

TM

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Institut für
Romanische
Philologie

*Crowdsourcing, Swarm Intelligence,
Data Mining* und die Wissenschaften

21.1.2013



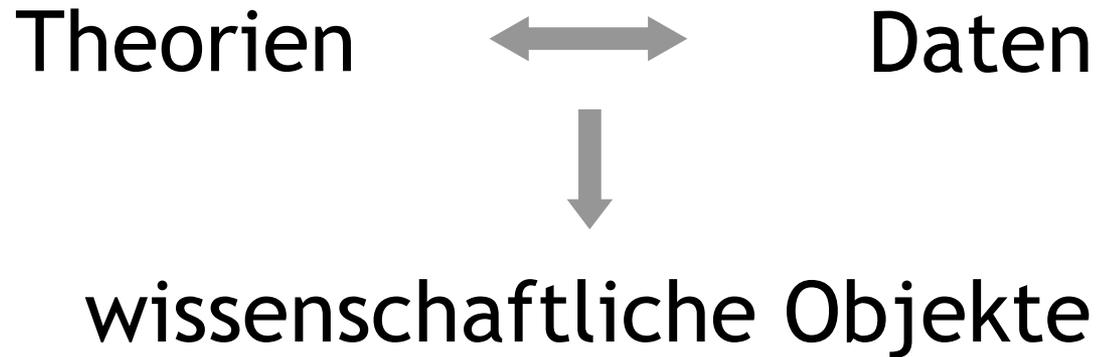
Morgenröte
eines Paradigmenwechsels -
im linguistischen Umgang mit der sprachlichen Variation

Thomas
Krefeld

Gliederung

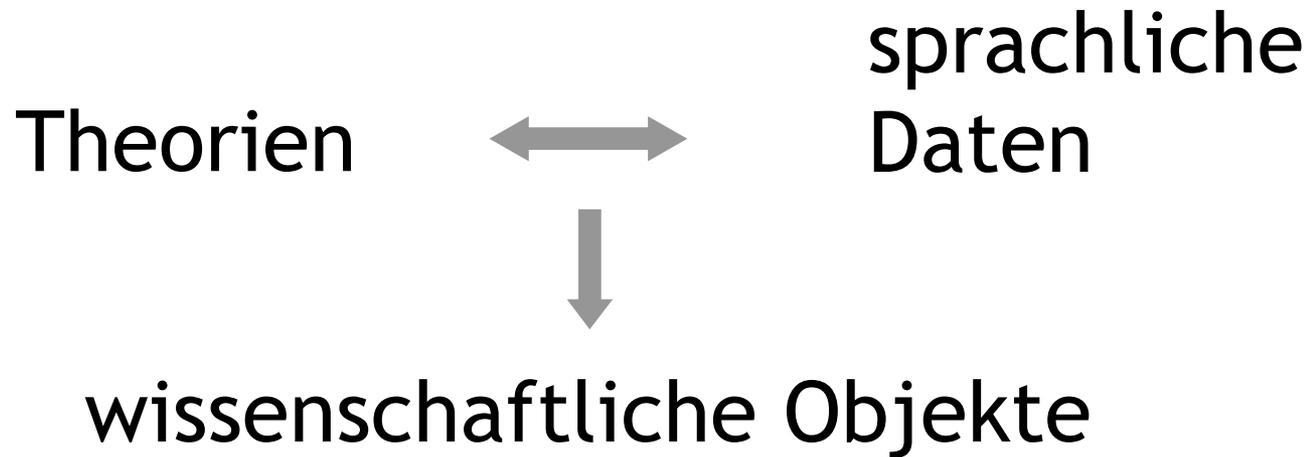
1. Sprache: *crowdproduced and crowdproducing*
2. emisches *crowdsourcing* in der Sprachwissenschaft
3. Beispiel: Matching von Ontologie und Bezeichnung

Wissenschaften



- Labordaten
- Spontandaten
eine Quelle:
crowdsourcing

unterschiedliche Wissenschaften



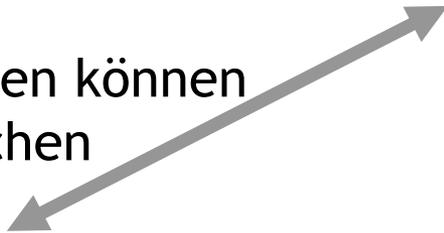


Sozialwissenschaften (Soziologie, Ethnologie)
Sprachen konstituieren Gemeinschaften ('sozial')



Sprach-
SPRACHE
wissenschaft

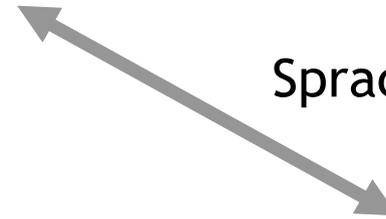
alle Menschen können
sprechen



Naturwissenschaften (Neurologie)



Sprachen verändern
sich



Geschichtswissenschaften



- jeder Mensch spricht eine (oder mehrere) spezifische Einzelsprache(n)
 - Einzelsprache(n):
 - gemeinschaftsbildend (identitätsstiftend)
 - historisch veränderlich (ändern sich im Gebrauch, ohne Änderungsabsicht ihrer Sprecher)

Heterogenität und Vielfalt der Sprachen

'äußere' Sprachverschiedenheit: Italienisch' vs. 'Englisch' vs. 'Französisch' usw.

'innere' S.: Heterogenität ('Variation') der Einzelsprachen: jede Sprache ist ein dynamisches **Bündel** von Varietäten (Dialekte, Stile, Soziolekte, gesprochene/geschriebene Sprache usw.)

Umgang mit der Variation

- Sprachen und ihre Varietäten sind ein Produkt der *crowd*
 - ▶ besondere Affinität von *crowd* und Sprachwissenschaft
- Variation: unvermeidbare Ausgangsbedingung der sprachwissenschaftlichen Beschreibung von Einzelsprachen

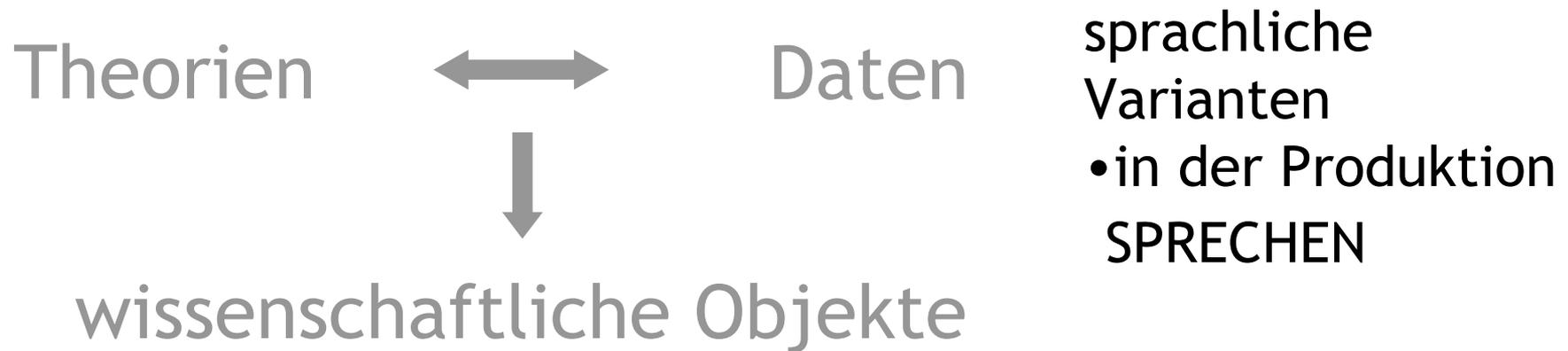
Sprachwissenschaft



Gliederung

1. Sprache: *crowdproduced and crowdproducing*
2. emisches *crowdsourcing* in der Sprachwissenschaft
3. Beispiel: Matching von Ontologie und Bezeichnung

Sprachwissenschaft (Varietätenlinguistik)



Sprachvariation
• Variablen
• Varietäten
SPRACHE

Analyse von Produktionsdaten
(SPRECHEN) nicht hinreichend
für die Beschreibung der
SPRACHE

→ Instanz des SPRECHERS

Sprache

- gehört zum Kern der Lebenswelt der SPRECHER
- lebensweltliches SPRECHERwissen
unhintergebares Fundament der
SPRACHwissenschaft
in philosophischer Terminologie: transzendental

(,Bedingung der Möglichkeit von...‘)
- SPRECHERwissen konstitutiv für das
wissenschaftliche Objekt SPRACHE

crowdsourcing als Königsweg zum SPRECHERwissen

- nicht SPRACH-, sondern SPRECHERbezogen
- Eröffnung einer (emischen) Innenperspektive
- Zugang zum Wissen des Sprechers
- Fortsetzung der Phänomenologie mit anderen Mitteln
- ‚vormediale‘ Autoritäten
 - Edmund Husserl (1992 [1976, unvollst.1936]): *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie*, Hamburg (= Gesammelte Schriften, hrsg. von Elisabeth Ströker, 8)

- die phänomenologisch begründete Wissenssoziologie
 - Berger, Peter L./ Luckmann, Thomas (1987 [1969]): *Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*, Frankfurt am Main
 - Schütz, Alfred/Luckmann, Thomas (1979, 1984): *Strukturen der Lebenswelt*, 2 Bde., Frankfurt am Main
- auf SPRECHEN bezogen
 - Wittgenstein, Ludwig (1971 [1952]): *Philosophische Untersuchungen*, Frankfurt am Main

crowdsourcing als Königsweg zum SPRECHERwissen

- Datenerhebung auf zwei Ebenen des Wissens
 - ▶ prozedurales Wissen: Daten produzieren können (Produktionsdaten)
 - ▶ deklaratives Wissen: perzipierte Daten kommentieren können (Perzeptionsdaten)

crowdsourcing

- löst grundsätzliche Probleme der sprachwissenschaftlichen Empirie
- ein Beispiel
 - Matching von Ontologie und Bezeichnung

Gliederung

1. Sprache: *crowdproduced and crowdproducing*
2. emisches *crowdsourcing* in der Sprachwissenschaft
3. Beispiel: Matching von Ontologie und Bezeichnung

sprachwissenschaftliche Traditionen der Datenerhebung → ontologisch problematisch

- Beispiel: Variation im Raum (Dialekte)
 - Sprachatlanten
 - Wörterbücher
- Mediale Revolution erlaubt umfassende Thesaurierung
 - aber
 - deckt Inkonsistenzen auf
 - erfordert Kompensation durch Neuerhebung

Steckbrief

- großräumige Stratigraphie des Alpenraums im Spiegel seiner Mehrsprachigkeit
- drei Datenquellen
 - Retrodigitalisierung (Atlanten, Wörterbücher)
 - Transfer aus aktuellen Projekten
 - Neuerhebung durch *crowdsourcing*
- georeferenzierte online-Kartierung

Sprachgeographische Erschließung des Alpenraums



- für Österreich (außer Tirol): Wörterbücher
- bereits publizierte Atlanten

Atlas linguistique et ethnographique de la Provence





Internet-Plattform

sprachliche &
ethnographische
Dokumentation

- Gelände
- Atmosphärisches
- Flora
- Fauna
- Brauereien
- Siedlungen
- Architektur
- Wirtschaft
- Wintersport

Datenquellen

1. Retrodigitalisierung
2. Transfer aus
anderen Projekten
3. Neuerhebung durch
crowdsourcing

wissenschaftliche
Interpretation

- Romanisierung
- Germanisierung
- Slavisierung
- Fragmentierung
- Kolonisation (Walser)
- Staatsbildung &
Nationalisierung

LA CASCINA DI MONTAGNA SENNHÜTTE-CHALET

26,2 = 18,7 = 0

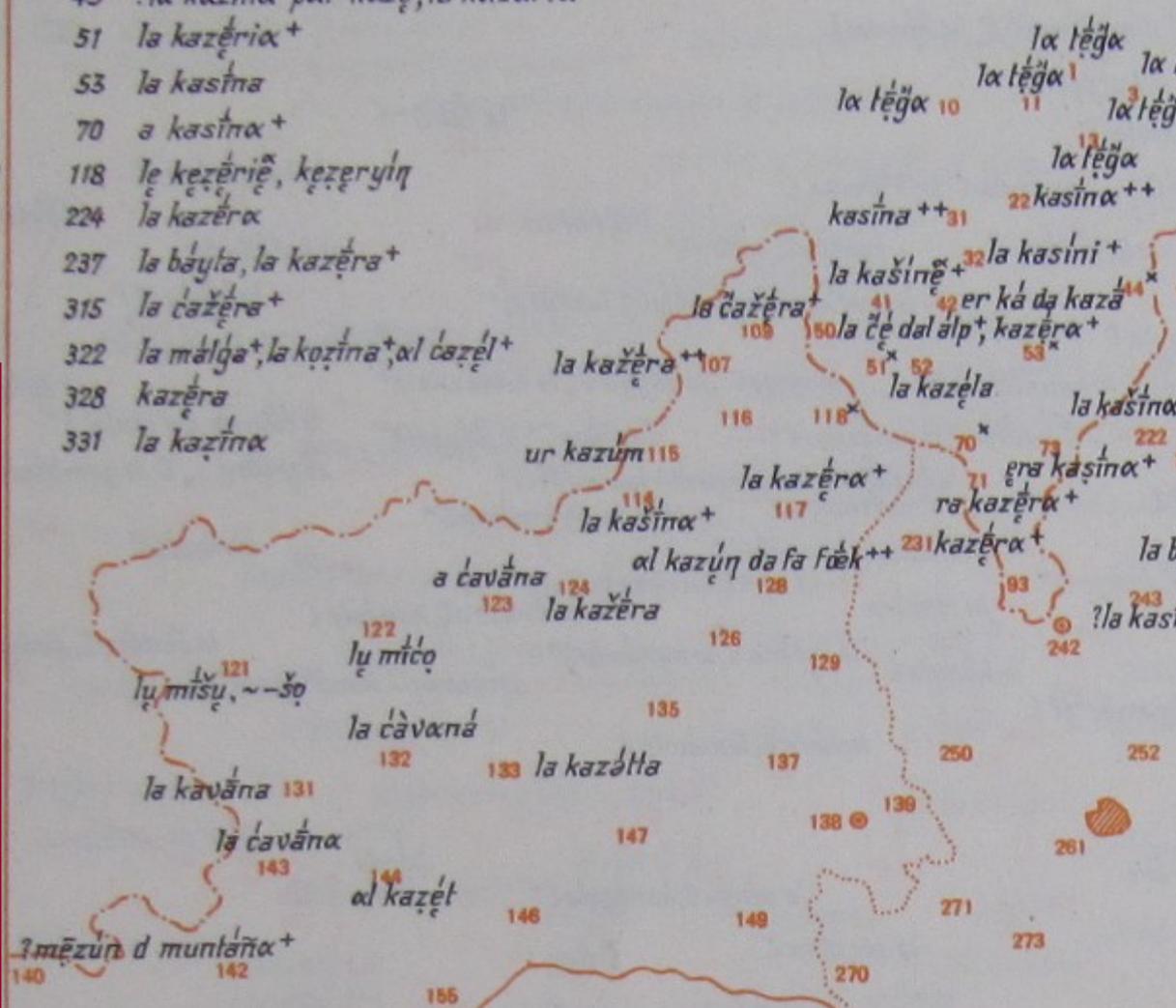
Legende:

Sachliches

Ausnutzung des Graswuchses auf der Alpweide durch das auf die Alp getriebene und unter der Hut von Hirten weidende Vieh sowie Verwertung der gewonnenen Milch an Ort und Stelle ist Zweck des alpwirtschaftlichen Betriebes. In den Alpen herrschen zwei Betriebsformen: 1) Einzelalpung und 2) Genossenschaftliche Betriebe.

Da, wo Einzelalpung besteht, besorgt jeder Alpgenosse selbst sämtliche Arbeiten. Besitzer eines eigenen auf Gemeindeboden stehenden Häuschens, melkt er seine Kühe, deren Milch er selbst weiter verarbeitet. Solche Einzelbetriebe herrschen noch vor in den Westalpen, im Ossolagebiet, im Velt-

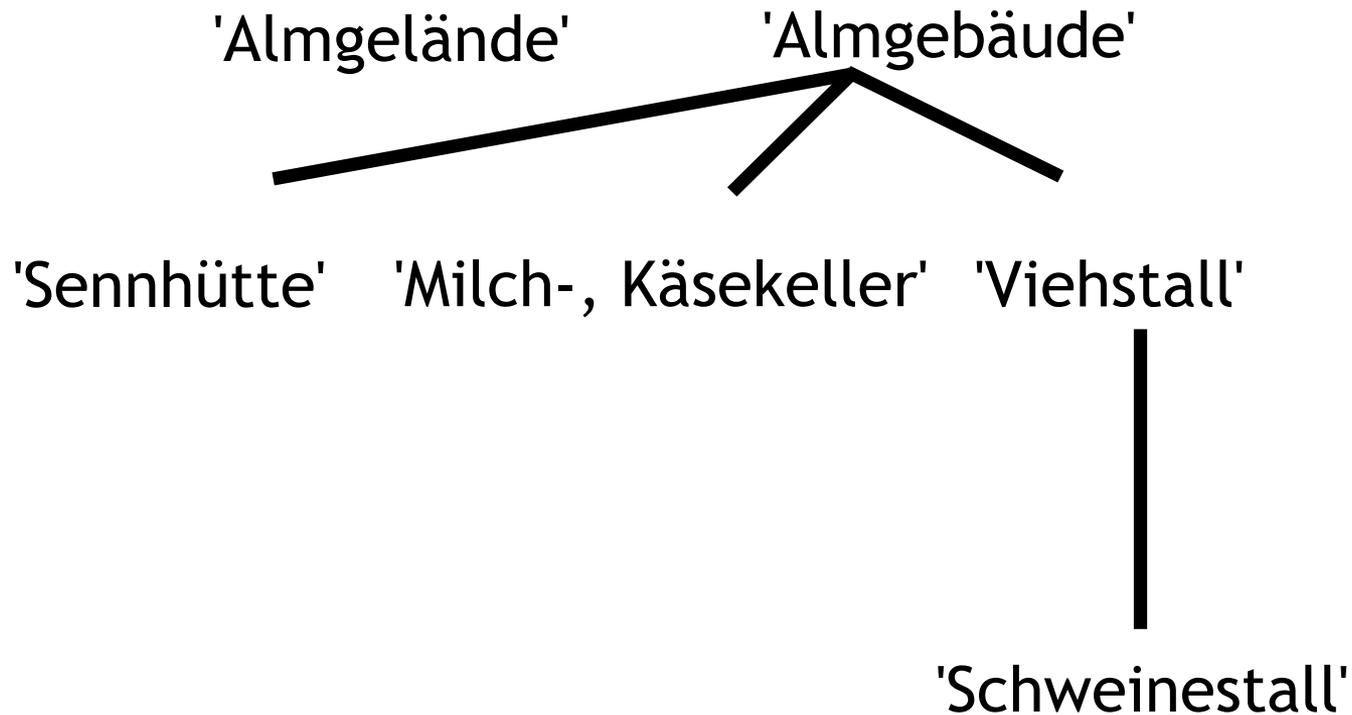
- 15 la tığa
- 17 la tiza
- 44 la kasina
- 45 : la kazina par kazê, la kazaria+
- 51 la kazeria+
- 53 la kasina
- 70 a kasina+
- 118 le kezëriê, kezëryî
- 224 la kazera
- 237 la bayta, la kazera+
- 315 la cazera+
- 322 la malga+, la kazina+, al kazêl+
- 328 kazera
- 331 la kazina



ontologisch relevante Stichwörter der Legende zu AIS 1192

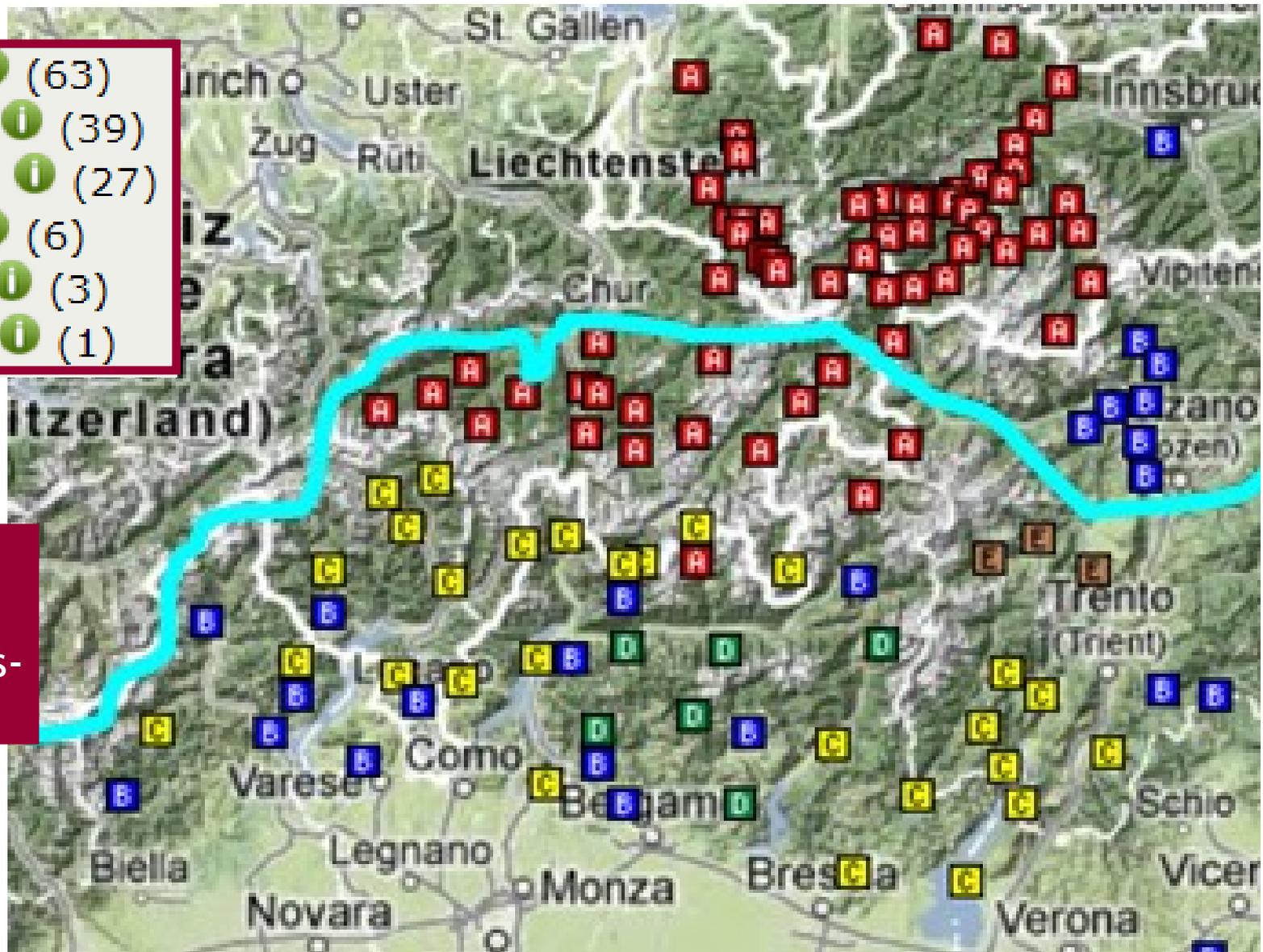
"Ausnutzung des Graswuchses auf der Alpweide ...
unter der Hut von Hirten weidendes Vieh ...
Verwertung der gewonnenen Milch an Ort und Stelle ...
Zweck des alpwirtschaftlichen Betriebs ...
Einzelalpfung ... genossenschaftliche Betriebe... Alpgenosse
... Häuschen ... melkt ... Kühe ... Milch ... meist massive
steinerne Sennhütte ... Käser ('Senn') ... Hilfspersonal ...
Käse ... Butter ... Zieger ... Milch-, Käsekeller ... Alpstall ...
Ziegermilch ... Herde ... Äplerfamilie ... Männer ...
Schutzhütte für das Vieh ... Schweinestall ... Milch-,
Käsekeller meist tiefliegend oder angebaut oder in einem
separaten Häuschen ... mit laufendem Wasser oder an
laufende Gewässer angebaut ..."

ontologischer Ausschnitt: Almgebäude



- A** => tegia **i** (63)
- B** => casara **i** (39)
- C** => cascina **i** (27)
- D** => baita **i** (6)
- E** => malga **i** (3)
- F** => mayra **i** (1)

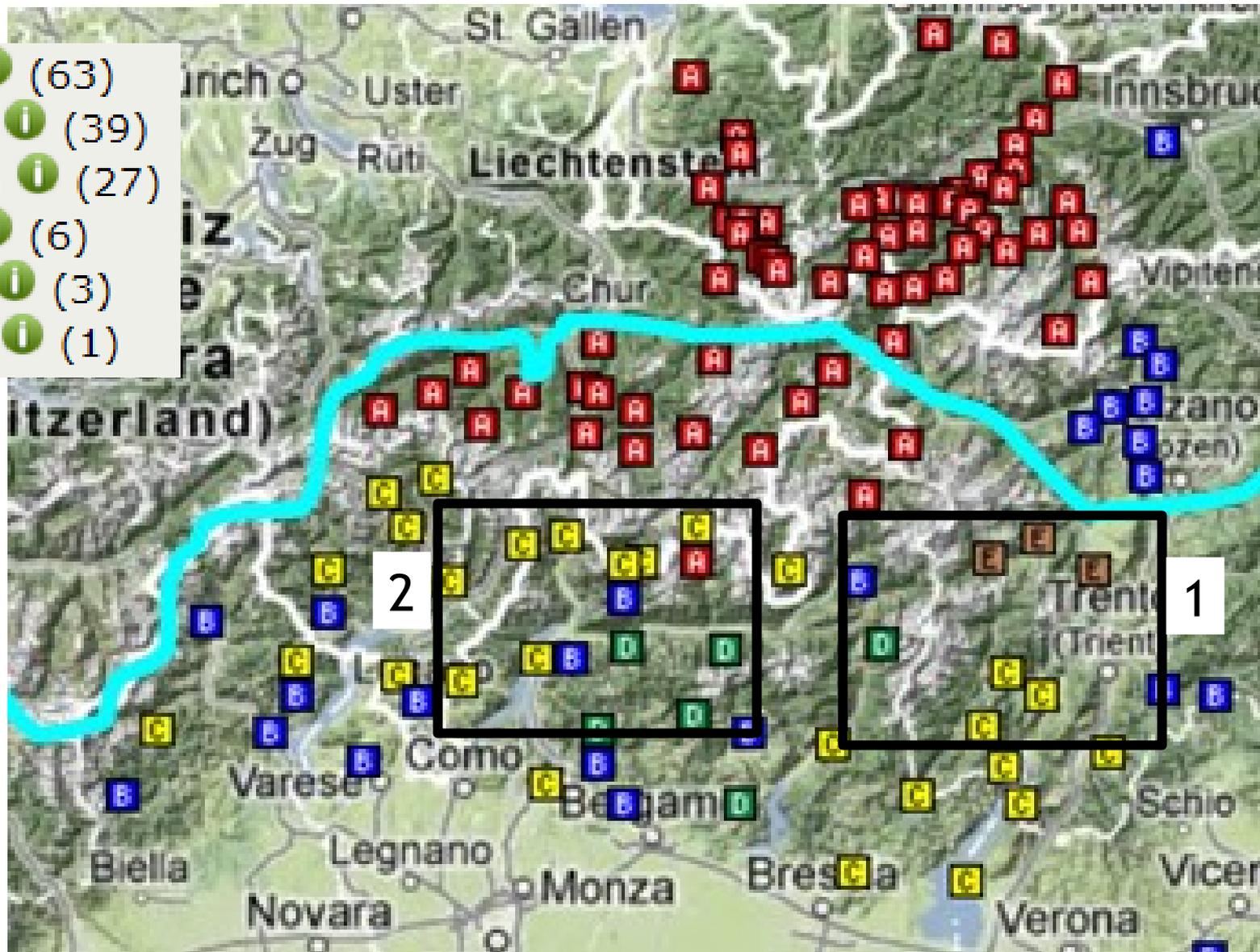
- einige Bez.
- Verbreitungs-
areale



Anreicherung durch Wörterbuchdaten

- Regionale Großwörterbücher: VDSI, DRG, insoweit georeferenzierbar
- Lokale Wörterbücher:
 - 1 Grassi, Corrado (2009): *Dizionario del dialetto di Montagne di Trento, San Michele all'Adige*
 - 2 De Agostini, Ferruccio (in Vorb.): *Vocabulari dal dialét de Gurdúna*
- digitale Wörterbücher:
<http://www.walser-alps.eu/mundart/woerterbuch> (noch sehr unvollständig, gerade im Blick auf Alpenwörter)

- A => tegia i (63)
- B => casara i (39)
- C => cascina i (27)
- D => baita i (6)
- E => malga i (3)
- F => mayra i (1)

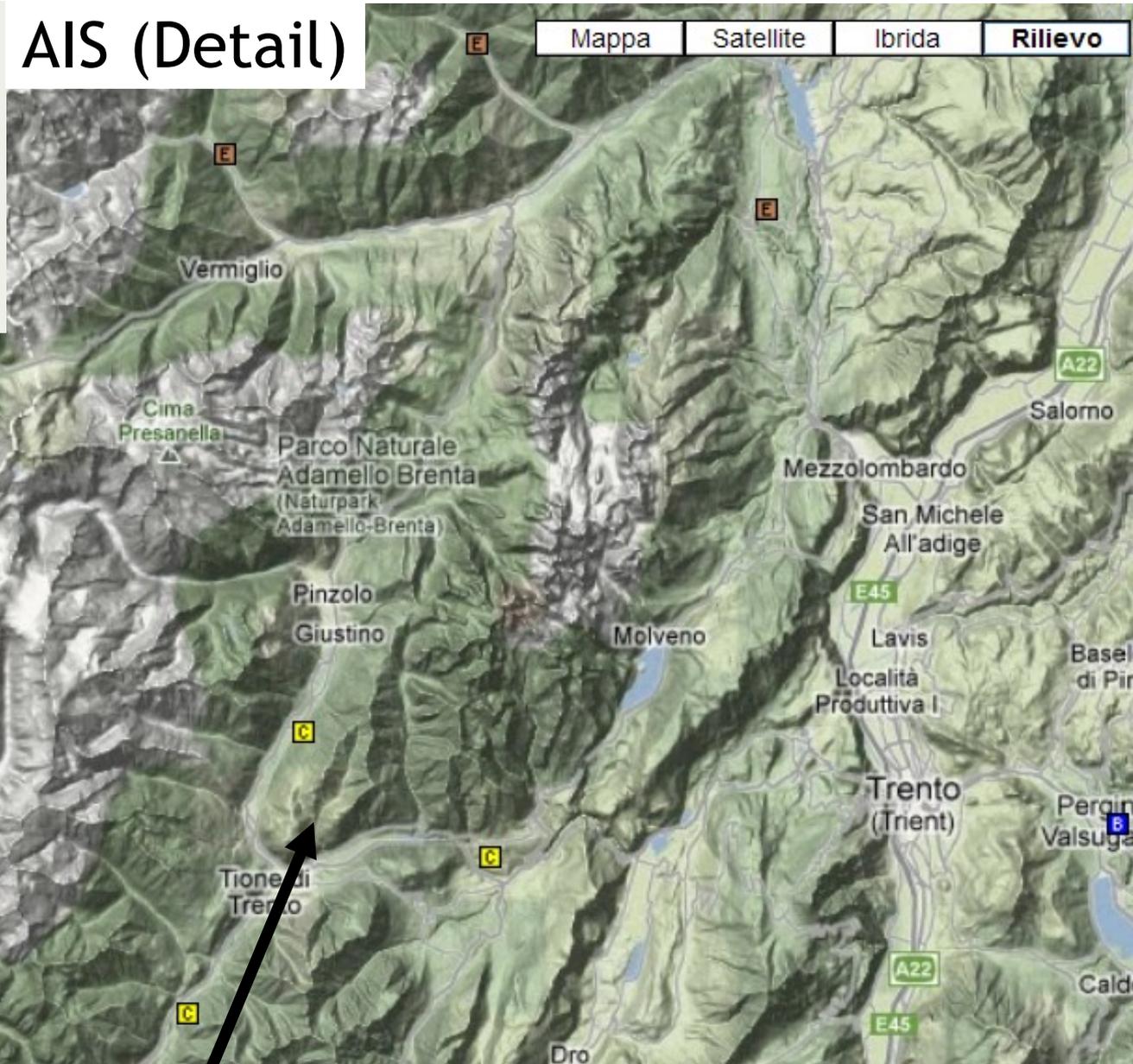


<http://www.adis.gwi.uni-muenchen.de/AIS.php?karte=true>

'Die Sennhütte' - Online-Kombinationskarte: Daten AIS & VALTS

AIS (Detail)

- A** => tegia **i** (63)
- B** => casara **i** (39)
- C** => cascina **i** (27)
- D** => baita **i** (6)
- E** => malga **i** (3)
- F** => mayra **i** (1)



Montagne di Trento



B

C

"**cašina**: sf. - (alp.) casera: *la cašina endó che stava i vachér e i lavorava el lat* (edificio della malga) dove stanno i malgari e lavoravano il latte" (Grassi 2009, 90)

→ dt. 'Sennütte'

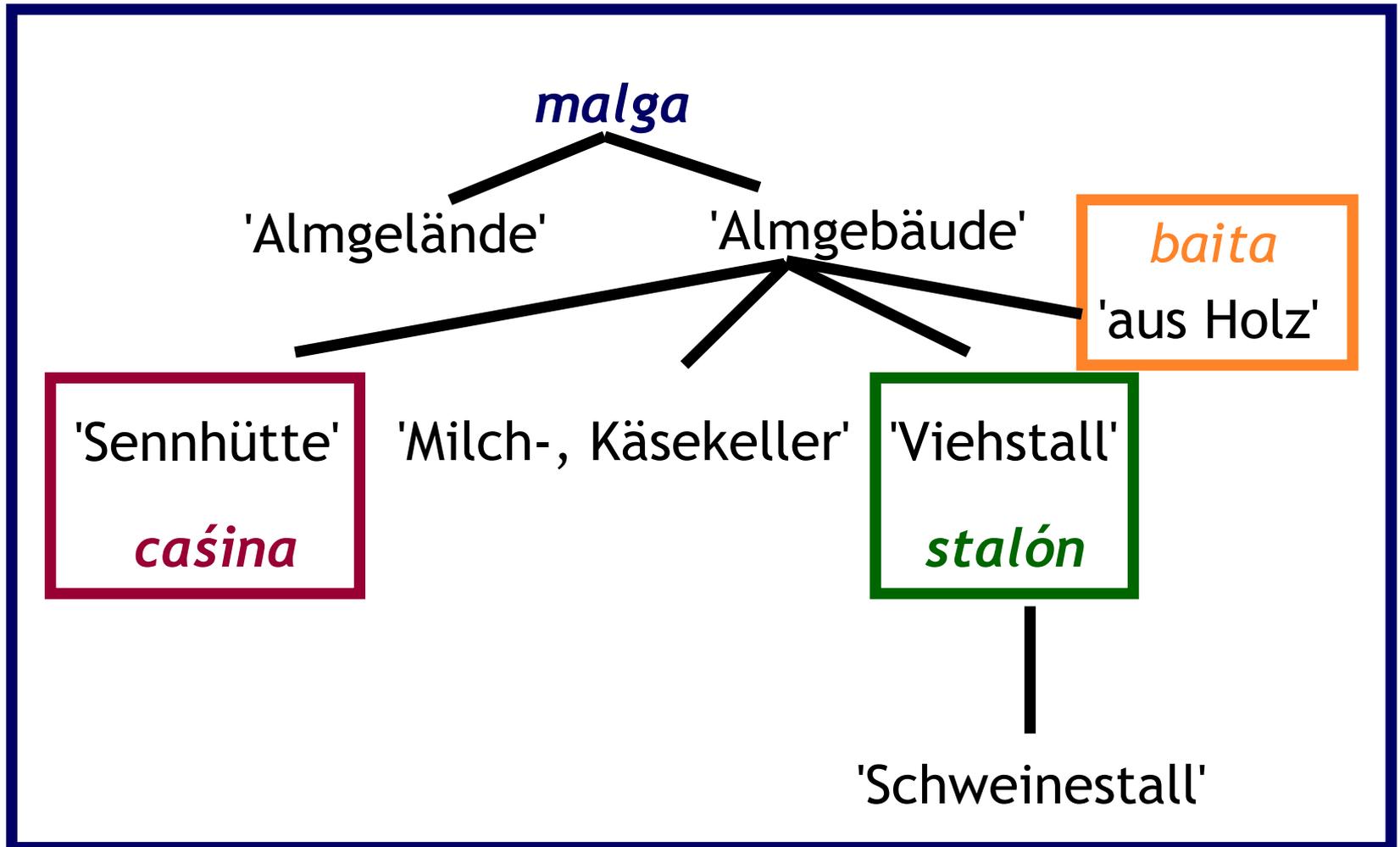
D

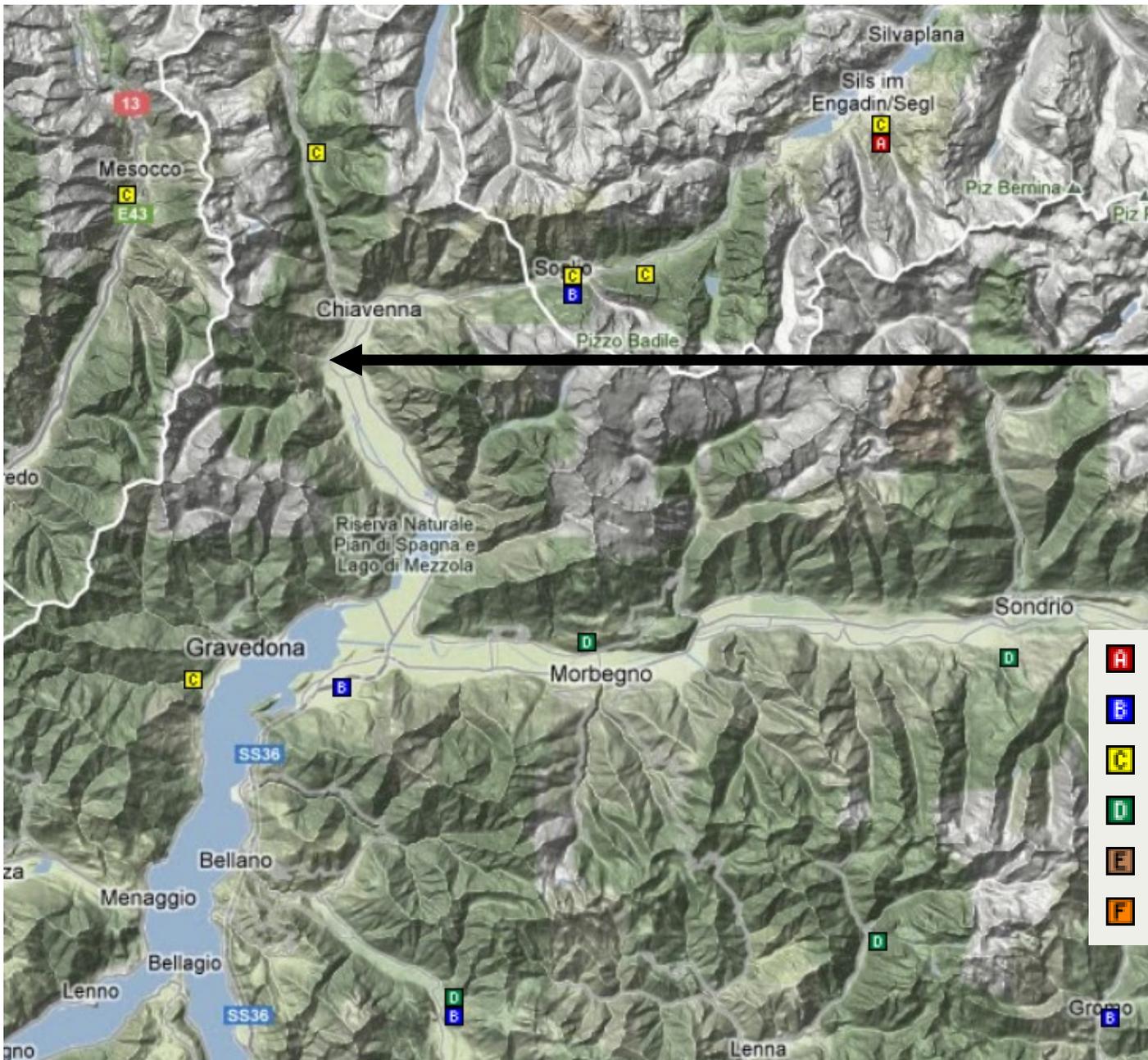
"**baita**: sf. - 1. capanna di montagna [...] 2. capanna solitamente di legno, anche semplice baracca, adibita a legnaia [...]" (Grassi 2009, 29)

"**malga**: sf. - (alp.) pascolo di montagna ed edifici di servizio ad esso annessi" (Grassi 2009, 274) - "Poste ad alta quota, spesso al di sopra della vegetazione d'alto fusto, sono costituite da terreno e pascolo, da una *cašina*, edificio dove si lavora il latte e dove trovano riparo gli addetti, e dallo *stalón*, l'edificio destinato ad accogliere il bestiame in caso di cattivo tempo." (Grassi 2009, 274)

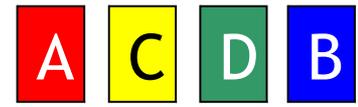
E

→ dt. *Alm* 1. 'Weidegelände im Gebirge', 2. 'Gebäude auf dem Weidegelände'





Gordona



- A** => tegia **i** (63)
- B** => casara **i** (39)
- C** => cascina **i** (27)
- D** => baita **i** (6)
- E** => malga **i** (3)
- F** => mayra **i** (1)

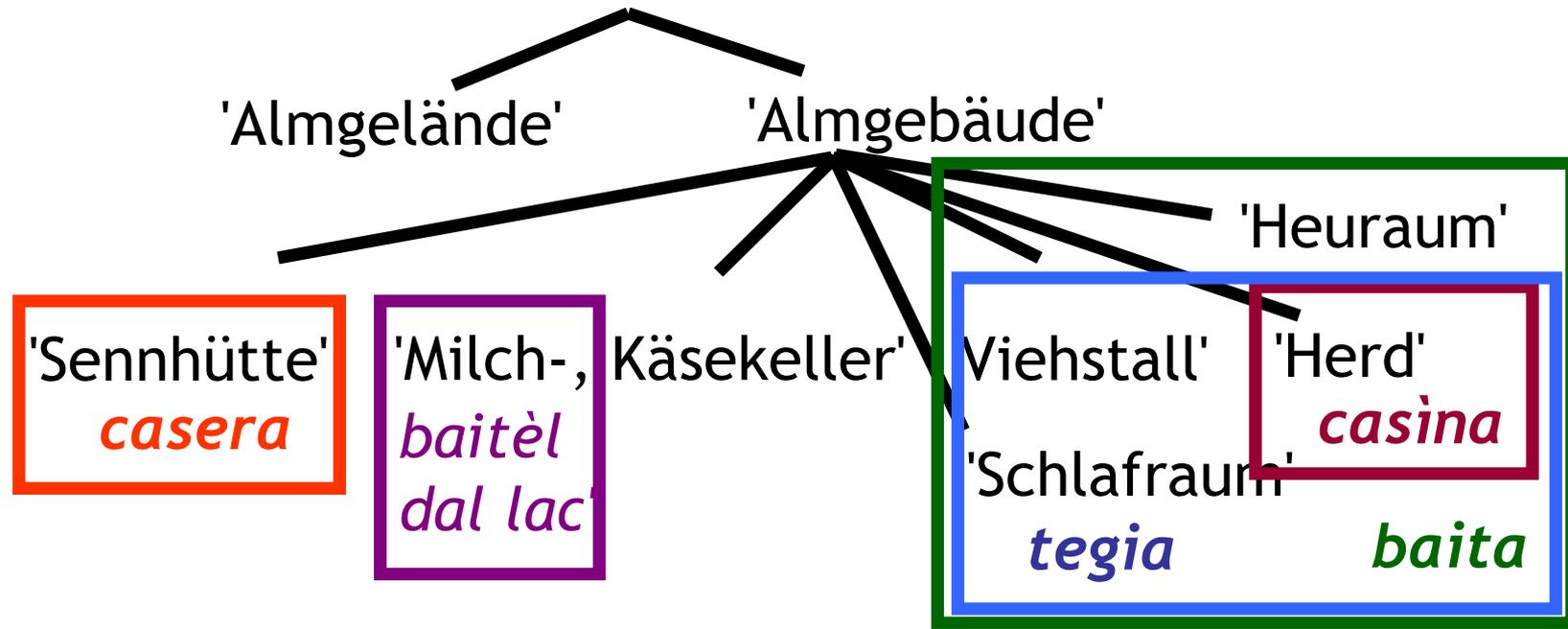
C "casìna (casìn) - sf. piccola baita. Sui monti era un locale, accanto ad una baita più grande che serviva da stalla e fienile, adibito a cucina." (De Agostini in prep., 18)

D "bàita (bàit) - sf. baita, casupola di montagna Il dim. *baitèl* sm. In genere era formata da un piano seminterrato con entrata a valle, che serviva da stalla, e da un vano superiore, con entrata laterale, che serviva da fienile, da camera da letto e, spesso, anche da angolo da cucina. Il *baitèl dal lac'* era invece una modesta costruzione sui monti, come un piccolo casolare, ad un solo piano, a volte seminterrato, dove scorreva un rigagnolo d'acqua in cui si mettevano a bagno *al còrch' de rem* per il latte." (De Agostini in prep., 7)

A "tegia - sf. baita d'alta montagna costituita da abitazione e stalla." (De Agostini in prep., 127)

B "caséra (casér) - sf. latteria, edificio adibito alla raccolta e alla lavorazione del latte." (De Agostini in prep., 18)

E andere Bedeutung (→ 'Herde'):
"malga (màlgh') - sf. mandria, gruppo di bestie al pascolo" (De Agostini in prep., 58)"



Kombination

- von tendenziell überspezifizierten Wörterbuchdaten
 - tendenziell unterspezifizierten Atlasdaten
- führt zu Inkonsistenzen
- *crowdsourcing* als Korrektiv

Interaktive Ontologie aus *crowdsourcing*

➤ Doppelte Bearbeitungsrichtung:

semasiologisch:

- Wort → Sache (Bild zuordnen)

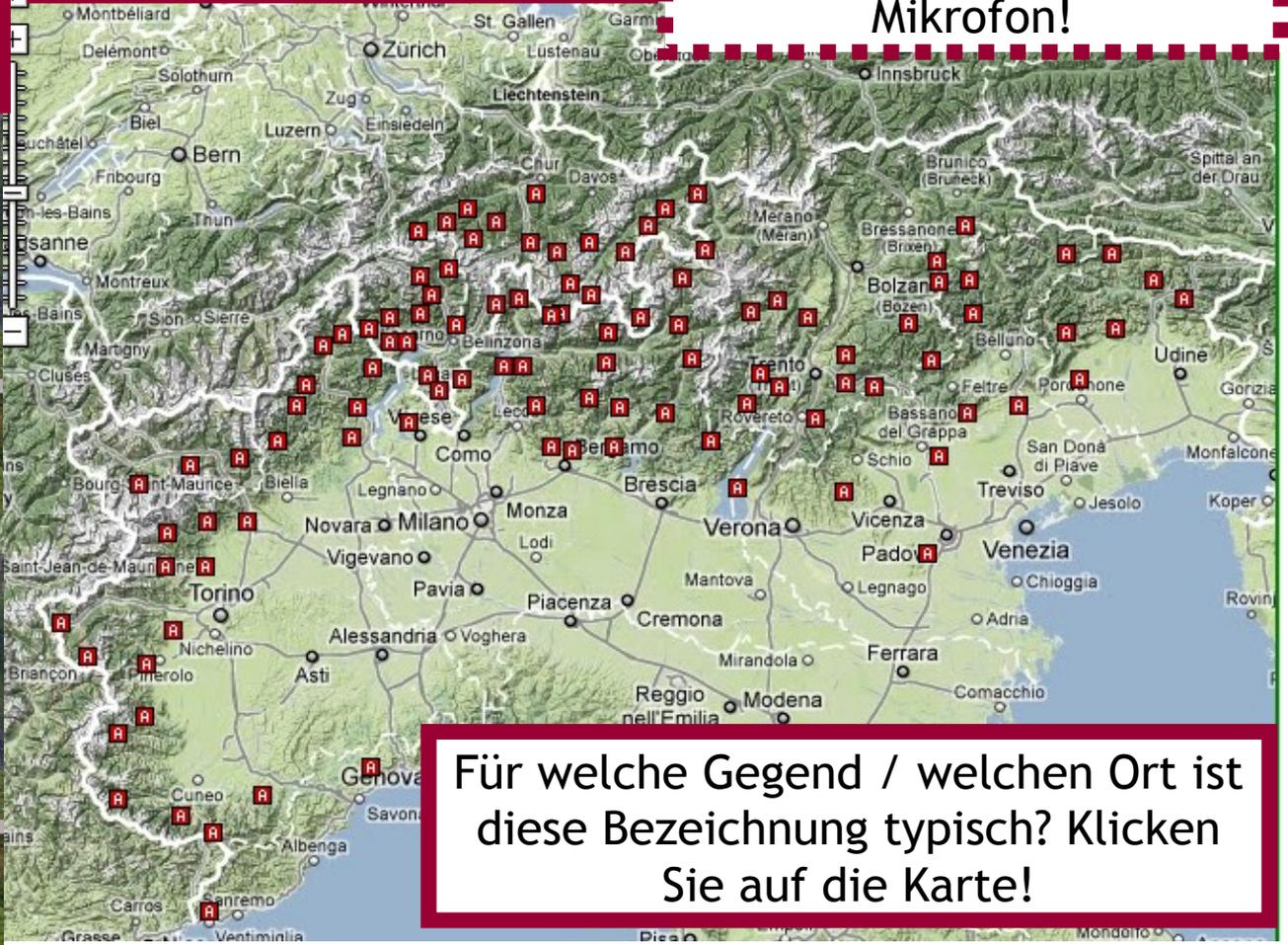
onomasiologisch

- Sache (Bild) → Wort zuordnen



Dies ist ein/e:
...
Bitte schreiben Sie die
Bezeichnung!

Dies ist ein/e:
?
...
Bitte sprechen Sie die
Bezeichnung jetzt in Ihr
Mikrofon!

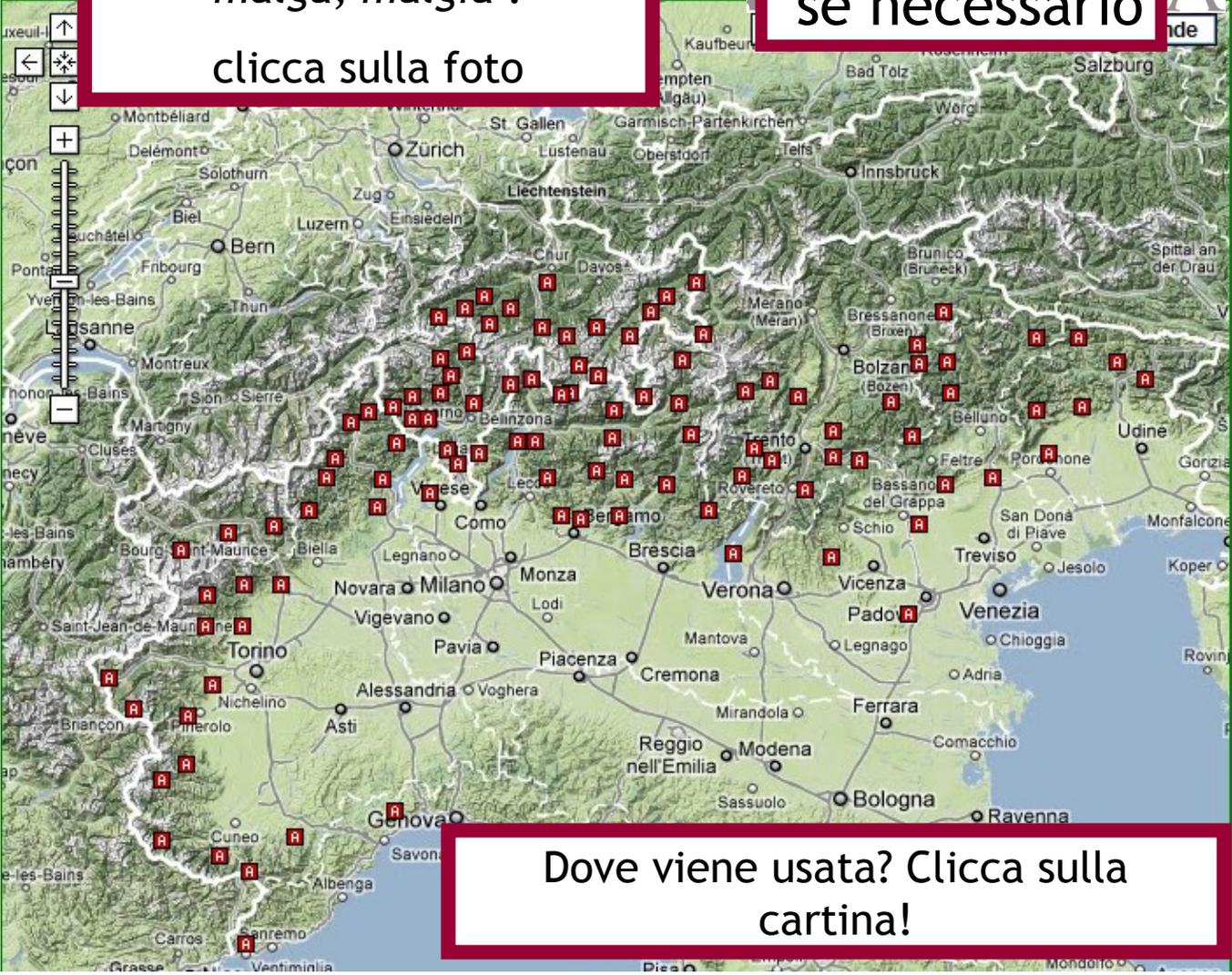


Für welche Gegend / welchen Ort ist
diese Bezeichnung typisch? Klicken
Sie auf die Karte!



Che cosa significa la parola *malga, malgia*?
clicca sulla foto

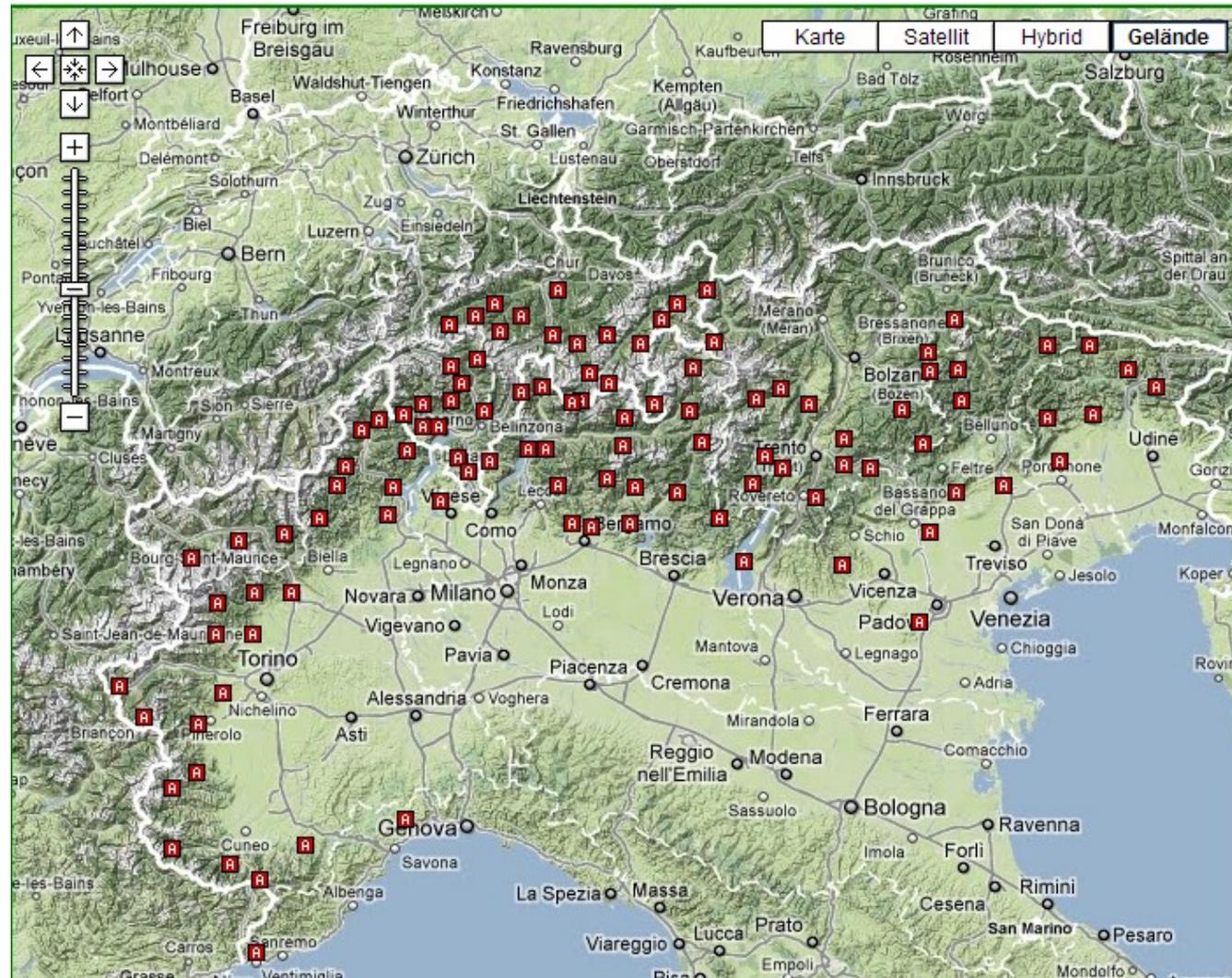
inserisci qui un'altra foto, se necessario



Dove viene usata? Clicca sulla cartina!

Endprodukt: Funktion 1 - Verbreitung von Bezeichnungen

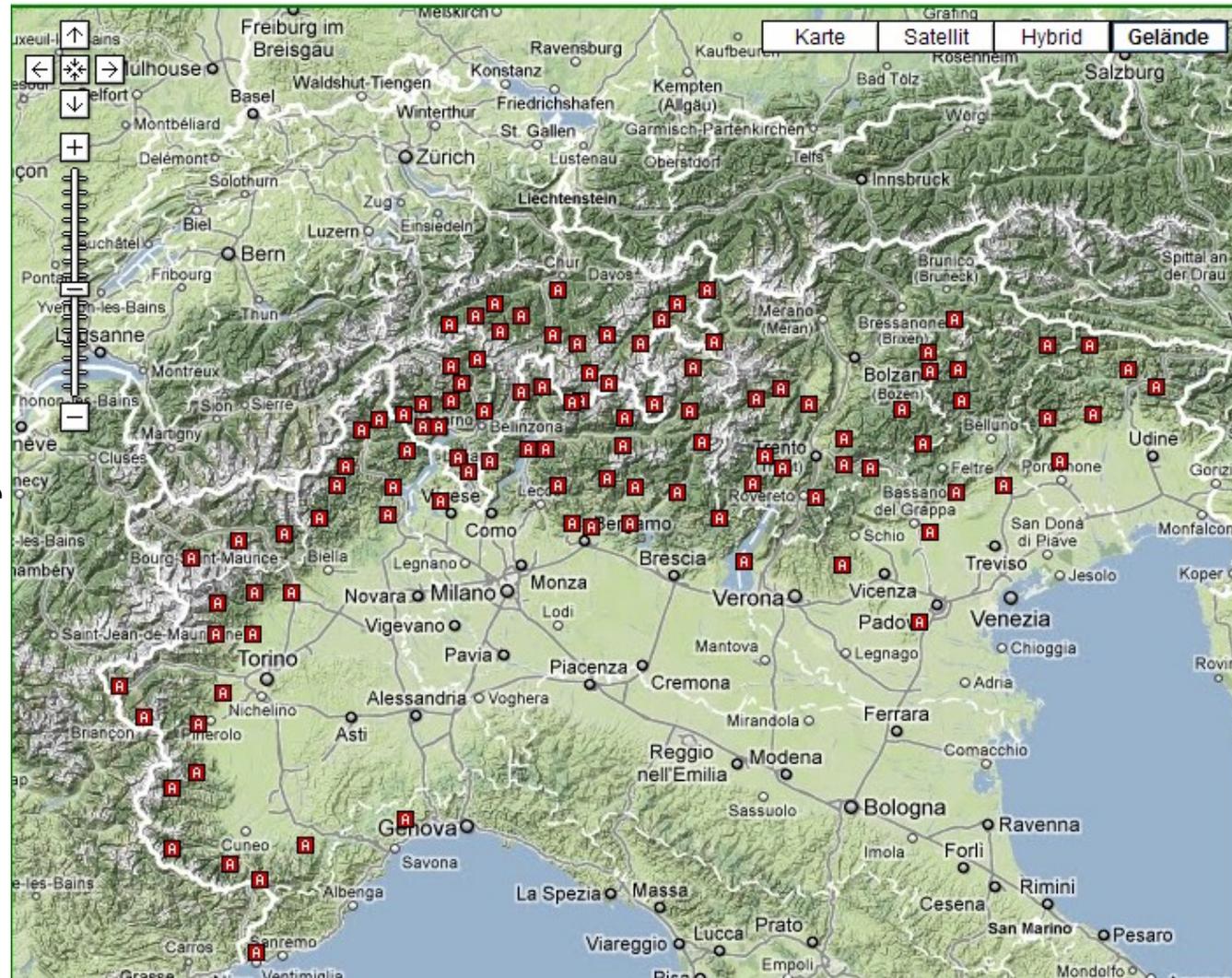
- alpeggio
- malga
- baita
- cascina
- casera
- ...



Endprodukt: Funktion 2 - Verbreitung von Bedeutungen

tipo *casera*

- alpeggio
 - pascolo
 - mandria
 - edificio con le seguenti funzioni
 - lavorazione del latte / formaggio
 - conservazione del latte fresco
 - fienile
 - stalla
 - porcile
 - camera
 - cucina



danke!
Ende