

The Meanings of the Generic Parts of Toponyms: Use and Limitations of Gazetteers in Studies of Landscape Terms

Curdin Derungs¹, Flurina Wartmann¹, Ross S. Purves¹, and David M. Mark^{2*}

¹ Geocomputation Unit, Geographic Institute, University of Zurich, Switzerland
{curdin.derungs, flurina.wartmann, ross.purves}@geo.uzh.ch

² Department of Geography, University at Buffalo, NY, USA

* Director, National Center for Geographic Information and Analysis (NCGIA), University at Buffalo, NY, USA
dmark@buffalo.edu

Abstract. Are the contents of toponyms meaningless, as it is often claimed in linguistic literature, or can the generic parts in toponyms, such as hill in *Black Hill*, be used to infer landscape descriptions? We investigate this question by, firstly, linking gazetteer data with topographic characteristics, and, secondly, by conducting analysis of how the use of landscape terms might have changed over time in a historic corpus. We thus aim at answering a linguistic, and ethnophysiographic, research question through digital input data and processing. Our study area is Switzerland and our main focus is on geographic eminences, and in particular on the use of the terms *Spitze*, *Horn* and *Berg*. We show that most prominent generic parts in toponyms show expected topographic characteristics. However, not all generic parts strictly follow this rule, as in the case of *Berg*. Some generic parts have lost their meaning in standard language over time (e.g. *Horn*). We therefore put a cautionary note on the use of generic parts in toponyms in landscape studies, but point out that the subtle details of these differences provide rich topics for future research.

Keywords: toponyms, proper names, generic parts, landscape terms, GIR, ethnophysiography, gazetteers

1 Introduction

Recently, research has accelerated on the topic of how landscape and its elements are conceptualized, on how these conceptualizations are expressed in human natural language, and on how and why the conceptualizations vary across languages and cultures. Mark and Turk coined the term ethnophysiography for this topic (Mark and Turk, 2003a, 2003b). A major development was the publication of a special issue of *adfa*, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

Language Sciences (Burenhult, 2008), followed by a book entitled “Landscape in Language” (Mark et al., 2011). Burenhult and Levinson listed some key research questions in this domain:

How are landscape features selected as nameable objects (‘river’, ‘mountain’, ‘cliff’)? 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 and Levinson, 2008, p. 136)

The detailed research agenda, and the potential of finding fundamental explanations for spatial particularities in conceptualizations, are a key strength of ethnophysiography. On the other hand, ethnophysiography up to now has concentrated on small samples of speakers, usually through ethnographic interviews or focus group discussions, and thus data collection is thus very time intensive.

In this context, Geographic Information Retrieval (GIR) can be considered another line of research. GIR is usually defined as a combination of methodologies from Geographic Information Science and Information Retrieval, where unstructured data, often in the form of natural language, is parsed for geographic information. Jones and Purves list some of the following tasks as being most relevant in GIR:

Detecting geographical references in the form of place names and associated spatial natural language qualifiers within text documents and in users’ queries; disambiguating place names to determine which particular instance of a name is intended; geometric interpretation of the meaning of vague place names, such as the ‘Midlands’ and of vague spatial language such as ‘near’; (Jones and Purves 2008, p. 220)

The two fields of research, namely GIR and ethnophysiography, can be usefully combined by using digitized input data to explore ethnophysiographic research questions. In this paper we analyze the use of landscape terms as parts of Swiss toponyms, as in for example the generic landscape term *Berg* (mountain) in the toponym *Uetliberg*, and their association with the use of landscape terms in everyday language. By incorporating data from gazetteers and digitized language corpora we aim to contribute to the ethnophysiographic research question:

What is the relation between landscape terms (common nouns) and place names (proper nouns)? (as outlined in Burenhult and Levinson 2008, p. 136)

We first introduce relevant theoretical background on proper names from different disciplinary angles including psychology, linguistics, semantics, and geography. We then provide an overview of the methods used. We investigated properties of generic parts in toponyms through a topographic analysis, i.e. geomorphometry (e.g. Pike et al. 2008), and explored change over time in the use of landscape terms in a digital text corpus and user generated content (UGC) (Goodchild 2007) through textual analysis and association of generic parts with regions.

In our study we therefore develop a framework that allows us to explore ways in which toponyms are used. In so doing, we demonstrate that caution is needed when using gazetteers as a corpus and, in particular, when using proper names as sources of generic information to describe landscapes. Moreover, we establish different approaches that exploit digital input data, automatic data processing, and thus, quantitative evaluations, in order to explore ethnophysiographic research questions, which so far have been primarily addressed by empirical ethnographic studies in the field. We argue that this combination of approaches has considerable potential, especially if we wish to broaden the spatial scope of ethnophysiographic research.

2 Theoretical Background

Domains of Proper Names. Proper names are used to refer to individual entities mainly in two domains of human experience: people, and places. It is important for our purposes to note that toponyms, also known as place names, are one of the most important subclasses of proper nouns or proper names. Recently, Levinson (2011) has pointed out that people and places, the main referents for proper nouns, are also two domains with specialized processing areas in the brain (faces and places).

The Psychology of Proper Names. Valentine et al. (1996) reviewed cognitive and psychological differences in how people process proper names, compared with how people process common nouns and noun-phrases. However, few rigorous psychology experiments with geographic proper names have actually been conducted. One exception is the work by Hollis and Valentine (2001) on effects of priming and associations between pictures and names, and between auditory priming and visual recognition of names. They found that the priming improved recognition for people names and landmark names (e.g. Levinson 2011), but not for country names which were processed in a way similar to common nouns.

In general proper names are often said to be pure referencing expressions, and therefore not to have semantic meaning or sense. Yet, as Hollis and Valentine point out, toponyms, at least in English, often have semantically-meaningful components, that is, parts of the names that at least *appear to make sense*:

Landmark names often contain a greater degree of meaning compared with people's names and country names that can be considered arbitrary. For example, the Eiffel Tower is a tower and Tower Bridge is a bridge, next to the Tower of London. (Hollis and Valentine 2001 p, 113)

However, as an example, of the 157 toponyms called “Mount Pleasant” in the US Geological Survey’s GNIS database (USGS, 2013), 76 percent are populated places or locales, rather than being topographic eminences¹². The sense in toponyms thus

¹ In this paper, we use the term *eminence* in its older landscape sense, as a superordinate term for a part of the landscape that stands above its surroundings.

appears to be neither absolutely true or absolutely false, but rather partial or probabilistic, to vary with circumstances, and to be different for different languages and cultures. Coates' (2006) article on properhood provides a basis for understanding this.

Properhood. Coates (2006) provides a very detailed account of how “properhood” works in language. Coates reviews previous efforts to define and distinguish proper names, mainly by philosophers and linguists, and to show how they differ from ordinary language. He then provides a compelling alternative theoretical treatment. Coates concludes that proper nouns are not a lexical category (type of noun), but rather are “a type of referring that discounts the sense of any lexical items (real or apparent) in the expression that is being used to do the referring” (Coates 2006 p. 378). Coates notes that the view that proper names have “no sense in the act of reference” does not exclude the possibility that they have senses that are “accessible during other (meta)linguistic activities” (Coates, 2006, p. 356). Conventions for understanding proper names license “probabilistic implications, nothing logically stronger, even though the probability in a given case could be extremely high—the implied categorization should always be taken as falsifiable in principle even if not yet falsified” (Coates 2006, p. 365).

Toponyms. Toponyms have referential meaning, that is, they denote individual geographic features. Toponyms are often compiled to form gazetteers, typically by government agencies, but sometimes as projects within volunteered geographic information (VGI), a current ‘hot’ topic in Geographic Information Science (Goodchild 2007).

Since, following from Coates’ theory of properhood, the non-referential ‘content’ of every toponym is to some extent ‘encapsulated’, that is, it may or may not “make sense”, or follow Grice’s maxims (Grice 1975), it should not be a surprise that the generic parts of toponyms, if present, may be from a different vocabulary than the general landscape terms of a language. This is an important point, since many researchers (including one of the present authors, e.g. Feng and Mark 2012) have uncritically treated lists of toponyms as corpora for studying general landscape terms. For the English language, the possibility that toponyms are not “normal language” is masked by the fact that, except for names of populated places, the generic parts of many toponyms in English involve the same terms as those that are used as landscape terms to denote classes of landscape features (e.g. hill, valley, lake, etc.).

One illuminating exception for natural landscape features in English is the word ‘Mount’, which is common in toponyms denoting mountains or other eminences, but

² We leave it to the reader to judge which of these places are “pleasant”!

which in contemporary English would not be used as a common noun outside of toponyms (i.e., “*I live near a mount”; “*what is the name of that mount?”)³.

2.1 Case Study of Toponyms for Mountains in Swiss German

Motivation. This research project began with the observation that equations of the form:

“the Eiffel Tower” is a “tower”

are not valid for many names of mountains in German, such as *Zugspitze* or *Matterhorn*. The following sentences, relating specific peaks to the generic parts of their names (*Spitze* and *Horn*⁴):

“*Die Zugspitze ist eine Spitze” or “*Das Matterhorn ist ein Horn“

are not well-formed sentences in standard German, because *Spitze* and *Horn* on their own refer to parts, but not whole, topographic features. However, *Spitze* and *Horn* are commonly used nouns in standard German, in contrast to *Mount*, *-ville* or *-ton*, popular generic terms in English and American toponyms, which do not form useful nouns in their own right. Furthermore, the meanings of the words in standard German suggest proximate association with topographic features, for instance in referring to topographic shape.

Coaz (1865) pragmatically explains the issue of *Horn* and other generic parts in toponyms in an article on place names in the Swiss Alps (“Ueber Ortsbenennung in den Schweizeralpen”) published by the Swiss Alpine Club.

From the canton of Bern, to the borders of the Canton of Vaud and the Oberwallis, most mountain tops have the generic name “Horn”, and, in fact, this part of the Alps is characterized by numerous pyramidal, rocky mountain tops. (Coaz 1865, p. 467, translated from German by the authors)

Coaz emphasizes that toponyms containing the generic part *Horn* belong to the feature class *Berg* (mountain), and that *Horn* is related to the summit of the mountain. This could also be true for *Spitze*, since the sentence “Die Spitze des Berges” makes perfect sense for German speakers. However, this would indicate that *Horn* and *Spitze* only refer to parts of a mountain, whereas the toponym, as for instance “Matterhorn”, certainly refers to the mountain as a whole. In the following, we will look at this divergence more closely.

³ Note: Following the practice in linguistics, an asterisk [*] is placed in front of an expression to indicate that it would not be considered to be a valid or well-formed utterance by speakers of the language

⁴ The primary meaning of *Horn* in German is the same as in English, namely a projection on the head of an animal, while the primary meaning of *Spitze* is similar to the meaning of the English word *point*, such as in “pencil point”.

In standard German, the association of *Spitze* and *Horn*, with *Berg*, is reflected by the high acceptance of the sentence

“Die Zugspitze und das Matterhorn sind Berge“

(e.g. de.wikipedia.org/wiki/Zugspitze). However, and this makes the issue much more complex, *Berg* is also prominently used as a generic part in Swiss toponyms. Nonetheless, the sentence:

“Zürichberg, Hönningerberg, Chäferberg und Uetliberg sind Berge“

would probably not be accepted by locally knowledgeable, native speakers. All four toponyms, although referring to geographic eminences, would not be considered as mountains, resembling the Matterhorn or Zugspitze, but would rather be classified as *Hügel* (hill), as is indeed the case in the feature classification of the gazetteer used in this study.

Problem Statement and Research Questions. From a small set of specific examples, we find the relation between landscape terms and toponyms, and generic parts in toponyms in particular, to be of a complex nature. *Spitze* and *Horn* are frequently used as generic parts in toponyms, but they do not seem to represent independent geographic feature types. Furthermore, these two generic parts are not productively used in language to refer to features that have these generics in their names. Instead, such features would rather be called *Berg*. However, a certain association of *Spitze* and *Horn* with geographic eminences is suggested by the use of the words in standard German. Derungs and Purves (2007), for instance, found evidence that *Spitze* was often associated with *Berg*. *Berg*, on the other hand, refers to a basic level geographic eminence type in standard German, but its occurrence as a generic part in toponyms appears to have little topographic overlap with prototypical mountains, such as the *Zugspitze* or *Matterhorn*. Toponyms containing *Berg* as a generic part seem to refer to geographic features that would not be called *Berg* in standard language.

From these observations three research questions emerge:

- Can generic parts in toponyms be associated with particular topographies?
- Are generic parts in toponyms fossilized, meaning that the productive use in standard German has changed over time?
- Is the use of generic parts in Swiss toponyms arbitrary, such that there is no meaning within toponyms, or, can we, for instance, deduce spatial relations or settings?

Topographic and Corpus based Study. To shed light on the relation between generic parts in Swiss mountain toponyms and their associated forms, we will link the locations of toponyms which contain generic parts to topographic characteristics, thus allowing comparison of topographic similarity between different classes of generic parts as used in toponyms.

In a corpus based study, we investigate the apparent mismatch of concepts between the generic part *Berg* and its underlying referent, as well as the use of *Horn* and *Spitze*, by analyzing whether the use and meanings of these terms has changed over 1975time. Our hypothesis is that change is of a different nature for *Horn/Spitze* than it is for *Berg*. In the case of *Horn* and *Spitze*, we suppose that these terms were productively used as category norms for geographic eminences in the past, and that they have become fossilized over time and are nowadays no longer productively used in standard language. Thus, the use of the generic parts has changed over time. In the case of *Berg* on the other hand we assume that the meaning itself has changed and that over time, different types of topographies have been referenced with the same generic part in their toponyms.

If, in both cases, change over time is not a useful explanatory variable, this indicates that, at least in the Swiss context, and examples from other European languages point in the same direction, generic parts in toponyms do not necessarily have full normal meaning.

In the following sections of this paper we present the set of input data and methods that we used for comparing topographic characteristics of generic parts in toponyms. Secondly, we report on the analysis on change of meaning over time, by using a large historic corpus on Swiss alpine landscape descriptions. We then present the results for these two empirical case studies and discuss them in view of the three research questions outlined above.

3 Methods

3.1 Topographic characteristics

Input Data. The input data consists of a digital elevation model (DEM) and a gazetteer. The digital elevation model we used in our analysis has a spatial resolution of 25m and its spatial coverage includes Switzerland in its entirety. Vertical accuracy in the Swiss Alps varies between 3 and 8 meters. This DEM is provided by the Swiss Federal Office of Topography Swisstopo (<http://www.swisstopo.admin.ch/>).

As a gazetteer we use “Swissnames”, the most extensive compilation of Swiss toponym data that provided by Swisstopo. It consists of more than 190,000 individual toponyms and the geographic coordinates of the locations that are associated with them. The georeferencing of toponyms reflects the location where the respective label is drawn on topographic maps from the National Mapping series (scales 1:25,000 and 1:50,000). Swissnames contains a feature classification that differentiates toponyms into types of geographic features. Roughly half of all toponyms refer to geographic features of natural feature type, with numbers of toponyms for different types of geographic eminences listed in Table 1. We also added separate columns for the generic parts *Berg*, *Spitze* and *Horn* and indicated their relative frequencies.

Table 1. Count of different types of geographic eminences in Swissnames and relative frequencies of *Berg*, *Spitze* and *Horn* as generic parts distributed over feature types.

Feature type	Description/ Example	Count	<i>Berg</i>	<i>Spitze</i>	<i>Horn</i>
Huegel	hill, e.g. <i>Uetliberg</i> , <i>Hönggerberg</i>	2543	81	16	4
KGipfel	small mountain peak, e.g. <i>Lochberg</i> , <i>Märensitz</i>	4414	17	68	72
GGipfel	big mountain peak, e.g. <i>Allalinhorn</i> , <i>Breithorn</i>	866	1	14	19
HGipfel	main mountain peak, e.g. <i>Matterhorn</i> , <i>Weisshorn</i>	165	0	1	5
Massiv	mountain range, e.g. <i>Engelhörner</i> , <i>Churfirten</i>	143	0	0	0

Analysis. By considering values of elevation and slope from a set of three buffer zones (200m, 400m and 2000m) around each toponym location, we make a simple association between toponym location and topographic properties. From the distribution of elevations within each buffer zone we store relief (the maximum difference between the elevation of two raster cells within the buffer zone) and standard deviation in elevation (which is related to surface roughness). From the distribution of slopes we retain mean slope and standard deviation. Since both types of measurements are computed for all three buffer zones, we generate 12 attributes that represent the topographic characteristics for each toponym location. The three buffer sizes can be seen as a very simplistic form of multi-scale analysis, which could be extended by, for example, identifying and generating term vectors of morphometric classes (e.g. pit, peak, pass, etc.) for each buffer at multiple scales (e.g. Fisher et al. 2004).

The topographic characteristics of neighborhoods of toponyms can thus be represented as a vector of features. Similarity between toponyms can then be compared either quantitatively, for example using a measure such as cosine similarity (e.g. Bayardo et al. 2007), or qualitatively, by reducing the multi-dimensional vector space to a two dimensional map (a “Self Organizing Map”, or SOM) (Mark et al. 2001; Skupin and Esperbé 2011). SOMs are a form of artificial neural networks, where training data is used to reduce the multidimensionality of feature vectors through a mapping to two dimensional space, whilst preserving the similarity between feature vectors as neighborhood relationships (Kohonen 1995).

To compute SOMs we used the *kohonen* package, as implemented in R (Wehrens and Buydens 2007). We selected, additionally to *Berg*, *Spitze* and *Horn*, sets of semantically comparable generic landscape terms from toponyms found in Swissnames that either have comparable meaning in different languages (*Pass*, *fuorcla* and *col*, all meaning pass in German, Rumantsch and French respectively or

champs and *Acker* meaning agricultural field in French and German respectively) and generics with differing spelling across dialects (*Acker*, *Acher*), and extracted feature vectors for all toponyms containing these generic parts.

3.2 Change of the meaning of generic parts in toponyms over time

Input Data. Historic uses of terms related to landscape features are investigated using Text+Berg (Volk et al. 2009), a large corpus (collection) of Swiss alpine landscape descriptions, published as yearbooks by the Swiss Alpine Club and covering the last 150 years. These books have been digitised and linguistically parsed (Sennrich et al. 2009), such that they can be automatically queried for the occurrence of particular terms. The descriptions cover all official Swiss languages, with a majority of the articles written in standard German. Articles cover a wide variety of topics ranging from scientific reports on the state of Swiss glaciers to reviews of newly introduced outdoor products. However, most common are reports on mountaineering trips in the Swiss mountains, where landscapes are often described in great detail. The authors of the articles are not necessarily residents from the regions that are being described, but can be mountaineers from other parts of Switzerland and the world.

In order to examine uses of landscape terms and their spatial distributions in a contemporary context, we use georeferenced Flickr photographs (www.flickr.com) that have been uploaded and tagged with different terms by users. For the bounding box of Switzerland we gathered a total number of about 4 million items using the Flickr API. In this case study we focus on analyzing the user-generated tags that were applied to describe the content or context of pictures.

Analysis. We split the analysis in two parts. First we relate the generic parts *Berg*, *Horn* and *Spitze* to regions. Using these initial results we analyze the relatedness of these terms to each other. In a second step, we investigate the occurrences and uses of these terms in a historic corpus.

Spatial footprints. We approximated footprints for generic parts in toponyms by mapping density peaks, calculated from all associated locations (e.g. Hollenstein and Purves 2010). We then compared these footprints with footprints gathered from more recent uses of landscape terms, outside of toponyms, by calculating spatial footprints from georeferenced Flickr photographs tagged with landscape terms, such as *Berg*. *Spitze* and *Horn* occur rarely as generic terms in their own right, and thus we did not generate spatial footprints. By comparing these footprints we approximate the relatedness of these concepts in terms of their spatial overlap.

Occurrences of terms in corpus. In order to explain patterns of relatedness, we compare occurrences of *Spitze* and *Horn* as nouns in early (prior to the year 1870) and recently (after 2000) published yearbooks from the Text+Berg corpus, with a particular focus on change of use over time. Since word sense disambiguation, for

instance in detecting if the term *Spitze* refers to mountain, part of mountain, or pencil point, is challenging, we relied on manual annotation. To detect change in meaning over time in the use of *Berg*, we again analyzed early published yearbooks of the Text+Berg corpus, this time with a particular focus on the types of toponyms that *Berg* is referring to. Through this approach we can investigate the use of generic parts in standard language outside toponyms.

4 Results

4.1 Topographic characteristics

First, we report on the topographic characteristics of the three generic parts *Horn*, *Spitze* and *Berg* and show the vector similarities of within-, and across-group comparisons. We then present the SOM of a set of nine generic parts that frequently occur in Swiss toponyms. These terms do not exclusively represent topographic eminences, but cover a broad selection of landscape terms.

SOM. We analyzed SOMs derived from feature vectors representing the topography of toponyms containing generic parts that frequently ($n > 200$) occur in Swissnames (Fig. 1). These nine generic parts represent very different topographies. The top row of Fig. 1 contains the three generic parts discussed in the introduction. We arranged the comparable generics in horizontal rows so that similarities in their SOM representations are clearly visible. The SOMs of *Spitze* and *Horn* are auto-correlated, and almost congruent, whereas the SOM of *Berg* shows a quite diffuse pattern, which is very different from *Spitze* and *Horn*. Equally, the terms describing passes through mountains, *Pass*, *fuorcla* and *col* are all rather similar and have some overlap with the feature space of *Horn* and *Spitze*. Finally, the generic parts describing cultivated fields (*champs*, *Acker*, *Acher*) occupy a different region of the SOM, interestingly overlapping slightly with the region occupied by *Berg*.

Cosine Similarities. Cosine similarities between all nine generics, computed from topographic vectors, are represented in Fig. 2. The grey boxes delineate groups of generic parts considered semantically similar, such as passes or cultivated fields that were arranged in comparable rows in Fig. 1. In general, generic parts that describe similar features have high cosine values in Fig. 2.

The (self-)similarity of *Horn* and *Spitze* is also reflected by cosine similarities, within and across groups (Fig. 2, eminences box). Toponyms containing *Horn* or *Spitze* have median cosine values of around 0.8, in both, within- and across-group comparisons, indicating topographic relatedness. *Berg* toponyms, on the other hand, are unrelated to other *Berg*, *Horn* or *Spitze* toponyms, manifested by cosine values of around 0.

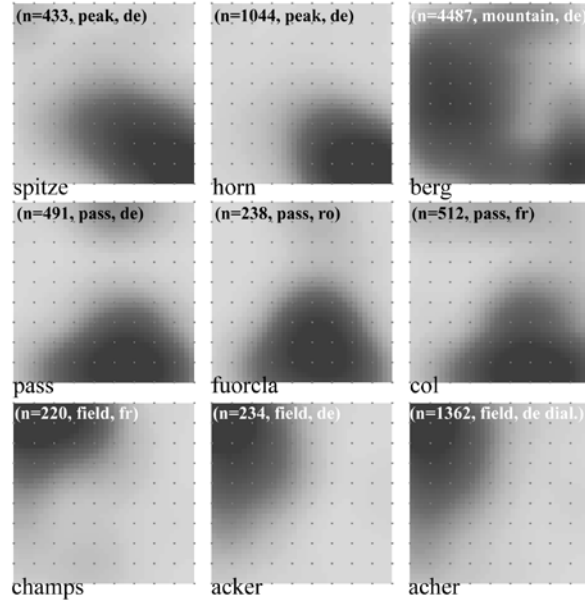


Fig. 1. SOM representation of the topographic vectors for nine frequent generic parts in Switzerland

4.2 Change of the meaning of generic parts in toponyms over time

Spatial footprints. We calculated the spatial footprints for the generic parts *Berg*, *Horn* and *Spitze* within toponyms (Fig. 3). For comparison, we also included the spatial footprints of the use of *Berg* and *mountain* as user-generated tags in georeferenced Flickr photographs. The footprint of the term *Berg*, used as a generic part (Fig. 3 upper right), is markedly different to the use of *Berg* (solid line), and *mountain* (dashed line), as a tag in Flickr (Fig. 3 upper left). The Flickr footprints reflect a spatial distribution of the term *Berg* and *mountain* that overlaps with what we would expect from its use in everyday language and “naïve” geography (Egenhofer and Mark 1995), covering Swiss alpine regions (Valais, the Bernese Oberland, parts of central Switzerland, Toggenburg, the Albula region and Oberengadin). On the other hand, toponyms containing *Berg* (Fig. 3 upper right) are mainly located in the *Swiss Mittelland*, an extensive plain. This does not imply that *Berg* toponyms refer to plains or valley floors: *Berg* toponyms do refer to geographic eminences. However, these features containing the term *Berg* in the Swiss Mittelland are topographically different from prototypical mountains, such as *Matterhorn* or *Zugspitze*.

Occurrences of terms in the corpus. *Spitze* and *Horn* are both prominently used in Text+Berg, with about 5,500 references to *Spitze* and 750 to *Horn*. These terms, however, are semantically ambiguous, with a significantly smaller number of references referring to actual geographic features. *Spitze* is used more often compared

to *Horn*, presumably because *Spitze* is also used to refer to different types of objects having a peaky shape (e.g. pencil points). We found that *Horn*, as a noun occurs more often in older yearbooks. A comparison between the 10 oldest and the 10 most recent yearbooks resulted in 34/mw (i.e. number of occurrence per million words) occurrences in old, compared to 10/mw occurrences in new yearbooks. This relation is even more marked for *Spitze*, with 500/mw in old compared to only 27/mw occurrences in the most recent yearbooks.

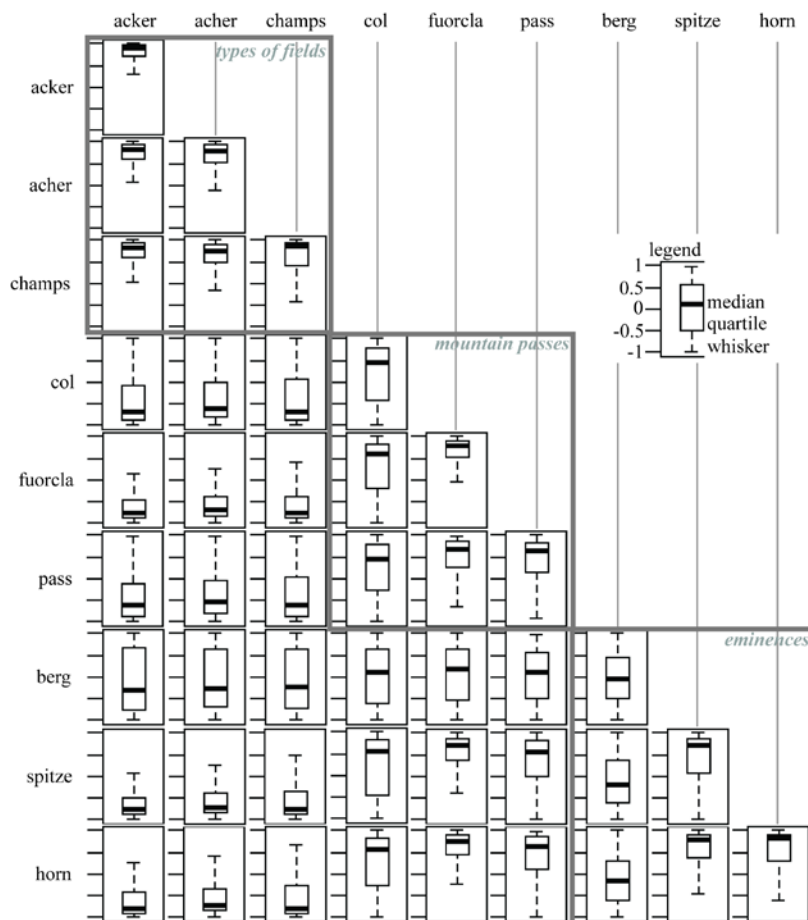


Fig. 2. Cosine similarities of topographic vectors for nine frequent generic parts in Swiss toponyms. Related generic parts are delineated (grey boxes) and labeled.

In a reading of early yearbooks, *Horn* appears to be most often used in reference to mountains, and mostly in reference to whole mountains. However, the mountain toponym does not necessarily need to contain the generic part *Horn*. For instance, in an article from 1865 (Weilenmann 1865), *Piz Rosegg* is referred to as ‘a *Horn* that shrinks, the more one approaches it’. In five yearbooks we found 15 cases where

Horn refers to a whole mountain and only 5 cases where *Horn* refers to the top part of a mountain.

By contrast, in recent yearbooks *Horn* is very rarely used to refer to mountains. In 10 yearbooks we only found three cases where *Horn* is used in this context, where in two cases it was used to reference a whole mountain.

For *Spitze*, we found few occurrences in early yearbooks that refer to whole mountains. The *Dufourspitze*, for instance, is called a *Spitze* in 1864 (Studer 1864), where the author regrets that this *Spitze* was not named after its first ascendant. Indeed, *Spitze* was, and is, prominently used to refer to the top part of a mountain. In a mountaineering context, for instance, it is well accepted to state: “Ich näherte mich der *Spitze*” (I approached the summit). Most interesting are the early occurrences, mainly of *Horn*, but also of *Spitze*, that refer to whole mountains. This suggests that it was acceptable to refer to a mountain by using the word *Horn*, and thus, the sentence:

“Das Matterhorn ist ein Horn“

was at some point in time valid. Although use of *Spitze* has decreased, its meaning in both old and new yearbooks appears to have primarily referred to the summit regions of a mountain.

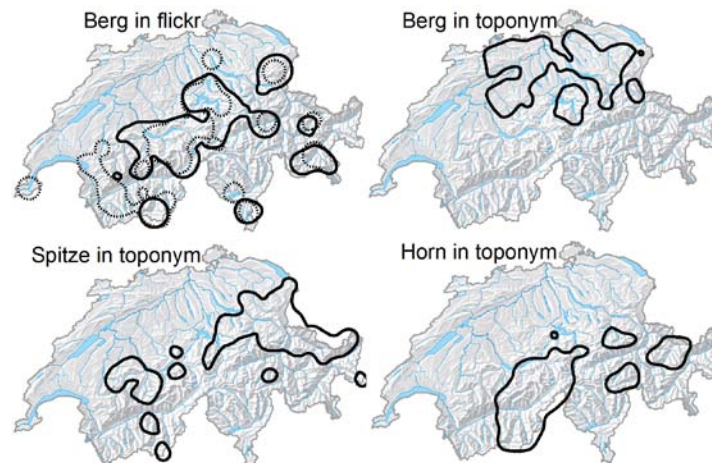


Fig. 3. The density contours enclosing 20% of the volume of Flickr photographs containing the tag *Berg* (upper left, solid line) and mountain (upper left, dashed line), and toponyms that contain the generics *Berg*, *Spitze* and *Horn*.

In the case of *Berg*, we did not detect a change in meaning over time using Text+Berg, such that *Berg* used to refer to topographic eminences in the *Swiss Mittelland* in early yearbooks, but now rather describes an eminence in the *Alps*, such as might be suggested by Fig. 3. In the very first yearbooks we find references to *Berg* where it is used to refer to prototypical mountains, such as the *Finsteraarhorn*, referred to as the “mächtige Berg” (majestic mountain, Lindt 1864), or, the *Aletschhorn*, which is called a “herrlicher Berg” (magnificent mountain, Fellenberg

1864). This indicates that the concept of *Berg* has not significantly changed over the last 150 years, at least in terms of its use in Text+Berg. One explanation for the mismatch between our observations in the corpus and the comparison of gazetteer data with UGC, may be that variations in the use of *Spitze* and *Horn* occurred within the 150 year lifespan of Text+Berg, whereas changes in the use of *Berg* date further back.

We found support for this hypothesis by searching for first occurrences of toponyms in historic documents. For three out of four *Berg* toponyms listed in section §1.2 we found historic references, dating back to the early 15th century or earlier, in a local, very fine grained gazetteer (www.ortsnamen.ch). *Hönggerberg* first occurs in 1432, *Zürichberg* is first mentioned in 1188, and the first reference to *Chäferberg* dates back to 1420. We could not find any historic reference to *Uetliberg*.

5 Discussion

We aimed to investigate an ethnophysiographic research question, namely the relation between toponyms and landscape terms, by exploiting a range of digital input data including gazetteers, text corpora, UGC and elevation models. To explore such data, we used a variety of methods allowing us to characterize regions in terms of geomorphometry, to calculate similarities between generic parts of toponyms and to identify parts of speech for a qualitative reading of a text corpus.

We argue that the combination of approaches used has considerable potential, especially if we wish to broaden the spatial and temporal scope of ethnophysiographic research. In the following, we discuss our findings with respect to the research questions set out in §1.2.

RQ1: Can generic parts in toponyms be associated with particular topographies?

For nine different generic parts that are frequently found in Swiss toponyms we used geomorphometric measurements of toponym locations as represented by relative drop and slope in different buffer zones. We have shown, by reducing feature vectors to two dimensions with SOM, that groups of toponyms with identical generic parts such as *Spitze* or *Horn* exhibit specific topographic patterns (Fig. 1). For instance, the patterns for *Spitze* and *Horn* were self-similar and similar to one another, while both differed from patterns exhibited by toponyms with generic terms related to agricultural fields and mountain passes, which we expect to have correspondingly different toponyms. Interestingly, we observed this pattern even for generic parts in toponyms in different languages (*Pass*, *col* and *fuorcla* for mountain pass in German, French and Rumantsch). Thus, our method can be used to compare landscape terms and their relation to morphometry across languages – thus potentially contributing to generating taxonomies and ontologies (Scheider et al. 2009). Using a quantitative measurement in the form of cosine similarity produced similar results (Fig. 2). These

findings are in accordance with our expectation that semantically similar terms are also topographically similar as measured by their SOMs and cosine similarities.

However, we did find an important exception to the more general pattern described above. The SOM of toponyms containing the generic part *Berg* showed a much more diffuse pattern than other generic parts, which was not similar to any other SOM of the terms investigated (Fig. 1). Furthermore, cosine similarity for *Berg* showed low self-similarity. This indicates that toponyms containing *Berg* include varying topographies that differ from the topographies associated with *Spitze* or *Horn* toponyms. This is, at first glance, puzzling because *Spitze*, *Horn* and *Berg* are all used to refer to topographic eminences (e.g. Zugspitze, Matterhorn, Chäferberg). In the following paragraph, we discuss the difference of the use of generic terms in toponyms as compared to the use of the terms in standard language.

RQ2: Are generic parts in toponyms fossilized, such that the productive use in standard German changed over time?

In order to investigate how generic parts in toponyms such as *Berg*, *Horn* and *Spitze* are spatially distributed in Switzerland we computed separate footprints for the three generic terms from the Swissnames gazetteer. The footprints for *Horn* and *Spitze* show that these terms are both most often used in the Swiss Alps, with slightly different distributions. *Spitze* is used more in Eastern Switzerland, while *Horn* seems to be primarily used in the Valais and the Central Alps. If we follow the argument that generic parts in toponyms have meaning (Hollis and Valentine 2001), this would be an indication that the topography in Eastern Switzerland is different from the topography in the Valais such as that *Spitze* and *Horn* are used to denote features with differing topographies. However, our results from SOM and cosine similarities clearly shown that *Spitze* and *Horn* indeed share similar topographies and both their spatial footprints overlap with mountainous terrain. Thus, we suggest that the spatial differences in the occurrence of toponyms with *Spitze* and *Horn* are probably more directly related to language use in dialect regions of Switzerland. For the generic term *Berg* the spatial footprint covers the *Swiss Mittelland*, but not the Alpine regions in Switzerland. This finding is counter-intuitive, as we would expect that the generic term *Berg* (“mountain”) is also used where the geographic features of mountains exist, i.e. in the Alps, which was not the case. However, this finding is consistent with our previous observations that the cosine similarities for *Berg* are distinct from *Horn* and *Spitze*, and the generic terms containing *Berg* have a more diffuse topography as represented by the SOM. Thus, through the spatial footprints we found a pattern that supports our first results, where *Spitze* and *Horn* are similar topographically, even though they do not have exact overlapping footprints, and *Berg* shows a distinct pattern.

In order to look more closely at this peculiarity of the use of *Berg*, we compared its use as a generic part in toponyms with its use in current standard language. The current use of the term *Berg* is, for example, encapsulated in user-generated and georeferenced Flickr tags. These two footprints of *Berg* show almost no spatial

overlap (Fig. 3 upper row), which strongly suggests that the generic part *Berg* as used in toponyms is different from the use of *Berg* in contemporary standard German. We thus suggest caution when using gazetteers to mine for contemporary uses of landscape terms.

Since we found that current use in standard language differed from use of generic terms, we looked at the historic uses of these terms in a corpus of Swiss Alpine yearbooks. Our results indicate that in a historic context, both *Horn* and *Spitze* were also used to refer to entire mountains. This was unexpected, since such an association of *Horn* and *Spitze* with an entire mountain is not reflected in the current use in standard German where *Spitze* and *Horn* are used as partonyms (names of parts) in reference to mountains. However, in some Swiss dialects, this association appears to still be found, but with our methodological approach, we could not investigate the details of this relationship, which should be examined using questionnaires or ethnographic methods.

For the use of the term *Berg* we could not find any evidence in our historic corpus that indicates a change of meaning of the term over time. On the contrary, many early occurrences of *Berg*, which date back to the late 19th century, refer to features that would still be classified as prototypical mountains. In this case, either the use of *Berg* in toponyms has always been independent from the use of *Berg* outside toponyms, or this change in meaning over time happened earlier than the time that our Text+Berg corpus covers. If we look at first historic references of a set of *Berg* toponyms dating back to the early 15th century, they support the assumption of earlier coining of toponyms with the term *Berg*. However, we currently lack more historic data to compare the use of the generic term and the use in standard language of the term *Berg* at the time these toponyms were coined.

In sum, we found that *Spitze* and *Horn* were indeed once used to refer to whole mountains in standard language as well, which may explain their high frequency as generic parts in toponyms associated with mountains, but that the use of these terms has changed over time. The case for *Berg* is different, since we found no change of meaning in the historic corpus and suggest that if a change has happened, it would date back further, to the time when the toponyms containing *Berg* were coined.

RQ3: Is the use of generic parts in Swiss toponyms arbitrary, such that there is no meaning in toponyms, or, can we, for instance, deduce spatial relations or settings from generic parts?

Do generic parts in toponyms have meanings? Based on our results that we discussed in the context of the first research question we argue that there is a relation of semantically similar groups of generic parts with topographic patterns. This result supports the finding of Hollis and Valentine (2001), namely that names for landmarks seem to have a greater degree of meaning (sense), compared to other types proper names. For toponyms denoting natural feature types, there seems indeed to be meaning in the generic parts; however, this meaning does not seem to be the meaning

of that same term in a non-toponyms context, but rather an older meaning. Also, even within our small sample of terms, we found exceptions. For example by comparing the use of terms inside and outside of toponyms we found that *Spitze*, used in general language, refers to the summit of a mountain, whereas *Spitze* as a generic part of a toponym, as in *Zugspitze*, refers to the whole mountain. Another exception is *Horn*, which changed over time, such that its original meaning inside toponyms can be considered fossilized. The examples of *Spitze* and *Horn* show that the relation between landscape terms, inside and outside toponyms, sometimes is complex, which is in accordance with Coates (2006) claim on “properhood”, that there is no *lexical* trust in the meaning of proper names beyond their function as merely referential

The example of *Berg* has shown that with our methodological approach in some cases no clear-cut explanation can be found for how the generic part in toponyms relates to the use of *Berg* in general language, and how the unexpected and diffuse topographic patterns can be explained (Fig. 1 and Fig. 2). In such cases we need to rely on additional information and historic background of specific toponyms. For instance in the case of *Uetliberg* we found historic accounts that the etymology of *Berg* goes back to *Burg* (castle). However, we refrain from considering this a general pattern, as for instance *Zürichberg* is translated from the Latin name *monte Turegicum*, and thus already contains the reference to mountain (www.ortsnamen.ch).

6 Conclusions and Future Work

Through our approach we aimed to contribute to the ethnophysiographic research question on the relation of generic parts in toponyms and landscape terms in general language. Current approaches to such research range from theorizing axiomatic frameworks, to empirical investigations and ethnographic enquiries. In our study, we applied a new palette of data and methods, and were able to fruitfully extend the state of the art in ethnophysiography. The strength of our approach is the large spatial and temporal coverage and the comparison of large groups of generic terms, whereas we are limited in finding explanations for specific instances, as we showed for the case of *Uetliberg*. Our results may thus contribute to the existing discussions on proper names in domains such as linguistics, geography and psychology, and provide a means of hypothesis generation for more detailed empirical studies.

Several future research questions arise from our work on the relation of generic parts in toponyms with general landscape terms, which used a combination of approaches from ethnography and GIScience. In the following, we only discuss a set of issues that are relevant in the context of the ethnophysiographic research question of the relation between toponyms and landscape terms.

The results we reported in this paper have made it clear that more detailed ethnographic studies are needed in the future. One such case study would be to investigate the process of fossilization, and in particular if fossilization of generic parts in toponyms means that such terms are not productively used for coining ‘new’ toponyms anymore. We did a small thought-experiment by thinking of Antarctica as a

tabula rasa for contemporary toponym coining. We found that around 1940, during the time of German expeditions in Antarctica and the discovery of *Neuschwabenland*, many new toponyms were coined, containing the generic parts *Spitze* (e.g. *Buckinghamspitze* or *Cladonspitze*) and *Horn* (e.g. *Schmidrighorn*) (Brunk 1986). In future research we aim to investigate this topic in more detail through field studies in Switzerland.

Our results show that both generic parts of *Spitze* and *Horn* have similar topographic characteristics (Fig. 1 and Fig. 2), however, the individual spatial footprints are, to large extents, different. Are these differences caused by the influence of language patterns, such as different dialects, or is there a physical parameter, other than topography, which is distinct for both these generic parts of toponyms? The underlying reasons for this fine granular difference, if there is indeed any, is a question that remains to be answered.

The question of whether the meaning of *Berg* has changed over time, which could not be fully answered in this paper, could be investigated by using a georeferenced corpus of historic landscape descriptions. In a recent paper two of the authors presented an approach for automatically gathering landscape specifications from natural language descriptions, which might be adapted for answering the present issue with *Berg* (Derungs and Purves, 2013).

Finally, it would be interesting to extend this analysis to different languages or settings. We would assume that in Europe, where place names have a long history, it is more likely that the generic parts in toponyms have become fossilized, or that there has been a change in the meaning of these generic parts, compared to the use in standard language. In other contexts, however, where colonization has replaced place names more recently, (e.g. United States, Latin America), fossilization might only have limited impact. However, the interpretation of New World toponyms is complex, since in many cases they were borrowed from the Old World, rather than being coined in more recent times.

Acknowledgments

The research reported in this paper was funded by the project FolkOnt supported by the Swiss National Science Foundation under contract 200021-126659. We would like to thank the Text+Berg team for giving us access to a pre-processed digital version of the corpus. We would also like to thank Werner Kuhn for valuable discussions of this topic.

7 References

- Bayardo, R.J. Yiming, M.A. Srikant, RR. Scaling up all pairs similarity search. In: Proceedings of the 16th international conference on World Wide Web. 2007. S. 131-140.

- Brunk, K.: Kartographische Arbeiten und deutsche Namengebung in Neuschwanden, Antarktis. *Geschichte und Entwicklung der Geodäsie* 24(1), (1986)
- Burenhult, N. (ed.): *Language and Landscape: Geographical Ontology in Cross-linguistic Perspective*. *Language Sciences* 30 (2/3). Elsevier, Amsterdam (2008)
- Burenhult, N., Levinson, S.C.: *Language and landscape: a cross-linguistic perspective*. *Language Sciences*, 30 (2/3), 135-150 (2008)
- Coates, R.A.: Properhood. *Language* 82(2), 356-382 (2006)
- Coaz, J.: Ueber Ortsbenennung in den Schweizeralpen. In *Jahrbuch des S.A.C von 1865*, pp. 461-478. S.A.C Verlag, Bern (1865)
- Derungs, C. and Purves, R. S.: Empirical experiments on the nature of Swiss mountains. In: *Proceedings of GISRUK'07*, pp. 348-352. Maynooth, Ireland (2007)
- Derungs, C. and Purves, R.S.: From text to landscape: Locating, identifying and mapping the use of landscape features in a Swiss Alpine corpus. *International Journal of Geographic Information Science*. (2013)
- Egenhofer, M.J. and Mark, D.M.: Naive geography. In: Frank A.U. and Kuhn W., (eds.) *Spatial Information Theory: A Theoretical Basis for GIS*, pp. 1-15, Springer-Verlag, Berlin (1995)
- Fellenberg, E.: Das Aletschhorn. In *Jahrbuch des S.A.C von 1864*, pp. 183-210. S.A.C Verlag, Bern (1864)
- Feng, C., Mark, D.M.: Exploring NGA GEOnet Names Server Data for Cross-Linguistic Research on Landscape Categories: A Case Study of Some Toponyms in Malaysia and Indonesia. *GIScience 2012*, Columbus, Ohio, September 19-21, Extended Abstracts (2012)
- Fisher, P., Wood, J. and Cheng, T.: Where is Helvellyn? Fuzziness of multi-scale landscape morphology. *Transactions of the Institute of British Geographers*, 29(1), pp.106-128 (2004)
- Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. *GeoJournal* 69 (4), 211–221 (2007)
- Grice, P.: Logic and conversation. In Cole, P.; Morgan, J. (eds.): *Syntax and semantics*. 3: Speech acts. pp. 41–58. Academic Press, New York (1975)
- Hollenstein, L. and Purves, R.S.: Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, 1, 21-48 (2010)
- Hollis, J., Valentine, T.: Proper-Name Processing: Are Proper Names Pure Referencing Expressions? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 99-116 (2001)
- Jones, C.B. and Purves, R.S.: Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3), 219-228 (2008)
- Kohonen, T.: *Self-Organizing Maps*. Springer-Verlag, Berlin (1995)
- Levinson, S. C.: Foreword. In Mark, D. M., Turk, A. G., Burenhult, N., and Stea, D., (eds.) *Landscape in Language: Transdisciplinary Perspectives*, p. ix-x. John Benjamins, Amsterdam (2011)

- Lindt, R.: Das Finsteraarhorn. In Jahrbuch des S.A.C von 1864, pp. 273-312. S.A.C Verlag, Bern (1864)
- Mark, D. M., Skupin, A., and Smith, B.: Features, Objects, and other Things: Ontological Distinctions in the Geographic Domain. In Montello, D. (ed.) Spatial Information Theory: Foundations of Geographic Information Science, Berlin: Springer-Verlag, Lecture Notes in Computer Science, 2205, pp. 489-502 (2001)
- Mark, D. M., Turk A. G.: Landscape categories in Yindjibarndi: Ontology, environment, and language. In: W. Kuhn, M. F. Worboys and S. Timpf (eds.) Spatial information theory: Foundations of geographic information science. Vol. 2825, Lecture notes in computer science, pp. 28-45, Springer, Berlin (2003a)
- Mark, D. M. and Turk, A. G.: Ethnophysiography. Paper presented at: Workshop on spatial and geographic ontologies on 23rd September 2003 (prior to COSIT03) - published on web site (2003b)
- Mark, D. M., Turk, A. G., Burenhult, N., and Stea, D. (eds.) Landscape in Language: Transdisciplinary Perspectives. John Benjamins, Amsterdam (2011)
- Pike, R.J., Evans, I., Hengl, T.: Geomorphometry: A Brief Guide. In: Geomorphometry - Concepts, Software, Applications, Hengl, T. and Hannes I. Reuter (eds.), Series Developments in Soil Science, 33, pp. 3-33 (2008)
- Scheider, S., Janowicz, K. and Kuhn, W.: Grounding Geographic Categories in the Meaningful Environment. In K.S. Hornsby et al. (eds.), Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrach, France, September 21-25, 2009 Proceedings, Springer LNCS 5756, 69-87 (2009)
- Sennrich, R., Schneider, G., Volk, M., Warin, M.: A New Hybrid Dependency Parser for German. In: Chiarcos, C., de Castilho, R.E., Stede, M. (eds.) Proceedings of Biennial GSCL Conference 2009, 115-124 (2009)
- Skupin, A. and Esperbé, A.: An Alternative Map of the United States Based on an n-Dimensional Model of Geographic Space. Journal of Visual Computing, 22(4), 290-304 (2011)
- Studer, G.: Das Mattwaldhorn. In Jahrbuch des S.A.C von 1864, pp. 226-232. S.A.C Verlag, Bern (1864)
- USGS, 2013. Geographic Names Information System (GNIS). <http://geonames.usgs.gov>, accessed 31 January 2013.
- Valentine, T., Brennen, T., Brédart, S.: The Cognitive Psychology of Proper Names: On the Importance of Being Ernest. Routledge, London (1996)
- Volk, M., Bubenhofer, N., Althaus, A., Bangerter, M.: Classifying Named Entities in an Alpine Heritage Corpus. Künstliche Intelligenz, 4, 40-43 (2009)
- Wehrens, R. and Buydens, L.M.C: Self- and Super-organising Maps in R: the kohonen package. Statistical Software, 30(6), 1-19 (2007)
- Weilenmann, J.J.: Der Piz Roseg. In Jahrbuch des S.A.C von 1865, pp. 86-128. S.A.C Verlag, Bern (1865)