



**University of  
Zurich**<sup>UZH</sup>

# Automatisches Extrahieren von Landschaftsbeschreibungen aus öffentlich zugänglichen Texten über den Kanton Schwyz

GEO 511 Master's Thesis

**Author**

Natalie Brandenburg  
14-707-632

**Supervised by**

Prof. Dr. Ross Purves  
Dr. Maarten van Strien (vanstrien@ethz.ch)

**Faculty representative**

Prof. Dr. Ross Purves

31.03.2021

Department of Geography, University of Zurich

UNIVERSITÄT ZÜRICH

MASTERARBEIT

---

**Automatisches Extrahieren  
von Landschaftsbeschreibungen  
aus öffentlich zugänglichen Texten  
über den Kanton Schwyz**

---

*Autorin:*

Natalie BRANDENBERG  
natalie.brandenberg@uzh.ch  
14-707-632

*Betreuer:*

Dr. Prof. Ross S. PURVES,  
Uni Zürich  
Dr. Maarten J. VAN STRIEN,  
ETH Zürich

Geocomputation  
Geographisches Institut

31. März 2021

## *Zusammenfassung*

Mit dem Wissen um die Bedeutung schöner Landschaften für das Wohlergehen der Menschen wird immer klarer, dass diese geschützt werden müssen. In der Schweiz werden hierzu Landschaftskonzepte verfasst, die die Landschaft charakterisieren und bewerten. Die Anforderung besteht, dass diese möglichst auch durch die allgemeine Bevölkerung mitgestaltet werden. In dieser Arbeit wird eine Methode vorgestellt, wie die Meinung der Bevölkerung zu verschiedenen Gebieten automatisiert und in grossen Mengen gesammelt werden kann. Dazu wird eine Textsammlung über den Kanton Schwyz erstellt. Mithilfe von Regex und einem räumlichen Algorithmus werden aus diesem Korpus Toponyme extrahiert und im Raum verortet. Ebenfalls werden Landschaftsbeschreibungen extrahiert und klassiert. Dafür werden Methoden des Natural Language Processing (NLP), Regex und ein Random Forest Modell genutzt. Die Resultate sind räumlich verortete Landschaftsbegriffe. Mit diesen wird die Landschaftsbeschreibung im ganzen Kanton Schwyz visualisiert und charakterisiert. Besondere, schützenswerte Gebiete werden in ihrer Beschreibung im Textkorpus mit ihrer offiziellen Beschreibung in einem Landschaftskonzept verglichen und es werden Möglichkeiten für zukünftige Landschaftskonzepte aufgezeigt.

## *Danksagung*

Vielen lieben Dank an alle, die es mir ermöglicht haben, mein Studium abzuschliessen und im Besonderen, diese Arbeit zu verfassen.

Ich bedanke mich ganz speziell bei:

- *Ross Purves*, durch den ich erst dieses spannende Forschungsfeld kennengelernt habe und der ein so geduldiger Betreuer während dem guten Jahr Masterarbeit war
- *Maarten van Strien*, der zentral war während der Ideenfindung und die Arbeit mitbetreut haben
- *Bettina Weibel*, die mich mit der Landschaftskonzeption Schwyz bekannt gemacht hat
- *Karin Schwitter*, *Raimund Rodewald*, *Remo Bianchi* und *Valentin Kessler*, die mir Ratschläge zu interessanten öffentlichen Texten über den Kanton Schwyz gegeben haben
- *Laura*, die mir mit der Kartengestaltung geholfen hat
- *Meine Eltern*, die mir das Studium ermöglichten und immer interessiert waren, was ich da eigentlich genau lernte
- *Freundinnen* und *Freunden*, die das Studium und die Zeit der Masterarbeit auch während dem Homeoffice gesellig machten
- *Simon*, für die technische und vor allem die moralische Unterstützung.

# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>i</b>
<b>Danksagung</b>	<b>ii</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Ausgangslage . . . . .	1
1.2 Motivation . . . . .	2
1.3 Aufbau der Arbeit . . . . .	3
<b>2 Wissenschaftlicher Hintergrund</b>	<b>4</b>
2.1 Landschaft . . . . .	4
2.1.1 Definition . . . . .	4
2.1.2 Charakterisierung der Landschaft . . . . .	4
Wahrnehmung . . . . .	4
Landschaftsbegriffe . . . . .	5
2.1.3 Landschaftskonzepte . . . . .	5
2.1.4 Partizipation . . . . .	6
2.2 Geographic Information Retrieval und Natural Language Processing . . . . .	7
2.2.1 Sprache, Landschaft und Raum . . . . .	7
2.2.2 GIR . . . . .	7
Gazetteer . . . . .	7
Georeferenzierung . . . . .	8
Natürliche Phänomene lokalisieren . . . . .	8
2.2.3 NLP . . . . .	9
2.2.4 Evaluation . . . . .	9
<b>3 Verortung der Arbeit</b>	<b>10</b>
3.1 Forschungslücke . . . . .	10
3.2 Forschungsfragen und Hypothesen . . . . .	10
3.3 Abgrenzung der Arbeit . . . . .	11
<b>4 Daten</b>	<b>13</b>
4.1 Texte . . . . .	13
4.2 Landschaftskonzeption . . . . .	14
4.3 Gazetteer . . . . .	16

4.4	Liste der Landschaftsbegriffe . . . . .	17
<b>5</b>	<b>Methoden</b>	<b>19</b>
5.1	Aufbereitung der Daten . . . . .	20
5.1.1	Annotation . . . . .	20
5.1.2	Korpus aufbereiten . . . . .	22
5.1.3	Toponym Recognition . . . . .	23
	Named Entity Recognition . . . . .	23
	Abgleich mit Gazetteer und Anwendung von Regex . . . . .	23
5.1.4	Landschaftsrelevanz bestimmen . . . . .	24
	LSB mit Regeln suchen . . . . .	24
	Landschaftskontext mit Machine Learning bestimmen . . . . .	24
5.1.5	Toponym Resolution . . . . .	26
5.2	Räumliche Analyse . . . . .	27
5.3	Inhaltliche Analyse . . . . .	28
5.4	Vergleich mit der Landschaftskonzeption Schwyz . . . . .	29
5.5	Fehlerberechnung . . . . .	30
<b>6</b>	<b>Resultate</b>	<b>32</b>
6.1	Textaufbereitung . . . . .	32
6.1.1	Precision und Recall . . . . .	32
6.1.2	Statistische Beschreibungen . . . . .	34
6.2	Räumlich: Toponyme . . . . .	37
6.3	Inhaltlich: Landschaftsbegriffe . . . . .	42
6.4	Fokus Schlüsselgebiete . . . . .	46
<b>7</b>	<b>Diskussion</b>	<b>52</b>
7.1	Interpretation . . . . .	52
7.1.1	Forschungsfragen . . . . .	53
	Forschungsfrage 1 . . . . .	53
	Forschungsfrage 2 . . . . .	54
	Forschungsfrage 3 . . . . .	54
7.2	Einschränkungen . . . . .	54
7.2.1	Datenwahl und Textaufbereitung . . . . .	55
7.2.2	GIR- und NLP-Methoden . . . . .	58
7.2.3	Interpretation . . . . .	61
7.2.4	Repräsentanz . . . . .	63
<b>8</b>	<b>Schlusswort</b>	<b>65</b>
8.1	Fazit . . . . .	65
8.2	Weiterführende Arbeiten . . . . .	65
	<b>Literatur</b>	<b>67</b>

<b>A</b>	<b>Vorgeschlagene Texte für den Korpus</b>	<b>73</b>
<b>B</b>	<b>Liste der Landschaftsbegriffe</b>	<b>76</b>
B.1	LSB-Liste . . . . .	76
B.1.1	Definitive Begriffe . . . . .	76
B.1.2	Begriffe im LS-Kontext . . . . .	79
B.1.3	Compound-Wörter . . . . .	80
B.1.4	Kontextwörter . . . . .	80
B.2	LSB in der Landschaftskonzeption und Erweiterungen mit LSB im Katalog . . . . .	81
<b>C</b>	<b>Kosinus-Ähnlichkeit zwischen den Landschaftsbegriffen</b>	<b>83</b>
	<b>Persönliche Erklärung</b>	<b>85</b>

## Abbildungsverzeichnis

3.1	Übersichtskarte Untersuchungsgebiet Kanton Schwyz. . . . .	11
4.1	Anteile der beiden Korpora an allen Sätzen. . . . .	14
4.2	Ausschnitt aus der Landschaftskonzeption Kanton Schwyz. . . . .	15
4.3	Plan Schlüsselgebiete aus der Landschaftskonzeption Kanton Schwyz. . . . .	16
4.4	Ausschnitt aus der LSB-Liste. . . . .	18
5.1	Arbeitsablauf der Datenaufbereitung . . . . .	19
5.2	Beispiel für die Auswahl der nächsten Koordinatenpaare dreier Toponyme. . . . .	27
5.3	Anteil relevanter Sätze am ganzen Korpus. . . . .	30
6.1	Anteil gefilterter Sätze pro Verarbeitungsschritt. . . . .	33
6.2	Häufigste Toponyme in beiden Korpora (geordnet nach der Toponymfrequenz im Boten der Urschweiz). . . . .	35
6.3	Häufigste Toponyme in beiden Korpora (geordnet nach der Toponymfrequenz in ausgewählten Sätzen im Y-Magazin). . . . .	35
6.4	PDF-Parsing: Zeitungsartikel (A) und Textdatei (B). . . . .	37
6.5	Vorkommen Toponyme aus beiden gefilterten Korpora pro Geometrietyp. . . . .	38
6.6	Dichtekarte aller Toponymvorkommnisse im Boten. . . . .	39
6.7	Dichtekarte aller Toponymvorkommnisse im Boten ohne Schwyz. . . . .	39
6.8	Dichtekarte gefilterter Toponymvorkommnisse im Boten ohne Schwyz. . . . .	40

6.9	Dichtekarte aller Toponymvorkommnisse im Y-Magazin ohne Schwyz. . .	40
6.10	Dichtekarte gefilterter Toponymvorkommnisse im Y-Magazin ohne Schwyz. . .	41
6.11	Häufigste Landschaftsbegriffe im Boten und im Y-Magazin (sortiert nach dem Vorkommen im Boten). . . . .	42
6.12	Word Cloud aus den LSB im Boten der Urschweiz. . . . .	43
6.13	Word Cloud aus den LSB im Y-Magazin. . . . .	43
6.14	Netzwerk der häufigsten Landschaftsbegriffe . . . . .	44
6.15	Chi <sup>2</sup> -Karten: Berge und Gipfel. . . . .	45
6.16	Chi <sup>2</sup> -Karten: Aussicht und Blick. . . . .	45
6.17	Dichtekarten zweier Landschaftsbegriffe: See und Ski & Skigebiet. . .	46
6.18	Häufigste Landschaftsbegriffe ausserhalb und innerhalb der Schlüsselgebiete kombiniert aus beiden Korpora. . . . .	47
6.19	Schlüsselgebiet Glattalp: LSB Word Cloud. . . . .	48
6.20	Schlüsselgebiet Wägital: LSB Word Cloud. . . . .	48
6.21	Schlüsselgebiet Riemenstalden: LSB Word Cloud. . . . .	48
6.22	Schlüsselgebiet Mythen: LSB Word Cloud. . . . .	49
6.23	Schlüsselgebiet Etzel - St. Meinrad: LSB Word Cloud. . . . .	49
6.24	Schlüsselgebiet Muotathal: LSB Word Cloud. . . . .	49
6.25	Word Cloud aus den LSB in der Landschaftskonzeption. . . . .	51
7.1	Chi <sup>2</sup> -Karte der Toponymdichte im Y-Magazin. . . . .	53
7.2	Punktoponyme im Gazetter Swissnames3D und im Boten der Urschweiz . . . . .	55
7.3	Falsch disambiguierte Toponyme . . . . .	60
7.4	BLN-Schutzgebiete im Kanton Schwyz. . . . .	63

## Tabellenverzeichnis

6.1	Precision und Recall für Datenaufbereitung . . . . .	33
6.2	Toponym- und LSB-Verteilung in gefilterten Sätzen . . . . .	36
6.3	Toponymdichten innerhalb und ausserhalb der Schlüsselgebiete. . . .	47
6.4	LSB-Dichten innerhalb und ausserhalb der Schlüsselgebiete. . . . .	47
6.5	Kosinus-Ähnlichkeiten zwischen den Schlüsselgebieten aufgrund der LSB-Kookurrenz. . . . .	50
6.6	Jaccard Index der LSB innerhalb der Schlüsselgebiete und ausserhalb der Schlüsselgebiete. . . . .	51
C.1	Kosinus-Ähnlichkeiten der LSB aufgrund ihrer Kookkurrenz. . . . .	84

# Abkürzungsverzeichnis

<b>BLN</b>	<b>B</b> undesinventar der <b>L</b> andschaften und <b>N</b> aturdenkmäler
<b>GIR</b>	<b>G</b> eographic <b>I</b> nformation <b>R</b> etrieval
<b>LSB</b>	<b>L</b> and <b>S</b> chafts <b>B</b> egriffe
<b>ML</b>	<b>M</b> achine <b>L</b> earning
<b>NER</b>	<b>N</b> amed <b>E</b> ntity <b>R</b> ecognition
<b>NLP</b>	<b>N</b> atural <b>L</b> anguage <b>P</b> rocessing
<b>UGC</b>	<b>U</b> ser <b>G</b> enerated <b>C</b> ontent

## Kapitel 1

# Einleitung

### 1.1 Ausgangslage

Mit der europäischen Landschaftskonvention 2000, die in der Schweiz 2013 in Kraft trat, wird der Landschaft in der Raumplanung einen hohen Stellenwert zugeordnet. Landschaft wird hier nicht bloss als die besonders schöne, photogene Landschaft verstanden, sondern auch als die Alltagslandschaft, die von der breiten Bevölkerung jeden Tag erlebt wird und deren Ästhetik und Wahrnehmung durch die Bevölkerung oft stark vernachlässigt wird ([European Landscape Convention](#)). Das Übereinkommen fordert unter anderem von den unterzeichnenden Nationen die Erfassung und Bewertung der eigenen Landschaftstypen. Dabei sollen vor allem auch Betroffene zur Sprache kommen und ihre Wahrnehmung der verschiedenen Landschaftsformen teilen ([European Landscape Convention](#)).

Die Umsetzung der Landschaftscharakterisierung ist offengelassen. In der Schweiz hat sich die Stiftung Landschaftsschutz Schweiz der Aufgabe angenommen und einen Katalog verschiedener Landschaftstypen aufgebaut: den Katalog charakteristischer Kulturlandschaften Schweiz (Rodewald, Schwyzer und Liechti, 2014). Dieser Katalog ist für die planerische Nutzung ausgelegt. Er soll Kantonen als Grundlage zur Ermittlung von Landschaftsentwicklungszielen dienen. Die Verantwortung für ein Landschaftskonzept liegt in der Schweiz bei den Kantonen. Nur wenige Kantone haben mit einer Charakterisierung begonnen. Darunter ist der Kanton Schwyz. Das Amt für Natur, Jagd und Fischerei hat mit der Stiftung Landschaftsschutz Schweiz und einer Begleitgruppe aus anderen kantonalen Ämtern die Landschaftskonzeption erarbeitet. Wichtig in der Erarbeitung der landschaftlichen Entwicklungsziele sind speziell auch die so genannten Schlüsselgebiete als Vertreter von besonders charakteristischen und qualitätsreichen Landschaften des Kantons. Neben diesen Schutzgebieten, die besonders genau beschrieben und auch bewertet und mit Entwicklungszielen bedacht wurden, gibt es den Rest des Kantons, der weniger genau beschrieben wird, aber dessen Oberflächen ebenfalls verschiedenen Landschaftstypen zugeteilt wurden ([Landschaftskonzeption Kanton Schwyz. Version vom Februar 2019](#)).

In meiner Arbeit soll die expertenbasierte Seite der Herangehensweise, wie sie

von der Stiftung Landschaftsschutz Schweiz vorgeschlagen wird, durch eine populäre ergänzt werden. Die Idee ist, nicht mit Expertenmeinungen zur physischen und historischen Beschaffenheit der Gegend zu arbeiten, sondern mit der Wahrnehmung und Bewertung der Landschaft wie sie von Nicht-Expertinnen beschrieben wird. Die Methoden der Datenakquirierung sind, wenn es um die Landschaftswahrnehmung geht, grösstenteils Interviews, sei es von Nutzern des Landes, wie Bauern, oder von Besucherinnen in einem Naherholungsgebiet (Hermes u. a., 2018). Die Popularität des Internets und einfache Erreichbarkeit von digitalen Texten ermöglicht eine hohe Quantität an Meinungen und die Möglichkeit, die subjektive Wahrnehmung von Personen, die tatsächlich in der Landschaft waren, in die Charakterisierung einer Gegend einfließen zu lassen. Automatisierte Textanalyse und Georeferenzierung machen dieses Vorgehen effizient und gut skalierbar und die Informationen daraus können die bestehende Datenakquirierung ergänzen und zur Beobachtung über grössere Zeitspannen genutzt werden (Koblet und Purves, 2020).

Ähnliche Projekte wurden von Koblet und Purves, 2020 in der Region des englischen Nationalpark *Lake District* durchgeführt. Sie haben eine Textsammlung aus dem Internet erstellt, in diesen Wahrnehmungen wie Geräusche und Gerüche gesucht und im Nationalpark verortet. Wartmann, Acheson und Purves, 2018 haben Fotografiebeschreibungen betrachtet. Dabei wurde untersucht, welche Landschaftselemente den Leuten eigentlich auffallen und mit welchen Adjektiven oder Wahrnehmungsverben, diese häufig auftraten. Mit den Wörtern konnte die Landschaft charakterisiert werden und festgestellt werden, welche Regionen als besonders attraktiv gelten (Chesnokova, Nowak und Purves, 2017).

## 1.2 Motivation

Landschaftskonzepte sind Neuland in der Schweiz. Trotzdem oder vielleicht auch deswegen steht die Landschaft in der Schweiz unter Druck durch Klimawandel und Zersiedlung, und Naherholungsgebiete werden seit Corona wichtiger denn je (Schwick, 2010, *Medienmitteilung. Freizeitaktivitäten erhöhen den Druck auf die Natur* 2020).

Es ist wichtig, dass zumindest eine Charakterisierung, eine Inventarisierung der bestehenden Verhältnisse in der Landschaft vorgenommen wird. Danach sollte eine möglichst ausgewogene Bewertung ihrer Qualität folgen, inklusive ökologischen Aspekten, Ansichten von Nutzern und Wahrnehmung von Besucherinnen ([European Landscape Convention](#)).

Bemerkenswert an der Landschaftskonzeption von Schwyz ist die Einführung der Schlüsselgebiete. Diese sollen bei der nächsten Richtplanüberarbeitung als kantonale Schutzgebiete ausgewiesen werden, analog zum Bundesinventar der Landschaften und Naturdenkmäler auf nationaler Ebene (*Bundesinventar der Landschaften und Naturdenkmäler (BLN)* 2020). Die Schlüsselgebiete werden am ausführlichsten aufgeführt und durch diese hat die Landschaftskonzeption auch den grössten Einfluss. Darum wird während der Arbeit ein besonderer Fokus auf diese Gebiete

gelegt.

Es handelt sich bei dieser Arbeit also auch um einen *proof of concept* mit einem vergleichsweise kleinen Datensatz, der die Möglichkeit dieser Methode für zukünftige Erarbeitungen von Landschaftskonzepten aufzeigen soll.

### **1.3 Aufbau der Arbeit**

Im Kapitel 2 wird erklärt, warum die Landschaft wichtig ist für die Bevölkerung, wie Landschaftskonzepte erstellt werden und mit welchen Methoden Teile davon automatisiert werden können. Es folgen meine Hypothesen, die Forschungsfragen und mit welcher Datengrundlage und Methoden diese beantwortet werden sollen (Kap. 3, 4 und 5). Die Resultate (Kap. 6) und wie sich diese mit der Landschaftskonzeption vergleichen lassen, werden in Kapitel 7 diskutiert und mit einem Schlusswort in Kapitel 8 abgeschlossen.

## Kapitel 2

# Wissenschaftlicher Hintergrund

Das grosse Potenzial an Wissen des Internets und allgemein elektronischen Dokumenten automatisch auszuschöpfen, ist nicht einfach. Das liegt daran, dass die Texte nicht wie in Datenbanktabellen üblich strukturiert aufgebaut sind und damit für Maschinen schwierig zu verstehen sind (Fujii und Ishikawa, 2000). Trotzdem wollen Forscher und Forscherinnen dieses Wissen abschöpfen und die grosse Datenflut übersichtlich darstellen (Fortuna, Grobelnik und Mladenic, 2005). Auch Texte über Landschaft werden schon lange geschrieben. Was man in ihr unternehmen kann, was man in ihr sieht, fühlt und erlebt und wie man sie nutzen kann.

## 2.1 Landschaft

### 2.1.1 Definition

Die Europäische Landschaftskonvention ([European Landscape Convention](#)) definiert Landschaft prägnant: „ein Gebiet, wie es vom Menschen wahrgenommen wird, dessen Charakter das Ergebnis der Wirkung und Wechselwirkung von natürlichen und/oder menschlichen Faktoren ist.“ Landschaft ist, wenn man diese Definition streng auslegt, also alles, was uns in einem Gebiet umgibt und beinhaltet von Stadtlanschaften, über landwirtschaftlich geprägte Regionen bis zu eisigen Gletscherlandschaften alles. Weiter handelt es sich bei der Landschaft um ein von Menschen wahrgenommenes Gebiet. Die Frage, wie Landschaft also von Menschen wahrgenommen und wie sie ihre Wahrnehmung ausdrücken, die Landschaft charakterisieren, ist also zentral und lässt sie sogar erst entstehen.

### 2.1.2 Charakterisierung der Landschaft

#### Wahrnehmung

Aus der Landschaftsdefinition der Europäischen Landschaftskonvention lässt sich ein konzeptionelles Modell der Landschaft aus drei Bereichen aufbauen. Die Bereiche subjektive Wahrnehmung (bestehend aus Emotionen, Aktivitäten), objektive Wahrnehmung (reale Objekte) und die kulturelle Wahrnehmung (Werte) bestimmen

die Wahrnehmung der Landschaft (Rodewald u. a., 2020). Innerhalb dieser drei Bereiche finden sich auch die unterschiedlichsten Ansätze, die im Lauf der Zeit versuchten unsere Präferenz für gewisse Landschaften zu erklären. Verfechter der objektivistischen Betrachtung weisen auf die Universalität der Bewunderung für die wilde Natur, Alpengipfel und Sonnenuntergänge hin (Seel, 1991). In der Habitattheorie und der Prospect-Refuge-Theorie von Appleton, 1996 kommt neben der Bewunderung für schöne Landschaften der objektive Nutzen der Landschaft hinzu. Menschen bevorzugen demnach Landschaften, in denen sie ihr Überleben sichern können mit Ressourcen (Wasser, Bäume) und in der sie sowohl die Übersicht (prospect, z.B. Aussicht) haben, wie auch die Möglichkeit sich zu schützen (refuge, z.B. eine Hecke im Rücken). Costonis, 1982 sieht den Wunsch nach Stabilität der kulturellen Identität als Grund, weshalb eine Gesellschaft ähnliche Landschaften mag und mit ihnen ein Heimatgefühl verbindet.

Unabhängig davon welcher Lehrmeinung man anhängt, es ist klar, dass Landschaften uns beglücken können, und dass besonders Alltagslandschaften, wo wir die meiste Zeit verbringen, für das allgemeine Wohlergehen der Menschen gefördert werden sollen (European Landscape Convention). Das gute an der Analyse der Landschaft via Sprache ist, dass sowohl die physische Manifestation des Betrachteten als auch die sozio-kulturelle Prägung in einer sprachlichen Beschreibung vertreten sind (Hunziker, Buchecker und Hartig, 2007, Mark u. a., 2011). Die Sprache muss dabei aber richtig interpretiert werden und auf einer gemeinsamen Grundlage aufbauen.

### **Landschaftsbegriffe**

Wenn Landschaften beschrieben werden sollen, werden sie selten als ein Ganzes benannt, sondern es werden kleine Einheiten, Landschaftselemente, beschrieben (Tversky und Hemenway, 1983). Diese greifbaren Elemente der Landschaft sind nützlich, da sie vergleichbar sind und als Indikatoren dienen können, wie sie sich verändert. Ausserdem gehören sie zu den wichtigsten Faktoren, die Landschaften beliebt machen (Zube, Sell und Taylor, 1982, Yang und Brown, 1992).

Unter den Landschaftsbegriffen, die am häufigsten von Teilnehmern in einer Studie von Garcia-Martin u. a., 2017 genannt wurden, kamen die Themen Freizeit und Ästhetik am meisten vor. Die Antworten waren aber unterschiedlich ja nach Alter, Beschäftigung und Beziehung des Teilnehmers zur Landschaft.

#### **2.1.3 Landschaftskonzepte**

Eine reine Ansammlung von Begriffen ohne Einordnung, Bewertung und Zielsetzung ist nur beschränkt nützlich. Deshalb braucht es Landschaftskonzepte. Das Ziel von diesen Konzepten ist, die Landschaft zu beschreiben, zu bewerten und ein Schutz- bzw. Zielzustand zu formulieren (Rodewald, Schwyzer und Liechti, 2014). Sie sollen

Planerinnen als Grundlage für die Siedlungsentwicklung dienen und können auch als Monitoringinstrument eingesetzt werden.

England und Schottland gehören zu den Vorreitern im Bereich der Landschaftscharakterisierung. Beim Landscape Character Assessment geht es darum, die Landschaft in möglichst eindeutige, d.h. charakteristische, Flecken einzuteilen und zu erklären, welche Elemente - Objekte, aber auch Ungreifbares wie Sense of Place - diese Landschaft einzigartig machen (Tudor, 2014).

In der Schweiz hingegen bildet der Katalog der Kulturlandschaften Schweiz ein Register an möglichen Landschaftstypologien, die ein Gebiet ausmachen können (Rodewald, Schwyzer und Liechti, 2014). Wie in Kapitel 1.1 beschrieben, ist der Kanton Schwyz einer der ersten Kantone, der für sich eine Gesamtbetrachtung der Landschaft erstellte. Die Landschaftskonzeption Schwyz ist noch nicht abgeschlossen, aber der Oberfläche des Kantons wurden bereits einzelne typische Kulturlandschaften zugeordnet und es wurden speziell charakteristische Gebiete eruiert, die in Zukunft unter stärkeren Schutz fallen sollen (s. Abb. 4.3) (*Landschaftskonzeption Kanton Schwyz. Version vom Februar 2019*).

Grundlegend in allen Landschaftskonzepten ist die Meinung der Bevölkerung. Eine möglichst breite Teilnahme zu versichern, ist allerdings nicht einfach (Jones und Stenseke, 2011). Die Methoden, die genutzt werden, um die Nutzerinnen und Nutzer der Landschaft zu Wort kommen zu lassen, reichen von ethnographischen Ansätzen wie Interviews oder Participatory Mapping bis zur kontaktlosen Informationsextraktion von User Generated Content (Wartmann, Derungs und Purves, 2016).

#### 2.1.4 Partizipation

Unter der Beteiligung der Öffentlichkeit wird verstanden, dass sich auch Personen und Gruppen eingeben können, die sonst nicht in formellen Entscheidungsprozessen involviert sind. Dabei sollen möglichst viele verschiedene Interessen vertreten sein, um die Erfahrung und das informelle Wissen der breiten Bevölkerung abzuholen (Jones und Stenseke, 2011). Die einfachste Form davon sind Informationsveranstaltungen, bei denen die Bürger und Bürgerinnen ihr Wissen weitergeben dürfen. Interaktivere Formen der Partizipation sind Workshops und Interviews. Dabei fallen Kosten an, die Apathie und mögliches Misstrauen der Beteiligten muss überwunden werden (Jones und Stenseke, 2011).

Unter diesen Umständen scheint der öffentliche und kostenlose Zugang zum Internet eine gute Alternative (Koblet und Purves, 2020). In diesen unkontrollierten Texten wird genau der Wunsch nach breiter Bevölkerungsbeteiligung möglich gemacht. Obwohl die Leute keine Expertinnen sind und sich die Meinungen von einzelnen Personen stark unterscheiden können, kann das Gesetz der grossen Zahlen zu einer kollektiven Intelligenz führen (Hill und Ready-Campbell, 2011). Das Internet wird allerdings sehr unterschiedlich verwendet. Verfasser machen nur 20% oder weniger der Nutzer aus (Ochoa und Duval, 2008).

## 2.2 Geographic Information Retrieval und Natural Language Processing

Schütze, Manning und Raghavan, 2008 beschreiben Information Retrieval als das Auffinden von Dokumenten an unstrukturiertem Text aus einer grossen Sammlung an Dokumenten, die einem eine gesuchte Information liefern (p. 1). Unter Geographic Information Retrieval (GIR) wird dementsprechend das Auffinden von Dokumenten, die geographische Informationen liefern, verstanden (Jones und Purves, 2008). Natural Language Processing (NLP) deckt zum Teil Aspekte der Information Retrieval ab, wenn es um Möglichkeiten geht, möglichst die richtigen Dokumente auf eine Abfrage zurückzugeben. Bei NLP geht es darum, die menschliche Sprache einer Maschine verständlich zu machen und so automatisch Information zu erlangen (Manning und Schütze, 1999).

### 2.2.1 Sprache, Landschaft und Raum

In einem Dokument aufgelistet helfen Landschaftsbeschreibungen nicht dabei, Landschaftskonzepte zu formulieren. Dazu müssen sie mit dem geographischen Raum verknüpft werden. Je mehr Textdokumente verortet werden können, desto besser können räumliche Muster und Muster in verschiedenen Arten von Texten und verschiedenen Urhebern von Texten erkannt werden (Cooper und Gregory, 2011). Denn Landschaftsbeschreibungen sind nicht zufällig im Raum verteilt, sondern bilden Cluster (Garcia-Martin u. a., 2017). Sie sind beeinflusst von der natürlichen Umgebung wie der Topographie, der Flora und Fauna und dem Klima. Aber sie unterscheiden sich auch je nach Sprache und Kultur, die, wie bereits beschrieben, unsere Wahrnehmung der Landschaft beeinflusst (Derungs und Purves, 2014).

### 2.2.2 GIR

Natürliche Sprache ist notorisch unspezifisch und so auch Sprache mit geographischer Information (Purves u. a., 2018). Die Aussage „Blick vom Wanderweg auf der Rigi hinab auf den See“ birgt verschiedene Herausforderungen. Wo auf dem grossen Rigi massiv mag der Wanderweg liegen? Welchen See kann man sehen? Tatsächlich, welches Rigi ist gemeint und wo liegt es?

#### Gazetteer

*Rigi* bildet die Georeferenz, die den Satz im Raum verortet. Mithilfe eines Gazetteers kann der Link zwischen dem Text und dem realen Ort gemacht werden. Gazetteers oder Ortsnamenslexika enthalten Toponyme oder Ortsnamen. Diese sind von verschiedenster Art, verschiedenster räumlicher Ausdehnung und Genauigkeit. Ein Eintrag eines Gazetteers muss zumindest aus einem Namen und Koordinaten

bestehen. Weiter enthält er Informationen zum Typ des Ortes - ist es ein Land, Quartier, Fluss, Bushaltestelle - und andere Details zur Ausprägung des Ortes, z.B. die Bevölkerungsgrösse (Goodchild und Hill, 2008). In einem Gazetteer können mehrere Namen pro Ort vorkommen. Das ist typischerweise in der mehrsprachigen und dialektreichen Schweiz der Fall. Der gleiche Namen benennt oft auch verschiedene Orte (z.B. Rüti). Verschiedene Gazetteers können unterschiedliche Informationen enthalten: das reicht von unterschiedlichen Schreibweisen der Ortsnamen über unterschiedliche geometrische Ausdehnungen und Skalen bis zu verschiedenen Herangehensweisen, die Art eines Ortes zu benennen (z.B. Gemeinde oder Stadt, Fluss oder Bach) (Acheson, Volpi und Purves, 2020).

### Georeferenzierung

Die Georeferenzierung umfasst das Identifizieren von Orten in Text und ihre Lokalisierung im Raum. Dabei geht es im Grundsatz darum, zwei Probleme zu lösen, die Geo/Non-Geo-Unsicherheit und die Geo/Geo-Unsicherheit. Die Geo/Non-Geo-Unsicherheit umfasst die Schwierigkeit, Wörter im Text als Toponyme zu erkennen. Die Mehrdeutigkeit von Wörtern bildet dabei die grösste Herausforderung (vergleiche Baden im Wasser und Baden im Kanton Aargau). Der Vorgang die Toponyme im Text zu erkennen heisst *Geoparsing* oder *Toponym Recognition* (Purves u. a., 2018). Dabei helfen Gazetteers und Methoden des NLP.

Beim *Geocoding* oder *Toponym Resolution* geht es darum, die Geo/Geo-Unsicherheit zu überwinden und aus verschiedenen Kandidaten mit dem gleichen Namen den richtigen auszuwählen. Dazu gibt es verschiedene Ansätze: distanzbasierte, wissensbasierte und Machine Learning Methoden (Buscaldi, 2011). Die distanzbasierten Methoden orientieren sich allein an der Nähe verschiedenen Toponyme zueinander. Wissensbasierte Methoden nehmen Kontextwissen wie Informationen aus Gazetteers dazu, z.B. zur Bevölkerungsgrösse eines Ortes. Machine Learning Methoden versuchen anhand statistischer Modelle noch mehr Kontext aus dem Text zu erkennen und z.B. bekannte Organisationen mit bekannten Orten zu verlinken.

Die letzte Herausforderung der Georeferenzierung sind Beziehungswörter wie *neben*, *in der Nähe von*, *entlang* und die unscharfe geometrische Ausdehnung von Objekten (Acheson, Volpi und Purves, 2020).

### Natürliche Phänomene lokalisieren

Wie bereits beschrieben, macht unter anderem die Unschärfe der Sprache Georeferenzierung schwierig (Mark u. a., 2011). Dazu gehört der Umstand, dass wir zwar allgemein verständliche Wörter haben für Objekte im Raum, aber eigentlich nicht ganz klar ist, wo diese Objekte beginnen und aufhören. Ein klassisches Beispiel dafür ist der *Berg* (Smith und Mark, 2003). Gazetteers enthalten weniger Zusatzinformationen zu so natürlichen Phänomenen wie zu den anthropogenen. Acheson,

Volpi und Purves, 2020 hatten Erfolg bei der Lokalisierung natürlicher Phänomene mithilfe von Machine Learning Methoden. Egorova u. a., 2018 haben untersucht wie dynamische Bewegungen erkannt und kartiert werden können. Dazu gehören Beziehungswörter in der Sprache wie *die Strasse führt entlang des Sees* oder *der Weg führt den Berg hinauf*.

### 2.2.3 NLP

Mithilfe von Natural Language Processing soll die Bedeutung der Wörter aus einem Text hervorgeholt werden. Ursprünglich funktionierte das mit einem Regelwerk, in dem die Bedeutung von Wörtern, ihre Endungen und mögliche Kombinationen miteinander festgehalten werden konnten. Um alle Feinheiten einer Sprache einzufangen (Wortspiele, seltene Wortbedeutungen in unterschiedlichen Kontexten) bräuchte es unendlich viele Regeln (Nadkarni, Ohno-Machado und Chapman, 2011). Regular Expressions (Regex) sind ein mächtiges Tool, um die erkannten grammatikalischen und syntaktischen Regeln umzusetzen. Eine Regex ist eine Schreibweise für Muster im Text, die zur Regex passen müssen (Gonzalez-Pardo und Camacho, 2011).

Mit leistungsstärkeren Computern und der Entwicklung von neuen Machine Learning Algorithmen wurden mehr und mehr probabilistische Sprachmodelle entwickelt, die anhand von ebenfalls schnell wachsenden elektronischen Dokumentsammlungen trainiert werden konnten (Hirschberg und Manning, 2015). Zu den typischen NLP Aufgaben gehören das Erkennen von einzelnen Tokens (einzelne Wörter, Satzzeichen), das Part of Speech Tagging von einzelnen Wortarten und die Named Entity Recognition (Erkennen von benannten Instanzen) (Nadkarni, Ohno-Machado und Chapman, 2011).

### 2.2.4 Evaluation

Die Evaluation von GIR- und NLP-Methoden soll zwei Dimensionen abdecken, die thematische und die räumliche (Clough, Joho und Purves, 2006). Gritta u. a., 2018 führen Evaluationsmetriken auf, die sowohl für die Bestimmung der thematischen wie auch der räumlichen Relevanz verwendet werden können. Zwei häufige Metriken sind *Precision* und *Recall*. Die Metriken gehören zusammen. Precision ist der Anteil der Toponyme oder Wörter, die korrekt erkannt oder lokalisiert wurden. Recall ist der Anteil der gefundenen Toponyme oder Begriffen von allen, die möglich gewesen wären. Der *F-score* ist eine Kombinationsmetrik aus Precision und Recall und ermöglicht auf einen Blick eine Bewertung. Eine Metrik, die nur die räumliche Genauigkeit bewerten kann, ist die mittlere Fehlerdistanz. Dabei wird der Median der Entfernung zwischen dem kartierten Toponym und der richtigen Koordinaten verwendet.

Damit die Metriken überhaupt berechnet werden können, muss ein Standard bekannt sein. Dieser wird durch annotierte Daten gegeben (Gritta u. a., 2018). Ein Teil der eigenen Daten sollte annotiert werden mit den Antworten, die man von den Methoden erwartet.

## Kapitel 3

# Verortung der Arbeit

Aus den bisherigen Arbeiten zum Thema erkenne ich die folgende Forschungslücke, aus der sich drei Forschungsfragen ableiten lassen.

### 3.1 Forschungslücke

In einem schweizerischen Kontext und in der deutschen Sprache wurden die Methoden der NLP und GIR noch nicht dazu angewendet, Landschaftsbeschreibungen der breiten Bevölkerung zu erforschen. Einen Vergleich zwischen einer offiziellen Landschaftscharakterisierung und der Meinung von Laien gibt es meines Wissens noch nicht. Die allermeisten Kantone stehen noch vor der Aufgabe, ein Landschaftskonzept zu erstellen, es ist daher auch von praktischem Interesse, einen Vergleich der Landschaftscharakterisierung von ExpertInnen und der breiten Bevölkerung zu haben.

### 3.2 Forschungsfragen und Hypothesen

Aufgrund früherer Arbeiten sind meine Hypothesen,

- dass ich von ExpertInnen als schützenswert eingestufte Gebiete vom Umland unterscheiden lassen, wenn es um die Anzahl und Art der Landschaftsbegriffe geht, mit denen sie beschrieben werden (Garcia-Martin u. a., 2017, Chesnokova und Purves, 2018b, Koblet und Purves, 2020)
- dass Schlüsselgebiete aufgrund der in den Korpora gefundenen Landschaftsbegriffen auch untereinander unterscheidbar sind (*Landschaftskonzeption Kanton Schwyz. Version vom Februar 2019*).

Daraus leite ich 3 Forschungsfragen ab.

1. Wie unterscheidet sich räumliche Verteilung von Toponymen in Sätzen mit einem Landschaftsbezug von der räumlichen Verteilung aller Toponyme im Kanton?

2. Wie unterscheiden sich die Dichten an verorteten Landschaftsbegriffen jeweils zwischen schützenswerten Gebieten (d.h. Schlüsselgebiete) und dem restlichen Kantonsgebiet?
3. Wie unterscheiden sich die Landschaftsbegriffe zwischen den Korpora und der Landschaftskonzeption in den Schlüsselgebieten?

### 3.3 Abgrenzung der Arbeit

Die Arbeit beschränkt sich räumlich auf das Gebiet des Kanton Schwyz. Bei den als schützenswert eingestuft Gebieten handelt es sich um die Schlüsselgebiete, die vom Kanton aufgrund ihrer charakteristischen und besonders vielfältigen, wertvollen und schönen Landschaftstypen ausgeschieden wurden (*Landschaftskonzeption Kanton Schwyz. Version vom Februar 2019*). Die Karte 3.1 zeigt die Lage des Untersuchungsgebietes in der Schweiz und die der Schlüsselgebiete im Kanton zusammen mit einigen der häufig genannten Ortschaften in den Texten.



ABBILDUNG 3.1: Übersichtskarte Untersuchungsgebiet Kanton Schwyz.

Zeitlich bilden Texte von 2012 bis 2020 das Korpus, wobei die Mehrheit aus den Jahren 2017 bis 2020 und die Landschaftskonzeption aus dem Jahr 2019 stammt.

Inhaltlich werden Landschaftsbegriffe (LSB) betrachtet (s. Kapitel 4.4). Es gibt darin eine Tendenz zur biologischen, geomorphologischen Seite der Landschaft und weg von den subjektiven Sinnen und Gefühlen des Beobachters oder der Beobachterin. Dies einerseits, weil die Grundlagen meiner Liste an LSB auf zwei Arbeiten im alpinen Bereich beruht (Huldi, 2015, Derungs und Purves, 2014). Andererseits ist der Bereich der fassbaren Objekte und Aktivitäten in der Landschaft einfacher mit den Methoden des NLP und dem Machine Learning zu finden, als weniger fassbare Wörter zur Atmosphäre oder zu den Sinnen, denen nicht direkt ein Kontext zur Landschaft zuzuordnen ist, da sich diese oft nicht an einem benannten Ort festmachen (Garcia-Martin u. a., 2017).

Schlussendlich soll mit dieser Arbeit nicht endgültig bestimmt werden, wie Nicht-Expertinnen die Schwyzer Landschaft wahrnehmen. Stattdessen soll eine skalierbare Möglichkeit aufgezeigt werden, wie diese Stimmen in Landschaftskonzepte integriert werden können und die Arbeit soll eine Machbarkeitsstudie für künftige Landschaftskonzepte in der Schweiz darstellen.

## Kapitel 4

# Daten

Die Grundlage dieser Arbeit bilden natürlich die Texte, in denen die Landschaft des Kanton Schwyz beschrieben werden, das so genannte Korpus. Daneben werden zwei weitere Datenquellen verwendet; nämlich die Landschaftskonzeption des Kanton Schwyz und ein Gazetteer, ein Ortsverzeichnis, welcher das Interessengebiet, den Kanton Schwyz abdeckt.

### 4.1 Texte

Die Suche nach Texten kann sowohl manuell, wie auch automatisch erfolgen. Manuell mit gezielter Suche nach relevanten Texten, automatisch z.B. durch eine Internetsuche mithilfe bestimmter Suchwörter aus Toponymen, Landschaftstypen und -elementen (Acheson, Wartmann und Purves, 2018). In dieser Arbeit werden die Texte manuell ausgewählt anhand von Herausgebern, von denen sowohl thematisch wie auch geographisch relevante Texte zu erwarten sind, ohne dass viel gefiltert werden müsste. Die Aussagen, die das Korpus aufbauen, sollen die Landschaftswahrnehmung einer möglichst breiten Bevölkerung abdecken. Meine eingeschränkten Kenntnisse über den Kanton Schwyz haben für Auswahl der Texte nicht gereicht, weshalb ich Personen<sup>1</sup>, die enger mit der Region vertraut sind, nach passenden Quellen befragt habe. Mögliche Herausgeber von Texten lassen sich in vier Bereiche einteilen:

- Zeitungen und Zeitschriften: Bote der Urschweiz, Höfner Volksblatt, March Anzeiger und das Y-Magazin
- Tourismus: Schwyz Tourismus, Tripadvisor
- Freizeitseiten: Hikr.org, Wegwandern.ch, Wandersite.ch, Gipfelbuch.ch
- (Umwelt-)Vereine: Naturforschende Gesellschaft SZ, Schwyzer Umweltrat, Landschaftsschutzverband Vierwaldstättersee

Die Auswahl fiel schliesslich auf die Kategorie *Zeitungen und Zeitschriften* mit den Exemplaren Bote der Urschweiz (*Bote der Urschweiz Januar 2017 - April 2020*) und

<sup>1</sup>Dr. Karin Schwitter, bis 2018 im Schwyzer Kantonsrat; Dr. Raimund Rodewald, Geschäftsleiter der Stiftung Landschaftsschutz Schweiz; Remo Bianchi, Amt für Natur, Jagd und Fischerei; Valentin Kessler, Staatsarchivar des Kanton Schwyz. Die Vorschläge sind in Anhang A aufgelistet.

Y-Magazin (*Y-Magazin Sommer 2012 - Winter 2020*). Für beide war es möglich, umfassende und digitale Ausgaben zu erhalten, während in allen anderen Kategorien eine feinere und aufwändigere Suche notwendig gewesen wäre, um ein gleichermaßen grosses Korpus aufzubauen.

Der Bote der Urschweiz, kurz der Bote, ist die grösste Zeitung im Kanton Schwyz und deckt mit seinen Berichten das ganze Kantonsgebiet ab, obwohl ein Fokus auf der inneren Kantonshälfte liegt. Es stehen alle Ausgaben der Zeitung im Zeitraum vom 03.01.2017 bis zum 17.04.2020 zu Verfügung. Das sind über 5 Millionen Sätze. Der Startpunkt wurde mit Anfang 2017 bewusst so gewählt, weil der neue Richtplan des Kantons ab Mai 2017 vom Bund genehmigt und ein Jahr früher vom Regierungsrat erlassen wurde (*Richtplan des Kanton Schwyz. Richtplantext 2020*). Die zweite Sammlung an Texten stammt aus dem Y-Magazin, welches vierteljährlich seit dem Sommer 2012 erscheint. In diesem Magazin, herausgegeben vom Amt für Wirtschaft, wird, was den Kanton so „vielseitig, spannend und lebenswert“ macht. Dazu gehören neben den Menschen, vor allem auch die herausstehende Landschaft (*Y-Magazin Nr. 1 Sommer 2012*, p. 5). Bis im Winter 2020 stehen 32 Magazine à je circa 85 Seiten zur Verfügung, was knapp 50'000 Sätzen oder weniger als 1% des Umfangs des Boten entspricht (s. Abb. 4.1).

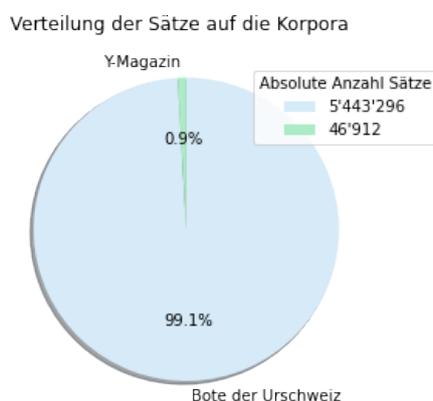


ABBILDUNG 4.1: Anteile der beiden Korpora an allen Sätzen.

## 4.2 Landschaftskonzeption

Gegenüber diesen beiden Korpora, die als Stimme der breiten Bevölkerung dienen, steht die bereits erwähnte Landschaftskonzeption des Kanton Schwyz (s. Kap. 2.1.3). Dieses Dokument wurde, anders als ein Grossteil von Artikeln in Zeitungen und Magazinen, von Fachpersonen verfasst. Insgesamt waren 11 Fachpersonen beteiligt, 3 aus der Stiftung Landschaftsschutz Schweiz und 8 aus diversen Kantonsdepartementen. So wurde das Dokument mit viel Fachverständnis verfasst, wieder spiegelt aber nicht zwingend die Wahrnehmung und Bedürfnisse einer breiten Bevölkerung. Das Dokument, herausgegeben vom Amt für Natur, Jagd und Fischerei,

ist noch nicht offiziell genehmigt worden. Zur Zeit der Arbeit steht die Arbeitsausgabe vom Februar 2019 zur Verfügung. Wie oben erwähnt, interessieren in der Landschaftskonzeption besonders die Schlüsselgebiete. In der Version vom Februar 2019 sind das sechs Gebiete, die:

1. „einen charakteristischen Landschaftstyp des Kantons Schwyz repräsentieren“ oder
2. „ausgeprägte Qualitäten mehrerer überlagerter Landschaftstypen aufweisen“ (*Landschaftskonzeption Kanton Schwyz. Version vom Februar 2019*, p. 5).

Die Schlüsselgebiete werden je auf zwei bis drei Seiten mit ihren Charakteristika und Qualitäten beschrieben (s. Abb. 4.2). Die Karte 4.3 zeigt, die Landschaftstypen im Kanton Schwyz und wo die Schlüsselgebiete liegen (*Öffentliche Mitwirkung Landschaftskonzeption Kanton Schwyz. 2019*).

2.2 Objektblätter

2.2.1 Mythen

Das kantonale Schlüsselgebiet „Mythen“ ist ein wichtiges **Naherholungsgebiet** für die Bevölkerung des Schwyzer Talkessels sowie die wohl wichtigste **Symbol- und Identifikationslandschaft** des Kantons. Erwähnenswert ist der hohe naturräumliche Wert dank des kantonalen Pflanzenschutzgebietes „Mythengebiet“ sowie des nationalen Jagdbanngbietes „Mythen“.



Perimeter	Mythen, Mythenbann, Haggenegg
Beschreibung	<p>Prägend für diese Landschaft sind der grosse und kleine Mythen, welche als schroffe Klippen auf der sanften Fylsch-Landschaft zu liegen kommen. Diese Schwyzer Wahrzeichen bieten nebst ihrer identitätsstiftenden Funktion auch einen hohen Naturelebniswert.</p> <p>Die bewaldete steil aufsteigende Bergflanke mit der eindrücklichen Bergkulisse erzeugt eine landschaftlich einprägsame Wirkung. Spirituelle Wirkung erhält diese Region, da sie einen Teilabschnitt des Jakobsweges beherrbergt und mit dem Haggenegg-Pass auch dessen höchstgelegenen Übergang nördlich der Pyrenäen aufweist. Teile des Jakobsweges sind auch die Kapellen Haggenepp und Muttergottes, welche im Kantonalen Inventar geschützter Bauten und Objekte (KIGBO) vertreten sind. An offenen Hanglagen ist das (extensiv bewirtschaftete) Kulturland von traditionellen Einzelhöfen durchsetzt und erzeugt, gemeinsam mit den kullissenartigen Wald- und Felspartien, ein intaktes (mosaikartiges) Landschaftsbild. Das Gebiet hat eine hohe Bedeutung für die Naherholung. Bereits im Richtplan 1986 wurde dem hohen landschaftsästhetischen Wert dieser Region Rechnung getragen, indem sie als Kernbereich eines Natur- und Landschaftsschutzinventares ausgewiesen wurde. Zentraler Bestandteil der Landschaftsan-</p>

ABBILDUNG 4.2: Ausschnitt aus der Landschaftskonzeption Kanton Schwyz. Beschreibung des Schlüsselgebietes Mythen.

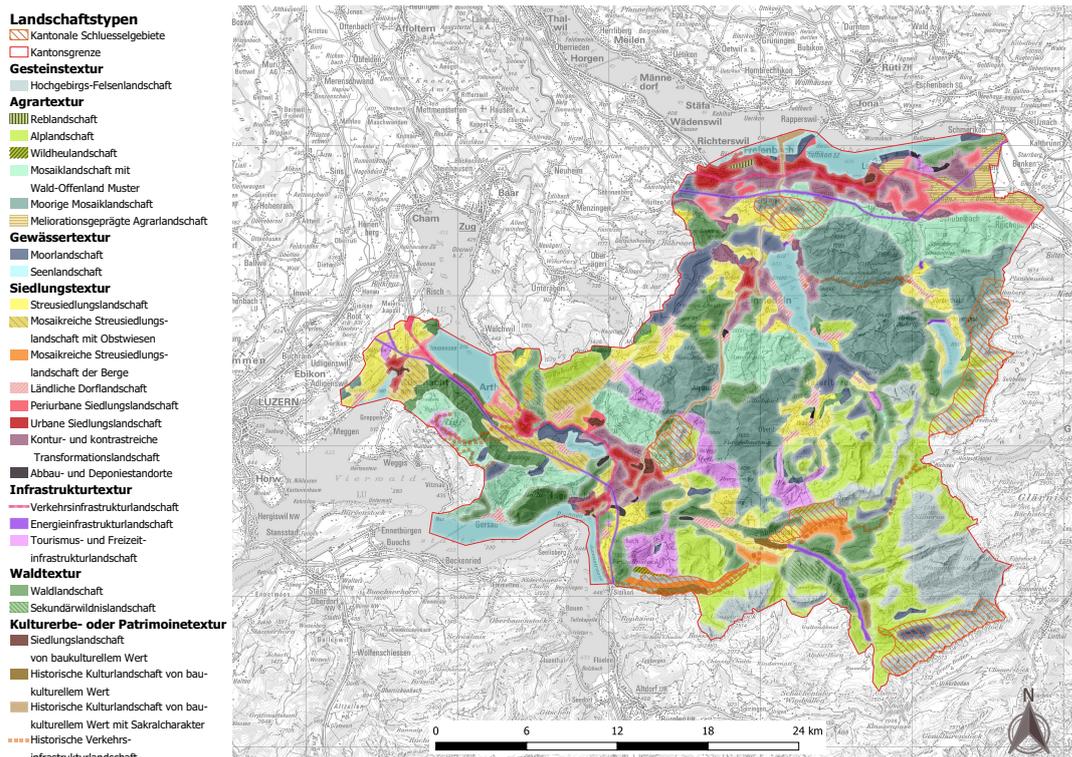


ABBILDUNG 4.3: Plan Schlüsselgebiete aus der Landschaftskonzeption Kanton Schwyz.

### 4.3 Gazetteer

Swissnames3D ist die umfangreichste Sammlung georeferenzierter Namen in der Schweiz und gehört zu den geographisch genaueren Namensverzeichnissen (Derungs und Purves, 2014, Landestopografie swisstopo, 2020). Es enthält alle Objekte, die im Topografischen Landschaftsmodell einen Namen besitzen (*Produktinformation swissNAMES3D 2020*) und deckt Kartenbeschriftungen im Massstab 1:25 000 bis 1:500 000 ab. Auf das Kantonsgebiet Schwyz und die deutsche Sprache begrenzt, gibt es:

- 6150 Punktobjekte. Dabei handelt es sich bei 68% um Flurnamen, die kleinste Fleckchen (z.B. Bergliwald, ein Waldstück in der Gemeinde Muoatathal) bis Städte (z.B. Einsiedeln oder Brunnen) beschreiben können.
- 372 Linienobjekte. Davon sind 60% Fliessgewässer, also Objekte mit hohem Landschaftsbezug. Es folgen Strassennamen und Skilifte, wobei letztere ebenfalls landschaftsprägend sind.
- 2254 Polygonobjekte. Davon benennen 54% Siedlungen, gefolgt von Quartieren und Schularealen. Objekte mit Landschaftsbezug machen hier nur gerade 18% aus (z.B. Täler, Massivnamen und Seeinseln).

In allen drei Datensätzen kommt geo/geo Mehrdeutigkeit vor, das heisst gut 25% der Objekte teilen sich den gleichen Namen. Wie mit dieser Unsicherheit umgegangen wird, behandelt das Kapitel 5.1.5. Damit Instanzen aus dem Attribut *NAME* im Gazetteer später automatisch im Text gefunden werden können, mussten punktuelle Anpassungen vorgenommen werden. Oft handelt es sich dabei um umgangssprachliche Namen, die sich von den offiziellen unterscheiden. Dazu gehören folgende Änderungen:

- Hoch Ybrig wurde ergänzt als häufiger Name für das Gebiet Iberig.
- Grosser und Kleiner Mythen: Im Gazetteer kamen nur die offiziellen Namen für die beiden berühmten Gipfel vor. Umgangssprachlich wird aber oft einfach von den Mythen oder sogar nur vom Mythen geschrieben.
- Im Norden des Kantons gibt es drei kleine Dörfer, die ausgerechnet Rigi, Brunnen und Zug heissen. Diese wurden aus dem Gazetteer bzw. den Resultaten gelöscht, weil überwiegend das sehr viel bekanntere Massiv, bzw. die Städte gemeint sind.
- Der Gipfel Rot Turen wird im Mundart Rot Turm genannt.
- Ebenfalls ist der Pass Furggelen im Mundart als Furggeli bekannt.

Diese Anpassungen passierten nur, weil die Fälle durch Stichproben aufgefallen sind. Flächendeckender könnte das mit einem Idiotikongazetteer wie dem Schwyzer Flurnamenbuch gelöst werden (Weibel und Hug, 2012). Zusätzlich wurden alle Kantonsendungen SZ gelöscht (z.B. Lachen SZ zu Lachen), da diese Unterscheidung in journalistischen Texten üblicherweise nicht gemacht wird.

Gewisse Toponyme müssen vollständig ignoriert werden. Bei diesen Toponymen handelt es sich um:

- Flurnamen oder sonstige Landschaftsbeschreibungen (z.B. Alp, Dorf, Tal, Post, Port, Quai, Steg, Stock, Strich, Tor)
- Personennamen und andere allgemeine Worte (z.B. Peter, Gisler, Suter, Vogler, Vogt, Länder, Kreuz, Gross, Hand, Paradies)
- Toponyme mit grösseren und bekannteren Entsprechungen in anderen Kantonen (z.B. Adelboden (BE vs. SZ), Albis (ZH vs. SZ), Saas (Fee) (VS vs. SZ), Wallis als Flurname)

#### 4.4 Liste der Landschaftsbegriffe

Die Grundlage für meine Sammlung an LSB bilden die LSB-Liste in Text+Berg von Derungs und Purves, 2014, extrahiert aus 10'000 digitalisierten SAC-Artikeln, und

eine Liste von Haldi, 2015, die im Rahmen einer Masterarbeit entstanden ist und aus einem Korpus aus Texten über Bergtouren stammt.

Diese beiden Listen wurden durch die Objektbeschreibungen aus der Schwyzer Landschaftskonzeption ergänzt, sowie mit LSB aus meinem Korpus. Genaueres zur Evolution der LSB-Liste findet sich im Kapitel 5.1.1.

Schlussendlich beinhaltet die Liste über 1000 LSB. Dabei kommen aber auch gleiche Wörter mit unterschiedlichen Endungen vor. Die Liste ist in vier Kategorien unterteilt (s. Abb. 4.4). Die grösste Kategorie beinhaltet Wörter, die unter allen Umständen als LSB gezählt werden. In der zweiten sind Wörter, die je nach Kontext die Landschaft beschreiben. Wenn ein Wort aus der dritten Kategorie (Verben und Adjektive) zusammen mit einem aus der zweiten auftritt, dann kann davon ausgegangen werden, dass es sich um eine Landschaftsbeschreibung handelt. In der vierten Kategorie sind Teilwörter. Damit nicht alle Kombinationsmöglichkeiten mit -landschaft oder Hügel- oder Alp- in die Liste aufgenommen werden müssen, werden Teilwörter, denen ich immer einen Landschaftskontext unterstelle, hier extra aufgeführt. Zusammengesetzte Wörter mit diesen Teilwörtern zählen ebenfalls als LSB.

KATEGORIE 1	KATEGORIE 2	KATEGORIE 3	KATEGORIE 4
<b>Definitive Begriffe</b>	<b>Begriffe im LS-Kontext</b>	<b>Kontextwörter</b>	<b>Komposita</b>
Abendglühen	Abendstunden	erleben	Alp
Abendgrauen	Abgrund	abwärts	Älp
Abendsonne	Absteigen	aufwärts	Berg
Abhang	Abstieg	biken	Birken
Allmeind	Abzweigung	fotografieren	Buchen
Allmend	Alpenschaf	landwirtschaftlich	Damm
Alp	Alpenschafe	neblig	Dämme
Alpen	Alpenschwein	regnerisch	Dunst
Alpenrose	Alpenschweine	schön	ebene
Alpenüberquerung	Alphorn	sonnig	Eis
Alphütte	Alpine	staunen	Fels
Äplerfest	Alpines	steil	Fluss
Alpweg	Alpines	talwärts	graben
Alpweide	Alpinismus	touristisch	gräben
Anhöhe	Alpinist	wandern	Hafen
Artenvielfalt	Alpinisten	weit	Hügel
Aue	Alpinistin	wunderschön	Kanal
Auerhahn	Alpkäse	bestaunen	Landschaft
Ausblick	Alpkäser	pittoresk	landwirtschaftliche/e +
Aussicht	Alpkäsetal	idyllisch	Nebel

ABBILDUNG 4.4: Ausschnitt aus der LSB-Liste.

Die komplette Liste findet sich im Anhang B.1.

## Kapitel 5

# Methoden

Die Methoden bilden eine Kombination aus den Bereichen des Geographic Information Retrieval (GIR) und dem Natural Language Processing (NLP) mit gezielten Anpassungen für die Landschaftswahrnehmung. Das Flussdiagramm in Abbildung 5.1 beschreibt den ganzen Arbeitsablauf der Methoden.

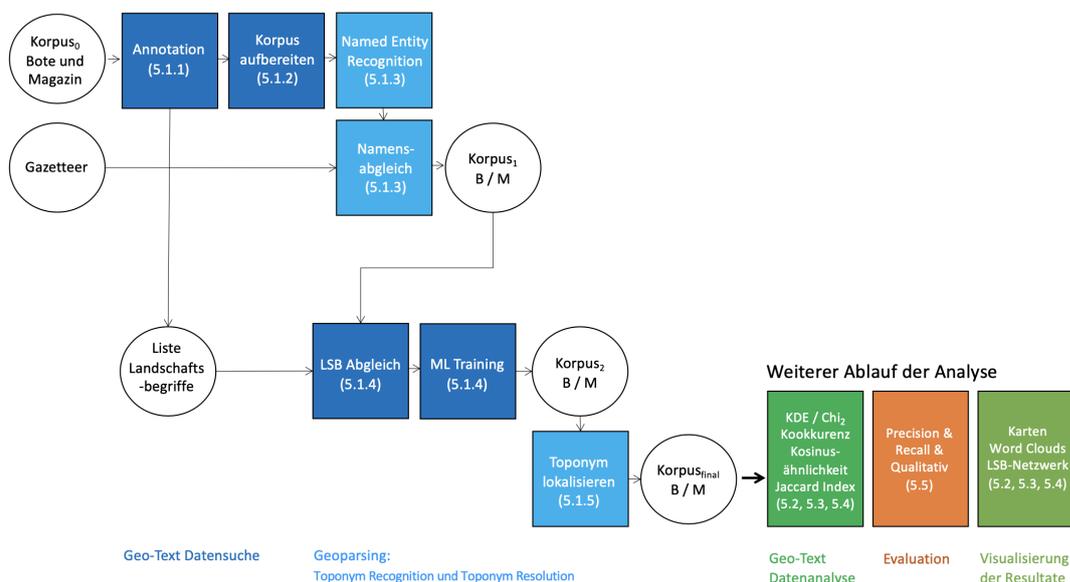


ABBILDUNG 5.1: Der Arbeitsablauf der Datenaufbereitung folgt dem verallgemeinerten Ablauf der Analyse von geographischen Textdaten in Hu, 2018. Kapitelnummern in Klammern angegeben.

Zufällig ausgewählte Artikel der Korpora<sub>0</sub> Bote der Urschweiz und Y-Magazin werden manuell annotiert, um ein besseres Verständnis für die Verwendung von Landschaftsbegriffen in diesen Texten zu gewinnen. Die Annotation dient auch als Grundlage für die Evaluation der angewandten Methoden. Dann wird der Korpus, der von PDF-Dateien in Textdaten umgewandelt wurde, aufgeräumt und mit NLP-Methoden in der gewünschte Datenstruktur gespeichert, hier in Sätzen mit mindestens einem Wort, welches als Örtlichkeit erkannt wurde. Diese Ortswörter werden mit den Toponymen im Gazetteer verglichen und die Sätze mit übereinstimmenden Toponymen bilden den Korpus<sub>1</sub>. In diesen Sätzen wird nach LSB gesucht und die

Sätze mit einem LSB werden mithilfe eines Machine Learning Algorithmus in tatsächlich landschaftsrelevant oder nicht landschaftsrelevant eingeteilt. Diese Sätze bilden den Korpus<sub>2</sub>. Den Toponymen in diesen Sätzen müssen noch die korrekten Koordinaten zugeordnet werden, bevor der finale Korpus aus Aussagen über die Landschaft im Kanton Schwyz steht. Diese Sätze bilden die Grundlage für die räumliche und inhaltliche Visualisierung und Analyse.

## 5.1 Aufbereitung der Daten

### 5.1.1 Annotation

Um ein besseres Verständnis für die Textdaten zu erhalten, ist es üblich, eine Auswahl davon zu annotieren (Chesnokova und Purves, 2018b). Mithilfe der Annotation werden die Struktur und Eigenheiten der Texte und der Sprache aufgedeckt, denen mit Regeln begegnet werden kann.

Für die Annotation von 60 Dokumenten wurden händisch die Toponyme in Schlüsselgebieten aus der Landschaftskonzeption aufgelistet und nach diesen im ganzen Korpus gesucht. Pro Toponym wurde gespeichert, an welcher Stelle es vorkam und die 100 Zeichen vorher und nachher wurden ausgegeben, um beurteilen zu können, ob es sich tatsächlich um landschaftsrelevante Artikel handelt. Insgesamt 60 Texte, 35 aus dem Boten und 25 aus dem Y-Magazin, wurden automatisch ausgewählt, wenn sie ein Toponym aus einem der Schlüsselgebiete enthielten. Diese Texte wurden annotiert, um ein besseres Verständnis für die Daten zu erhalten. Gekennzeichnet wurden Sätze, die als landschaftsrelevant eingestuft wurden, sowie die Landschaftsbegriffe, die in den Sätzen vorkamen. Ausserdem müssen die Sätze durch ein Toponym im Kanton Schwyz lokalisierbar sein.

Um die Annotation reproduzierbar zu machen, müssen Regeln festgelegt werden, denen ein anderer Annotator folgen könnte und zu einer möglichst gleichen Auswahl käme. Um zu testen, wie nachvollziehbar die Annotationsregeln formuliert sind, sollten weitere Personen eine Stichprobe an Texten annotieren und eine Interannotatorübereinstimmung berechnet werden (Chesnokova und Purves, 2018b). Das wurde im Umfang dieser Arbeit nicht gemacht.

Annotiert wurden Aussagen, d.h. Sätze mit einem Landschaftsthema, Landschaftsbegriffe und Toponyme, die im Kanton Schwyz liegen. Die Annotation erfolgte auf folgenden Regeln, die im Lauf der Arbeit verfeinert wurden:

#### **Landschaftsrelevante Aussage**

*Regel:* Die Einheit einer Aussage bildet ein Satz. Der Satz soll sowohl ein Toponym als auch einen LSB beinhalten und - das ist die Schwierigkeit - er soll landschaftsrelevant sein.

*Umsetzung:* Wie die Landschaftsrelevanz automatisch bestimmt wurde, wird im Teil 5.1.4 erklärt.

### Landschaftsbegriff

*Regel:* Als LSB sollen nur Nomen zählen. Verben, Adjektive und Präpositionen können aber helfen, einen Satz als landschaftsrelevant zu verstehen; z.B. wandern, steil, hinauf. Substantivierte Verben hingegen zählen als LSB (z.B. das Hinaufsteigen). Laut Smith und Mark, 2003 nutzen Laien überwiegend Objekte, um die Realität zu beschreiben und diese entsprechen in der Sprache den Nomen. Bei der Annotation wird ein spezielles Augenmerk auf LSB gelegt, die noch nicht aus den Listen von Derungs und Purves, 2014 und Huldi, 2015 bekannt sind. Dazu gehören Wörter, die ohne Kontext nicht direkt mit Landschaft assoziiert werden; z.B. Strasse (auf einer Wanderung), Autobahn (Sicht ins Tal hinunter), Aufweitung oder Verbauung (eines Baches), Gärten (als wichtiger Teil von Alltagslandschaften), Identität und Heimat (beim Antreffen von Kühen mit Glocken) und dann konkreter Alphorn und Schwyzerörgeli, Klima (körperliche Wahrnehmung der Umgebung), Licht (Beschreibung einer Atmosphäre auf einem Spaziergang), Strom (Wasser, nicht Elektrizität), Gasthaus und Unwetterereignis (beides Dinge, die man in der Freizeit draussen antreffen könnte und die die Wahrnehmung der Landschaft prägen).

Die LSB lassen sich in verschiedene Kategorien sortieren, inspiriert von kulturellen Ökosystemdienstleistungen (Conrad und Hilchey, 2011, Stephenson, 2008):

- **Elemente:** beinhalten natürliche Landformen, Kulturbauten.
- **Tätigkeiten:** beinhalten Bräuche, Freizeitaktivitäten, Gewohnheiten, Verhalten.
- **Atmosphäre:** beinhalten Gefühle, Beziehungen zu Orten, immaterielle Qualitäten und visuelle Ästhetik.

Diese Einteilung erwies sich nicht als nützlich und wurde daher nicht weiter verwendet. Um eine Übersicht über die Kombinationen von LSB zu erhalten, werden stattdessen Word Clouds und Netzwerke verwendet (s. Kap. 5.3).

*Umsetzung:* Obwohl für als Annotator klar ist, nach welchem Kontext gesucht wird, sind nicht alle Wörter (z.B. Strasse, Autobahn) semantisch eindeutig und brauchen daher später einen angepassten Umgang im GIR (Derungs und Purves, 2014). Wie das umgesetzt wurde, wird im Teil 5.1.4 erklärt.

### Toponym im Untersuchungsgebiet

*Regel:* Ortsnamen sollen gefunden werden, wenn sie irgendwie mit dem Untersuchungsgebiet interagieren. Das heisst, es zählen nicht nur geographische Bezeichnungen, die sich ganz im Kantonsgebiet befinden (z.B. der Lauerzersee), sondern auch Gebiete, die nur teils im Kanton liegen (z.B. der Zugersee) und auch Gebiete, die den Kanton grossräumig überlappen (z.B. die Voralpen). Auch Subtypen von Toponymen zählen (Van, Tien Nguyen and Mauro Gaio, and Ludovic Moncla, 2013). Das bedeutet, nicht nur reine Toponyme (Schwyz), sondern auch die Beschreibung eines Toponyms (Bahnhof Schwyz, Restaurant Hirschen) sollen genutzt werden. Ausserdem sind auch Adjektive und Adverben erlaubt, da auch diese auf einen Ort

und dessen Qualitäten verweisen (z.B. Riemenstaldener Tal, Glattalps Kartsgebiete).

*Umsetzung:* Wie das umgesetzt wurde, wird im Teil 5.1.3 erklärt.

### 5.1.2 Korpus aufbereiten

Die PDFs wurden mithilfe des Pythonmodul *pdftotext* in Textdateien transformiert. Bevor mit der eigentlichen Geokodierung und der inhaltlichen Analyse fortgefahren werden kann, müssen die Textdatei zur weiteren Verarbeitung aufbereitet werden. Da meine Analyse auf Aussagen beruht, die ich als einen Satz umfassend definiere, wird der Text jeweils bei folgenden Satzzeichen gesplittet: Punkt (.), Fragezeichen (?), Ausrufzeichen (!). Bei häufige Abkürzungen wie Nr. für Nummer, z.B./zB. für zum Beispiel und St. für Sankt werden die Punkte entfernt, damit sie nicht irrtümlich als Satzende gefunden werden. Danach werden die Sätze durch eine Pipeline von NLP-Schritten geführt. Hierzu wird die Pythonbibliothek *spacy* genutzt (Honnibal u. a., 2020). Die deutschen Modelle wurden auf annotierten Texten von Wikipedia und dem Tiger Korpus der Universität Stuttgart trainiert (Nothman u. a., 2013, TIGER-Project, 2003).

Der Text wird in Tokens verwandelt, Part-of-Speech Tags werden verteilt und - für diese Arbeit am wichtigsten - es geschieht eine Named Entity Recognition (NER). In der Tokenisierung wird der Text aufgeteilt in einzelne Wörter und Satzzeichen. Das geschieht nicht naiv, sondern ist sprachspezifisch. So weiss das deutsche Modell, das „z.B.“ nicht in Buchstaben und Satzzeichen getrennt werden sollen. Danach sollen den Tokens ihre linguistische Bedeutung im Text zugewiesen werden (Subjekt, Adjektiv, Akkusativobjekt etc.). Eine Named Entity ist alles, was einen Namen hat: Personen, Organisationen, Marken, Länder. In meinem Fall bin ich an allen Objekten interessiert, die vom Modell als Standort (LOC) identifiziert werden (Honnibal u. a., 2020). Sätze, die laut NER ein LOC-Wort enthalten, werden später mit den Toponymen im Gazetteer abgeglichen.

Weiter müssen die Sätze von Artefakten bereinigt werden, die bei der Konvertierung von PDF zu Text entstanden sind. Dazu gehört das Entfernen von Spezialzeichen, einzelstehenden Buchstaben und so genannten Stopwörtern (z.B. *ein* oder *die*).

Im üblichen NLP-Ablauf wäre das Reduzieren der Wörter auf ihre Lemmata ein wichtiger Schritt (Plisson, Lavrac, Mladenec u. a., 2004). Die Worte werden dabei nicht auf ihren Stamm ohne Endung gekürzt, sondern eher auf ihre normalisierte Form gebracht. Das wäre besonders im Deutsch mit den vielen Beugungen nützlich. So würde das Bächlein nicht zum Bäch, sondern zum Bach und Schlitteln nicht zum Schlittel sondern Schlitten. Der Lemmatizer von *spacy* beruht auf Tiger Stuttgart und funktioniert auf Deutsch nicht befriedigend. Die Precision für den Lemmatizer performt tatsächlich am schlechtesten im Vergleich zu den anderen Funktionalitäten des Modells (0.73 vs. im Schnitt 0.92) (Honnibal u. a., 2020). Mit *spacy* könnte ich

einen eigenen Lemmatizer trainieren. Da dies nicht Teil meiner Forschung ist, hätte das allerdings den Rahmen dieser Arbeit gesprengt.

Eine weitere Möglichkeit, neben einem trainierten Lemmatizer, wäre eine eigene Liste von Lemmas, bzw. Stämmen zu erstellen. Wegen der Möglichkeit von Komposita in der deutschen Sprache müsste diese Liste müsste allerdings eine unmessbare Menge an Kombinationen von Teilwörtern enthalten und somit keine umsetzbare Alternative. Deshalb habe ich mich dafür entschieden neben einer sehr umfangreichen LSB-Liste auch mit Regex zu arbeiten. Ein Lemmatizer hätte die Berechnung aber schneller gemacht, da Regex eher eine Brute-Force-Methode ist, bei der für jedes einzelne Wort mehrere Regex geprüft werden müssen.

### 5.1.3 Toponym Recognition

Um Texte im Raum zu verankern, müssen als erstes mögliche Kandidaten für Georeferenzen im Text gefunden werden. Dieser Schritt wird Geoparsing oder Toponym Recognition genannt. Die semantische Mehrdeutigkeit, dass Ortsnamen auch nicht geographische Bedeutungen haben können, stellt eine Schwierigkeit dar (Purves u. a., 2018). Ein gutes Beispiel für diese Geo/Non-Geo-Ambiguität ist der Ort und das Objekt *Brunnen*.

Die Toponyme im Text zu finden, läuft in drei Schritten ab.

1. Named Entity Recognition (NER)
2. Abgleich mit Gazetteer
3. Anwendung von Regex

#### Named Entity Recognition

Nach der NER von spacy wird nur mit Sätzen weitergefahren, denen ein LOC Tag vergeben wurde, die also laut NER einen Ortsnamen enthalten. Das ist eine enorme Optimierung gegenüber der Variante, in der alle Sätze mit Regex auf Matches mit dem Gazetteer durchsucht werden müssten. Da einige Wörter sowohl Toponyme als auch verbreitete Familiennamen (z.B. Suter, Steiner) oder allgemeine Flurbeschreibungen (z.B. Tal, Ort) sein können, werden diese automatisch als mögliche Toponyme ausgeschlossen.

#### Abgleich mit Gazetteer und Anwendung von Regex

In den gefilterten Sätzen wird nach ganzen Matches mit dem Gazetteer gesucht. Ganze Matches sind nötig, um nicht versehentlich kurze Namen in längeren Worten zu finden (wie Akti in Aktien oder Farb in Farben). Dabei werden nicht nur ganz genaue Übereinstimmungen erlaubt, stattdessen sollen diverse Endungen möglich sein (z.B. Ried+er, Lachen+s). Beim Annotieren wurden alle mit einem Toponym vorkommenden Endungen notiert. Deshalb kommen auch Endungen für Adjektive vor

(z.B. isch, ig). Tatsächlich kommen in den Aussagen kaum solche Varianten von Toponymen vor, entweder weil sie nicht als Orte erkannt wurden im NER oder weil sie so selten sind. Abgesehen von den Endungen wird nicht auf variable Rechtschreibung geachtet. Gewisse verschieden gebräuchliche Namen wurden bereits manuell dem Gazetteer hinzugefügt (s. Kap. 4.3). Für jedes Toponym werden alle möglichen Kandidaten gespeichert, welche später disambiguiert werden (s. Kap. 5.1.5).

Aus diesem Schritt ergibt sich ein Datensammlung aus Sätzen und den darin gefundenen Toponyme mit ihren IDs, der Objekt- und Geometrieart und den Koordinaten. Für Toponyme, die linienförmige Objekte benennen, wurde der Mittelpunkt des Linienabschnitts gespeichert, für Polygone der Mittelpunkt innerhalb der Fläche.

#### **5.1.4 Landschaftsrelevanz bestimmen**

Aus diesen Sätzen mit Ortsbezug sollen jetzt jene gefiltert werden, die um Landschaft handeln. Das passiert in zwei Schritten:

1. LSB mit Regeln suchen
2. Landschaftskontext mit Machine Learning bestimmen

##### **LSB mit Regeln suchen**

In einem ersten Schritt wird ähnlich wie bei den Toponymen mithilfe von Regex nach LSB (s. 4.4) inklusive verschiedener Endungen gesucht, die anhand der annotierten Dokumente ermittelt wurden. Dann wurde noch automatisch ein Fünftel aller Dokumente durchsucht (entspricht etwa 6500 Seiten des Boten und 6 Ausgaben des Y-Magazin), um zu sehen, ob Endungen vergessen wurden. Daraus ergaben sich weitere 2000 Wörter, die ich schnell händisch durchgehen konnte und entscheiden konnte, ob es sich dabei um Wörter handelt, die in die LSB-Liste aufgenommen werden sollten oder nicht. Unregelmässige Deklinationen und Pluralformen werden händisch ergänzt. Da diese Formen nicht mit Regex gesucht werden, verschnellert sich die Berechnungszeit deutlich (ca. Faktor 10) und stellt sicher, dass nicht nur nach Endungen und Komposita gesucht wird, die bei der Annotation der ersten 50 Dokumente aufgefallen sind. Als weitere Regel wird nach eventuellen LSB gesucht (vgl. Kategorie 2 in Kap. 4.4). Das sind Begriffe, die für sich genommen nicht eindeutig in einen Landschaftskontext gehören. Kombiniert mit einer Reihe von landschaftstypischen Adjektiven und Verben (wie talwärts, wandern, steil, idyllisch) hingegen, werden Sätze mit solchen Begriffen dennoch gespeichert.

##### **Landschaftskontext mit Machine Learning bestimmen**

Die Sätze, in denen sowohl ein oder mehrere Toponyme als auch ein oder mehrere LSB gefunden wurden, sollen nun auf ihren Landschaftskontext geprüft werden. Ein

Modell soll erkennen, ob in einem Satz Landschaft thematisiert wird oder nicht. Dazu wurde ein Random Forest Modell aus der Pythonbibliothek *scikit-learn* trainiert (Pedregosa u. a., 2011). Random Forests kombinieren mehrere Entscheidungsbäume, die auf zufällig gewählte Subsets der Daten passen. Die Kombination soll im Schnitt die Voraussage verbessern und zu möglichst wenig Überanpassung führen (Criminisi, Shotton, Konukoglu u. a., 2012, Pedregosa u. a., 2011). Dieser Typ Klassifizierer zeigt laut Onan, Korukoğlu und Bulut, 2016 gute Resultate bei Textdaten.

Die Trainings- und Testdatensätze werden aus 400 der bisher gefilterten Sätzen erstellt und in echt positive (d.h. Aussagen, die tatsächlich um Landschaft gehen) und falsch positive (d.h. Sätze mit einem LSB, deren Aussage aber nichts mit Landschaft zu tun haben) Klassen eingeteilt. Einige Beispiele für falsch positive Aussagen sind:

- Beispiele für Metaphern: *Jeder Künstler muss bei seiner Lebenswanderung riskieren dass ein Weg eine Sackgasse sein kann auf der er nicht weiterkommt.* und *Es kann nicht sein, dass die ganze Zentralschweiz umgegraben oder mit einem Wald aus Schildern versehen wird.*
- Beispiel für eine Wettervorhersage: *Morgen überwiegen die Wolken, es regnet zeitweise, Schnee gibt es ab 700 bis 900 Metern.*
- Beispiele für Werbung: *Tal der Dinosaurier 041 726 10 03 14.15/17.00/20.00 D Premiere 12/10 J kinozug.ch Maxx 7 Seetalplatz 1 Maria Magdalena kitag.com* und *Mietzins Fr. 1150.–, exkl. NK Zu vermieten in Steinen mit Seesicht ab 1. Oktober*
- Beispiele ohne Landschaftskontext: *Am Karfreitag dürfte es wenn schon einen Stau Richtung Einsiedeln gegeben haben: An diesem ersten grossen Pilgertag des Jahres kamen die Menschen gleich zu Tausenden zu uns.* und *Muotathal 30 Uhr im Restaurant Alpenblick in Muotathal statt*

Schlussendlich handelt es sich um 483 Sätze, die in mehreren Iterationen zusammen gekommen sind. Das Verhältnis zwischen positiven (212) und negativen (271) Aussagen ist hier fast 1:1, was nicht dem Verhältnis der realen Inputdaten entspricht, welche gemäss meiner Annotationen noch viel weniger positive Aussagen enthalten. Laut (Acheson, Volpi und Purves, 2020) gibt es keine übereinstimmende Meinung in der Literatur und es werden Verhältnisse von 1:1 bis 30:1 verwendet. Gemessen an den unzähligen Themenbereichen, die die Landschaft umfasst, ist das nicht viel und das Modell kann natürlich nur klassieren, was es in den Trainingsdaten angetroffen hat.

Das Modell erhält drei erklärende Variablen, anhand derer die Antwortvariable, die Klassierung der Aussagen, bestimmt werden:

- Klassierte Aussagen, davon 80% als Trainingsdaten
- Term frequency - inverse document frequency Mass (Tf-idf-Mass)
- Bi- und Trigramme

Das Tf-idf-Mass ist ein übliches Mass in NLP und Information Retrieval. Es ermöglicht, Dokumente nicht nur mit so genannten Bag-of-Words zu charakterisieren, sondern die Worthäufigkeiten in ein Verhältnis mit ihrer allgemeinen Häufigkeit im Korpus zu setzen. So werden häufige, aber semantisch nicht unbedingt reiche Wörter, nicht übermässig gewichtet (Schütze, Manning und Raghavan, 2008). Diese Methode beachtet jedoch die Syntax der Sätze nicht. Hier kommen die Bi- und Trigramme ins Spiel, wobei häufige Wortfolgen (zwei bzw. drei aufeinander folgende Wörter) ebenfalls zur Charakterisierung der Texte beitragen (Joulin u. a., 2016).

Das beste Resultat mit einer Gesamt-Precision von 0.82, unabhängig von der zufälligen Auswahl der Trainingsdaten, wurde erzielt mit folgenden Hyperparamtern:

- 150 Entscheidungsbäumen
- Maximale Anzahl Entscheidungen pro Baum: 3000
- Tf-idf mit max. 2000 Begriffen und Wörter werden als korpuspezifische Stopwörter angesehen, wenn sie in mehr als 80% der Dokumente auftauchen

Die resultierenden positiv und negativ klassierten Aussagen wurden zusammen mit den darin enthaltenen LSB und allen Toponymkandidaten gespeichert.

### 5.1.5 Toponym Resolution

In der Toponym Resolution geht es darum, die Geo/Geo-Unsicherheit eines Toponyms zu lösen. Bisher wurde für jedes Toponym die Koordinaten aller Orte mit dem gleichen Namen gespeichert. Nun soll bestimmt werden, welcher Ort desselben Namens denn gemeint war und diesen im Raum festmachen.

Der Algorithmus, um dieses Problem zu lösen, basiert auf der räumlichen Verteilung aller Toponyme in einer Aussage und minimiert die Summ der Distanzen zwischen den ausgewählten Namen (s. Abb. 5.2). Für alle Kombinationen von Koordinaten der Toponyme wird diejenige mit der kleinsten Gesamtdistanz zu anderen Koordinatenpaaren gewählt. So ist mein einziges Auswahlkriterium die Nähe. Es wird keine weitere Information beachtet, weil mir beim Annotieren keine Regelmäßigkeiten auffielen bezüglich der häufigsten Objektarten der Toponyme.

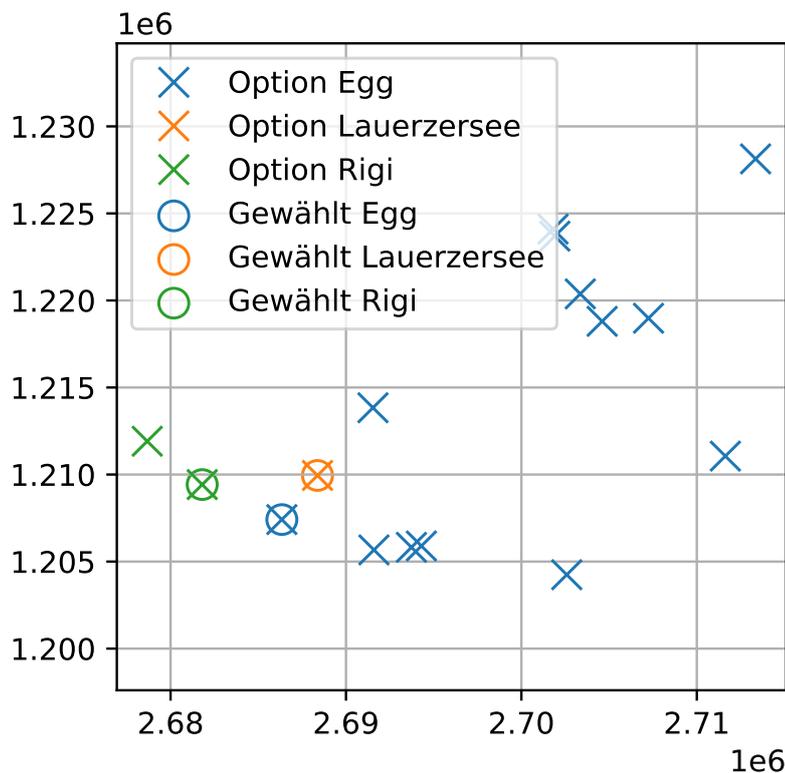


ABBILDUNG 5.2: Beispiel für die Auswahl der nächsten Koordinatenpaare dreier Toponyme.

So liegt jetzt eine Liste vor, welche pro eindeutigem Toponym und Aussage alle darin enthaltenen LSB auflistet. Auf diese Art wird jedes LSB mit allen Toponymen in einem Satz assoziiert, was die höchstmögliche Anzahl an Assoziationen bedeutet.

## 5.2 Räumliche Analyse

Punkteverteilungen zu analysieren ist eine gängige GIS-Operation. Die assoziierten Wörter mit jedem Punkt geben dieser Operation aber noch eine semantische Dimension, die sonst fehlt (Hu, 2018).

Räumliche kontinuierliche Werte ermöglichen den Vergleich verschiedener Ebenen und zwischen Regionen (Derungs und Purves, 2014). Dazu müssen die Punktdaten in eine Dichteoberfläche verwandelt werden. Vom Gazetteer werden jeweils nur ein Punkt pro Toponym gespeichert. Für Linien ist das der Mittelpunkt eines Linienabschnittes, für Polygone ist es der Mittelpunkt *innerhalb* des Polygons. Der kombinierte Punktdatensatz aus Punkt, Linien und Polygonen wird mit einer Kernel Density Estimation (KDE) geglättet. Explorativ zeigt sich, dass ein Kernelradius von 5 (für das Y-Magazin), bzw. 4 (Bote) Kilometer optimal ist, um weder an einzelnen Punkten Peaks zu erhalten, noch einzelne Dichteregionen verschmelzen zu lassen.

Die beiden Korpora werden getrennt untersucht, einerseits um Unterschiede in verschiedenen Textquellen aufzuzeigen und andererseits, weil Resultate aus dem Y-Magazin als viel kleinerer Datensatz untergehen würden. Dichtekarten werden je erstellt für alle Toponyme im Boten und im Y-Magazin, sowie für diejenigen Toponyme, die mit einem LSB zusammen in einem Satz stehen. Mit diesen beiden Karten je Korpus können Chi<sup>2</sup>-Karten erstellt werden. Mit dem Pearson's Chi<sup>2</sup>-Test wird pro Rasterzelle berechnet, ob die LSB-Toponymdichte (O) höher oder tiefer ist, als erwartet durch die Verteilung der Toponyme im ganzen Korpus (E) (Chernoff und Lehmann, 1954):

$$\chi^2 = \frac{(O - E)^2}{E}$$

Mit den LSB wird gleich vorgegangen. Hier handelt es sich bei der Dichtekarte um eine Approximation. Dies weil ein Punkt ein Toponym und einen Satz repräsentiert, und in einem Satz können LSB mehrfach vorkommen, werden aber in der KDE nur einmal gezählt.

### 5.3 Inhaltliche Analyse

Die Resultate werden sowohl quantitativ als auch qualitativ analysiert. Dabei spielen sowohl Worthäufigkeiten, wie auch die Ähnlichkeiten zwischen Beschreibungen eine Rolle. Ausserdem wird ein Netzwerk aus Begriffen gebaut, um deren Verbindungen aufzuzeigen.

Die deskriptive Statistik zu den gefilterten Aussagen gibt Auskunft über die durchschnittliche Anzahl Toponyme und LSB pro Satz, sowie die maximale und modale Anzahl. Die Häufigkeiten der LSB und Toponyme je Korpus werden gespeichert, sowie die Kookkurrenz für die häufigsten 21 LSB je Korpus, was je 30% aller Vorkommnisse entspricht bzw. etwa einem Vorkommen von über 1% für ein einzelners LSB. Das heisst für die häufigsten LSB wird gespeichert, mit welchen anderen LSB und wie oft sie zusammen in einer Aussage auftreten. Kookkurrenz impliziert eine gewisse Assoziation zwischen den Begriffen (Brezina, 2018). Dies muss hier etwas vorsichtig betrachtet werden, weil das Suchfenster sehr breit ist, nämlich Begriffe die nur schon im gleichen Satz sind, werden zusammen aufgeführt. Die Häufigkeiten, in Word Clouds visualisiert, geben einen schnellen Überblick über die Objekte, die ein Korpus und die sechs verschiedenen Schlüsselgebiete repräsentieren (Hu, 2018).

Um die Beziehungen der LSB und der Schlüsselgebiete untereinander besser zu verstehen, soll ihre Ähnlichkeit berechnet werden. Mit den LSB-Häufigkeiten und der Kookkurrenz kann der Kosinus-Ähnlichkeitsindex für die Schlüsselgebiete und die häufigsten LSB berechnet werden. Die Kosinus-Ähnlichkeit wird mithilfe der Pythonbibliothek *scikit-learn* gerechnet (Pedregosa u. a., 2011). Die Idee ist, dass Texte - in meinem Fall Listen von LSB - als numerische Vektoren repräsentiert miteinander verglichen werden, indem der Kosinus des Winkels zwischen den beiden Vektoren

berechnet wird, wie folgt:

$$\cos(\alpha) = \frac{a * b}{|a| * |b|}$$

wobei  $a$  und  $b$  die beiden Vektoren sind mit den LSB-Begriffen und ihren Häufigkeiten. Die Vektoren beinhalten den zu Zahlen konvertierten Text. Die Zahlen beruhen wieder auf dem Tf-idf-Mass und werden normalisiert, damit das Mass zwischen zwei ähnlichen aber unterschiedlich langen Vektoren nicht zu unterschiedlich ausfällt (Schütze, Manning und Raghavan, 2008). Sind zwei Texte, d.h. Vektoren, identisch, wird die Ähnlichkeit 1 sein, wenn zwei Vektoren hingegen überhaupt keine Übereinstimmung haben, dann ist das Ähnlichkeitsmass 0. Als Grundlage für die Berechnung der Kosinus-Ähnlichkeit zwischen den Schlüsselgebieten dienen die darin vorkommenden LSB. Der Input sind also 6 Vektoren je mit einer Liste der LSB. Das Resultat ist eine 6x6-Matrix mit Werten zwischen 0 und 1. Die Kookkurrenzen der kombinierten häufigsten LSB aus beiden Korpora hingegen bilden die Vektoren für die Kosinus-Ähnlichkeit hier, die in einer 19x19-Matrix resultiert.

Mit den Kookkurrenzen der LSB kann auch ein Netzwerk zwischen den Begriffen gebaut werden, um noch besser zu visualisieren wie diese zusammenhängen. Das Netzwerk wird mit dem Algorithmus von Fruchterman-Reingold aus dem Pythonmodul *python-igraph* aufgebaut, der die Knoten basiert auf der Stärke der Kanten anordnet (Gajdoš u. a., 2016).

## 5.4 Vergleich mit der Landschaftskonzeption Schwyz

In diesem Kapitel soll herausgefunden werden, wie ähnlich sich die Beschreibungen der Schlüsselgebiete in der Landschaftskonzeption und in der Zeitung und dem Magazin sind.

Da die Beschreibungen zu den Schlüsselgebieten in der Landschaftskonzeption nur 13 Seiten ausmachen, werden die LSB händisch aufgelistet. Dabei werden die gleichen Regeln angewandt wie bereits fürs Annotieren. Die Gesamtheit and Beschreibungen nur in den Schlüsselgebieten fällt knapp aus. Deshalb werden die 10 häufigsten LSB in der Landschaftskonzeption (ca. 30% der Vorkommnisse) mit Begriffen aus dem Katalog charakteristischer Kulturlandschaften der Schweiz ergänzt (Rodewald, Schwyzer und Liechti, 2014). Dabei werden die Nomen gewählt, die mit den LSB zusammen im Satz vorkommen und falls das für ein LSB nie zutrifft, dann wird zumindest der Landschaftstyp genannt, in dem der LSB vorkommen soll. Im Anhang B.2 sind die häufigsten LSB der Landschaftskonzeption und die Erweiterungen durch den Katalog aufgeführt.

Die Übereinstimmung zwischen den Beschreibungen der Schlüsselgebiete in der Landschaftskonzeption und in den Korpora wird mit dem Jaccard Index berechnet. Der Jaccard Index beachtet bei der Ähnlichkeit keine Häufigkeiten, sondern die Überlappung zweier Dokumente aufgrund der Menge an einzelnen, gemeinsamen



ABBILDUNG 5.3: Anteil relevanter Sätze am ganzen Korpus.

Wörtern (Schütze, Manning und Raghavan, 2008). Das ist sinnvoll, da mit der Erweiterung durch den Katalog die Häufigkeiten der verschiedenen LSB nicht mehr repräsentativ für die Landschaftskonzeption sind. Wie die Kosinusähnlichkeit ist auch der Jaccard Index vor allem im relativen Vergleich interessant.

## 5.5 Fehlerberechnung

Die Evaluation der erfolgreichen Geotextextraktion sollte sowohl quantitative als auch qualitative Elemente beinhalten (Hu, 2018). Hier werden zuerst robuste Auswertungen erklärt, in der Diskussion werden Karten und Ähnlichkeitsmasse visuell verglichen.

Zur Evaluation der Datenaufbereitung werden die Standardindikatoren Precision und Recall berechnet (Gritta u. a., 2018). Precision ist die Antwort auf die Frage, wieviele der gefundenen Resultate richtig identifiziert wurden oder anders ausgedrückt

$$P = \frac{\text{echt positiv}}{\text{echt positiv} + \text{falsch positiv}}$$

Recall ist die Antwort auf die Frage, wieviele von allen möglichen Resultaten gefunden wurden oder anders ausgedrückt

$$R = \frac{\text{echt positiv}}{\text{echt positiv} + \text{falsch negativ}}$$

Während es für die Precision relativ einfach wäre, eine grosse Anzahl Dokumente zu evaluieren, stellt sich das Recall ungleich schwieriger heraus. Um das Recall über den ganzen Prozess zu berechnen müssten über 1000 Sätze annotiert werden, bis einmal einer ein LSB und ein Toponym beinhaltet und als landschaftsrelevant gilt. Das, weil solche Sätze höchstens ein halbes Prozent des ganzen Korpus ausmachen (s. Abb. 5.3). Dieses Vorgehen wäre zu zeitintensiv (Chesnokova und Purves, 2018b).

Stattdessen wird die Precision und das Recall jeweils an 20 Dokumenten berechnet. Zuerst für die Toponym Recognition je mit NER und nur mit dem Gazetteer-Matching ohne vorherige NER. Fehlerhafte Toponym Recognition ist laut Acheson, Wartmann und Purves, 2018 für einen grossen Teil von weiterführenden Fehlern verantwortlich. Auch in dieser Arbeit ist dieser Teil besonders wichtig, weil es einen starken ersten Filter darstellt. Danach werden die Masse für das Auffinden von LSB und die Klassifikation der Aussagen mit und ohne Random Forest berechnet. Diese werden nur mit Sätzen gerechnet, die bereits durch die ersten Filter gegangen sind. Somit fehlt eine Evaluation der gesamten Pipeline, was in einem Verfahren mit realen Daten sehr wichtig wäre (Acheson, Volpi und Purves, 2020). Diese Evaluationsmasse wurden immer wieder berechnet und die falsch positiv klassierten Resultate halfen, übliche Fehler mithilfe von neuen Regeln zu beheben.

Die Plausibilität der häufigsten Toponyme und der häufigsten LSB (die je 30% aller Vorkommnisse ausmachen bzw. etwa einem Vorkommen von über 1% pro Wort) wurde qualitativ überprüft. Die häufigsten Toponyme konnten einfach in einem GIS selektiert werden und die Lage mit einer Karten vergleichen. Danach musste der Gazetteer nochmal angepasst werden (s. Kapitel 4.3). Für die häufigsten LSB lese ich je 5-10 Sätze, in denen sie vorkommen, um zu bestimmen, ob der Landschaftskontext richtig erkannt wurde. Dadurch wurden einige LSB aus der Liste gestrichen (z.B. Haus, Restaurant, Weg und Wetter).

*Der Pythoncode für die einzelnen Abschnitte finden sich im folgenden Git Repository: [https://github.com/natal1905/MA\\_code](https://github.com/natal1905/MA_code).*

## Kapitel 6

# Resultate

Bevor die Resultate in Bezug auf die Fragestellungen betrachtet und ausgewertet werden, sollen zuerst die Ergebnisse der Datenaufbereitung beschrieben werden, da diese als Grundlage für die Erforschung der Fragestellung dienen.

### 6.1 Textaufbereitung

Die Kuchendiagramme (s. Abb. 6.1) zeigen wie signifikant die Filterung der Sätze ist. Die grösste Einschränkung passiert bei der Suche nach LSB und der Klassierung. Die unterschiedliche Ausrichtung der beiden Korpora ist gut sichtbar. Das Y-Magazin mit thematischem Fokus auf dem Kanton weist mehr ortsgebundene Sätze auf. Interessant ist, dass sich diese Sätze auch vier Mal häufiger mit Landschaft beschäftigen. Die Unterschiede zwischen den Sätzen mit Toponym und relevantem LSB und den resultierenden Sätzen kommen davon, dass gewisse Toponyme erst hier gelöscht werden anstatt bereits im Gazetteer und daher einige Sätze keinen Raumbezug mehr haben.

#### 6.1.1 Precision und Recall

Wie sicher diese Extraktion von Sätzen ist, wird mit der Berechnung von Precision und Recall nach einzelnen Schritten bestimmt (s. Tab. 6.1). Die Zahlen repräsentieren Durchschnitte von 2-3 Durchläufen mit verschiedenen annotierten Dokumenten. Sowohl die Toponym Recognition wie auch die Suche nach landschaftsrelevanten Sätzen sind wesentlich genauer geworden durch die Verwendung von Machine Learning (NER und Random Forest).

Für die LSB-Suche, die nur mit Regeln passierte, ohne dass diese mithilfe des Klassifikators weiter gefiltert wurden, wurde kein Recall berechnet. Man kann annehmen, dass durch die hohe Prozentzahl an Sätzen, die extrahiert wurde (ca. 25%), der Rücklauf an möglicherweise relevanten Sätzen recht hoch ist.

Am meisten Fehler entstehen bereits am Anfang der Pipeline, bei der Toponym-suche. Dabei ist der grösste Fehlertyp das PDF-Parsing. Die Schriftsetzung führt dazu, dass Anfangsbuchstaben oft vom Rest des Wortes getrennt werden. So werden *L*

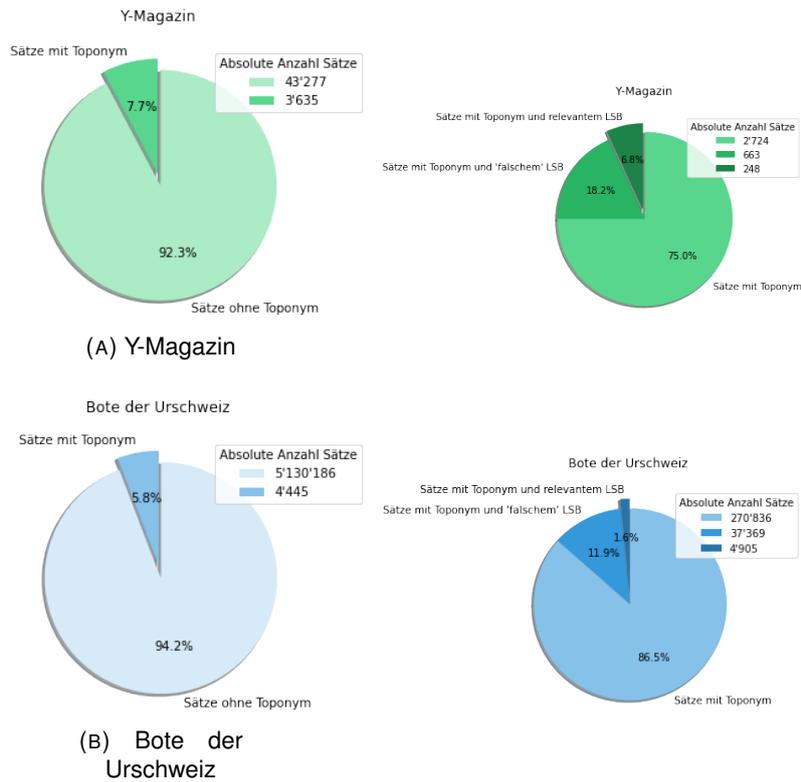


ABBILDUNG 6.1: Anteil gefilterter Sätze pro Verarbeitungsschritt.

Kapitel	Schritt	P	R	Annotierte Grundlage
5.1.3	Toponym Recognition ohne NER	0.57	0.69	60 Artikel
5.1.3	Toponym Recognition mit NER	0.74	0.63	60 Artikel
5.1.4	LSB-Suche ohne ML	0.35	-	150 Sätze
5.1.4	LSB-Suche mit ML	0.64	0.87	150 Sätze

TABELLE 6.1: Precision und Recall für Datenaufbereitung

achen oder H öfe nicht als Toponyme erkannt. Auch Worttrennungen können problematisch sein, z.B. *Wildiheuregi-on*. Gewisse Artefakte bleiben trotz der Bearbeitung noch übrig, z.B. *yWeiden*. Für Toponyme kommt die Geo/Non-Geo-Disambiguierung hinzu, wobei gerade im Kanton Schwyz zwei grosse Gemeinden mit Verwechslungspotenzial liegen: Brunnen und Lachen.

Durch restriktive Regeln wird auch nicht das ganze Potenzial an ortsbezogenen Aussagen gefunden. Toponyme in Wortzusammensetzungen z.B. das *Höllochdrama* oder die *Sihlseebrücke* werden nicht gefunden. Weiter kommen die Grenzen des NER von spacy zum Vorschein, welches oft Personennamen als Örtlichkeiten identifiziert hat. Diese wurden im Verlauf der Pipeline aber entfernt (s. Kapitel 4.3).

Für die LSB ist vor allem der Landschaftsbezug schwierig; war eine Wanderung in der Sonne schön und der Föhn hat eine gute Aussicht ermöglicht oder sind das die Wettervorhersagen für nächstes Wochenende? Manche Fehlerquellen konnten

in mehreren Iterationen gut beseitigt werden, so z.B. die Ergänzungen im Gazetteer und das Hinzufügen von häufigen Kompositateilen. Tatsächlich scheinen am Ende sowohl die Geo/Non-Geo-Disambiguierung, wie auch der Landschaftsbezug für LSB, zumindest für die häufigsten 30% der Toponyme und LSB, nicht mehr eine Einschränkung gewesen zu sein. Was das Erkennen von Aussagen mit LSB aber klar einschränkt, ist der Umfang der LSB-Liste. Eine umfassendere Liste würde den Recall erhöhen, das aber auf Kosten der Precision. Etwas, was für das Random Forest Modell schwierig schien, sind Metaphern und Sprichwörter, wie „im Fluss sein“, „den Wald vor lauter Bäumen nicht mehr sehen“. Das wäre etwas, was einem Modell aber einfach mitgegeben werden könnte und in bemerkten Fällen auch gemacht wurde (z.B.: „Es kann nicht sein, dass die ganze Zentralschweiz umgegraben oder mit einem Wald aus Schildern versehen wird.“).

### 6.1.2 Statistische Beschreibungen

Die Toponymverteilungen folgen dem Zipfgesetz; eine quantitative Verteilung, die die Worthäufigkeiten von Sprachen beschreibt. Dabei sind wenige Wörter extrem häufig und die allermeisten sind extrem selten (Engelberg, 2015). In den Abbildungen 6.2 und 6.3 kann man sehen, dass eine solche Verteilung resultiert, egal wie die Inputdaten gefiltert wurden. Die Abbildungen zeigen nur Toponyme, die mindestens 1% aller Nennungen ausmachen. Mehr als 2000 Toponyme bilden den Schwanz der Verteilung. Eine ähnliche Verteilung sieht man bei den Häufigkeiten der LSB in beiden Korpora (s. Abb. 6.11).

Die beiden ungefilterten Korpora zeigen einen sehr viel stärkeren Abfall der Frequenz. Bei beiden ist Schwyz mit 20 bis 25% das dominanteste Toponym. Interessant ist, dass die Untergruppen der beiden ungefilterten Korpora weniger der Verteilung des eigenen Korpus folgen und mehr Gemeinsamkeiten untereinander aufweisen. Zum Beispiel ist bei beiden gefilterten Korpora eindeutig nicht mehr Schwyz das häufigste Toponym.

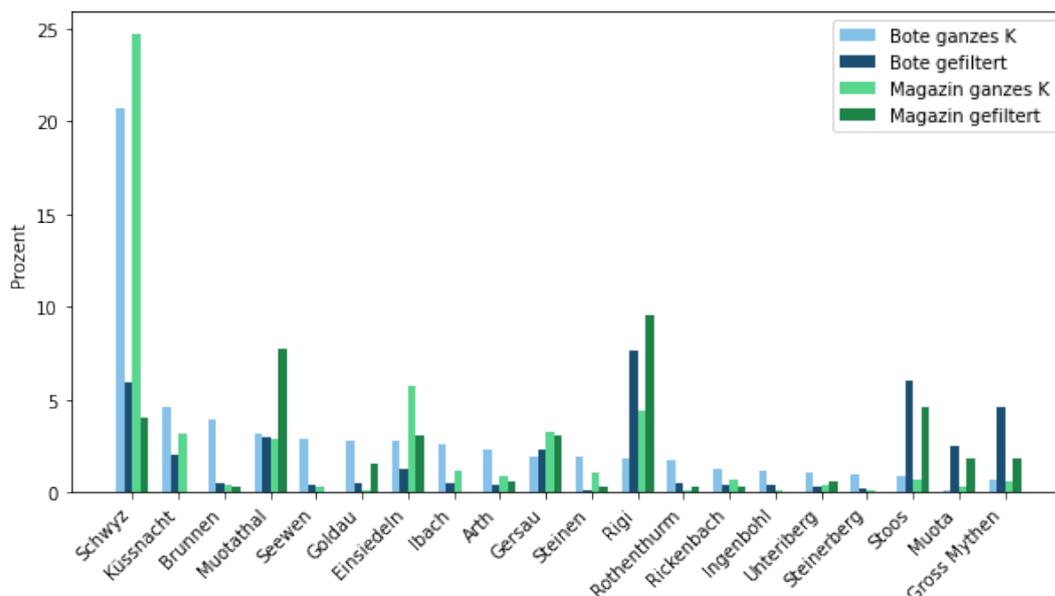


ABBILDUNG 6.2: Häufigste Toponyme in beiden Korpora (geordnet nach der Toponymfrequenz im Bote der Urschweiz).

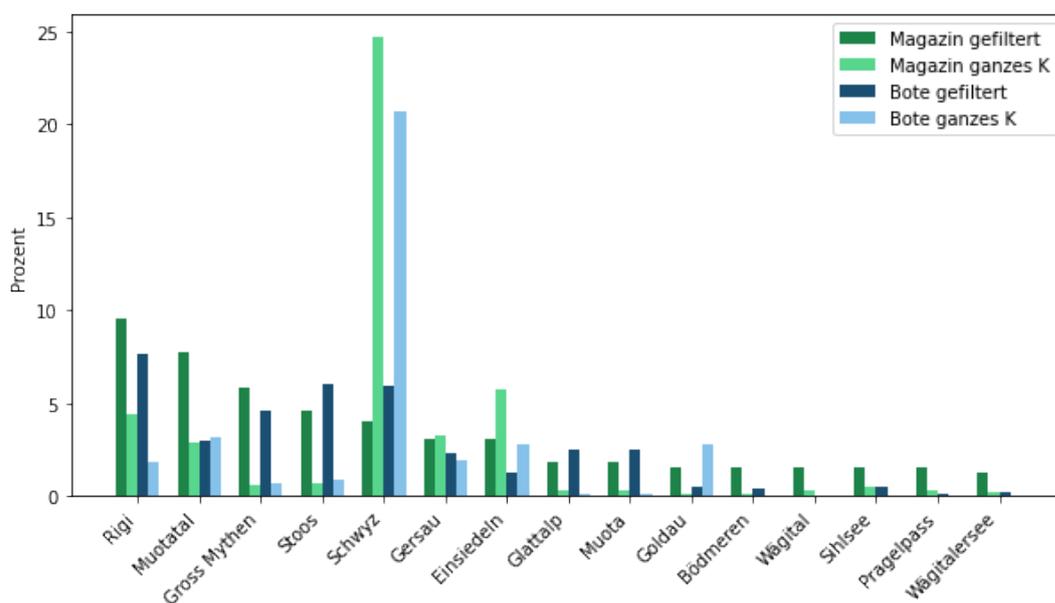


ABBILDUNG 6.3: Häufigste Toponyme in beiden Korpora (geordnet nach der Toponymfrequenz in ausgewählten Sätzen im Y-Magazin).

Bemerkenswert ist auch die Tabelle 6.2. Die Verteilungen der Toponyme und auch der LSB pro Satz zwischen den beiden Dokumentensammlungen sind sehr ähnlich. Das obwohl, wie in den Kuchendiagrammen (s. Abb. 6.1) ersichtlich, die Filterung die beiden Korpora unterschiedlich betraf. Das deutet auf eine Generalität in der Sprache hin. Toponyme und LSB folgen einer gewissen Häufigkeit in der Sprache, die sich in der Auswahl hier widerspiegelt.

	Bote				Y-Magazin			
	∅	median	modus	max	∅	median	modus	max
# Toponyme	1.7	1.0	1	11	1.7	1.0	1	7
# mehrdeutige Topos	1.6	1.0	0	20	1.4	1.0	0	15
# LSB	1.5	1.0	1	9	1.7	1.0	1	6

TABELLE 6.2: Toponym- und LSB-Verteilung in gefilterten Sätzen

Der Ort mit den meisten Sätzen (362) und assoziierten (535) LSB im Boten ist Schwyz. Im Magazin wird die Rigi Staffel am meisten mit LSB beschrieben (42 LSB), während das ganze Muotatal am häufigsten in einem Satz vorkommt (25 Sätze). Ein perfektes Beispiel liefert das Y-Magazin mit 6 LSB:

*Wenn die Sonne tief steht und alles in goldgelbes Licht taucht die Schatten weicher und länger werden und Flora und Fauna die Milde des Abends in sich tragen dann ist es geradezu bezaubernd über Auen und Matten zu gehen und die erhabene Höhe der Rigi zu verlassen um wieder auf der Erde im Tal anzukommen (LSB: Auen, Höhe, Licht, Matten, Sonne, Tal).*

Der Satz mit den meisten LSB im Boten hingegen ist ein gutes Beispiel für Schwierigkeiten, die beim Übersetzen vom PDF-Format in Textdateien entstehen und die Fallstricke, die sich durch meine Definitionen ergeben. Die Abbildung 6.4 zeigt, wie die Seite in der Zeitung aussieht. Darunter steht der Satz, den meine Pipeline identifiziert hat. Weil Sätze nur durch Satzzeichen beendet werden können, wird er viel zu lange und umfasst die ganze Aufzählung sowie den Titel und Untertitel des nächsten Artikels. Weitere Regeln zum Layout im Textfile könnten dem Abhilfe schaffen.

Zusage für den Standort liegt aber vor, im Idealfall werde schon 2021 mit dem Bau begonnen.

Morschach, Muotathal und Illgau haben als Eigenregierenden der Tourismus GmbH bereits einen Beitrag in die Kosten von etwa 9 Millionen Franken in Aussicht gestellt. Sachgeschäfte dazu sollen nun ausgeschrieben und im Idealfall in einem Jahr dem Bürger präsentiert werden. Da die Gemeindeführer aber weniger als die Hälfte der Kosten decken werden, sind auch bereits Gespräche mit dem Kanton sowie Wirtschaftspartnern am Laufen, um die Restfinanzierung sichern zu können.

**WWW.**  
Projektbeschreibung im Animations-Video unter: [bbch.ch](#)

### Diese Projekte sind geplant

**Wetter-Messstationen**

- Dutzende Wetter-Messstationen sollen in Muotathal, Morschach, Illgau und auf dem Stoos erstellt werden
- Das dichteste Netz an Wetter-Messstationen in der Schweiz ist das Ziel
- Einzelne Wetterstationen sollen begehbar werden
- Wetterunterschiede und -extreme in der Region werden mess- und visualisierbar gemacht

**Erlebniszentrum in Muotathal**

- Daten werden in einer App für Touristen und Einheimische öffentlich zugänglich gemacht
- Wetterphänomene werden auf eine wissenschaftliche, interaktive und spielerische Weise erfahrbar gemacht (zum Beispiel, indem im Windkanal die Wucht des Föhnwindes erlebt werden kann)
- Die Besucher sollen durch klimatische Extreme der Region und der Erde geführt werden

**Konkrete Themen-Beispiele: Wind und Luftdruck, Erwärmung und Kälte, Niederschlag, Nebel oder die Auswirkungen des Wetters auf die Menschen**

- Aktuelles Wettergeschehen in der Welt soll dargestellt werden
- Meteorologische Moderne und Tradition sollen verbunden werden
- Wetterschmöckerlounge mit persönlichen Einblicken in die humorvolle, traditionsreiche Kultur der Muotathaler Wetterschmöcker und Zusammen

**menspiel von Tier- und Pflanzenwelt mit dem lokalen Wetter**

- Wetterspielplatz mit weiteren spielerischen Elementen zum Thema Wetter

**Aussichtsplattformen**

- In Morschach, Illgau und auf dem Stoos sollen Aussichtsplattformen gebaut werden
- Gestalterische und inhaltliche Verbindung zum Wetter und dem Erlebniszentrum in Muotathal

## In reichen oder abgelegenen Orten tummeln sich viele Offroader

Bei der SUV-Verbreitung macht sich im Kanton Schwyz ein breiter Graben bemerkbar.

(A) Artikel im Boten der Urschweiz vom  
05.12.2019

Diese Projekte sind geplant

**Wetter-Messstationen**

- Dutzende Wetter-Messstationen sollen in Muotathal, Morschach, Illgau und auf dem Stoos erstellt werden
- Das dichteste Netz an Wetter-Messstationen in der Schweiz ist das Ziel

– Daten werden in einer App für Touristen und Einheimische öffentlich zugänglich gemacht  
Erlebniszentrum in Muotathal

- Einzelne Wetterstationen sollen begehbar werden
- Wetterphänomene werden auf eine wissenschaftliche, interaktive und spielerische Weise erfahrbar gemacht (zum Beispiel, indem im Windkanal die Wucht des Föhnwindes erlebt werden kann)
- Wetterunterschiede und -extreme in der Region werden mess- und visualisierbar gemacht
- Die Besucher sollen durch klimatische Extreme der Region und der Erde geführt werden

– Konkrete Themen-Beispiele: Wind und Luftdruck, Erwärmung und Kälte, Niederschlag, Nebel oder die Auswirkungen des Wetters auf die Menschen

- Aktuelles Wettergeschehen in der Welt soll dargestellt werden
- Meteorologische Moderne und Tradition sollen verbunden werden
- Wetterschmöckerlounge mit persönlichen Einblicken in die humorvolle, traditionsreiche Kultur der Muotathaler Wetterschmöcker und Zusammen

menspiel von Tier- und Pflanzenwelt mit dem lokalen Wetter

- Wetterspielplatz mit weiteren spielerischen Elementen zum Thema Wetter

**Aussichtsplattformen**

- In Morschach, Illgau und auf dem Stoos sollen Aussichtsplattformen gebaut werden
- Gestalterische und inhaltliche Verbindung zum Wetter und dem Erlebniszentrum in Muotathal

In reichen oder abgelegenen Orten tummeln sich viele Offroader  
Bei der SUV-Verbreitung macht sich im Kanton Schwyz ein breiter Graben bemerkbar.

(B) Ausschnitt Textfile

ABBILDUNG 6.4: PDF-Parsing: Zeitungsartikel (A) und Textdatei (B).

## 6.2 Räumlich: Toponyme

Auf der Karte 6.5 sieht man, dass Toponyme eben nicht nur Namen für Gemeinden sind, die als Punkte abstrahiert werden können, sondern auch für Flüsse, Skilifte, Täler, Seen und Berge. Besonders auffällig ist dabei die Muota mit 157 Nennungen und das Rigi-Massiv und das Muotathal mit je um 200 Nennungen. Trotzdem sind insgesamt die Toponyme, welche als Punkte im Gazetteer waren, am meisten vertreten mit 10'908 Vorkommnissen. Im Gazetteer werden 6150 Punktoponyme aufgelistet, in den Korpora kommen 564 unterschiedliche davon vor. Linien werden 353 genannt, aufgeteilt auf 35 einzigartige Toponyme (im Gazetteer gab es 372). Bei den Flächen sind es 2387 Vorkommnisse, wovon 353 einzigartige Nennungen von

insgesamt 2254 im Gazetteer auftreten. Relativ zum Gazetteer gesehen, werden also am meisten Namen von Polygonen gefunden (15%).

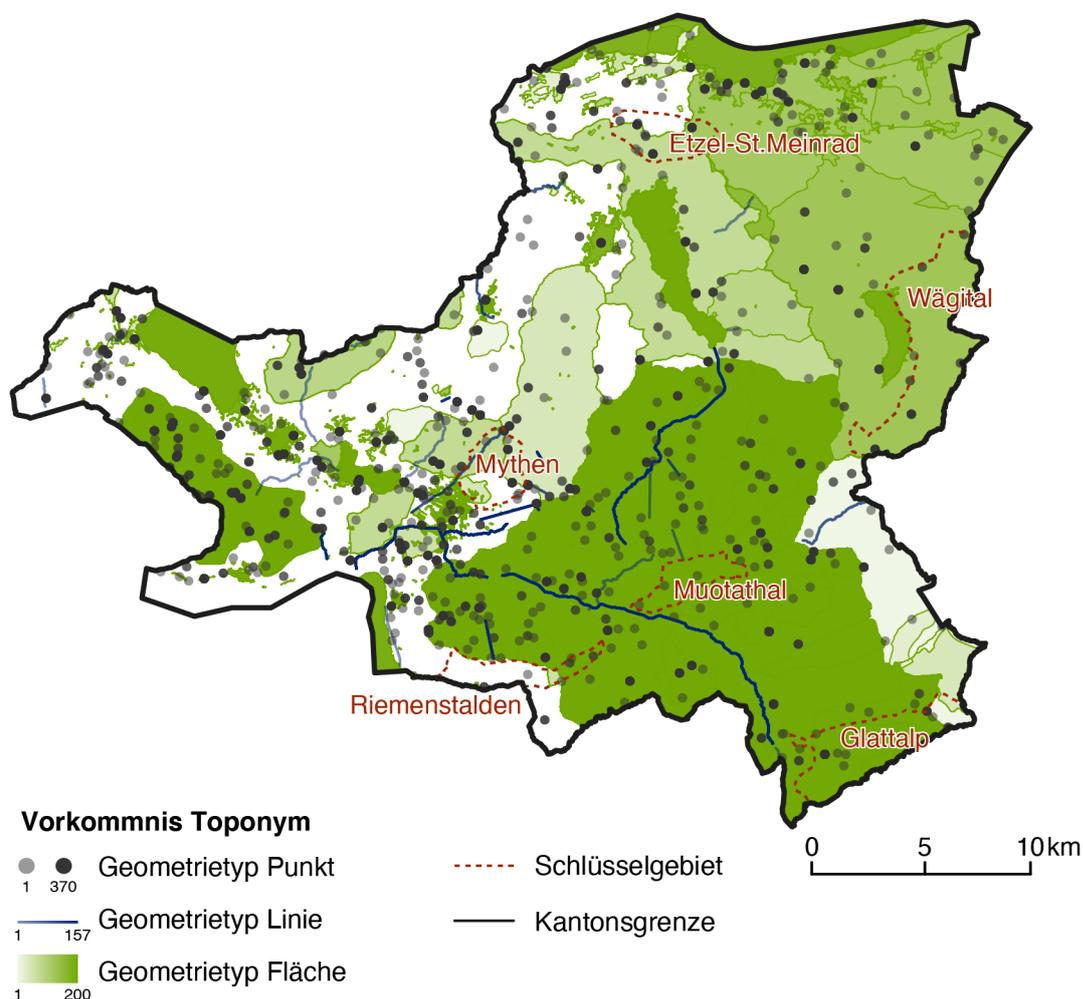


ABBILDUNG 6.5: Vorkommen Toponyme aus beiden gefilterten Korpora pro Geometrietyp.

Die Transparenz in der Karte 6.5 reicht nicht aus, um besonders dicht beschriebene Gebiete zu erkennen und Unterschiede auszumachen. Deshalb folgen nun Dichtekarten von Toponymen aus dem Y-Magazin und dem Boten (s. Abb. 6.7 bis 6.10), sowie im nächsten Abschnitt zu ausgewählten LSB (s. Abb. 6.15 bis 6.17).

Weil Schwyz besonders die ungefilterten Korpora dominiert (s. als Beispiel die Dichtekarte 6.6 des Boten), werden Punktnennung von Schwyz in den Dichtekarten nicht mitgerechnet. Da die genauen Werte weniger wichtig sind als die qualitativen Unterschiede wird auf eine Legende verzichtet. Die Tabellen 6.3 und 6.4 geben eine Übersicht über die durchschnittlichen Dichten pro km<sup>2</sup>.

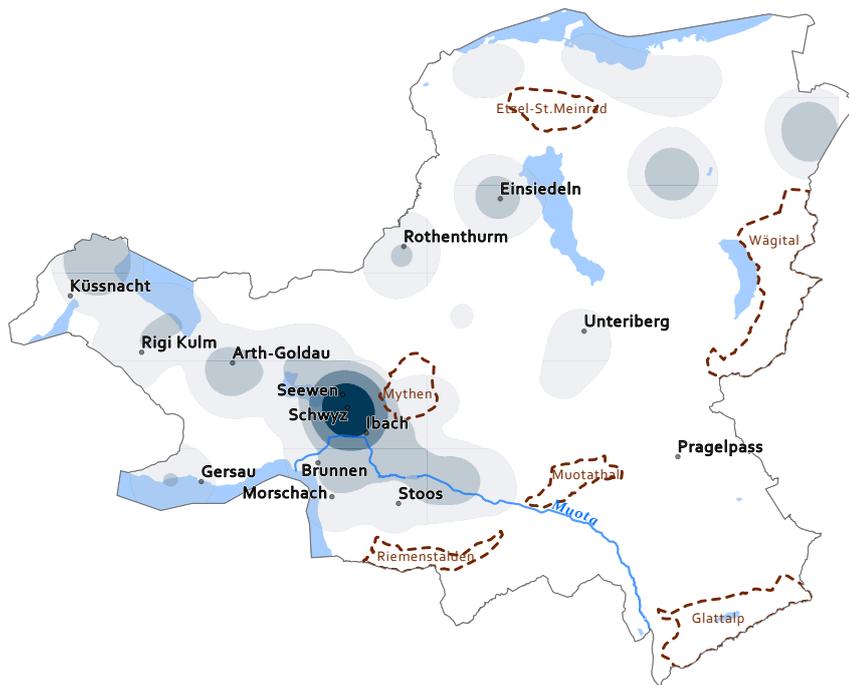


ABBILDUNG 6.6: Dichtekarte aller Toponymvorkommnisse im Boten.

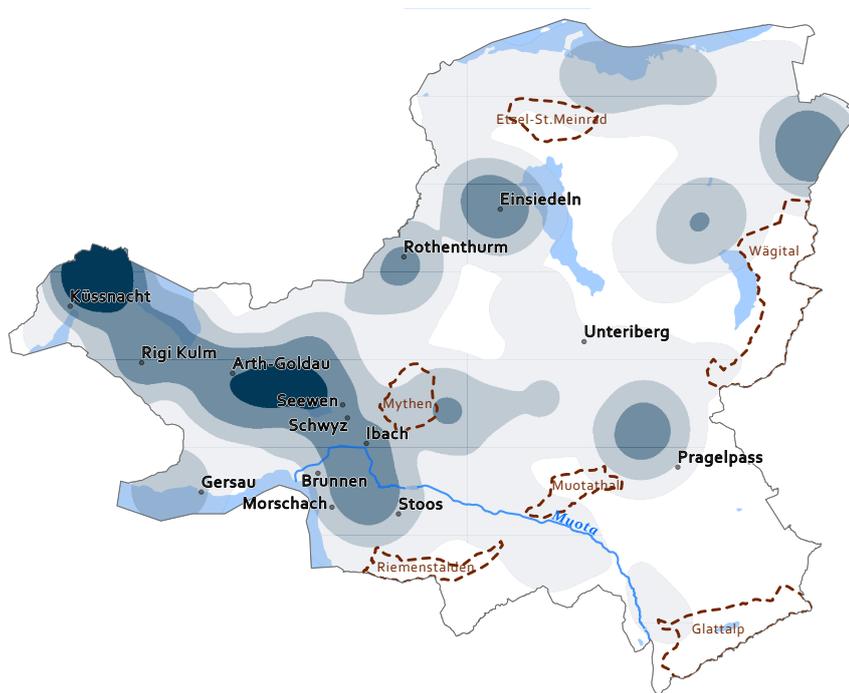


ABBILDUNG 6.7: Dichtekarte aller Toponymvorkommnisse im Boten ohne Schwyz.

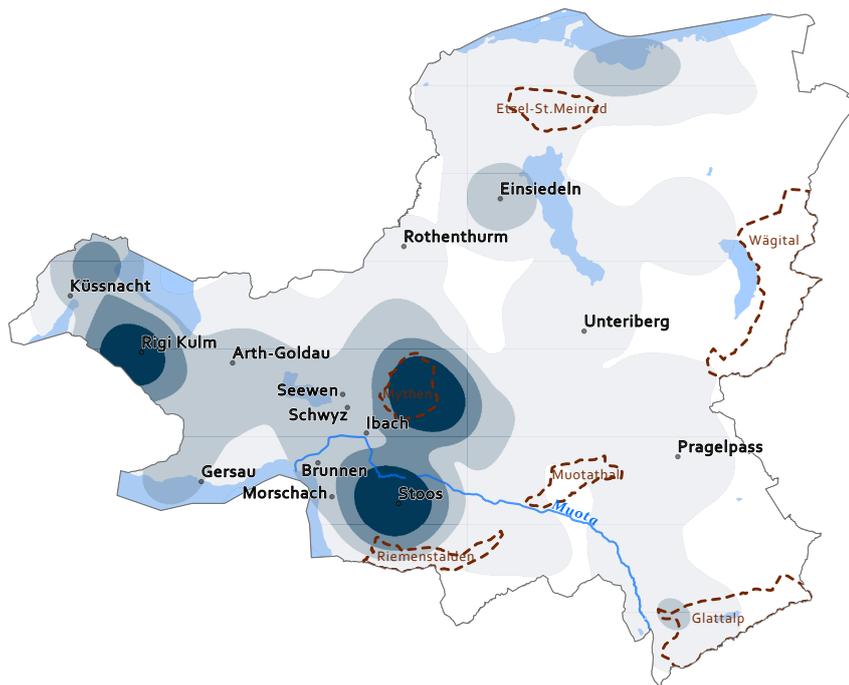


ABBILDUNG 6.8: Dichtekarte gefilterter Toponymvorkommnisse im Boden ohne Schwyz.

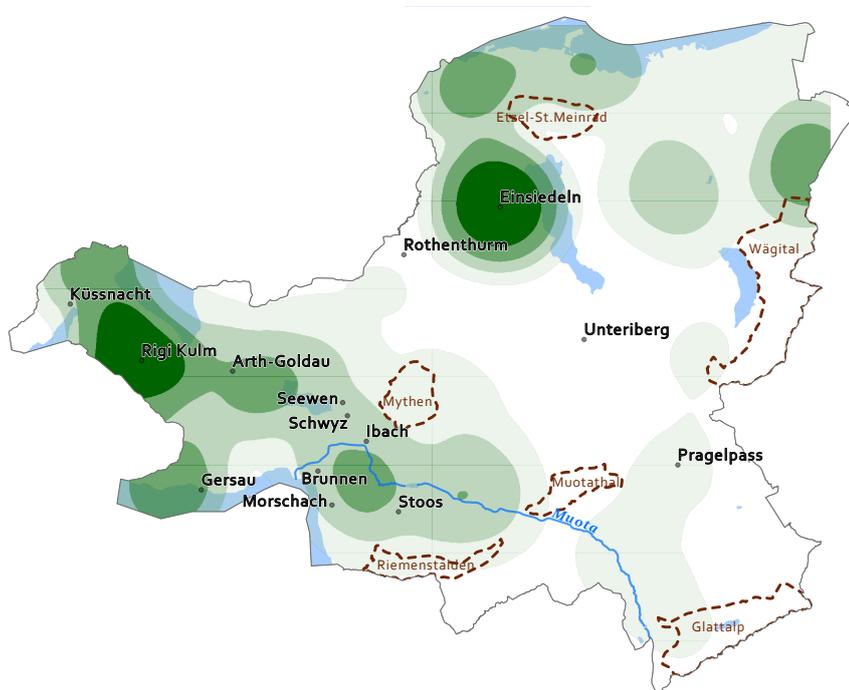


ABBILDUNG 6.9: Dichtekarte aller Toponymvorkommnisse im Y-Magazin ohne Schwyz.

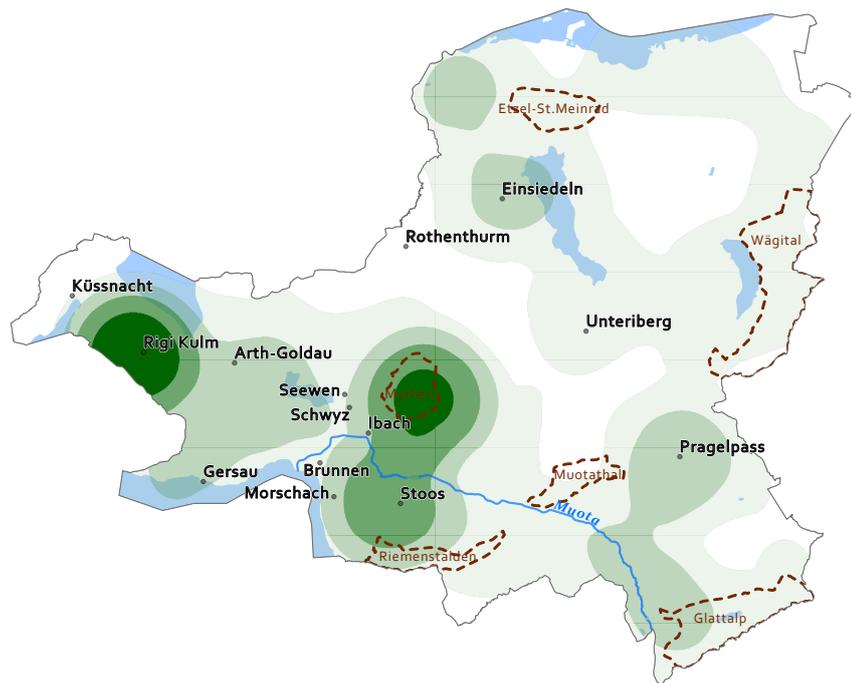


ABBILDUNG 6.10: Dichtekarte gefilterter Toponymvorkommnisse im Y-Magazin ohne Schwyz.

Es ist ersichtlich, dass die Verteilungen der gefilterten Toponyme für beide Korpora nicht einfach der zugrundeliegenden Punktverteilung des gesamten Korpus folgt.

Beiden Korpora ist gemein, dass sich ein Band zieht vom Stoss über den Talkessel Schwyz bis in die Ebene von Arth-Goldau, auf das Rigimassiv hinauf und wieder runter nach Küssnacht. Im Norden ist die Stadt Einsiedeln auffällig, sowie Lachen und die Autobahnausfahrt Lachen (im Nordosten), was auf suboptimale Toponymdisambiguierung zurückzuführen ist, da vermutlich öfters die Stadt anstatt der Ausfahrt gemeint war.

Die gefilterten Toponyme des Boten verteilen sich weg vom Zentrum Brunnen-Schwyz-Arth-Goldau ins Umland und konzentrieren sich auf die Rigispitze und auf die Gipfel beim Stoos und den Mythen. Im Y-Magazin ist von Anfang an ein weniger starkes Cluster in der Region Schwyz-Arth-Goldau sichtbar. Dafür stechen hier Einsiedeln und der Sihlsee besonders hervor, sowie das auffällige Cluster am westlichen Spitz der Rigi um Rigi Kulm herum. Dieses bleibt noch deutlicher bestehen, wenn nur die nach LSB-Vorkommen gefilterten Toponyme betrachtet werden. Hinzu kommt ganz auffällig die Mythenregion und der Stoos. Etwas stärker tritt auch die Region vom Pragelpass über die Glattalp bis entlang der Muota im Süden auf. Die Regionen im Norden sind hingegen kaum noch vertreten.

### 6.3 Inhaltlich: Landschaftsbegriffe

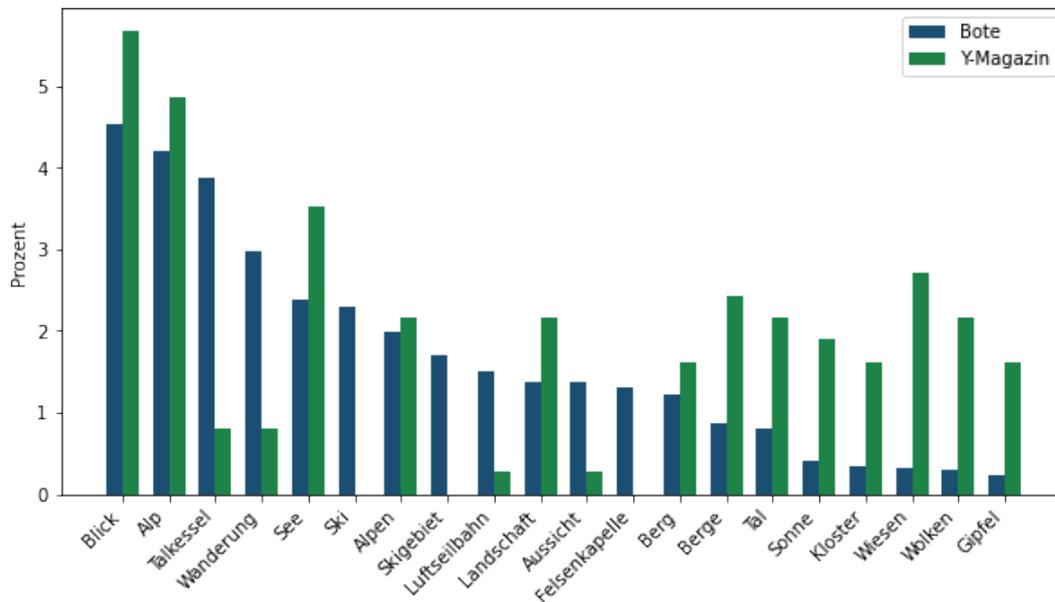


ABBILDUNG 6.11: Häufigste Landschaftsbegriffe im Boten und im Y-Magazin (sortiert nach dem Vorkommen im Boten).

Die Abbildung 6.11 und die beiden Word Clouds 6.12 und 6.13 visualisieren das gleiche, nämlich die LSB-Frequenzen in den beiden Korpora. Die Tabelle weist die LSB mit mehr als 1% Vorkommen in je einem der beiden Korpora auf. Die prozentuale Verteilung der LSB weist wie die der Toponyme eine rasch abnehmende Häufigkeit auf. Aber es ist auffällig, dass es weniger Extreme gibt bei den LSB.

Die unterschiedlichen Häufigkeiten zwischen den beiden Korpora lassen sich teils räumlich teils inhaltlich erklären, obwohl nicht klar ist, was im spezifischen Fall überwiegt. Sind *Gipfel*, *Sonne* und *Kloster* häufiger im Y-Magazin anzutreffen, weil im Y-Magazin öfter über die Mythen und Einsiedeln geschrieben wird oder wird wegen dieser Eigenschaften häufiger über die Orte geschrieben? Die *Aussicht* wird besonders oft von der Rigi aus beschrieben (s. Abb. 6.16), die ihrerseits besonders häufig im Boten vorkommt. Für andere LSB wie *Wolken* oder *Wanderung* wurden keine besonderen räumlichen Unterschiede zwischen den Korpora festgestellt. Im Boten werden aber häufig Gruppenwanderungen beworben, was vermutlich zu der höheren Häufigkeit im Vergleich zum Y-Magazin führt. Wolken kommen häufiger im Y-Magazin vor. Das mag daran liegen, dass sie im ungefilterten Boten oft im Zusammenhang mit Wetterprognosen genannt werden. Diese wurden in den Trainingsdaten des Random Forest Modell als nicht landschaftsrelevant klassiert. Trotzdem bedeutet das, dass im Y-Magazin häufiger über das Wetter (*Sonne* und *Wolken*) geschrieben wird. Die Zeitung konzentriert sich in ihrer Berichterstattung etwas stärker auf die urbanen Gebiete als auf die ländlichen (s. Abb. 6.6 und 6.7), was das



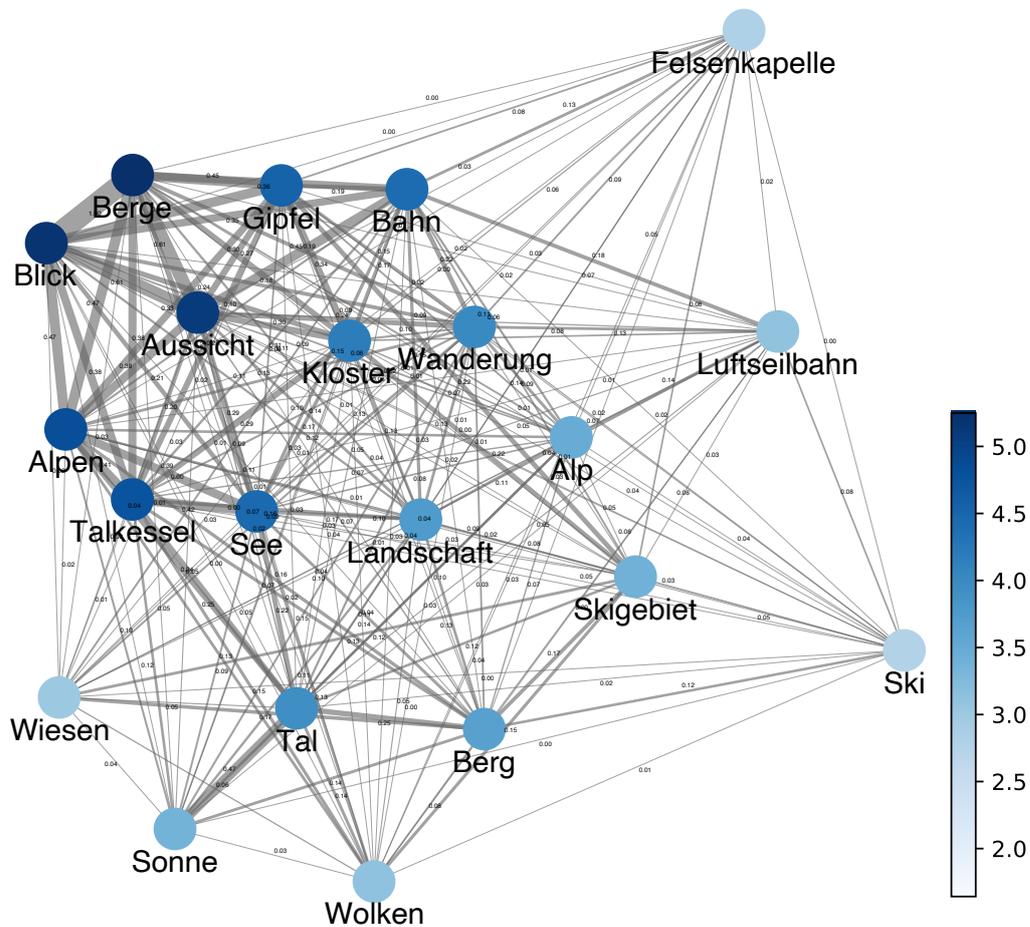
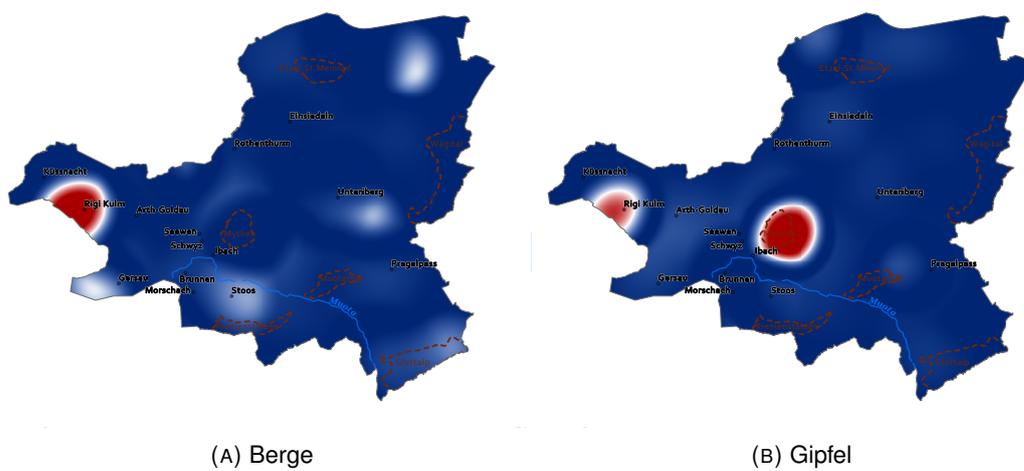


ABBILDUNG 6.14: Netzwerk der häufigsten Landschaftsbegriffe (eingefärbt nach der Stärke ihrer aufsummierten Kosinus-Ähnlichkeit, die Kanten verbinden LSB, die mit Kookkurrenzen mit gleichen LSB teilen).

Die häufigsten LSB können inhaltlich in drei Kategorien eingeteilt werden, nämlich in anthropogene Objekte (Kloster, Luftseilbahn), in geomorphologische (Berg, See) und in Wetterphänomene (Sonne, Wolken). Es ist aber nicht so, dass diese drei Kategorien besonders häufig zusammen auftreten. Die grösste Kategorie sind die geomorphologischen Begriffe, die wohl auch klassisch als Teil der Landschaft gesehen werden. Diese sind es auch, die am meisten mit anderen LSB zusammen vorkommen.

Das Cluster Aussicht-Blick-Berge-Gipfel, welches sich in den Texten so ähnlich ist, wird in den Karten 6.15 und 6.16 räumlich verglichen. Da diese LSB zu den häufigeren gehören, ist in den Dichtekarten schwierig zu erkennen, wo sie überdurchschnittlich oft vorkommen.  $\chi^2$ -Karten visualisieren solche Sachverhalte besser. Es wird ersichtlich, dass Kookkurrenz in Texten überhaupt nicht mit Kookkurrenz im Raum korrelieren muss. Sie korrelieren allerdings, nicht überraschenderweise,

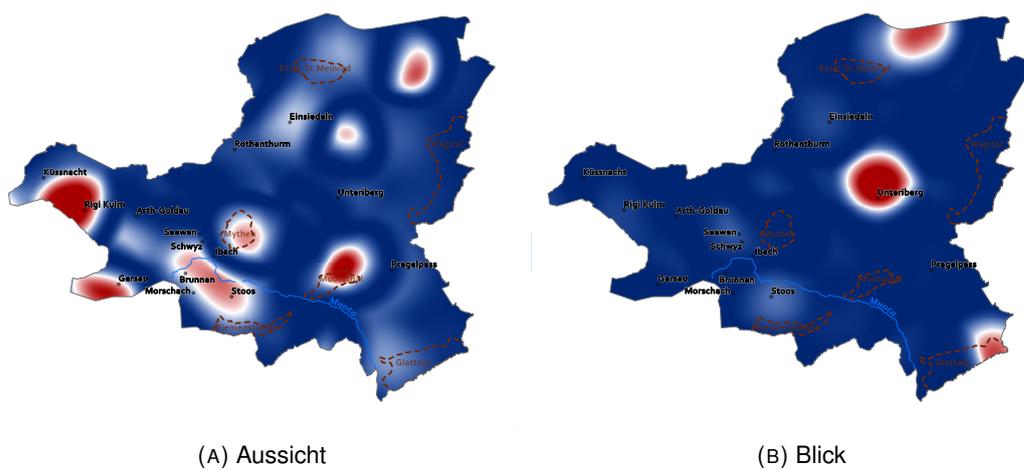
mit der Topographie. Die Rigi scheint ganz stark mit *Bergen* assoziiert zu werden, etwas weniger stark der Fronalpstock beim Stoos. Die Spitzen der Mythen sind ikonische *Gipfel*, aber auch damit wurde die Rigi öfters als durchschnittlich genannt. Die Orte Lachen und Etzel im Norden werden oft mit *Blick* genannt. Der Hotspot um Unteriberg bezieht sich auf das Ybrig-Gebiet, welches von Unteriberg erreicht werden kann. Die Form des Hotspots dehnt sich nach Norden aus, weil da ein Ort namens Hau liegt. Das ist eines jener Toponyme, die so allgemein sind, dass ihre Disambiguierung schwierig ist. Die beiden Gipfel nördlich des Schlüsselgebiets Muotathal, von denen man wohl eine *Aussicht* hat, sind der Forstberg und der Druesberg. Klassischerweise werden auch der Mythen und die Rigi oft mit *Aussicht* genannt.



(A) Berge

(B) Gipfel

ABBILDUNG 6.15: Chi²-Karten: Berge und Gipfel.



(A) Aussicht

(B) Blick

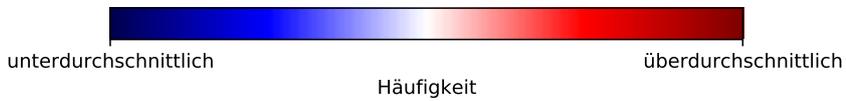


ABBILDUNG 6.16: Chi²-Karten: Aussicht und Blick.

Die beiden Dichtekarten in Abbildung 6.17 sind anschauliche Beispiele, wo die Verteilung der LSB selbst ohne  $\chi^2$ -Test interessante Muster zeigt. Der See wird auffälliger Richtung Süden am Vierwaldstättersee, bei Einsiedeln am Sihlsee und auch im Norden dem Zürichsee entlang. Vom Mythen herab hat man eine gute Aussicht auf den See. Der Stoos scheint gute Werbung für sein *Skigebiet* zu machen. Vergleicht man diese Dichtekarte mit der Dichtekarte 6.8 der gefilterten Toponyme im Boten, dann erklärt das auch, warum *Ski* und *Skigebiete* in der Abbildung 6.11 nur dem Boten zugeordnet sind.

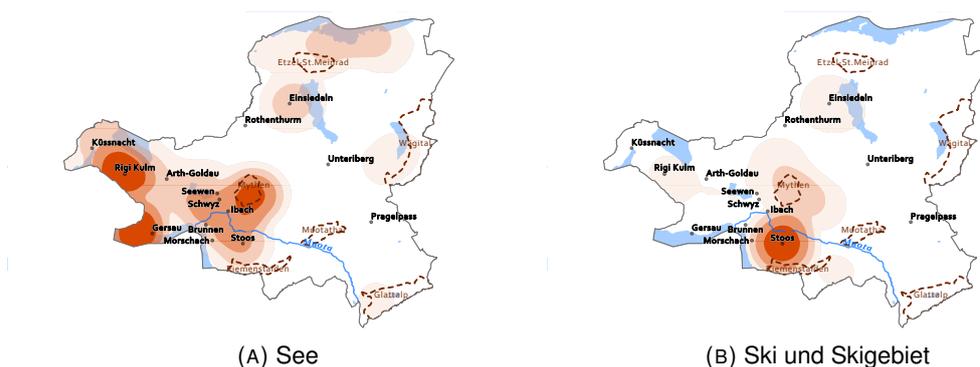


ABBILDUNG 6.17: Dichtekarten zweier Landschaftsbegriffe: See und Ski & Skigebiet.

## 6.4 Fokus Schlüsselgebiete

Nun zu den sechs Schlüsselgebieten aus der Landschaftskonzeption Schwyz: aus der Landschaftskonzeption stehen uns für diese Gebiete besonders detaillierte Beschreibungen zur Verfügung. Bevor diese mit den Beschreibungen der Korpora verglichen werden, folgen zuerst einige Resultate zu den Schlüsselgebieten. Die Resultate sind aus dem Boten und dem Y-Magazin aggregiert, weil hier die Schlüsselgebiete im Vergleich zum restlichen Kanton interessieren und nicht die Unterschiede zwischen den Korpora.

Wie bereits in den Karten im Kapitel 6.2 ersichtlich wurde, sind die Toponymdichten der gefilterten Toponyme auch innerhalb der Schlüsselgebieten tiefer als in anderen Gebieten des Kantons. Tatsächlich handelt es sich um einen Faktor drei (s. Tab. 6.3). Ein anderes Bild bietet die Tabelle 6.4. Ein Toponym innerhalb eines Schlüsselgebietes wird im Schnitt 1.5 mal öfter mit einem LSB beschrieben als ein Toponym ausserhalb!

Nicht nur räumlich, sondern auch inhaltlich unterscheiden sich die Schlüsselgebiete vom restlichen Kantonsgebiet. In der Abbildung 6.18 sind die LSB, die entweder innerhalb oder ausserhalb der Schlüsselgebiete mehr als 1% der Vorkommnisse ausmachen. Diejenigen ausserhalb der Schlüsselgebiete sind bereits bekannt aus

Toponym pro km <sup>2</sup>	
innerhalb der SG	7.36
ausserhalb der SG	24.5
ganzer Kanton	23.15

TABELLE 6.3: Toponymdichten innerhalb und ausserhalb der Schlüsselgebiete.

LSB pro	km <sup>2</sup>	Toponym
innerhalb der SG	9.94	14.82
ausserhalb der SG	38.62	9.86
ganzer Kanton	36.31	9.93

TABELLE 6.4: LSB-Dichten innerhalb und ausserhalb der Schlüsselgebiete.

den allgemeinen Resultaten der Korpora. Fokussiert man jedoch auf die Schlüsselgebiete, treten neue Begriffe auf wie *Föhn*, *Mythengebiet*, *Naturschutzgebiet*, *Sonnenaufgang* und *Burg*.

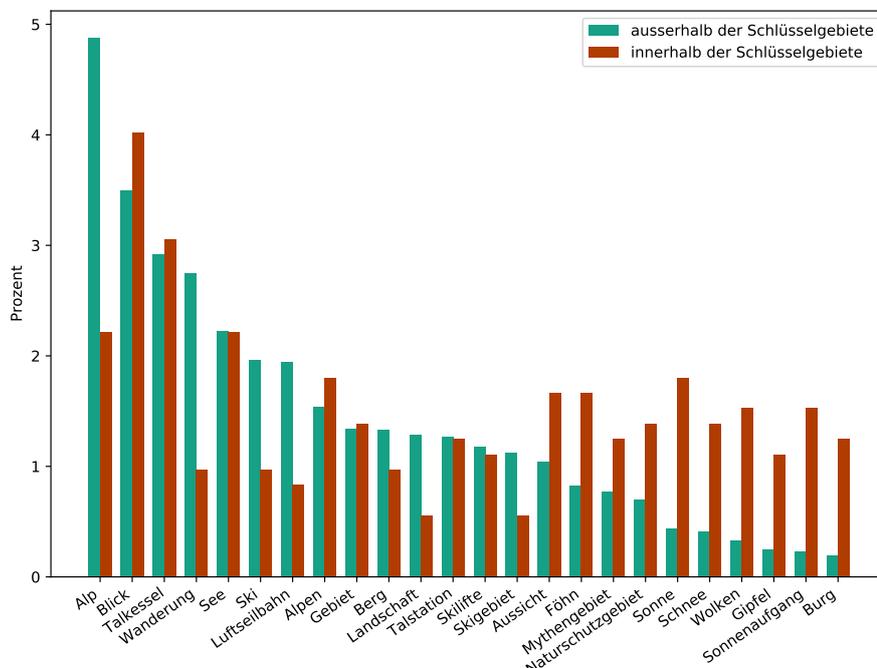


ABBILDUNG 6.18: Häufigste Landschaftsbegriffe ausserhalb und innerhalb der Schlüsselgebiete kombiniert aus beiden Korpora.





Schlüsselgebiete gut charakterisieren (s. Abb. 6.24 zum Muotathal).

Am ähnlichsten sind sich das Mythengebiet und die Glattalp mit einem Kosinusindex von 0.21 (vergleiche Tab. 6.5). Das ist wohl dem Fakt geschuldet, dass beide viele LSB beinhalten, was eine Überlappung wahrscheinlicher macht. Denn liest man die kleineren Wörter in den Word Clouds, dann erschliesst sich einem ein anderes Bild der beiden Schlüsselgebiete. Obwohl in der Glattalp *Tourismus* recht gross geschrieben steht, scheint sie doch alpiner und abgeschiedener (Hochgebirge, Felsenlandschaft, Karstgebiet, Abgeschiedenheit). Beim Mythen fällt die *Aussicht*, möglicherweise auf den *See* und den *Talkessel* auf. Hier findet man relativ mehr anthropogene LSB zum Thema Skifahren und Wandern, sowie meteorologische LSB wie *Wolken*, *Sonnenaufgang*, *Nebelmeer*.

Die nächstähnlichen Schlüsselgebiete sind das Wägital und Riemenstalden (Kosinusindex 0.19), sowie die Glattalp und Riemenstalden (Kosinusindex 0.14). Dafür verantwortlich ist wohl die *Alp*. Weiter wird sowohl im Wägital, wie in Riemenstalden über Tiere und Wanderungen geschrieben und es scheint Seilbahnen zu geben, auch wenn in Riemenstalden nur anhand der Tal- und Bergstation darauf geschlossen werden kann.

Am unähnlichsten zu allen anderen und untereinander sind das Muotathal und Etzel-St.Meinrad. Das Muotathal ist geprägt von LSB, die mit Gewässern zu tun haben. *Kloster* wird eher klein geschrieben in Etzel-St.Meinrad, obwohl die Kapelle St. Meinrad zum Kloster Einsiedeln gehört. Die *Burg* fällt hingegen auf. Es handelt sich dabei um eine Verwechslung eines Toponyms mit einem LSB, eine Geo/Non-Geo-Verwechslung.

	Glattalp	Wägital	Riemenst.	Mythen	Etzel-St.M.	Muotathal
Glattalp	1.000	0.052	0.138	0.214	0.030	0.056
Wägital	0.052	1.000	0.196	0.067	0.043	0.000
Riemenstalden	0.138	0.196	1.000	0.073	0.034	0.000
Mythen	0.214	0.067	0.073	1.000	0.038	0.093
Etzel-St.Meinrad	0.030	0.043	0.034	0.038	1.000	0.032
Muotathal	0.056	0.000	0.000	0.093	0.032	1.000
Summe	1.489	1.358	1.440	1.485	1.176	1.180

TABELLE 6.5: Kosinus-Ähnlichkeiten zwischen den Schlüsselgebieten aufgrund der LSB-Kookurrenz.

Abschliessend soll die Landschaftskonzeption mit seinen Beschreibungen der Schlüsselgebiete den Beschreibungen in den Korpora gegenübergestellt werden. Meine Analyse der Landschaftskonzeption beschränkt sich auf die Schlüsselgebiete. In der Tabelle 6.6 werden trotzdem auch die Gebiete ausserhalb der Schlüsselgebiete mit der Charakterisierung der Landschaftskonzeption verglichen, um auch



## Kapitel 7

# Diskussion

Die Balkendiagramme der häufigsten Toponyme und die Dichtekarten zeigen klar, dass die Auswahl der Aussagen nicht bloss der zugrundeliegenden Verteilung der Toponyme im Korpus entspricht. Es ist mit meiner Pipeline folglich möglich, Orte, die mit einem gewissen Thema assoziiert werden, in einem Korpus zu finden und richtig zu verorten. Somit kann auch eine Aussage darüber gemacht werden, wie diese Orte von einer gewissen Textquelle charakterisiert werden. Die Inputdaten sind einfach anpassbar; sowohl der Gazetteer wie auch die Themenliste können je nach Frage ausgetauscht werden.

Es folgt, wie sich die anfangs gestellten Forschungsfragen mit diesem Gazetteer und dieser LSB-Liste beantworten lassen und welche Vorbehalte und Nachteile der verwendeten Methoden und gemachten Annahmen bei der Interpretation beachtet werden müssen.

### 7.1 Interpretation

Meine beiden Hypothesen, welche zu den Forschungsfragen führten, sind:

- dass sich die Schlüsselgebiete in der Anzahl und Art ihrer Landschaftsbegriffe vom Umland unterscheiden
- dass sie sich auch untereinander unterscheiden

Grundsätzlich können beide Hypothesen bestätigt werden. Im Schnitt haben die Schlüsselgebiete weniger Toponyme pro km<sup>2</sup>, aber mehr LSB pro Toponym. Die Toponymdichten unterscheiden sich aber je nach Lage des Schlüsselgebietes. Die Art der Toponyme unterscheidet sich für vier der sechs Schlüsselgebiete. Glattalp und Mythen folgen in der Anzahl und Art eher dem Gesamtkorpus im restlichen Kantonsgebiet. Auch die zweite Hypothese kann bestätigt werden. Gemäss des Kosinusindexes unterscheiden sich alle Schlüsselgebiete von einander.

### 7.1.1 Forschungsfragen

#### Forschungsfrage 1

*Wie unterscheidet sich räumliche Verteilung von Toponymen in Sätzen mit einem Landschaftsbezug von der räumlichen Verteilung aller Toponyme im Kanton?*

Wie im Kapitel 6.2 beschrieben, kann für beide Korpora grob gesagt werden, dass die Toponyme sich weg vom urbanen Gebiet hin zu ländlicheren, natürlicheren Gebieten bewegen. Die Chi<sup>2</sup>-Karte 7.1 verdeutlicht repräsentativ für das Y-Magazin, dass aber nicht das komplett ländliche oder alpine Kantonsgebiet besonders häufig mit LSB beschrieben wird. Stattdessen werden bekannte Freizeitgebiete wie die Mythenregion, der Stoss mit dem Klingenstock und Fronalpstock, der Pragelpass und die Ybergregion. Die meisten dieser Regionen sind im Winter wie im Sommer beliebt zum Wandern, Skifahren und für Skitouren. Die Rigi als die Königin der Berge und Tourismushotspot schlechthin wird trotzdem nicht häufiger mit LSB genannt. Das kann daran liegen, dass die Rigi zwar schweizweit und international beliebt ist, die Schwyzer und Schwyzerinnen selber aber andere lokale Naherholungsgebiete nutzen. Sowohl der Bote der Urschweiz wie auch das Y-Magazin adressieren die Schwyzerbevölkerung als ihr Leserschaft.

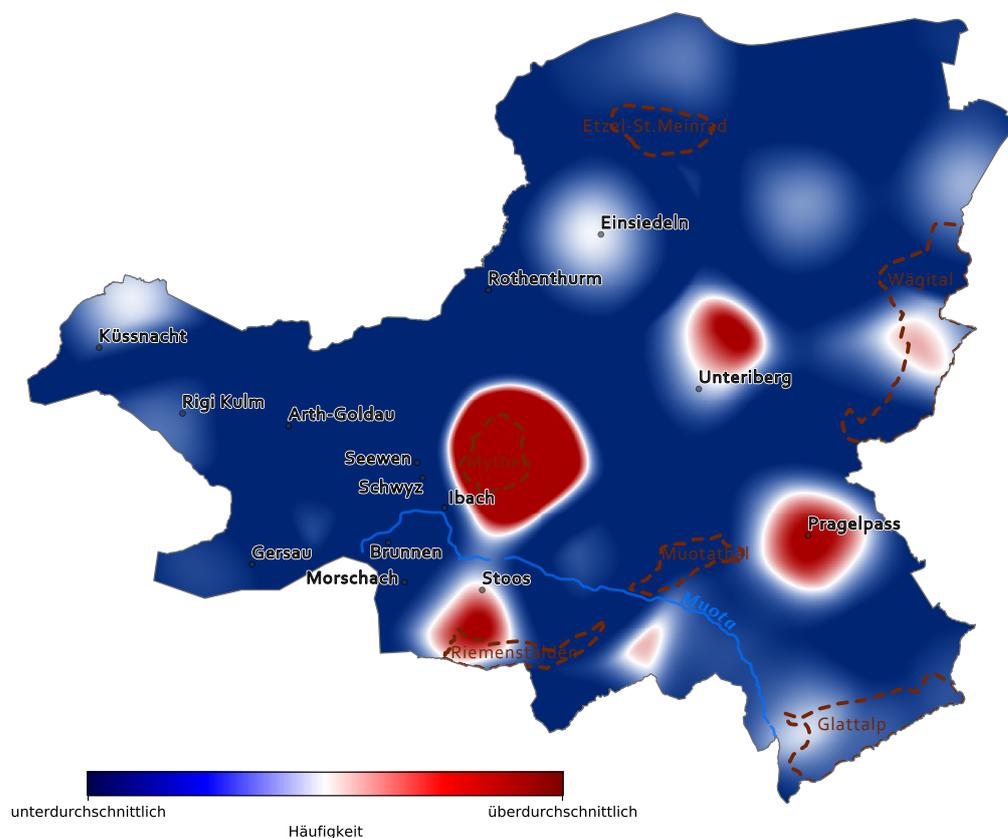


ABBILDUNG 7.1: Chi<sup>2</sup>-Karte der Toponymdichte im Y-Magazin.

## Forschungsfrage 2

*Wie unterscheiden sich die Dichten an verorteten Landschaftsbegriffen jeweils zwischen schützenswerten Gebieten (d.h. Schlüsselgebiete) und dem restlichen Kantonsgebiet?*

Die Dichte an *verorteten* LSB ist in den Schlüsselgebieten tiefer als ausserhalb (s. Kap. 6.4). Das ist aber der tieferen Dichte an Toponymen in den Schlüsselgebieten geschuldet, die bereits in den ungefilterten Korpora auftritt (s. Abb. 6.7 und 6.9). Sieht man sich aber die durchschnittliche Anzahl LSB pro Toponym an, dann sind es viel mehr in den Schlüsselgebieten. Das lässt die Vermutung zu, dass die Schlüsselgebiete tatsächlich eine aussergewöhnliche Landschaft besitzen, die es sich lohnt ausführlich zu beschreiben in Medien.

## Forschungsfrage 3

*Wie unterscheiden sich die Landschaftsbegriffe zwischen den Korpora und der Landschaftskonzeption in den Schlüsselgebieten?*

Die Beschreibung der Schlüsselgebiete unterscheiden sich stark (s. Tab. 6.6). Es scheint, die professionelle Sprache der Landschaftskonzeption lässt sich nicht mit der Alltagssprache in einer Zeitung oder einem Magazin vergleichen. So lange alle, die mit der Konzeption arbeiten, die gleiche Vorstellung unter den benutzten LSB haben, scheint es nicht notwendig, mehr Alltagssprache einfliessen zu lassen. Aber es wäre sicher spannend für die Nutzerinnen und Nutzer der Landschaftskonzeption, zu sehen, wie diese eher abstrakten Beschreibungen der Schlüsselgebiete mit einfacheren Worten für die Menschen konkret werden.

## 7.2 Einschränkungen

Es ist wichtig, sich der Annahmen und Limitierungen in der Datenauswahl und den Methoden bewusst zu sein, um die Resultate richtig zu interpretieren. Besonders grundsätzliche Einschränkungen noch vor dem Beginn oder am Anfang der Arbeit, wandern immer weiter und nachgelagerte Analysen werden davon beeinflusst (Hu, 2018).

### Grundsätzliche Prämisse

Daher muss auch die grundsätzliche Prämisse, dass Artikel eine gute Grundlage bieten, um die Landschaft zu beschreiben, hinterfragt werden. Wie sehr können Texte unsere gesamte Wahrnehmung wiedergeben? Und können wir uns zuhause am Computer noch daran erinnern, was uns draussen begegnet ist? Tatsächlich findet

eine Reduktion der Landschaftswahrnehmung statt, wenn nur einzelne Wörter genutzt werden, um sie zu beschreiben. Landschaft wird nicht nur als eine Sammlung an Begriffen wahrgenommen, sondern auch als eine Atmosphäre oder als etwas, dass nur in unserem Kopf erst entsteht und sich einem anderen Beobachter an der gleichen Stelle nicht erschliessen würde (Rees, 2016, Lingiardi, 2020). Weiter, ist es erwiesen, dass uns je nach Umgebung unterschiedliche Dinge in den Sinn kommen (Wartmann u. a., 2015). So macht es einen Unterschied, ob Leute im Feld nach Landschaftsbegriffen befragt werden oder auf einem Foto oder zuhause. Es sind folglich verschiedene Arten, der Datenerhebung nötig, um möglichst viele Aspekte der Landschaft abzudecken (Wartmann, Acheson und Purves, 2018). Neben diesen grundsätzlichen Einschränkungen werden nun weitere beschrieben, deren Einfluss aber durch verschiedene Erwägungen gemildert werden kann.

### 7.2.1 Datenwahl und Textaufbereitung

#### Auswahl

Die Quelle der Daten bestimmt gänzlich, welche Muster überhaupt je erkannt werden können. Dabei beeinflusst die Quelle sowohl den Inhalt, die Auswahl der Themen und die Sicht darauf, wie auch die räumliche Abdeckung. Der Bote der Urschweiz deckt laut Selbstbeschreibung seiner Berichterstattung den ganzen Kanton ab, laut Dr. Schwitler eher den inneren Kanton. Wobei weitere, aber kleinere Zeitungen andere Bezirke stärker abdecken würden, z.B. das Höfner Volksblatt und der March Anzeiger. Die Verteilung der Toponyme innerhalb der Zeitung folgt tatsächlich im Südwesten vorwiegend der Verteilung der Namen im Gazetteer (s. Abb. 7.2). Im Norden und entlang des Sihlsees scheinen die Punktoponyme des Boten relativ seltener zu sein. aber vorwiegend der Verteilung der Namen im Gazetteer, weshalb davon ausgegangen werden kann, dass die Berichterstattung zumindest auf die benannten Gebiete bezogen fair verteilt ist.

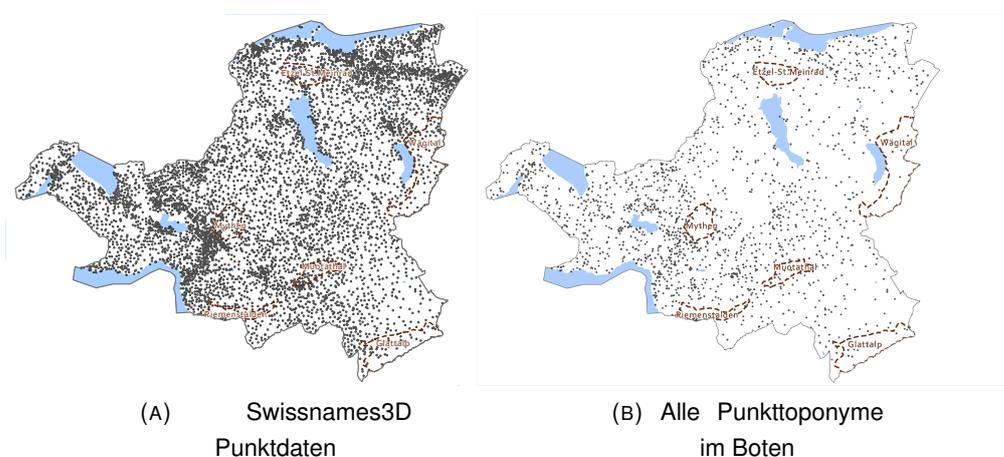


ABBILDUNG 7.2: Punktoponyme im Gazetteer Swissnames3D und im Boten der Urschweiz

Inhaltlich würden mehr Quellen vermutlich auch zu mehr Diversität und zu neuen Sichtweisen führen, da Beschreibungen innerhalb eines Korpus sich gerne ähnlicher sind als zu einer anderen Datenquelle, auch wenn es sich um den gleichen Landschaftstyp handelt (Wartmann, Acheson und Purves, 2018, Koblet und Purves, 2020). Die zeitliche Komponente wird nicht beachtet. Zwar habe ich die Ausgaben so ausgewählt, dass sie etwa in die Zeit des neuen Raumplanungsgesetzes fallen (s. Kap. 4.1), es werden aber keine historischen Aussagen rausgefiltert, wie z.B. diese Beschreibung von Goethe über Schwyz und den Mythen (damals Haken): „Ueber den See hinweg sieht man die grünen Matten, Dörfer und Höfe von Schwyz im hellen Sonnenschein liegen. Zur Linken des Zuschauers zeigen sich die Spitzen des Haken, mit Wolken umgeben; zur rechten im fernen Hintergrund sieht man die Eisgebirge“ (*Y-Magazin Nr. 1 Sommer 2012*).

Die Daten zeichnen sich nicht durch eine grosse Vielfalt verschiedenen Partizipatoren aus. Es sind immer die gleichen Artikelschreiber, ab und zu gespickt mit einem Interview oder einem Meinungsbeitrag oder Lesebrief. Um diesen Bias zu verhindern, hätte ich die Anzahl Beiträge pro Autorin begrenzen können oder, wie in Chesnokova und Purves, 2018b, besonders ähnliche Beschreibungen eines Autors oder einer Autorin rausfiltern. Das würde allerdings nicht die Frage beantworten, ob es eine gewisse Zirkularität in den Texten gibt, falls z.B. die Landschaftskonzeption als Grundlage für einen Artikel verwendet wurde.

Auch wenn viele einzelne Aussagen aus den Texten extrahiert wurden, muss das also nicht ein grosses Spektrum an möglichen Aussagen abdecken. Die effektive Grösse von UGC-Daten muss nicht zwingend der Summe aller Datenpunkte entsprechen. Eine grosse Datenmenge ohne Variabilität in den "Befragten"(oder hier den Schreibenden) lässt trotzdem nicht Rückschlüsse auf eine grosse Menge der Bevölkerung zu. Meng u. a., 2018 beschreibt wie die effektive Fallzahl berechnet werden kann und somit auch eine Aussage darüber ermöglicht wird, wie repräsentativ die Daten sind.

Die zweite wichtige Datenquelle ist die Liste mit allen Landschaftsbegriffen. Die Liste mit den Landschaftswörtern stammt aus zwei Quellen und aus manuell gefundenen Tags bei der Annotation der Daten. Besonders die beiden Quellen von Huldi, 2015 und Derungs und Purves, 2014 bringen durch ihren Forschungskontext einen Bias besonders für natürliche Formen rein, während Tätigkeiten (z.B. Freizeitaktivitäten) und Atmosphärisches (z.B. Abendrot) seltener sind. Weiter kommen in der ganzen LSB-Liste nur Nomen vor. Diese könnten ergänzt werden mit Wahrnehmungsverben und bestimmten Adjektiven, die oft im Landschaftskontext vorkommen und Hinweise auf Tätigkeiten in und Qualitäten von der Landschaft geben (Chesnokova, Nowak und Purves, 2017). Die ganze Landschaftscharakterisierung kann nur innerhalb der Begrenzung dieser Liste stattfinden. Das stellt natürlich für diese Analyse eine Einschränkung dar. Andererseits kann die Liste jederzeit um weitere Begriffe erweitert werden. Wenn weiterhin ein Machine Learning Algorithmus angewendet wird, um die Sätze mit neuen Begriffen zu klassieren, dann müsste dieses wieder

frisch trainiert werden. Dabei muss man sich des Zeitaufwandes bewusst sein, der der Aufbau von neuen Trainingsdaten mitbringt.

Die Auswahl der Begriffe und das Annotieren von landschaftsrelevanten und nicht landschaftsrelevanten Sätzen für das Modell beruhen auf selber aufgestellten Regeln und die Resultate sind daher einem starken Bias unterworfen. Die Regeln für das Annotieren sollten durch weitere AnnotatorInnen auf ihre Übertragbarkeit geprüft werden, was in dieser Arbeit nicht gemacht wurde. Selbst dann ist die Übertragbarkeit auf andere Korpora nicht gegeben, da es schwierig ist beim Annotieren den gegebenen Kontext aussen vor zu lassen (Chesnokova und Purves, 2018b). Um den Bias zu reduzieren, wären, neben der Übereinstimmung zwischen den AnnotatorInnen, auch Sensitivitätsanalysen eine gute Möglichkeit die Robustheit der Annotation zu testen. Dabei werden Entscheidungen, d.h. Annahmen zu was alles ein LSB ist oder welche Trainingssätze verwendet werden, geändert und die Resultate beobachtet. So können auch die entscheidenden Steuergrößen im Ablauf gefunden werden (Dinkelbach, 2013).

### **Textaufbereitung**

Wie bereits in der Abbildung 6.4 gezeigt, führt die Konvertierung der PDF-Dateien zu reinem Text zu Problemen wie unvollständigen Sätzen und vermischten Absätzen aus verschiedenen Artikeln. Dem könnte mit einer breiteren Textverarbeitung mit Regeln begegnet werden, z.B. wie Zeilenumbrüche gehandhabt werden sollten. Mit einer stärkeren Anwendung von NLP-Methoden wie dem Part-of-Speech Tagging könnten unsinnige Sätze besser erkannt und gefiltert werden. Dadurch, dass ich eine *Aussage* auf einen Satz beschränke, wird eine Menge Information bereits im Vorhinein weggelassen. Schliesslich können Landschaftsbeschreibungen und der zugehörige Ort auch über mehrere Sätze verteilt sein. Aber diese Vorgehensweise bedeutet eine einfache Implementation und verkleinert das Risiko, dass ein Toponym und ein LSB fälschlicherweise kombiniert werden. Trotzdem kann man sich bei der automatischen Textanalyse nie ganz sicher sein, ob der Link zwischen dem Ort und dem Text tatsächlich besteht. Diese Einschränkung könnte mit Part-of-Speech Tagging gemildert werden. Weiter können pro Satz sowohl mehrere LSB als auch mehrere Toponyme in jeder Kombination textuell zusammengehören (Hu, 2018). Als Beispiel die folgende Aussage aus dem Y-Magazin: „Man sieht einen Streifen des Vierwaldstätter Sees beschneite Gebirge jenseits der Eingang ins Muotatal aus dem Tal erscheint links“. In dieser Arbeit ging ich von einem maximalen Kontext aus, wobei in einem Satz alle Toponyme mit allen LSB beschrieben werden, d.h. dass sowohl das Muotatal als auch der Vierwaldstättersee mit See, Gebirge und Tal assoziiert werden.

## 7.2.2 GIR- und NLP-Methoden

### Toponym Recognition

Die Named Entity Recognition klassiert nicht immer nachvollziehbar. Klar wurde in Tests mit Fantasiewörtern, dass Präpositionen wie *entlang*, *zwischen*, *unweit von*, *hinter*, *ausserhalb* von spacy genutzt werden, um Örtlichkeiten zu erkennen. Aber die beiden Aussagen „Mit 80 km durch die Kuhherde Suter kennt die Situation mit den Wildunfälle“ und „Damit bleibt der Stauffacher weiterhin in der Familie Steiner Bissig erhalten“ enthalten zwar wohl Schwyzer Toponyme (*Suter* und *Steiner*), aber für einen Menschen wäre klar, dass es sich in diesem Zusammenhang um Nachnamen handelt. Solchen Problemen wurde mit stärkeren Filtern bei den Toponymen begegnet.

Im deutschen Modell steht für die Erkennung von Orten der Tag LOC für alle Örtlichkeiten zur Verfügung, während im Englisch zusätzlich die Unterscheidung zwischen LOC (Berge, Gewässer, etc.) und GPE (Geopolitical entity, d.h. Länder, Städte) genutzt werden könnte (Honnibal u. a., 2020). Eine solche Erweiterung auch für die deutsche Sprache könnte verhindern, dass z.B. Alp oder Tal als Örtlichkeiten anstatt als Flurnamen klassiert werden.

Ausser der verwendeten Matchingmethode (s. Kap. 5.1.3), in der die Ortsnamen abgesehen von ihrer Endung genau der Schreibweise im Gazetteer entsprechen mussten, würde ein weniger rigides Matching wie z.B. Fuzzy-Matching mit der Levenshteindistanz vielleicht bessere Resultate liefern (Chesnokova und Purves, 2018a). Gazetteerergänzung könnte mit den schweizerdeutschen Namen helfen (Acheson, Volpi und Purves, 2020). Der Gazetteer von Swisstopo enthält nur offizielle Namen, die auch auf Karten auftauchen. Gerade in der Schweiz mit vielen Dialekten und einer unregelmässigen Rechtschreibung der Umgangssprache sind andere Schreibweisen häufig. Diese wurden hier nur punktuell ergänzt. Für den Kanton Schwyz bestünde schon ein umfangreicherer Gazetteer, das Schwyzer Namenbuch (Weibel und Hug, 2012). Mit Beziehungswörtern (wie *hinter*, *in der Nähe von*, *unweit von*) könnte versucht werden, sowohl die Toponym Recognition als auch die Toponym Resolution genauer zu gestalten. Beziehungswörter weisen auf Toponyme hin und helfen so bei der Geo/Non-Geo-Disambiguierung. Die Verortung würde wohl nicht genauer, da sich das kontextuelle Verständnis der Beziehungswörter nur schwer codieren lässt (Egorova u. a., 2018). Aber es würde die Unschärfe integriert werden, die zeigt, dass es sich eben nicht punktgenau um das beschriebene Toponym handelt.

### Toponym Resolution

Die grösste Einschränkung davon, wie einem Toponym die richtigen Koordinaten zugewiesen werden, ist offensichtlich, dass ich jegliche kontextuelle Hinweise ignoriere und eine rein distanzbasierte Methode anwende (Buscaldi, 2011).

Es gibt verschiedene Möglichkeiten, wie die Toponym Resolution angegangen

werden kann. Eine naive Methode wäre eine Auswahl nach dem Zufallsprinzip. Das macht Sinn, wenn nur keinerlei andere Anhaltspunkte vorhanden sind oder die Texte vor allem einzigartige Toponyme beinhalten, so dass sich der Aufwand einer spezifischeren Methode nicht lohnt. Da der *Swissnames3D-Gazetteer* bereits eine gute Reihenfolge der Kandidatennamen liefert, erhält man auch gute Resultate, wenn davon die oberste Option gewählt wird (Acheson, Wartmann und Purves, 2018).

Als einfachstes Kontextmittel könnte auch der Geometrietyp des Toponyms beachtet werden. Dabei wäre es sinnvoll, Polygonen den Vorrang vor Punkten zu geben, weil Punkte oft Haltestellen oder Ausfahrten referenzieren, anstatt wie Polygone ein ganzes Gemeindegebiet oder ein ganzes Tal. Weiter enthält der *Swissnames3D-Gazetteer* nicht nur Namen und Koordinaten, sondern auch weitere Attribute, die die Örtlichkeiten beschreiben, wie z.B. um was es sich eigentlich handelt (Gebäude, Gipfel, Bahnen, Fliessgewässer). Mit mehr Wissen über das Korpus können also auch Regeln verfasst werden, die aufgrund dieser Attribute ein Toponym auswählen. Innerhalb der Polygone wäre ein Default zur Fläche mit den meisten Einwohnerzahlen zielführend, so lange dieses Attribut vorhanden ist (Buscaldi, 2011). Bei Punkt- und Linienobjekten hilft die Objektart weiter. War in dem Satz bereits von einem Fluss die Rede, so kann man das zweite Toponym ebenfalls dem Kandidaten vom Typ Fluss zuordnen anstatt dem Typ Tunnel.

Wenn vor allem die Verteilung im Raum als Basis verwendet wird, dann wird davon ausgegangen, dass es sich beim richtigen Toponym um einen eher zentraler gelegenen Punkt handeln muss und dass je weiter entfernt ein Punkt liegt, desto unwahrscheinlicher, dass es sich um den gesuchten Ort handelt (Acheson, Volpi und Purves, 2020). Koblet und Purves, 2020 erreichen gute Resultate sowohl mit dem DBSCAN-Clustering Algorithmus, der Ausreisser identifiziert, als auch mit einem distanzbasierten Filter vom Zentroid aller Punkte her. Zuletzt liesse sich auch ein Machine Learning Modell trainieren, was allerdings wieder einen gewissen Zeitaufwand für die Aufbereitung von Trainingsdaten bedeuten würde (Acheson, Volpi und Purves, 2020).

Zumindest ein gewisses Kontextwissen zum Korpus und der Gegend muss mitgebracht werden. Das sieht man daran, dass in einer ersten Runde der Koordinatenzuweisung für die Rigi häufig ein kleines Dorf im Norden des Kantons anstatt des Massives gewählt wurde (s. Abb. 7.3). Ausserdem gibt es auch ein Zug im Kanton Schwyz, welches später aus dem *Gazetteer* entfernt werden musste, da die meisten Nennungen sich auf die grössere Stadt und den Kanton bezogen. Der Erfolg der Toponym Resolution wurde nur qualitativ für die 20 häufigsten Toponyme getestet. Die mittleren Fehlerdistanz zwischen den kartierten und den richtigen Toponymen wäre eine passende Metrik für die quantitative Auswertung (Gritta u. a., 2018).

Trotzdem hat meine Art der Koordinatenzuweisung ziemlich gut funktioniert. Das erkennt man an einem auffälligen Unterschied zwischen einer ersten Toponymdichtekarte des ganzen Korpus Bote (s. Abb. 7.3) und der Toponymdichtekarte der ausgewählten Sätze. Die Toponyme des ganzen Korpus wurden nämlich nicht mit dem

distanzbasierten Algorithmus disambiguiert, da das zu aufwändig gewesen wäre. Stattdessen habe ich ihnen zufällig ein Koordinatenpaar zugewiesen. Dabei entstand ein Cluster im Nordosten des Kantons. Dort liegt wieder ein kleines Dorf namens Brunnen, welches mit dem viel grösseren Brunnen im Norden verwechselt wurde. Bei den gefilterten Sätzen, die mit dem Algorithmus disambiguiert wurden, fällt dieses Problem aber nicht auf.

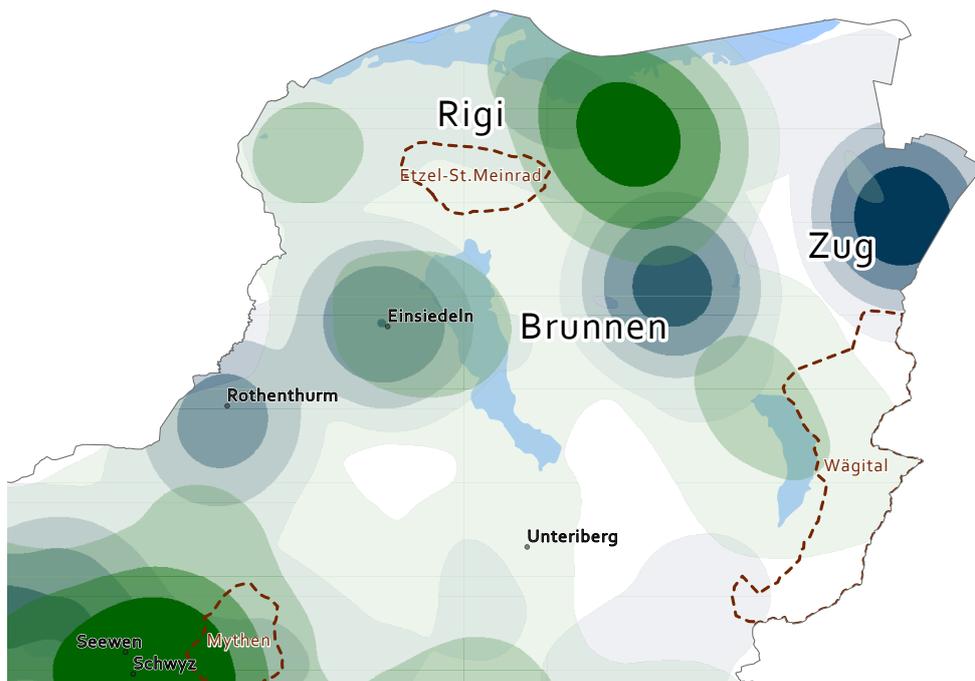


ABBILDUNG 7.3: Falsch disambiguierte Toponyme Rigi, Brunnen und Zug im Norden des Kantons.

### Landschaftskontext

Gewisse Schwierigkeiten entstehen durch die selber erstellte und erweiterte LSB-Liste und durch die Anwendung von Regex. Die sehr ausführliche Liste ermöglichte zwar, dass auf viel rechenintensives Regex verzichtet werden konnte, aber dieses Hardcoding von Komposita schränkt auch ein. Landschaftsbegriffe die in Teilwörtern vorkommen und nicht in der Liste sind, werden einfach nicht gefunden. Einige besonders häufige Start- und Endkomponenten wurden deshalb trotzdem noch mit Regex gesucht, z.B. -landschaft. Dabei treten Fehler auf. Z.B. der *Blüenberg* in Riemensalden oder *Eishockey* im Wägital (s. Abb. 6.20 und 6.21). Gespeichert wird aber das ganze Wort und dadurch zählen diese Begriffe nicht zu Berg oder Eis dazu. Dafür fallen gewisse Landschaftsbegriffe weniger auf, weil sie als Teil eines Kompositums weniger häufig sind als der reine Wortteil. Weiter sollte eine Regel geschrieben werden, dass Toponyme nicht LSB sein dürfen. In Etzel - St. Meinrad gibt es einen kleinen Feld namens *Deponi* neben der Gemeinde Etzelweid. Dieses

wurde korrekterweise als Toponym erkannt und fälschlicherweise zusammen mit der Endung +e als *Deponie* (s. Abb. 6.23).

Da zuerst im Text nach LSB gesucht wird, bevor ein Modell bestimmt, ob diese Landschaftskontext beinhalten oder nicht, soll die LSB-Liste doch vor allem klar landschaftsnahe Begriffe enthalten. Ansonsten müssten viel mehr Sätze durch das Modell laufen und die Precision würde gesenkt oder die Trainingsdaten müssten vergrössert werden. Hier sind weniger automatisierte Vorgehen wie Free-Listing im Vorteil (Wartmann und Purves, 2018). Es können Wörter vorkommen, die auf den ersten Blick nichts mit Landschaft zu tun haben (z.B. Asylheim oder Mücken), die aber für die Person wichtig sind und eine ein Gefühl für den Ort vermitteln.

Obwohl das Random Forest Modell zur Einteilung der Landschaftsrelevanz einzelner Aussagen auf den Testdaten gut funktionierte (82% overall precision), hat es - und NLP-Methoden allgemein - noch Schwierigkeiten, Textkontexte gut zu verstehen. Das Modell baut auf tf-idf und Bag of Words auf. Dabei wird angenommen, dass wichtige Wörter öfter vorkommen (tf-idf), obwohl auch seltene Begriffe wichtige Konzepte beinhalten können (Hu, 2018). Kleine Sensitivitätstests mit verschiedenen Kombinationen von Training- und Testdaten lieferten robuste Ergebnisse (s. Tab. 6.1). Aber die Themen zur Landschaft sind schier unendlich und es ist zeitintensiv, einen grossen Trainingsdatensatz aufzubauen, weshalb die Sensitivitätstests nicht erweitert wurden. Wichtig scheint mir die Erweiterung des Modells, Sätze mit Negierungen und bestimmte Beziehungswörtern, die auf grosse Entfernung hindeuten, zu erkennen. Eine Aussage zum Schlüsselgebiet Etzel-St.Meinrad lautet nämlich „... weit entfernt von Wohngebiet in den Bergen...“. Die Word Cloud (s. Abb. 6.23) gibt hier ein falsches Bild wieder.

### 7.2.3 Interpretation

#### Dichtekarten

Die verschiedenen Dichtekarten erwecken den Schein, als seien Toponyme und LSB vor allem im Südwesten des Kantons anzutreffen, während besonders im Südosten und in der Kantonsmitte wenig Spannendes anzutreffen ist. Natürlich stimmt es, dass dort die Toponyme sind, wo auch die Menschen sind und je dichter die Menschen desto dichter die Toponyme (Hunn, 1994). Es entsteht auch eine gewisse Zirkularität; Wo es Menschen hat, kann Landschaft überhaupt wahrgenommen werden und da wo es Menschen hat, setzt ein Landschaftskonzept überhaupt erst an. Unerreichbare Gletscher müssen nicht reguliert werden.

Bezogen auf die Verteilung der Toponyme mit Landschaftskontext gäbe es zwischen den oben genannten Hotspots wie dem Mythen oder dem Pragelpass noch ganz viel Landschaftsgebiete, von Landwirtschaftsflächen über Seeufer bis zu den Glarneralpen. Die Karte 6.5 gibt einen möglichen Hinweis darauf, warum die Punktverteilung so unregelmässig ist. Das gesamte südöstliche Gebiete mit den drei Schlüsselgebieten Glattalp, Muotathal und Riemenstalden ist dunkelgrün eingefärbt, enthält

also viele Toponymvorkommnisse. Aber diese Toponyme sind mit Polygonen assoziiert und die KDE läuft über Punktverteilungen. Da jedes Polygon, egal wie gross, nur mit einem Punkt repräsentiert ist, werden sie in der Dichtekarte unterschätzt. Die weniger dicht besiedelten Regionen haben weniger Toponyme, aber diese decken grössere Gebiete ab und das wird mit meinen Methoden nicht richtig wiedergegeben.

Die Ausbreitung und Dichte der Toponyme zu bestimmen mit Flächendaten, hätte weitere Fragen aufgeworfen. Während für Siedlungen zumindest politisch, wenn auch nicht funktional, recht klare Begrenzungen bestehen, ist die Ausdehnung von natürlichen Phänomenen viel schwieriger zu bestimmen. Auch wenn ein Gazetteer die Ausdehnung eines Berges oder eines Tales ganz klar definiert, so bleibt sie doch vage in Realität und muss nicht der alltäglichen Wahrnehmung von Beschreibenden entsprechen. Die Form und Repräsentation wird auch nicht gleich bleiben bei Gazetteers von verschiedenen räumlichen Auflösungen (Smith und Mark, 2003).

Selbst wenn also die Auflösung und Ausdehnung der Toponyme auf Seiten des Gazetteers eindeutig wären, bleibt die Unbekannte auf Seiten des Textes. Wir wissen nicht, ob ein Toponym im Ganzen gemeint ist oder, vermutlich meistens, nur ein Teil davon, z.B. die „Ansammlung prachtvoller Herrenhäuser in der Ortschaft Schwyz“ (*Landschaftskonzeption Kanton Schwyz. Version vom Februar 2019*, p. 15). Die Annahme, dass alle Toponyme und alle LSB in einer Aussage miteinander verknüpft sind, muss natürlich überhaupt nicht der Realität entsprechen. Die Disambiguierung der Toponyme basiert hier auf einem Clustering-Algorithmus und Ausreisser werden gemieden (Koblet und Purves, 2020). In Landschaftsbeschreibungen zur Aussicht könnten aber auch gerade diese Ausreisser gemeint sein.

All diese Unsicherheiten betreffend der räumlichen Fussabdrücke der Aussagen, könnten durch passende Visualisierungsmethoden wie Sättigungsgrade, Transparenz oder Überlagerungen kommuniziert werden (Hu, 2018).

### **Vergleich zur Landschaftskonzeption**

Der Vergleich zwischen der Landschaftskonzeption und den Aussagen aus der Zeitung und Magazin gestaltete sich schwierig. Ab wann sind Wörter *ähnlich*? Wie im Kapitel 6.4 beschrieben, differieren Alltagssprache in Laintexten und Expertensprache in der Konzeption so stark voneinander, dass eigentlich kaum eine Überlappung festgestellt werden kann. Hier wäre es interessant, die LSB in den Artikeln und in der Landschaftskonzeption Überkategorien zuzuordnen und diese miteinander zu vergleichen. Mithilfe von so genannten Hypernymen würden möglicherweise gemeinsame Konzepte sichtbar und die unterschiedliche Sprache stünde weniger im Weg (Manning und Schütze, 1999).

Aber die Schlüsselgebiete, die in der Landschaftskonzeption herausgestrichen werden, stellen sich auch in dieser Arbeit als speziell heraus, wenn es um die Anzahl LSB pro Toponym geht. Ausserdem liessen sie sich durch die gefundenen LSB

charakterisieren. Es wäre spannend zu sehen, ob durch eine Umkehrung der Herangehensweise weitere Gebiete mit höheren LSB-Beschreibungen pro Toponym und räumliche Cluster von ähnlichen Toponymen gefunden werden können. Diese könnten als mögliche Schlüsselgebiete weiter untersucht werden.

Die Toponymdichte ist in den meisten Schlüsselgebieten tiefer als im Rest des Kantons. Schaut man aber, wo die BLN-Inventare, die nationalen Schutzgebiete, liegen (s. Abb. 7.4) würden einige besonders dichte Gegenden durch diese abgedeckt werden (v.a. die Gegend des Zugersees bei Küsnacht, das Rigi-Kulm und der Lauerzersee östlich von Goldau). Eine Arbeit, die sich auf die Unterschiede in LSB in den BLN-Schutzgebieten fokussiert, kann hier weitere Einblicke liefern.

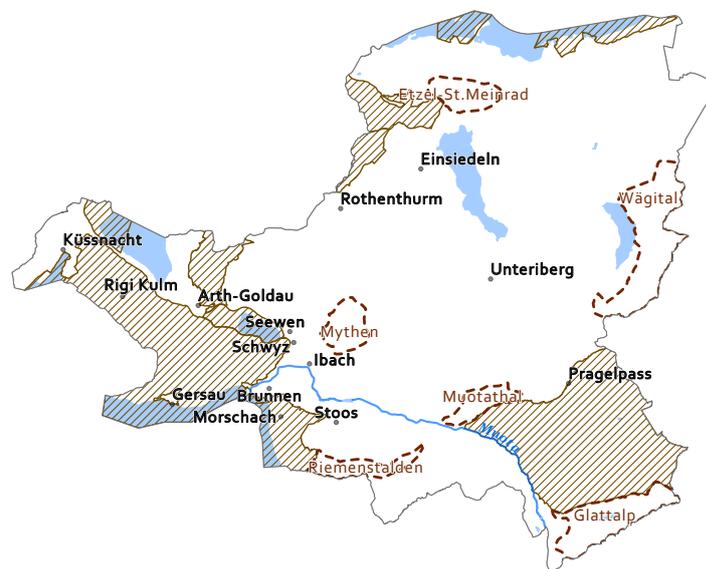


ABBILDUNG 7.4: BLN-Schutzgebiete im Kanton Schwyz.

## 7.2.4 Repräsentanz

Eine wichtige Frage, die sich VerfasserInnen eines Landschaftskonzept stellen müssen, und die sich auch in dieser Arbeit stellt, ist die Repräsentanz. Wessen Wahrnehmung der Landschaft wird wiedergegeben?

Zwei Ebenen machen die Nähe zur Quelle und Qualität der Landschaftsbeschreibung aus. Erstens kann eine Aussage von einer unpersönlichen Organisation gemacht werden oder sie kann direkt aus dem Kopf eines Beobachters zu Papier oder zu Worten gebracht werden. Zweitens können die Quellen eine möglichst hohe Durchmischung an BeobachterInnen aufweisen und so verschiedene Stakeholder repräsentieren.

Aussagen nur aus der Ich-Perspektive würde die Meinung von realen Personen wiedergeben, die möglicherweise sogar gerade kürzlich in der Gegend waren und mit frischem Gedächtnis schreiben, und nicht eine idealisierte Beschreibung von Landschaften aus einer Werbung für den Tourismus. Meine Korpora boten mir diese Möglichkeit nicht, da die Datengrundlagen extrem ausgedünnt worden wären. Das Y-Magazin enthält mehr Erlebnisberichte und direkte Zitate. Ich würde es aber nicht als eine unbeeinflusste Quelle bezeichnen, da es vom Amt für Wirtschaft des Kantons Schwyz herausgegeben wird und definitiv die attraktiven Seiten des Kantons betont. Der Bote der Urschweiz hingegen ist eher trocken und kann kaum eine sentimentale, gefühlvolle Verbindung mit der Landschaft aufzeigen, die so wichtig ist für Leute, die mit ihr interagieren (Campelo u. a., 2014).

Auch was die Darstellung möglichst vieler Perspektiven angeht, ist meine Arbeit stark eingeschränkt. Beinahe 100% der Aussagen kommen aus dem Boten der Urschweiz. Weitere Quellen, die stärker von Beginn weg einen landschaftlichen Hintergrund mitbringen, können z.B. von Umweltvereinen stammen oder von Webseiten, die mit konkreten Toponym- oder Landschaftssuchwörtern gefunden werden.

Der letzte Bias, der nicht ausgeglichen werden kann, besteht darin, dass nicht alle Leute über ihre Erlebnisse in der Landschaft schreiben und diese veröffentlichen. Das Alter, das Geschlecht, der soziokulturelle Hintergrund spielen alle eine Rolle darin, wie wir uns in der Umwelt *und* im Internet bewegen, und da eine ausgewogene Partizipation zu erreichen, ist schwierig. Und dennoch nötig, denn es gibt je nach kulturellem Hintergrund ganz andere Konzepte wie Objekte in der Landschaft beschrieben werden und wie sehr die Landschaft überhaupt in Objekten oder in einem Kontinuum wahrgenommen wird (Burenhult und Levinson, 2008). Die Beschäftigung in der Landschaft verändert die Sicht und Bewertung auch signifikant. Hier ist die Partizipation im Internet extrem ungleich. Vor allem besser ausgebildete und urbane, tech-affine Personen teilen ihre Fotos z.B. mit einer Beschreibung einer Skitour (Oteros-Rozas u. a., 2018). So gibt es eine starke Tendenz zur Besuchersicht und zur Freizeitbeschäftigung. Dabei fallen professionelle Beschäftigungen in der Land- oder Forstwirtschaft zurück, die ein grosses Wissen und eine andere Beziehung zur Landschaft repräsentieren (Garcia-Martin u. a., 2017).

Trotz der Vorteile der automatischen Textextraktion, wie die hohe Skalierbarkeit und eben doch auch die Möglichkeit einer breiten Abdeckung an Meinungen, sollte diese Methode eine Datenerhebung vor Ort und die Diskussion mit Expertinnen nicht ersetzen, sondern höchstens ergänzen.

## Kapitel 8

# Schlusswort

### 8.1 Fazit

Mit dieser Arbeit wurde gezeigt, dass es möglich ist mit öffentlichen Texten und relativ einfachen Methoden des Natural Language Processing und der Geographic Information Retrieval etwas über die Beschaffenheit der Landschaft in einem Gebiet herauszufinden. Zumindest grossräumige Trends in der Verteilung von bestimmten Landschaftsbegriffen können ausgemacht werden.

Die Schlüsselgebiete, die in der Landschaftskonzeption einen grossen Stellenwert haben, sind in den Korpora Bote der Urschweiz und Y-Magazin nicht besonders stark vertreten. Trotzdem wurde anhand der Word Clouds sichtbar, dass sie sich charakterisieren und voneinander unterscheiden lassen. Wenn etwas über kleine, bzw. selten genannte Orte herausgefunden werden soll, dann muss konkreter nach ihnen gesucht werden, sonst werden sie überschattet durch die häufigeren Toponyme.

Schlussendlich kann dieser Ansatz nicht die Einschätzung von Experten und die Landschaftscharakterisierung im direkten Austausch mit Betroffenen ersetzen. Aber er bietet eine Möglichkeit den Prozess zu demokratisieren und grosse Mengen an diversen und potenziell ungehörten Meinungen zu integrieren.

### 8.2 Weiterführende Arbeiten

Am naheliegendsten ist eine Folgearbeit mit einem breiter angelgten Korpus. Dabei sollten die Daten nicht nur quantitativ vergrössert werden, sondern auch qualitativ die Meinungsvielfalt verbreitert werden. Dazu könnte man sowohl gezielt bekannte Quellen nutzen, wie jene im Anhang A, oder es wird mit spezifischen Suchwörtern das schier unendliche Internet durchsucht (Koblet und Purves, 2020). Für die Suchwörter könnte man mit dem Netzwerk in Abbildung 6.14 arbeiten, um mehr über die nicht geomorphologischen Kategorien zu erfahren. Abgesehen von dieser passiven Art der Datenakquirierung würde die aufwändigere aktive Art via Partizipation von Interessensgruppen dabei helfen, die Aussagen besser zu interpretieren und eine Grundlage liefern, wonach weiter gesucht werden könnte (Wartmann u. a., 2015).

Eine Erweiterung der Arbeit was die Lage und Bewertung der LSB angeht, wäre ebenfalls spannend. Garcia-Martin u. a., 2017 haben beobachtet, dass besonders sichtbare Orte als ästhetischer bewertet werden. Allgemein tauchten LSB vor allem an gut zugänglichen Orten wie in der Nähe von Flüssen, Strassen und Wegen auf. Ausserdem gab es in dieser Studie LSB, die sich gegenseitig ausschlossen, z.B. Landwirtschaft und Naturschutz. Es wäre spannend zu sehen, ob solche Mechanismen auch in Schwyz auftreten.

Um die Bewertung der LSB besser zu verstehen, müssten Gefühle und Meinungen aus dem Text extrahiert werden. Um die LSB mit den richtigen Kontextwörtern zu verbinden, gibt es neben dem NLP-Programm von spacy noch weitere Sprachparser, die den Aufbau der deutschen Sprache noch besser verstehen, z.B. der Hannover Tagger (Wartena, 2019), der TreeTagger von Schmid, 1999 oder die Gewinner der 2018 Conference on Computational Natural Language Learning (Zeman u. a., 2018). Aber damit Qualität der LSB richtig verstanden wird, muss die Worteinbettung des Adjektivs oder des Gefühls speziell auf den Landschaftskontext trainiert werden (Chesnokova und Purves, 2018b).

Albers und James, 1988 haben auf Landschaftsfotografien bezogen geschrieben, dass selbst die umfangreichste und ausgeklügeltste Analyse nicht in der Lage ist, die symbolische Bedeutung, die die Ansicht einer Landschaft für jeden und jede einzelne bedeutet, einzufangen. Das stimmt bestimmt auch für die textuelle Beschreibung der Landschaft. Und trotzdem ist es spannend, mit der Analyse von Texten so nahe wie möglich an diese gefühlte, symbolische Bedeutung der Landschaft heranzukommen.

# Literatur

- Acheson, Elise, Michele Volpi und Ross S. Purves (2020). „Machine learning for cross-gazetteer matching of natural features“. In: *International Journal of Geographical Information Science* 34.4, S. 708–734. ISSN: 13623087. DOI: [10.1080/13658816.2019.1599123](https://doi.org/10.1080/13658816.2019.1599123).
- Acheson, Elise, Flurina M. Wartmann und Ross S. Purves (2018). „Generating spatial footprints from hiking blogs“. In: *Lecture Notes in Geoinformation and Cartography* 198809, S. 5–7. ISSN: 18632351. DOI: [10.1007/978-3-319-63946-8](https://doi.org/10.1007/978-3-319-63946-8).
- Albers, Patricia C und William R James (1988). „Travel photography: A methodological approach“. In: *Annals of tourism research* 15.1, S. 134–158.
- Appleton, Jay (1996). *The experience of landscape*. Wiley Chichester.
- Bote der Urschweiz* (Januar 2017 - April 2020). Bd. Jahrgang 158 (1) - 162 (75).
- Brezina, Vaclav (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Bundesinventar der Landschaften und Naturdenkmäler (BLN)* (2020). URL: <https://www.bafu.admin.ch/bafu/de/home/themen/landschaft/fachinformationen/landschaftsqualitaet-erhalten-und-entwickeln/landschaften-von-nationaler-bedeutung/bundesinventar-der-landschaften-und-naturdenkmaeler-von-national.html>.
- Burenhult, Niclas und Stephen C Levinson (2008). „Language and landscape: a cross-linguistic perspective“. In: *Language Sciences* 30.2-3, S. 135–150.
- Buscaldi, Davide (2011). „Approaches to disambiguating toponyms“. In: *Sigspatial Special* 3.2, S. 16–19.
- Campelo, Adriana u. a. (2014). „Sense of place: The importance for destination branding“. In: *Journal of travel research* 53.2, S. 154–166.
- Chernoff, Herman und E. L. Lehmann (1954). „The Use of Maximum Likelihood Estimates in  $\chi^2$  Tests for Goodness of Fit“. In: *The Annals of Mathematical Statistics* 25.3, S. 579–586. DOI: [10.1214/aoms/1177728726](https://doi.org/10.1214/aoms/1177728726). URL: <https://doi.org/10.1214/aoms/1177728726>.
- Chesnokova, Olga, Mario Nowak und Ross S. Purves (2017). „A crowdsourced model of landscape preference“. In: *Leibniz International Proceedings in Informatics, LIPIcs* 86.19, S. 1–13. ISSN: 18688969. DOI: [10.4230/LIPIcs.COSIT.2017.19](https://doi.org/10.4230/LIPIcs.COSIT.2017.19).

- Chesnokova, Olga und Ross S. Purves (2018a). „Automatically creating a spatially referenced corpus of landscape perception“. In: *Proceedings of the 12th Workshop on Geographic Information Retrieval, GIR 2018*, S. 3–4. DOI: [10.1145/3281354.3281356](https://doi.org/10.1145/3281354.3281356).
- (2018b). „From image descriptions to perceived sounds and sources in landscape: Analyzing aural experience through text“. In: *Applied Geography* 93, February, S. 103–111. ISSN: 01436228. DOI: [10.1016/j.apgeog.2018.02.014](https://doi.org/10.1016/j.apgeog.2018.02.014). URL: <https://doi.org/10.1016/j.apgeog.2018.02.014>.
- Clough, Paul D, Hideo Joho und Ross Purves (2006). „Judging the spatial relevance of documents for GIR“. In: *European Conference on Information Retrieval*. Springer, S. 548–552.
- Conrad, Cathy C. und Krista G. Hilchey (2011). „A review of citizen science and community-based environmental monitoring: issues and opportunities“. In: *Environmental Monitoring and Assessment* 176, S. 273–291. ISSN: 1. DOI: [10.1007/s10661-010-1582-5](https://doi.org/10.1007/s10661-010-1582-5).
- Cooper, David und Ian N Gregory (2011). „Mapping the English lake district: a literary GIS“. In: *Transactions of the Institute of British Geographers* 36.1, S. 89–108.
- Costonis, John J (1982). „Law and Aesthetics: A Critique and a Reformulation of the Dilemmas“. In: *Michigan Law Review* 80.3, S. 355–461.
- Criminisi, Antonio, Jamie Shotton, Ender Konukoglu u. a. (2012). „Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning“. In:
- Derungs, Curdin und Ross S. Purves (2014). „From text to landscape: Locating, identifying and mapping the use of landscape features in a Swiss Alpine corpus“. In: *International Journal of Geographical Information Science* 28.6, S. 1272–1293. ISSN: 13623087. DOI: [10.1080/13658816.2013.772184](https://doi.org/10.1080/13658816.2013.772184).
- Dinkelbach, Werner (2013). *Sensitivitätsanalysen und parametrische Programmierung*. Bd. 12. Springer-Verlag.
- Egorova, Ekaterina u. a. (2018). „Fictive motion extraction and classification“. In: *International Journal of Geographical Information Science* 32.11, S. 2247–2271.
- Engelberg, Stefan (2015). „Quantitative Verteilungen im Wortschatz. Zu lexikologischen und lexikografischen Aspekten eines dynamischen Lexikons“. In:
- European Landscape Convention (2000). *European Landscape Convention*. URL: [https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/176?\\_coconventions\\_WAR\\_coeconventionsportlet\\_languageId=de\\_DE](https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/176?_coconventions_WAR_coeconventionsportlet_languageId=de_DE).
- Fortuna, Blaz, Marko Grobelnik und Dunja Mladenic (2005). „Visualization of text document corpus“. In: *Informatica* 29.4.
- Fujii, Atsushi und Tetsuya Ishikawa (2000). „Utilizing the World Wide Web as an encyclopedia: Extracting term descriptions from semi-structured texts“. In: *arXiv preprint cs/0011001*.

- Gajdoš, Petr u. a. (2016). „A parallel Fruchterman–Reingold algorithm optimized for fast visualization of large graphs and swarms of data“. In: *Swarm and evolutionary computation* 26, S. 56–63.
- Garcia-Martin, Maria u. a. (2017). „Participatory mapping of landscape values in a Pan-European perspective“. In: *Landscape Ecology* 32.11, S. 2133–2150. ISSN: 15729761. DOI: [10.1007/s10980-017-0531-x](https://doi.org/10.1007/s10980-017-0531-x).
- Gonzalez-Pardo, Antonio und David Camacho (2011). „Analysis of grammatical evolutionary approaches to regular expression induction“. In: *2011 IEEE Congress of Evolutionary Computation (CEC)*. IEEE, S. 639–646.
- Goodchild, Michael F und Linda L Hill (2008). „Introduction to digital gazetteer research“. In: *International Journal of Geographical Information Science* 22.10, S. 1039–1044.
- Gritta, Milan u. a. (2018). „What’s missing in geographical parsing?“ In: *Language Resources and Evaluation* 52.2, S. 603–623.
- Hermes, Johannes u. a. (2018). „Assessment and valuation of recreational ecosystem services of landscapes“. In: *Ecosystem Services* 31, S. 289–295. ISSN: 22120416. DOI: [10.1016/j.ecoser.2018.04.011](https://doi.org/10.1016/j.ecoser.2018.04.011). URL: <https://doi.org/10.1016/j.ecoser.2018.04.011>.
- Hill, Shawndra und Noah Ready-Campbell (2011). „Expert stock picker: the wisdom of (experts in) crowds“. In: *International Journal of Electronic Commerce* 15.3, S. 73–102.
- Hirschberg, Julia und Christopher D Manning (2015). „Advances in natural language processing“. In: *Science* 349.6245, S. 261–266.
- Honnibal, Matthew u. a. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. DOI: [10.5281/zenodo.1212303](https://doi.org/10.5281/zenodo.1212303). URL: <https://doi.org/10.5281/zenodo.1212303>.
- Hu, Yingjie (2018). „Geo-text data and data-driven geospatial semantics“. In: *Geography Compass* 12.11, S. 1–19. ISSN: 17498198. DOI: [10.1111/gec3.12404](https://doi.org/10.1111/gec3.12404).
- Huldi, Ramón (2015). „Extrahierung von Landschaftswahrnehmungen aus Tourenberichten Vorstellung einer neuen Analyseverfahren basierend auf“. In: September.
- Hunn, Eugene (1994). „Place-names, population density, and the magic number 500“. In: *Current Anthropology* 35.1, S. 81–85.
- Hunziker, Marcel, Matthias Buchecker und Terry Hartig (2007). „Space and place—two aspects of the human-landscape relationship“. In: *A changing world*. Springer, S. 47–62.
- Jones, Christopher B. und Ross S. Purves (2008). „Geographical information retrieval“. In: *International Journal of Geographical Information Science* 22.3, S. 219–228. DOI: [10.1080/13658810701626343](https://doi.org/10.1080/13658810701626343).
- Jones, Michael und Marie Stenseke (2011). *The European landscape convention: challenges of participation*. Bd. 13. Springer Science & Business Media.

- Joulin, Armand u. a. (2016). „Bag of tricks for efficient text classification“. In: *arXiv preprint arXiv:1607.01759*.
- Koblet, Olga und Ross S. Purves (2020). „From online texts to Landscape Character Assessment: Collecting and analysing first-person landscape perception computationally“. In: *Landscape and Urban Planning* 197. February, S. 103757. ISSN: 01692046. DOI: [10.1016/j.landurbplan.2020.103757](https://doi.org/10.1016/j.landurbplan.2020.103757). URL: <https://doi.org/10.1016/j.landurbplan.2020.103757>.
- Landestopografie swisstopo, Bundesamt für (2020). *swissNAMES3D*. URL: <https://shop.swisstopo.admin.ch/de/products/landscape/names3D>.
- Landschaftskonzeption Kanton Schwyz. Version vom Februar* (2019). Techn. Ber.
- Lingiardi, Vittorio (2020). *Mindscapes: psiche nel paesaggio*. Raffaello Cortina Editore.
- Manning, Christopher und Hinrich Schütze (1999). *Foundations of statistical natural language processing*. MIT press.
- Mark, David M u. a. (2011). „Landscape in language: An introduction“. In: *Medienmitteilung. Freizeitaktivitäten erhöhen den Druck auf die Natur* (2020). URL: <https://www.sz.ch/behoerden/information-medien/medienmitteilungen/medienmitteilungen.html/72-416-412-1379-1377-4603/news/13852>.
- Meng, Xiao-Li u. a. (2018). „Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election“. In: *Annals of Applied Statistics* 12.2, S. 685–726.
- Nadkarni, Prakash M, Lucila Ohno-Machado und Wendy W Chapman (2011). „Natural language processing: an introduction“. In: *Journal of the American Medical Informatics Association* 18.5, S. 544–551.
- Nothman, Joel u. a. (2013). „Learning multilingual named entity recognition from Wikipedia“. In: *Artificial Intelligence* 194, S. 151–175.
- Ochoa, Xavier und Erik Duval (2008). „Quantitative analysis of user-generated content on the web“. In: *Proceedings First International Workshop Understanding Web Evolution*, S. 19–26.
- Onan, Aytuğ, Serdar Korukoğlu und Hasan Bulut (2016). „Ensemble of keyword extraction methods and classifiers in text classification“. In: *Expert Systems with Applications* 57, S. 232–247. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2016.03.045>. URL: <https://www.sciencedirect.com/science/article/pii/S0957417416301464>.
- Oteros-Rozas, Elisa u. a. (2018). „Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites“. In: *Ecological Indicators* 94, S. 74–86. ISSN: 1470160X. DOI: [10.1016/j.ecolind.2017.02.009](https://doi.org/10.1016/j.ecolind.2017.02.009). URL: <http://dx.doi.org/10.1016/j.ecolind.2017.02.009>.
- Pedregosa, Fabian u. a. (2011). „Scikit-learn: Machine learning in Python“. In: *Journal of machine learning research* 12. Oct, S. 2825–2830.

- Plisson, Joël, Nada Lavrac, Dunja Mladenic u. a. (2004). „A rule based approach to word lemmatization“. In: *Proceedings of IS*. Bd. 3, S. 83–86.
- Produktinformation swissNAMES3D* (2020). Techn. Ber.
- Purves, Ross S u. a. (2018). „Geographic information retrieval: Progress and challenges in spatial search of text“. In: *Foundations and Trends in Information Retrieval* 12.2-3, S. 164–318.
- Rees, Anke (2016). „Das Gebäude als Akteur“. In: *Architekturen und ihre Atmosphären*.
- Richtplan des Kanton Schwyz. Richtplantext* (2020). Techn. Ber.
- Rodewald, Raimund, Yves Schwyzer und Karina Liechti (2014). „Katalog der charakteristischen Kulturlandschaften der Schweiz. Grundlage zur Ermittlung von Landschaftsentwicklungszielen“. In: *Stiftung Landschaftsschutz Schweiz slfp*.
- Rodewald, Raimund u. a. (2020). *Landscape Aesthetics: Theory and Practice of the Sensuous Cognition of Landscape Qualities – Lecture Script*. PLUS – ETH Zurich.
- Schmid, Helmut (1999). „Improvements in part-of-speech tagging with an application to German“. In: *Natural language processing using very large corpora*. Springer, S. 13–25.
- Schütze, Hinrich, Christopher D Manning und Prabhakar Raghavan (2008). *Introduction to information retrieval*. Bd. 39. Cambridge University Press Cambridge.
- Schwick, Christian (2010). *Zersiedelung der Schweiz-unaufhaltsam?: quantitative Analyse 1935 bis 2002 und Folgerungen für die Raumplanung*. Bd. 26. Haupt Verlag AG.
- Seel, Martin (1991). *Eine Aesthetik der Natur*. Suhrkamp.
- Smith, Barry und David M Mark (2003). „Do mountains exist? Towards an ontology of landforms“. In: *Environment and Planning B: Planning and Design* 30.3, S. 411–427.
- Stephenson, Janet (2008). „The Cultural Values Model: An integrated approach to values in landscapes“. In: *Landscape and Urban Planning* 84.2, S. 127 – 139. ISSN: 0169-2046. DOI: <https://doi.org/10.1016/j.landurbplan.2007.07.003>. URL: <http://www.sciencedirect.com/science/article/pii/S0169204607001661>.
- TIGER-Project (2003). „TIGER Annotationsschema. Manuscript“. In: *Universität des Saarlands, Universität Stuttgart Universität Potsdam*.
- Tudor, Christine (2014). „An Approach to Landscape Character Assessment“. In: *Natural England* October, S. 56. DOI: NE579. URL: <http://www.programmeofficers.co.uk/Cuadrilla/CoreDocuments/CD40/CD40.20.PDF>.
- Tversky, Barbara und Kathleen Hemenway (1983). „Categories of environmental scenes“. In: *Cognitive psychology* 15.1, S. 121–149.
- Van, Tien Nguyen and Mauro Gaio, and Ludovic Moncla (2013). „Topographic subtyping of place named entities: a linguistic approach“. In: *AGILE 2013*, S. 1–5.

- Wartena, Christian (2019). „A Probabilistic Morphology Model for German Lemmatization“. In: *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)*, S. 40–49.
- Wartmann, Flurina, Curdin Derungs und Ross Purves (2016). „Characterizing place: an empirical comparison between user-generated content and freelisting data“. In: *International Conference on GIScience Short Paper Proceedings 1*, S. 336–339. DOI: [10.21433/b3111bw2c801](https://doi.org/10.21433/b3111bw2c801).
- Wartmann, Flurina M., Elise Acheson und Ross S. Purves (2018). „Describing and comparing landscapes using tags, texts, and free lists: an interdisciplinary approach“. In: *International Journal of Geographical Information Science* 32.8, S. 1572–1592. ISSN: 13623087. DOI: [10.1080/13658816.2018.1445257](https://doi.org/10.1080/13658816.2018.1445257).
- Wartmann, Flurina M. und Ross S. Purves (2018). „Investigating sense of place as a cultural ecosystem service in different landscapes through the lens of language“. In: *Landscape and Urban Planning* 175, S. 169–183. ISSN: 01692046. DOI: [10.1016/j.landurbplan.2018.03.021](https://doi.org/10.1016/j.landurbplan.2018.03.021).
- Wartmann, Flurina M. u. a. (2015). „More than a list: What outdoor free listings of landscape categories reveal about commonsense geographic concepts and memory search strategies“. In: *Lecture Notes in Computer Science (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9368, S. 224–243. ISSN: 16113349. DOI: [10.1007/978-3-319-23374-1\\_{\\\_}11](https://doi.org/10.1007/978-3-319-23374-1_{\_}11).
- Weibel, Viktor und Albert Hug (2012). *Schwyzer Namenbuch. Die Orts- und Flurnamen des Kantons Schwyz*. Bd. 6 Bände.
- Y-Magazin* (Sommer 2012 - Winter 2020). Bd. 1 - 35. Amt für Wirtschaft Kanton Schwyz.
- Y-Magazin Nr. 1* (Sommer 2012). Bd. 1. Amt für Wirtschaft Kanton Schwyz.
- Yang, Byoung-E und Terry J Brown (1992). „A cross-cultural comparison of preferences for landscape styles and landscape elements“. In: *Environment and behavior* 24.4, S. 471–507.
- Zeman, Daniel u. a. (2018). „CoNLL 2018 shared task: Multilingual parsing from raw text to universal dependencies“. In: *Proceedings of the CoNLL 2018 Shared Task: Multilingual parsing from raw text to universal dependencies*, S. 1–21.
- Zube, Ervin H, James L Sell und Jonathan G Taylor (1982). „Landscape perception: research, application and theory“. In: *Landscape planning* 9.1, S. 1–33.
- Öffentliche Mitwirkung Landschaftskonzeption Kanton Schwyz*. (2019). URL: <https://www.sz.ch/behoerden/umwelt-natur-landschaft/natur-und-landschaft/landschaftsschutzgebiete.html/72-416-397-396-4020>.

## Anhang A

# Vorgeschlagene Texte für den Korpus

Weitere mögliche Korpora vorgeschlagen von SchwyzerInnen und dem Geschäftsführer der Stiftung Landschaftsschutz Schweiz:

- von Dr. Karin Schwitter: ev. Zeitung ergänzen mit March-Anzeiger und Höfner-Volksblatt (Abo pflichtig):
  - Ein weiterer guter Fundus sind unsere Lokalzeitungen. Da hat jede Region ihre eigene Zeitung mit Online-Archiv, zu dem sie dir für Forschung auf Anfrage sicher Zugang geben: March-Anzeiger, Höfner Volksblatt, Einsiedler Anzeiger, Bote der Urschweiz (innere Kantonshälfte), Rigi Post (Arth-Goldau), Freier Schweizer (Küssnacht)
  - Die Gemeinde Wollerau hat zu ihrem 800-Jahr Jubiläum übrigens kürzlich ein ganzes Buch herausgegeben, wo's was drin haben könnte: Wollerau 1217-2017 : Geschichte(n) eines Dorfs, von Manuela Gili Sidler, Claudia Hiestand, Manuel Hiestand, Andreas Meyerhans, und anderen.
  - Dann haben wir eine (derzeit leider geschlossene) Kantonsbibliothek, die praktisch alle Schriften hat, die zu unserem Kanton erscheinen: <https://www.sz.ch/staatskanzlei-departemente/bildungsdepartement/amt-fuer-kultur/kantonsbibliothek/kantonsbibliothek.html/72-416-387-380-2480-2501-2498>
- von Remo Bianchi (Verantwortlicher für Natur und Landschaft Kanton Schwyz):
  - Schwyzerische Naturforschende Gesellschaft (diese hat diverse Berichte zu verschiedenen Natur- und Landschaftsthemen herausgegeben und u.a. auch das Inventar der Geotope erstellt, <http://www.szng.ch/>)
  - Landschaftsschutzverband Vierwaldstättersee ([www.lsvv.ch](http://www.lsvv.ch), setzt sich für die Interessen der Landschaft rund um den Vierwaldstättersee ein)
  - Schwyzer Umweltrat (Zusammenschluss vieler Schutzorganisationen, deren Mitglieder vielfach Landschaftskenner sind)

- SAC- Sektionen Mythen, Zindelspitz und Einsiedeln (haben insbesondere einen Blick auf die alpinen Landschaften im Kanton und kennen diese gut)
  - Objektbeschreibungen für BLN-Gebiete und Moorlandschaften (Der Kanton Schwyz ist stark von diesen Inventaren betroffen, siehe Webseite BA-FU)
  - Landwirtschaftliche Vernetzungsprojekte und Landschaftsqualitätsprojekte gemäss Literaturverzeichnis in der Landschaftskonzeption (die Träger-schaften sind i.d.R. Bauernvereinigungen, welche sich vertieft mit ihrer Landschaft auseinandergesetzt haben)
  - Betroffene Gemeinden - über die Gemeinden werden sie evtl. auf lokale Landschaftskenner aufmerksam
  - Kommunale Inventare der schützenswerten Biotope und Landschaften (liegt für jede Gemeinde vor, meistens nur in Papierform, könnte bei uns eingesehen werden)
  - Schwyz Tourismus (Manuela Gilli, haben das Thema SSwyzer Land-schaften aufgegriffen und vermarkten dieses seit kurzem, es ist auch ein Faltblatt dazu in Erarbeitung - Beilage)
  - Falls Sie zudem Einzelpersonen suchen, welche die Schlüsselgebiete be-sonders gut kennen, könnten wir allenfalls einige Namen nennen (vgl. beiliegendes Beispiel eines Riemenstaldner Bauern).
- von Valentin Kessler (Staatsarchivar Kanton Schwyz):
    - Schwyz. Portrait eines Kantons, hg. von Niklaus Flüeler, Schwyz 1991, (insbesondere Kapitel "Land und Landschaften").
    - Eine Reise durch den Kanton Schwyz, hg. vom Regierungsrat des Kan-ton Schwyz, 2005.
    - 3 Orte, 2 Seen, 1 Berg. Bezirk Küssnacht, hg. vom Bezirk Küssnacht, Immensee 2006 (?).
    - Ännet em Jordan. Wie sich die Kantone St. Gallen, Schwyz und Glarus beid-seits der Linth begegnen, Kaltbrunn 2013.
    - Wasserschwyz. Porträt der Wasserwelten im Kanton Schwyz, von ihren Menschen und von ihrer Nutzung, Alex B. Bott, Wollerau 2015.
    - Schwyzer Erinnerungsorte, Schwyzer Hefte Band 100, hg. von der Kul-turkommission des Kantons Schwyz, Lachen 2013.
    - ROTHENTHURM – mehr als „Turpnä“ und „Ischä“, Albert Marty-Gisler, Schwyz 2016.
    - Muotatal. Geschichten und Geschichte. Ein weiteres Stück Heimatkunde, Walter Imhof, Schwyz 2016.

- Der Sihlsee, eine Landschaft ändert ihr Gesicht, hg. von Karl Saurer, Zürich 2012.
- Die Schlacht am Morgarten aus geo-wissenschaftlicher Sicht. Rekonstruktion des Landschaftsbildes mit Hilfe eines multidisziplinären Ansatzes..., Markus Egli/Max Maisch, in: Der Geschichtsfreund. Mitteilungen des Historischen Vereins Zentralschweiz, Bd. 168, 2015, S. 117 ff.
- Schwyzer Namenbuch. Die Orts- und Flurnamen des Kantons Schwyz, hg. vom Kuratorium Orts- und Flurnamenbuch des Kantons Schwyz, 5 Lexikonbände und ein Registerband, Schwyz 2012.
- Vom Dräckloch i Himel. Namenbuch des Kantons Schwyz, Viktor Weibel, Schwyz 2012.
- Institutionen: Arbeitsgemeinschaft Höllochforschung AGH, Tourismusorganisationen

## Anhang B

# Liste der Landschaftsbegriffe

## B.1 LSB-Liste

### B.1.1 Definitive Begriffe

Abendglühen	Einsattelung	Gondelbahn	Magerheuwiese
Abendgrauen	Einzugsgebiet	Graben	Mähen
Abendsonne	Eisfläche	Gras	Massiv
Abhang	Eiswand	Grasflanke	Matten
Allmeind	Ernte	Grashalde	Meer
Allmend	Extensivierung	Grashang	Mittelgipfel
Alp	Extensivwiese	Grasrampe	Mittelstation
Alpen	Fahrsträsschen	Grat	Monte
Alpenrose	Feld	Gratkante	Moor
Alpenüberquerung	Feldlager	Gratrücken	Moorwald
Alphütte	Fels	Gratweg	Moräne
Äplerfest	Felsaufschwung	Halde	Morgengrauen
Alpweg	Felsband	Hang	Morgensonne
Alpweide	Felsbastion	Hänge	Mulde
Anhöhe	Felsblock	Hängebrücke	Mündung
Artenvielfalt	Felsbrocken	Hauptgipfel	Murgang
Aue	Felsgrat	Heu	Natur
Auerhahn	Felskopf	Hirte	Naturpyramide
Ausblick	Felsplatte	Hochebene	Naturschutzgebiet
Aussicht	Felsriegel	Hochgebirge	Naturwunder
Aussichtspunkt	Felsstufe	Hochmoor	Pass
Aussichtspunkt	Felsturm	Höcker	Passhöhe
Bach	Felswand	Höhenweg	Passo
Bachbett	Fettwiese	Höhle	Pfad
Bachlauf	Feuerstelle	Hölloch	Pfadspur
Bächlein	Firn	Holzbrücke	Pilger
Bächli	Firnfeld	Hügel	Pilgern
Bauernhaus	Firngrat	Hüttenweg	Pilzsammeln
Bauernhäus	Fischen	Insel	Piste
Baum	Fischer	Jagd	Piz
Bäume	Flanke	Jagen	Plateau
Bauzonen	Flue	Jagdbann	Quelle
Berg	Fluss	Jagdbanngebiet	Renaturieren
Bergbahn	Föhn	Jäger	Renaturierung
Berggasthaus	Forstweg	Jakobspilger	Route
Berggrat	Gebirge	Jakobsweg	Ruine
Berghang	Gebirgswelt	Kalkstein	Scharte
Berghaus	Gelände	Karrenfeld	Blätter
Bergrestaurant	Geröll	Karst	Schilf
Bergrücken	Geröllfeld	Klettersteig	Schleusen
Bergschrund	Geröllhalde	Klippe	Schlitteln
Bergstation	Gestein	Kraxelei	Schlitten
Bergwanderweg	Gewässer	Krete	Schlittenhunde
Bergweg	Gipfel	Kuppe	Schlucht
Biken	Gipfelbereich	Küste	Schnee
Biker	Gipfelchen	Lac	Schneedecke
Blockgrat	Gipfelfels	Lago	Schneefeld
Blüemberg	Gipfelgrat	Landschaft	Schneeegrat
Brücke	Gipfelhang	Landwirtschaft	Schneien
Burg	Gipfelkopf	Laub	Schotter
Couloir	Gipfelkreuz	Lawine	Schrund
Delta	Gipfelplateau	Lawinenkegel	Schutt
Dunst	Gletscher	Lawinenverbauung	Schutthalde
Ebene	Gletscherzunge	Luftseilbahn	See

Seedamm	Äplerfests	Waldreichtum	Skigegebiete
Seelein	Wildzaun	Waldreservat	Skigebiet
Seilbahn	Wüsten	Waldreservatskonzept	Skigebiets
Seilbahnstation	Wüstengebieten	Waldlichtung	Skihütte
Senke	Wildtierkorridor	Waldmann	Skiinfrastrukturanlagen
Sesselbahn	Wildtierkorridors	Waldmeisterart	Skilifte
Sessellift	Wildwarnanlage	Waldhängen	Skiliften
Siedlung	Wildwarner	Waldhüsli	Skiliftes
Ski	Wildwechsel	Waldgebiete	Skipisten
Skifahren	Wildtierarten	Waldgottesdienst	Skiplausch
Skifahrer	Wildtiere	Waldhaus	Skiregion
Skigebiet	Wildtieren	Waldfläche	Skis
Skilift	Wildtieres	Waldfriedhof	Skisport
Skipiste	Wildsträucher	Uferabschnitte	Skisprungschanze
Skitour	Wildsüü	Uferbereich	Skisprungschanzen
Sommerweg	Wildkräuter	Ufergebiet	Skistationen
Sonne	Wildlife	Ufermauern	Skitouren
Spaltenzone	Wildlilienart	Ufern	Skitourenfahrer
Spitzkehren	Wildnis	Uferpromenade	Skitourengänge
Standseilbahn	Wildnisgebiet	Uferrand	Skitourenrennen
Station	Wildnisgebiete	Ufers	Skitourenwoche
Staumauer	Wildpflanzen	Ufersanierung	Skiweltmeisterschaft
Stausee	Wildregion	Urwaldentwicklung	Skiwettrennen
Steg	Wildregionen	Urwaldes	Skiwochenendes
Steig	Wildruhezonen	Urwaldholz	Skiabfahrten
Steilhang	Wildblumen	Urwaldreservat	Skiausflugs
Steinbruch	Wildblumenpflanzen	Verbauungen	Skier
Steinmännchen	Wilde	Verbauungsarbeiten	Skierlebnis
Tal	Wilden	VierSeen	Siedlungsgebiet
Talboden	Wilder	Viertausendermarke	Siedlungsraum
Tälchen	Wilderei	Waldbesitzer	Siedlungen
Talgrund	Wilderern	Waldbesitzerin	Seilbahnstützen
Talkessel	Wildern	Waldbewachsene	Sesselbahnanlage
Talseite	Wilderness	Waldbewirtschaftung	Sesselbahnen
Talstation	Wildes	Waldboden	Sees
Teich	Wildhaus	Waldbrand	Seeschiff
Terrain	Wildheuen	Waldbrandgefahr	Seesicht
Teufelsbrücke	Wildheuer	Waldbrände	Seestrände
Thal	Wanderer	Waldeigentümer	Seeufer
Tier	Wanderern	Waldes	Seeuferbereich
Tobel	Wandernden	Tobelbachbrücke	Seeufermauer
Trampelpfad	Wanderungen	Tierpark	Seeweg
Traverse	Wanderungs	Tierparks	Seezugang
Überhang	Wanderwege	Tierreichs	Seezugangs
Ufer	Wanderwegen	Tierwelt	Seilbahnen
Urwald	Wanderwegmitarbeitende	Talstrasse	Seen
Val	Wanderwegnetz	Talweg	Seepromenade
Verbauung	Wanderwegs	Teiche	Seeblick
Verbindungsgrat	Wegen	Terrainanpassung	Seebücke
Viertausender	Wegweisern	Terrainverlaufs	Seefront
Vorgipfel	Weideflächen	Terrainveränderung	Seeli
Wald	Weidegebiet	Tierchen	Seenebel
Wäldchen	Weidel	Tiere	Seebad
Waldgrenze	Weideli	Tieren	Seebades
Waldrand	Weiden	Talflanken	Seebads
Waldstück	Weideunterstand	Talgegend	Seebecken
Waldweg	Wiesen	Talkessels	Seeanlage
Wanderer	Wiesenbord	Steinbruchareals	Schuttfluren
Wandern	Wildbestand	Steinbruchs	Schuttkegel
Wanderung	Wildbestandes	SuworowBrücke	Schuttwalds
Wanderweg	Wildbestände	Talbecken	Schneewände
Wandfluss	Wildbienen	Talbodens	Schnee?ächen
Wasseraue	Waldspielgruppe	Taldünsten	Schotterweg
Wasserfall	Waldspielplatz	Tale	Schneesport
Wasserscheide	Waldsterben	Tales	Schneesportler
Wegmarkierung	Waldstrasse	Stegweidli	Schneesportlerin
Wegspur	Waldstriche	Steige	Schneesporttag
Wegweiser	Waldstätte	Steigeisen	Schneeschuhtour
Weide	Waldstücke	Stegs	Schneeschuhtouren
Weiler	Waldungen	Staumauerstrasse	Schneeschuhwanderern
Weite	Waldvogel	Stege	Schneeschuhwandern
Wiese	Waldwege	Stegen	Schneeschuhwanderung
Wild	Waldwegen	Sonnenuntergang	Schneeschuhläufer
Wildiheuen	Waldweiher	Sonnenuntergängen	Schneeschuhläuferin
Wildiheuer	Waldniederlegungen	Staumauern	Schnees
Winterwanderweg	Waldnutzung	Sonnenaufgang	Schneepass
Wüste	Waldpark	Sonnenaufgängen	Schneeplaushtag
Zuckerhut	Waldpartien	Skies	Schneehöhen
Hölloch	Waldparzelle	Skifahrerin	Schneehöhlen
zAlp	Waldpflege	Skifahrerinnen	Schneehalde
Äplerfestes	Waldregionen	Skifahrern	Schneeberge

Ruinen	Jagdhaus	Bergwiese	Baumstamm
Schilffelder	Jagdhütte	Bergwiesen	Baumstrunk
Schilfgürtel	Jakobsweegen	Bikergruppe	Baumstrünke
Schlittenfahrten	Jägerinnen	Bikern	Baumstrünken
Renaturierungen	Jägern	Bikerreise	Baumstumpf
Renaturierungsprojekt	Jägersteg	Bikers	Baumwipfel
Pilgerweg	Karstgebiet	Brücken	Baumwipfelpfad
Pilgerwege	Karstregion	Brückenpfeiler	Baumgrenze
Pisten	Höhlen	Brückenquerung	Aussichtsbänkli
Pistenfahrzeug	Höhleneingang	Brückenrundweg	Aussichtshütte
Pistenrand	Höhlenfluss	Brückenverbreiterung	Aussichtskanzel
Pistenrestaurant	Hirten	Burgen	Aussichtsplattform
Pilgerin	Hochmoore	Bäumen	Aussichtsplattformen
Pilgerort	Hochmoorebene	Bergtouren	Aussichtsturm
Passes	Hochmoorstall	Bergwanderer	Bachdelta
Passstrasse	Heuet	Bergwanderin	Bacheinläufe
Pfaden	Heuwagen	Bergwanderung	Bacheinschwemmungen
Naturweihern	Hanglage	Bergweges	Bachholz
Naturstrasse	Hauptgipfels	Bergwegs	Bachmündung
Naturschutzgebiete	Heubeeripflücken	Bergwelt	Auen
Naturschutzgebieten	Heublumen	Bergsturzgelände	Auerhuhn
Naturschönheit	Heuen	Bergsturzmasse	Alpwirtschaftsbetriebe
Naturschönheiten	Grate	Bergsturmateriale	Alpweiden
Naturlandschaft	Grates	Bergsee	Alpwirt
Naturparadies	Gratwanderung	Bergseen	Alpwirtschaft
Naturreservat	Halden	Bergsicht	Alpstall
Naturerlebnis	Gipfelhöhe	Bergsonne	Alpstrasse
Naturerlebnissen	Gipfeln	Bergspitzen	Alpsiedlung
Natureldorado	Gipfelrestaurant	Bergsteigen	Alps
Naturdenkmäler	Gipfels	Bergsteiger	Alpsees
Moorwälder	Gipfeltour	Bergsteigerin	Alpkäserei
Mulden	Gletschergarten	Bergsteigern	Alpkräuter
Muldenlagen	Gletschers	Bergsteigers	Allmeindli
Murgangs	Gletscherspur	Berggänger	Allmeindwald
Mündungsbereich	Gletschervorfeld	Berghotel	Rebe
Moorsees	Gondelbahnen	Berghütte	Reben
Moorwaldlandschaft	Gewässern	Bergkamm	Rebberg
Moorweg	Geröllsammler	Bergkante	Rebberge
Moorwäld	Gebirgen	Bergkapelle	Rebbergen
Moore	Geländekammer	Bergkirche	Streusiedlung
Moorfläche	Geländemulde	Bergkulisse	Streusiedlungen
Moorflächen	Föhnstimmung	Berglagen	Deponie
Moorgebiete	Flussbett	Berglandschaft	Deponien
Moorgebieten	Flusses	Berglandwirtschaftsbetrieb	Obstwiese
Moorlandschaft	Flusslandschaft	Berglauf	Obstwiesen
Moorlandschaften	Flusslauf	Berglaufmeisterschaft	Obstbaum
Lawinenschutzmauer	Flussläufe	Bergli	Obstbäume
Lawinenverbauten	Flusssystem	Bergluft	Obstbäumen
Luftseilbahnen	Flussufer	Bergläufen	Hochstammbaum
Lawinenhang	Felsvorsprung	Bergläufer	Hochstamm bäume
Landwirtschaftsbetrieb	Felswanderweg	Bergläuferin	Hochstämme
Landwirtschaftsbetriebe	Felswänden	Bergmassiv	Abbaustandort
Landwirtschaftsfläche	Fettwiesen	Bergorte	Wildtieranlage,
Landwirtschaftsflächen	Fischerboot	Bergpanora	Waldweihnacht,
Landwirtschaftsgebäude	Fischerhaus	Bergpanorama	Waldföhre
Landwirtschaftsland	Fischern	Bergregion	Talfahrt,
Landwirtschaftszone	Felssicherungen	Bergregionen	Stauseeprojekt
Laubbäumen	Felsgewölbe	Bergrennen	Sonnenrain
Landschaft	Felsklotz	Bergrückens	Sonnenberg
Landschaftspanoramen	Felskluft	Bergschwingens	Sonnenhängen
Landschaftstypen	Felsmassen	Bergschwinget	Seerestaurant
Landschaftswerten	Felsnasen	Berggasthauses	Seerundwegs
Landschaften	Felspfad	Berggebiet	Schutteingang,
Klippen	Felsberg	Berggemeinde	Seeablagerungen,
Klippenspringen	Felsblöcke	Berggemeinden	Seeanstoss,
Klippenweg	Felsblöcken	Berggipfel	Seebereich,
Kreten	Felsen	Berggipfeln	Seebistro,
Küsten	Felsenkapelle	Bergflanke	Naturräume
Küstenberge	Felsenkirche	Bergdorf	Moorschutz
Küstengebiete	Felsenweg	Berge	Landschaftsbild
Küstengebieten	Felsflühen	Bergen	Naturbelag,
Küstenstadt	Felder	Berges	Lawinenabgang,
Höhlengruppe	Feldern	Bergbahnen	Lawinenabgängen,
Höhlensystem	Feldes	Bergbahnstation	Lawinengefahr,
Höllochs	Feldgehölz	Bergbauer	Lawinnenedergang,
Hügellandschaften	Feldstrassen	Bergbauern	Luftseilbahnfahrt
Inselgruppe	Feldweg	Bergbeizen	Naturgefahren,
Inseli	Feldwegen	Bergblick	Naturgewalten,
Inseliquai	Feldbach	Baumkronen	Naturereignisse,
Inseln	Bergwelten	Baums	Naturereignissen
Jagdbanngebieten	Bergwetter	Baumschulen	Gewässerräume

Gewässerraum	Rebberge	Moorbiotop	Bannkreuz
Felsstürzen	Biodiversität	Moorbiotope	Wegkreuz
Erntedankfest	Alplandschaft	Biotop	Passroute
Erntedankgottesdienst	Waldgrenze	Biotope	Passrouten
Ernten	Waldgrenzen	Passlage	Biotops
Felsabsturz	Weidelandschaft	Passlagen	Biotores
Felsrutsch	Weidelandschaften	Hochebene	Talkessels
Felssturz	Alplandschaften	Hochebenen	Waldpartien
Felssturzes	Flachmoor	Moorwald	Waldpartie
Felssturzgebiet	Flachmoore	Moorwälder	Felspartien
Berggottesdienst	Flachmooren	Produktionslandschaft	Felspartie
Bergsturz	Trockenwiesen	Produktionslandschaften	Jakobsweges
Bergsturzes	Alphütten	Wildnisinsel	Flysch
Bergsturzgebiet	Ställe	Wildnisinseln	Klippen
Bergstürzen	Stall	Seenlandschaft	Naturerlebniswert
Bergchilbi	Käselager	Stillegewässer	Einzelhof
Bergfest	Hochgebirgsrelief	Fliessgewässer	Einzelhöfe
Bergfeste	Bestossung	Südhang	Pflanzenschutzgebiet
Bauzonengrenze	Alpweiden	Südhänge	Pflanzenschutzgebietes
Bauzonrand	Alpweide	Stauwehr	Flysch-Landschaft
Baumfällaktionen	Weideflächen	Stauwehre	Kapelle St. Johann
Ausblicke	Weidefläche	Landschaftsansicht	Passlandschaft
Ausblicken	Vergandung	Landschaftsansichten	Kapelle St. Meinrad
Alphorntönen	Wildheulandschaft	Hofgruppe	Bacheinschnitt
Alpstaeg	Wildheulandschaften	Hofgruppen	Bacheinschnitten
Alphornklängen	Sömmerungsgebiet	Graswirtschaft	Bacheinschnitten
Alpgottesdienst	Sömmerungsgebiete	Obstbäumen	Lebensraumvielfalt
Alpgebäuden	Mähnutzung	Strukturvielfalt	Wanderggebiet
Alpgebäude	Hanglagen	Obstwiese	Wanderggebiete
Alpentöne	Fusswege	Obstwiesen	Bewirtschaftungsform
Alpenweg	Verbrachung	Obstgärten	Bewirtschaftungsformen
Alpenpanorama	Fussweg	Bewirtschaftung	Bachgraben
Alpenpässe	Heuseil	Industriestandort	Bachgräben
Alpenklänge	Heuseile	Industriestandorte	Wildheuplängen
Alpbetriebe	Mosaiklandschaft	Waldeinheit	Besiedlungsmuster
Alpbetriebe	Mosaiklandschaften	Gebirgselement	Frauenkloster St. Josef
Alpbetriebe	Hügelgebiet	Gebirgselemente	Pfarrkirche St. Sigismund und Wal-
Alpchäsmärcht	Topographie	Bauernhof	burga
Alpchäsmärchts	Offenland	Bauernhöfe	Polje
Alpdankgottesdienst	Grünland	Bauernhöfen	Verebnungsfläche
Alpabfahrt	Allee	Bachlauf	Gebirgszug
Alpabfahrten	Alleen	Bachläufe	Gebirgszüge
Alpabzug	Trockensteinmauer	Hangtopographie	Karstlandschaft
Alpauhfahrt	Trockensteinmauern	Waldstreifen	Karst
Alpauhfahrten	Hecke	Landsitz	Karstlandschaften
Kulturlandschaft	Hecken	Landsitze	Karrfeld
Hochgebirge	Flur	Herrenhaus	Karrfelder
Felsenlandschaft	Bachuferbestockung	Herrenhäuser	Alpbetrieb
Felsenlandschaften	Bachuferbestockungen	Ortsbild	Alpbetriebe
Felsformation	Bachufer	Ortsbilder	Erholungspotenzial
Felsformationen	Landschaftskammer	Siedlungslandschaft	Erholungswert
Gestein	Landschaftskammern	Siedlungslandschaften	Alpenpflanzen
Waldgebiete	Futterbau	Siedlungsnutzung	Alpenflora
Waldnutzung	Viehhaltung	Sakralbaute	Alpenfauna
Waldnutzungen	Forstwirtschaft	Sakralbauten	Habitatvielfalt
Waldbauform	Waldbestand	Gewerbelandschaften	Artenvielfalt
Waldbauformen	Waldbestandes	Gewerbelandschaft	Poljen
Schutzwald	Vernetzungsqualität	Deponie	Waldhütte
Schutzwälder	Vernetzungsqualitäten	Deponiestandort	Kuhgebimmel
Holzproduktion	Melioration	Deponien	Naherholung
Naturschutz	Meliorationen	Hochspannungsleitung	Erholungsgebiet
Landschaftsschutz	Gewässerkorrektio	Hochspannungsleitungen	Naherholungsgebiet
Wiederbewaldung	Gewässerkorrekturen	Baudenkmal	Naherholungsgebiete
Reben	Anbauvielfalt	Baudenkmäler	Erholungsgebiete
Reblandschaft	Moorlandschaft	Bildstock	
Rebberg	Moorlandschaften	Bildstücke	

## B.1.2 Begriffe im LS-Kontext

Abendstunden	Alpenschweine	Alpinistin	Alpsommers
Abgrund	Alphorn	Alpkäse	Anlage
Absteigen	Alpine	Alpkäser	Area
Abstieg	Alpinen	Alpkäsetal	Aufsteigen
Abzweigung	Alpines	Alpmilchproduktion	Aufstieg
Alpenschaf	Alpinismus	Alpsegen	Ausflug
Alpenschafe	Alpinist	Alpsegnung	Ausgleichsbecken
Alpenschwein	Alpinisten	Alpsommer	Auslugsziel

Aussichten	Foto	Lift	Spitze
Auto	Fotograf	Loch	Stadt
Autobahn	Fototour	Lücke	Stall
Autobahn	Gasthaus	Marsch	Steigung,
Autobahnausfahrt	Gebäude	Mauer	Steigungen
Autobahnen	Gebiet	Mäuerchen	Stein
Autofahrer	Geburtsaus	Mond	Steinschlag
Automobilist	Gefahr	Nebel	Stimmung
Axenstrasse	Geiss	Offenheit	Stollen
Bahn	Geissen	Ort	Strässchen
Bahnhof	Geländes	Panzersperre	Strukturvielfalt
Bähnlein	Gemeinde	Passage	Talschaft,
Bahnstation	Gesteine	Pferd	Talschaften,
Bank	Gewitter	Platte	Terrasse
Bänkli	Grossvieh	Platz	Tour
Bauer	Handmähen	Psalm	Tourismus
Bäuerin	Hängebirke	Rast	Tourist
Baute	Haus	Restaurant	Tradition
Beiz	Heimatbildung	Riese	Treppe
Bergweltbegeisterte	Hindernis	Rinne	Tunnel
Besichtigung	Hof	Rot-Weiss	Turm
Bike	Höhe	Sammeln	Vernetzung
Blau-Weiss	Höhe	Sammler	Verzweigung
Blick	Horn	Sattel	Waldkauz,
Blume	Hotel	Schneefälle	Waldschnefpe
Boden	Hütte	Schneefällen	Waldschnepfen
Boden	Joch	Schneegestöber	Wand
Brauchtum	Kälte	Schneehasen	Wasserkraftwerk
Brauchtümer	Kamm	Schneehöhe	Weg
Busch	Kante	Schneemassen	Weggabelung
Büsche	Kapelle	Schneemenge	Weglein
Dorf	Käser	Schneemengen	Weite
Dörfchen	Kirche	Schneeschnelze	Weitläufigkeit
Dorfkirche	Kleinvieh	Schneestaub	Wetter
Ebenen	Kletterer	Schneesturm,	Wolke
Eis	Kloster	Schneetreiben	Zaun
Erhebung	Kraftwerk	Sehenswürdigkeit	Zäune
Fahrt	Kreuz	Sicht	Ruhe
Ferne	Kuh	Sonnenplätzli	Heimatgefühl
Föhns	Kühe	Sonnenseite	Heimatgefühle
Föhnsturm	Kurhaus	Spalt	
Föhnstürme	Landwirt	Spalte	
Föhntälern	Licht	Sperre	

### B.1.3 Compound-Wörter

Alp	Fels	park	Wasser
Älp	Fluss	Rasen	Werk/werk
Berg	graben	Schnee	Wiesen
Birken	gräben	Schwing	Wild
Buchen	Hafen	See	wolken
Damm	Hügel	Sonne	Wolken
Dämme	Kanal	Stein	Heu
Dunst	Landschaft	Ufer	
ebene	Nebel	Wander	
Eis	nebel	wanderung	

### B.1.4 Kontextwörter

erleben	neblig	talwärts	pittoresk
abwärts	regnerisch	touristisch	idyllisch
aufwärts	schön	wandern	hangwärts
biken	sonnig	weit	bewandern
fotografieren	staunen	wunderschön	bestaunen
landwirtschaftlich/e	steil	bestaunen	

## B.2 LSB in der Landschaftskonzeption und Erweiterungen mit LSB im Katalog

### Streusiedlungslandschaft

Weide	Siedlung	Gewässer	Wasserfall
Landwirtschaft	Forstwirtschaft	Kulturland	Aussichtspunkt
Bauernhof	Gras	Heimat	Baum
Zaun	Flur	Brauchtum	Einzelbaum
Wege	Weiler	Bauern	Wiese
Wald	Moor	Stille	Magenwiese
Stall	Waldrand	Aussicht	

### Mosaiklandschaft

Wald	Waldrand	Aussicht	Lesesteinhaufen
Offenheit	Trockenwiesen	Aussichtslage	Nutztier
Hügel	Weiden	Sicht	Nutzpflanzen
Berge	Feuchtgebiete	Zerschneidung	Terrain
Berggebiet	Naturinseln	Weite	Flur
Topographie	Waldung	Weideland	Tourismus
Landwirtschaft	Relief	Wiese	Erholung
Hof	Kulturland	Hecken	Naturpark
Bauernhof	Licht	Ufergehölz	

### Kulturlandschaft

Dorf	Infrastruktur	Parkanlage	Lesesteinhaufen
Weiler	landwirtschaftliche Nutzung	Agrartextur	Weidemauern
Hochstammbaum	Terrassenflur	Obstwiese	Holzzaun
Obstgarten	Sakralbauten	Baum	Weiden
Hecken	Selven	Einzelbaum	
Terrassen	Ernte	Fest	
Trockensteinmauern	Fruchternte	Alp	
Einzelbaute	Park	Alpen	

### Alplandschaft

Vieh	Hirte	Fauna	Abgeschiedenheit
Sömmerung	Weide	Wiesen	Aussicht
Stall	Saumweg	Trockenwiesen	Aussichtslage
Käseproduktion	Transportseilbahn	Trockenweiden	Weite
Zäune	Pferche	Wasserfall	Offenheit
Wege	Lesesteinhaufen	Brauchtum	Naturnähe
Brunnen	Senken	Natürlichkeit	Wildnis
Hochgebirge	Gewässer	Wanderziel	
Hochgebirgsrelief	Moore	Landwirtschaft	
Alpweiden	Wald	Naturnähe	
Alpgebäude	Flora	Ruhe	

### Jagdbanngebiet

Wald	Licht	Fischerei	Waldhütte
Ruhe	Jahreszeitenwechsel	Wanderweg	Picknickplatz
Abgeschiedenheit	Naturerlebnis	Feuerstelle	

### Berge

Verkehrsinfrastruktur	Rebberg	Wiese	Weinberg
Passhöhe	Wald	Weide	Berghütte
Bergreise	Offenland	Flurbereinigung	Tourismus
Brücke	Mosaiklandschaft	Streuhöfe	
Berggasthof	Bewirtschaftung	Wege	

## Wald

Waldbewirtschaftung	Wasserfälle	Mittelwald	Kulturwald
Waldnutzung	Wildnis	Plenterwald	Flur
Waldrand	Ruhe	Waldweide	Terrassen
Waldgesellschaft	Abgeschiedenheit	Kulturlandschaft	Wiesen
Hohlweg	Lichtungen	Weide	Offenland
Baum	Jahreszeitenwechsel	Vieh	Verwaldung
Totholz	Naturerlebnis	Holzzaun	Schutzwald
Artenvielfalt	Wanderweg	Trockensteinmauern	Waldstrasse
Einzelbaum	Jagd	Baumarten	Lawinenverbauung
Kreten	Holznutzung	Weidenutzung	Vergandung
Lichtungen	Feuerstellen	Wildnisinseln	Wildbrücke
Felsen	Waldhütte	Gelände	
Gewässer	Picknickplatz	Park	

## Felsenlandschaft

Wald	Baum	Einzelbaum	Weg
Gewässer	Alplandschaft	Bäume	Saumpfad
Lichtung	Alpen	Felsblock	Felsenweg
Wasserfall	Alp	Senke	

## Pass

Moor	Kunststrasse	Hochspannungsleitungen	Gasthaus
Hochebene	Schutzwälder	Wegabschnitte	Hospiz
Hochtal	Sakralbauten	Tourismus	Susten
Verkehrsinfrastruktur	Reisewege	Brücke	
Passroute	Bergreise	Wege	
Passhöhe	Rauheit	Wegnetz	

## Flachmoor

Gewässer	Flachmoor	Hochtäler	Heuwiesen
Moorlandschaft	Biotop	Flusstäler	Kulturland
Moorlandschaft	Pass	Seen	
Moorwald	Pässen	Streuwiesen	
Hochmoor	Hochebene	Magerwiese	

## Anhang C

# Kosinus-Ähnlichkeit zwischen den Landschaftsbegriffen



## Persönliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und die den verwendeten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Unterschrift:

N. Brandenburg

Datum: 31.03.2021, Baden