Exploring microtoponyms through linguistic and geographic perspectives

Julia Villette
Department of Geography
University of Zurich
Zürich, Switzerland
julia.villette@geo.uzh.ch

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

Abstract

Toponyms are an important part of our cultural heritage and are thus more than names, they also encode history. Flurnamen, or microtoponyms, are names given to natural features and they have been argued to be relatively closely linked to properties of landscape. In this paper we analyse a large databases of microtoponyms using a mix of linguistic information and simple text analysis to explore patterns of naming of natural features in the canton of St. Gallen in Switzerland. We observe that meaningful elements grounded in descriptive cultural and natural readings of landscapes are common, and that usage of microtoponyms closely follows Zipf's law.

Keywords: Flurnamen, ethnophysiography, microtoponyms, landscape

1 Introduction

Toponyms, or place names, have been the subject of study in a range of fields including linguistics, geography, history and political science (e.g. Burenhult & Levinson 2008; Zelinsky 1955; Derungs et al. 2013; Fagúndez & Izco 2016; Feng & Mark 2017; Fagúndez & Izco 2016; Radding & Western, 2010). From a geographic perspective they offer, at first glance, a beguiling window to understanding the physical environment and its development through time (Zelinsky 1955). Linguistically, naming is clearly not arbitrary, since the action of naming a place is a conscious one (Radding & Western 2010). Naming also has a purpose, allowing us to distinguish one salient location from another, facilitating communication and reducing ambiguity (Coates 2006). In this paper we concern ourselves with *Flurnamen* (microtoponyms), defined by Tyroller (1996) as toponyms given to non-populated places such as fields, mountains, forests and so on.

Microtoponyms are particularly interesting for geographers, since they refer to natural features, and thus may allow us to understand ways in which landscapes are partitioned into meaningful elements, while also providing clues as to the history of, for example, land use. However, microtoponyms, like all toponyms can act as referents without conveying any meaning (Coates 2006). For example, over time language can evolve, spelling may be normalised, events can be forgotten, and landscapes and their usage can change making the link between a name and its origins opaque. In linguistics, the field of onomastics explicitly seeks to etymologically disentangle the original meanings of toponyms, typically by exploring historical sources to find and explain the first documented usage of a toponym as referent to a place. Geographical investigations of toponyms have often focussed on a few concepts, captured through generic terms in compound toponyms, and explored, for instance, their spatial distribution (e.g. Zelinsky 1955; Derungs et al. 2013; Fagúndez & Izco 2016; Feng & Mark 2017). To date, few studies have sought to combine deeper linguistic etymological studies with automated analysis of large numbers of toponyms in space.

Making this link more explicit, and carrying out multi- and interdisciplinary research on toponyms is one of the explicit aims of the emerging field of ethnophysiography, as set out by Mark & Turk (2003). Indeed, in this context, Burenhult & Levinson (2008, p. 136) posed the question 'What is the relation between landscape terms (common nouns) and place names (proper nouns)?'

Understanding patterns of toponym usage, and relationships between, for example, generic landscape terms and place names requires some form of classification related to ways in which names are given. Tent & Blair (2011), in a detailed review, explore many of the criteria which can be used in naming, while Tyroller (1996) does so specifically in the context of microtoponyms. Tyroller makes a useful distinction between natural and culturally determined influences on naming (for example *Rotberg* (Red Mountain) versus *Rüti* (A place cleared of trees)). We assume that microtoponyms are, *sensu* Tent & Blair (2011, p. 85), more likely to be either descriptive ('indicating an inherent characteristic of the feature') or associative ('indicating something which is always or often associated with the feature or its physical context') and thus offer a relatively direct link to landscape.

In this paper, in contrast to previous work, we start from a spatially contiguous set of microtoponyms for the canton of St Gallen in Switzerland, and analyse these with respect to a detailed lexicon prepared for the same region by linguists. In doing so, we argue that we come closer to an interdisciplinary study bridging the gap between linguistics and geography.

2 Data and Methods

2.1 Study Area and Data

Study area. Our study area is the canton of St Gallen located in north-eastern Switzerland which has an area of 1951km^2. Some 48% of the canton is used for agriculture, 32% is forested and only around 10% urban. The topography ranges from plains to high mountains (from 398m to 3247m) (Kanton of St Gallen 2017). The official language of the canton is German, but the inhabitants speak a range of Swiss German dialects, and historically the language of Romansh was also spoken and has influenced toponyms.

Microtoponyms of St Gallen. Microtoponym data were provided by the canton itself and consist of names assigned to individual polygons for the whole canton. These names are collected in a bottom-up process by individual communes, and provided to the cantonal authorities for a variety of purposes. Importantly, the data are spatially contiguous, meaning that microtoponyms are assigned even to land parcels in urban areas. A total of 17598 individual names are contained in the dataset, of which some 54% (9489) are unique. On average, parcels associated with microtoponyms have an area of 11.2±2.3 hectares.

Lexicon of meaningful elements. To analyse the microtoponyms data we used a lexicon of meaningful elements. This lexicon was created by linguists tracing the etymology of individual toponyms and their components. Terms included range from generic landscape terms (e.g. *Berg* (mountain) or *Wald* (forest), adjectives (e.g. *lang* (long) or rot (red)) and spatially modifying terms which Leino (2005) postulated were used in inductive toponyms (e.g. *ober* (upper) whose usage implies another similar toponym (e.g. Upper Town implies a nearby place named simply Town or Lower Town)). The lexicon consists of 3378 meaningful elements, with links allowing meaningful elements with the same meaning to be analysed together (e.g. *Berg* and *Bärg* are alternative spellings or forms with the same roots). It is important to note that an individual microtoponym may contain no, one or more than one of these meaningful elements. Furthermore, the lexicon is not exhaustive, meaning that not all toponyms or parts thereof are contained.

2.2 Analysing microtoponyms using a lexicon of meaningful elements

Our analysis of microtoponyms data focussed on using the lexicon of meaningful elements to, firstly, explore which meaningful elements were most commonly used and, secondly, explore with what other terms the most common meaningful elements were associated. Thirdly, by performing a frequency analysis using the lexicon, we were also able to identify microtoponyms with no meaningful elements currently contained in the lexicon. Our basic approach to frequency analysis used simple string matching. However, we first removed all two letter strings from the lexicon, since these led to a high proportion of false positives.

Our matching process was iterative, and since the same microtoponyms could match onto multiple meaningful elements, we first sorted meaningful elements by length such that the longest possible matching meaningful element from the lexicon was matched onto a microtoponym. Having found a match, the matching part of a microtoponym was deleted, and the process repeated allowing further matches to be made. This is especially important since German contains many compound nouns (e.g. *Rotberg* would match both *rot* and *Berg*).

To explore the relationships between meaningful elements, we used collocation frequencies to identify combinations of meaningful elements occurring together more or less than one would expect by chance. We therefore compute χ-squared values for collocated frequent meaningful elements. χ-squared essentially measures whether or not co-occurrence is more or less than we would expect by random chance, given the underlying number of occurrences of each term. The 26 most commonly occurring meaningful elements, used as a basis for the exploration of collocation, included different sorts of terms such as: natural features (e.g. *Berg*, *Wald*), cultural features (e.g. *Hof* (farmyard), *Dorf* (village), adjectives (e.g. *lang*, *neu* (new)), spatially modifying terms (e.g. *ober*, *hinder* (behind)) and animals (e.g. *Geiss* (goat)).

3 Results and Discussion

After analysing the microtoponyms using the lexicon, we found that 15153 (86%) of microtoponyms contained at least one lexicon term. These were matched by 1409 different lexicon terms. Thus, we can surmise that meaningful elements, which are interpretable in terms of landscape properties (e.g. used in descriptive and associative ways Tent & Blair (2011)) are indeed common in St Gallen's microtoponyms. Furthermore, microtoponym uniqueness is the result of combinations of meaningful elements, rather than the use of a single term in isolation (though indeed the most common microtoponyms are meaningful elements used in isolation (compare Figure 1 and Figure 2)).

Figure 1: Frequency against rank of microtoponyms.

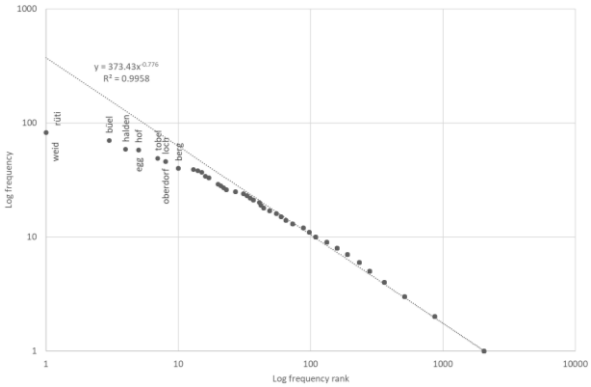
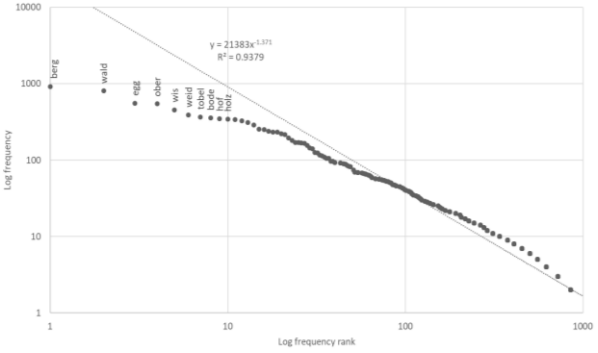


Figure 2: Frequency against rank of meaningful elements.

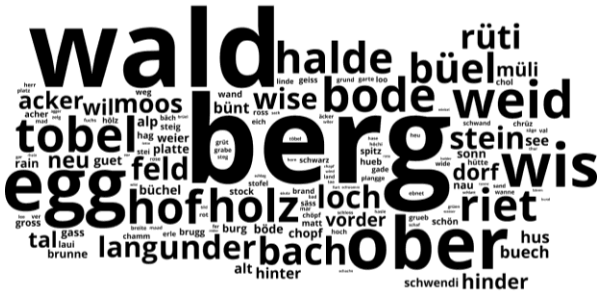


To illustrate this result we show in Figures 1 and 2 plots of meaningful element frequency and microtoponym frequency as a function of rank on a logarithmic scale. Both plots are relatively linear, thus following (more or less) Zipf's law. However, the relationship is clearly stronger for microtoponyms (r^2=0.99) than meaningful elements (r^2=0.94)

and we postulate that this has two potential reasons. Firstly, microtoponyms follow general patterns of language, which have been shown to follow a power law in many examples (c.f. Zipf 1935; Piantadosi 2014). Secondly, as we have shown, many microtoponyms use meaningful elements, which in turn reflect landscape elements and their properties. These meaningful elements however, are not likely to be distributed according to a power law (i.e. we have no reason to assume there are twice as many mountains as forests in St Gallen), but rather according to the inherent properties of the landscape. Thus, though microtoponyms follow broadly properties of language, the lexicon of meaningful elements is constrained by the properties of landscape, in other words as put by Piantadosi (2014, p. 9): 'The lexicon did not have much freedom in how it labelled the terms in these categories since the referents of these terms are salient, fixed natural binds.'

In Figures 1 and 2 we mostly concentrate on the frequencies of microtoponyms and meaningful elements, without paying much attention to their semantics. In Figure 3, we show a wordcloud of the 200 more frequent meaningful elements according to frequency.

Figure 3: 200 most frequent meaningful elements.



The ten most frequent terms are a mix of natural (*Berg* (mountain), *Wald* (forest), *Egg* (edge), *Tobel* (ravine), *Bode* (level place on or between slopes), cultural (*Wis* (meadow), *Weid* (pasture), *Hof* (farmyard), *Holz* (timber) and one term used in inductive toponyms (*ober* (upper)). Indeed, in general this list is composed of many terms which appear to be related to the physical properties of the canton of St Gallen, again suggesting the strong link between the microtoponyms and the physical environment. Note that a small number of terms in fact refer to, for example, person names (e.g. *nau* or *gar*). Their appearance in the word cloud is most likely the result of both their misclassification in the lexicon and erroneous matching on very short terms in our matching process, and demonstrates one of the limitations of our approach. Furthermore, some terms are associated with a high degree of semantic ambiguity – thus *val*, a Romansh term can refer to valley, forest or a ditch.

Figure 4 illustrates our final result, co-occurrence of a selection of meaningful elements in our microtoponyms list. Importantly, we illustrate not only raw co-occurrence counts which are influenced in turn by the overall frequency of occurrence of a term, but also the statistical significance of the co-occurrence according to a χ-squared test. Thus, all bold values in Figure 4 occur significantly more than would be expected by chance. An examination of these co-occurrences quickly demonstrates that they are also semantically meaningful choices in naming. For example *Buech* (beech) relates a particular type of tree to *Wald* (forest) or *Holz* (timber) which is particularly common in the area. *Egg* (edge) is associated with length (*lang*) and *Dorf* (village) with *neu* (new) and a range of spatial qualifiers (*under*, *hinter*, *vorder*, *ober*). It is also worth noting that *Dorf* would typically be associated with settlement names, which are not considered by linguists to be microtoponyms, however, the example allows us to show the utility of our approach well.

Underrepresented co-occurrences are also instructive. Thus, *Feld* (field) is specifically not chosen with *Berg*, *Wald* or *Holz* presumably because these features exclude one another. Furthermore, *Feld* is also note associated with *Wis* (meadow) in this case more likely because these are related categories.

4 Conclusions

In this paper we set out to show how a multidisciplinary approach could allow us to explore toponyms from both a geographic and linguistic perspective. However, it is also important to set out some key limitations. Firstly, we analyse microtoponyms in the aggregate, that is to say our approach does not include any examination of the etymologies of individual usages. Thus, although we base our analysis on a detailed, linguistic lexicon, our methods are relatively crude and rely on their application to a relatively large dataset. This in turn means we are limited to exploring, at least semantically, relatively common toponym occurences. Secondly, we analysed a contiguous dataset for the canton of St. Gallen. These data include settlement areas, something which would not typically be the case in data collected by linguists. Thirdly, our results are specific to a particular region and its landscape, and thus the occurrence of particular meaningful elements could be argued to be trivial.

However, we believe that there are also some important and useful results in this work, which will form the basis for our ongoing research. Firstly, we were able to show that microtoponyms in St. Gallen are dominated by meaningful elements, and that these meaningful elements are combined to produced microtoponyms which broadly follow Zipf's law, even though the landscape producing them (and thus the supply of concepts used in their production) probably does not. In future work we will test this hypothesis by more directly linking geographic properties to microtoponyms, and exploring their distributions.

Acknowledgments

The first author would like to thank Olga Chesnokova and Pawel Biernacki for their help getting started with Python. We would also like to thank Linda Steiner, Stefan Wuerth and Elvira Glaser for access and explanations of the lexicon, and valuable discussions on this topic. This research is part of a project 'Local Toponyms of the Canton of St. Gallen' funded by the Swiss National Science Foundation and the Canton of St. Gallen (SNF project no. 100015_162591). We also acknowledge the support of the university research priority programme language and space.

Figure 4: Co-occurrence counts for selected meaningful elements. Counts marked ** occur significantly more ($p < 0.01$) than expected by chance, *- and **- significantly less ($p < 0.05$ and $p < 0.01$ respectively).

	buech	feld	sonn	neu	lang	under	hinter	vorder	ober
wald	**20****	1**-	0*-	4	8	6	3	4	11**-
berg	7	1**-	**26****	5	1	15	**17****	10	31
egg	2	5	4	4	**15****	9	4	5	16
wis	1	1*-	0	7	**14****	4	4	3	14
holz	**15****	0*-	0	1	6	7	3	3	16
hof	2	8	**12****	**21****	1*-	1	1	1	10
dorf	0	0	0	**6****	0	**11****	**11****	**7****	**51****

References

Burenhult, N. & Levinson, S.C. (2008) Language and landscape: a cross-linguistic perspective. *Language Sciences*. [Online] 30 (2–3), 135–150. Available from: doi:10.1016/j.langsci.2006.12.028.

Coates, R. (2006) Properhood. *Language*. 356–382.

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: International Conference on Spatial Information Theory*. 2013 Springer, 261–278.

Fagúndez, J. & Izco, J. (2016) Spatial analysis of heath toponymy in relation to present-day heathland distribution. *International Journal of Geographical Information Science*. 30 (1), 51–60.

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*. 1–12.

Leino, A. (2005) In search of naming patterns: A survey of Finnish lake names. *Denominando il mondo. Dal nome comune al nome proprio/Naming the world. From Common Nouns to Proper Names*. 355–367.

Mark, D.M. & Turk, A.G. (2003) Ethnophysiography. *In: Workshop on Spatial and Geographic Ontologies*, 23 September 2003. Kartause Ittingen, Switzerland.

Piantadosi, S.T. (2014) Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*. 21 (5), 1112–1130.

Radding, L. & Western, J. (2010) What's in a name? Linguistics, geography, and toponyms. *Geographical Review*. 100 (3), 394–412.

Tent, J. & Blair, D. (2011) Motivations for naming: The development of a toponymic typology for Australian placenames. *Names*. 59 (2), 67–89.

Tyroller, H. (1996) Typologie der Flurnamen (Mikrotoponomastik): Germanisch. *Eichler et alii*. 1434–1441.

Zelinsky, W. (1955) Some Problems in the distribution of generic terms in the Place- names of the northeastern United States. *Annals of the Association of American Geographers*. 45 (4), 319–349.

Zipf, G.K. (1935) *The psycho-biology of language*. Oxford, Houghton, Mifflin.

Kanton of St Gallen. (2017) *Kopf und Zahl* [Online] Available from: https://www.statistik.sg.ch/home/publikationen/ksgmz/2017.html [Accessed 20th January 2018].