and in Italian, including their variants. After that, the comparisons of each ‘normal’ one with the diminutive forms were covered.



a)



b)

Figure 4.6: a) Distribution of the objects whose histograms were partitioned into cluster 6 with different colors representing different generic parts of those objects' names. b) The histogram of cluster center for cluster 6.

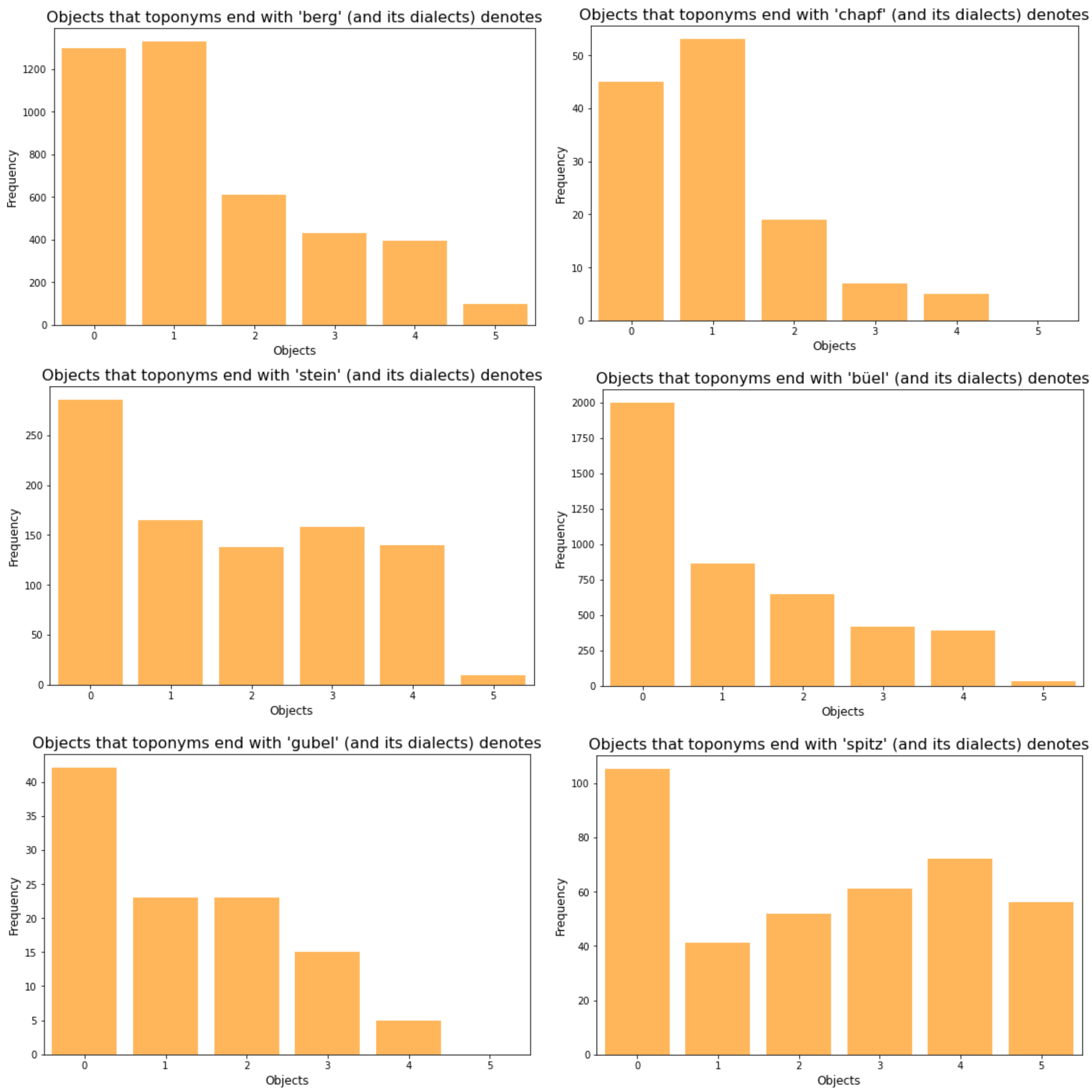
Among those 19 'normal' generic parts denoting convex landforms, I observed that 'Berg' and 'Chapf' share nearly identical patterns, which, roughly speaking, are also similar to patterns of 'Stein', 'Büel', 'Gubel' and 'Spitz', as shown in Figure 4.7 a). Specifically, 'Berg' and 'Chapf' frequently denote convex landforms assigned to clusters 1 or 2, characterized by relatively flat terrain with gentle slopes, while occasionally they represent objects assigned to other clusters as well, where the altitude is higher, and the terrain is more undulated. As for 'Stein' 'Büel' and 'Gubel', they could less frequently denote objects partitioned into cluster 2, compared to 'Berg' and 'Chapf'. 'Spitz' is, among these 6 generic parts, the most special one. The pattern of it has another 'peak' at cluster 5, means that it could also denote Pre-alps and Alps areas frequently.

Given the similarity in patterns of these generic parts, and the fact that the clustering used for each of them indicated their grouping into 2 similar clusters as outlined in Appendix B, cosine similarities were calculated for further comparisons. Nevertheless, only the self-comparison of 'Chapf' showed a median cosine value of around 0.7, showing autocorrelation of it, suggesting concentrated objects 'Chapf' often denote, while for other self- and cross-group comparisons, the median cosine values were relatively low, as shown in Figure 4.7 b). These results are not really counter-intuitive, as from their patterns it had already been clear that, the other 5 generic parts' patterns are somehow 'scattered'. And note that, the results could not suggest that these generic parts fail to represent similar objects, as for those generic parts with scattered patterns, even self-comparisons could not yield high median cosine values, let alone comparing them to other generic parts.

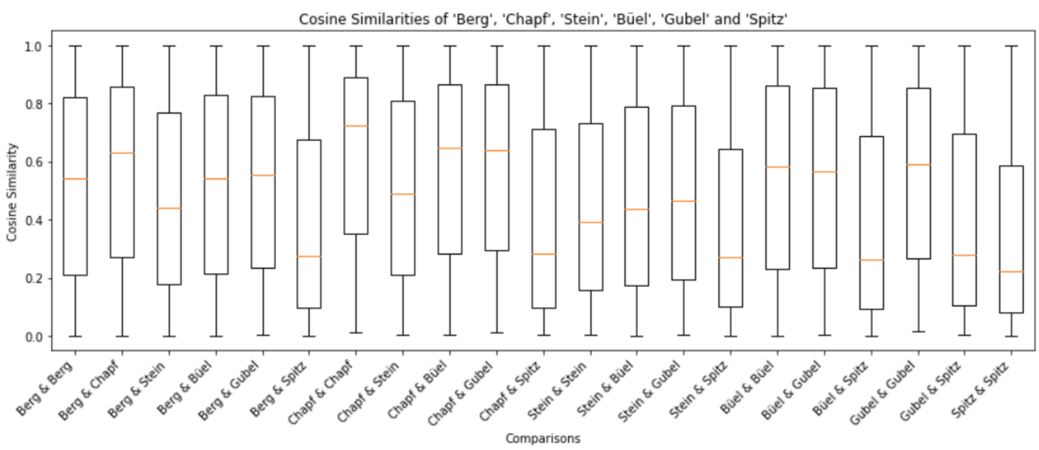
Besides, I observed that 'Bool', 'Buck' and 'Hoger' share nearly identical patterns, as shown in Figure 4.8 a). Specifically, they almost always denote convex landforms on Swiss Plateau, with flat terrain and gentle slopes. Given the similarity in their patterns, and the fact that the clustering conducted for each of them got similar results as summarized in Appendix B, cosine similarities were calculated for further comparisons. All the self- and cross-group comparisons here got the median cosine values to be greater than 0.8, suggesting similarities not only within these generic parts themselves, but also among them, about the landscape that they denote, as shown in Figure 4.8 b).

Additionally, I also found that 'Horn' in German, and 'Cima' and 'Pizzo' in Italian show nearly identical patterns, as Figure 4.9 a) illustrates. Specifically, they almost always denote convex landforms in the major Alps region, with high altitudes and undulated slopes, and covered with unproductive vegetation or without any vegetation. Occasionally, 'Cima' could denote objects in 'Voralpen' areas as well. Given the similarity in their patterns, I calculated cosine similarities for further comparisons as well. Only the self-comparison of 'Pizzo' showed a median cosine value of greater than 0.7, showing autocorrelation of it, suggesting less scattered objects 'Pizzo' denote, while for other self- and cross-group comparisons, the median cosine values were often not greater than 0.5, as shown in Figure 4.9 b).

Then, 'Chopf' and 'Nolle' were found to show nearly identical patterns, according to Figure 4.10 a). Specifically, they could denote all kinds of convex landforms in Switzerland, with the peak of their patterns at cluster 4-5, means they denote Pre-alps areas and those Alps areas that are close to Pre-alps regions more frequently. Given the similarity in their patterns, I calculated cosine similarities for further comparisons as well. However, the results shown in Figure 4.10 b) indicates that, all the comparisons got low median cosine values. Actually, they should be expected to get such results, as they were found to denote various kinds of objects, leading to low median cosine values while calculating cosine similarities.

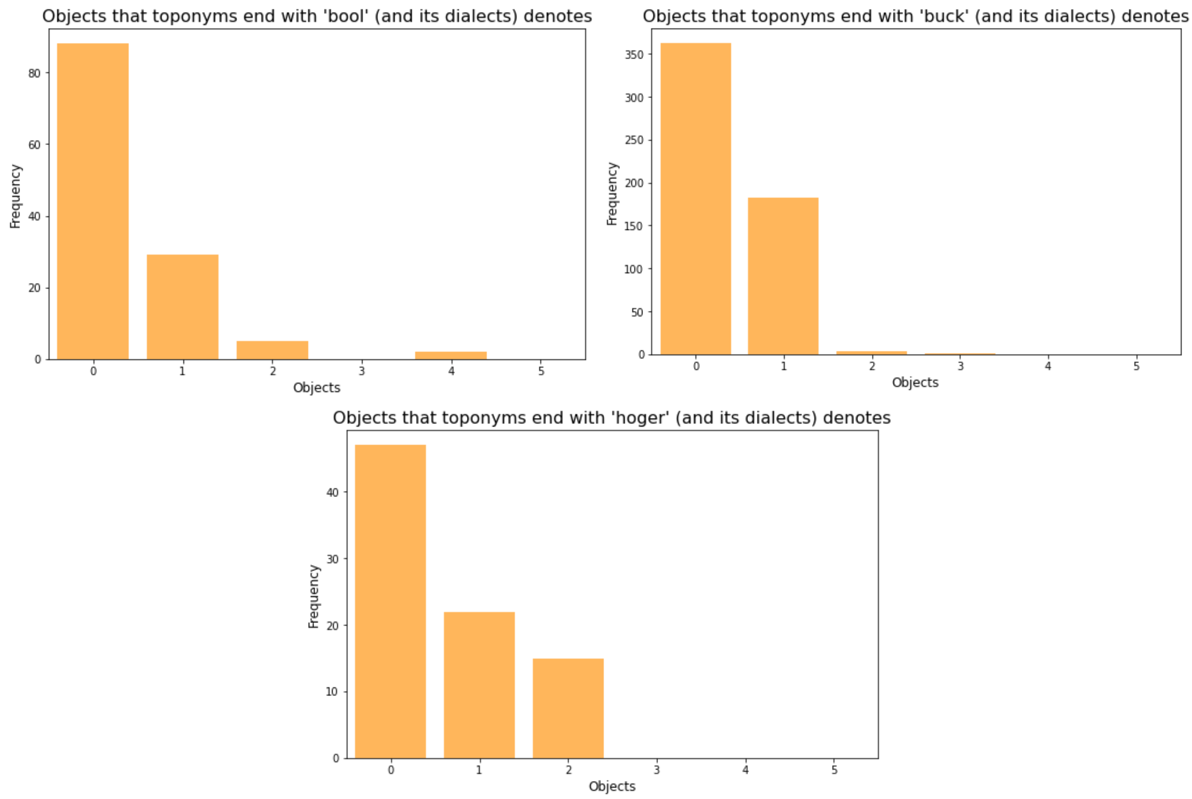


a)

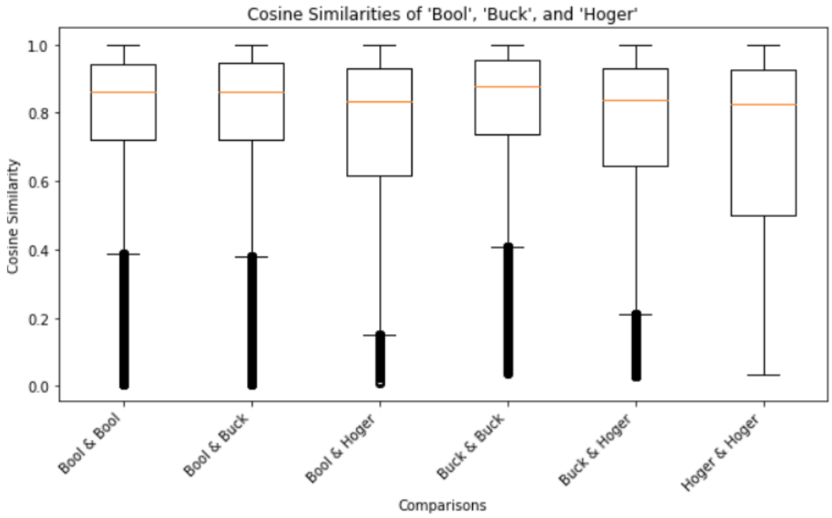


b)

Figure 4.7: a) Patterns of ‘Berg’, ‘Chapf’, ‘Stein’, ‘Büel’, ‘Gubel’ and ‘Spitz’. b) Cosine similarities of ‘Berg’, ‘Chapf’, ‘Stein’, ‘Büel’, ‘Gubel’ and ‘Spitz’.



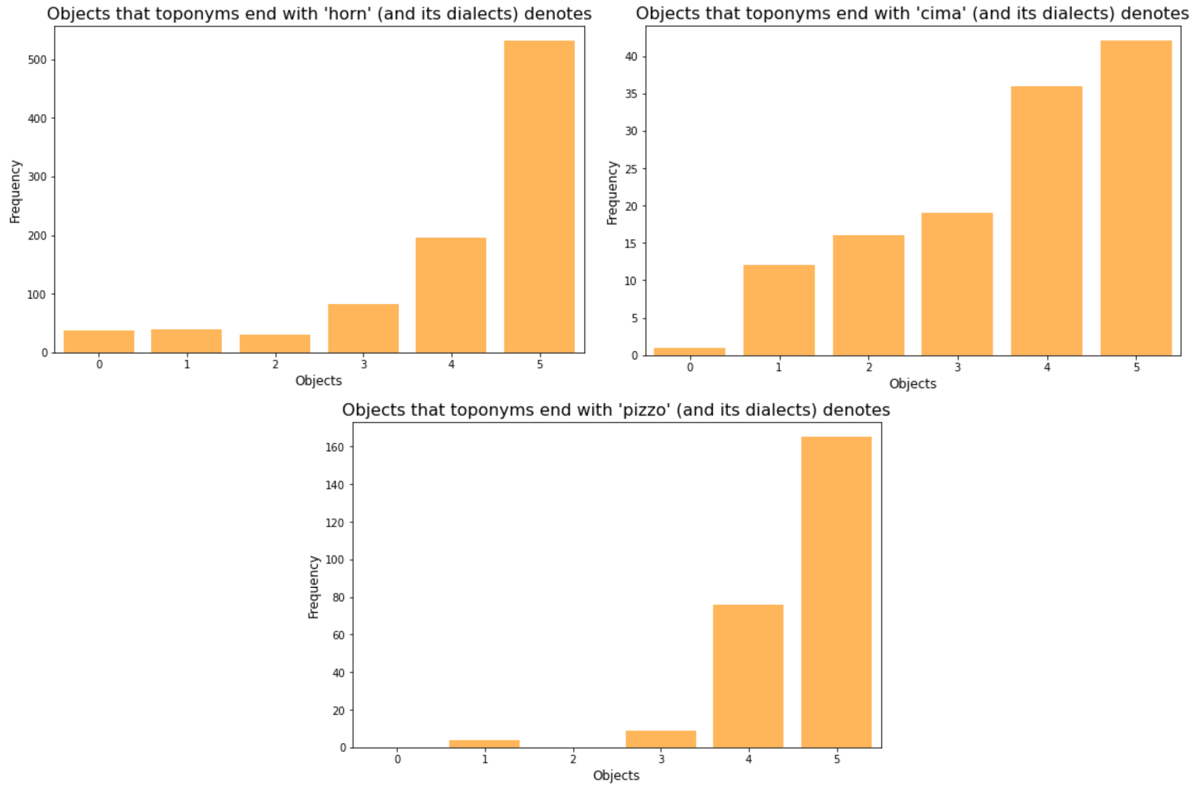
a)



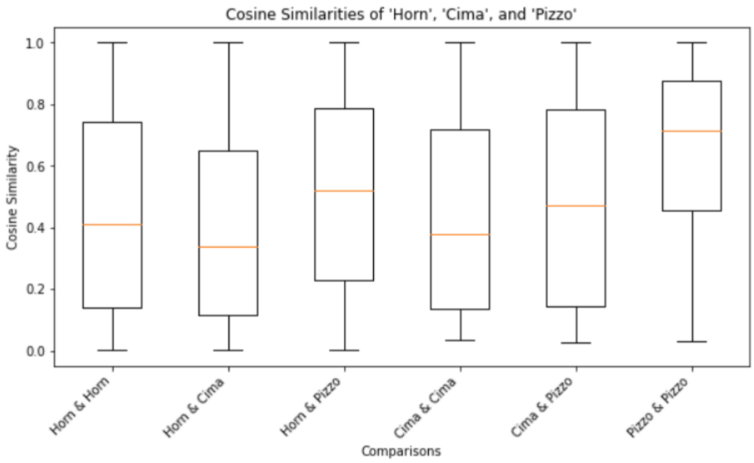
b)

Figure 4.8: a) Patterns of 'Bool', 'Buck', and 'Hoger'. b) Cosine similarities of 'Bool', 'Buck', and 'Hoger'.

Finally, 'Flue' and 'Gütsch' were found to show similar patterns, according to Figure 4.11 a). They could almost denote all kinds of convex landforms in Switzerland as well, and without a clear 'peak' regarding their patterns. Given this similarity, I also calculated cosine similarities for further comparisons, though I did not expect to find high median cosine values through this, as they had already been found to denote various landforms. Just like I expected, the results in Figure 4.11 b) here shows that all the comparisons got low median cosine values.



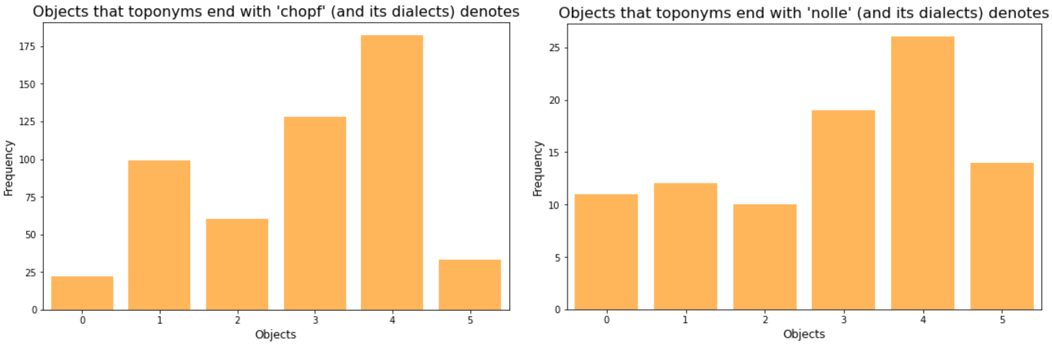
a)



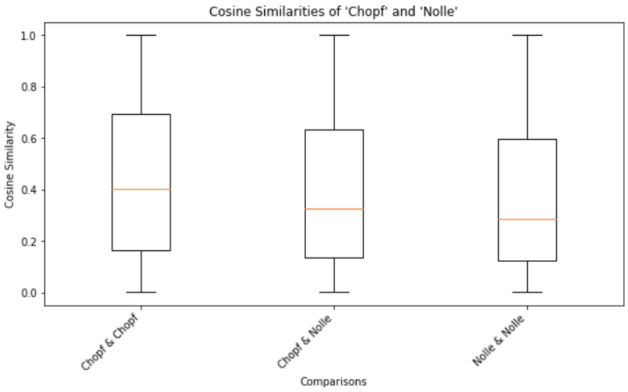
b)

Figure 4.9: a) Patterns of 'Horn', 'Cima', and 'Pizzo'. b) Cosine similarities of 'Horn', 'Cima', and 'Pizzo'.

Among these 19 generic parts, 'Monte', 'Stock' and 'Tschugge' were found to be three distinct cases, as no other generic parts show similar patterns to each of them, according to Figure 4.12. Specifically, 'Monte' usually represent convex landforms in valleys and in 'Voralpen' areas in Italian-speaking areas. 'Stock' on the other hand, could denote all kinds of landforms and has 2 'peaks' regarding its pattern, one of which is at cluster 1 (objects in Swiss Plateau with flat terrain and used by human a lot), the other at cluster 6 (objects around the main ridge of the Alps). Finally, 'Tschugge' also shows a 'peak' around cluster 4 and 5 but differs from 'Chopf' and 'Nolle' as it rarely denotes convex landforms in other clusters.

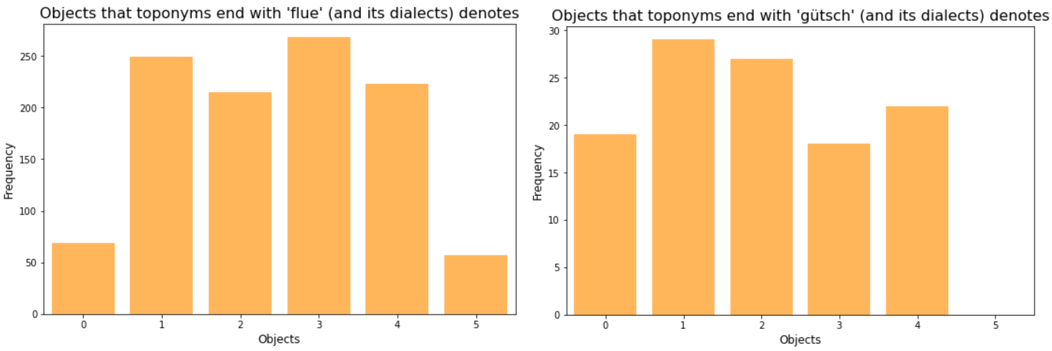


a)

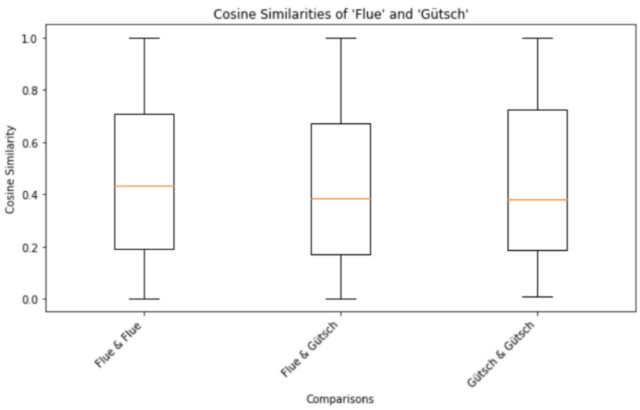


b)

Figure 4.10: a) Patterns of ‘Chopf’, and ‘Nolle’. b) Cosine similarities of ‘Chopf’, and ‘Nolle’.



a)



b)

Figure 4.11: a) Patterns of ‘Flue’ and ‘Gütsch’. b) Cosine similarities of ‘Flue’ and ‘Gütsch’.

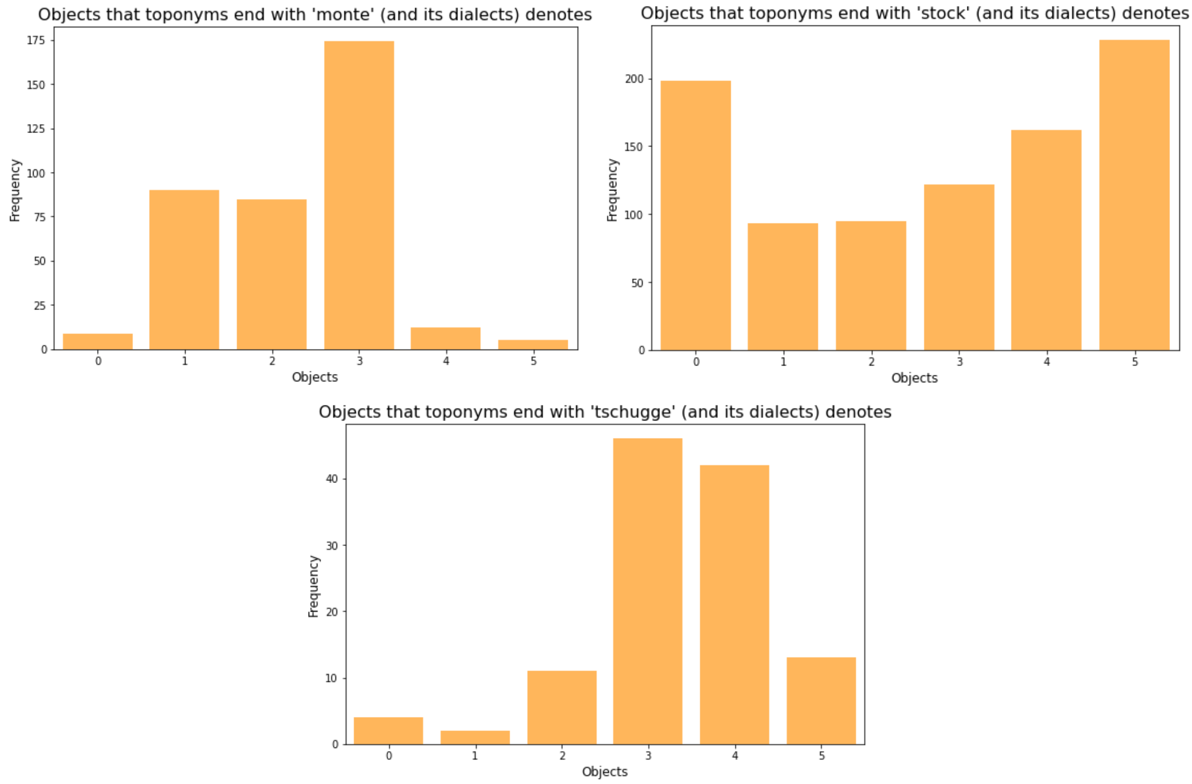


Figure 4.12: Patterns of ‘Monte’, ‘Stock’ and ‘Tschugge’.

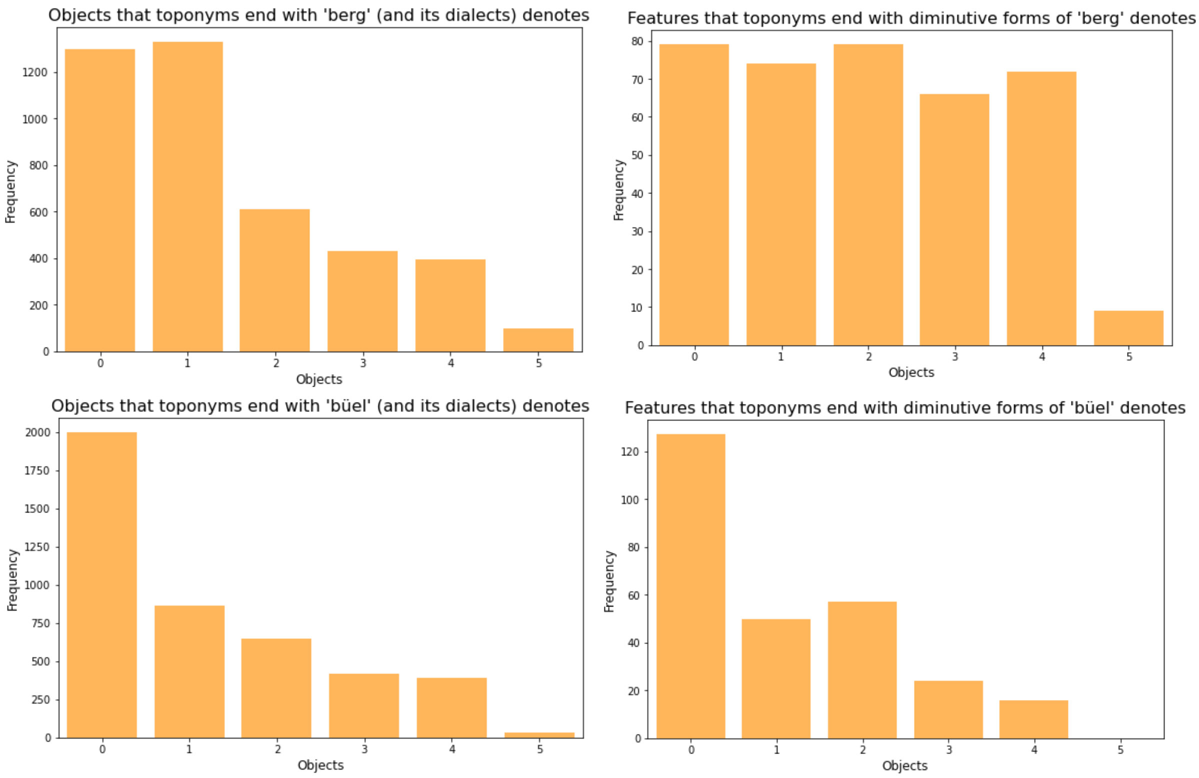


Figure 4.13: Comparisons of ‘Berg’, ‘Büel’ with their diminutive forms.

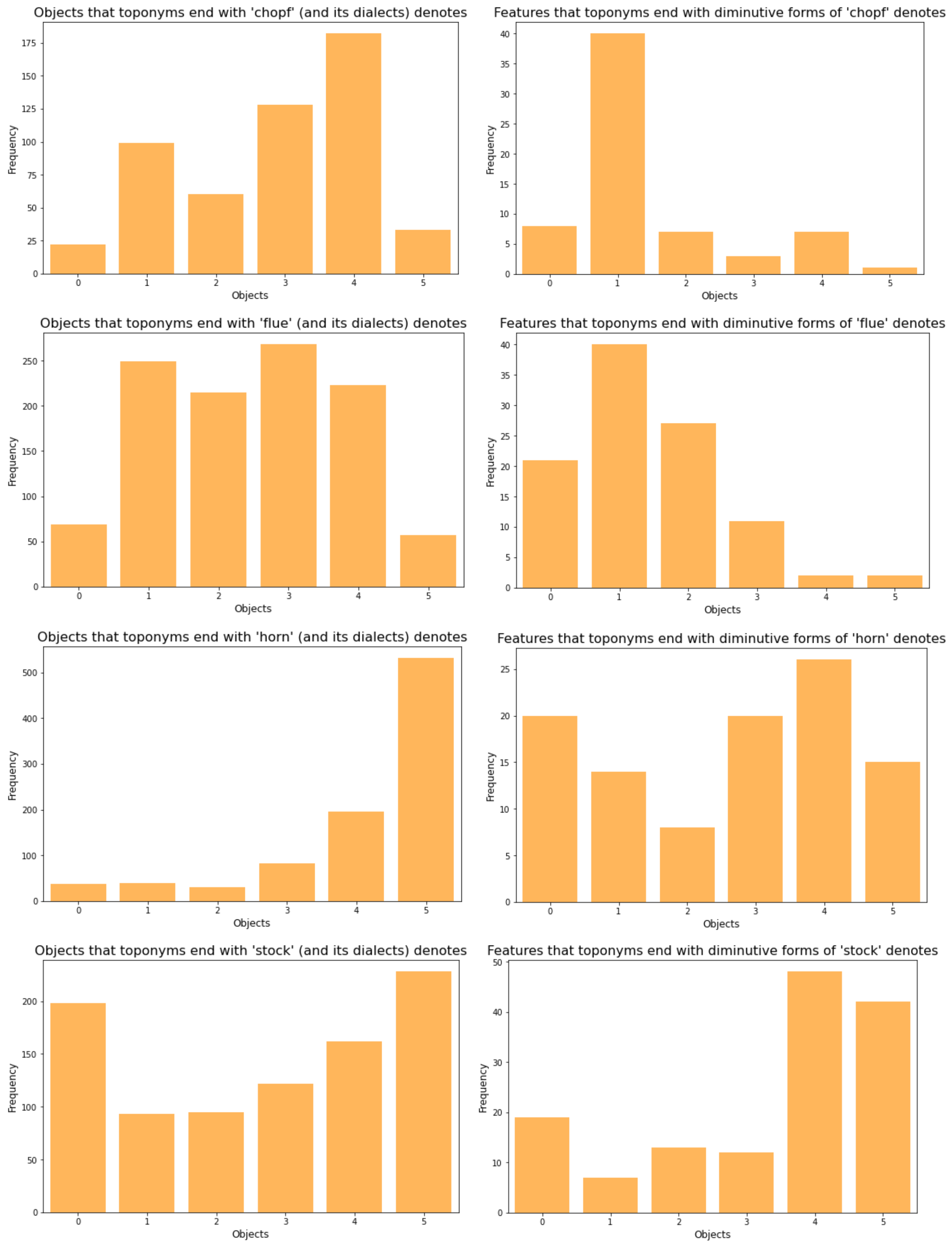


Figure 4.14: Comparisons of ‘Chopf’, ‘Flue’ , ‘Horn’, ‘Stock’ with their diminutive forms.

The differences of some ‘normal’ generic parts with their diminutive forms were also studied. For ‘Berg’, ‘Büel’, however, their diminutive forms do not really denote convex landforms that are flatter with lower altitudes, according to Figure 4.13. The diminutive forms of ‘Berg’ even seems like denoting higher and steeper convex landforms, which is somehow counter-intuitive.

As for ‘Chopf’, ‘Flue’ and ‘Horn’, their diminutive forms were found to denote landforms that are relatively flatter with lower altitudes, compared to their ‘normal’ ones, as Figure 4.14 shows, while for the diminutive forms of ‘Stock’, the ‘peak’ at cluster 1 for the pattern of ‘Stock’ seems disappeared here, where the ‘peak’ at cluster 6 seems moved a bit to the left hand side, to cluster 5, hence it’s complicated to compare them regarding their landscape.

In 5.1, I will come back to these results to further discuss them combining their meanings from different sources, compare with other researchers’ works, and answer the research questions.

4.2.2 Comparison of Generic Parts Denoting ‘Open Areas’

Table 4.3 concludes all the generic parts (normal forms) denoting open areas with the amounts of toponyms containing each of them used for the analysis. Their variants, diminutive forms, and concrete meanings from ‘Ortsnamen.ch’ (Schweizerisches Idiotikon, 2023), please refer to Appendix A.

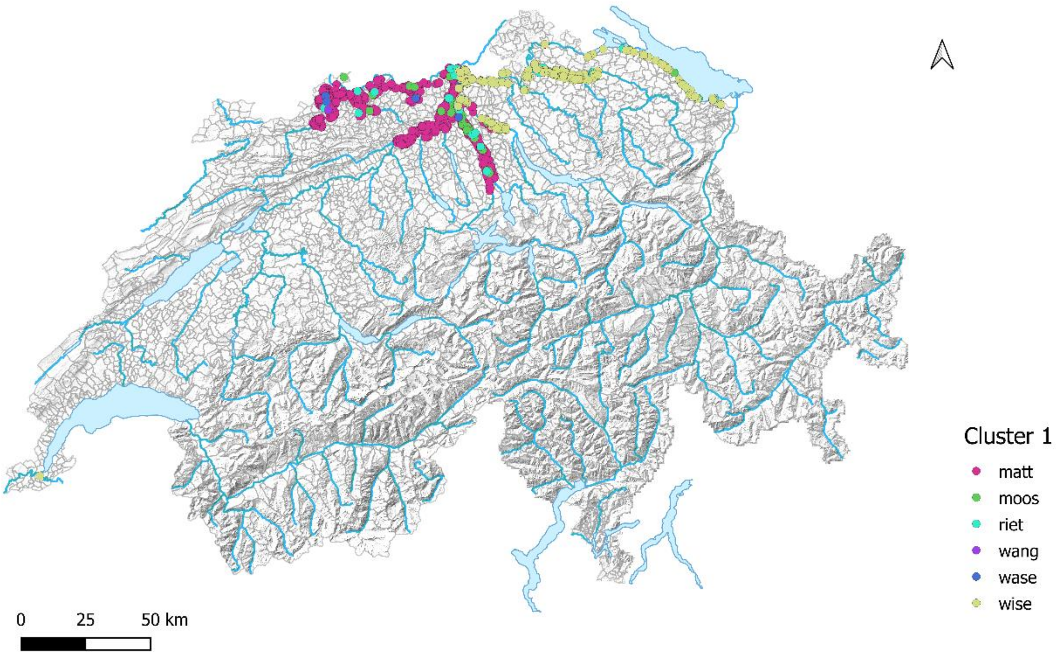
Table 4.3: Generic parts denoting open areas.

Generic Part	Amount	Language	Generic Part	Amount	Language
‘Matt’	10026	GER	‘Wang’	670	GER
‘Moos’	2190	GER	‘Wase’	67	GER
‘Riet’	1544	GER	‘Wise’	4690	GER

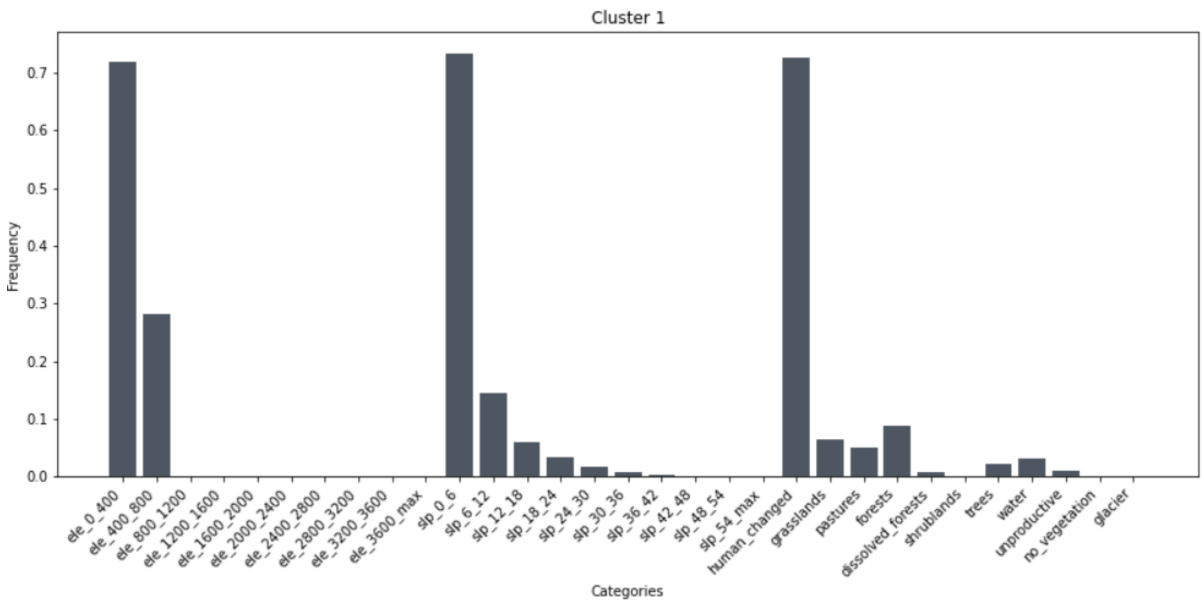
After implementing K-means clustering to the aggregated bag of words for open areas, it was partitioned into the following 5 clusters, which is a bit different from convex landforms:

- Objects in cluster 1 are mostly located in valleys near rivers and lakes in Swiss Plateau, with elevations in their peripheral areas typically below 400m, and are characterized by very gentle slopes. On average, over 70% of their surrounding areas have been changed by people. No cluster like this exist for ‘convex landforms’, since they should inherently not be found in valleys, almost only where the altitudes could be below 400m in Swiss Plateau. Figure 4.15 shows the conditions of this cluster, including the distributions of objects and the cluster center. Overall, this cluster contains 1057 records.
- Cluster 2 is like cluster 1 of convex landforms. Objects here are located in Swiss Plateau and occasionally in valleys, with elevations in their peripheral areas typically between 400m and 800m and are characterized by gentle slopes. On average, over 70% of their surrounding areas have been changed by people. Figure 4.16 shows the conditions of this cluster, including the distributions of objects and the cluster center. Overall, this cluster contains 8758 records.
- Objects in cluster 3 are mostly located in Swiss Plateau and also occasionally in valleys, with elevations in their peripheral areas mostly between 400m and 800m, characterized

by more or less gentle slopes. The peripheral areas of them could be covered by human-built things, grasslands, forests and pastures. Figure 4.17 shows the conditions of this cluster, including the distributions of objects and the cluster center. Overall, this cluster contains 5691 records.



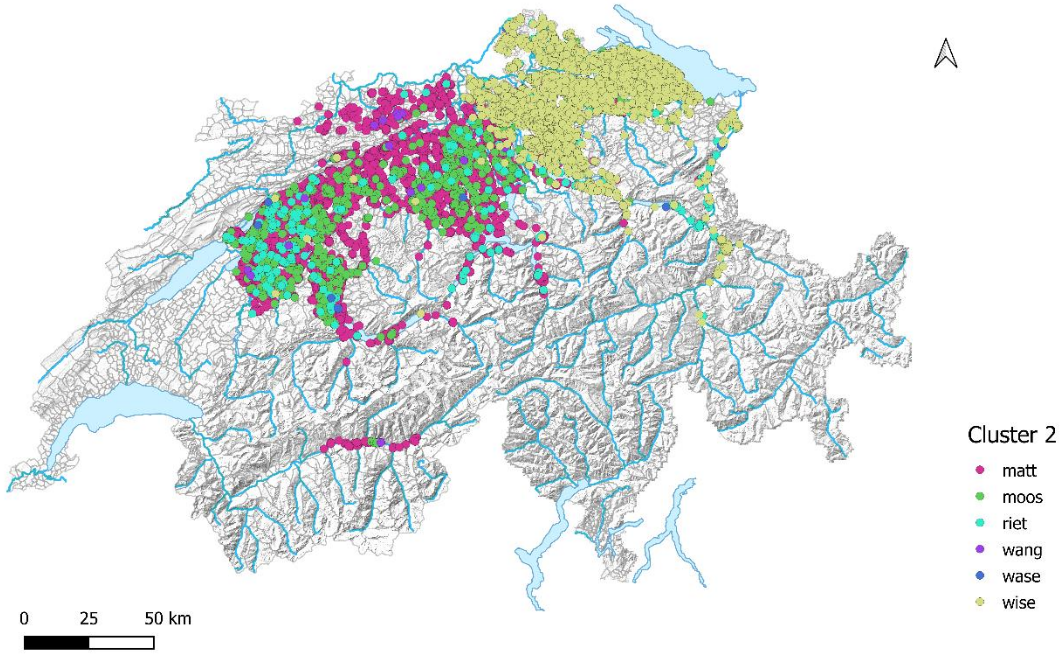
a)



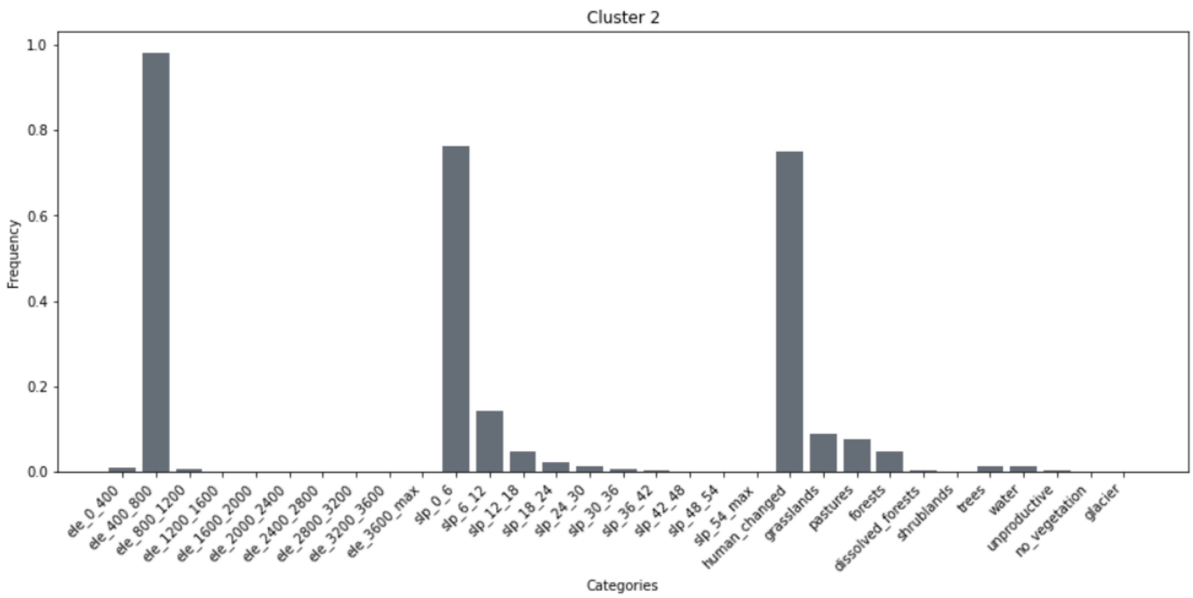
b)

Figure 4.15: a) Distribution of objects whose histograms were partitioned into cluster 1 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 1.

- Cluster 4 is similar to cluster 3 of convex landforms. Objects here are mostly located in 'Voralpen' ('Pre-Alps') or in Jura areas, with elevations in the peripheral areas typically between 800m and 1200m, and with slopes that are not so gentle. Grasslands and forests are on average the main land cover types in the surrounding areas, followed by pastures and human-related features. Figure 4.18 shows the conditions of this cluster, including the distributions of objects and the cluster center as well. Overall, this cluster contains 1915 records.

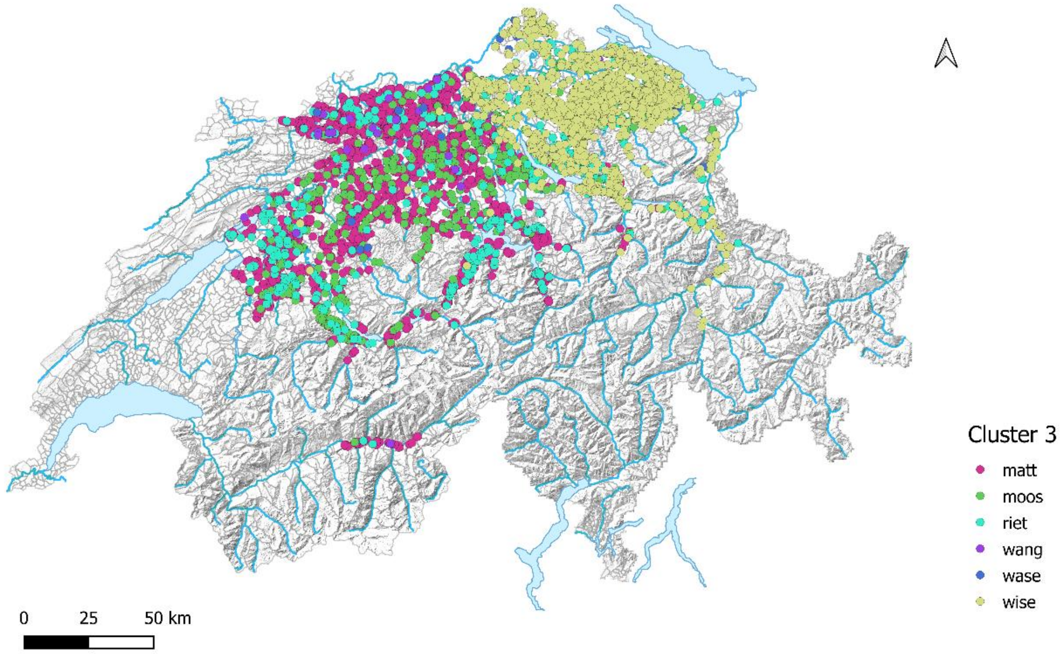


a)

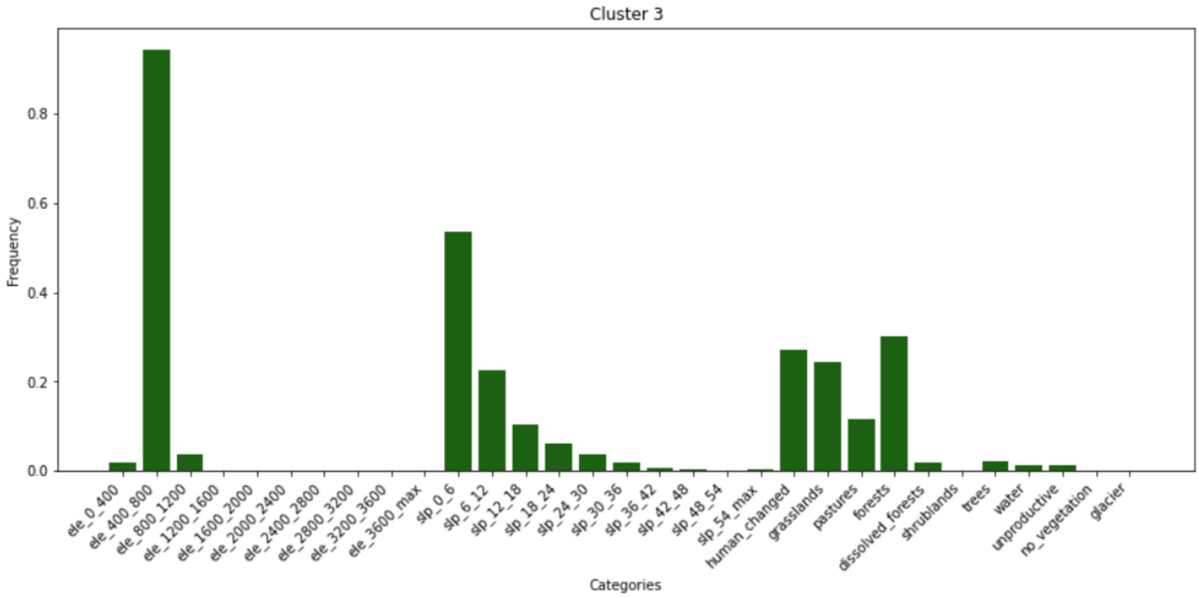


b)

Figure 4.16: a) Distribution of objects whose histograms were partitioned into cluster 2 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 2.



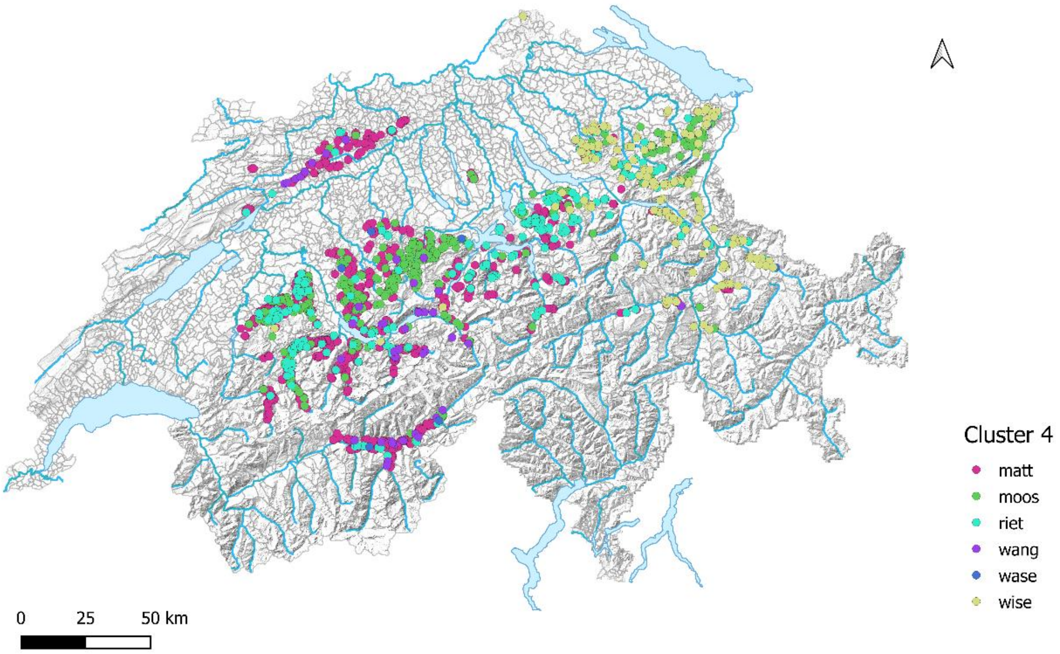
a)



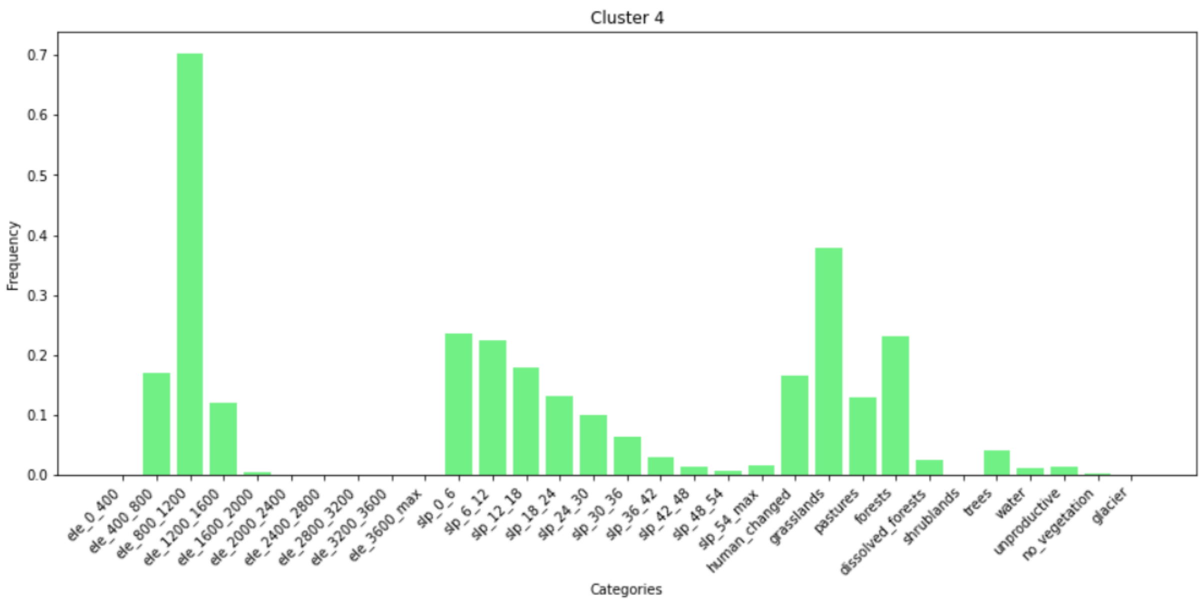
b)

Figure 4.17: a) Distribution of objects whose histograms were partitioned into cluster 3 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 3.

- Objects in cluster 5 are mostly located in 'Voralpen' areas, near the major Alps region, with elevations in the peripheral areas typically between 1200m and 2000m, sometimes also greater than 2000m or lower than 1200m, and with steep slopes. The most prevalent land cover type on average in the peripheral areas is pastures, followed by grasslands and forests there. Figure 4.19 shows the conditions of it, including the distributions of objects and the cluster center. Overall, this cluster contains 1766 records.



a)

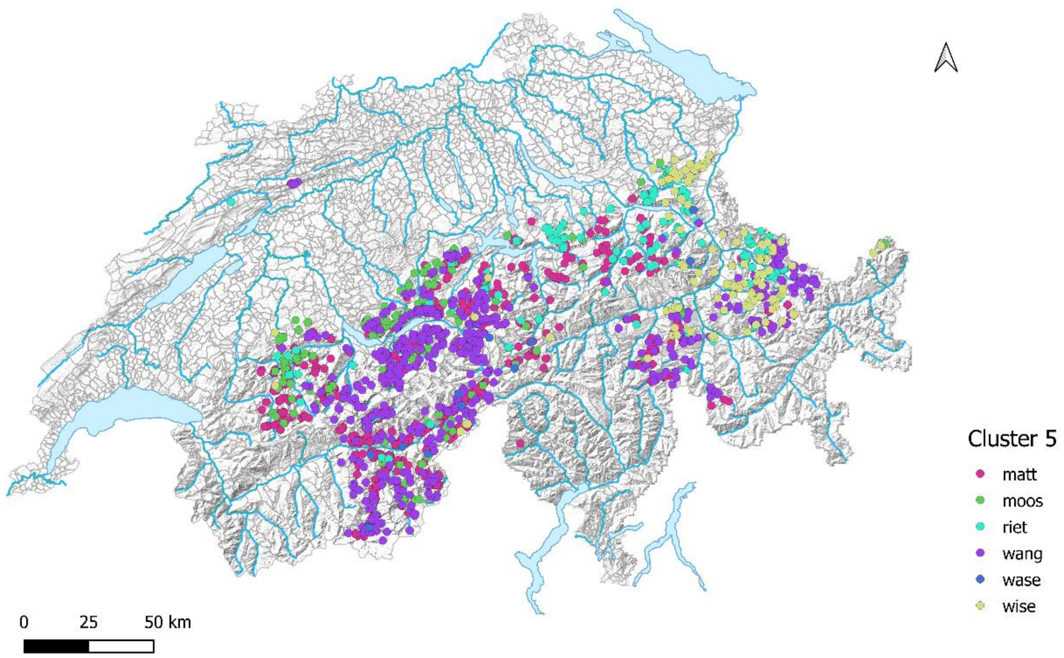


b)

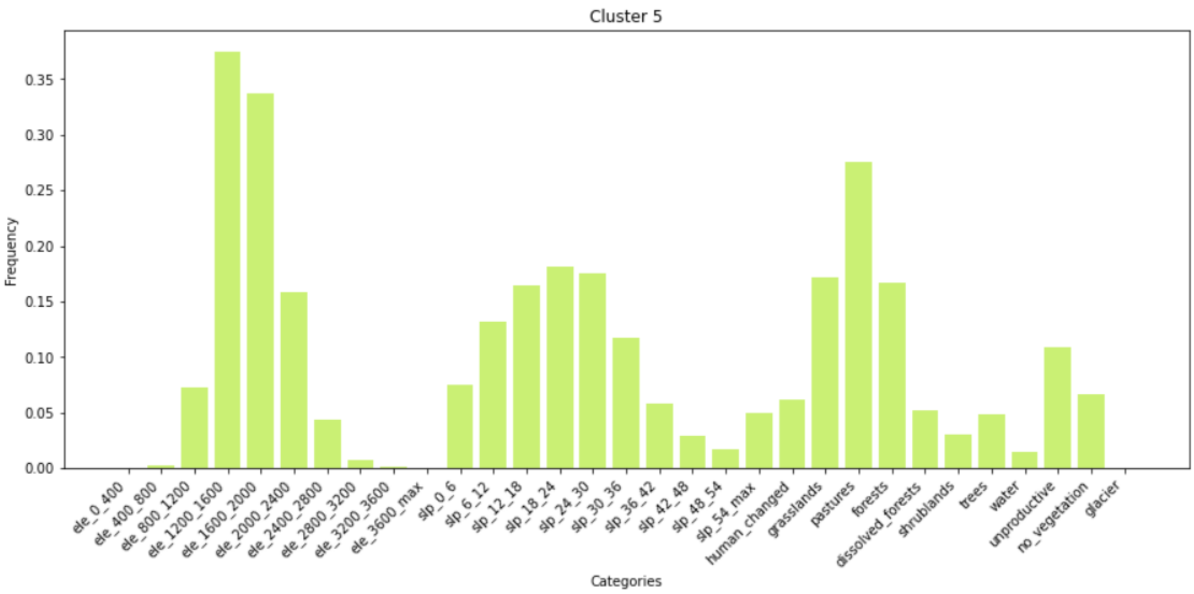
Figure 4.18: a) Distribution of objects whose histograms were partitioned into cluster 4 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 4.

- No cluster exists for open areas where objects are located around the main Alps ridge, like the cluster 6 for convex landforms. This is reasonable, as land cover types of objects' peripheral areas in that cluster are usually 'no vegetation', as introduced in 4.2.1, with which it's normal that they don't have names about meadows or grasslands.

With these clustering results, I investigated the patterns of those generic parts representing open areas, by finding the distributions of the local descriptors within each of those 6 bags of words. In the following paragraphs, I will discuss the patterns of these generic parts. Meanwhile, for pairs or sets of generic parts that show similar patterns, I also conducted cosine similarity analysis, which are also included here. Note that, in the visualizations illustrating patterns, these clusters are marked by 'cluster 0' to 'cluster 4', where the 'cluster 0' there corresponds to cluster 1 mentioned above, 'cluster 1' means cluster 2 mentioned above, and so on and so forth.

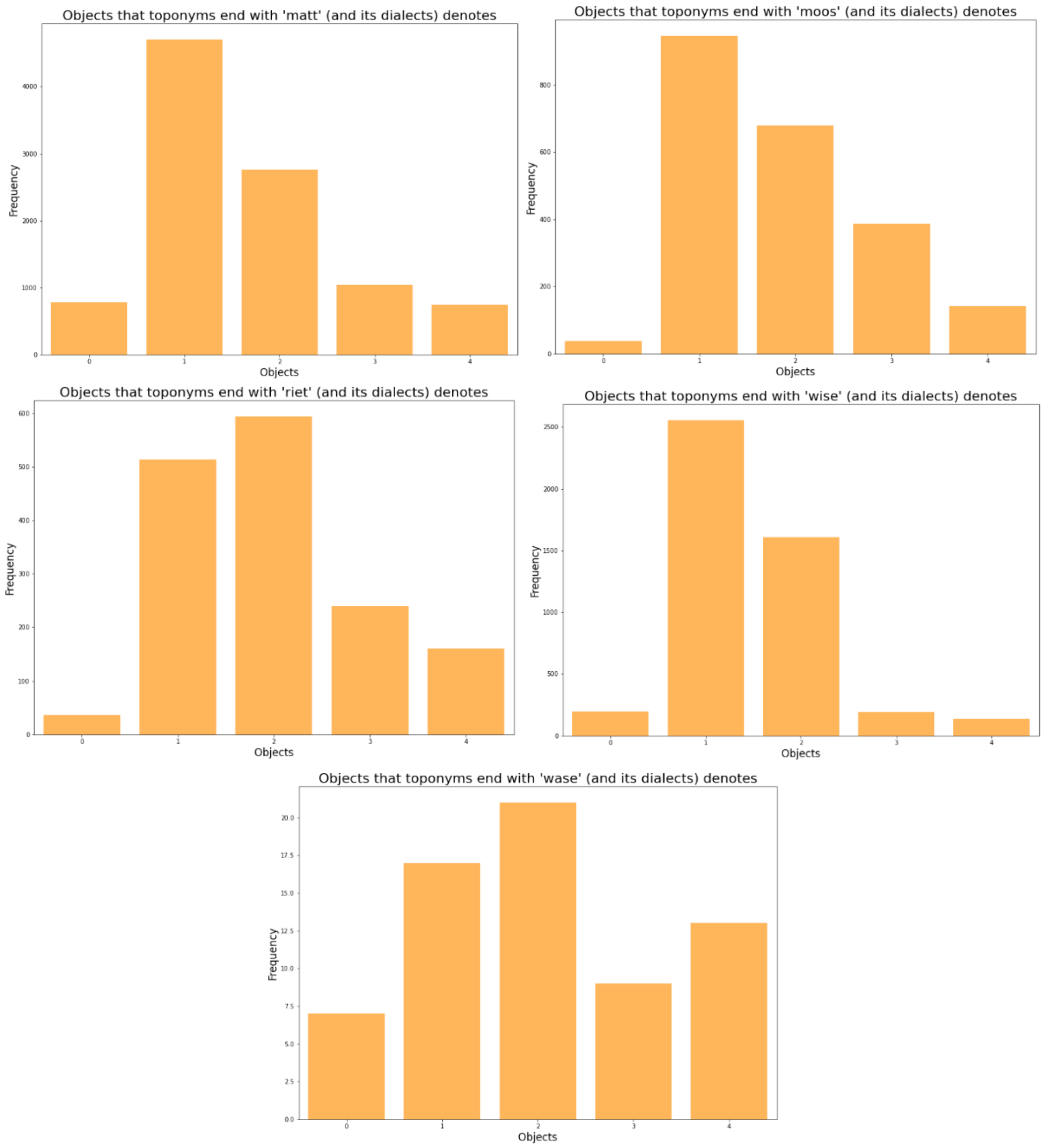


a)

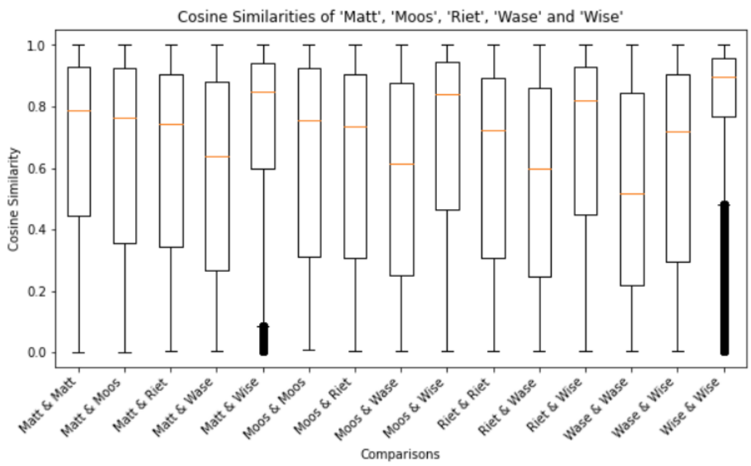


b)

Figure 4.19: a) Distribution of objects whose histograms were partitioned into cluster 5 with different colors representing different generic parts of their names. b) The histogram of cluster center for cluster 5.



a)



b)

Figure 4.20: a) Patterns of ‘Matt’, ‘Moos’, ‘Riet’, ‘Wase’ and ‘Wise’. b) Cosine similarities of ‘Matt’, ‘Moos’, ‘Riet’, ‘Wase’ and ‘Wise’.

Among those 6 generic parts denoting open areas, I observed that 'Matt' & 'Moos' share nearly identical patterns, which are also similar to patterns of 'Riet', 'Wise' and 'Wase', as shown in Figure 4.20 a). Specifically, all of them frequently denote open areas assigned to clusters 2 and 3, with relatively flat terrain and gentle slopes, sometimes used by people while sometimes grasslands and forests are mostly kept. Occasionally they represent objects assigned to other clusters as well (for 'Wase' this could be happened more frequently than others). Given the similarity in patterns of them, cosine similarities were calculated for further comparisons. Self- and cross-group comparisons among 'Matt', 'Moos', 'Riet' and 'Wise' showed median cosine values of greater than 0.7, meaning that they denote similar kinds of landscape, and they are autocorrelated regarding the landscape they denote. Especially, 'Wise' was found to be strong autocorrelated. While for 'Wase', it's not autocorrelated, which could also be assumed through its pattern, which is more or less scattered. Figure 4.20 b) illustrates the results of the cosine similarities.

Besides, 'Wang' was found to be a distinct case, as no other generic parts show similar pattern as it, according to Figure 4.21. Specifically, it mostly denotes objects partitioned in cluster 5, which means in the peripheral areas the terrain is undulated with higher altitudes, and the land is usually covered with grasslands, pastures and forests. Though it seems that 'Wang' is strong autocorrelated regarding the landscape they denote according to its pattern, the self-comparison shows that it's actually not auto-correlated, as illustrated by Figure 4.21. The reason could be that, cluster 5 also contains various kinds of objects itself.

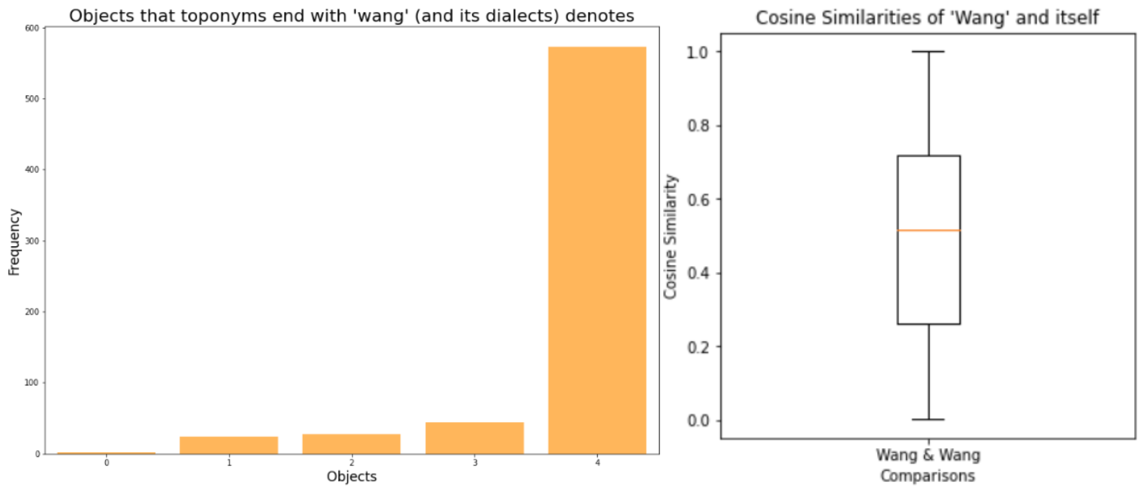


Figure 4.21: Patterns and self-similarity of 'Wang'.

In 5.1, I will come back to these results as well, to further discuss them regarding their meanings from sources, compare with other researchers' works, and answer the research questions.

Chapter 5 Discussions

5.1 Analysis of the Results

Recall that, in 2.4.2 I proposed the following two research questions:

RQ1: Does a specific generic part of Swiss toponyms in German always relate to the same kind of landscape (including topography and land cover)? Considering other generic parts that are identified to denote same categories of geographical objects with it, do the sets of objects they denote show similar patterns regarding landscape?

RQ2: How translatable are some of the generic parts of Swiss toponyms in German compared to those in Italian, still from the perspective of landscape?

In this section, I will come back to discuss them with the results I got in the research, combining the meanings of related generic parts from various sources.

5.1.1 Generic Parts Denoting 'Convex Landforms'

Among those 19 'normal' generic parts related to convex landforms studied in this research, 14 could be regarded as denoting diverse landscape, given their scattered patterns, and the lack of autocorrelation observed through cosine similarities calculations. In contrast, 'Bool', 'Buck', 'Chapf', 'Hoger' and 'Pizzo' consistently represent the same kind of landscape, individually. According to Ortsnamen.ch (Schweizerisches Idiotikon, 2023), the first 3 generic parts in these five, together with 'Büel', 'Chopf' and 'Gubel', contain the semantic meaning 'hills', however, they are not always the same regarding their patterns. Specifically, 'Bool' & 'Buck' were found to be more similar. Objects whose names contain them could in general indeed be regarded as 'hills', since usually the altitudes are low and the terrain is flat in their peripheral areas. Though the meaning of 'Hoger' was not really identified as 'hills', it was also found to be similar with these two in its pattern. On the other hand, 'Chapf', 'Büel', and 'Gubel' occasionally denote convex landforms with higher altitudes and steeper slopes, which is different from those three, but more similar to 'Berg', 'Stein' and 'Spitz', and 'Chopf' was discovered to denote this kind of convex landforms more frequently, which is even similar as 'Nolle' which contains semantic meaning 'peak' according to Ortsnamen.ch (Schweizerisches Idiotikon, 2023). Additionally, 'Horn', 'Cima' and 'Pizzo' were also found to represent peaks (Olivieri, 1931; Schweizerisches Idiotikon, 2023), but they were verified to mostly denote peaks around the main ridge of Alps in this study, with higher altitudes steeper slopes compared to 'Nolle'. For some other generic parts about convex landforms (but not 'hills' or 'peaks'), they were also found to have various patterns, like 'Flue', 'Gütsch', 'Stock' etc. Overall, generic parts related to convex landforms really display diverse patterns.

Those 6 diminutive forms studied in this research should also be discussed separately. Among them, some were found to denote convex landforms which are on average of lower altitudes and flatter terrain compared to their ‘normal’ forms. Nonetheless, diminutive forms of ‘Büel’ and ‘Berg’ are quite unique, especially for ‘Berg’ whose diminutive forms even seems to denote grander convex landforms compared to normal forms. Investigating the reasons of these might also provide some interesting insights to the naming of geographical objects, though it was not studied in this research.

5.1.2 Generic Parts Denoting ‘Open Areas’

Among those 6 generic parts related to open areas studied in this research, 4 could be regarded as representing similar landscape individually, given their clustered patterns and high median cosine values, including ‘Matt’, ‘Moos’, ‘Riet’ and ‘Wise’. They were discovered to be located in areas with flat terrain and low altitude, sometimes mainly covered by human-built objects, while sometimes grasslands, forests and pastures are still remained. ‘Wang’, on the other hand, could usually be found in areas with undulated terrain with higher altitudes, and covered with grasslands, pastures and forests frequently. This is in line with the explanation of ‘Wang’ from Ortsnamen.ch (Schweizerisches Idiotikon, 2023), where it denotes ‘grasslands on steep slopes’. And probably because of this, it has not changed a lot by people to urban areas or to farmlands, compared to the others.

5.1.3 Back to the Research Questions

RQ1.1: Does a specific generic part of Swiss toponyms in German always relate to the same kind of landscape (including topography and land cover)?

There is no universal answer to this question. In this research, I found some generic parts that could mostly denote the same kinds of objects, like ‘Buck’. On the contrary, some could denote various kinds of objects, like ‘Stock’. Probably, this is due to the fact that in Switzerland, there indeed exist different kinds of terrain and land cover. Consequently, people in different regions in Switzerland might name different objects with same terms, according to different definitions of such terms in their mind. However, this is just a hypothesis. Further investigations are needed to determine the specific reasons. Hence, here comes a new research question for future studies: ‘Why do some generic parts in Swiss toponyms denote different landscapes?’ waiting to be answered.

RQ1.2: Considering generic parts that are identified to denote same categories of geographical objects, do the sets of objects they denote show similar patterns regarding landscape?

There is no universal answer to this question as well. In this research, some generic parts were found to denote similar objects compared to each other, even if sometimes they were not found to contain same semantic meanings according to various sources. Conversely, some genetic parts share the same semantic meanings according to sources, but actually they were discovered to denote different kinds of objects. The reasons for such discrepancies are also waiting to be mined. Is it due to specific usage by people living in different regions of Switzerland, each with various perception of landscape? These need to be further investigated. Together, the reasons that some diminutive forms of generic parts about convex landforms were found to denote even grander objects than ‘normal’ ones, are also needed to be further studied, which is, probably the result of people's different perceptions to landscape, combined with their different usage of such generic parts as well.

RQ2: How translatable are some of the generic parts of Swiss toponyms in German compared to those in Italian, still from the perspective of landscape?

According to the patterns of those three generic parts in Italian about convex landforms, I found that 'Pizzo' and 'Cima' are, to some extent, 'translatable' to 'Horn' in German, since they show similar patterns regarding landscape. However, 'Monte' is not translatable to any of the generic parts in German. This is also the reason why exonyms were excluded from this research, as in this case, they might not reveal how people initially perceived landscape. Therefore, exonyms that were directly translated could lead to biased results.

5.1.4 Back to the Research Gaps and Related Works

This research filled the research gaps by investigating more generic parts in Swiss toponyms, with more 'global' methods, and included cross-language studies. Compared to the previous research about Swiss toponyms, this study generated richer results about various generic parts, provided more information for the future studies about people's perception to the landscape. Additionally, this research got some different results compared to previous studies. Derungs et al. (2013) found that, 'Spitz' and 'Horn' share similar patterns considering the landscape they denote, which is totally different from the pattern of 'Berg'. However, in my study, 'Spitz' was found to be more similar to 'Berg' than to 'Horn'. One possible explanation to this difference is that we used different methods for investigation, with data of different sample size.

5.2 Decisions Made

In this section, I would like to elaborate on some of the decisions made in the study, regarding toponyms and landscape, along with discussions of some possible alternative approaches.

5.2.1 Toponyms

In line with the criteria outlined in 2.1.3, I specifically chose certain types of toponyms for the analysis. In detail, only nature-related micro-endonyms were considered to reduce the risk of obtaining biased results. Hydronyms were also excluded due to the limited variation in generic parts in Switzerland. These decisions made could improve the accuracy and specificity, as well as overall persuasiveness of the research.

In addition to the types of toponyms, I also considered their structural components. As stated in 2.1.4, they typically consist of generic parts that classify features, and qualifier parts, which describe objects. In this research, I only studied on their generic parts, to compare the kinds of landscape they could represent. However, I must admit that qualifier parts deserve attention as well, according to the results I got, as such study might shed light on why some specific generic parts could denote various landscape. For example, the qualifier parts 'Klein' and 'Gross' could describe the size of objects, while 'Ober' and 'Unter' may correlate to the objects' elevation (Schweizerisches Idiotikon, 2023). They might therefore imply a connection with the objects' characteristics. What's more, some other aspects of the features might also be revealed by them. Take the example of 'Milchbuck' used throughout this thesis again: 'Buck' in this toponym has the classification function, meaning that the related object is a hill, while 'Milch' as the qualifier part, probably means the area is covered with juicy grass, which on the other hand, have some relationship with land cover. The associations of them with generic parts are also worth being studied, as what Villette (2021) once did. In her research, the meaningful elements combined with 'Wald', 'Holz', 'Riet' and 'Moos' were investigated, to reveal the semantic information of them. In all, qualifier parts in toponyms are not 'useless'. Investigating them might reveal interesting patterns as well.

5.2.2 Land Cover and Land Use

As claimed in 2.2.2 and 3.2.3.2, I only considered land cover data for analysis, as land use data mostly reflects how people use the related areas, which is not in line with the landscape-related research. And even within land cover data, there also exists multiple human-related categories. Therefore, I did a reclassification to classify the land cover into 11 main classes: human-built or human-changed objects, grasslands, pastures, forests, dissolved forests, bushes, other trees, water areas, unproductive vegetation, no vegetation and glacier. There could have alternative approaches to reclassify them, which could potentially yield different results. For instance, I grouped 'farmlands' with other human-built features like buildings and roads, while 'pastures' were remained separate. This decision was based on the understanding that, while pastures are utilized by humans as well, they predominantly retain their natural state as grasslands. For the needs of different approaches, the reclassification of these two categories, probably as well as other categories, should be reconsidered.

5.3 Methodologic Approaches

This research was conducted in two main phases, including the extraction of generic parts and the comparisons of them. In this section, I will discuss the approaches, compare with others' works, and talk about some limitations.

5.3.1 Extraction of Generic Parts

To extract the generic parts in German toponyms, I employed a somehow 'exhaustive' approach in this study, as outlined in 3.3.1. Specifically, the recursive process did not operate in a fully automatic way but involved manual filtering and determination referring to sources containing information about toponyms. In doing so, I ensured that I could uncover as many generic parts as possible for the research, including many variants in plural forms or in different dialects, as well as diminutive forms. However, this approach is aa bit time-consuming. Probably, machine learning methods like word embedding could be helpful to find more variants of generic parts, or information retrieval might be useful to make the searching of information about toponyms easier, which were not implemented in this study. In that case, one might be able to conduct the related research in a more rapid way.

5.3.2 Comparison of Generic Parts

In this study, I implemented zonal methods to represent the landscape of the geographic objects, which was similar to what Derungs et al. (2013) and Villette (2021) did. In this case, the objects could be investigated from a more 'global' view, which using merely the data at exact positions could not realize. Imagine that we know the elevation of an object in Switzerland to be 1000m. While it might represent a mountain in the Swiss Plateau, it could also be situated in a valley within the Voralpen regions. Therefore, we could not get accurate results via that way. Using the data in the peripheral areas of objects, I calculated histograms to describe their landscape, considering the distribution of elevations and slopes, as well as the frequencies of land cover types. Then I constructed a bag of words for each generic part and aggregated those groups of bags of words I wanted to compare into comprehensive ones. Employing K-means clustering, I segmented each comprehensive 'bag' into some clusters, and comparisons were made among the bags of words within each comprehensive bag, based on the distribution of the histograms contained in them according to those clusters. This process has been used for classifying texts and images for many years, while as analyzing the raster files of elevation, slope and land cover is just like analyzing images with RGB values for colors, I applied this analytical method in my research. Through this way richer information for representing the landscape of a position could be got, compared to using specific indices like mean slope, relief, standard deviation of elevation etc. which was considered by Derungs et al. (2013), hence the results of this study might reflect the reality better, which could make it more persuasive.

Nevertheless, for the implementation of clustering and results of the study, there are something needed to be noted. For those generic parts showing similar pattern according to the distribution of their histograms based on the clustering results, we can not be arbitrary to say that they are similar without the help of cosine similarity, as groups of data points showing different patterns could actually be classified into the same cluster, as Figure 5.1 illustrates. On the other hand, we should also not determine that sets of generic parts showing low median cosine values but similar patterns denote different objects arbitrarily, as they might also denote scattered objects individually, making the median cosine values of self-comparisons low as well. Whenever this kinds of issues happen, more other studies would better be done to get more precise results.

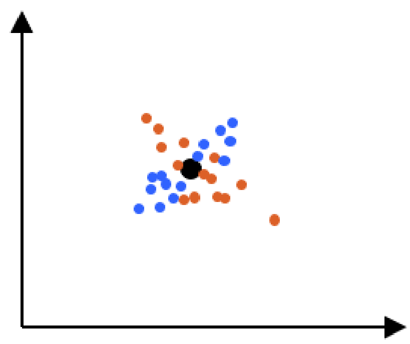


Figure 5.1: An example of 2-D data showing different patterns but are partitioned into the same cluster. The yellow dots and the blue dots here denote two different dataset, while the black dot is the center of this cluster.

5.4 Limitations

Besides of the limitations of this research discussed in 5.2 and 5.3, there still exists more aspects that needed to be pointed out, which are contained in this section.

5.4.1 Buffer Areas

In this research, I determined the buffer distances used for investigating an object's peripheral areas to be 1000m and 200m for topography and land cover respectively, after doing some tests. However, in this way I ignored the fact that objects size could also be revealed by their names. Future research could be conducted to investigate this, by considering size as another index, as what Villette (2021) did.

5.4.2 Cut-off Values

In this study, I determined the cut-off values of topography (including elevations and slopes) based on the distribution of them in Switzerland, without any tests. By fine-tuning such values, probably some different results could be generated, as the clustering processes totally depended on them.

5.4.3 Cross-dialect Analysis

In this research, I conducted cross-language analysis, via which I found that the generic parts are sometimes not translatable from one language to another. However, as there exists various dialects in Switzerland, cross-dialect analysis might also reveal some interesting patterns. For example, do 'Berg' and 'Bärg' denote similar landscape? Note that, even if the answer is yes, one could not arbitrarily say that this is because of the cultural difference as well, as it might be caused by various landscape features there are in different regions as well.

Chapter 6 Conclusion

6.1 Summary

This research set out to investigate the relation of toponyms and landscape, shedding light on the way people named geographic objects. Specifically, I first extracted nature-related generic parts (excluding those related to water areas or very tiny landforms) in toponyms from gazetteer with the help of multiple sources and classified them into several categories according to their semantic meanings identified by those sources. Then, comparisons were made among generic parts denoting convex landforms and those representing open areas as case studies. Landscape including topography and land cover associated with these generic parts were investigated via the 'bag-of-words' and the clustering methods. For each case study, the patterns of generic parts were compared, together with the help of cosine similarities. The research found that, first, some of the generic parts can represent multiple kinds of landscapes individually. What's more, even though some generic parts were identified to denote similar kinds of landscapes according to the sources, they were actually found to show different patterns. Diminutive forms of generic parts were also found to not denoting smaller landforms occasionally. Besides, cross-language comparisons were also conducted, via which it was discovered that generic parts in different languages are sometimes not translatable. The investigations of the meanings of such generic parts could potentially pave the way for further investigations, about the way people perceived landscapes and named geographic objects. This might let us get closer to the cultural heritages.

6.2 Future Works

Multiple future works could be conducted based on the results of this research. First, a lot of generic parts denoting different kinds of landscape were found, however, only those about convex landforms and those related to open areas were investigated. Other categories including slopes, valleys etc. are also worth being studied, though they contain fewer generic parts, while comparisons of those generic parts might also reveal some interesting insights. Second, I only conducted cross-language comparisons among generic parts in German and Italian. As there are other language regions in Switzerland (and more broadly, in Europe), far more this kind of investigations could be done. And as suggested in the previous chapter, cross-dialect research might also reflect some differences in naming geographic objects among people living in various regions in, for example, German-speaking Switzerland. Third, qualifier parts could also be combined together for analysis. Though they do not have classifying functions, they have descriptive functions and could also be related to landscape. Last, the diversity of some generic parts in their patterns, and the differences of generic parts about landscape they denote, as found by this research, could also be studied from cultural side, to reveal more about the reasons of such patterns.

References

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*. United States Government Printing Office.

Antrop, M., & Van Eetvelde, V. (2019). Territory and/or Scenery: Concepts and Prospects of Western Landscape Research. In L. Mueller & F. Eulenstein (Eds.), *Current Trends in Landscape Research* (pp. 3–39). Springer International Publishing. https://doi.org/10.1007/978-3-030-30069-2_1

Atik, M., Kanabakan, A., Ortaçeşme, V., & Yildirim, E. (2022). Tracing Landscape Characters Through Place Names in Rural Mediterranean. *CATENA*, *210*, 105912. https://doi.org/10.1016/j.catena.2021.105912

Atik, M., & Swaffield, S. (2017). Place Names and Landscape Character: A Case Study from Otago Region, New Zealand. *Landscape Research*, *42*(5), 455–470. https://doi.org/10.1080/01426397.2017.1283395

Bastian, O. (2008). Landscape Classification – Between Fact and Fiction. *Paper Presented at the Conference Warzaw University and Polish Association of Landscape Ecology. Warzawa: Polska Asocjacja Ekologii Krajobrazu.*

Blair, D., & Tent, J. (2021). Revised Typology of Place-Naming. *Names*, *69*(4), 30–47. https://doi.org/10.5195/names.2021.2260

Brabyn, L. (2009). Classifying Landscape Character. *Landscape Research*, *34*(3), 299–321. https://doi.org/10.1080/01426390802371202

Bredart, S., Brennen, T., & Valentine, T. (2002). *The Cognitive Psychology of Proper Names*. Taylor & Francis e-Library.

Burenhult, N., & Levinson, S. C. (2008). Language and Landscape: A Cross-Linguistic Perspective. *Language Sciences*, *30*(2–3), 135–150. https://doi.org/10.1016/j.langsci.2006.12.028

Cablitz, G. H. (2008). When “What” is “Where”: A Linguistic Analysis of Landscape Terms, Place Names and Body Part Terms in Marquesan (Oceanic, French Polynesia). *Language Sciences*, *30*(2–3), 200–226. https://doi.org/10.1016/j.langsci.2006.12.004

Calvo-Iglesias, M. S., DÍaz-Varela, R. A., MÉndez-MartÍnez, G., & Fra-Paleo, U. (2012). Using Place Names for Mapping the Distribution of Vanishing Historical Landscape Features: The *Agras* Field System in Northwest Spain. *Landscape Research*, *37*(4), 501–517. https://doi.org/10.1080/01426397.2011.604716

Clark, I. (2009). Reconstruction of Aboriginal microtoponymy in western and central Victoria: Case studies from Tower Hill, the Hopkins River, and Lake Boga. In H. Koch & L. Hercus (Eds.), *Aboriginal Placenames. Naming and re-naming the Australian landscape*. ANU Press. https://doi.org/10.22459/AP.10.2009.09

Coates, R. A. (2006). Properhood. *Language*, *82*, 356–382. https://doi.org/10.1353/lan.2006.0084

Conedera, M., Vassere, S., Neff, C., Meurer, M., & Krebs, P. (2007). Using Toponymy to Reconstruct Past Land Use: A Case Study of ‘brüsáda’ (Burn) in Southern Switzerland. *Journal of Historical Geography*, *33*, 729–748.

Council of Europe Landscape Convention, Pub. L. No. ETS No. 176, 1 (2000).

Defries, R. S., & Townshend, J. R. G. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing*, *15*(17), 3567–3586. https://doi.org/10.1080/01431169408954345

DeMers, M. N. (2001). *GIS Modeling in Raster*. John Wiley & Sons.

Derungs, C., Wartmann, F., Purves, R. S., & Mark, D. M. (2013). The Meanings of the Generic Parts of Toponyms: Use and Limitations of Gazetteers in Studies of Landscape Terms. In T. Tenbrink, J. Stell, A. Galton, & Z. Wood (Eds.), *Spatial Information Theory* (Vol. 8116, pp. 261–278). Springer International Publishing. https://doi.org/10.1007/978-3-319-01790-7_15

Feng, C.-C., & Mark, D. M. (2017). Cross-Linguistic Research on Landscape Categories Using GEOnet Names Server Data: A Case Study for Indonesia and Malaysia. *The Professional Geographer*, *69*(4), 567–578. https://doi.org/10.1080/00330124.2017.1288575

Fisher, P. F., Comber, A. J., & Wadsworth, R. (2005). Land use and Land cover: Contradiction or Complement. In *Re-Presenting GIS*. John Wiley & Sons. https://www.wiley.com/en-ca/Re+Presenting+GIS+-p-9780470848470

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*(3), 399–409. https://doi.org/10.1016/S0034-4257(97)00049-7

Gammeltoft, P. (2005). *In Search of the Motives Behind Naming. A Discussion of Name-semantic Model of Categorisation*. Proceedings of the 21st International Congress of Onomastic Sciences, Uppsala 19-24 August 2002.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Goodchild, M. F., & Hill, L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, *22*(10), 1039–1044. https://doi.org/10.1080/13658810701850497

Hard, G. (1970). *Die 'Landschaft' der Sprache und die 'Landschaft' der Geographen: Semantische und forschungslogische Studien zu einigen zentralen Denkfiguren in der deutschen geographischen Literatur*. In Kommission bei F. Dümmler.

Helleland, B. (2002). *The Social and Cultural Value of Place Names*. Eighth United Nations Conference on the Standardization of Geographical Name, Berlin.

Hollis, J., & Valentine, T. (2001). Proper-name Processing: Are Proper Names Pure Referencing Expressions? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 99–116. https://doi.org/10.1037/0278-7393.27.1.99

Hough, C. (2006). Place Names. In *Encyclopedia of Language & Linguistics* (2nd ed., pp. 613–620).

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323

Huber, J. (1885). *Schweizerisches Idiotikon*. https://digital.idiotikon.ch/

Jett, S. C. (1997). Place-Naming, Environment, and Perception among the Canyon de Chelly Navajo of Arizona. *The Professional Geographer*, *49*(4), 481–493. https://doi.org/10.1111/0033-0124.00094

Jones, M. C. (2015). *Policy and Planning for Endangered Languages*. Cambridge University Press. https://books.google.ch/books?id=v-ZBCgAAQBAJ

Jordan, P. (2012). Place Names As Ingredients of Space-Related Identity. *Oslo Studies in Language*, *4*(2), 117–131. https://doi.org/10.5617/osla.314

Kadmon, N. (2002). *Glossary of terms for the standardization of geographical names*. United Nations. https://unstats.un.org/unsd/ungegn/pubs/documents/Glossary_of_terms_rev.pdf

Levinson, S. C. (2008). Landscape, Seascape and the Ontology of Places on Rossel Island, Papua New Guinea. *Language Sciences*, *30*(2–3), 256–290. https://doi.org/10.1016/j.langsci.2006.12.032

Li, Z., Zhu, Q., & Gold, C. (2005). *Digital terrain modeling: Principles and methodology*. CRC Press.

Light, D., & Young, C. (2015). Toponymy as Commodity: Exploring the Economic Dimensions of Urban Place Names. *International Journal of Urban and Regional Research*, *39*(3), 435–450. https://doi.org/10.1111/1468-2427.12153

Manning, C. D., Raghavan, P., & Schütze, H. (2009). *Introduction to Information Retrieval*. Cambridge University Press. https://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf

Mark, D. M., Skupin, A., & Smith, B. (2001). Features, Objects, and other Things: Ontological Distinctions in the Geographic Domain. In D. R. Montello (Ed.), *Spatial Information Theory* (pp. 488–502). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45424-1_33

Mark, D. M., & Smith, B. (2004). A Science of Topography: Bridging the Qualitative-Quantitative Divide. *Geographic Information Science and Mountain Geomorphology*, 75–100.

Mark, D. M., & Turk, A. G. (2017). Ethnophysiography. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Eds.), *International Encyclopedia of Geography* (1st ed., pp. 1–11). Wiley. https://doi.org/10.1002/9781118786352.wbieg0349

Miccoli, P. (2019). Colonial place-names in Italian East Africa (AOI) (with additional data from Tripoli). The linguistic heritage of colonial practice. In B. Weber (Ed.), *The Linguistic Heritage of Colonial Practice* (pp. 75–92). De Gruyter. https://doi.org/10.1515/9783110623710-004

Miller, C., & Laflamme, R. (1958). The digital terrain model—Theory and applications. *Photogrammetric Engineering*, *24*, 433–442.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461. https://doi.org/10.1007/s10708-014-9602-6

Mitkov, R. (2014). *The Oxford Handbook of Computational Linguistics* (2nd ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199573691.001.0001

Mücher, C. A., Klijn, J. A., Wascher, D. M., & Schaminée, J. H. J. (2010). A new European Landscape Classification (LANMAP): A Transparent, Flexible and User-oriented Methodology to Distinguish Landscapes. *Ecological Indicators*, *10*(1), 87–103. https://doi.org/10.1016/j.ecolind.2009.03.018

Olivieri, D. (1931). *Dizionario di toponomastica lombarda* (2nd ed.). https://archive.org/details/toponomastica-lombarda/page/n195/mode/2up

Penko Seidl, N. (2008). Significance of Toponyms, with Emphasis on Field Names, for Studying Cultural Landscape. *Acta Geographica Slovenica*, *48*(1), 33–56. https://doi.org/10.3986/AGS48102

Perono Cacciafoco, F., & Cavallaro, F. (2023). *Place Names*. Cambridge University Press. https://www.cambridge.org/core/books/place-names/B6C6CB54DF0896D31CAE470C710D22D1

Qian, S., Kang, M., & Wang, M. (2016). An Analysis of Spatial Patterns of Toponyms in Guangdong, China. *Journal of Cultural Geography*, *33*(2), 161–180. https://doi.org/10.1080/08873631.2016.1138795

Qiu, W., & Joe, H. (2009). *clusterGeneration: Random cluster generation (with specified degree of separation).* [Computer software]. https://cran.r-project.org/web/packages/clusterGeneration/clusterGeneration.pdf

Radding, L., & Western, J. (2010). What's in a Name? Linguistics, Geography, and Toponyms. *Geographical Review*, *100*(3), 394–412. https://doi.org/10.1111/j.1931-0846.2010.00043.x

Redmonds, G. (2004). *Names and History: People, Places and Things.* A&C Black.

Rennick, R. (2005). How to Study Placenames. *Names*, *53*(4), 291–308. https://doi.org/10.1179/nam.2005.53.4.291

Rigaux, P., Scholl, M., & Voisard, A. (2022). *Spatial Databases—With application to GIS*. Morgan Kaufmann.

Sarstedt, M., & Mooi, E. (2019). *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56707-4

Savage, V. R. (2020). Place names. *International Encyclopedia of Human Geography*, *10*, 129–138. https://doi.org/10.1016/B978-0-08-102295-5.10852-2

Schweizerisches Idiotikon. (2023). *Ortsnamen.ch*. Ortsnamen.Ch. https://www.ortsnamen.ch/

Senft, G. (2008). Landscape Terms and Place Names in the Trobriand Islands – the Kaile'una Subset. *Language Sciences*, *30*(2–3), 340–361. https://doi.org/10.1016/j.langsci.2006.12.001

Simensen, T., Halvorsen, R., & Erikstad, L. (2018). Methods for Landscape Characterisation and Mapping: A Systematic Review. *Land Use Policy*, *75*, 557–569. https://doi.org/10.1016/j.landusepol.2018.04.022

Sivic, J., & Zisserman, A. (2009). Efficient Visual Search of Videos Cast as Text Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(4), 591–606. https://doi.org/10.1109/TPAMI.2008.111

Swiss Federal Council. (2023). *Geography about Switzerland*. https://www.eda.admin.ch/aboutswitzerland/en/home/umwelt/geografie.html

Swiss Federal Council. (2024). *Languages and dialects*. https://www.eda.admin.ch/aboutswitzerland/en/home/gesellschaft/sprachen/sprachen-und-dialekte.html

Swiss Federal Office of Topography Swisstopo. (2004). *Product information of DHM25*. https://backend.swisstopo.admin.ch/fileservice/sdweb-docs-prod-swisstopoch-files/files/2023/11/14/c4f59d00-5935-4de3-88dc-fa699bf685ea.pdf

Swiss Federal Office of Topography Swisstopo. (2020). *Produktinformation swissNAMES3D*. https://backend.swisstopo.admin.ch/fileservice/sdweb-docs-prod-swisstopoch-files/files/2023/11/14/7068474d-a976-4168-932a-63dd7c3a95d7.pdf

Swiss Federal Office of Topography Swisstopo. (2022). *swissALTI3D* [dataset]. https://www.swisstopo.admin.ch/en/geodata/height/alti3d.html

Swiss Federal Office of Topography Swisstopo. (2023a). *SwissNAMES3D* [dataset]. https://www.swisstopo.admin.ch/en/geodata/landscape/names3d.html

Swiss Federal Office of Topography Swisstopo. (2023b). *Produktinformation swissALTI3D*. https://backend.swisstopo.admin.ch/fileservice/sdweb-docs-prod-swisstopoch-files/files/2023/11/14/6d40e558-c3df-483a-bd88-99ab93b88f16.pdf

Swiss Federal Office of Topography Swisstopo. (2024). *DHM25* [dataset]. https://www.swisstopo.admin.ch/en/height-model-dhm25

Swiss Federal Statistical Office. (2021). *Swiss Land Use Statistics* [dataset]. https://www.bfs.admin.ch/bfs/en/home/services/geostat/swiss-federal-statistics-geodata/land-use-cover-suitability/swiss-land-use-statistics.html

Swiss Federal Statistical Office. (2022). *Die 4 Sprachgebiete der Schweiz nach Gemeinden*. https://www.bfs.admin.ch/bfs/en/home/statistics/population/languages-religions/languages.assetdetail.23907271.html

Swiss National Center for Climate Services NCCS. (2023). *Major Regions*. https://www.nccs.admin.ch/nccs/en/home/regions/grossregionen.html

Tent, J. (2015). Approaches to Research in Toponymy. *Names*, *63*(2), 65–74. https://doi.org/10.1179/0027773814Z.000000000103

Thornton, T. F. (1997). Anthropological Studies of Native American Place Naming. *American Indian Quarterly*, *21*(2), 209–228.

Tomlin, D. C. (1990). *GIS and Cartographic Modeling*. Prentice Hall.

Ursini, F.-A. (2016). On the structure of toponyms. In *Word formation universals* (pp. 375–393). Cambridge Scholars.

Villette, J. (2021). *Exploring Landscape through Place Names: An Interdisciplinary Approach*. University of Zurich.

Wayback Machine. (2018, February 18). *Mfnames.com*. https://web.archive.org/web/20160304120036/http://www.mfnames.com/

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2019). Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, *221*, 173–187. https://doi.org/10.1016/j.rse.2018.11.014

Appendix A Summary of Generic Parts

‘Normal’ form	Variants	Diminutive forms	Meaning	Language
Berg	Berge, Bergen, Berger, Bärg, Bärge, Bärgen, Bärger, Bärga, Perg, Pärg	Bergli, Bärgli, Bergerli, Bärgerli	High terrain elevation	DEU
Bode	Boden, Böde, Böden, Boeden, Bodma, Bodme, Bodmen, Bodmer, Bodmeren, Bödma, Bödme, Bödmen, Bödmer, Bödmeren, Bodmi, Boda, Böda, Bodini, Bodu, Bodo	Bödeli, Bödemli, Bodeli, Bödali, Bödili	Lower land, flat terrain, small plateau, valley floor	DEU
Bool	Boole, Booler, Bol, Boll, Bolle, Bollen, Boller, Bollere, Bolleren, Bohl		Rounded, dome-shaped hills, height, elevation	DEU
Bord	Bort, Borter, Börter		The rising terrain, the small slope	DEU
Buck	Bück, Bücken		Rounded elevation, hill	DEU
Büel	Büele, Büelen, Büeler, Büela, Biel, Biela, Bielu, Biele, Bielen, Bieler, Bühl, Bühel, Büehl, Büchel, Büechel, Büechle, Büechlen, Büechler, Böchel, Bööl, Bööler, Böhl, Böhler, Böl, Böler, Büül, Bül, Büler, Bial, Hügel, Hübel, Hüble, Hüblen, Huble, Hublen, Hubel, Hubil	Hübeli, Büeli, Bieli, Büheli, Bücheli, Büelti, Böli, Bööli, Büüli, Hügeli, Hügli, Hübli, Hubli, Hubeli, Hugeli, Büechli, Büecheli, Hubelti, Büelti, Bielti	Hills, small mountains	DEU
Chanzel	Chanzle, Chanzlen, Chanzler		Outstanding, promising part, ridges or slabs, hills between other outstanding hills	DEU
Chäle	Chälen, Chälle, Chäller, Kehle		Gorge, mountain cut	DEU
Chapf	Kapf, Chapfen, Chäpf		Hill, hilltop	DEU

‘Normal’ form	Variants	Diminutive forms	Meaning	Language
Chopf	Chöpf, Chöpfe, Chöpfer, Kopf, Köpf, Chopfle	Chöpfli, Köpfli	Head-like hills	DEU
Ebnet	Ebnit, Ebni, Äbnit, Äbni, Äbeni, Ebeni, Ebene, Äbnet, Aebnet, Aebnit		Plain area	DEU
Egg	Egge, Eggen, Egger, Eggle, Egglen, Eggler, Egga, Eggu	Eggli, Eggi, Eggeli, Eggelti	Ridge-shaped hills, or sloping ridges, sometimes corners or edges	DEU
Feld	Felden, Felder, Fäld, Fälden		Plain, flat land	DEU
First	Fürst		Mountain ridge	DEU
Flue	Fluh, Flüe, Flüh, Flueh, Fluech, Fluo, Fliela, Flüela, Flüö, Flöö, Flua	Flueli, Flüeli, Flühli	Rocky cliffs, steep rocks, rock walls	DEU
Furgge	Furgg, Furggen, Furgga, Furka, Furkel	Fürggli, Furggle, Furggla, Furggele, Furkel, Furggi, Furggelti, Furggeli	Pass, mountain saddle	DEU
Grund	Gründ, Gründe, Gründen, Gründle, Gründlen, Grünge, Grüngen		Valley bottom, lowland, plain	DEU
Gubel	Gibel, Giebel		Hill	DEU
Gummen	Gumme, Gumm, Chummen, Chumme, Chumm, Gumen, Gume, Gum, Chumen, Chume, Gumma, Chumma, Gomma, Kumma	Gummli, Gummeli, Chummli, Gummi, Chumi	Bowl-shaped valleys, gorges, hollows	DEU
Gütsch	Gütsche, Gütschen, Gutsch		Small, roundish rock head, rocky peak	DEU
Halde	Halden, Halder, Halte, Halter, Halten, Holde, Holden, Hälten, Hälter, Halle, Hallen, Holle, Hollen, Halda, Halta, Holta, Haltu, Haltä, Haalde, Haalte, Haaltu	Haldi, Häldi, Halti, Holti, Haaldi, Haalti, Häldeli, Hälteli, Höldeli, Hölleli, Häldili, Häldli, Haldi, Häldi, Halti, Holti, Haaldi, Haalti	Slope of hills or mountains	DEU
Hell	Höll, Held		The places where they go into the depths, or near the depth	DEU

‘Normal’ form	Variants	Diminutive forms	Meaning	Language
Hoger	Höger		Rounded terrain elevation	DEU
Hol	Hole, Holen, Hohl, Hola		Depressions, hollows, empty inside	DEU
Holz	Holze, Holzen, Holzer, Holzere, Holzeren, Hölzer, Hölzere, Hölzeren, Hölzler	Holzli, Hölzli, Holzerli, Hölzeli	Wood, forests	DEU
Horn	Hore, Horen	Hörnli, Höreli	Rock towers, pointed peaks, sometimes locations that protrude into a lake	DEU
Laub	Loub, Loube		Places where have hardwood	DEU
Loch	Loche, Lochen, Locher, Lochere, Locheren, Löch, Löcher, Löchere, Löcheren	Lochli, Löchli	Deepening of the terrain, hole, cave	DEU
Loo	Löö, Löw, Loh	Löli, Lööli	Groves, bushes, forests	DEU
Matt	Matte, Matten, Matter, Matta, Mattä, Mattu	Mattli, Mättli, Mätteli, Mättili, Matti, Mätti	Flat grassy area, meadow, especially in the valley floor	DEU
Moos	Moosen, Moss, Mos, Mose, Mosen, Moser, Mosler, Mosi, Möser, Mösere, Mösle, Mösler, Möslen, Mies, Mieser, Miesen	Möösli, Mösli, Müüsli, Müsli	Moor, damp, swampy land	DEU
Nolle	Noll, Nollen, Nell, Nelle, Nellen, Nöllen		Rounded mountain peak	DEU
Plangg	Plangge, Planggen, Plangger, Plangga		Steep slopes	DEU
Platte	Platt, Platten, Platti, Blatt, Blatte, Blatten, Blattu, Plattelen, Platta, Blatta, Blatti, Plattis		Rock plateau, wide flat rocks, slope surface	DEU
Riet	Riete, Rieten, Rieter, Rietere, Ried, Riede, Rieden, Rieder, Riedere, Rietle, Rietlen	Riedli, Rietli	Places covered with marsh grass	DEU
Sack	Seck		Sack-shaped terrain	DEU
Sattel			Saddle	DEU

‘Normal’ form	Variants	Diminutive forms	Meaning	Language
Schache	Schach, Schachen		Isolated piece of forest	DEU
Site	Sit, Siten, Seite, Seiten		More or less steep slopes, flanks or entire valley sides	DEU
Spitz	Spitze, Spitzen, Spitzig, Spitza, Spitzler		Pointed, wedge-shaped or narrow course of field shapes	DEU
Stein	Steien, Stei, Steine, Steinen, Steiner, Steinere, Steineren, Steinen, Staa, Steini, Steinig, Steinler, Steindler		Stones, rocks, boulders, sometimes borders or castles	DEU
Stock	Stocke, Stocken, Stocker, Stockere, Stockeren, Stöck, Stöcke, Stöcken, Stöckere, Stöckeren, Stocki, Stogg, Stogge, Stoggen, Stögg, Stögge, Stöggen, Stockera	Stockli, Stockeli, Stockerli, Stöckli	Tree trunks, wooden blocks, cone-shaped heaps	DEU
Stutz	Stütz, Stütze, Stützen, Stotz, Stotzig	Stutzli, Stützli	Steep slope, steep path	DEU
Stuude	Stuud, Stud, Stude, Studen, Studa, Studer, Stüde		Perennial, woody bush, shrub	DEU
Tal	Bachtale, Bachtalen, Taler, Taal, Thal, Bachtel, Bachtele, Bachtelen, Telle, Tellen, Teller, Täl, Täle, Tälen, Täler, Däl, Däle, Dell, Delle, Dellen, Deller, Dellere		The deepening of the terrain	DEU
Tobel	Tobele, Töbel, Töbler	Töbeli	Forest valley, gorge	DEU
Tschugge	Tschugg, Tschuggen, Tschugga, Tschuggi, Tschuggu		Boulders, rock heads	DEU
Wald	Walde, Walden, Wäld, Wälde, Wälden, Waldi, Wäldi, Walda, Walder, Wold, Waald	Wäldli, Wäldeli	Smaller or larger dense stands of trees	DEU
Wang	Wange, Wangen, Wanger, Weng, Wenge, Wengen, Wäng, Wänge, Wängen, Wanger, Wenger		Unforested, grassy, often steeply sloping slopes	DEU

'Normal' form	Variants	Diminutive forms	Meaning	Language
Wanne	Wann, Wannen, Wanner, Wanna, Wannel, Wannelen, Wandel, Wandele, Wandelen, Wanni		Trough-shaped depressions, indentations, larger depressions	DEU
Wase	Wasen		Pieces of lawn	DEU
Wise	Wis, Wisen, Wiser, Wiss, Wisse, Wissen, Wyss, Wyssi, Wysse, Wyssen, Wies, Wiese, Wiesen, Wes, Wese, Wesen, Wees, Weesen, Wasa, Wesa, Wisa, Wiser, Wiis, Wiisse, Wissa	Wisli, Wiesli, Wesli	Grassland, meadow that is regularly mown and temporarily grazed	DEU
Bosch			Bush	ITA
Bosco			Woodlands	ITA
Cima			Tops	ITA
Costa	Costi		More or less steep slopes, sometimes 'towards the plain'	ITA
Monte	Mont, Monti, Mött		Mountains	ITA
Passo	Pass		Pass, mountain saddle	ITA
Pianca			smooth slopes	ITA
Piano	Pian, Piani, Piàn		Less inclined surfaces, which interrupt the steepness of slopes	ITA
Pizzo	Piz, Pizzi		Sharp peaks	ITA

Appendix B Cluster Centers

The cluster centers for each generic part denoting convex landforms or open areas were used sometimes for auxiliary purposes. They are concluded here. In the main body, I used histograms for visualizing the cluster centers of some 'comprehensive' bags of words, however, using that here would be too space-consuming, with more than 60 histograms. Hence, I would rather use arrays alternatively, for representing the cluster centers.

Class	Cluster	Cluster Center
Berg	1	[0.06, 0.91, 0.03, 0, 0, 0, 0, 0, 0, 0, 0.56, 0.22, 0.10, 0.06, 0.03, 0.02, 0.01, 0, 0, 0, 0.37, 0.14, 0.09, 0.35, 0.02, 0, 0.02, 0.01, 0, 0, 0]
	2	[0, 0.11, 0.35, 0.21, 0.16, 0.11, 0.05, 0, 0, 0, 0.10, 0.15, 0.17, 0.17, 0.16, 0.11, 0.05, 0.03, 0.02, 0.05, 0.03, 0.17, 0.26, 0.29, 0.04, 0.02, 0.04, 0.01, 0.07, 0.07, 0]
Horn	1	[0.04, 0.50, 0.28, 0.15, 0.03, 0, 0, 0, 0, 0, 0.32, 0.18, 0.13, 0.11, 0.10, 0.07, 0.03, 0.02, 0.01, 0.03, 0.17, 0.10, 0.10, 0.47, 0.07, 0, 0.02, 0.05, 0.01, 0.01, 0]
	2	[0, 0, 0.04, 0.21, 0.40, 0.30, 0.05, 0, 0, 0, 0.04, 0.09, 0.13, 0.18, 0.20, 0.14, 0.08, 0.04, 0.03, 0.08, 0, 0, 0.29, 0.05, 0.03, 0.04, 0.03, 0, 0.31, 0.23, 0]
	3	[0, 0, 0, 0, 0.03, 0.17, 0.43, 0.24, 0.09, 0.03, 0.04, 0.10, 0.14, 0.17, 0.17, 0.12, 0.07, 0.05, 0.04, 0.12, 0, 0, 0, 0, 0, 0, 0, 0.07, 0.79, 0.14]
Spitz	1	[0.10, 0.88, 0.02, 0, 0, 0, 0, 0, 0, 0, 0.69, 0.16, 0.07, 0.04, 0.02, 0, 0, 0, 0, 0, 0.49, 0.13, 0.08, 0.23, 0.02, 0, 0.02, 0.03, 0.01, 0, 0]
	2	[0, 0.03, 0.22, 0.24, 0.21, 0.14, 0.10, 0.04, 0.01, 0, 0.07, 0.12, 0.15, 0.17, 0.17, 0.12, 0.07, 0.04, 0.02, 0.07, 0.02, 0.08, 0.22, 0.17, 0.05, 0.01, 0.03, 0.01, 0.15, 0.23, 0.02]
Stein	1	[0.05, 0.89, 0.05, 0, 0, 0, 0, 0, 0, 0, 0.58, 0.19, 0.09, 0.05, 0.04, 0.02, 0.01, 0, 0, 0, 0.46, 0.14, 0.08, 0.24, 0.02, 0, 0.02, 0.03, 0, 0, 0]
	2	[0, 0.07, 0.31, 0.27, 0.21, 0.11, 0.02, 0, 0, 0, 0.10, 0.14, 0.16, 0.17, 0.16, 0.11, 0.06, 0.03, 0.02, 0.05, 0.04, 0.14, 0.23, 0.29, 0.06, 0.03, 0.04, 0.01, 0.09, 0.06, 0]

Class	Cluster	Cluster Center
Stock	1	[0.05, 0.92, 0.03, 0, 0, 0, 0, 0, 0, 0, 0.69, 0.17, 0.07, 0.04, 0.02, 0.01, 0, 0, 0, 0, 0.52, 0.12, 0.09, 0.23, 0.01, 0, 0.01, 0.01, 0, 0, 0]
	2	[0, 0.08, 0.33, 0.33, 0.21, 0.04, 0, 0, 0, 0, 0.11, 0.15, 0.17, 0.17, 0.16, 0.11, 0.05, 0.03, 0.02, 0.04, 0.04, 0.13, 0.25, 0.29, 0.08, 0.03, 0.05, 0.01, 0.09, 0.04, 0]
	3	[0, 0, 0, 0.03, 0.18, 0.37, 0.30, 0.09, 0.02, 0, 0.04, 0.09, 0.14, 0.17, 0.19, 0.12, 0.07, 0.04, 0.03, 0.12, 0, 0, 0.07, 0, 0, 0, 0, 0, 0.17, 0.65, 0.09]
Büel	1	[0.05, 0.93, 0.03, 0, 0, 0, 0, 0, 0, 0, 0.63, 0.20, 0.08, 0.04, 0.03, 0.01, 0.01, 0, 0, 0, 0.50, 0.16, 0.10, 0.20, 0.01, 0, 0.02, 0.01, 0, 0, 0]
	2	[0, 0.11, 0.36, 0.22, 0.18, 0.09, 0.02, 0, 0, 0, 0, 0.15, 0.18, 0.17, 0.16, 0.14, 0.09, 0.04, 0.02, 0.01, 0.03, 0.09, 0.25, 0.24, 0.20, 0.04, 0.02, 0.04, 0.01, 0.06, 0.03, 0]
Buck	1	[0.09, 0.90, 0.01, 0, 0, 0, 0, 0, 0, 0, 0.68, 0.18, 0.07, 0.04, 0.02, 0.01, 0, 0, 0, 0, 0.52, 0.08, 0.06, 0.30, 0.02, 0, 0.01, 0.01, 0.01, 0, 0]
Flue	1	[0.02, 0.55, 0.36, 0.07, 0, 0, 0, 0, 0, 0, 0.29, 0.20, 0.15, 0.12, 0.10, 0.07, 0.03, 0.02, 0.01, 0.02, 0.13, 0.11, 0.09, 0.56, 0.04, 0, 0.02, 0.02, 0.01, 0.01, 0]
	2	[0, 0.01, 0.11, 0.31, 0.33, 0.16, 0.07, 0.02, 0, 0, 0.06, 0.11, 0.15, 0.18, 0.19, 0.13, 0.07, 0.04, 0.02, 0.07, 0.01, 0.04, 0.24, 0.20, 0.08, 0.03, 0.05, 0.01, 0.18, 0.16, 0.01]
Chopf	1	[0.04, 0.46, 0.30, 0.17, 0.04, 0, 0, 0, 0, 0, 0.25, 0.17, 0.15, 0.13, 0.12, 0.08, 0.05, 0.02, 0.01, 0.03, 0.09, 0.07, 0.08, 0.62, 0.08, 0.01, 0.02, 0.01, 0.01, 0.02, 0]
	2	[0, 0, 0.04, 0.21, 0.39, 0.26, 0.08, 0.01, 0, 0, 0.05, 0.10, 0.15, 0.18, 0.19, 0.13, 0.07, 0.04, 0.02, 0.07, 0, 0, 0.32, 0.08, 0.05, 0.07, 0.03, 0, 0.25, 0.19, 0.01]
Bool	1	[0.09, 0.86, 0.03, 0, 0.01, 0.01, 0, 0, 0, 0, 0.64, 0.19, 0.08, 0.04, 0.03, 0.02, 0, 0, 0, 0, 0.50, 0.17, 0.08, 0.20, 0.01, 0, 0.02, 0.01, 0.01, 0, 0]

Class	Cluster	Cluster Center
Chapf	1	[0.06, 0.92, 0.02, 0, 0, 0, 0, 0, 0, 0, 0.54, 0.22, 0.10, 0.06, 0.04, 0.02, 0.01, 0, 0, 0, 0.36, 0.17, 0.10, 0.33, 0.02, 0, 0.02, 0, 0, 0, 0]
	2	[0, 0.12, 0.56, 0.16, 0.09, 0.06, 0, 0, 0, 0, 0.15, 0.20, 0.19, 0.16, 0.13, 0.08, 0.04, 0.02, 0.01, 0.03, 0.05, 0.25, 0.18, 0.32, 0.05, 0.01, 0.04, 0.01, 0.05, 0.04, 0]
Gubel	1	[0.01, 0.94, 0.05, 0, 0, 0, 0, 0, 0, 0, 0.60, 0.20, 0.09, 0.05, 0.03, 0.01, 0.01, 0, 0, 0, 0.46, 0.17, 0.12, 0.20, 0.01, 0, 0.02, 0.01, 0, 0, 0]
	2	[0, 0.14, 0.50, 0.26, 0.09, 0.01, 0, 0, 0, 0, 0.13, 0.17, 0.17, 0.16, 0.15, 0.10, 0.05, 0.03, 0.01, 0.03, 0.07, 0.24, 0.21, 0.35, 0.03, 0.01, 0.05, 0.01, 0.02, 0.01, 0]
Gütsch	1	[0.03, 0.93, 0.04, 0, 0, 0, 0, 0, 0, 0, 0.52, 0.23, 0.11, 0.06, 0.04, 0.02, 0.01, 0, 0, 0, 0.29, 0.17, 0.09, 0.38, 0.03, 0, 0.02, 0.01, 0, 0, 0]
	2	[0, 0.07, 0.37, 0.28, 0.20, 0.07, 0.01, 0, 0, 0, 0.14, 0.17, 0.19, 0.17, 0.14, 0.09, 0.04, 0.02, 0.01, 0.03, 0.03, 0.15, 0.33, 0.25, 0.05, 0, 0.04, 0.01, 0.07, 0.06, 0]
Hoger	1	[0.02, 0.77, 0.21, 0, 0, 0, 0, 0, 0, 0, 0.53, 0.21, 0.11, 0.07, 0.04, 0.02, 0.01, 0.01, 0, 0, 0.41, 0.16, 0.15, 0.23, 0.01, 0, 0.02, 0, 0, 0]
Nolle	1	[0.06, 0.51, 0.28, 0.14, 0.01, 0, 0, 0, 0, 0, 0.34, 0.18, 0.13, 0.11, 0.09, 0.06, 0.03, 0.01, 0.01, 0.03, 0.20, 0.17, 0.10, 0.41, 0.05, 0.01, 0.02, 0.01, 0.01, 0.01, 0]
	2	[0, 0, 0.06, 0.18, 0.33, 0.25, 0.12, 0.03, 0.03, 0.01, 0.06, 0.10, 0.14, 0.17, 0.18, 0.12, 0.07, 0.03, 0.01, 0.01, 0, 0, 0.30, 0.05, 0.03, 0.05, 0.02, 0.02, 0.22, 0.27, 0.04]
Tschugge	1	[0, 0.10, 0.23, 0.33, 0.29, 0.06, 0, 0, 0, 0, 0.12, 0.12, 0.14, 0.17, 0.17, 0.13, 0.06, 0.03, 0.02, 0.05, 0.09, 0.16, 0.10, 0.40, 0.10, 0.01, 0.06, 0, 0.04, 0.03, 0]
	2	[0, 0, 0, 0.02, 0.29, 0.42, 0.19, 0.06, 0.02, 0, 0.04, 0.12, 0.17, 0.20, 0.19, 0.12, 0.05, 0.03, 0.02, 0.05, 0, 0.02, 0.38, 0.03, 0.01, 0.04, 0.03, 0.01, 0.24, 0.22, 0.01]

Class	Cluster	Cluster Center
Cima	1	[0.10, 0.27, 0.37, 0.19, 0.07, 0.01, 0, 0, 0, 0, 0.11, 0.07, 0.10, 0.14, 0.18, 0.15, 0.08, 0.05, 0.03, 0.09, 0.09, 0.07, 0.06, 0.57, 0.09, 0.04, 0.03, 0.01, 0.03, 0.02, 0]
	2	[0, 0, 0.01, 0.13, 0.39, 0.42. 0.05, 0, 0, 0, 0.02, 0.05, 0.10, 0.15, 0.21, 0.17, 0.09, 0.05, 0.03, 0.12, 0, 0, 0.12, 0.02, 0.01, 0.08, 0.01, 0, 0.43, 0.31, 0]
	3	[0, 0, 0, 0, 0.06, 0.32, 0.46, 0.16, 0, 0, 0.02, 0.07, 0.12, 0.17, 0.20, 0.13, 0.08, 0.05, 0.03, 0.13, 0, 0, 0, 0, 0, 0, 0, 0, 0.13, 0.82, 0.05]
Monte	1	[0.22, 0.44, 0.28, 0.06, 0, 0, 0, 0, 0, 0, 0.15, 0.07, 0.10, 0.14, 0.18, 0.14, 0.08, 0.04, 0.03, 0.07, 0.10, 0.04, 0.02, 0.75, 0.04, 0, 0.02, 0.01, 0.01, 0.01, 0]
	2	[0, 0.07, 0.35, 0.36, 0.17, 0.04, 0.01, 0, 0, 0, 0.04, 0.07, 0.11, 0.16, 0.21, 0.17, 0.09, 0.04, 0.03, 0.08, 0.04, 0.08, 0.05, 0.55, 0.08, 0.04, 0.05, 0.01, 0.06, 0.04, 0]
Pizzo	1	[0.01, 0.01, 0.01, 0.06, 0.20, 0.37, 0.28, 0.06, 0, 0, 0.04, 0.07, 0.12, 0.16, 0.19, 0.14, 0.08, 0.05, 0.03, 0.12, 0, 0, 0.05, 0.03, 0.01, 0.03, 0.01, 0, 0.27, 0.58, 0.02]
Berg (dim)	1	[0.02, 0.89, 0.09, 0, 0, 0, 0, 0, 0, 0, 0.47, 0.21, 0.13, 0.08, 0.05, 0.03, 0.01, 0.01, 0, 0, 0.39, 0.19, 0.13, 0.23, 0.01, 0, 0.04, 0.01, 0, 0, 0]
	2	[0, 0.07, 0.29, 0.28, 0.22, 0.10, 0.03, 0, 0, 0, 0.09, 0.14, 0.17, 0.18, 0.16, 0.11, 0.05, 0.03, 0.02, 0.05, 0.03, 0.17, 0.32, 0.24, 0.05, 0.01, 0.04, 0.01, 0.06, 0.06, 0]
Büel (dim)	1	[0.03, 0.94, 0.04, 0, 0, 0, 0, 0, 0, 0, 0.60, 0.20, 0.09, 0.05, 0.03, 0.02, 0.01, 0, 0, 0, 0.52, 0.15, 0.11, 0.17, 0.01, 0, 0.02, 0.01, 0, 0, 0]
	2	[0, 0.11, 0.49, 0.22, 0.13, 0.04, 0, 0, 0, 0, 0.19, 0.19, 0.17, 0.15, 0.12, 0.08, 0.04, 0.02, 0.01, 0.03, 0.14, 0.35, 0.16, 0.21, 0.03, 0.01, 0.05, 0.01, 0.03, 0.01, 0]

Class	Cluster	Cluster Center
Chopf (dim)	1	[0.12, 0.86, 0.02, 0, 0, 0, 0, 0, 0, 0, 0.44, 0.27, 0.14, 0.08, 0.04, 0.02, 0.01, 0, 0, 0, 0.19, 0.09, 0.06, 0.59, 0.03, 0, 0.02, 0.02, 0, 0, 0]
	2	[0, 0.06, 0.29, 0.21, 0.22, 0.15, 0.06, 0, 0, 0, 0.05, 0.12, 0.19, 0.21, 0.19, 0.11, 0.05, 0.03, 0.02, 0.04, 0.01, 0.01, 0.36, 0.31, 0.05, 0.02, 0.02, 0, 0.09, 0.11, 0.02]
Flue (dim)	1	[0.03, 0.89, 0.08, 0, 0, 0, 0, 0, 0, 0, 0.42, 0.25, 0.14, 0.09, 0.06, 0.03, 0.01, 0.01, 0, 0.01, 0.30, 0.15, 0.08, 0.37, 0.03, 0, 0.03, 0.04, 0, 0, 0]
	2	[0, 0.14, 0.53, 0.22, 0.06, 0.05, 0.01, 0, 0, 0, 0.14, 0.18, 0.19, 0.16, 0.14, 0.09, 0.04, 0.02, 0.01, 0.03, 0.06, 0.20, 0.17, 0.39, 0.05, 0, 0.04, 0, 0.02, 0.05, 0]
Horn (dim)	1	[0.04, 0.88, 0.08, 0, 0, 0, 0, 0, 0, 0, 0.55, 0.21, 0.10, 0.06, 0.04, 0.02, 0.01, 0, 0, 0.01, 0.40, 0.13, 0.06, 0.30, 0.02, 0, 0.02, 0.06, 0.01, 0, 0]
	2	[0, 0.01, 0.13, 0.24, 0.28, 0.21, 0.10, 0.02, 0.01, 0, 0.05, 0.11, 0.15, 0.18, 0.19, 0.13, 0.07, 0.04, 0.02, 0.07, 0.01, 0.05, 0.25, 0.15, 0.04, 0.02, 0.03, 0, 0.19, 0.23, 0.01]
Stock (dim)	1	[0.05, 0.58, 0.27, 0.09, 0, 0, 0, 0, 0, 0, 0.35, 0.22, 0.15, 0.10, 0.08, 0.05, 0.02, 0.01, 0.01, 0.02, 0.30, 0.21, 0.14, 0.27, 0.03, 0, 0.03, 0.01, 0, 0, 0]
	2	[0, 0, 0.02, 0.17, 0.30, 0.28, 0.19, 0.04, 0, 0, 0.04, 0.10, 0.05, 0.18, 0.19, 0.12, 0.06, 0.04, 0.03, 0.09, 0, 0, 0.26, 0.02, 0.02, 0.01, 0, 0.21, 0.41, 0.04]
Matt	1	[0.09, 0.90, 0.02, 0, 0, 0, 0, 0, 0, 0, 0.65, 0.19, 0.08, 0.04, 0.02, 0.01, 0, 0, 0, 0, 0.62, 0.12, 0.10, 0.12, 0.01, 0, 0.02, 0.02, 0, 0, 0]
	2	[0.01, 0.17, 0.41, 0.23, 0.13, 0.04, 0, 0, 0, 0, 0.18, 0.19, 0.17, 0.15, 0.13, 0.08, 0.04, 0.02, 0.01, 0.03, 0.16, 0.33, 0.17, 0.20, 0.03, 0.01, 0.05, 0.01, 0.03, 0.01, 0]
Moos	1	[0.02, 0.96, 0.02, 0, 0, 0, 0, 0, 0, 0, 0.72, 0.16, 0.06, 0.03, 0.01, 0.01, 0, 0, 0, 0, 0.54, 0.14, 0.08, 0.20, 0.01, 0, 0.01, 0.01, 0.01, 0, 0]
	2	[0.01, 0.09, 0.61, 0.20, 0.06, 0.02, 0, 0, 0, 0, 0.22, 0.23, 0.19, 0.14, 0.10, 0.06, 0.03, 0.01, 0.01, 0.01, 0.10, 0.33, 0.18, 0.27, 0.04, 0, 0.03, 0.01, 0.03, 0.01, 0]

Class	Cluster	Cluster Center
Riet	1	[0.03, 0.95, 0.01, 0, 0, 0, 0, 0, 0, 0, 0.69, 0.16, 0.06, 0.04, 0.02, 0.01, 0.01, 0, 0, 0, 0.48, 0.18, 0.07, 0.18, 0.01, 0, 0.02, 0.02, 0.04, 0, 0]
	2	[0, 0.16, 0.43, 0.26, 0.13, 0.03, 0, 0, 0, 0, 0.17, 0.22, 0.20, 0.15, 0.12, 0.07, 0.03, 0.02, 0.01, 0, 0.06, 0.29, 0.19, 0.30, 0.05, 0.01, 0.04, 0.01, 0.04, 0, 0]
Wang	1	[0.05, 0.88, 0.07, 0, 0, 0, 0, 0, 0, 0, 0.55, 0.21, 0.10, 0.06, 0.04, 0.02, 0.01, 0, 0, 0, 0.46, 0.13, 0.09, 0.26, 0.02, 0, 0.03, 0.01, 0.01, 0, 0]
	2	[0, 0.03, 0.18, 0.40, 0.32, 0.07, 0, 0, 0, 0, 0.06, 0.10, 0.14, 0.16, 0.18, 0.14, 0.08, 0.04, 0.02, 0.07, 0.03, 0.09, 0.19, 0.30, 0.08, 0.08, 0.06, 0.01, 0.12, 0.05, 0]
	3	[0, 0, 0, 0.06, 0.33, 0.40, 0.17, 0.03, 0, 0, 0.04, 0.09, 0.14, 0.18, 0.19, 0.14, 0.07, 0.04, 0.03, 0.08, 0.01, 0.03, 0.39, 0.02, 0.01, 0.04, 0.02, 0.01, 0.27, 0.21, 0]
Wase	1	[0.14, 0.84, 0.02, 0, 0, 0, 0, 0, 0, 0, 0.63, 0.19, 0.08, 0.05, 0.03, 0.01, 0, 0, 0, 0, 0.55, 0.09, 0.05, 0.26, 0.04, 0.02, 0.05, 0.02, 0.07, 0.06, 0]
	2	[0, 0.09, 0.31, 0.20, 0.22, 0.12, 0.06, 0, 0, 0, 0.13, 0.16, 0.17, 0.16, 0.14, 0.10, 0.05, 0.03, 0.02, 0.04, 0.07, 0.17, 0.26, 0.24, 0.04, 0.02, 0.05, 0.02, 0.07, 0.06, 0]
Wise	1	[0.04, 0.96, 0, 0, 0, 0, 0, 0, 0, 0, 0.74, 0.15, 0.05, 0.03, 0.02, 0.01, 0, 0, 0, 0, 0.56, 0.18, 0.08, 0.14, 0.01, 0, 0.02, 0.01, 0.01, 0, 0]
	2	[0, 0.23, 0.40, 0.21, 0.13, 0.03, 0, 0, 0, 0, 0.16, 0.21, 0.19, 0.16, 0.12, 0.08, 0.04, 0.02, 0.01, 0.02, 0.10, 0.34, 0.19, 0.23, 0.05, 0.01, 0.04, 0.01, 0.02, 0.01, 0]

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.

Songlin Wang

Songlin Wang
Zürich, 30.04.2024