

Der Audio-Atlas Siebenbürgisch-Sächsischer Dialekte
„ASD“

Die Datenbasis: Das ASD-Corpus

- Über 350 Stunden Audio-Material
- Ursprünglich auf Tonband
- Jetzt digitalisiert
- 2212 Dateien, 141 GB wav, 11 GB mp3
- Aufnahmen zw. 1968 und 1973
- Universitäten Bukarest, Hermannstadt, Klausenburg
- Über IKGS aus Gundelsheim

ASD-Corpus: Inhalt

- Inhaltliche Erschließung: grobe Inhaltsangaben auf Tonbändern
- Übertragen in Excel-Tabelle (wohl in Gundelsheim; nicht immer ganz zuverlässig)
- Informanten aus über 200 siebenbürgischen Ortschaften
- Informanten im Alter zw. 5 und 93 Jahren

ASD-Corpus: Inhalt

- Gelenkte Interviews: 140 Wenkersatzaufnahmen
- Rest: Spontansprachliche Interviews
- Inhaltliche Vielfalt
- Aktuell rund 2200 Stichwörter abrufbar, z.B.:
 - Ackerbau, Kollektivierung,
 - Hochzeitsbräuche, Märchen,
 - Deportationen, Osterbräuche,
 - Auswanderung ...

- März 2009: Einrichtung der Internetplattform
<http://www.asd.gwi.uni-muenchen.de/>
- 2009: 1. Förderung (BKM, IKGS):
phonetische Transkription der Wenkersatzdateien
- 2011: 2. Förderung (BKM, IKGS):
standardnahe Transkription
ausgewählter spontansprachlicher
Interviews

Projektverlauf

- Ab März 2012 bis Febr. 2014:
3. Förderung (BKM, IKGS): Weitere inhaltliche Erschließung, Erstellung einer Ontologie, morphosyntaktische Etikettierung, sprachwissenschaftliche Analyse, Erweiterung des Datenbestands

Projektverlauf: Aktueller Bearbeitungsstand

- **Wenkersatz-Interviews:**
 - 137 Dateien phonetisch transkribiert
 - 14 Audio-Stunden, ca. 140.000 Wörter
- **Spontansprachliche Interviews:**
 - 364 Dateien standardnah transkribiert
 - 54 Audio-Stunden, ca. 450.000 Wörter (entspricht ca. 1100 DIN-A4-Seiten)
- **Transkriptionsleistung insgesamt:**
 - 68 Audio-Stunden
 - 590.000 Wörter → ca. 1450 DIN-A4-Seiten

Technische Konzeption

- Seit 2009 Internetplattform
<http://www.asd.gwi.uni-muenchen.de/>
- Client-Server-Prinzip
- Backend: MySQL-Datenbank
- Frontend: Webseite
- Schnittstelle zw. Backend und Frontend:
PHP-Skripts



An der Oberfläche: Frontend

- Aktuell im wesentlichen zwei Module:
 - Corpus
 - Wenkersatzanalyse

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR DEUTSCHE KULTUR UND GESCHICHTE SÜDOSTEUROPAS
INSTITUT FÜR ROMANISCHE PHILOLOGIE
IT-GRUPPE GEISTESWISSENSCHAFTEN

**AUDIO-ATLAS
SIEBENBÜRGISCH-SÄCHSISCHER DIALEKTE**

ASD

LMU-Startseite asduser - Logout

ÜBER DAS PROJEKT

CORPUS

WENKERSATZANALYSE

DOKUMENTE

Leitung:
Thomas Krefeld | Stephan Lücke | Stefan Sienerth

Transkription:
Grete Klaster-Ungureanu | Horst Schuller | Johannes Sift | Ute Maurer

Verfügbare Optionen

Frontend: Corpus

➤ „Corpus“ bietet Zugriff auf Sound und Transkriptionen (sofern vorhanden); Filterung nach verschiedenen Kriterien möglich:

- Herkunft und Alter der Informanten
- Jahr der Erhebung
- Schlagwörter auf Basis der Gundelsheimer Excel-Tabelle

Datenbestand: 2212 | Anzahl gefundene Dateien: 35
[Alle Dateien](#) | [Nur transkribierte Dateien](#) | [Alle Filter entfernen](#) | [Karte bei Google Earth](#)

Audio	Ort	Jahr	Alter	Corpus	Stichwörter
Filter	Filter	Filter	Filter	Filter	Hochzeit
 	Arkedon (S)		49	Bukarest	Hochzeit
 	Birk (S)		73	Bukarest	Hochzeit in Birk
 	Brenndorf (S)		62	Bukarest	Hochzeit
 	Felldorf (S)		42	Bukarest	Kinder; Landwirtschaft; Verlobung und Hochzeit; Hochzeitsbräuche; Erzählung
 	Girelsau (S)	1968	60	Bukarest	Hochzeitsbrauch
 	Großeidau (S)	1972	68	Bukarest	Hexen; Truden; Hochzeit; Bräuche; Aufgaben einer Ehefrau; Gemüseverarbeitung
	Großeidau (S)	1972	54	Bukarest	Jugenderinnerungen; Bräuche; Feste;

Frontend: Corpus

- Klick auf Audiosymbol zeigt Transkription und startet Ton-Wiedergabe
- Paßwortgeschützt
- Synoptische Darstellung phonetischer und orthograph. Transkr., wo vorhanden

Audio: 1438 | Jahr: 1438 | Alter: 66 | Corpus: Bukarest | Stichwörter: Jugenderinnerungen; Fest in Kronstadt; Rosenauer Burg; Tracht und Stadtkleidung; Theaterspiel; Museumseinrichtung; Tanz früher; Spinnstube; Rockenstube

Phonetische Transkription
Anmerkungen | Textgrid-Datei

Orthographische Transkription
Anmerkungen | Textgrid-Datei

Transkription noch nicht verfügbar

[1]: #1# Mich heizt man ### aus Arkeden
[2]:
[3]: #2# {?unverstaendlich}
[4]:
[5]: #1# Und ich will Ihnen gern erzahlen as fuer eine schoene<huersch> Jugendzeit ich mit habe<hun> gemacht
[6]:
[7]: #1# Geboren bin ich in meinem Elternhaus
[8]: #2# In welchem<welem> Jahr
[9]: #1# Im Jahr Tausendneunhundertsechs
[10]:
[11]: #1# Und ich hatte nur einen<ein> Bruder noch
[12]: #1# Da habe<hun> ich ein Leben gehabt mit einem schoenen<huerschen> Rosengarten
[13]:

Audio: 1113-05 | Ort: Abtsdorf bei Agnetheln (S) | Jahr: 1972 | Alter: 32 | Corpus: Bukarest | Stichwörter: Wenkersätze

Phonetische Transkription
Anmerkungen | Textgrid-Datei

Orthographische Transkription
Anmerkungen | Textgrid-Datei

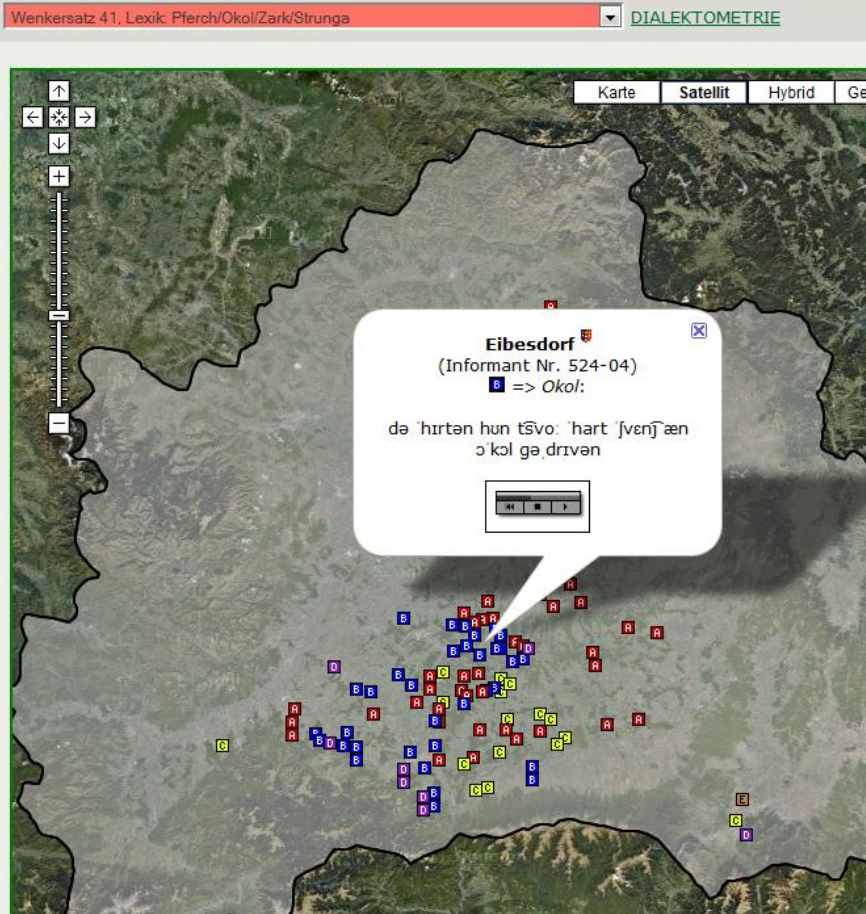
[1]: ɪç h̥i̯sə fʊstər johana bæ fun ɔpəsdrɛf
[2]: fɔ vənaiɓmhær
[3]: vɛŋkɛrɛtʂɛ
[4]: {WS Anfang}
[5]: {WS 1} u̯ɛnt
[6]: ## æm væŋtɛr 'flejn də gə'drɛçt 'bladɛr æn dɛr 'loft ə ram
[7]: {WS 2} tʂvæ:
[8]: ## et hi̯rt glɛç 'oɔf fʊm 'no:ən drɛ vɪd dət 'vader 'vedɛr bɛisɪr
[9]: {WS 3} dro

Wenkersätze

1. Im Winter fliegen die trockenen Blätter in der Luft herum.
2. Es hört gleich auf zu schneien, dann wird das Wetter wieder besser.
3. Tu Kohlen in den Ofen, dass die Milch anfängt zu kochen.
4. Der gute alte Mann ist mit dem Pferd durchs Eis gebrochen und in das kalte Wasser gefallen.
5. Er ist vor vier oder sechs Wochen gestorben.
6. Das Feuer ist zu stark, der Kuchen ist ja unten

Die Hirten haben zwei
Herden Schweine in den
Pferch getrieben

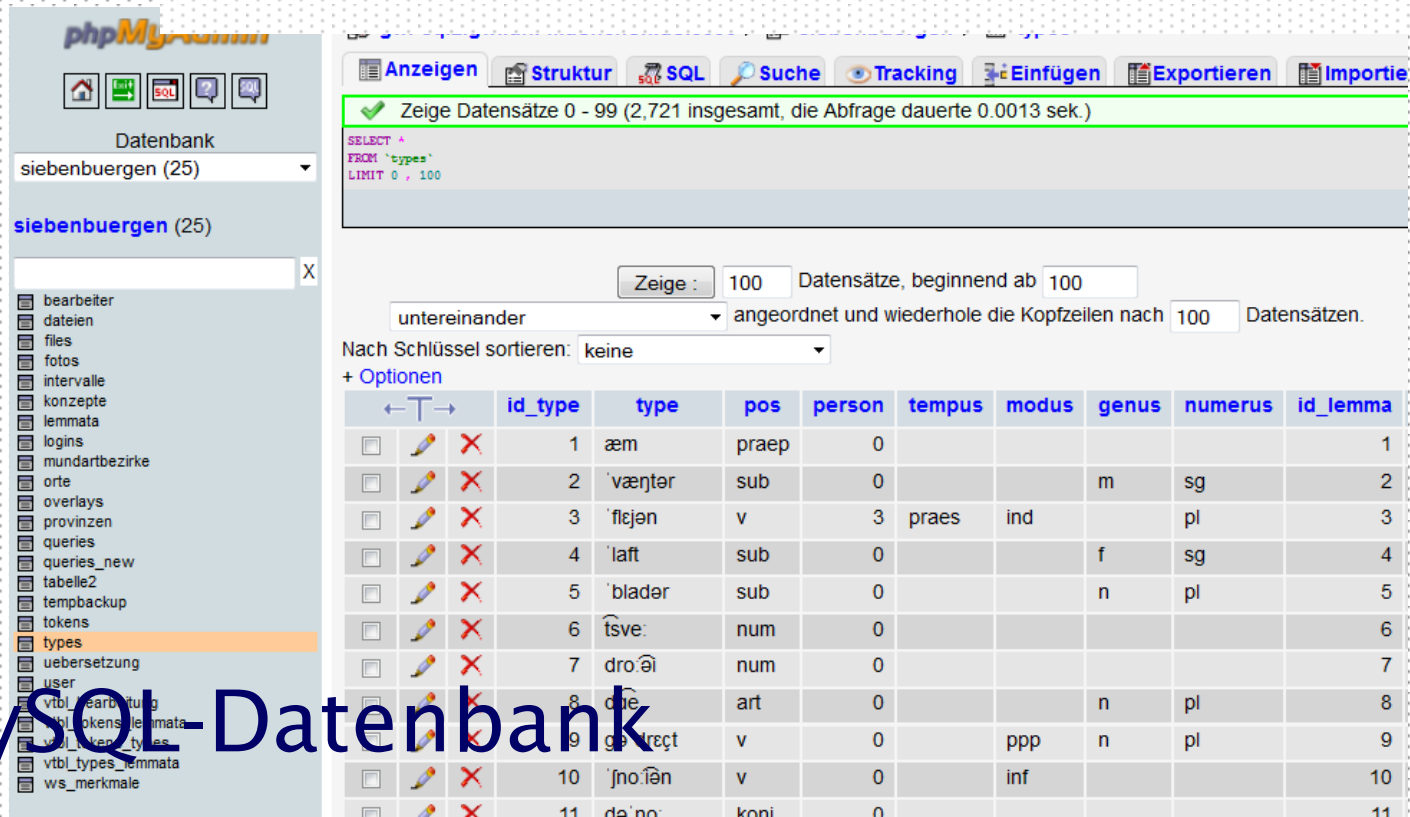
- => Pferch
- => Okol
- => Zark
- => Strunga
- => Umzäunung



Frontend: Wenkersatzanalyse

- Repräsentation in Gestalt eines Atlas (aktuell: Google Maps)
- Analyse auf Basis von Textmustersuche
- Abfrage definierter Merkmale aus den Bereichen Phonetik, Lexik und Morphosyntax
- Ergebnispräsentation durch farbige Symbole auf der Karte
- Anklicken der Symbole startet Wiedergabe des Audio-Belegs

Im Hintergrund: Die Datenbank



phpMyAdmin

Datenbank
siebenbuergen (25)

Die Datenbank **siebenbuergen** (25) enthält folgende Tabellen:

- bearbeiter
- dateien
- files
- fotos
- intervalle
- konzepte
- lemmata
- logins
- mundartbezirke
- orte
- overlays
- provinzen
- queries
- queries_new
- tabelle2
- tempbackup
- tokens
- types
- uebersetzung
- user
- vtbl_bearbeitung
- vtbl_lemmata
- vtbl_lemmata
- vtbl_lemmata
- vtbl_types_lemmata
- ws_merkmale

Anzeigen Struktur SQL Suche Tracking Einfügen Exportieren Importieren

Zeige Datensätze 0 - 99 (2,721 insgesamt, die Abfrage dauerte 0.0013 sek.)

```
SELECT *  
FROM `types`  
LIMIT 0 , 100
```

Zeige : 100 Datensätze, beginnend ab 100

untereinander angeordnet und wiederhole die Kopfzeilen nach 100 Datensätzen.

Nach Schlüssel sortieren: keine

+ Optionen

	id_type	type	pos	person	tempus	modus	genus	numerus	id_lemma
<input type="checkbox"/>	1	æm	praep	0					1
<input type="checkbox"/>	2	'væn̥tər	sub	0			m	sg	2
<input type="checkbox"/>	3	'flejən	v	3	praes	ind		pl	3
<input type="checkbox"/>	4	'laft	sub	0			f	sg	4
<input type="checkbox"/>	5	'blader	sub	0			n	pl	5
<input type="checkbox"/>	6	ʃsve:	num	0					6
<input type="checkbox"/>	7	dro:əi	num	0					7
<input type="checkbox"/>	8	die	art	0			n	pl	8
<input type="checkbox"/>	9	gə'ʁɛçt	v	0		ppp	n	pl	9
<input type="checkbox"/>	10	'jno:ʔən	v	0		inf			10
<input type="checkbox"/>	11	de'no	koni	0					11

➤ MySQL-Datenbank

➤ Open source

➤ Kostenlos

➤ leistungsfähig

➤ bewährt

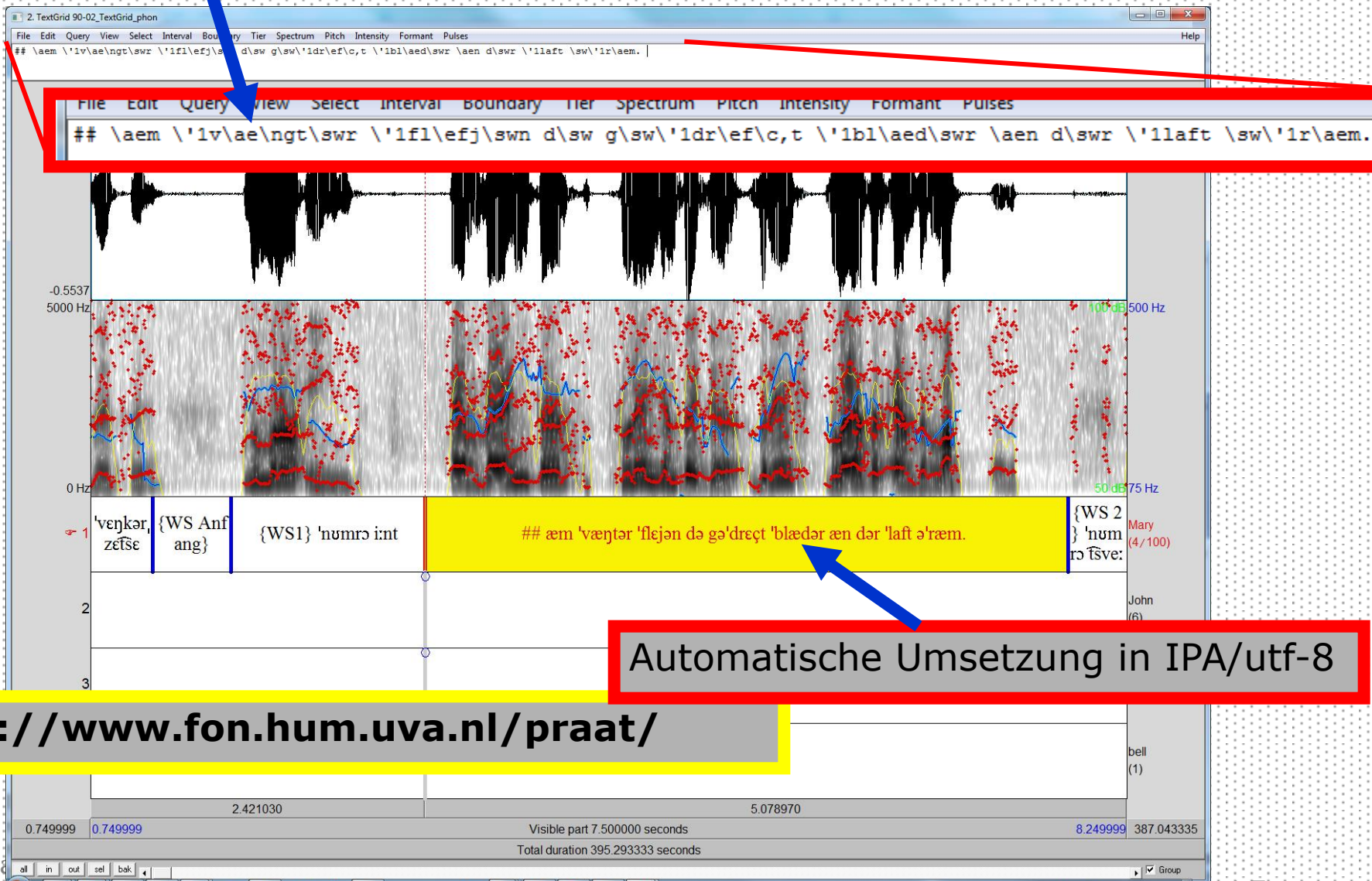
➤ internetfähig

Transkriptionsverfahren

- Verwendung des Programms „Praat“
(kostenfrei, Uni Amsterdam)
- Phonetische Transkription des
Wenkermaterials
- Standardnahe Transkription des
Spontanmaterials

Tastatureingabe

Transkription: Das Programm Praat

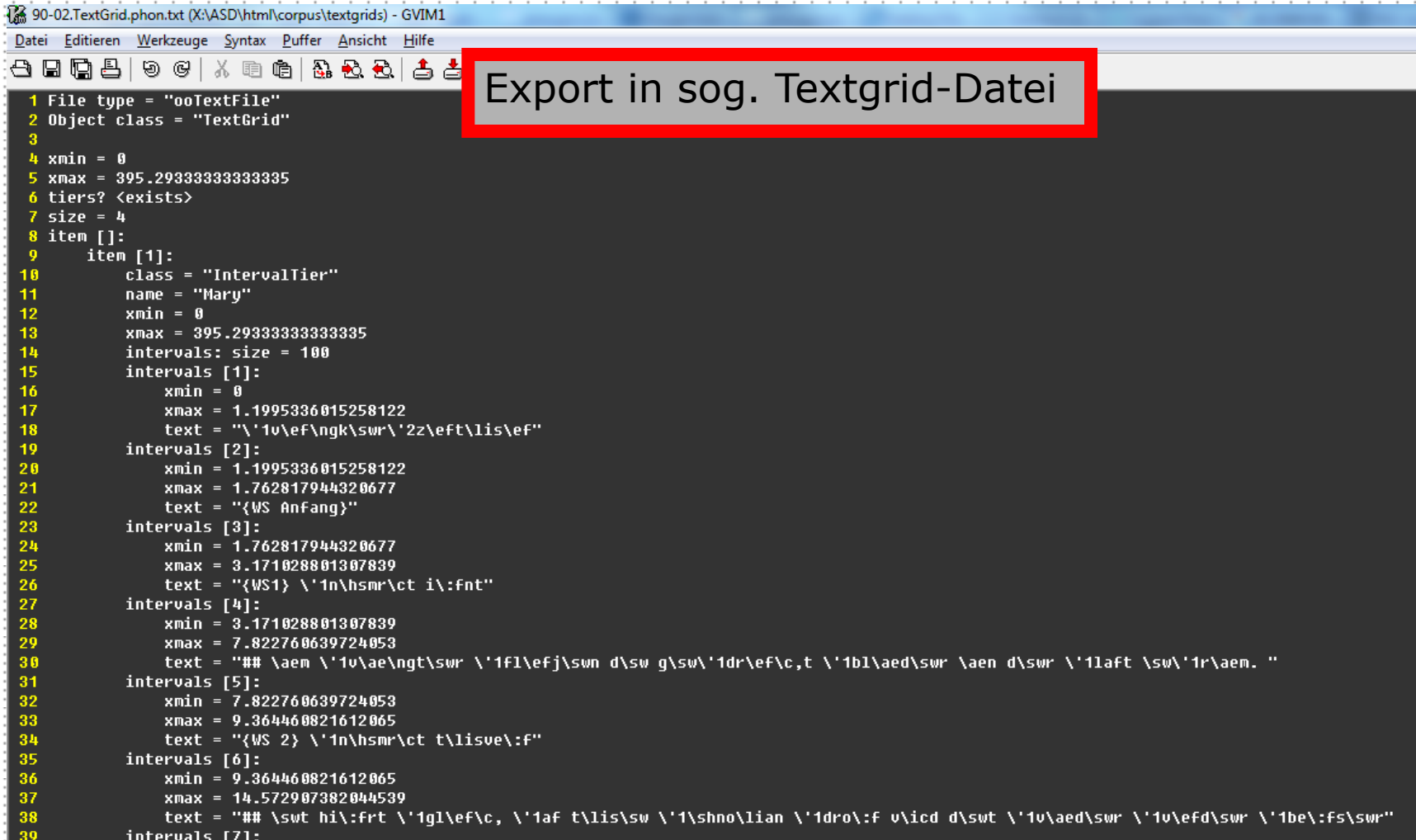


Automatische Umsetzung in IPA/utf-8

<http://www.fon.hum.uva.nl/praat/>

Von Praat in die Datenbank

Beispiel: phonetische Transkription einer Wenker-Aufnahme



```
90-02.TextGrid.phon.txt (X:\ASD\html\corpus\textgrids) - GVIM1
Datei Editieren Werkzeuge Syntax Puffer Ansicht Hilfe
Export in sog. Textgrid-Datei
1 File type = "ooTextFile"
2 Object class = "TextGrid"
3
4 xmin = 0
5 xmax = 395.29333333333335
6 tiers? <exists>
7 size = 4
8 item []:
9   item [1]:
10    class = "IntervalTier"
11    name = "Mary"
12    xmin = 0
13    xmax = 395.29333333333335
14    intervals: size = 100
15    intervals [1]:
16      xmin = 0
17      xmax = 1.1995336015258122
18      text = "'1u\ef\ngk\swr'2z\eft\lis\ef"
19    intervals [2]:
20      xmin = 1.1995336015258122
21      xmax = 1.762817944320677
22      text = "{WS Anfang}"
23    intervals [3]:
24      xmin = 1.762817944320677
25      xmax = 3.171028801307839
26      text = "{WS1} '\1n\hsmr\ct i:\fnt"
27    intervals [4]:
28      xmin = 3.171028801307839
29      xmax = 7.822760639724053
30      text = "## \aem '1u\ae\ngt\swr '1f1\efj\swm d\sw g\sw'1dr\ef\c,t '1b1\aed\swr \aen d\swr '11aft \sw'1r\aed. "
31    intervals [5]:
32      xmin = 7.822760639724053
33      xmax = 9.364460821612065
34      text = "{WS 2} '\1n\hsmr\ct t\lisue:f"
35    intervals [6]:
36      xmin = 9.364460821612065
37      xmax = 14.572907382044539
38      text = "## \swt hi:\fnt '1gl\ef\c, '1af t\lis\sw '1\shno\lian '1dro:f u\icd d\swt '1u\aed\swr '1u\efd\swr '1be:\fs\swr"
39    intervals [7]:
```


Von der Transkription in die Datenbank

Audio-Datei

Transkription mit Praat

Erzeugung der Textgrid-Datei

Umformung in Tabellengestalt mit awk-Skript

Import in Datenbank



Funktionsweise der Datenbank: SQL

Beispiel: Verwendung des Wortes „Gurre“ in Wenkersatz 4

Select

token, -- Wort, auf das das in der where-Klausel formulierte Muster paßt
nr, -- Nummer des Interviews
intervall, -- Nummer des Intervalls innerhalb des Interviews
`alter`, -- Alter des Informanten
dateien.ort, -- Name der Ortschaft
latitude, -- Breitengrad
longitude -- Längengrad

From

tokens -- Tabelle namens „tokens“

Join

dateien using (**nr**) -- Tabelle mit Infos zu den Tonaufnahmen

Join

orte using (**ort**) -- Tabelle mit Orten und deren Geokoordinaten

Where

token rlike 'g.*r.*' -- Wörter, beginnend mit „g“, außerdem „r“ enthaltend
and core=1 -- Informant formuliert Wenkersatz
and sprecher=1 -- „Sprecher 1“ ist stets der Informant
and ws=4 -- Beschränkung auf Wenkersatz Nr. 4

Ergebnis der Abfrage: Verwendung des Wortes „Gurre“ in Wenkersatz 4

Zeige Datensätze 0 - 16 (17 insgesamt, die Abfrage dauert 0.3230 sek.)

token	phonetische Transkription	nr	intervall	alter	ort	latitude	longitude
'gøren		1004	22	NULL	Almen	46.0476318	24.43398839
'gøren		1005a	11	16	Almen	46.0476318	24.43398839
'gør		1282	16	54	Baaßen	46.19980269	24.28294165
'gør		1185a-02	11	44	Großprobstdorf	46.13999744	24.29290734
'güer		320-10	11	38	Großscheuern	45.85044968	24.15956847
'gør		255-04	12	48	Kerz	45.7887385	24.56354744
'gør		539b	11	42	Leschkirch	45.89577402	24.45653004
'güer		554a	16	34	Marpod	45.8718039	24.50157643
'gør		486-04	17	55	Martinsberg	45.85469669	24.67094128
'gør		486-04	18	55	Martinsberg	45.85469669	24.67094128
'gør		1369	10	NULL	Martinsdorf	46.0201028	24.40532664

intervals [17]:

xmin = 64.26604224948727

xmax = 77.80399006192299

Sekunden in
Audiodatei

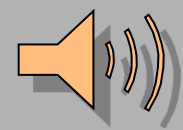
text = "## der 'geât 'ɔ:lt 'mu:n æs mæt 'ər mæt æi'gør durç dət 'ɛi:s æn 't 'kɔ:lt 'vɔ:sər gə'fɔln"

intervals [18]:

xmin = 77.80399006192299

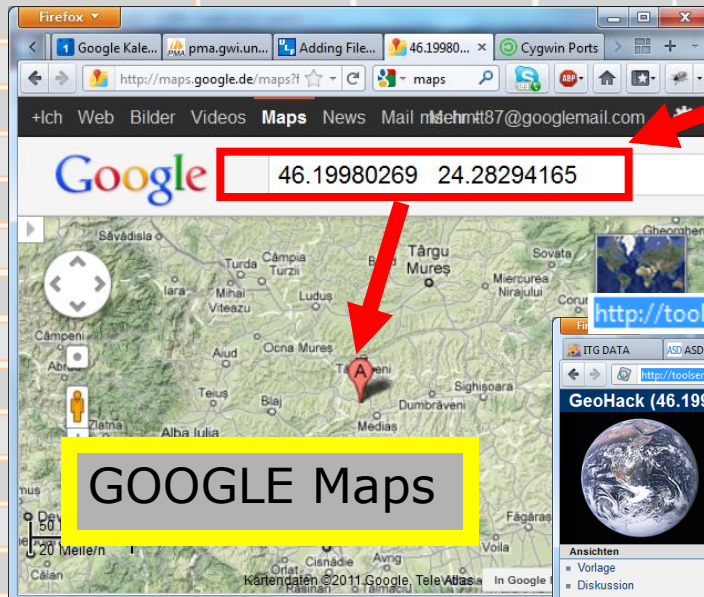
xmax = 93.76230348165966

text = "## der 'geât 'ɔ:lt 'mu:n æs mæt 'm mæt æi'gør durç dn 'ɛi:s gəb 'ænəgəbro:xən æn 't 'kɔ:lt vɔ ən æs æn 't 'kɔ:lt 'vɔ:sər gə'fɔln"



Georeferenzierung: Funktionsweise

token	phonetische Transkription	nr	intervall	alter	ort	latitude	longitude
'gøren		1004	22	NULL	Almen	46.0476318	24.43398839
'gøren		1005a	11	16	Almen	46.0476318	24.43398839
'gør		1282	16	54	Baaßen	46.19980269	24.28294165
'gør					roßprobstdorf	46.13999744	24.29290734
'güær					roßscheren	45.85044968	24.15956847
'gør					schkirch	45.7887385	24.51354744
'gør					arpod	45.89577402	24.45613004
'gær					artinsberg	45.8718039	24.50157613
'gør					artinsberg	45.85469669	24.6709412
'gør						45.85469669	24.6709412
'gør							
'gør							
'gør							
'gør							
'gøren							
'gør							
'gør							
'gør		834-03		12			



Georeferenzierung: Erzeugung einer kml-Datei

SQL-Abfrage

```
c:\cygwin\bin\bash --rcfile /cygdrive/z/.bashrc
1 select
2   dateien.ort -- Name der Ortschaft
3   latitude, -- Breitengrad
4   longitude, -- Längengrad
5   concat('Interview ',nr,':',token) -- Wort, auf das das in der where-Klausel
6 from
7   tokens -- Tabelle namens „tokens“
8 join
9   dateien using (nr) -- Tabelle mit Infos zu den Tonaufnahmen
10 join
11   orte using (ort) -- Tabelle mit Orten und deren Geokoordinaten
12 where
13   token rlike 'g.*r.*' -- Wörter, beginnend mit „g“, außerdem „r“ enthaltend
14   and core=1 -- Informant formuliert Wenkersatz
15   and sprecher=1 -- „Sprecher 1“ ist stets der Informant
16   and ws=4; -- Beschränkung auf Wenkersatz Nr. 4
```

SQL-Ergebnis

```
c:\cygwin\bin\bash --rcfile /cygdrive/z/.bashrc
1 Almen 46.0476318 24.43398839 Interview 1004: ggrdn
2 Almen 46.0476318 24.43398839 Interview 1005a: ggrdn
3 Baaßen 46.19980269 24.28294165 Interview 1282: ggr
4 Großprobstdorf 46.13999744 24.29290734 Interview 1185a-02: ggr
5 Großscheuern 45.85044968 24.15956847 Interview 320-10: ggr
6 Kerz 45.7887385 24.56354744 Interview 255-04: ggr
```

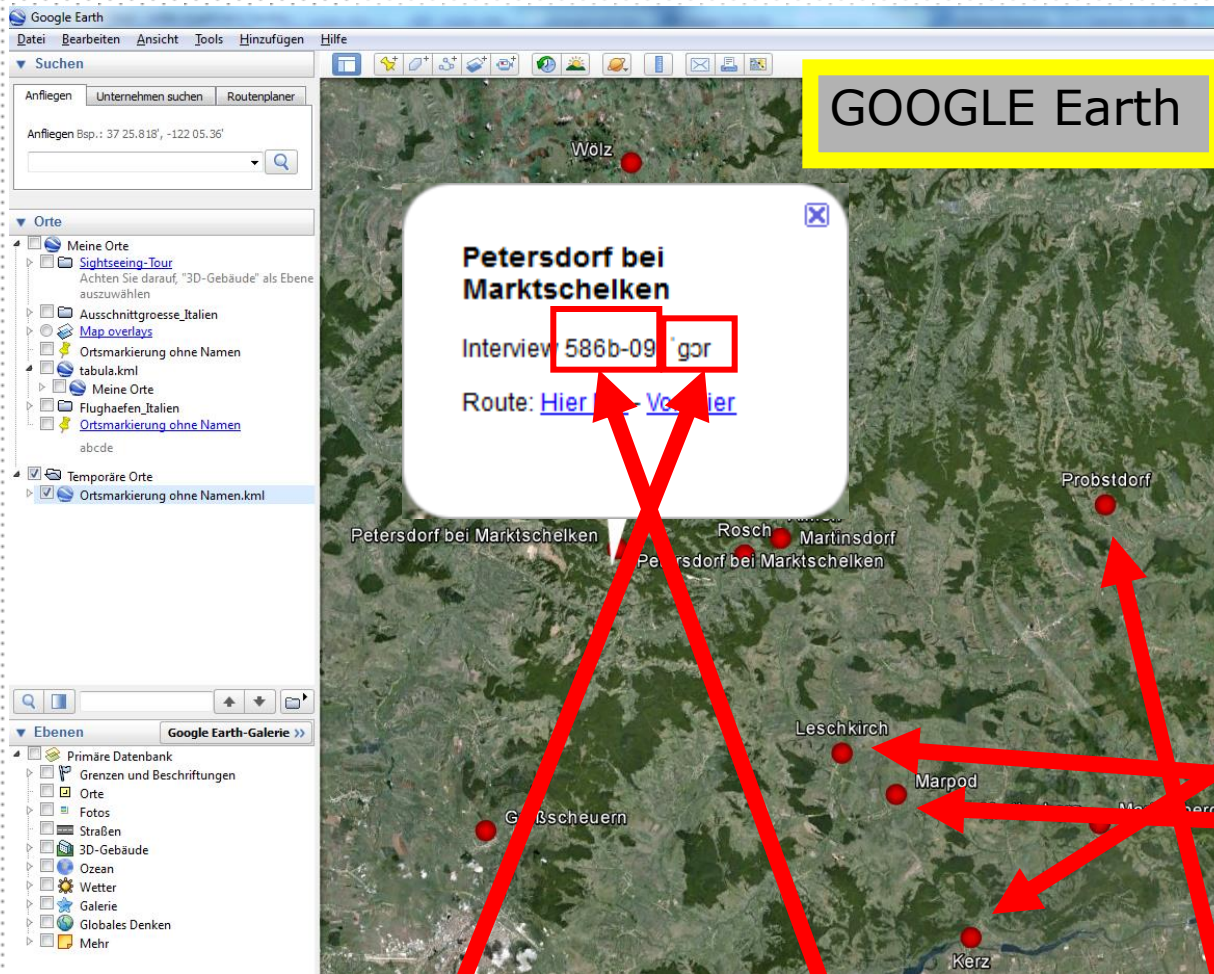
Awk-Skript

```
{
  Name=$1;
  lat=$2;
  long=$3;
  description=$4;
  print "<Placemark>
  <name>'Name'</name> \
  <styleurl>'msn_ylw-pushpin</styleurl> \
  <Point>
  <altitudeMode>clampToGround</altitudeMode> \
  <gx:altitudeMode>clampToSeaFloor</gx:altitudeMode> \
  <coordinates>'long','lat',0</coordinates> \
  </Point>
  <description>'description'</description> \
  </Placemark>";
}
END {
  print "</Document> \
  </kml>";
}
```

KML-Datei

```
c:\cygwin\bin\bash --rcfile /cygdrive/z/.bashrc
1 <?xml version="1.0" encoding="UTF-8"?> <kml xmlns="http://www.opengis.net/kml/2.2" xmlns:gx="http://www.google.com/kml/ext/2.2" xmlns:atom="http://www.w3.org/2005/Atom"> <Document>
2   <Placemark> <name>Almen</name> <coordinates>24.43398839,46.0476318,0</coordinates> </Point>
3   <Placemark> <name>Almen</name> <coordinates>24.43398839,46.0476318,0</coordinates> </Point>
4   <Placemark> <name>Baaßen</name> <coordinates>24.28294165,46.19980269,0</coordinates> </Point>
5   <Placemark> <name>Großprobstdorf</name> <coordinates>24.29290734,46.13999744,0</coordinates> </Point>
6   <Placemark> <name>Großscheuern</name> <coordinates>24.15956847,45.85044968,0</coordinates> </Point>
7   <Placemark> <name>Kerz</name> <coordinates>24.56354744,45.7887385,0</coordinates> </Point>
8   <Placemark> <name>Leschkirch</name> <coordinates>24.45653004,45.89577402,0</coordinates> </Point>
9   </Document>
10 </kml>
```

Georeferenzierung: KML-Datei in Google Earth



	ort	latitude
	Almen	46.0476318
	Almen	46.0476318
	Baaßen	46.19980269
	Großprobstdorf	46.13999744
	Großscheuern	45.85044968
	Kerz	45.7887385
	Leschkirch	45.89577402
	Marpod	45.8718039
	Martinsberg	45.85469669
	Martinsberg	45.85469669
	Martinsdorf	46.0201028
	Mortesdorf	46.06098516
'gnr'	Petersdorf bei Marktschelken	46.01412
'gnr'	Petersdorf bei Marktschelken	46.01412
'gnr'	Probstdorf	46.03917601
'gnr'	Rosch	46.010522
'gnr'	Wölz	46.23824455

'gnr'	20	11	5
'gnr'	586b-09	10	20
'gnr'	586b-09	10	20
'gnr'	852b-04	11	45
'gnr'	1389a-07	10	60
'gnr'	834-03	12	47

Kombination unterschiedlicher Ergebnisse

Gurre

Hengst

Wenkersatz 4 + `,g.*r'`

Wenkersatz 4 + `h.*t'`

SQL

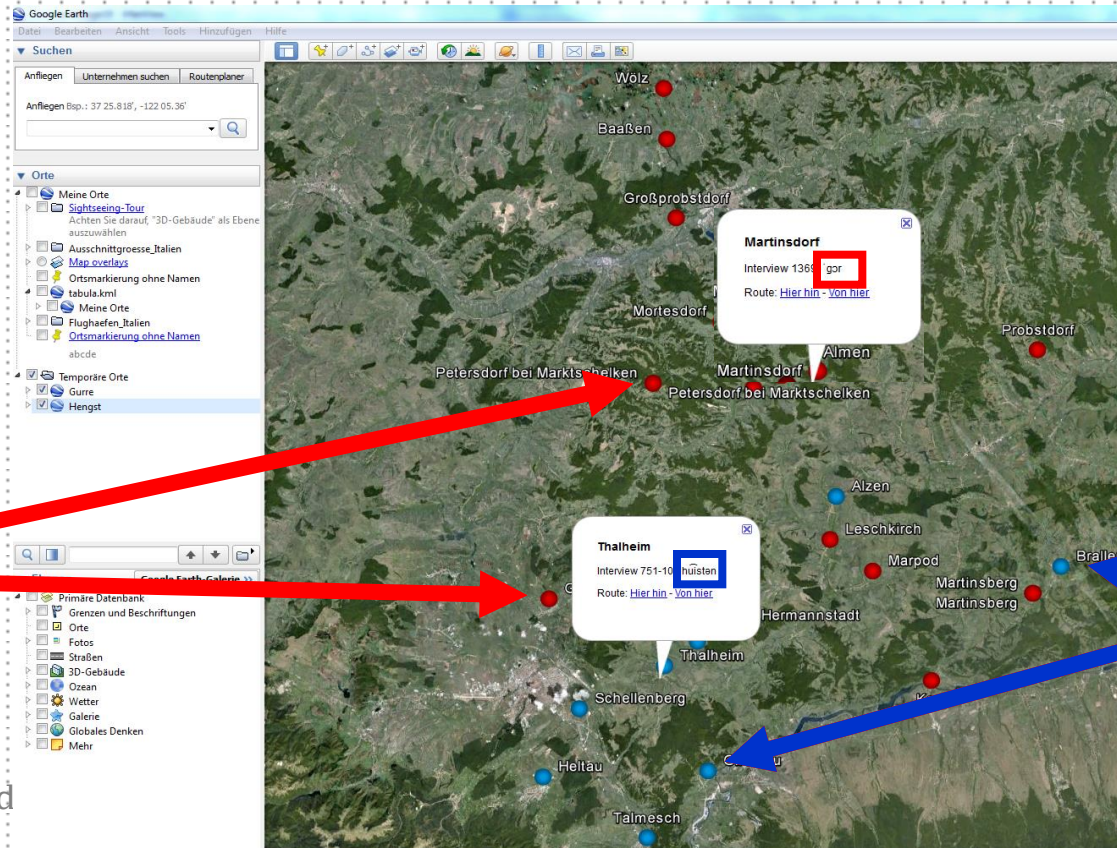
SQL

AWK

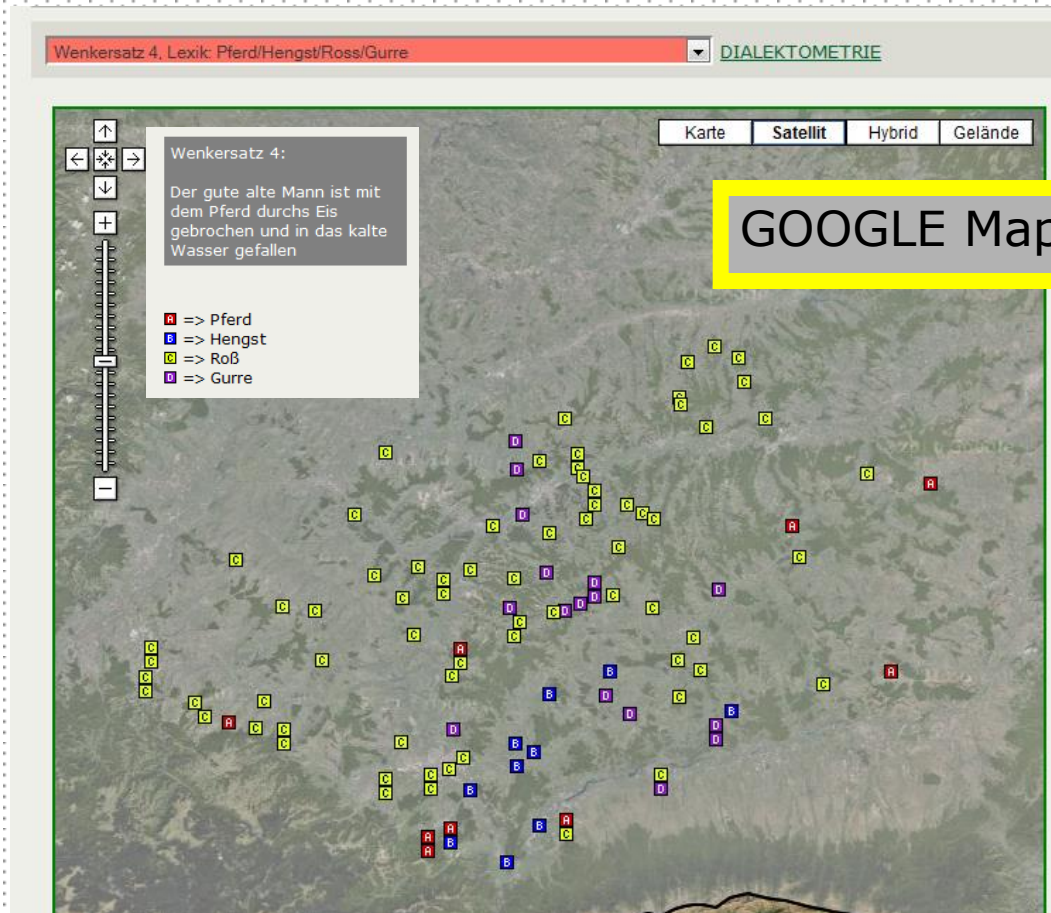
AWK

KML

KML



Kombination mehrerer Varianten auf einer Karte



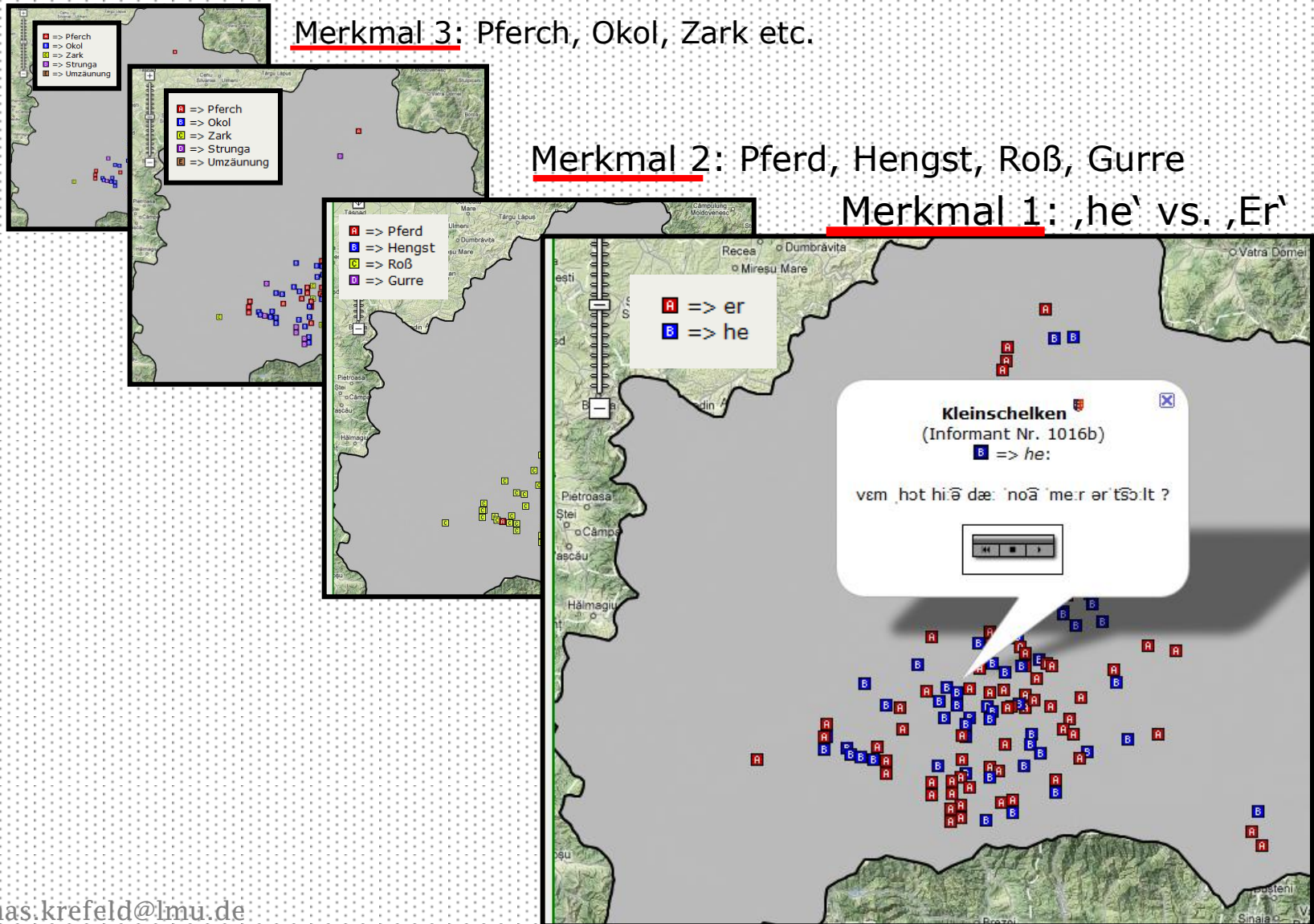
Analyse einer beliebigen Anzahl von Merkmalen

Merkmal n: beliebig aus den Bereichen Phonetik, Lexik, Morphosyntax

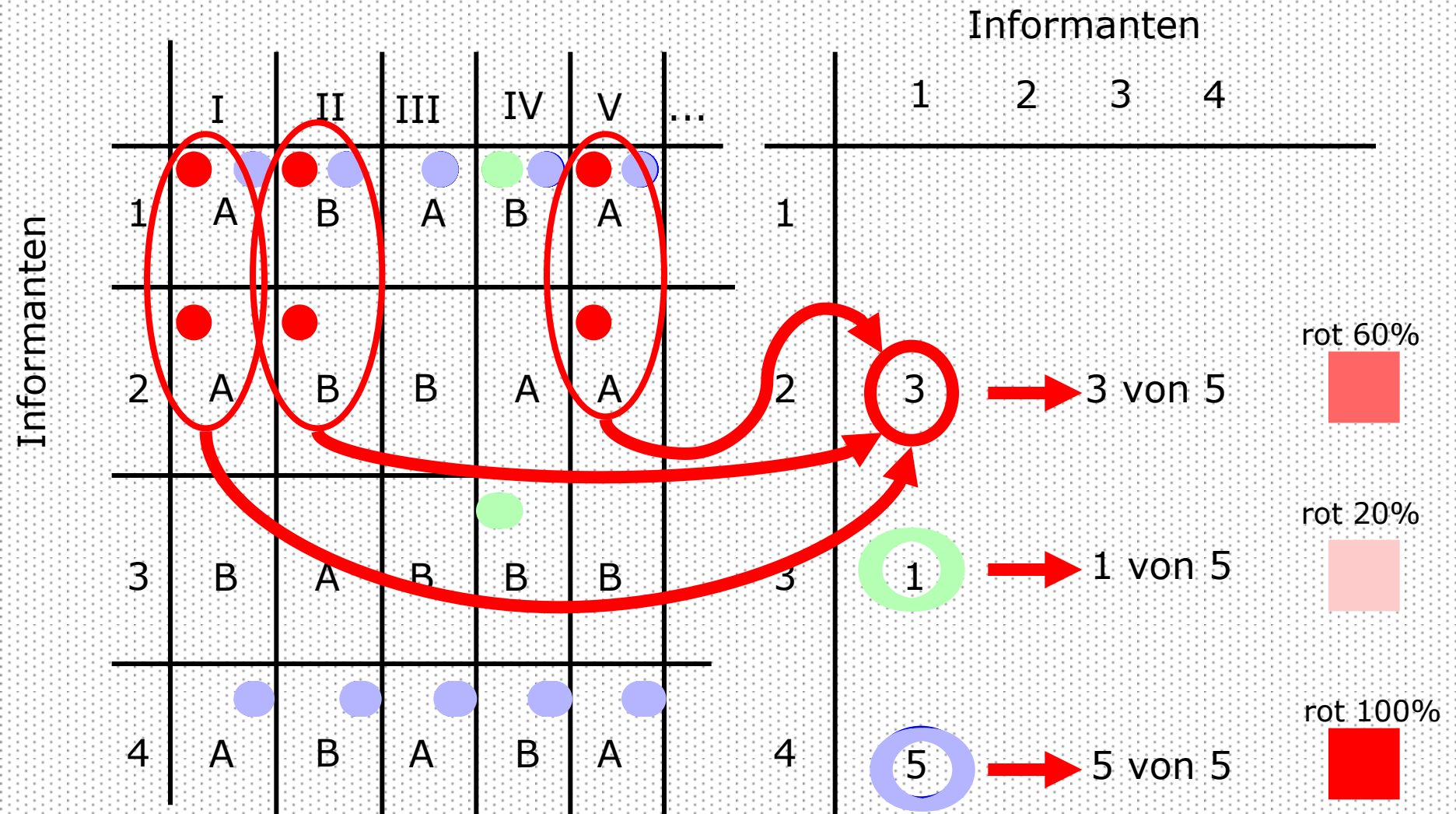
Merkmal 3: Pferch, Okol, Zark etc.

Merkmal 2: Pferd, Hengst, Roß, Gurre

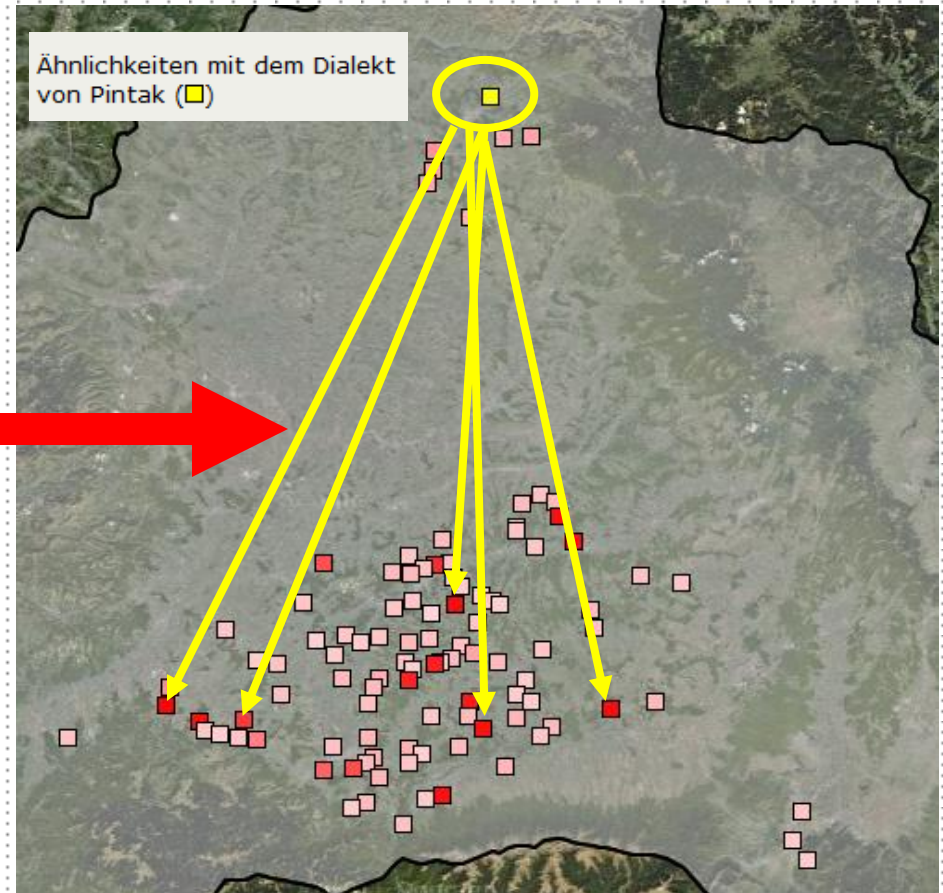
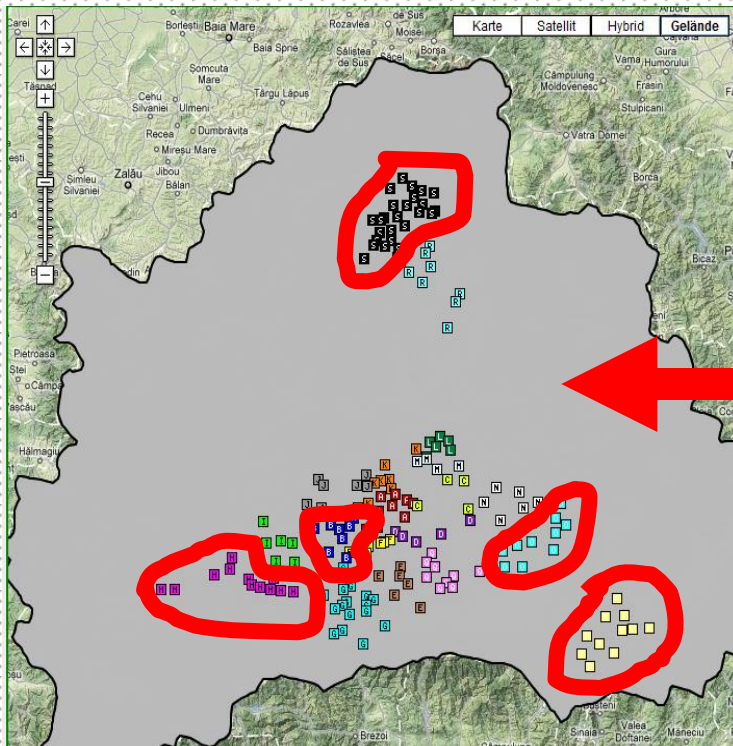
Merkmal 1: ‚he‘ vs. ‚er‘



Dialektometrie – Ermittlung des Ähnlichkeitsindex‘ Quantitative Analyse qualitativer Übereinstimmungen

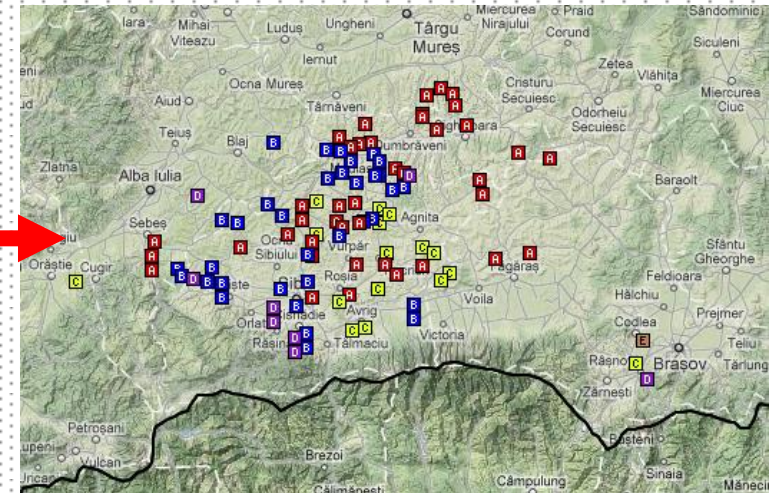
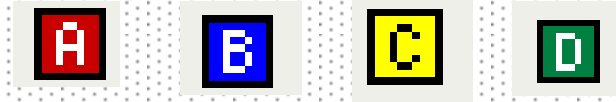
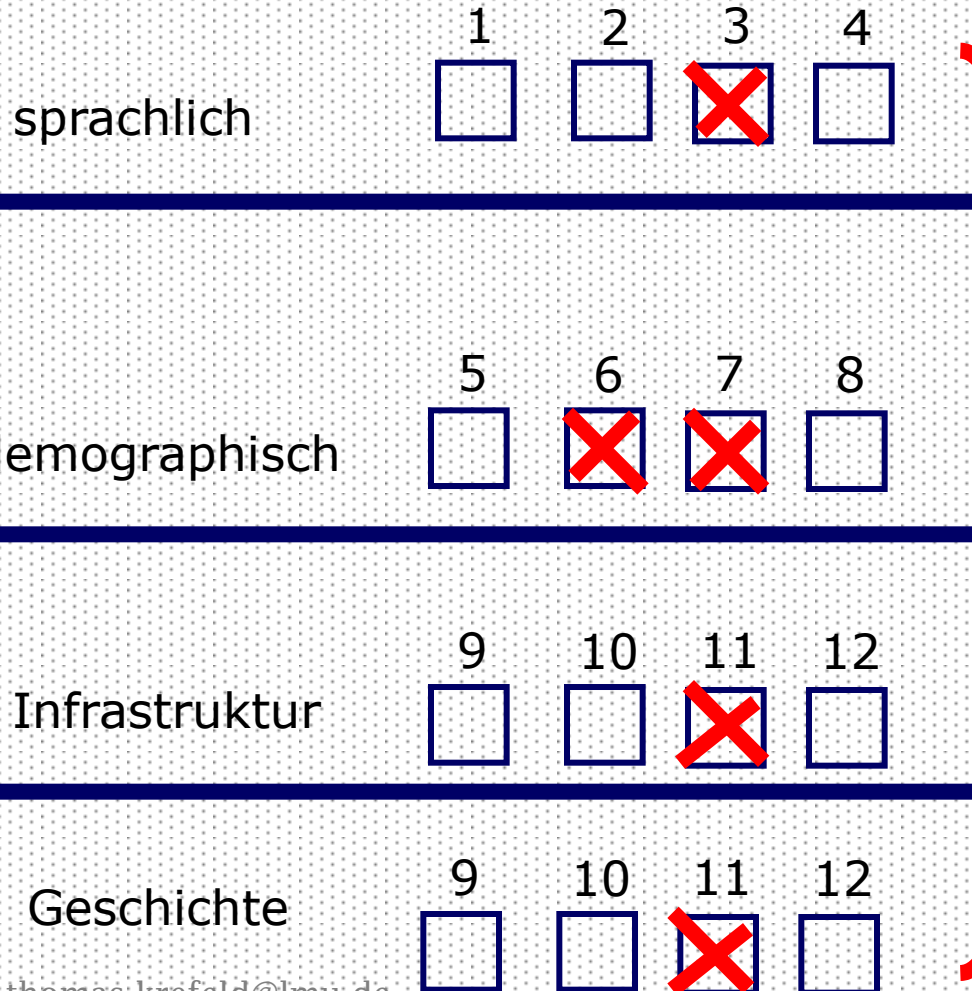


Visualisierung des Ähnlichkeitsindex (Beispiel Pintak)



Optionen I

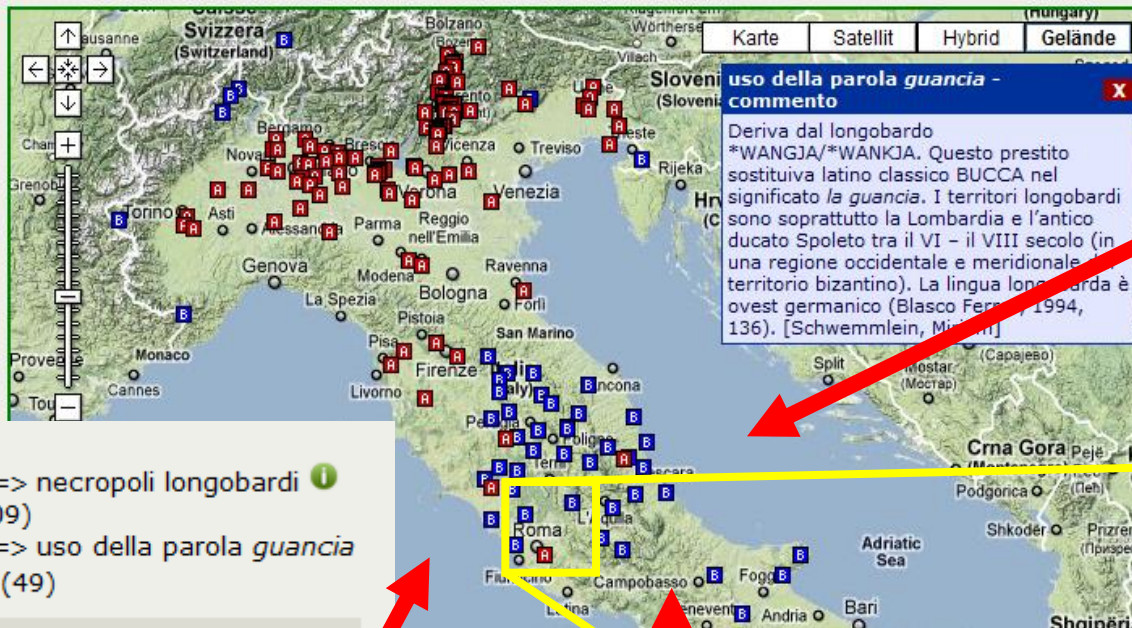
Freie Kombination von Merkmalen



Optionen I

Beispiel: Kombination sprachlicher mit außersprachlichen Realitäten

sinossi: uso della parola "gancia" e necropoli longobardi

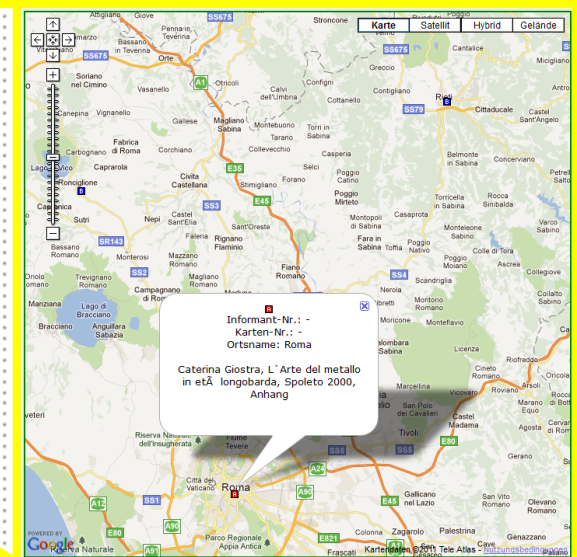


Demographie

- A** => necropoli longobardi (109)
- B** => uso della parola *gancia* (49)

Geschichte

Infrastruktur



Optionen II

Erweiterte Dokumentation des Kulturräumes: Bilder & Ton Einbeziehung von Experten und Laien mittels „social software“

The screenshot shows the Panoramio website interface. At the top, there's a navigation bar with 'Panoramio' logo, 'Registrieren', 'Hochladen', 'Gruppen', 'Wettbewerb', 'Orte', and 'Tags'. A search bar contains 'siebenbürgen' and a 'Ort suchen' button. Below the navigation, there's a banner for 'Panoramio Groups'. The main content area shows a photo of a street in Gura Raului, Romania, with a map overlay. The map shows the location of the photo, with a red pin and a red box around the 'Falsch platziert? Bessere Platzierung vorschlagen' button. A red arrow points from the photo to the map, and another red arrow points from the map to the button. A third red arrow points from the button to a red box at the bottom of the slide.



Falsch platziert? Bessere Platzierung vorschlagen

Kontakt & Links

- Johannes.Sift@web.de
- thomas.krefeld@lmu.de
- luecke@lmu.de

- <http://www.asd.gwi.uni-muenchen.de/>