



Semantics for Mountaineering History - SEMOHI

Gerald Hiebel, Claudia Posch, Gerhard Rampl, Elisabeth Gruber, Klaus Hanke, Eva Zangerle, University of Innsbruck / Institute for Languages and Literature: Linguistics, Surveying and Geoinformation Unit, Department of Computer Science

Overview

Project Information

Sources

- ⌘ The Austrian Alpine Journal / Alpenwort corpus
- ⌘ Photo Index of the Austrian Alpine Club

Methodology

- ⌘ Semantic Representation of Information
- ⌘ Named Entity Recognition and Named Entity Linking
 - ⌘ Gazetteers
 - ⌘ Manual Name Identification: Gold Standard
 - ⌘ Automated Name Matching

SEMOHI – Semantics of Mountaineering History

🎗️ Start: 03/2017

🎗️ Funding: go!digital call of the Austrian Academy of Sciences 2016

🎗️ Team: Gerald Hiebel (Surveying and Geoinformation Unit), Claudia Posch, Gerhard Rampl, Elisabeth Gruber (Institute for Languages and Literature: Linguistics), Eva Zangerle (Department of Computer Science)

Recognition of place names and personal names in the Alpenwort corpus and the Photo Index of the Austrian Alpine Club.
Adding of semantic information.

Photo Index of the Austrian Alpine Club



A. Örtliche Gebiete
(Gebirge und Länder)

I. Ostalpen

<p>1. Bregenzer Wald, Allgäuer Vor- und Hochalpen</p> <p style="text-align: center;">Umgrenzung:</p> <p>Bregenzer Wald und Allgäuer Voralpen: Bregenz — Rheintal — Walgau — Walsertal — Illertal — Bodensee.</p> <p>Allgäuer Alpen: Illertal — Kl. Walsertal — Starzloch — Oberstes Bregenzerachtal — Sattel von Hochkrumbach — Lechtal — Schwäbisches Alpenvorland</p> <p style="text-align: center;">Bregenzer Wald und Allgäuer Voralpen:</p> <p>1 Alberschwende 2 Andelsbuch 3 Au 4 Au mit Canisfluh 5 Muenalpe, untere, und Walmendingerhorn 6 Muenfeldergruppe zwischen Wilden-</p>	<p>28 Hochgrat, Brunnenauer Alpfessel 29 Hochgrat von D. 30 Hochkrumbach, Haus, mit Biberkopf 31 Hochkrumbach, Kirche, mit Biberkopf 32 Hohenems gegen S.O. 33 Hopfreen 34 Isen, Hoher 35 Isen, Hoher 36 Isen, Hoher 37 Isen, Hoher 38 Isen, Hoher, a. Biberkopf a. Junciere Gottesackerwand vom Gällers W. 39 Isen, Hoher 40 Isen, Hoher 41 Isenalp geg 42 Isen, Nordplato 43 Isenfeld mit 44 Isenfeld, B 45 Immenstädte 46 Ingolstädter städter Horn</p>
---	--

5 volumes: 1927-1928, 1931, 1935, 1936, 1941

containing descriptions for approx. 32.000 photos of which 85 % depict places of some kind

Alpenwort:

The Austrian Alpine Journal as a Corpus (1869-today)

- Scientific essays, expedition and mountaineering reports, route descriptions, descriptions of mountains and mountain ranges, alpine literature
- Total: 42.118 pages
- The go!digital project Alpenwort (2014 - 2017) turned the Austrian Alpine Club Magazine documenting 146 years of mountaineering history into the linguistically annotated **Alpenwort corpus**.

Session Building Bridges – Corpus Linguistics and Digital Humanities
Wednesday 06.12.2017, 11-12.30

What is Semantic Annotation?

Grosse Geiger.

141

person name

place name

Richard Issler (in **Wien**). Erste Besteigung des **Grossen Geiger**. (10600 Fuss = 3350 Meter Keil.) Sonntag den 20. August marschierte ich mit **Michel Groder** von **Kals** wenige Minuten nach 6 früh durch das **Virgenthal** aufwärts nach **Prägraten**. Hier bei **Isaias Steiner** verproviantirten wir uns und stiegen Nachmittags zur **Johann'shütte** empor.

Gegen 6 Uhr Nachmittags langten wir in der Johann'shütte an und trafen hier einen Soldaten, welcher mit Holz-

Richard Issler (in **Vienna**). First Ascent of the **Grossen Geiger**. (10600 feet = 3350 meter). Sunday the 20th of August I walked with **Michel Groder** from **Kals** some minutes...

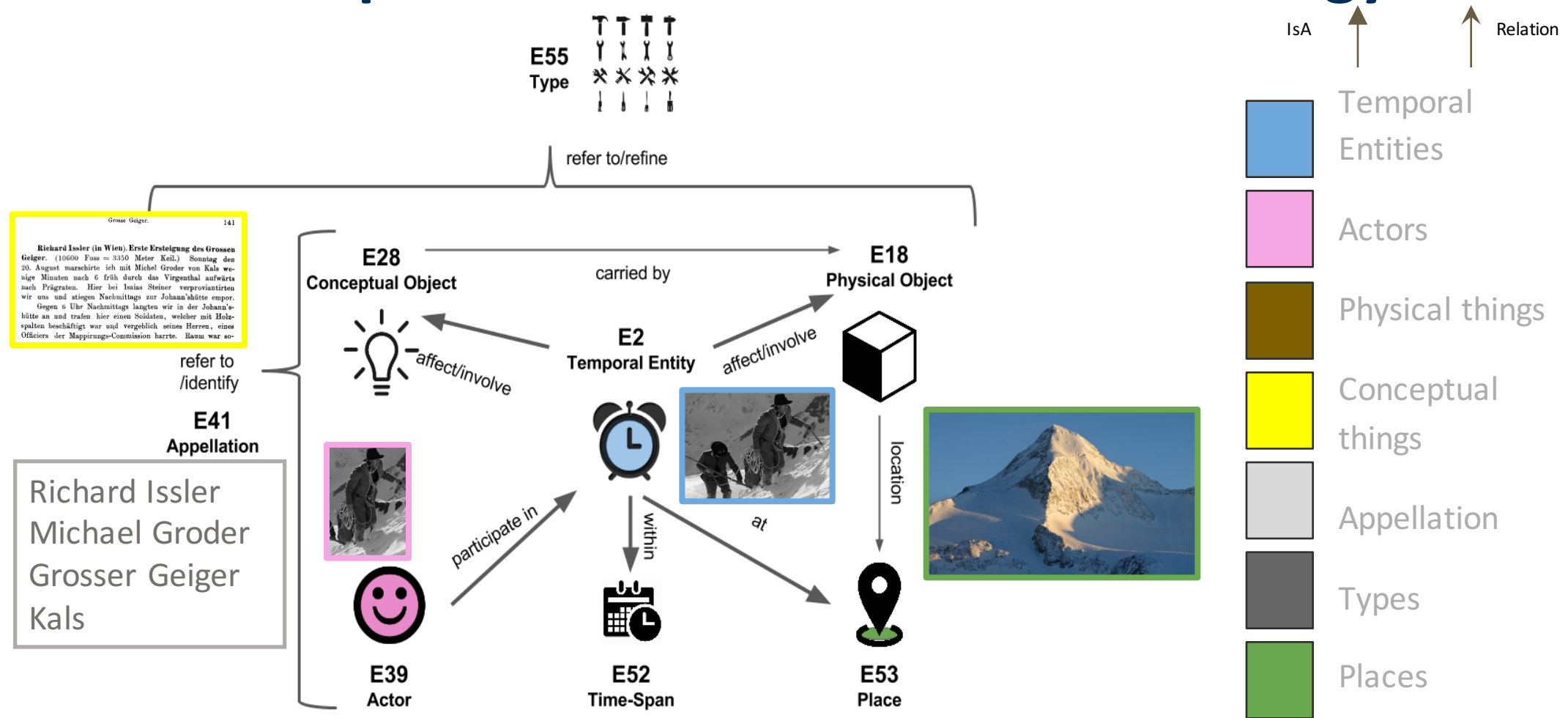
people



place



Semantic Representation: CIDOC CRM Ontology



Semantic Web Standards / Linked Open Data



GeoNames



WIKIPEDIA
Die freie Enzyklopädie

Richard Issler (in Vienna). First Ascent of the Grosser Geiger. (10600 feet = 3350 meter). Sunday the 20th of August I walked with Michel Groder from Kals some minutes

Großer Geiger



Der Große Geiger (links) von Norden, vom Keeskogel aus gesehen

Höhe	3360 m ü. A.
Lage	Salzburg und Tirol, Österreich
Gebirge	Venedigergruppe
Dominanz	2,7 km → Großvenediger
Schartenhöhe	285 m ↓ Breiter Sattel
Koordinaten	Ö 47° 5' 37" N, 12° 18' 29" O

Issler, Richard (1842-1896), Alpinist, Herausgeber und Photograph

Issler Richard, Herausgeber, Lichtbildner und Alpinist. * Berlin, 1842 (43); † Wien, 17. 5. 1896. Gründete 1875 die „Neue Deutsche Alpenzeitung“ und redigierte sie bis 1880. Ein vorzüglicher Lichtbildner. Auf seinen zahlreichen Bergfahrten (Dachsteingebiet, Hohe Tauern, Dolomiten und Steiner Alpen) stieß er immer wieder — mit wechselndem Erfolg — 1877 versuchte er mit Faschingbauer die 1. Besteigung der Dachsteinsüdwand, 1878 mit Führer Dimai die der Kl. Zinne, 1879 zweimal die der Gr. Bischofsmütze mit Sattler (beide Führer Knauss) ohne Erfolg. Dagegen gelangen ihm: 1868 Rax, Gamseck, 1. Erst. mit Löttsche; 1871 Gr. Eiser, 1. Besteig., mit Brandtner und Gr. Geiger, 1. Besteig., mit Führer Hoher Nuvolao, 1. tourist. Best., mit Führer Siorpaës; 1876 Brana (Steiner Alpen), neuer Anstieg mit Frischauf und Ratmoser; 1877 Froppa (Marmarolegruppe), W-Grat, 1. Besteig. und Führer Lacedelli. — Bei einer winterlichen Besteigung des Antelao erfror sich I. beide Füße und erkrankte, ohne je wieder seine volle Gesundheit zu erlangen. I. gehört zu den Führern, die in den 70er Jahren durch ihr Wirken und Vorbild der Wr. Bergsteigerschule einen bedeutenden Aufschwung gaben.

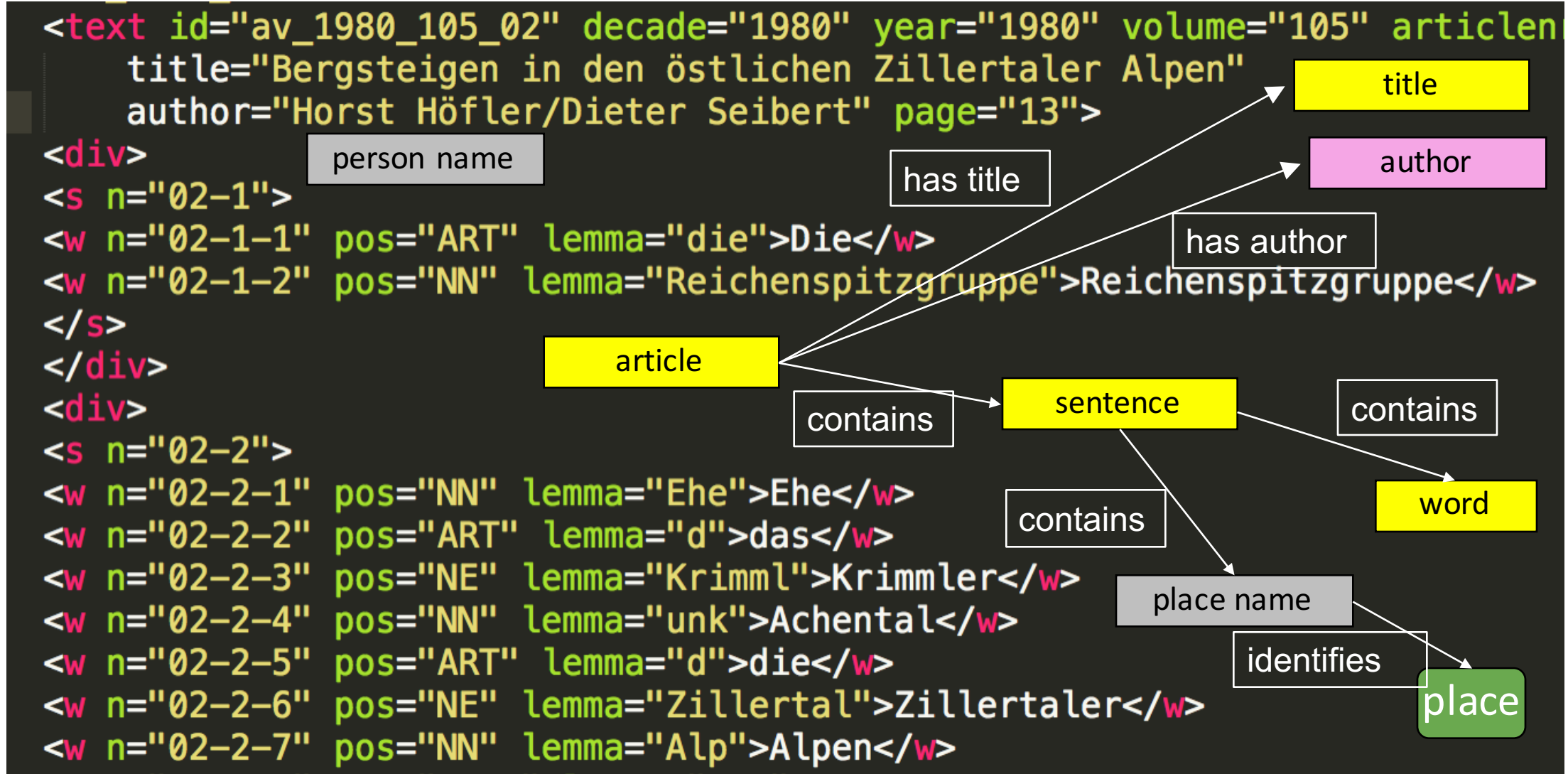
L.: Mitt. DÖAV 1896, S. 140; ÖAZ 1896; EOAI, III; WB.

PUBLIKATION: ÖBL 1815-1950, Bd. 3 (Lfg. 11, 1961), S. 45f.

Alpenwort Corpus: Structured & Part of Speech (POS) tagged

```
<text id="av_1980_105_02" decade="1980" year="1980" volume="105" articlen  
  title="Bergsteigen in den östlichen Zillertaler Alpen"  
  author="Horst Höfler/Dieter Seibert" page="13">  
<div>  
<s n="02-1">  
<w n="02-1-1" pos="ART" lemma="die">Die</w>  
<w n="02-1-2" pos="NN" lemma="Reichenspitzgruppe">Reichenspitzgruppe</w>  
</s>  
</div>  
<div>  
<s n="02-2">  
<w n="02-2-1" pos="NN" lemma="Ehe">Ehe</w>  
<w n="02-2-2" pos="ART" lemma="d">das</w>  
<w n="02-2-3" pos="NE" lemma="Krimml">Krimmler</w>  
<w n="02-2-4" pos="NN" lemma="unk">Achental</w>  
<w n="02-2-5" pos="ART" lemma="d">die</w>  
<w n="02-2-6" pos="NE" lemma="Zillertal">Zillertaler</w>  
<w n="02-2-7" pos="NN" lemma="Alp">Alpen</w>
```

Alpenwort Corpus: Structured & Part of Speech (POS) tagged



Named Entity Recognition and Named Entity Linking

- POS Tagger does name recognition, but badly.
- In particular place names consisting of more than one word
 - <w n="02-2-13" pos="ADJA" lemma="hoch">Hohen</w>
 - <w n="02-2-14" pos="NN" lemma="Tauern">Tauern</w>
- Use of gazetteers (lists of place names) to identify names
 - ⌘ Recognition – identify place names in text that exist in a gazetteer
 - ⌘ Linking – identify the place in a gazetteer that is referenced by the place name in the text (disambiguation)

Gazetteers

Usually consist of:

- identifier
- name
- place type
- coordinate information

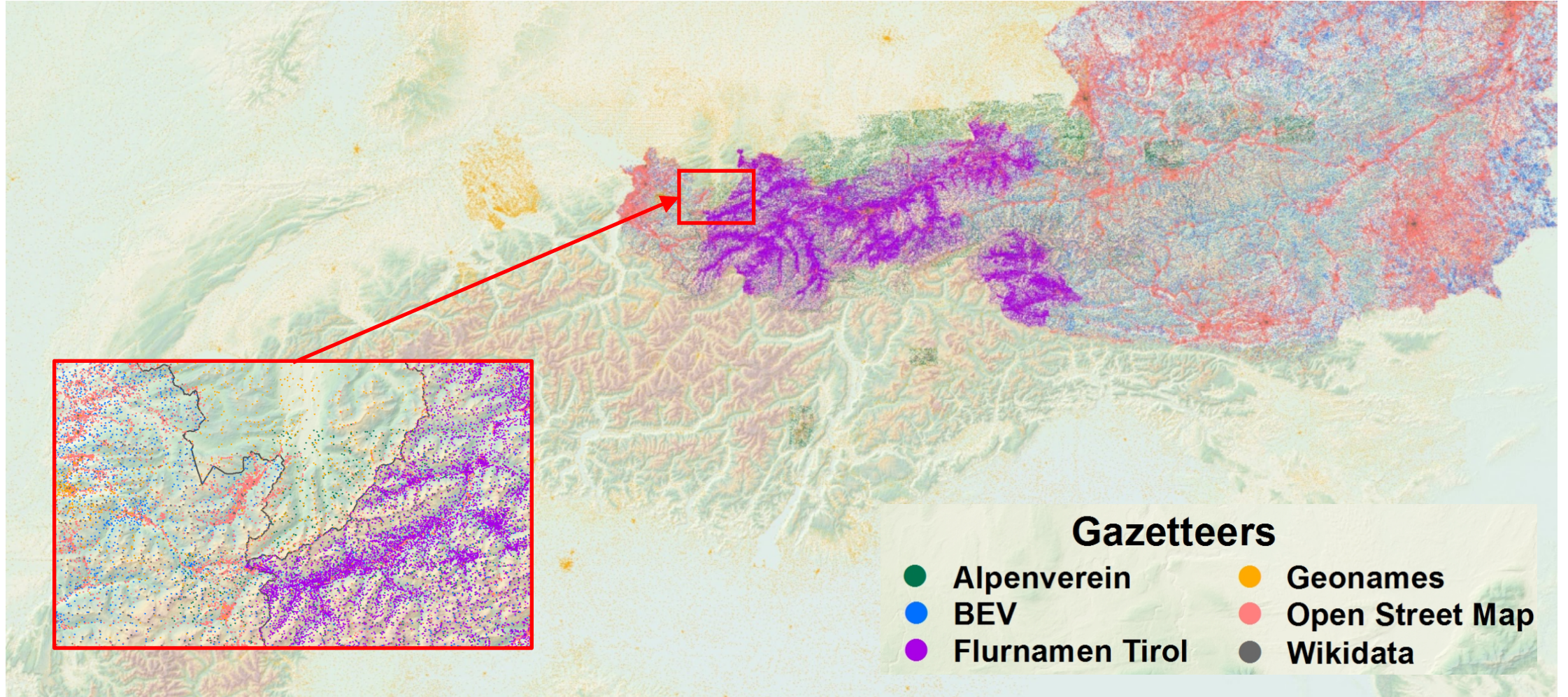
May consist of:

- alternate names
- relations to other places

List of gazetteers used:

- Austria National Survey (BEV)
- Austrian Alpine Club
- Field Name Survey in the Tyrol
- Open Street Map
- Geonames
- Wikidata

Integrated Gazetteers



Manual Identification of Places

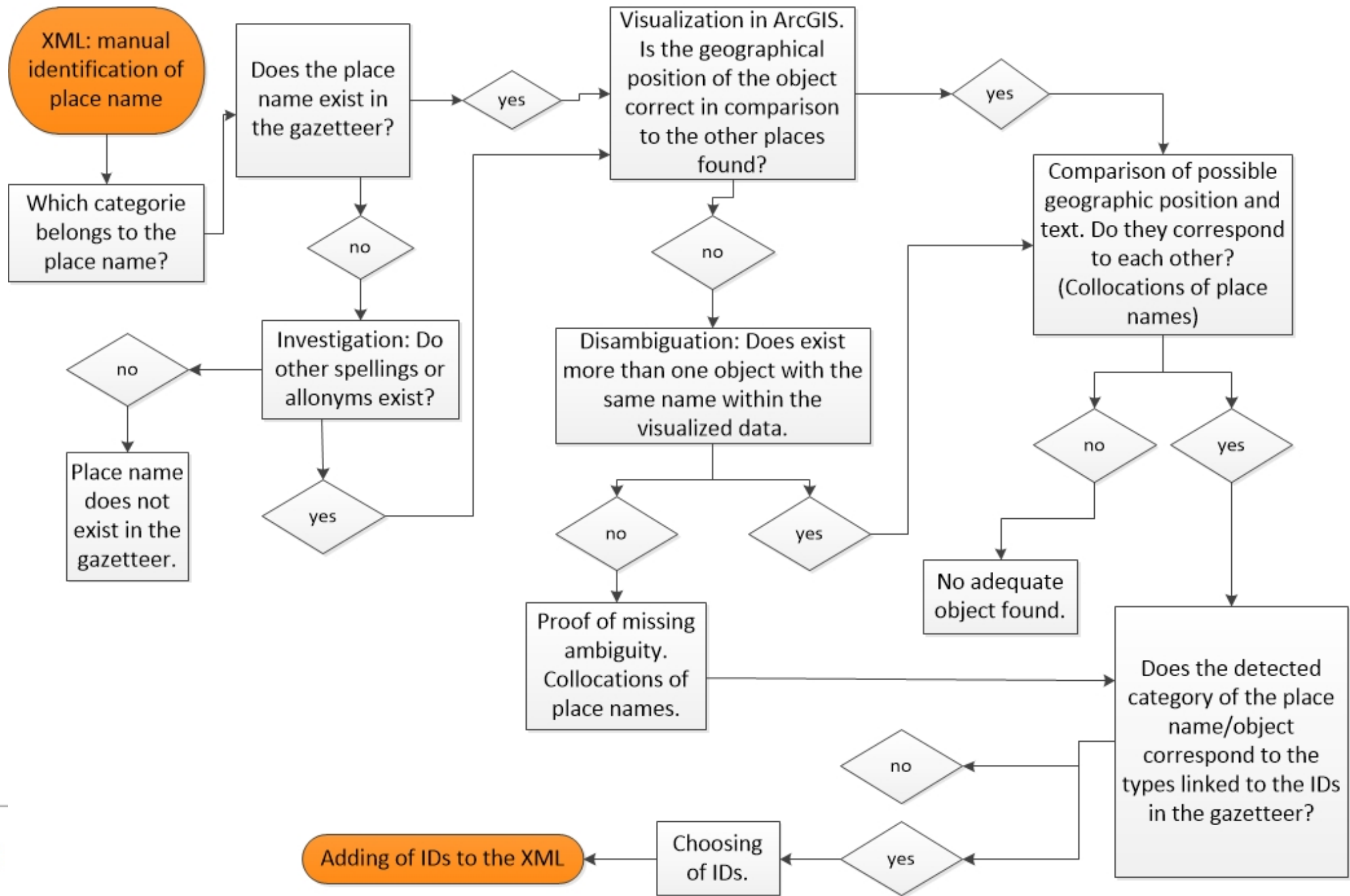
- Recognition of place names in a text – place names are tagged

Der `<placeName>Roßkopf</placeName>`, 2845 m, kann dabei ohne große Mühe, mit kurzweiligem Aufstieg [...]

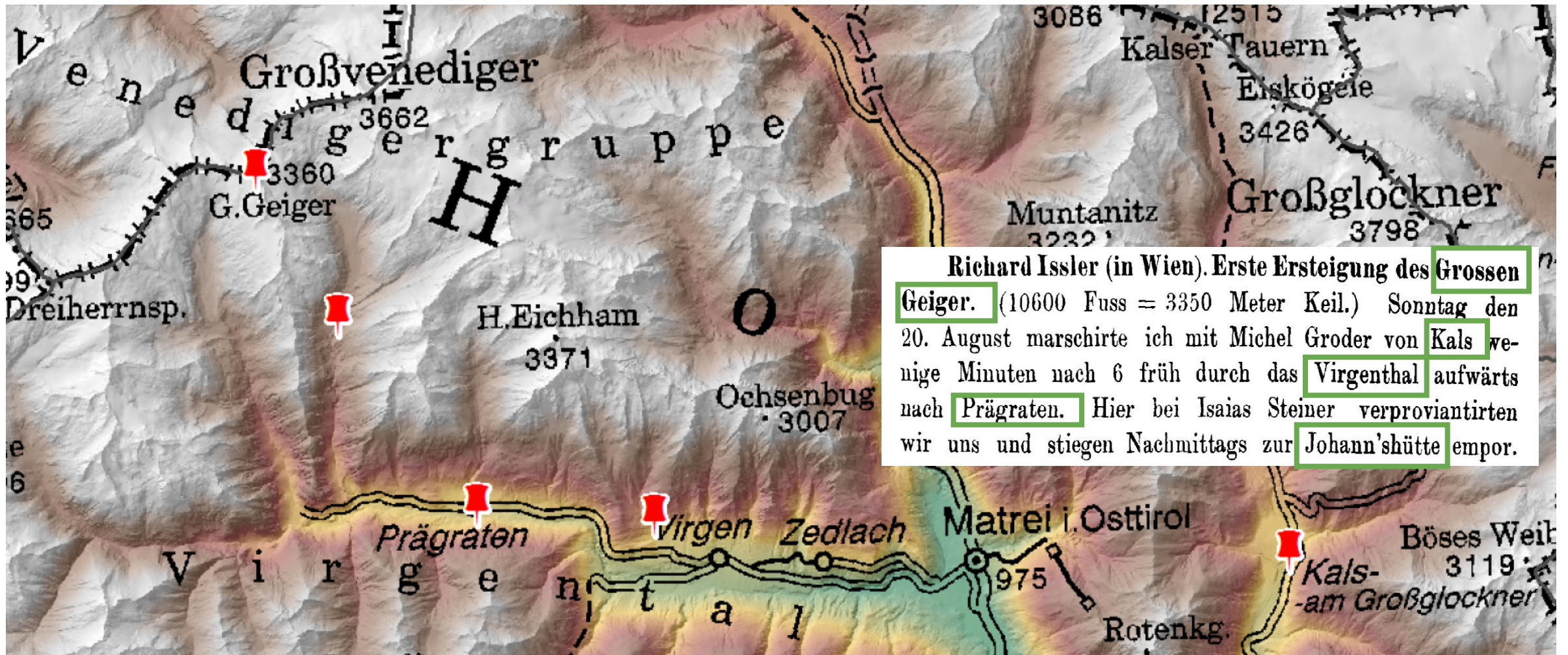
- Manual matching to corresponding entries in the gazetteer – identifier of gazetteer is added to the tag

Der `<placeName gn='gn_6939049'>Roßkopf</placeName>`, 2845 m, kann dabei ohne große Mühe [...]

Creation of ground truth (gold standard) articles
for automated place identification

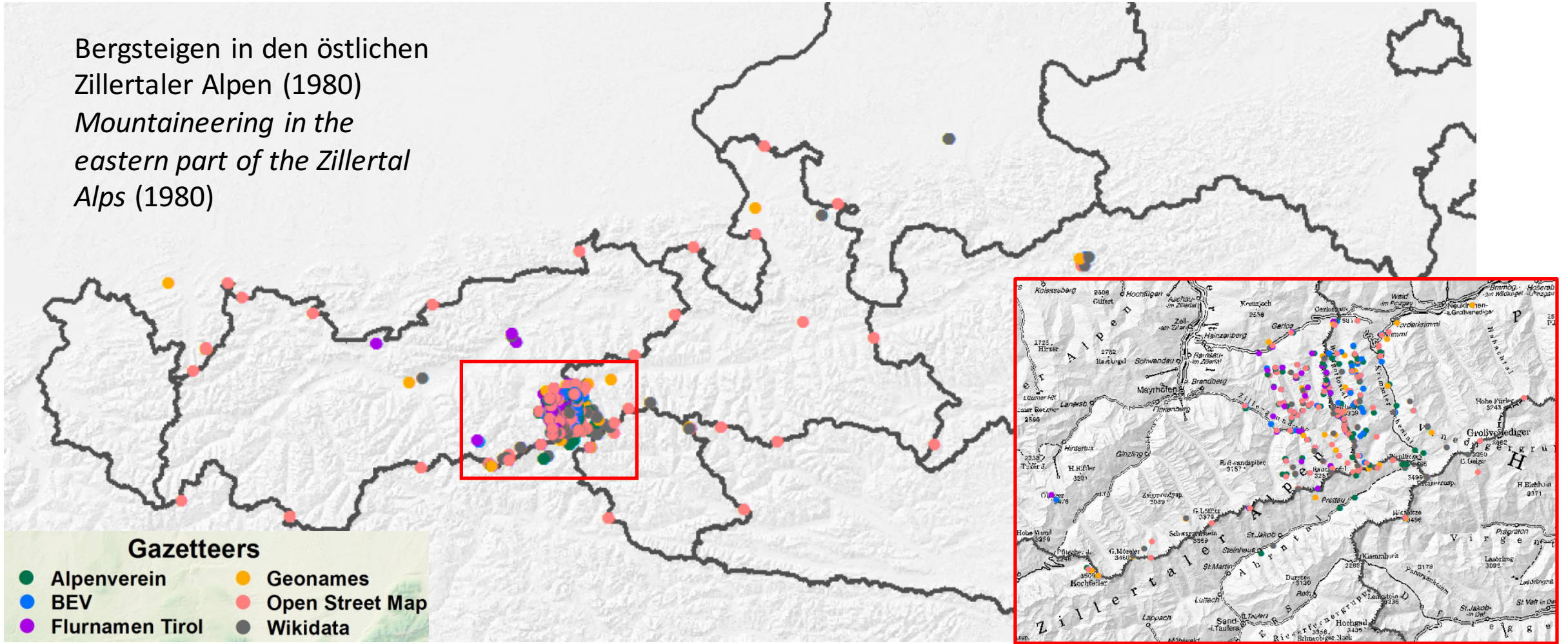


Manual Identification → Geographic Localisation



Manual Identification

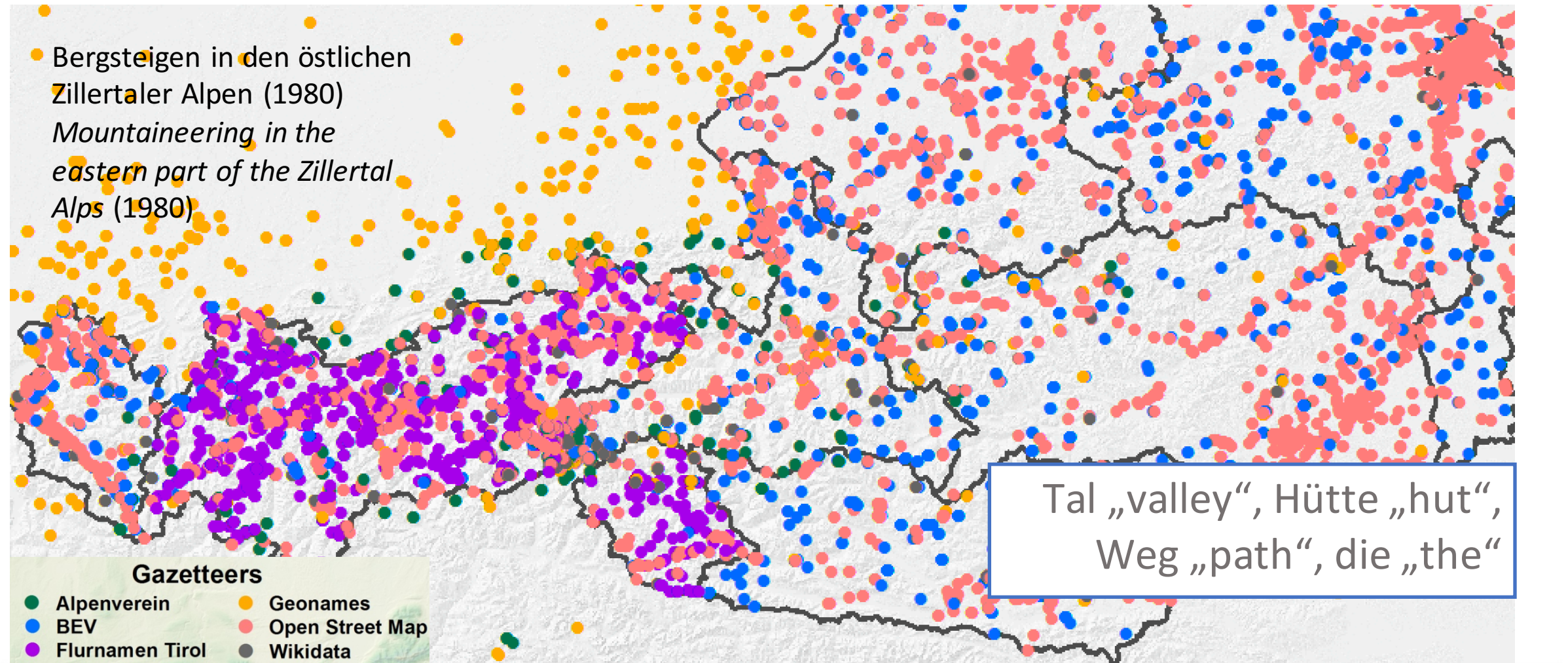
Bergsteigen in den östlichen
Zillertaler Alpen (1980)
*Mountaineering in the
eastern part of the Zillertal
Alps (1980)*



Automated Place Name Identification Problems & Approach

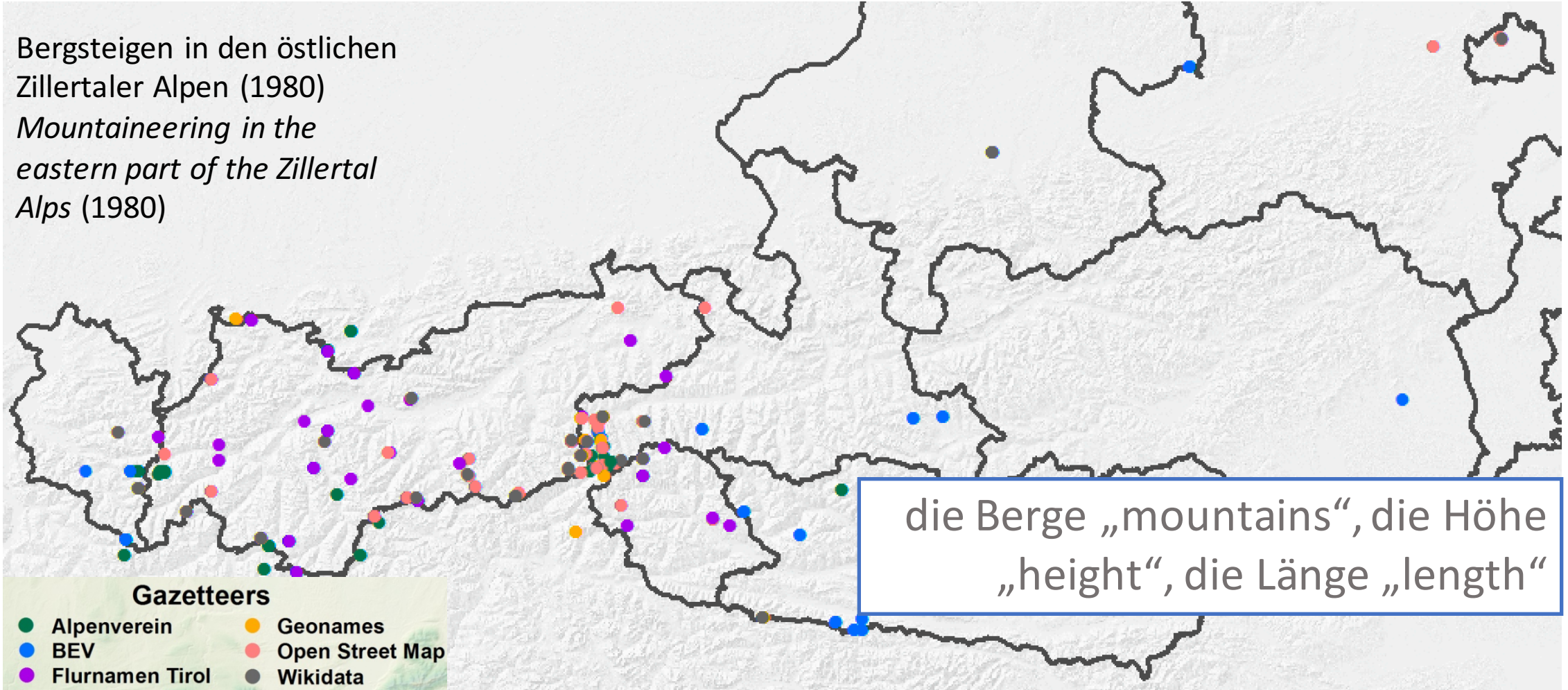
- Place Name Recognition
- Declension of place names
- Disambiguation

Place Name Recognition & Disambiguation: One Word Place Names



Place Name Recognition & Disambiguation: Two Word Place Names

Bergsteigen in den östlichen
Zillertaler Alpen (1980)
*Mountaineering in the
eastern part of the Zillertal
Alps (1980)*



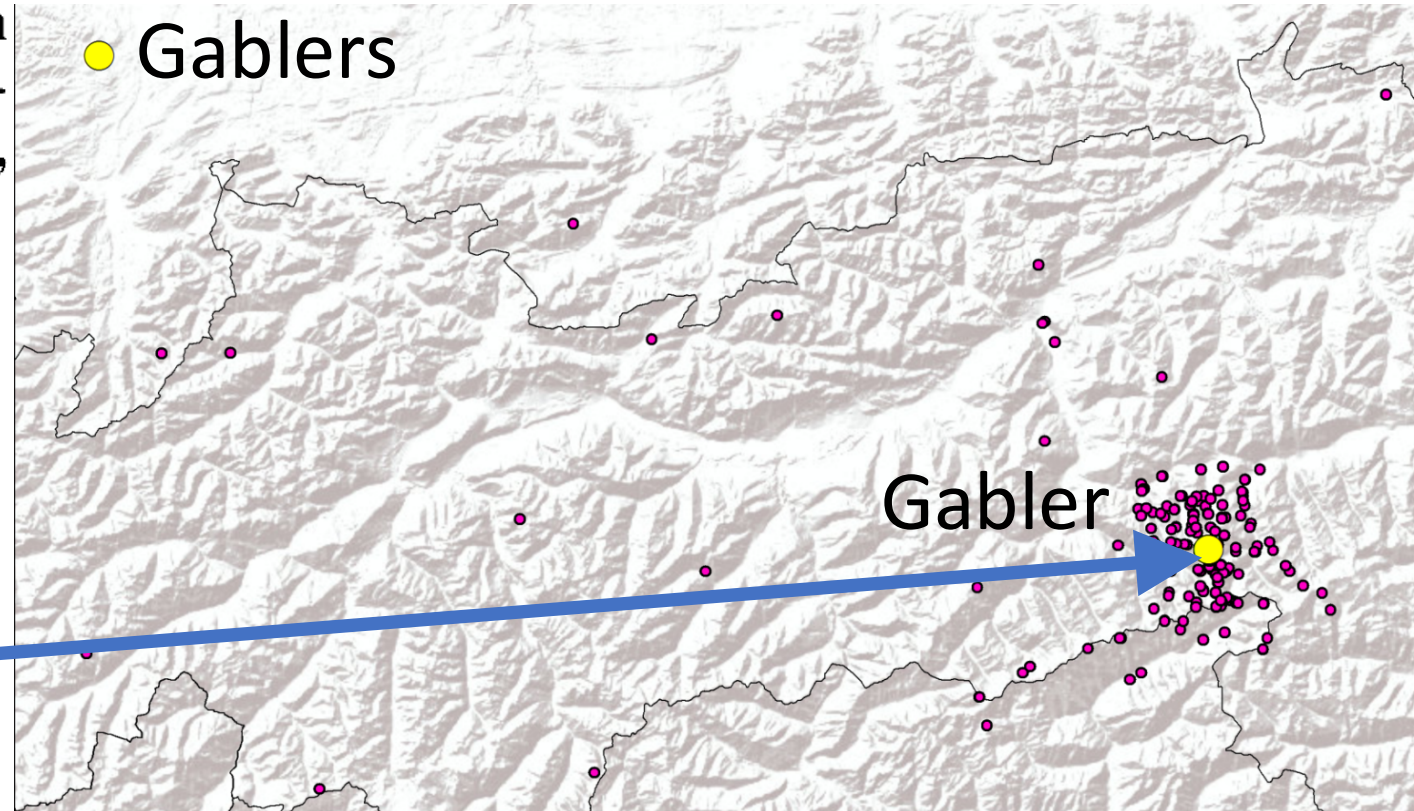
Declension of Place Names

Gabler

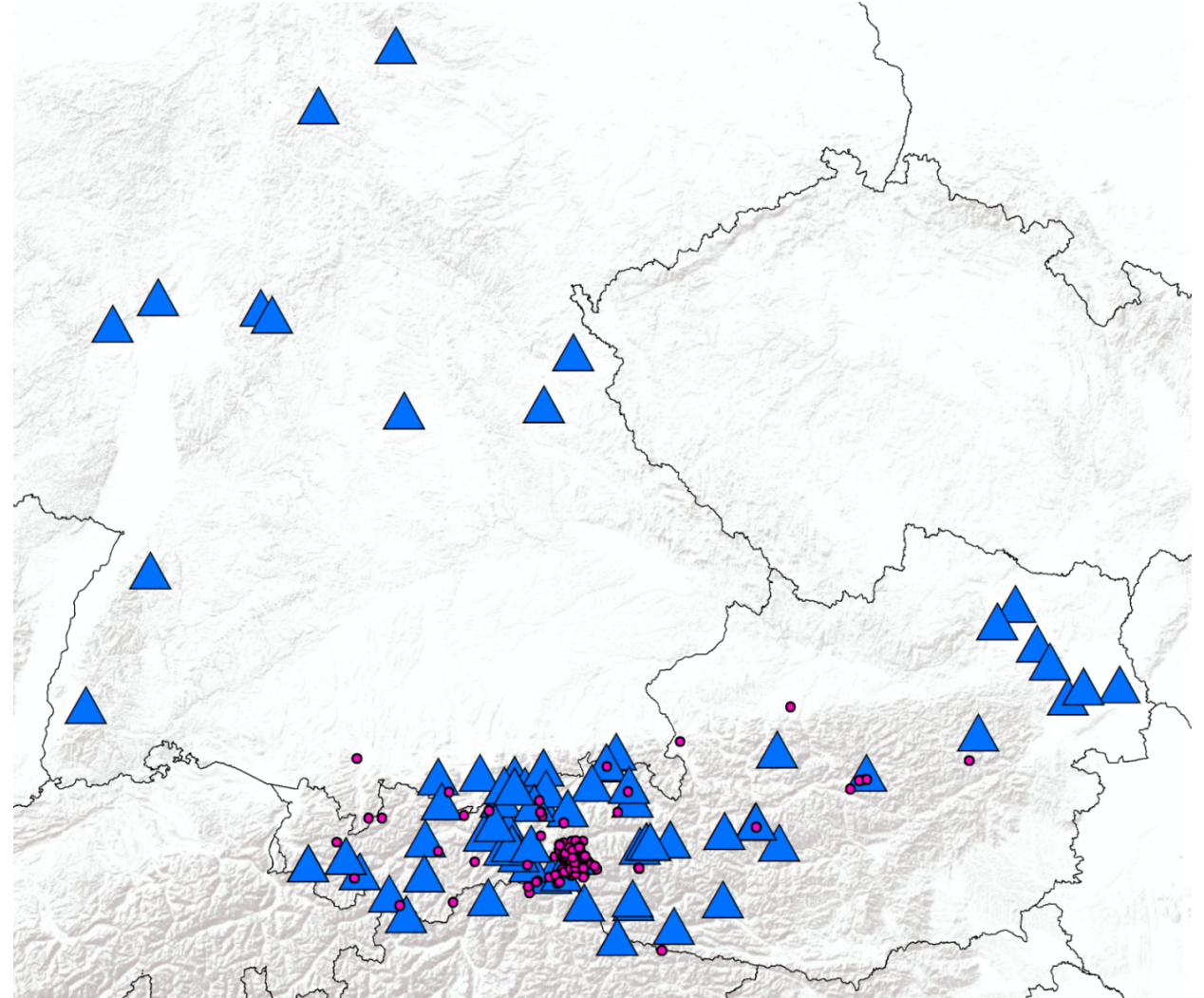
Da lockt das Felshorn des **Gablers**, der meist von Reichenspitzbesteigern als Zugabe »mitgenommen« wird. **Keeskarkopf**, **Zillerplattenspitze**,

O. Nafe und O. Fischer erreichten 1885 den Gipfel des **Gablers**, 3260 m. Er ist der nordöstliche

die vielbesuchte **Zittauer Hütte**, 2330 m. Im unmittelbaren Bereich des Wildgerloskeeses ist sie Ausgangspunkt sowohl für die üblichen und häufig begangenen Nordanstiege auf Reichenspitz und **Gabler**, jedoch auch für eine Reihe selten ausgeführter kombinierter Fahrten. Über die

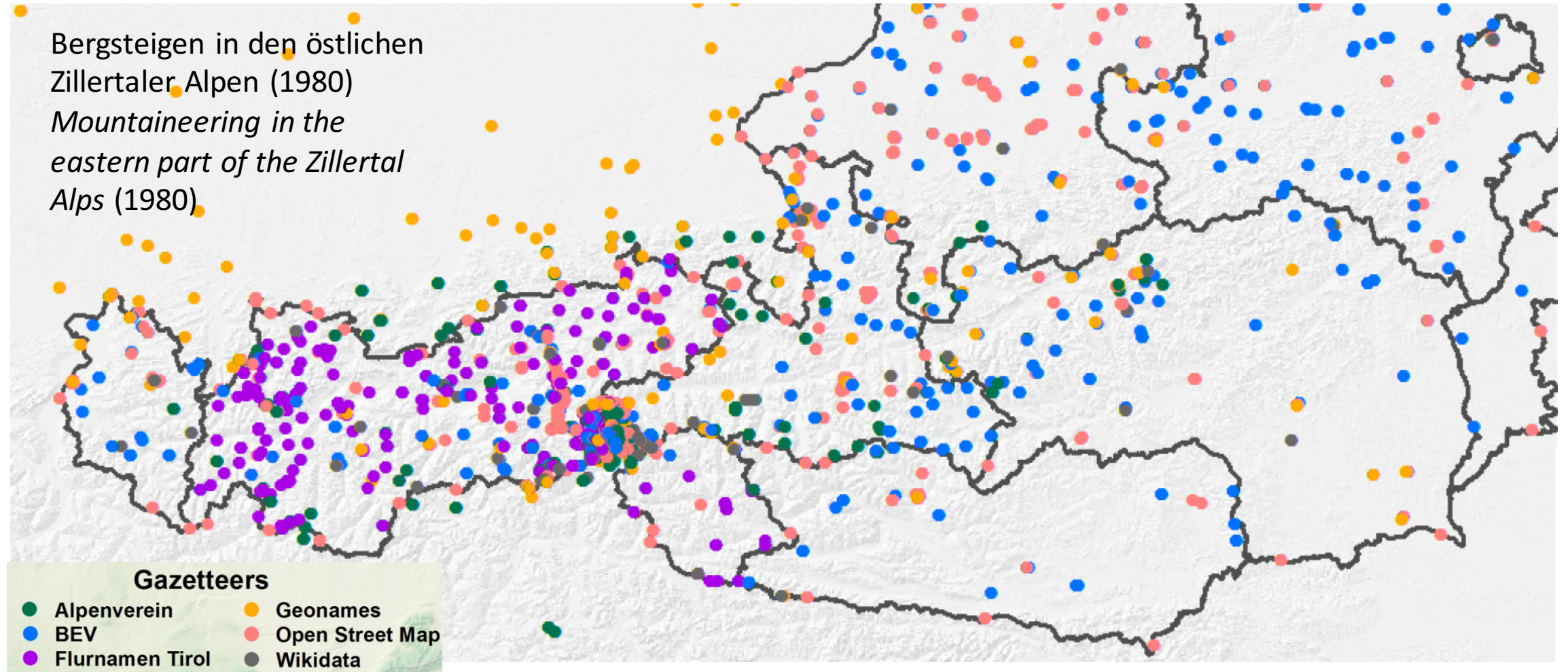


Disambiguation: *Rosskopf*



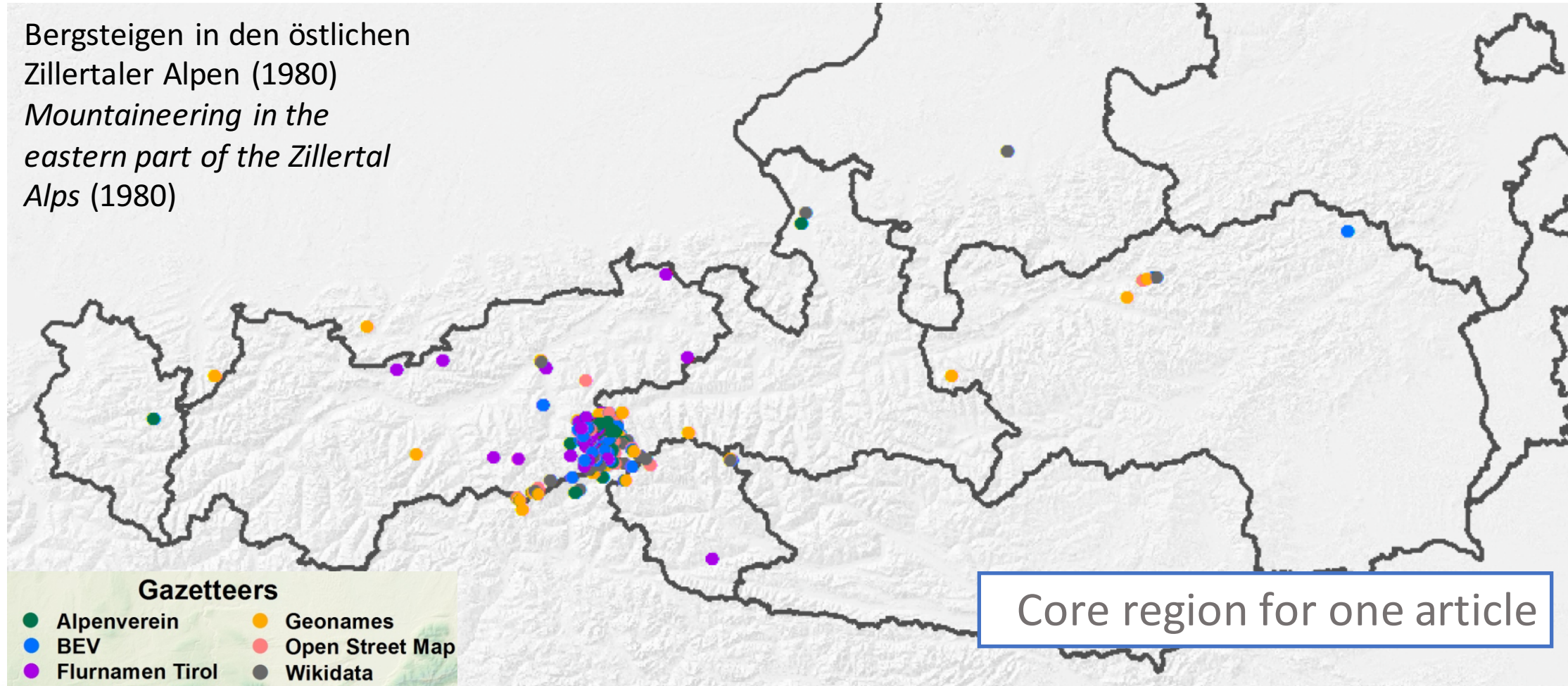
Disambiguation Problems with Recognized Places

Bergsteigen in den östlichen
Zillertaler Alpen (1980)
*Mountaineering in the
eastern part of the Zillertal
Alps (1980)*



Only Unique Names in Gazetteer (Recognised Place Names)

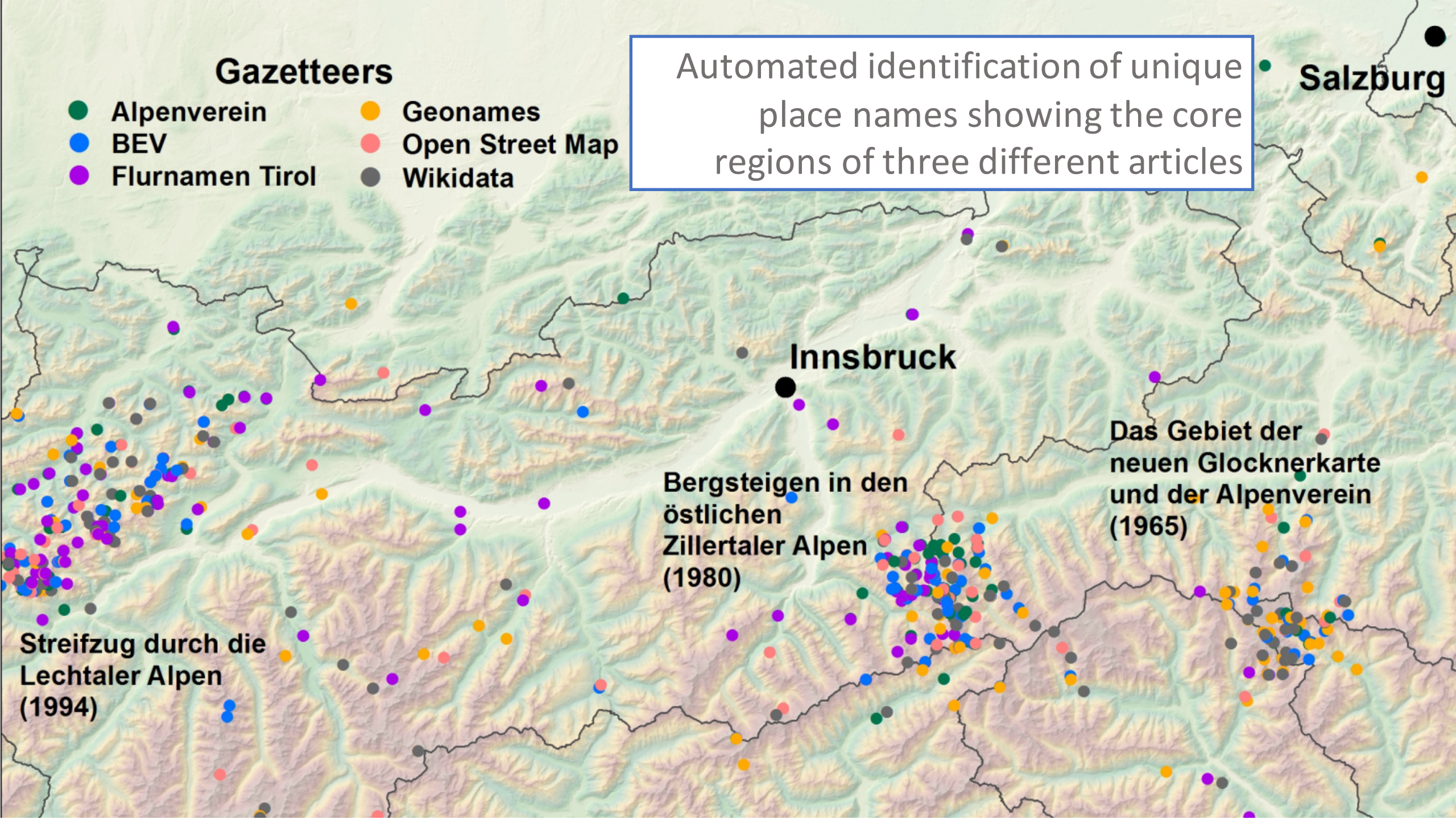
Bergsteigen in den östlichen
Zillertaler Alpen (1980)
*Mountaineering in the
eastern part of the Zillertal
Alps (1980)*



Gazetteers

- Alpenverein
- BEV
- Flurnamen Tirol
- Geonames
- Open Street Map
- Wikidata

Automated identification of unique place names showing the core regions of three different articles



Salzburg

Innsbruck

Das Gebiet der neuen Glocknerkarte und der Alpenverein (1965)

Bergsteigen in den östlichen Zillertaler Alpen (1980)

Streifzug durch die Lechtaler Alpen (1994)

Strategy for the automated Recognition and Identification of Place Names in the Corpus (to be tested)

- Identify place name candidates through exact string matches and possible variations (declension)
- Try to exclude „false positive“ candidates through statistics on word frequency in alpine corpus, pos tags and word frequency in gazetteer
- Identify core regions for articles based on unique names (or rare names) in gazetteers
- Create parameter for geographic proximity to core region
- Analysis of place types: preferred place types in- and outside core region

Summing up...

- Semantic annotations allow the computational analysis of categories like places, people and events and enable linking to other sources
- The geographic visualization of identified places creates a new mean of communicating the content of a text.
- The used Gazetteer has to be very detailed (integration of various gazetteers) to cover the occurring place names
- Manual Identification is essential to understand the human process and create a ground truth (Gold Standard)
- Automated place name identification can only create probable solutions as statistical information is used for the process, but semantics can improve the results

Thank You!

