
Markus Kunzmann | Christina Mutter

Dialektdatenerhebung und Internet: Chancen und Grenzen. Ein Werkstattbericht.

<https://www.verba-alpina.gwi.uni-muenchen.de/>

6. IGGD-Kongress, 13. – 15. September 2018, Marburg

Gliederung

- 1) Kurze Projektvorstellung von VerbaAlpina
- 2) Konzeption Crowdsourcing
- 3) Resonanz Crowdsourcing
- 4) Einblick in die Crowdsourcing-Daten
 - a) Phonetik/Phonologie
 - b) Lexik
- 5) Fazit



Projektvorstellung

- *VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit*
- Förderung durch die DFG seit 2014 als Langfristvorhaben
- 2. Laufzeit seit 11/2017-11/2020 (Perspektive bis 2025)
- Untersuchung des ethnographisch und topographisch weitgehend homogenen, sprachlich aber stark heterogenen Alpenraums: D, A, CH, I, F, FL, SLO, MC



Projektvorstellung

Projektjahre

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|

Kalenderjahre

| | | | | | | | | | |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
| i, ii, iii, iv |

Quartale

| | | | |
|--------------|--|---|--|
| Projektphase | I | II | III |
| Schwerpunkt | <p>Kultur</p> <ul style="list-style-type: none"> • Almwirtschaft • volkstümliche Medizin • traditionelle Küche | <p>Natur</p> <ul style="list-style-type: none"> • Landschaftsformen • Wetter • Fauna • Flora | <p>moderne Lebenswelt</p> <ul style="list-style-type: none"> • Ökologie • Tourismus |



Projektvorstellung

- Kombination von (Geo-)Linguistik mit *Digital Humanities*
- Datenquellen:
 - Analoge Daten aus Sprachatlanten und Wörterbüchern (z.B. SDS, BWB,...)
 - Digitale Daten aus Partnerprojekten (z.B. BayDat)
 - Datenerhebung mittels Crowdsourcing



Projektvorstellung

Nachteil des Sprachmaterials aus Sprachatlanten + Wörterbüchern:

Inhomogener Datenbestand



Ausgleich und Ergänzung der Daten durch Crowdsourcing



Chancen + Grenzen



Konzeption Crowdsourcing

Zielgruppenorientierte Akquirierung von Sprechern von Alpensprachen (Mailingaktion Februar-Mai 2017)

Phase I: Schwerpunkt Almwirtschaft/Milchverarbeitung

Kategorien

Gewerbe

- Käsereien
- Molkereien
- Almen

Kultur

- Vereine (z.B. Almwirtschaftlicher Verein Oberbayern (AVO), Alpwirtschaftlicher Verein im Allgäu e.V. (AVA))
- Museen (z.B. Museum Glentleiten)

Ausbildung

- Berufsschulen und sonstige Ausbildungsstätten

Portale

- Fachportale der Milchbranche (z.B. z'alp.ch)

Medien

- Zeitschriften aus dem Bereich Almwirtschaft/Milchverarbeitung (z.B. „Deutsche Molkereizeitung“, „Der Almbauer“)
- (Online-) Zeitungen
- Fernsehen
- Radio
- Verteiler
- Social Media: Facebook + Twitter

Private Kontakte



Konzeption Crowdsourcing



VERBA ALPINA
Sprich die Sprache der Alpen!

DU KOMMST AUS DEN ALPEN?
DU SPRICHST EINEN ALPENDIALEKT?

HILF DER WISSENSCHAFT!

Mach mit beim größten alpinen Sprachprojekt auf
www.lmu.de/verbaalpina

DFG
VERBA ALPINA

LMU
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN


© VERBAALPINA

VERBAALPINA
Ludwig-Maximilians-Universität
Hauspostfach 152
Geschwister-Scholl-Platz 1
80539 München



VERBA ALPINA
Parlez la langue des Alpes!

TU VIENS DES ALPES?
TU PARLES UN DIALECTE ALPIN?

AIDE LA SCIENCE!

Participe au plus grand projet linguistique des Alpes sur
www.lmu.de/verbaalpina

DFG
VERBA ALPINA

LMU


© VERBAALPINA

VERBAALPINA
Ludwig-Maximilians-Universität
Hauspostfach 152
Geschwister-Scholl-Platz 1
80539 München



VERBA ALPINA
Parlez la langue des Alpes!

VIENI DALLE ALPI?
PARLI UN DIALETTO ALPINO?

AIUTA LA SCIENZA!

Participa al più grande progetto linguistico delle Alpi su
www.lmu.de/verbaalpina

DFG
VERBA ALPINA

LMU


© VERBAALPINA

VERBAALPINA
Ludwig-Maximilians-Universität
Hauspostfach 152
Geschwister-Scholl-Platz 1
80539 München



VERBA ALPINA
Govorite jezik jezika Alpe!

SEIŠ ALP?
GOVORIŠ KAKŠEN ALPSKI DIALEKT?

POMAGAJ ZNANOSTI!

Sodeluj pri največjim alpskem jezikovem projektu na
www.lmu.de/verbaalpina

DFG
VERBA ALPINA

LMU


© VERBAALPINA

VERBAALPINA
Ludwig-Maximilians-Universität
Hauspostfach 152
Geschwister-Scholl-Platz 1
80539 München



Konzeption Crowdsourcing

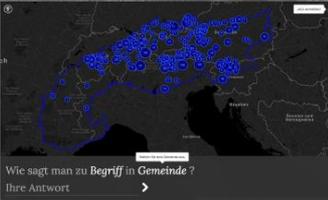
VerbaAlpina Pressetext



VERBA ALPINA
Sprich die Sprache der Alpen!

Das Projekt VerbaAlpina der Ludwig-Maximilians-Universität in München hat sich zum Ziel gesetzt, länderübergreifend den Sprach- und Kulturraum des gesamten Alpengebiets zu untersuchen. Ausgewählte und für diese Gegend charakteristische Dinge sowie die entsprechenden Dialektwörter werden systematisch erfasst und mit modernen informatischen Methoden analysiert und präsentiert. In der aktuellen Projektphase dreht sich dabei alles um den Wortschatz aus dem Bereich der Almwirtschaft, insbesondere der Milchverarbeitung. Das Projekt ist sowohl für Wissenschaftler als auch für interessierte Laien, die einen Alpendialekt sprechen, von Interesse. Der Großteil der Sprachdaten, mit denen das Projekt arbeitet, stammt aus Sprachatlanten der vergangenen hundert Jahre. Um diese Daten durch aktuelle Dialektwörter zu ergänzen, wurde eine groß angelegte Sprecherbefragung (Crowdsourcing) gestartet, bei der das Wissen von Experten auf dem Gebiet der Almwirtschaft und Milchverarbeitung gefragt ist. Mitmachen beim größten alpinen Sprachprojekt kann man unter www.lmu.de/verbaalpina. Die Seite ist Smartphone kompatibel und kann somit problemlos auch von unterwegs aufgerufen werden.

VerbaAlpina wird seit Oktober 2014 – mit einer Perspektive bis 2025 – von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Das Projekt wird in einer Kooperation des Instituts für Romanische Philologie mit der IT-Gruppe Geisteswissenschaften (ITG; LMU Center for Digital Humanities; <http://www.itg.lmu.de/>) durchgeführt und ist eine Verbindung aus Sprachwissenschaft, Volkskunde und Informationstechnologie im Sinne der Digital Humanities. VerbaAlpina wird durch zahlreiche internationale Partner aus dem gesamten Alpenraum unterstützt. Das Untersuchungsgebiet ist länderübergreifend und umfasst die Sprachen Deutsch, Französisch, Italienisch, Slowenisch und Ratoromanisch und ihre jeweiligen Dialekte.



Wie sagt man zu Begriff in Gemeinde?
Ihre Antwort >

Abb. 1: VerbaAlpina - Crowdsourcing Tool

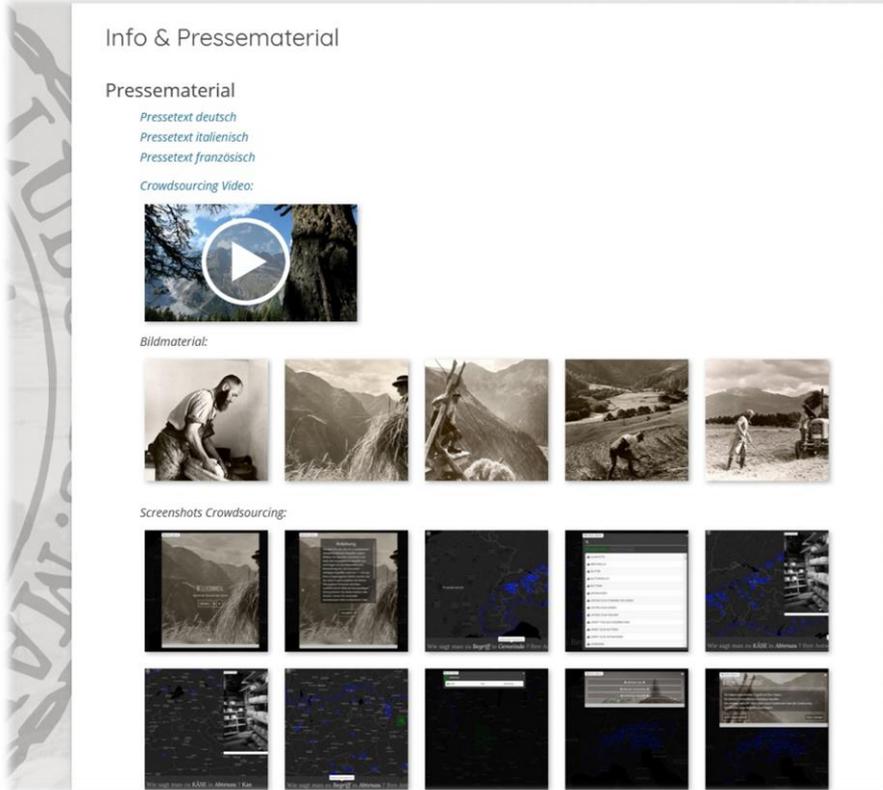
Das Projekt wird von Prof. Dr. Thomas Krefeld und Dr. Stephan Lücke geleitet. Weitere Mitarbeiter des Projekts sind: Annette Antonial (Hilfskraft), David Engelmayer (Wiss. Mitarbeiter Informatik), Monika Hausmann (Hilfskraft), Filip Hristov (Hilfskraft Informatik), Markus Kunzmann (Wiss. Mitarbeiter Germanistik), Christina Mutter (Wiss. Mitarbeiterin Koordination), Alexander Wlatr (Wiss. Mitarbeiter Romanistik) und Florian Zacherl (Wiss. Mitarbeiter Informatik).

Kontakt
Telefon: +49 89 2180 5917
E-Mail: VerbaAlpina@itg.lmu-muenchen.de
<https://www.facebook.com/verbaalpina/>
<https://twitter.com/VerbaAlpina>

VerbaAlpina
Ludwig-Maximilians-Universität Hauspostfach 152
Geschwister-Scholl-Platz 1 80539 München



Info- & Pressematerial auf www.lmu.de/verbaalpina



Info & Pressematerial

Pressematerial

- [Presstext deutsch](#)
- [Presstext italienisch](#)
- [Presstext französisch](#)

Crowdsourcing Video:



Bildmaterial:



Screenshots Crowdsourcing:



Bindung der Crowd

Im CS-Tool

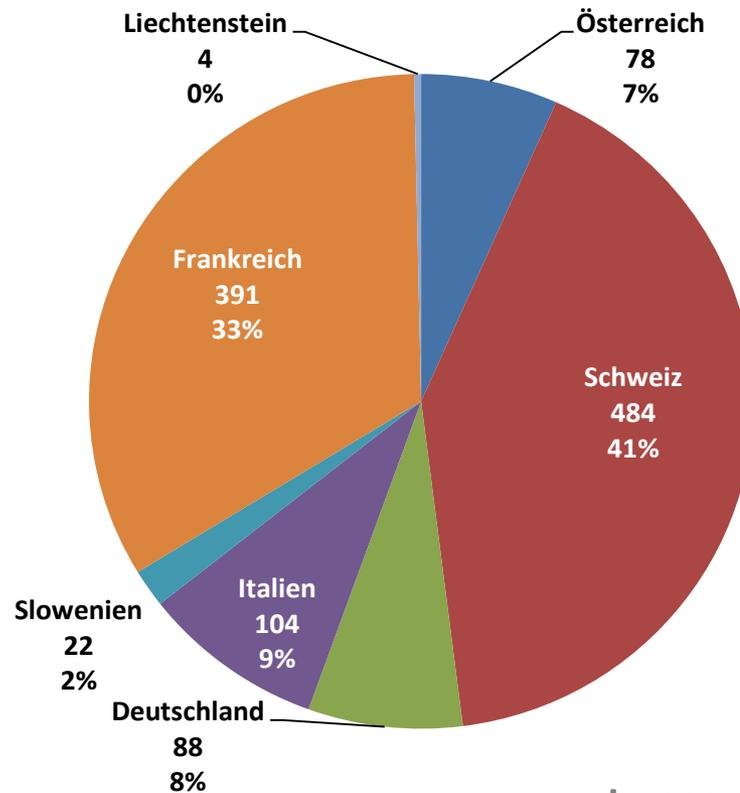
- Möglichkeit zur Registrierung
- Bestenliste: Aktivste User
Aktivste Gemeinde

Auf Social Media

- Posts auf Facebook/Twitter über neue Features und Aktionen

Resonanz Crowdsourcing

Kontaktierte Adressen nach Staaten (in Projektphase I, Februar 2017-Oktober 2017)

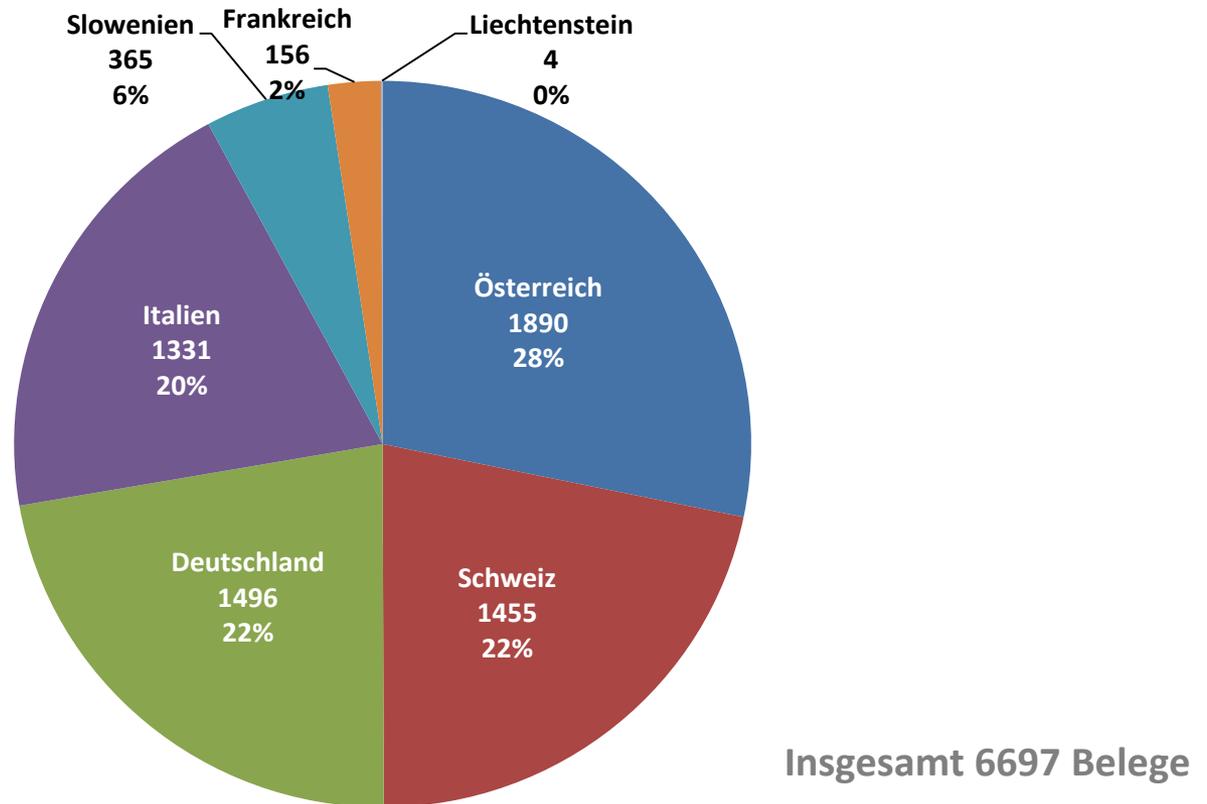


Insgesamt 1171 kontaktierte Adressen



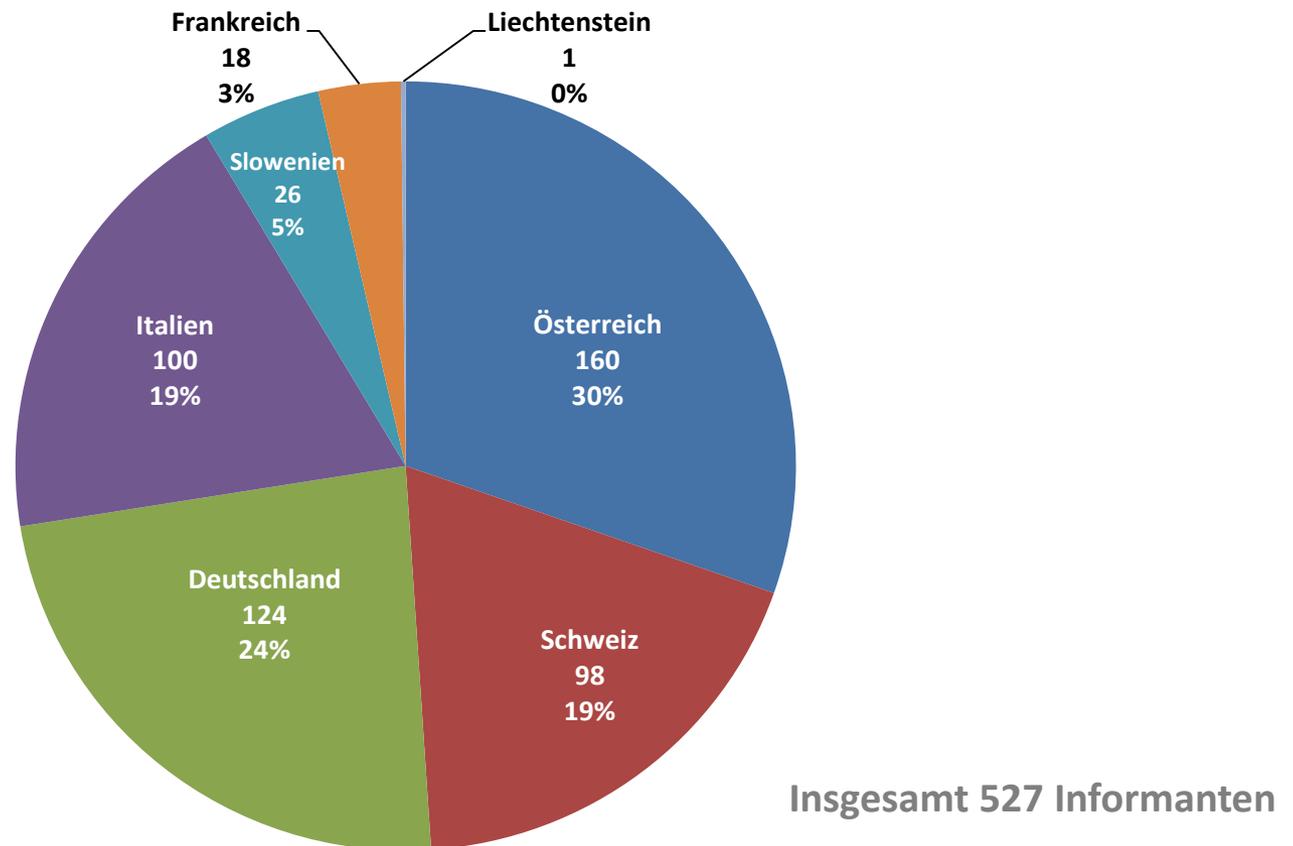
Resonanz Crowdsourcing

Beleganzahl nach Staaten (bis 31.10.17 – Ende Projektphase I)

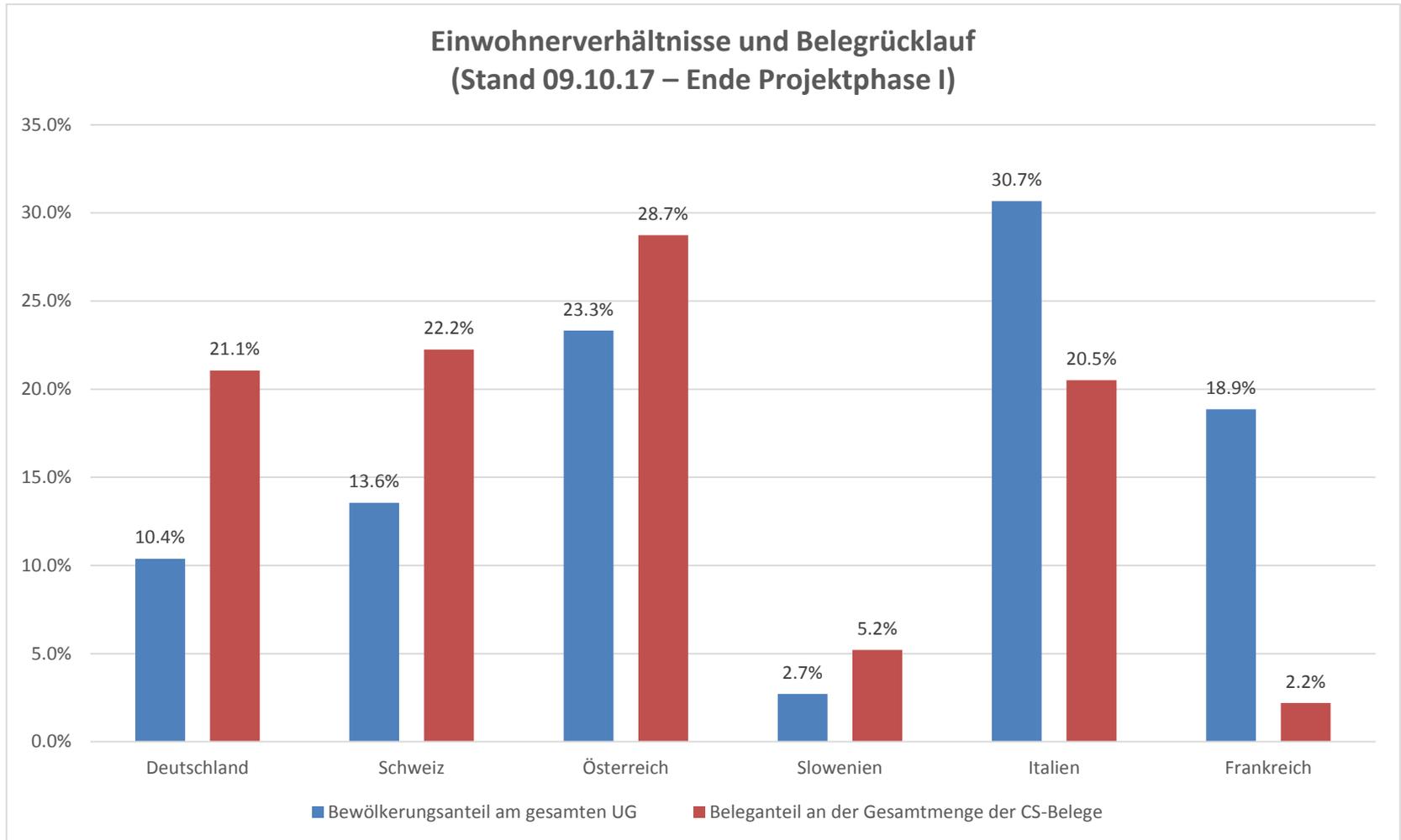


Resonanz Crowdsourcing

Anzahl der Informanten nach Staaten (bis 31.10.17 – Ende Projektphase I)

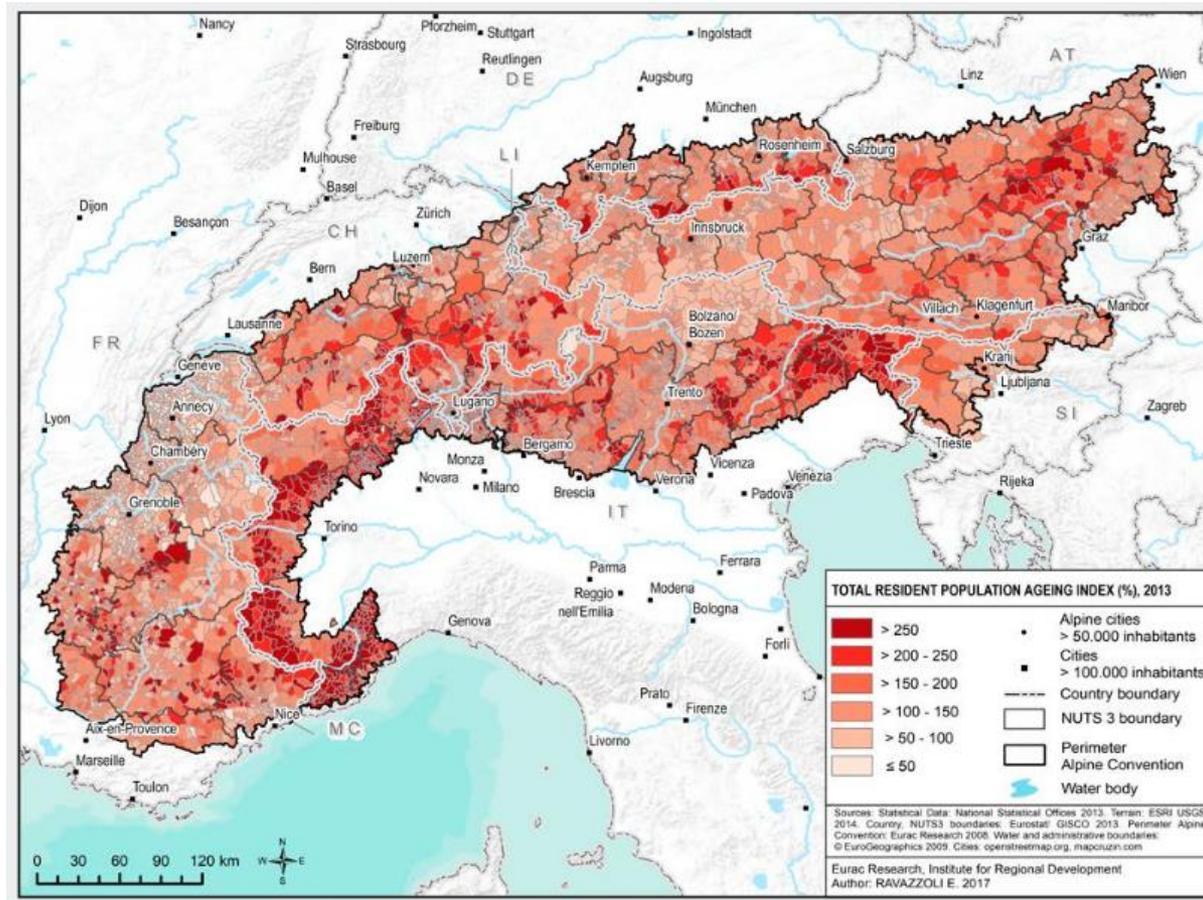


Resonanz Crowdsourcing



Resonanz Crowdsourcing

Bevölkerungsalterung (Einwohner pro km²)

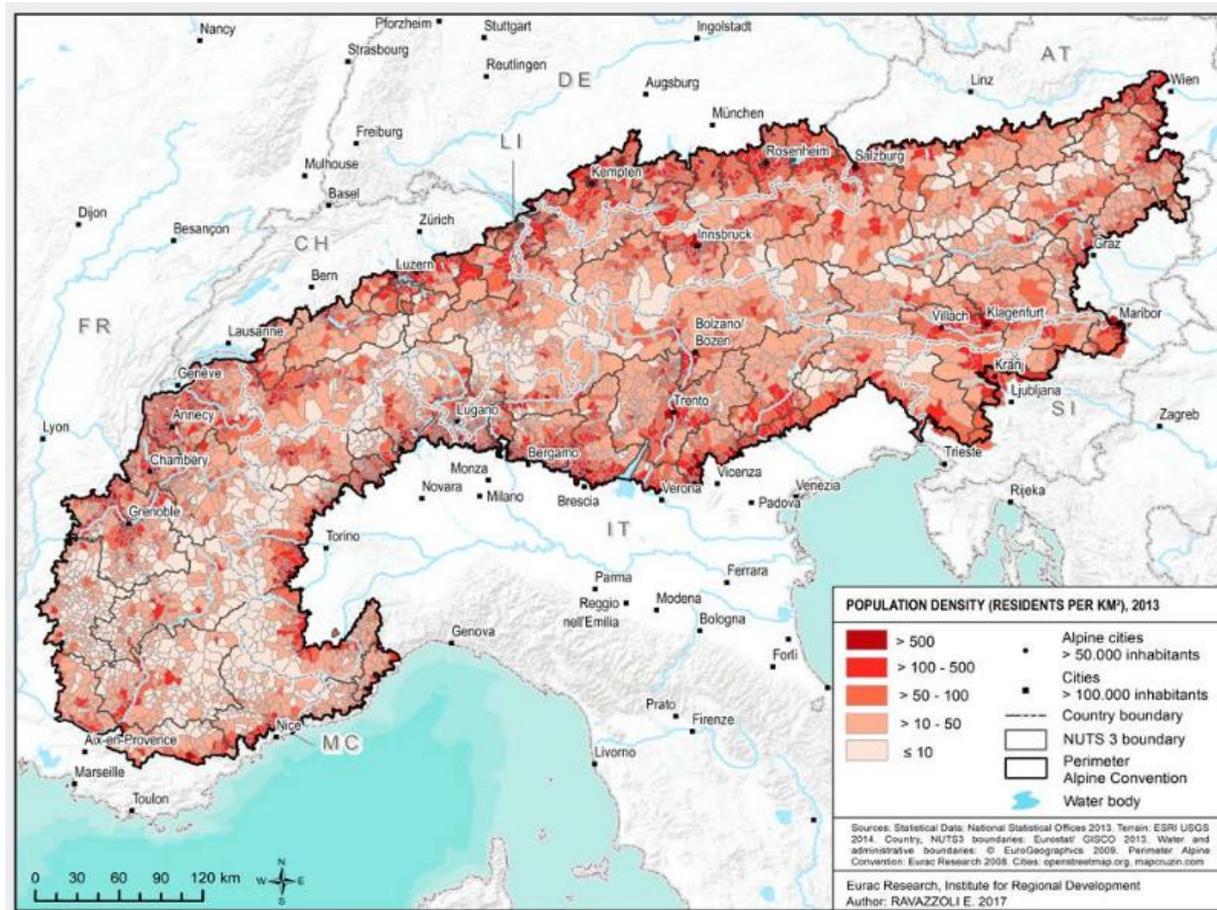


ALPENKONVENTION (2018), S. 15



Resonanz Crowdsourcing

Bevölkerungsdichte (Einwohner pro km²)



ALPENKONVENTION (2018), S. 14



Resonanz Crowdsourcing

CS-Aktionen mit geringer Resonanz

- Vorträge bei Veranstaltungen von Vereinen

Vortrag beim Almbauerntag des Almwirtschaftlichen Vereins Oberbayern,
ca. 40 Personen anwesend



2 deutsche Informanten aus 2 Gemeinden, insgesamt 23 Belege

- Posts auf den Facebook-Seiten entsprechender Gruppen/Vereine

Post in der Facebook-Gruppe „Allevatori italiani“, 40.172 Mitglieder



7 italienische Informanten aus 5 Gemeinden, insgesamt 146 Belege

- Textbeiträge in Printmedien
- Radio-Interviews

Resonanz Crowdsourcing

CS-Aktionen mit der größten Resonanz

- Teilnahme an der Dialekt-Themenwoche im BR

Dialekt-Themenwoche im BR vom 27.04.18-04.05.18



183 deutsche Informanten aus 126 Gemeinden, insgesamt 2082 Belege

- Artikel auf OVB24.de

Textbeitrag auf den News-Portalen von OVB24.de am 19.06.18

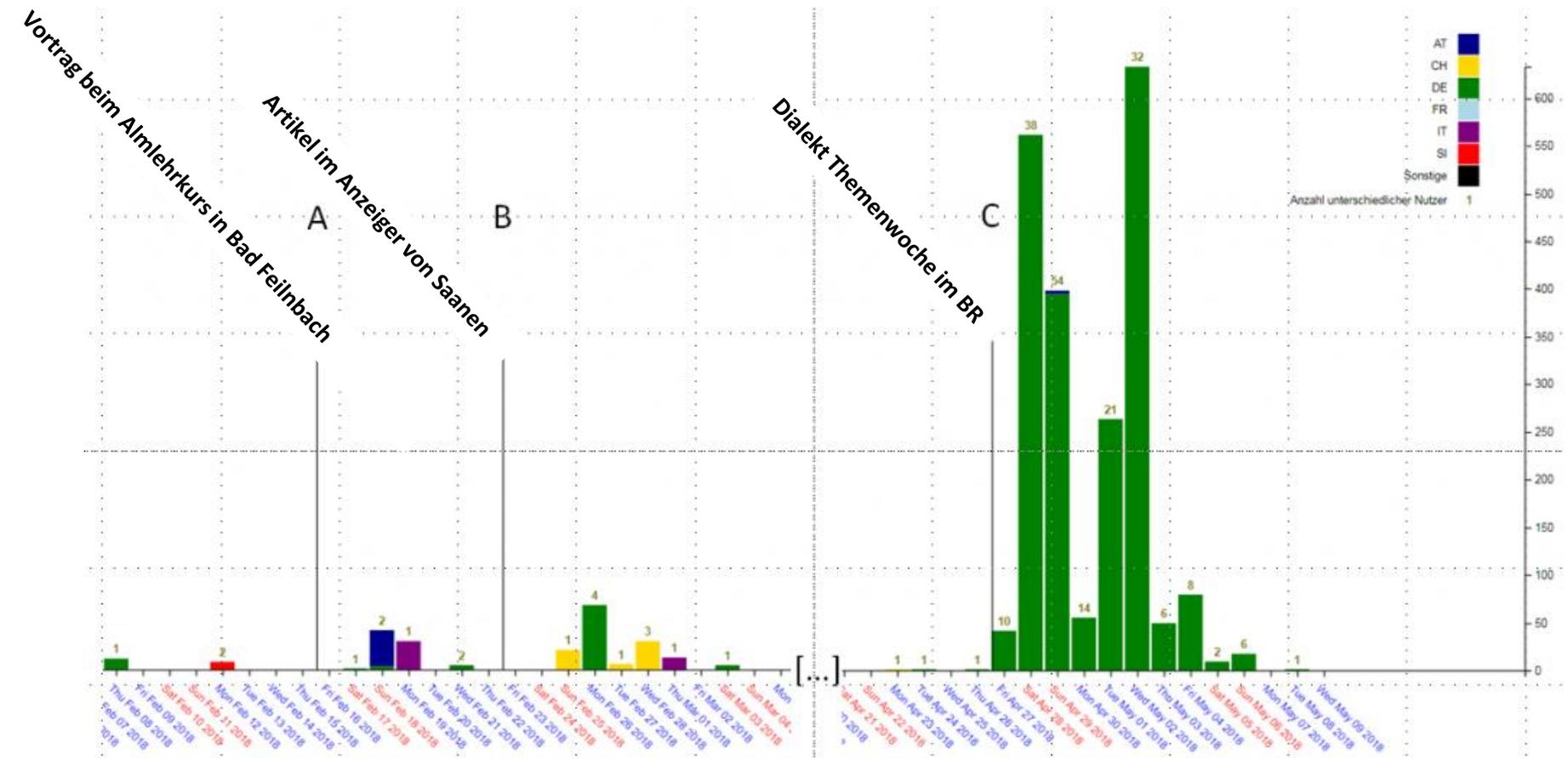


66 deutsche Informanten aus 43 Gemeinden, insgesamt 657 Belege

→ beiden Aktionen ist gemeinsam, dass eine direkte Weiterleitung zum Crowdsourcing erfolgte: **kein Wechsel des Mediums erforderlich**



Resonanz Crowdsourcing



Einblick in die Crowdsourcing-Daten

Welche Daten erhebt VerbaAlpina mittels Crowdsourcing?

a. Sprachdaten

- dialektale Lexik (Stimulus: „Begriff“, z.T. Bild)
- laienverschriftet

b. Metadaten

- Ort (Angabe des Informanten)
- Zeit
- E-Mail-Adresse (nur bei Anmeldung)
- Dialektbezeichnung (Angabe des Informanten)



Welche Daten **fehlen** VerbaAlpina?

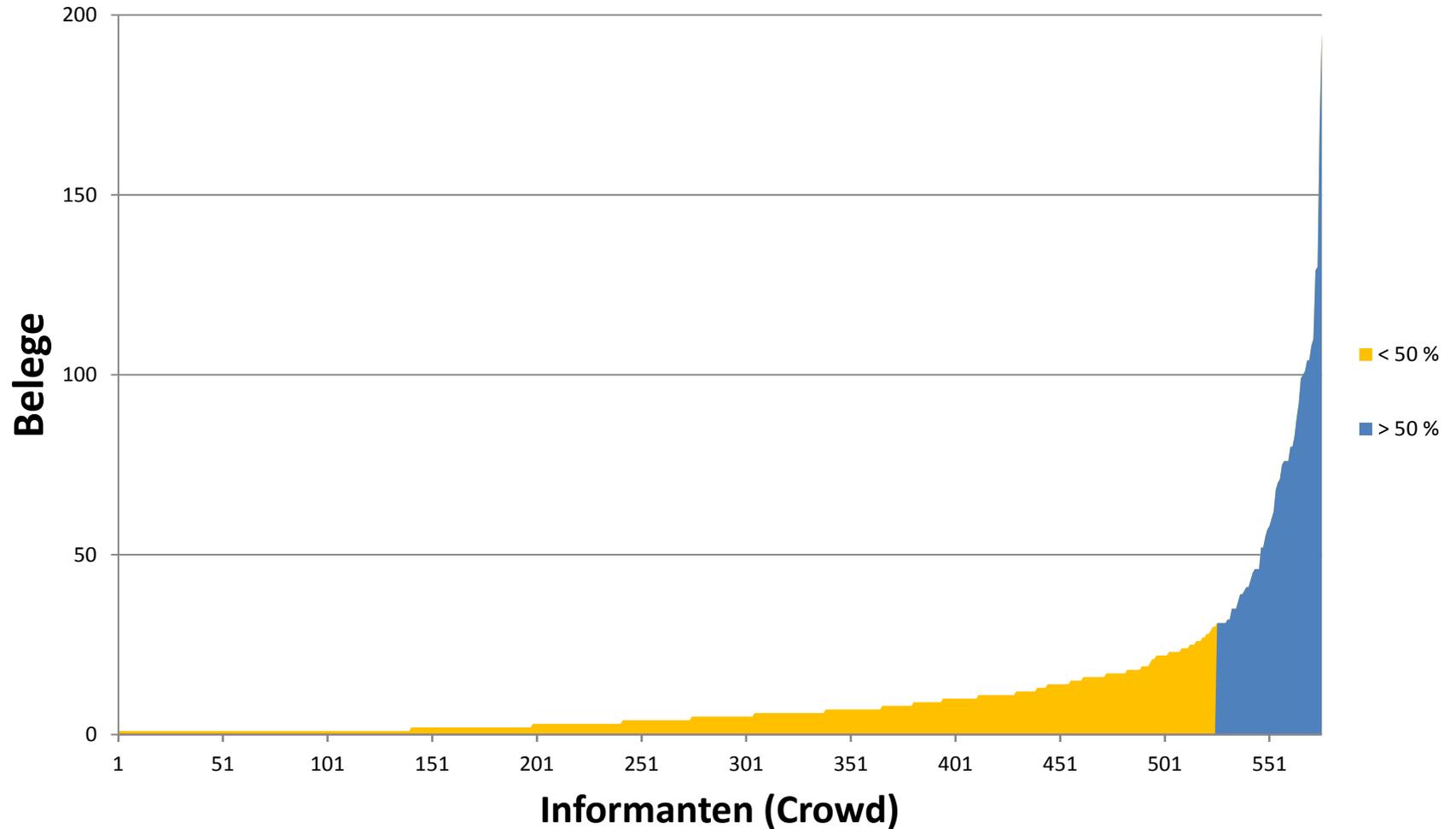
- a. Sprachdaten
 - phonetische Transkription
 - Audioaufnahmen

- b. Metadaten
 - Name
 - Alter
 - Geschlecht
 - Ausbildung
 - ...



Einblick in die Crowdsourcing-Daten

Anzahl der Belege pro Informant (aufsteigend)



Phonetik/Phonologie

Vokalqualität <a>

nhd. *Käse*

[ka:s]

[ka:s]

...

The screenshot shows a linguistic data entry for the word 'Kas' from the location 'Reith im Alpbachtal'. The entry includes phonetic and morpho-lexical information, as well as a source attribution to a crowdsource user.

| Quelle | Konzept |
|--|---------|
| CROWD 1#1 anonymousCrowder_374 (Reith im Alpbachtal) | KÄSE |

Phonetischer Typ (nicht typisiert) VA

Morpho-lexikalischer Typ Käse (ger.) **D** VA

Basistyp caseu(m) **F** **G** VA

Phonetik/Phonologie

Vokalqualität <e>

nhd. *Bett*

[bet]

[bɛt]

...

The screenshot shows a map interface with a pop-up window for the word "Bett". The map background shows the location of Meran, with a red circle 'A' indicating the location. The pop-up window contains the following information:

Bett
(Einzelbeleg)

| | | |
|--------------------------|----------------------|----|
| Phonetischer Typ | (nicht typisiert) | VA |
| Morpho-lexikalischer Typ | Bett (ger.) D | VA |
| Basistyp | *badja- K | VA |

| Quelle | Konzept |
|--|-----------------|
| CROWD 1#1 anonymousCrowder_8 (Meran) | BETT (Wikidata) |

Phonetik/Phonologie

Realisierung -<er>

nhd. *Butter*

[-ɛr]

[-a]

[-e]

...

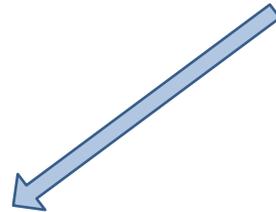
| Token | morph.-lex. Typ | Anzahl |
|--------|-----------------|--------|
| Budder | Butter | 2 |
| Butter | Butter | 39 |

Phonetik/Phonologie

Realisierung -<er>

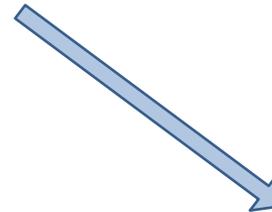
nhd. *Butter*

vokalisiert (-)



| Token | morph.-lex. Typ | Anzahl |
|-------|-----------------|--------|
| Buttr | Butter | 6 |
| Buddr | Butter | 1 |
| Puttr | Butter | 1 |

vokalisiert (+)

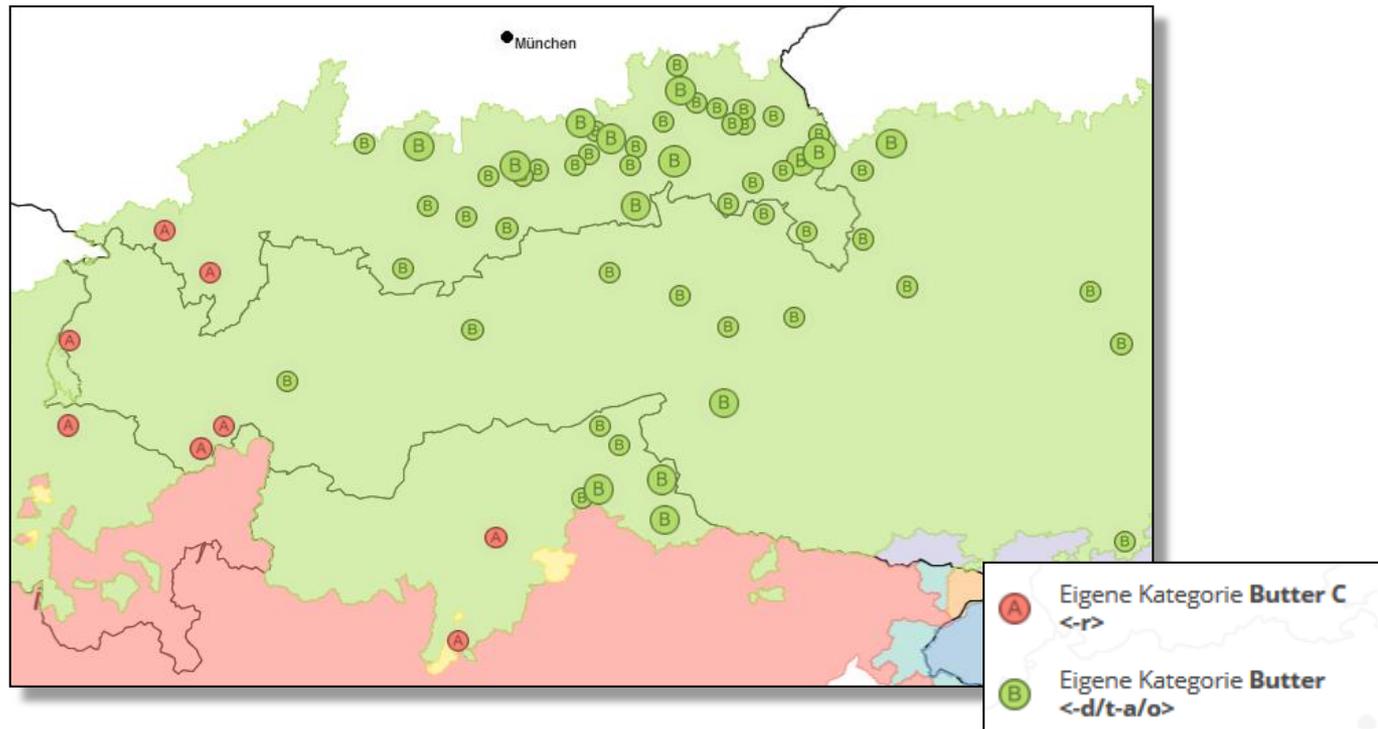


| Token | morph.-lex. Typ | Anzahl |
|-------|-----------------|--------|
| Budda | Butter | 38 |
| Butta | Butter | 26 |
| Butto | Butter | 6 |
| Buda | Butter | 3 |
| Putto | Butter | 2 |
| Buto | Butter | 1 |

Phonetik/Phonologie

Realisierung -<er>

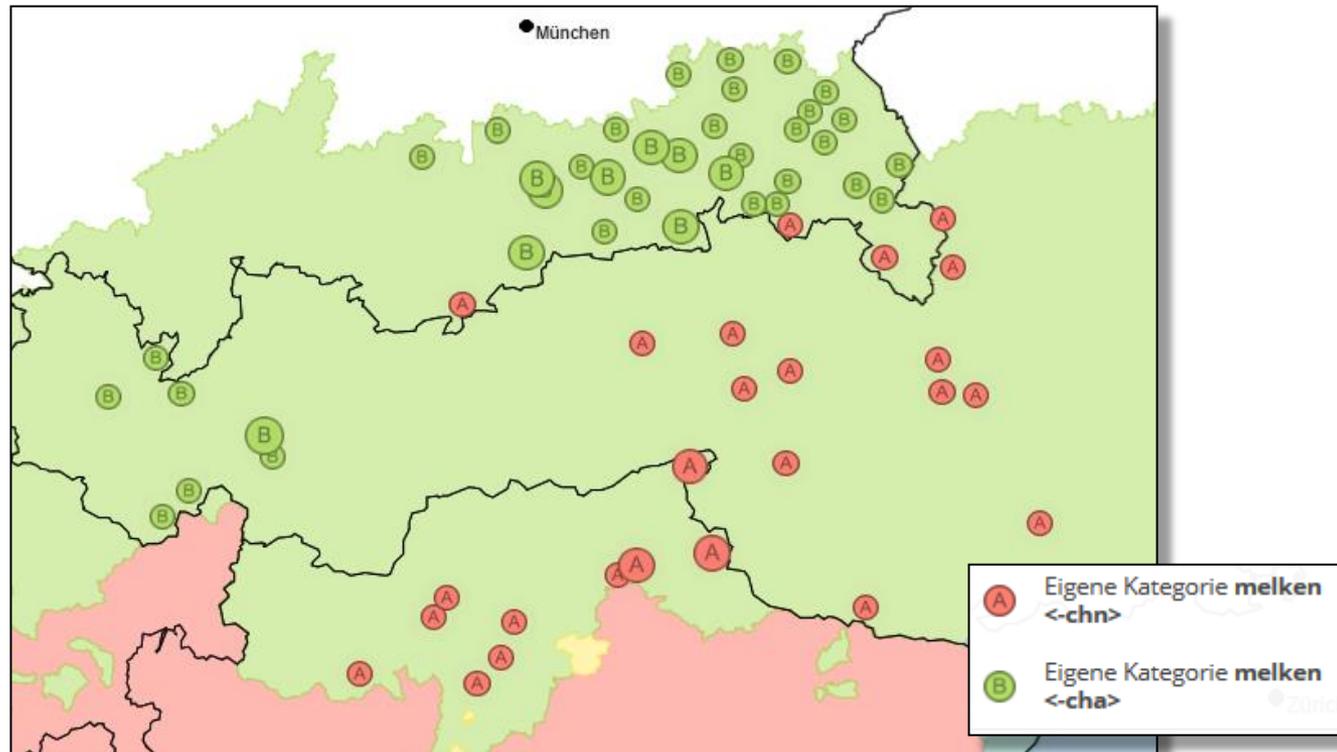
nhd. *Butter*



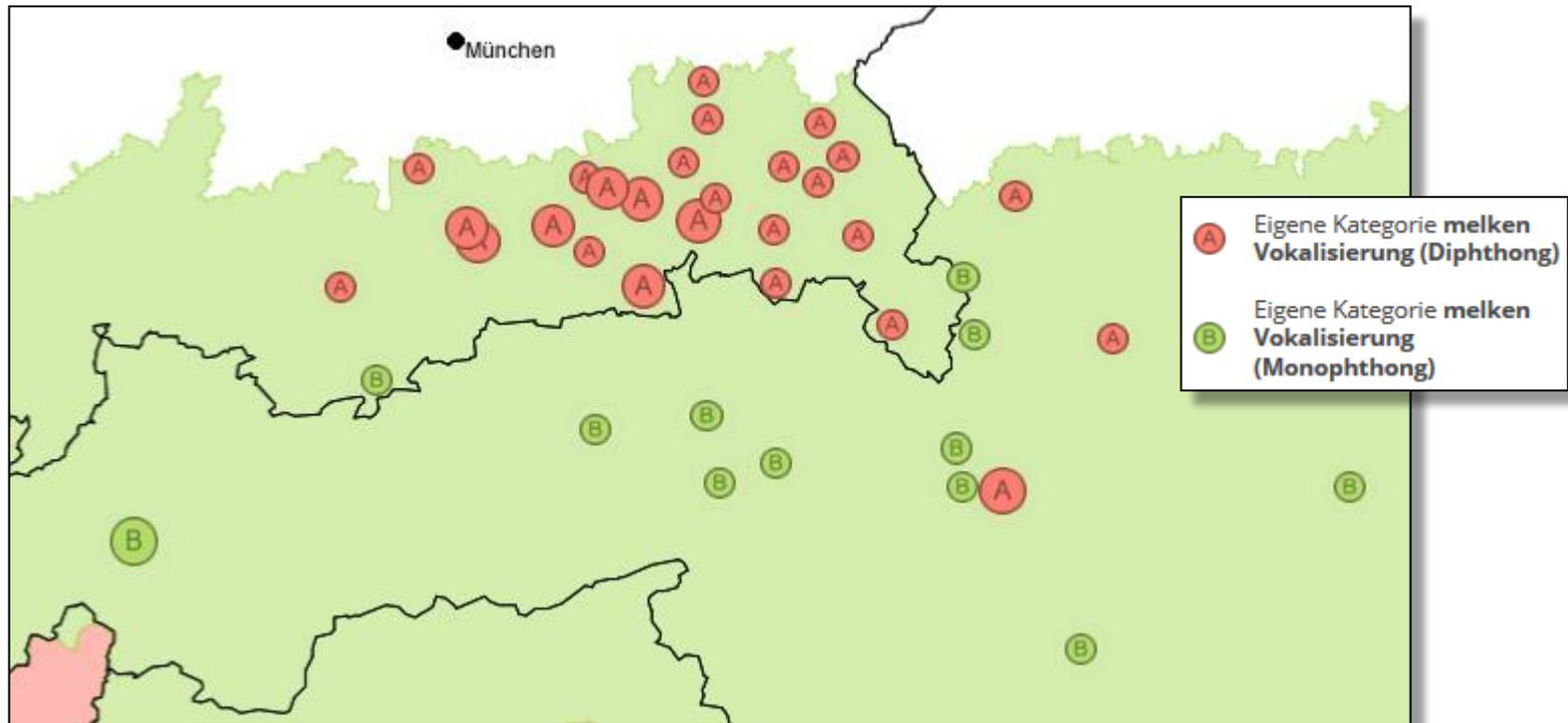
Phonetik/Phonologie

Realisierung -<en>

nhd. *melken*



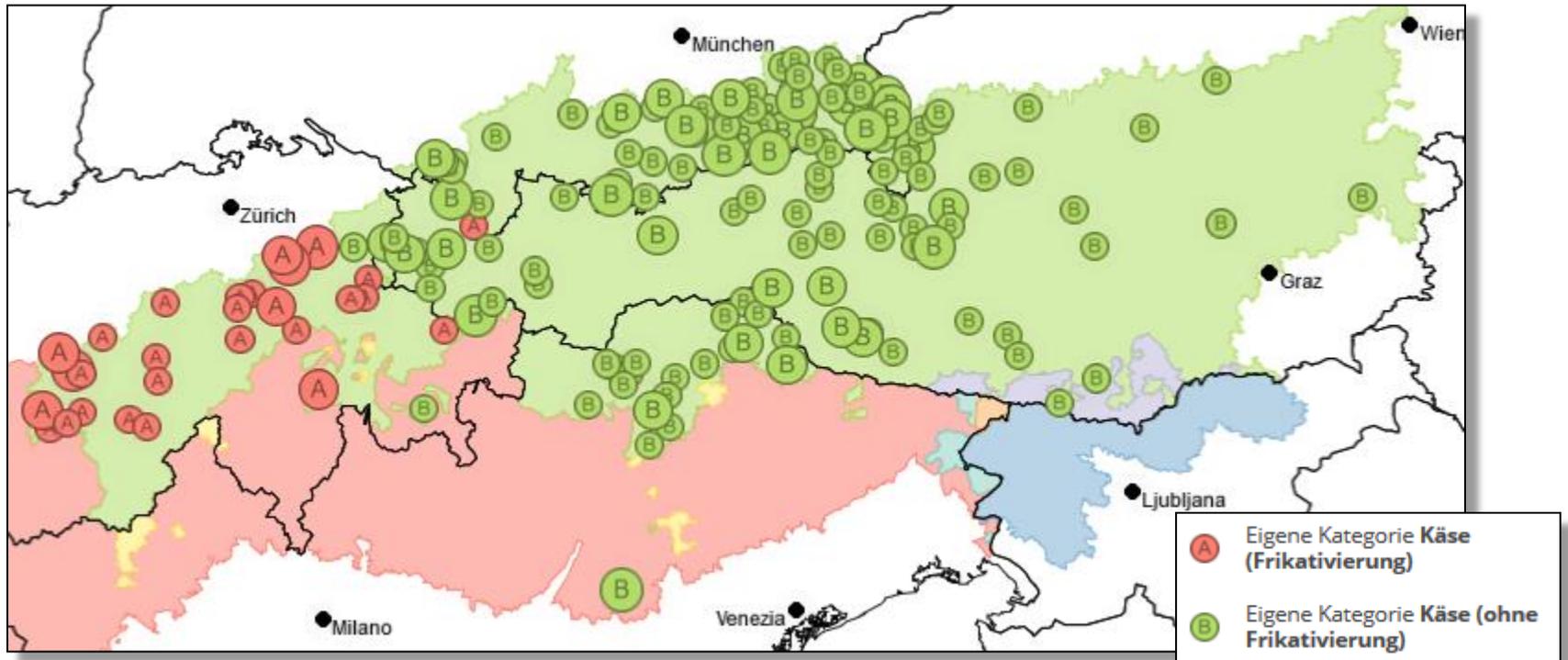
Phonetik/Phonologie
Vokalisierung -<el>-
nhd. *melken*



Phonetik/Phonologie

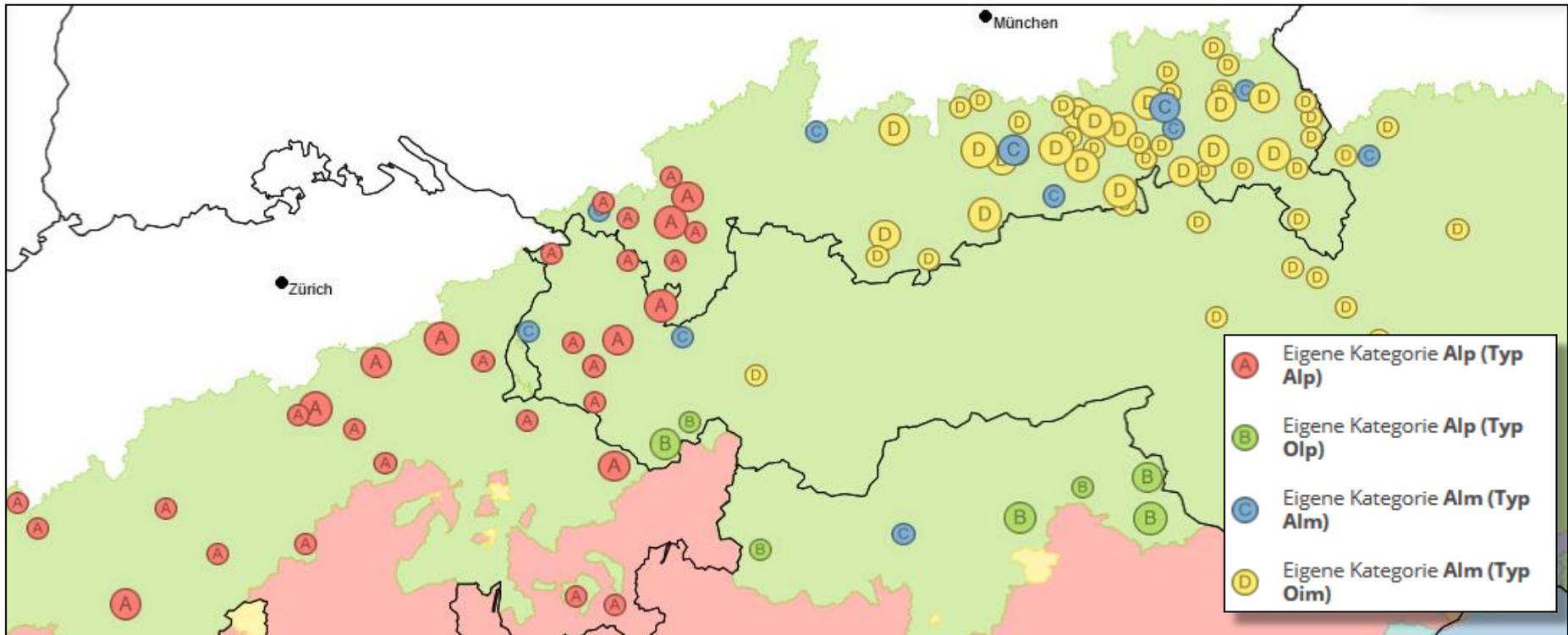
Spriantisierung

nhd. *Käse*

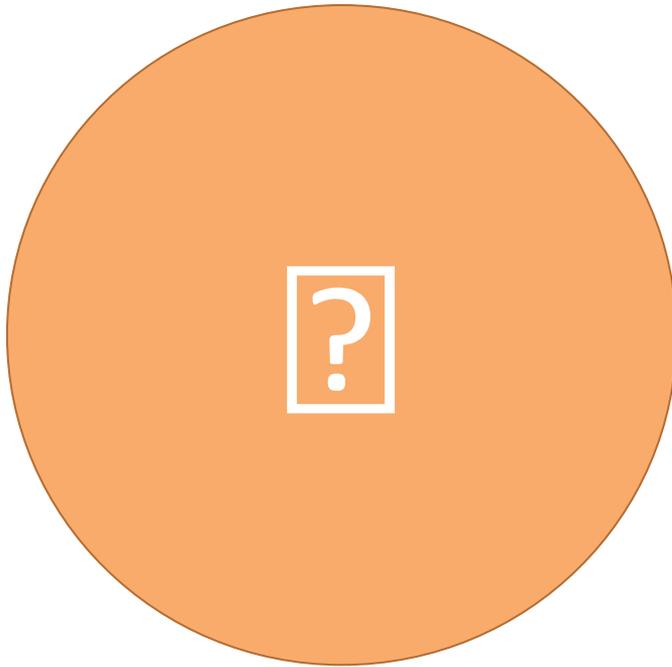


Phonetik/Phonologie

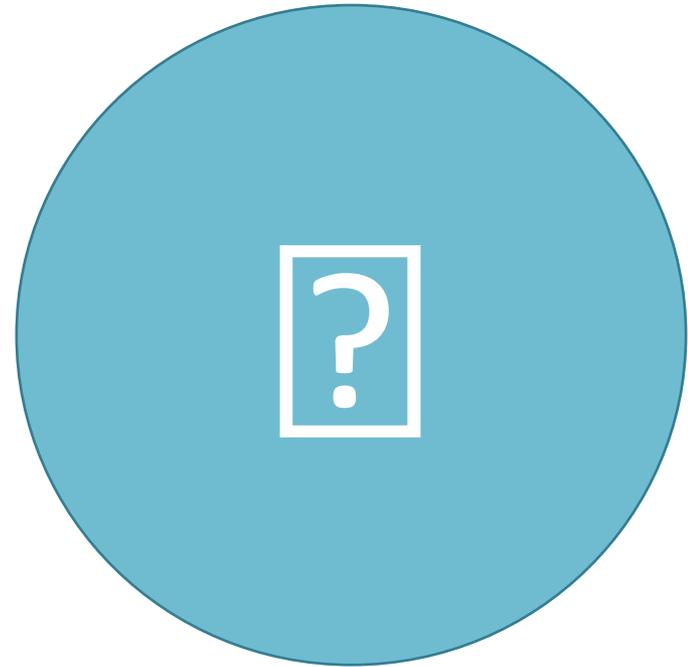
nhd. *Alp* vs nhd. *Alm*



Herkunft der Daten

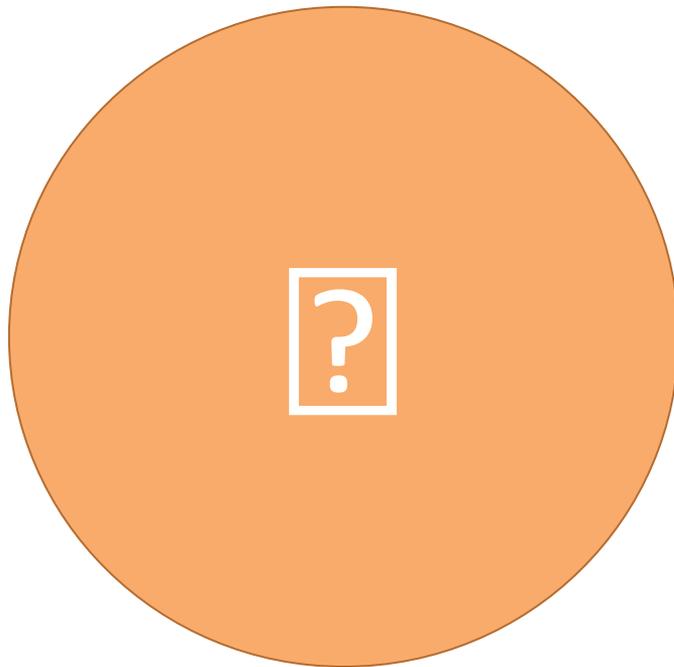


Atlanten
Wörterbücher



Crowdsourcing

Anzahl der deutschsprachigen Belege (typisiert)

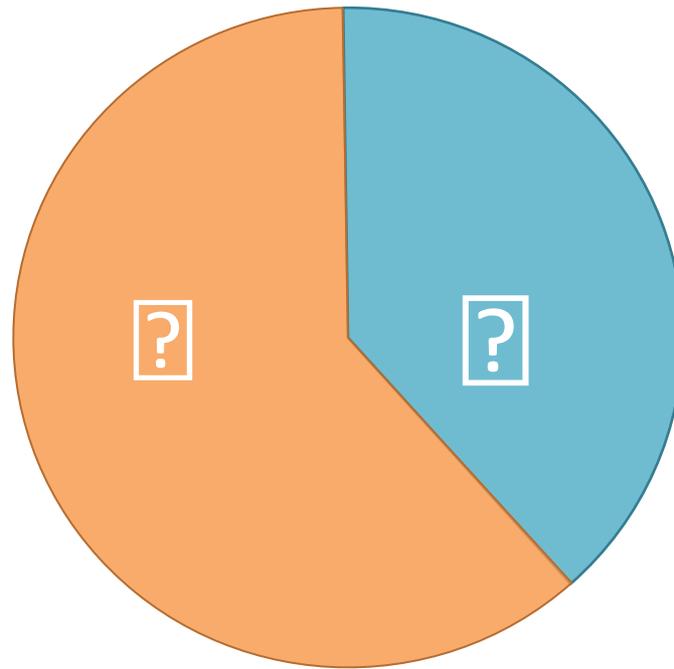


~ 29.000



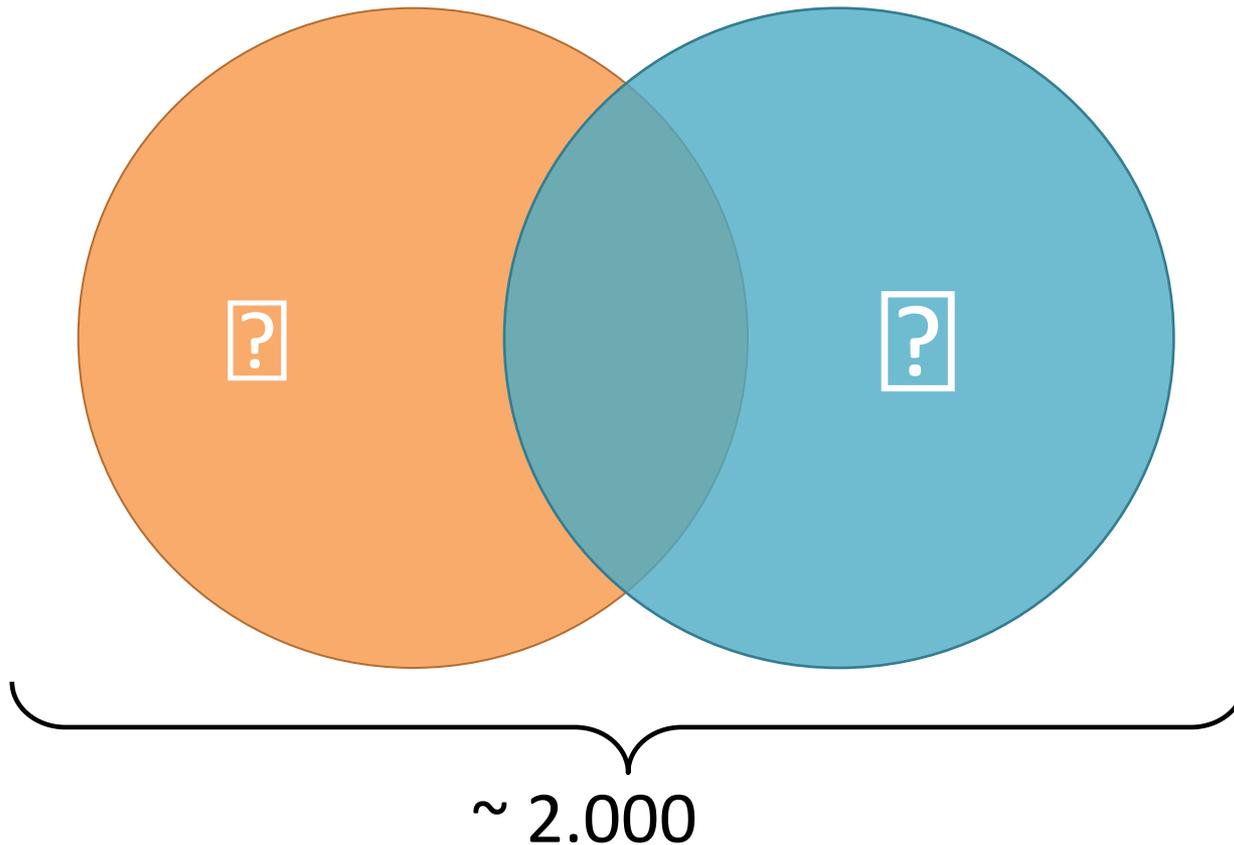
~ 7.000

Anzahl der deutschsprachigen Belege (typisiert)

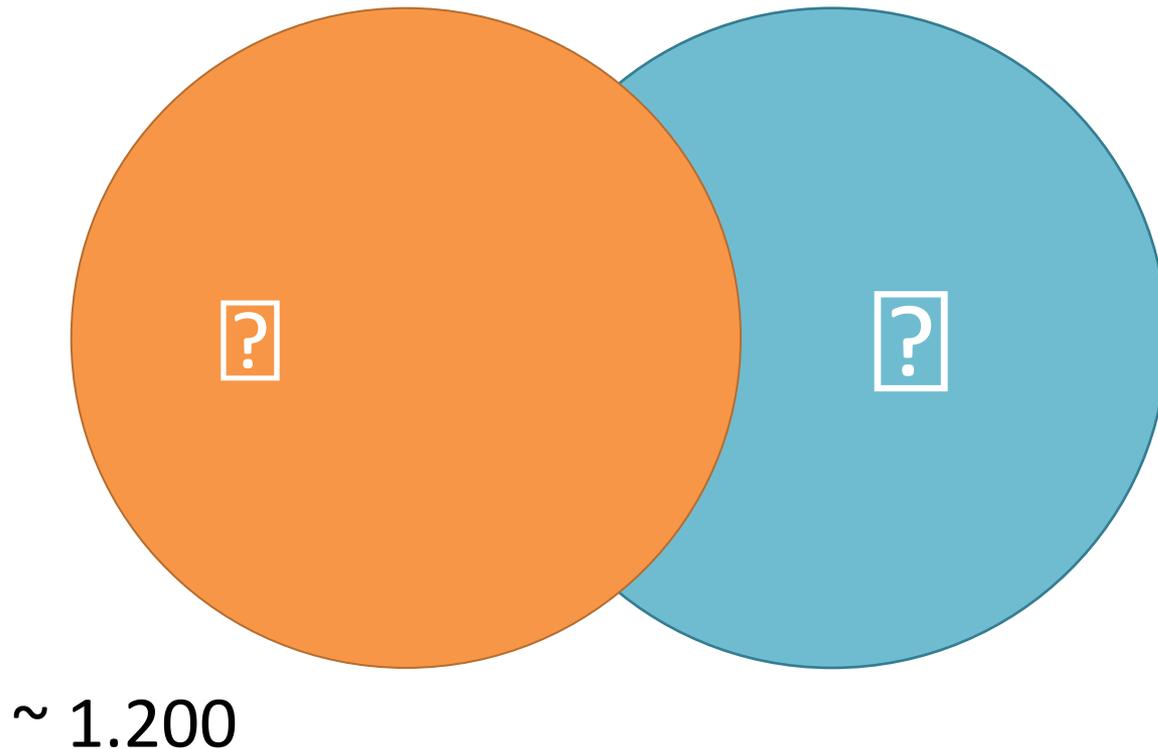


~ 36.000

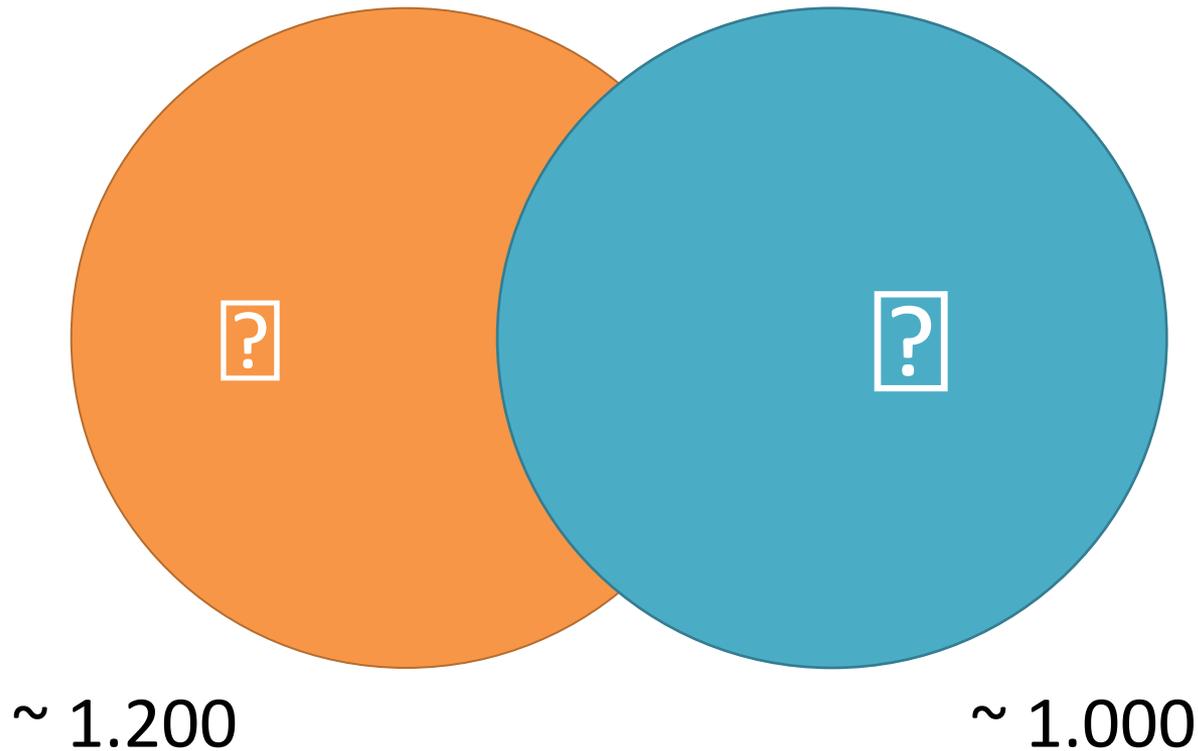
Anzahl der morpho-lexikalischen Typen gesamt (ger.)



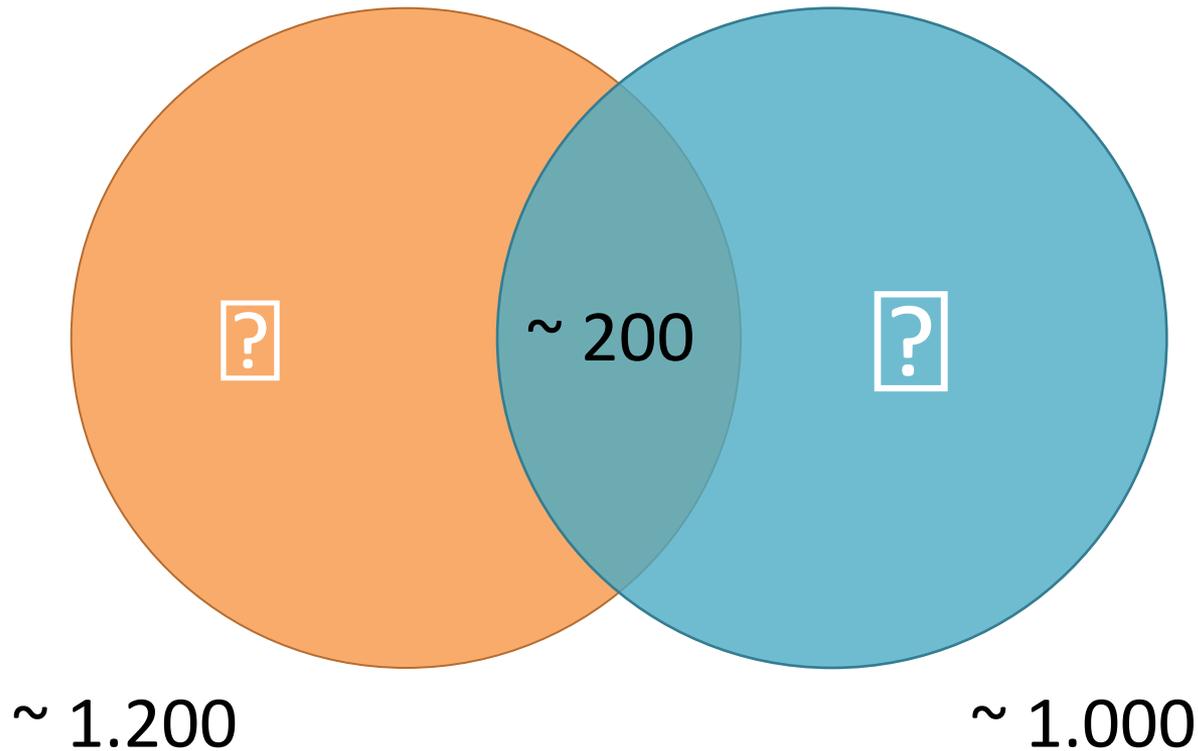
Anzahl der morpho-lexikalischen Typen verteilt (ger.)



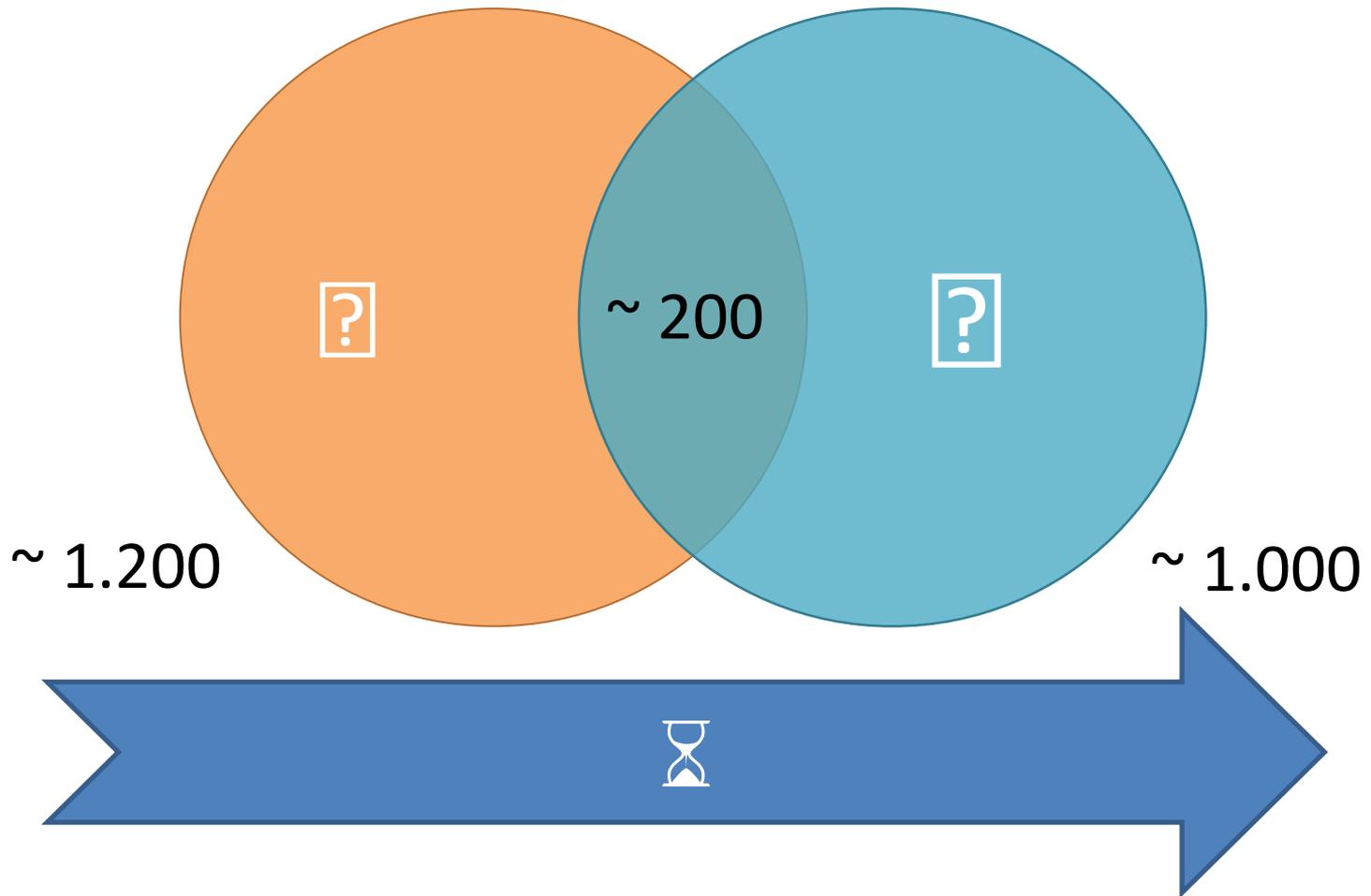
Anzahl der morpho-lexikalischen Typen verteilt (ger.)



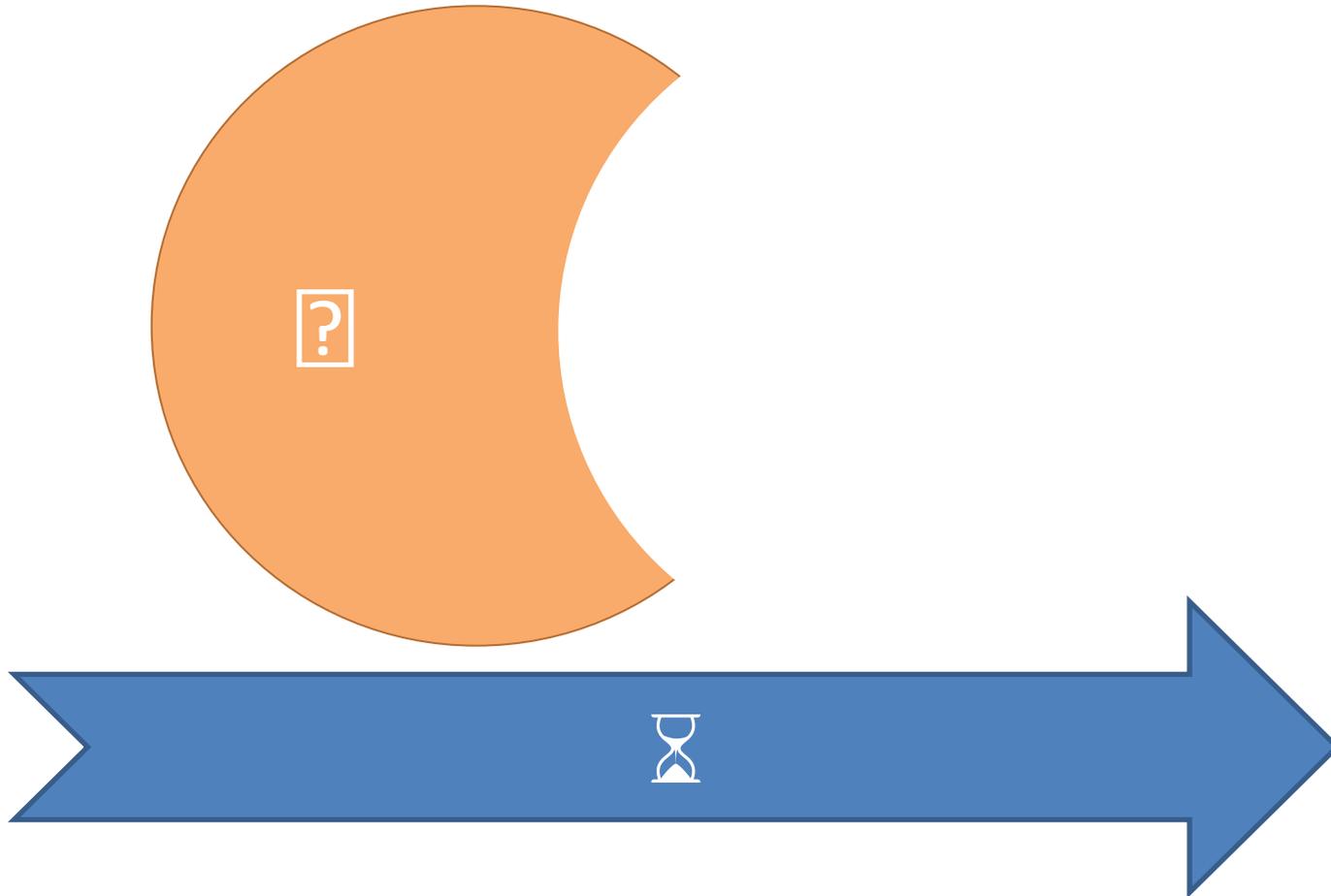
Anzahl der morpho-lexikalischen Typen verteilt (ger.)



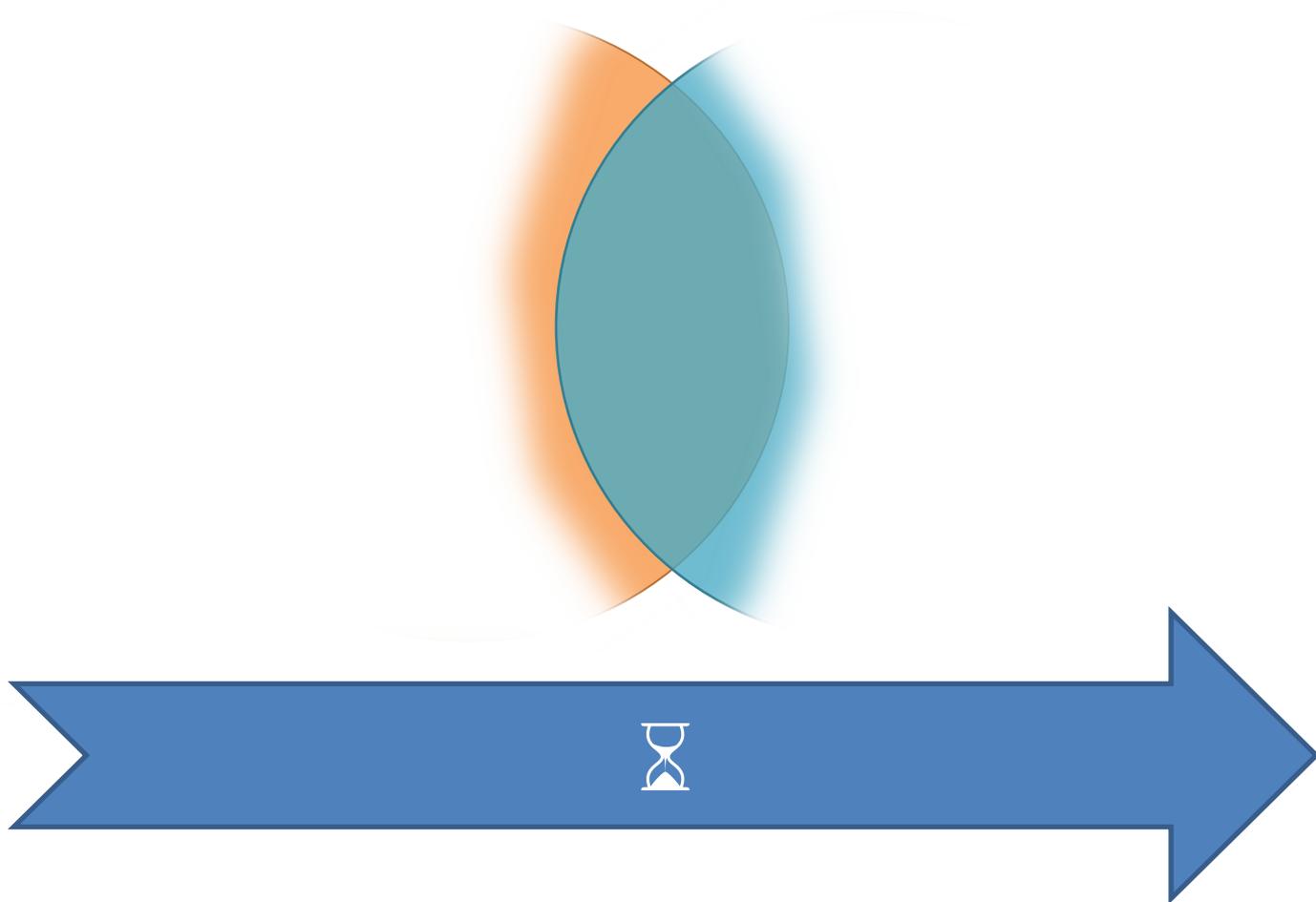
Verteilung der Daten aus diachronen Gesichtspunkten



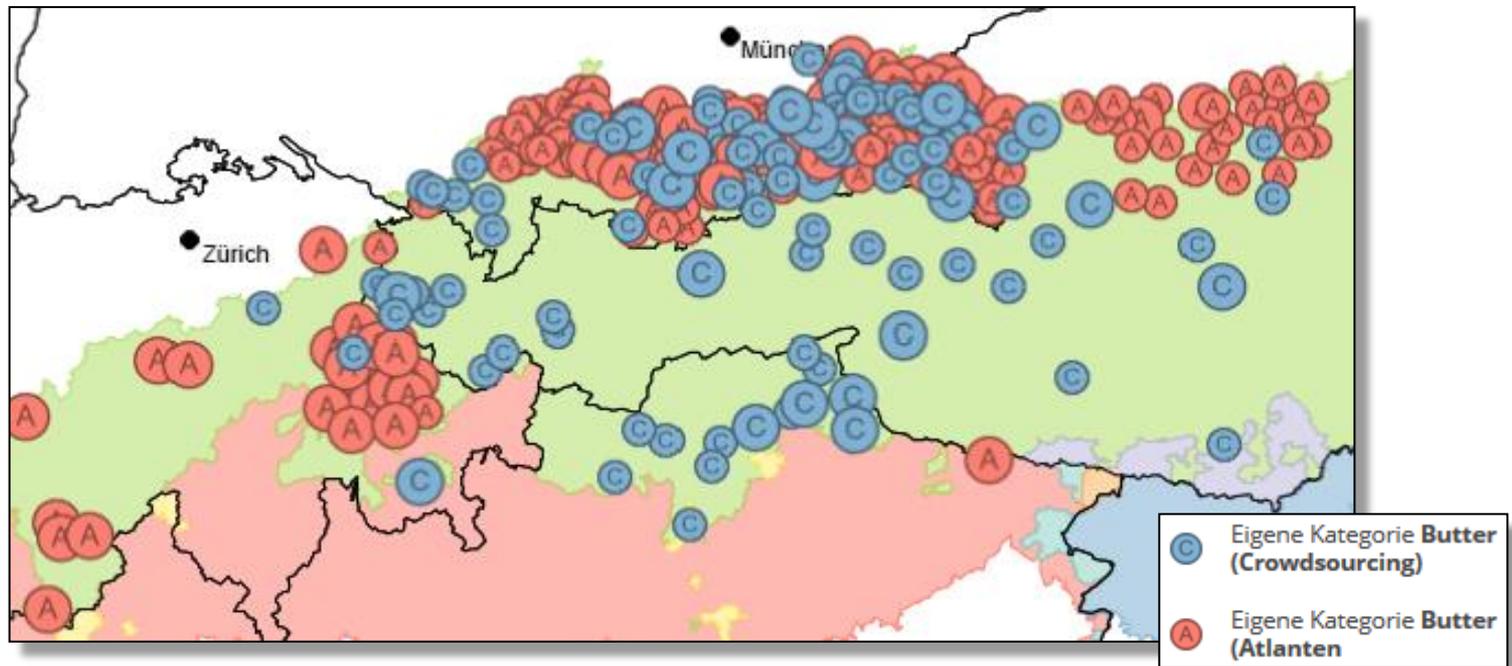
Alter Datenbestand, in Crowd-Daten nicht mehr belegt



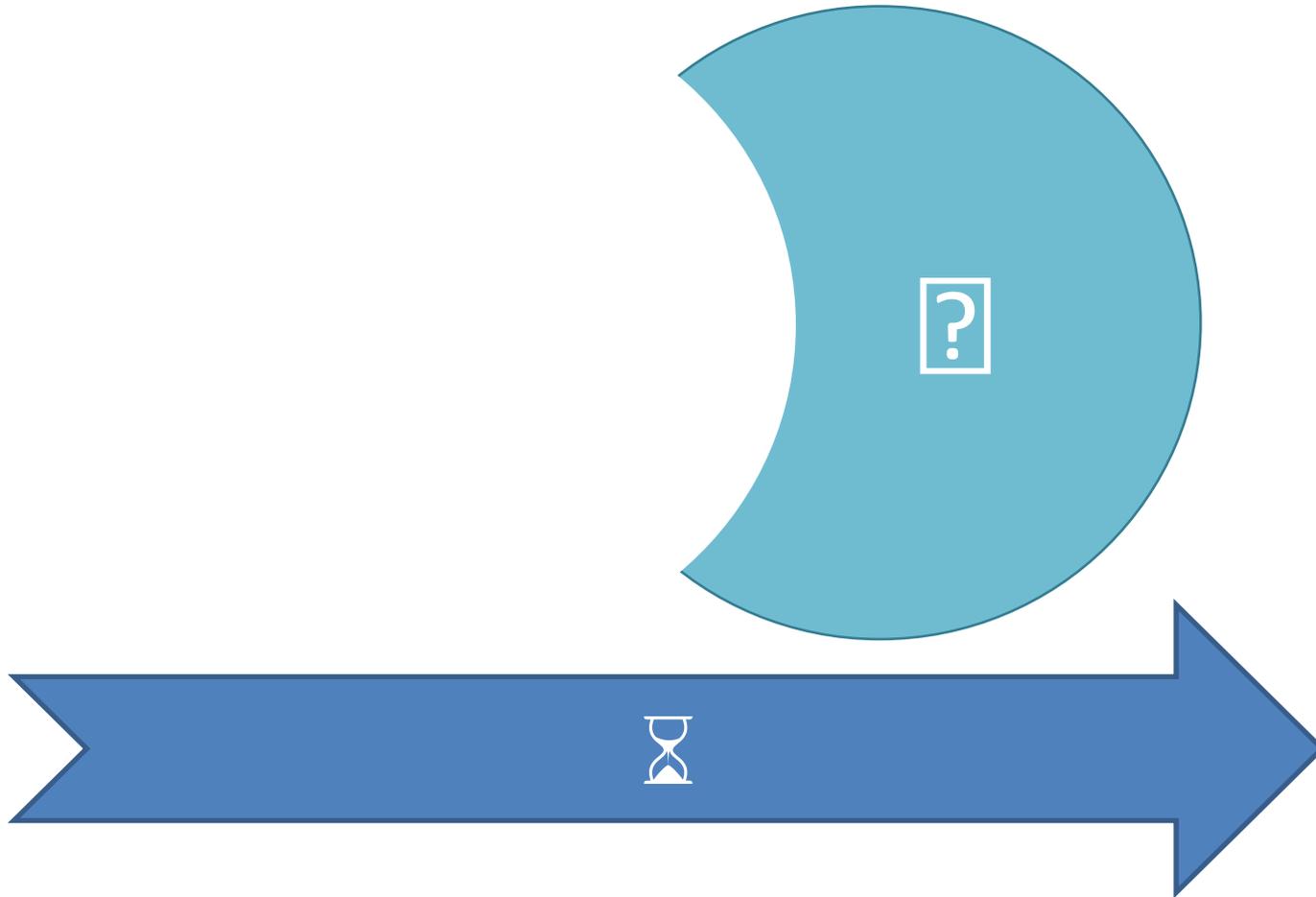
Alte und Crowd-Belege zu einzelnen Worttypen



Alte und Crowd-Belege zu einzelnen Worttypen

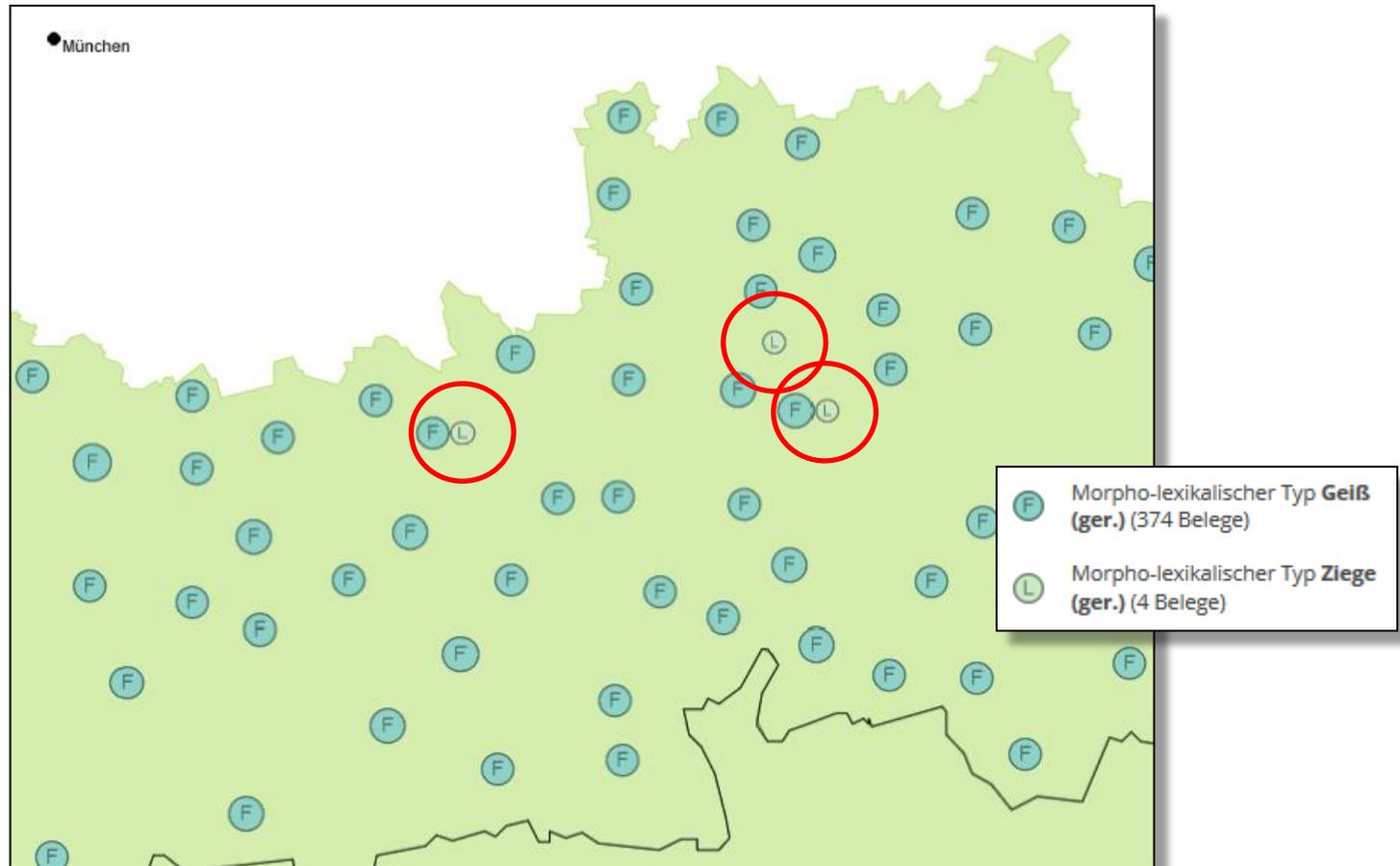


Neuer Datenbestand, Innovationen/Neologismen/...



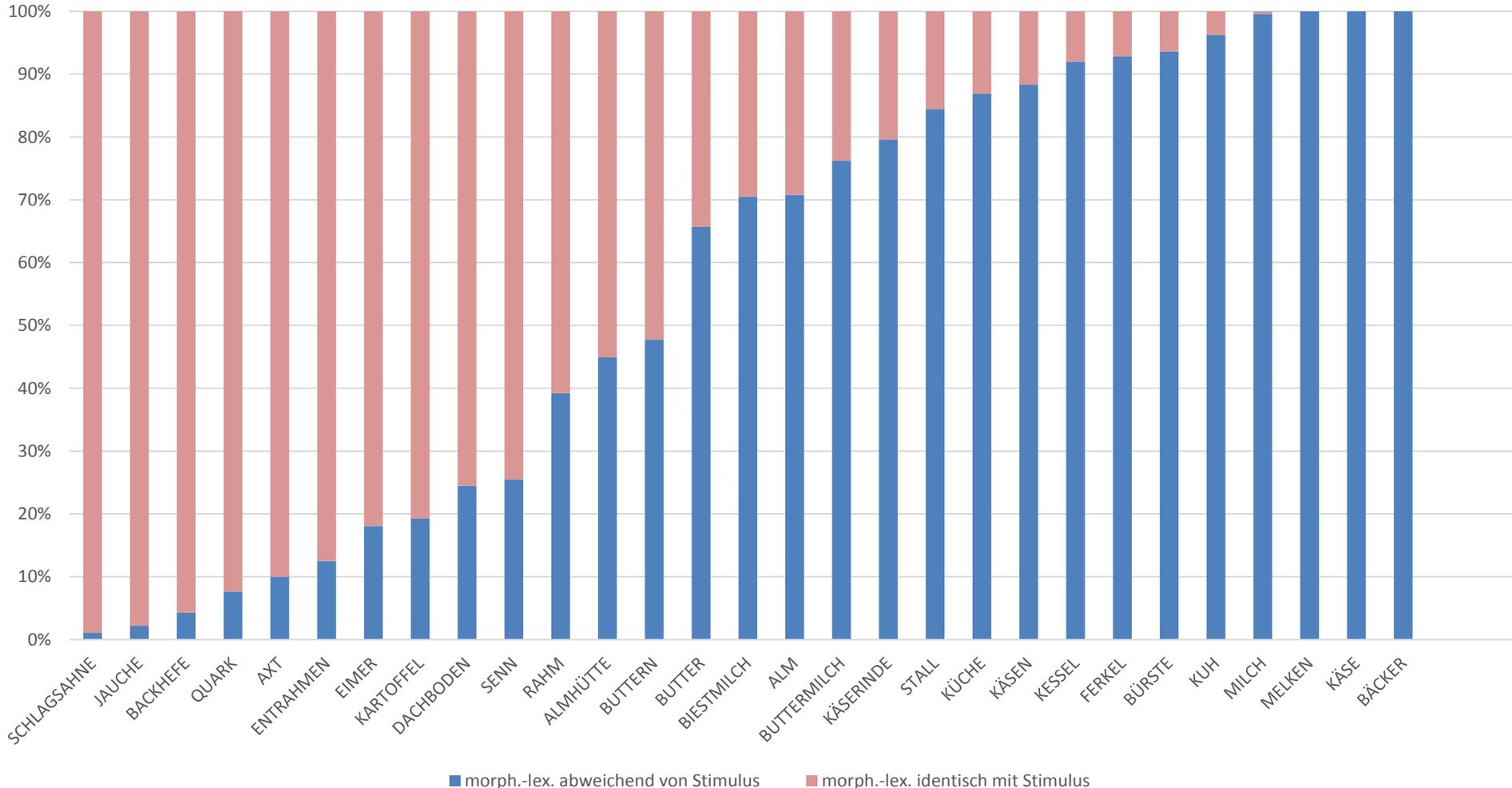
Einblick in die Crowdsourcing-Daten

Neuer Datenbestand, Innovationen/Neologismen/...



Einblick in die Crowdsourcing-Daten

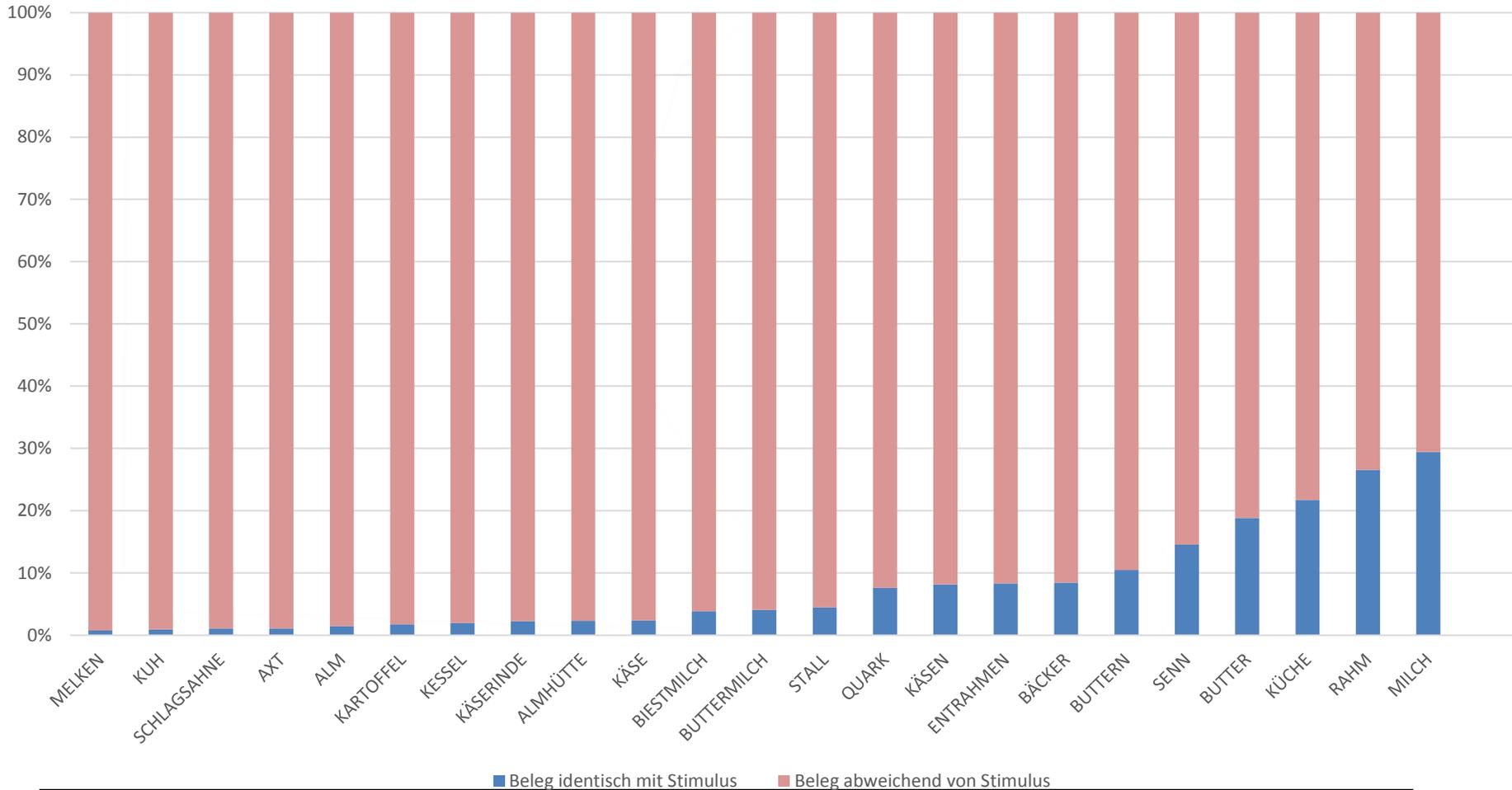
Verhältnis von Stimulus zu morpho-lexikalischem Typ



■ morph.-lex. abweichend von Stimulus ■ morph.-lex. identisch mit Stimulus



Verhältnis von Stimulus zu morpho-lexikalischem Typ



Fazit

■ Grenzen

- Resonanz der CS-Aktionen schwer vorhersehbar
- Verhältnis von erreichten Personen und Rücklauf nicht kalkulier- und nachvollziehbar (Schneeballeffekt)
- Daten aus lautlicher Perspektive schlecht verwertbar
- Wenig Metadaten der Informanten



Fazit

■ Chancen

- Schließen arealer und diachroner Lücken
- Erweiterte Analysemöglichkeiten im Bereich der Lexik
- Datenerhebung zum aktuellen Sprachstand
- Ökonomische Methode der Datenerhebung
- Aufdecken neuer lokaler Varianz



Vielen Dank für Ihre Aufmerksamkeit!

www.verba-alpina.gwi.uni-muenchen.de

