

Arbeitstagung des
Audioatlas **S**iebenbürgisch-Sächsischer **D**ialekte

Symposion am 7./8.10.2013
Internationales Begegnungszentrum München (IBZ)

Die informatische Seite des Audio-Atlas Siebenbürgisch-Sächsischer Dialekte (ASD)

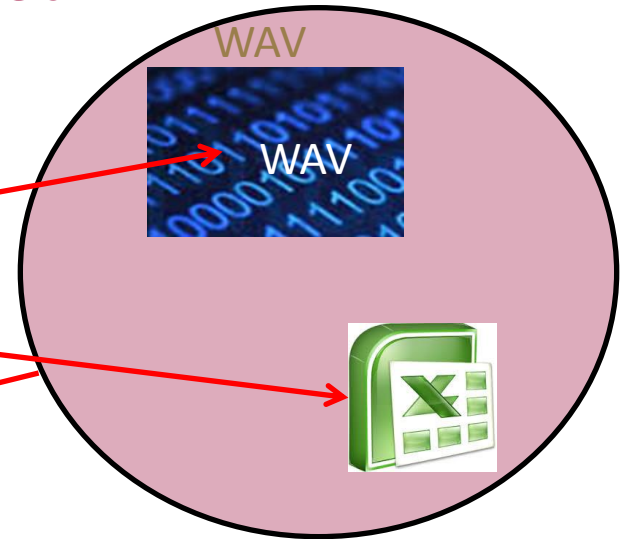
Stephan Lücke | Emma Mages

Die informatische Seite des ASD

1. Projektgeschichte
2. Erschließung des Materials – Von der Audiodatei zur Tabelle
3. Datenstrukturierung und -analyse
4. Benutzerfunktionen
5. Möglichkeiten und Perspektiven

Der Weg der Daten Von Gundelsheim nach München, vom Tonband ins Netz

Siebenbürgen-Institut
Gundelsheim

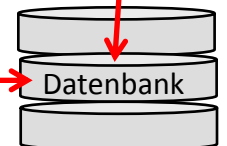


IKGS: Institut für Kultur und Geschichte Südosteuropas
München

IT-Gruppe Geisteswissenschaften der LMU
München



Transkription

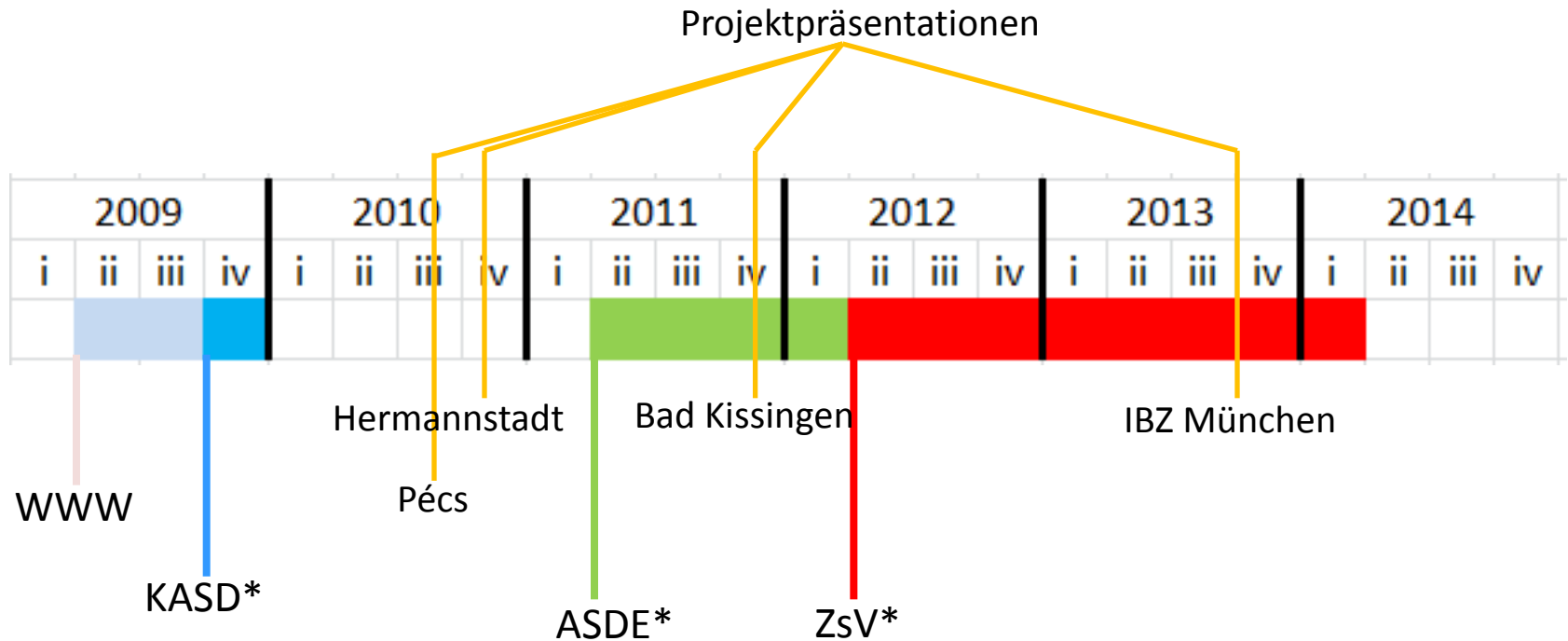


Die IT-Gruppe Geisteswissenschaften (ITG) der LMU

- Zentrale Einrichtung der 6 geisteswissenschaftlichen Fakultäten
- Seit über 15 Jahren Know-How in Projektentwicklung im Bereich der Digital Humanities
- Ausstattung mit modernen Serveranlagen
- Bestand langfristig abgesichert durch bislang 7 unbefristete Stellen
- Dadurch Garantie von „Nachhaltigkeit“ und „Langzeitarchivierung“
- www.itg.lmu.de



Projektphasen



* = Förderung durch den Bundesbeauftragten für Kultur und Medien (BKM)

Projektphasen:

WWW:

- Präsentation der Audiodateien im Internet
- einfache Suchfunktion auf Basis der Gundelsheimer Excel-Tabelle
- Georeferenzierung der Erhebungsorte

KASD:

- „Kleiner Audioatlas Siebenbürgisch-Sächsischer Dialekte“
- phonetische Transkription der Wenkersatzaufnahmen (Frau Klaster-Ungureanu)

ASDE:

- „Audioatlas Siebenbürgen – lexikalische Erweiterung“
- standardnahe Transkription spontansprachlicher Texte (Herr Schuller, Frau Maurer; Fließtext, insgesamt 475000 Wörter, entsprechend über 300 Normseiten)
- spezielles Transkriptionsverfahren: Balance zwischen Standardnähe und Dokumentation sprachlicher, d.h. dialektaler Besonderheiten
- Ziel: inhaltliche Tiefenerschließung durch Stichwortsuche

ZsV:

- „Zeitgeschichte und sprachliche Variation im Spiegel von Dialektaufnahmen aus Südosteuropa“
- phonetische Transkription von Spontansprache
- inhaltliche Tiefenerschließung durch Verschlagwortung („Ontologie“)
- geographische und sprachliche Ausweitung
- Mittelbayerisches Material aus dem Wassertal (Märchensammlung von Anton Ilk aus Alkoven)

Projektpräsentationen:

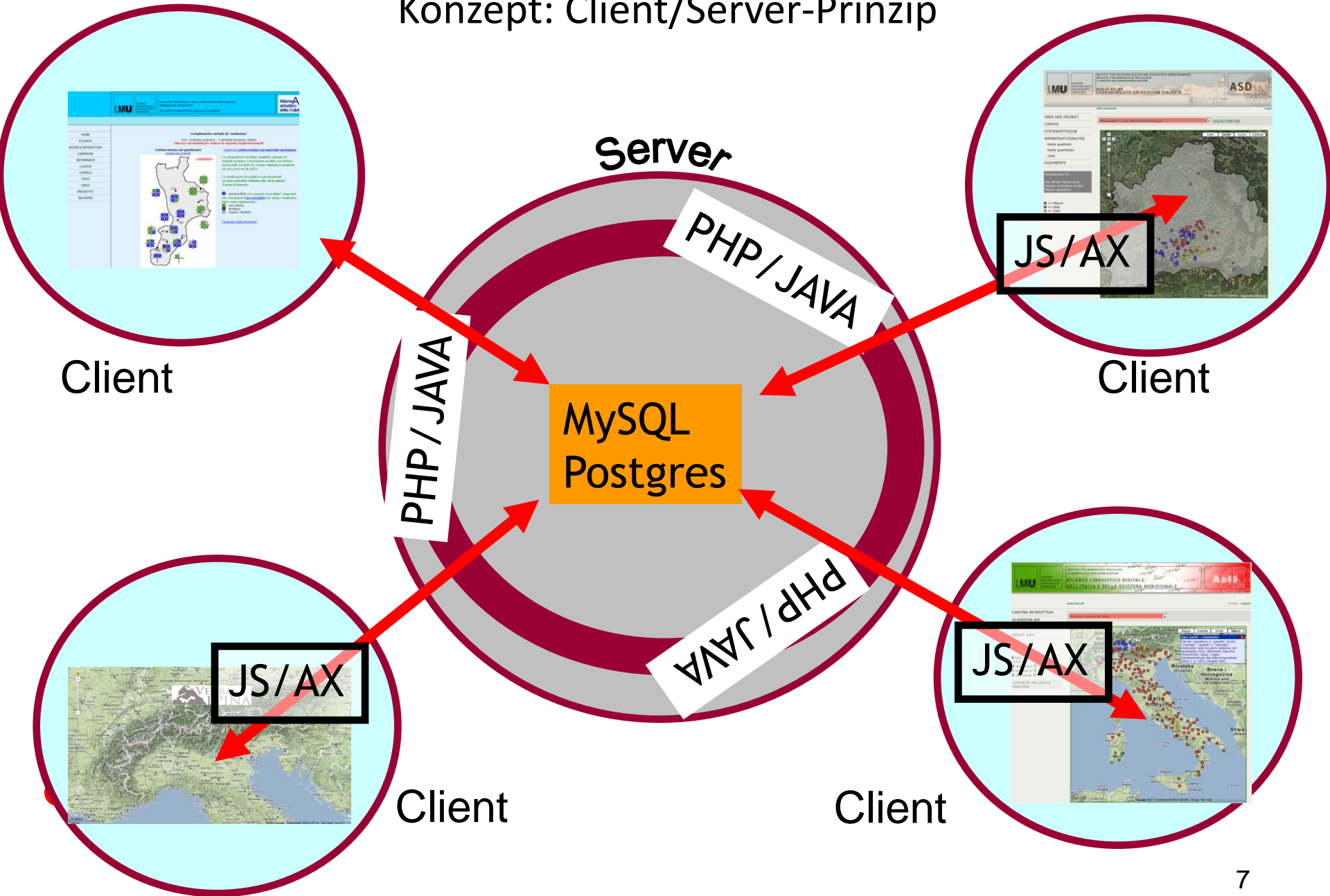
Pécs, 9.-11.09. 2010; Der Audio-Atlas Siebenbürgischsächsischer Dialekte (ASD-online) - und die Sprachgeographie nach der digitalen Revolution (Vortrag im Rahmen des Kongresses "Deutschsprachige Kultur und Literatur im regionalen und internationalen Kontext")

Hermannstadt: 18.-20.11.2010; Lucian-Blaga-Universität; Internationale Tagung "Interkulturelle Beziehungen im Spiegel der Sprache und Literatur. Dokumentation und Deutung"

Bad Kissingen: 2.-4.12.2011; Akademie Mitteleuropa e.V.; Internetseminar „Homepage und Heimat – Die Präsenz deutscher Geschichte und Kultur aus dem östlichen Europa im Internet“

IBZ München: 7./8.10.2013; Arbeitstagung des Audioatlas Siebenbürgisch-Sächsischer Dialekte

Konzept: Client/Server-Prinzip



Die informatische Seite des ASD

1. Projektgeschichte
2. Erschließung des Materials – Von der Audiodatei zur Tabelle
3. Datenstrukturierung und -analyse
4. Benutzerfunktionen
5. Möglichkeiten und Perspektiven

361 h in 2212 Dateien
aus 199 Orten

62 h orth.
transkribiert

48 h phon.

ca. 470.000
Tokens

ca. 204.000
Tokens

phonetisch

gelenkt

ca. 80.000 Tokens

spontansprachlich

ca. 124.000
Tokens

Wenkersätze
ca. 75.000 Tokens

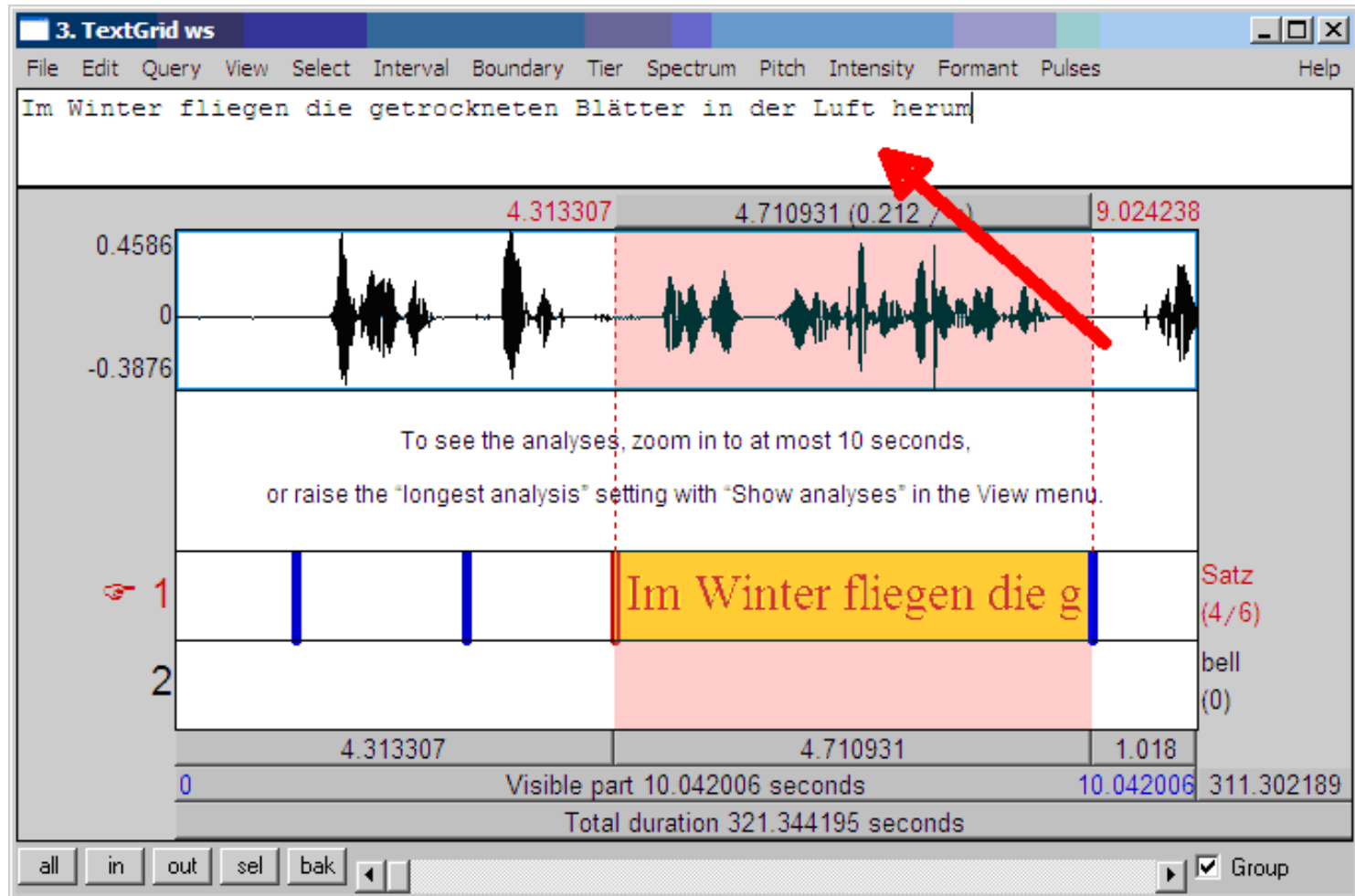
Hermannstädter
Mustersätze
ca. 4.500 Tokens

orthographisch

gelenkt
Hermannstädter
Mustersätze
ca. 2.700 Tokens

spontansprachlich
ca. 467.000 Tokens

Transkribieren mit Praat



Transkription

orthographisch

a-zA-Z# (< ASCII)

#1# Demnach sind unsere<as>
Sachsen vor dem <r>camin
cultural</r> gewesen<gewest> und
haben<hun> getanzt

Kennzeichnung von Dialekt- und
fremdsprachigen Passagen durch
xml-artige „Tags“

phonetisch

nur ASCII-Zeichen

```
\ef\c, b\æn d\swt \'1\iclz\ef  
\'1ki\:f\sh o\lius \'1h\ærm\sw\shtat  
\ef\c, \'2v\ærdn  
\'1\æ\c,t\sw\'2zi\li\swst\lis\ic\c,  
\'1jo\:fr \'2\ct\:flt \hsnt
```

```
εç bæ̃n dət 'ɪlzε 'ki:f ō̃us 'hæ̃rmə̃ftat εç  
,væ̃rdn 'æçtə ,zĩæ̃st̃s̃iç 'jo:r ,ɔ:lt unt
```

- einfache Eingabe über die Tastatur
- Nachhaltigkeit, Langzeitarchivierung

Vokale

	front		central		back	
<i>close</i>	i i	y y	ɨ ɨ	ɥ ɥ	ɯ ɯ	u u
<i>close centralized</i>	ɪ ɪc	ʏ ɥc			ʊ ʊs	
<i>close-mid</i>	e e	ø ø/	ə ə	ɘ ɘ	ɤ ɤ	o o
			ə ə			
<i>open-mid</i>	ɛ ɛf	œ œ	ɜ ɜ	ɞ ɞ	ʌ ʌ	ɔ ɔ
			ɚ ɚ			
	æ æ		ɐ ɐ			
<i>open</i>	a a	ɶ ɶe			ɑ ɑ	ɒ ɒ

Konsonanten

	bilabial	labiodental	dental	alveolar	alv. lateral	postalveolar	retroflex	alveolo-palatal	palatal	labial-palatal	labial-velar	velar	uvular	pharyngeal	epiglottal	glottal
<i>voiceless plosive</i>	p p			t t	t ^l t ^l		ʈ ʈ	c c				k k	q q	ʕ ʕ	ʔ ʔ	
<i>voiced plosive</i>	b b			d d	d ^l d ^l		ɖ ɖ	ɟ ɟ				g g	ŋ ŋ			
<i>nasal</i>	m m	ɱ ɱ		n n			ɳ ɳ	ɲ ɲ				ŋ ŋ	ɴ ɴ			
<i>voiceless fricative</i>	ɸ ɸ	f f	θ θ	s s	ʃ ʃ	ʂ ʂ	ç ç	ç̟ ç̟		ɬ ɬ	x x	ç ç	ħ ħ	ħ ħ	ħ ħ	
<i>voiced fricative</i>	β β	v v	ð ð	z z	ʒ ʒ	ʐ ʐ	ʝ ʝ	ʝ̟ ʝ̟			ɣ ɣ	ʁ ʁ	ʕ ʕ	ʕ ʕ	ʕ ʕ	
<i>approximant</i>				ɹ ɹ	ɻ ɻ		ɻ̟ ɻ̟	j j	ɥ ɥ	w w	ɥ ɥ					
<i>trill</i>	ʙ ʙ			r r												R R
<i>tap or flap</i>				ɾ ɾ	ɽ ɽ		ɽ̟ ɽ̟									
<i>lateral approx.</i>				l l	l ^l l ^l		ɭ ɭ	ɻ̟ ɻ̟				ɭ ɭ				
<i>implosive</i>	ɓ ɓ			ɗ ɗ				f̟ f̟				ɠ ɠ	ɠ̟ ɠ̟			
<i>click</i>	ǀ ǀ			ǃ ǃ	ǂ ǂ	ǁ ǁ	ǁ̟ ǁ̟									

TextGrid-Datei

```
1 File type = "ooTextFile"
2 Object class = "TextGrid"
3
4 xmin = 0
5 xmax = 2038.4892063492064 Länge der Aufnahme
6 tiers? <exists>
7 size = 1
8 item []:
9   item [1]:
10     class = "IntervalTier"
11     name = "satz"
12     xmin = 0
13     xmax = 2038.4892063492064
14     intervals: size = 127 Intervallzahl, Intervall-Nr.
15     intervals [1]:
16       xmin = 0
17       xmax = 20.299709172629175 Zeitmarken Intervalltext
18     text = "#1# ich ### bin geboren im Jahr 1911 den elften Monat und den elften Tag und heute hier in der Gemeinde
19     Oberneudorf<Noandrof> und heute sind wir bin ich im einundsechzigsten Lebensjahr"
20     intervals [2]:
21       xmin = 20.299709172629175
22       xmax = 39.04992892200463
23     text = "#1# wenn ich nun ein wenig<ekait> u#ber meinen Lebenslauf erza#hlen will ist das elfer Jahr wie mir mei
24     ne Eltern erza#hlt haben ein sehr reiches Jahr gewesen viel Obst hu#bsches Korn und alles damals war noch die o#sterr
25     eich ungarische Monarchie"
26     intervals [3]:
27       xmin = 39.04992892200463
28       xmax = 57.155116486499224
29     text = "#1# aber meine Kindheit war ja gar wenig in der Zeit dann ist der erste Krieg eingetroffen und wo die K
30     raft der Gemeinde hinausziehen musste ins Feld und ich als ein kleines Kind bei der Mutter und bei den Gros#eltern da
31     heim blieb"
32     intervals [4]:
33       xmin = 57.155116486499224
34       xmax = 74.54466691134206
35     text = "#1# auch das ist vergangen die Schule war schwach in der Zeit denn es war immer Mangel an Lehrern und w
36     enn man wollte gehen in die Schule ging man und wenn man nicht wollte blieb man eben daheim und durch das ist die Leh
37     re etwas zuru#ckgeblieben"
```

Vom TextGrid zur Tabelle

```
1
2 # gawk -f grid2tab.awk *TextGrid.utf8.orth.txt
3 # Skript durchläuft sukzessiv alle TextGrid-Dateien mit orthographischen Transkriptionen
4 BEGIN {
5   OFS = "\t";
6
7   printf "" >"intervalle.csv"; # leere Tabellen werden bereitgestellt
8   printf "" >"notizen.csv";
9   printf "" >"tags.csv";
10  printf "" >"tokens_standard.csv";
11  sprache="d";
12 }
13
14 # Textgrids werden an Tag-Grenzen gespalten, die dabei entstehenden Gruppen werden mit unterschiedlichem Glue zu Einheiten zusammengebunden
15 # (freie Tags mit '@', gebundene Tags mit '+', <D>-Gruppen und Tag-Übersetzungen mit '_', Notizen mit '$')
16 # und anschließend wieder zu ganzen Intervallen zusammengesetzt
17
18 {
19  ## Dialekt-Tag
20  #frei#
21  n=patsplit($0, arr, /[ \":]+<[^>*>*>+>[ \":]+/, seps);
22  for (i=0;i<n;i++) {
23    gsub(/ /,"@",arr[i]); # Bsp.: <kriglich@miglich>
24  }
25  $0="";
26  for (i=0;i<n;i++) {
27    $0=$0 arr[i] seps[i];
28  }
29
30  ## Dialekt-Tag
31  #gebunden#
32  n=patsplit($0, arr, /<[^>*>*>+>/, seps);
33  for (i=0;i<n;i++) {
34    gsub(/ /,"+",arr[i]); # Bsp.: <en+maschin+ge>in_einen_Lkw_zugeteilt</>
35  }
36  $0="";
37  for (i=0;i<n;i++) {
38    $0=$0 arr[i] seps[i];
39  }
40
41  ## Sprach-Tag: r,u,rus
42  #frei#
43  n=patsplit($0, arr, /[ \":]+<([ru])|(rus))>+<[^<\/>*>*>+<\/([ru])|(rus))>+/, seps);
44  for (i=0;i<n;i++) {
45    gsub(/ /,"@",arr[i]); # Bsp.: <r>secretar@de@baza</r>
46  }
47  $0="";
48  for (i=0;i<n;i++) {
49    $0=$0 arr[i] seps[i];
50  }
51 }
```

Intervalle werden an Tag-Grenzen gespalten

Vom TextGrid zur Tabelle

```
106 ### Tabellen ###
107
108 # Dateiname wird gelesen
109 if(FNR==1) {
110     split(gensub(/^.*[\/]/,"","g",FILENAME),tmp1,".");
111     my_FILENAME=tmp1[1];
112     print "Bearbeite Datei " my_FILENAME " ...";
113     # position=0;
114 }
115 # Intervall-Nr. wird extrahiert
116 if ($0 ~ /intervals/) {
117     gsub(/intervals /,"", $0);
118     gsub(/[\/]/,"", $0);
119     gsub(/[\/]/,"", $0);
120     gsub(/:/,"", $0);
121     interval=gensub(/ */,"","g", $0);
122 }
123 # Zeitmarken werden extrahiert
124 if ($0 ~ /xmin/) {
125     gsub(/xmin = /,"", $0);
126     xmin=gensub(/ */,"","g",sprintf("%.3f", $0));
127 }
128 if ($0 ~ /xmax/) {
129     gsub(/xmax = /,"", $0);
130     xmax=gensub(/ */,"","g",sprintf("%.3f", $0));
131 }
132 # Intervalltext wird extrahiert und in Tabelle 'intervalle' geschrieben
133 if ($0 ~ /text = /) {
134     gsub(/\\s*text = /,"", $0);
135     gsub(/"/,"", $0);
136     gluetext=gensub(/</," <","g",gensub(/>/,"> ","g",gensub(/@</," <","g",gensub(/>@/,"> ","g", $0))););
137     print my_FILENAME, interval, xmin, xmax, gensub(/[\+@_]/," ","g",gluetext) >>"intervalle.csv"; # Schreiben in Tabelle intervale
138     gsub(/\\s*/,"",gluetext);
139     gsub(/\\s+/, " ",gluetext);
140     gsub(/\\s*$/,"",gluetext);
141     # intervallinterne Zähler zur Zuordnung von Tokens und Tags werden gesetzt
142     positionImIntervall=0;
143     idGruppe=0;
144     # das Intervall wird an Leerzeichen gespalten, das entstehende Array durchlaufen und jeweils 'gluetoken' und 'cleantoken' (ohne Tags) erstellt
145     n=split(gluetext,arr,/ [ ]/);
146     for (i=1;i<=n;i++) {
147         glossiert=0;
148         # gluetoken und cleantoken werden parallel durchlaufen, ihre Zählung entspricht sich, so dass Tag- und Tokengruppen einander zugeordnet werden können
149         gluetoken=arr[i];
150
151         cleantoken=gensub(/<\/>/,"","g",gensub(/<[^>\\V]+$/,"\\1","g",gensub(/([^{1})<[^>\\V]+$/,"\\1","g",gensub(/([^{1})<([rud])(rus)+>[<^>]+<\/([rud])(rus)+>/,"\\1","g",gensub(/<<([rud])(rus)+>[<^>\\V]+<\/([rud])(rus)+>>/,"","g",arr[i])))););
152     }
```

Metadaten werden extrahiert

Tabelle 'intervalle' wird geschrieben

Intervalle werden tokenisiert

Vom TextGrid zur Tabelle

```
153
154 # Notizen werden extrahiert
155 if(match(gluetoken,/{[^{}]+}/)) {
156     ntz=substr(gluetoken,RSTART,RLENGTH);
157     if(ntz~/</>/) {
158         notiz=gensub("/{}/", "", "g", gensub("/{}/", "\\1", "g", gensub("/{}/", " ", "g", substr(ntz,RSTART,RLENGTH)))
159         print my_FILENAME, interval, notiz >>"notizen.csv"; # Schreiben in Tabelle notizen
160     }
161 }
162 # Tag-Gruppen, die mit Übersetzung stehen, werden glossiert und als idGruppe gezählt (die entsprechenden Tokens erhalten die gleiche idGruppe)
e)
163 # glossierte Teile werden wiederum zerlegt und Gruppen-Positions-Nummern vergeben
164 # das Ergebnis wird in die Tabelle 'tags' geschrieben
165 else if(match(gluetoken,/<<([rud])(rus)+>>[<>]+<\/([rud])(rus)+>>/)) {
166     glossiert=0;
167     sprache="0";
168     idGruppe="0";
169     positionInDerGruppe=0;
170 }
171 else if(match(gluetoken,/<<([rud])(rus)+>>[<>\\V]+<\/([rud])(rus)+>>/)) {
172     glossiert=1;
173     sprache="0";
174     ++idGruppe;
175     positionInDerGruppe=0;
176     tmp=substr(gluetoken,RSTART,RLENGTH);
177     o=split(tmp,glosspart,"+");
178     for (ii=1;ii<=o;ii++) {
179         tag=gensub("/{}/", "<<", "g", gensub("/{}/", ">>", "g", gensub("/{}/", " ", "g", glosspart[ii]));
180         print my_FILENAME, interval, idGruppe, ++positionInDerGruppe, tag, sprecher, sprache >>"tags.csv"; # Schreiben in Tabelle tags
181     }
182 }
183 else if(match(gluetoken,/[^ ]{1}<<([rud])(rus)+>>[<>]+<\/([rud])(rus)+>>/)) {
184     glossiert=1;
185     sprache="0";
186     ++idGruppe;
187     positionInDerGruppe=0;
188     tmp=substr(gluetoken,RSTART+1,RLENGTH-1);
189     o=split(tmp,glosspart,"+");
190     for (ii=1;ii<=o;ii++) {
191         tag=gensub("/{}/", " ", "g", glosspart[ii]);
192         print my_FILENAME, interval, idGruppe, ++positionInDerGruppe, tag, sprecher, sprache >>"tags.csv"; # Schreiben in Tabelle tags
193     }
194 }
195
196 else if(match(gluetoken,/<[<>\\V]+>/)) {
197     glossiert=1;
198     sprache="s";
199     ++idGruppe;
200     positionInDerGruppe=0;
201     tmp=substr(gluetoken,RSTART,RLENGTH);
202     o=split(tmp,glosspart,"+");
203     for (ii=1;ii<=o;ii++) {
204         tag=gensub("/{}/", "<>", "g", gensub("/{}/", " ", "g", glosspart[ii]));
205         print my_FILENAME, interval, idGruppe, ++positionInDerGruppe, tag, sprecher, sprache >>"tags.csv"; # Schreiben in Tabelle tags
206     }
207 }
```

Gruppen-Positions-Nr. werden vergeben

Tabelle 'tags' wird geschrieben

Vom TextGrid zur Tabelle

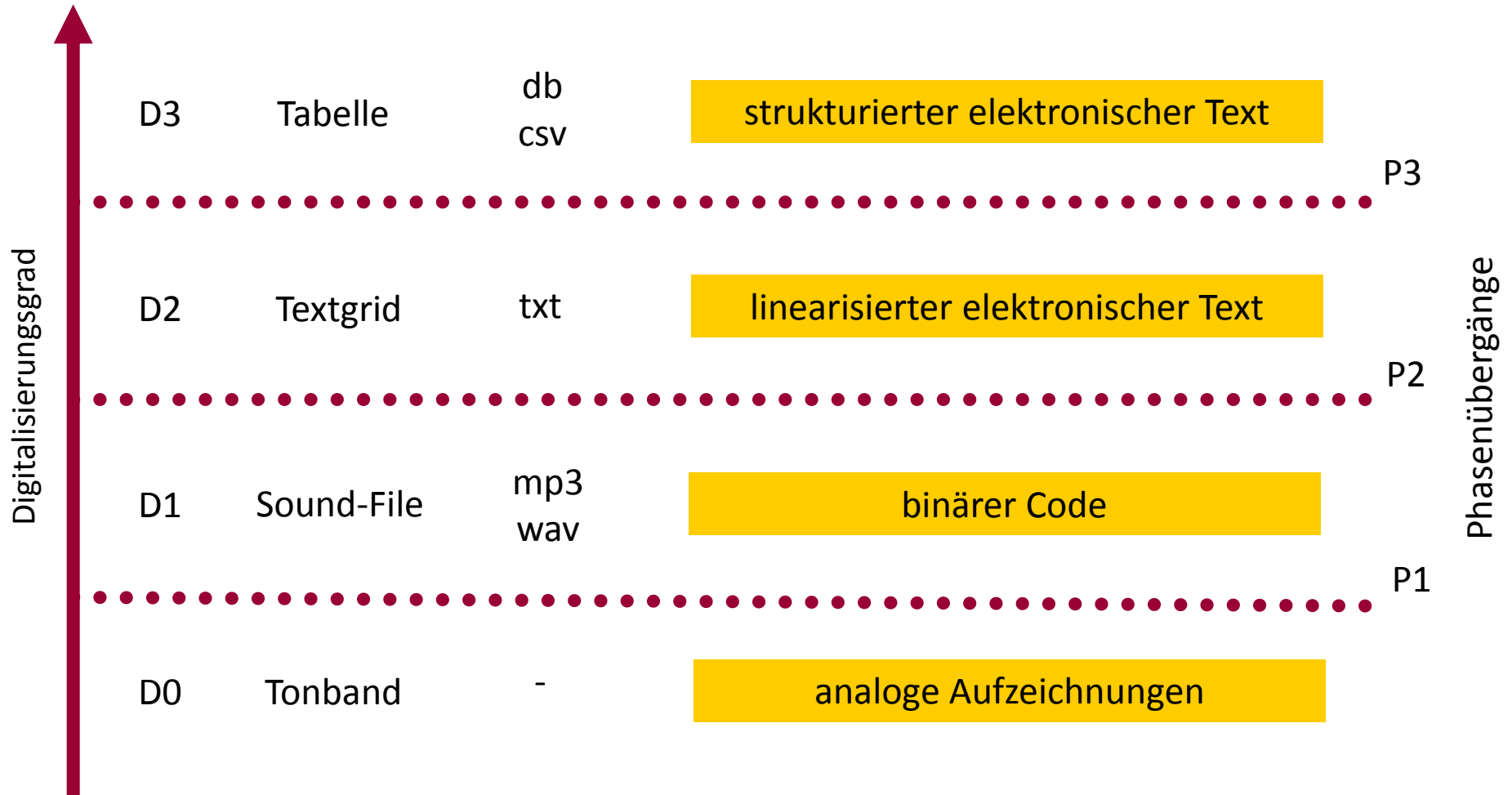
```
208
209 sprache="d"; # Defaultsprache in orth. Transkriptionen deutsch, wird in weiterem Schritt angepasst (setLanguageTokens.awk, setLanguageTags.a
wk)
210
211 # SPRECHER
212 if(match(cleantoken,/^[^ #]+#/)) {
213     sprecher=substr(cleantoken,RSTART+1,RELENGTH-2);
214 }
215 else if(match(cleantoken,/([^aousAOUS]{1})#[^ #]+#/)) {
216     sprecher=substr(cleantoken,RSTART+2,RELENGTH-2);
217 }
218 else if((match(cleantoken,/{[^{}]+/))) {
219     ntz=substr(cleantoken,RSTART,RELENGTH);
220     if(ntz!~/::/) {
221         notiz=gensub(/$/, " ", "g", gensub(/[{}/]/, "", "g", gensub(/?([{}])/, "\\1", "g", substr(cleantoken,RSTART,RELENGTH))));
222         print my_FILENAME, interval, notiz >>"notizen.csv"; # Schreiben in Tabelle notizen
223     }
224 }
225 else if(cleantoken~/[_@]/) {
226     nn=split(cleantoken,tk,/[_@]/);
227     for(ii=1;ii<=nn;ii++) {
228         cleantoken=tk[ii];
229         print my_FILENAME, interval, idGruppe, ++positionImIntervall, cleantoken, sprecher, sprache >>"tokens_standard.csv"; # Schreiben in Tabell
e tokens
230     }
231 }
232 else if (cleantoken!~/[_@]/) {
233     if(glossiert==1) {
234         print my_FILENAME, interval, idGruppe, ++positionImIntervall, cleantoken, sprecher, sprache >>"tokens_standard.csv"; # Schreiben in Tabell
e tokens
235     }
236     else {
237         print my_FILENAME, interval,          0, ++positionImIntervall, cleantoken, sprecher, sprache >>"tokens_standard.csv"; # Schreiben in Tabelle
tokens
238     }
239 }
240 }
241 }
242 }
243 }
```

Tabelle 'tokens_standard' wird geschrieben

csv-Dateien > sql-Dateien > Datenbank

nr	intervall	idGruppe	positionImIntervall	token	sprecher	sprache
1	109	1	0	ich	1	d
2	109	1	1	heis#e	1	d
3	109	1	2	###	1	d
4	109	1	3	und	1	d
5	109	1	4	bin	1	d
6	109	1	5	aus	1	d
7	109	1	6	Heltau	1	d
8	109	1	7	mein	1	d
9	109	1	8	Alter	1	d
10	109	1	9	ist	1	d
11	109	1	10	65	1	d
12	109	1	11	Jahre	1	d
13	109	1	12	ich	1	d
14	109	2	0	will	1	d
15	109	2	1	einen	1	d
16	109	2	2	kleinen	1	d
17	109	2	3	U#berblick	1	d
18	109	2	4	u#ber	1	d
19	109	2	5	die	1	d
20	109	2	6	Gemeinde	1	d
21	109	2	7	Heltau	1	d
22	109	2	8	aus	1	d
23	109	2	9	alter	1	d
24	109	2	10	Zeit	1	d
25	109	2	11	geben	1	d
26	109	2	12	die	1	d
27	109	2	13	Gemeinde	1	d
28	109	2	14	Heltau	1	d
29	109	2	15	hatte	1	d
30	109	2	16	die	1	d
31	109	2	17	zweitgro#s#te	1	d
32	109	2	18	Gemarkung	1	d
33	109	2	19	neben	1	d
34	109	2	20	der	1	d
35	109	2	21	Gemeinde	1	d
36	109	2	22	Zeiden	1	d
37	109	2	23	aus	1	d
38	109	2	24	Siebenbu#rgen	1	d
39	109	2	25	die	1	d
40	109	3	1	Gemarkung	1	d
41	109	3	2	war	1	d
42	109	3	3	verha#ltnisma#sig	1	d
43	109	3	4	arm	1	d
44	109	3	5	und	1	d
45	109	3	6	aus	1	d
46	109	3	7	dem	1	d
47	109	3	8	Grund	1	d
48	109	3	9	waren	1	d
49	109	3	10	die	1	d
50	109	3	11			

Grade der Sprachdigitalisierung



Die informatische Seite des ASD

1. Projektgeschichte
2. Erschließung des Materials – Von der Audiodatei zur Tabelle
3. Datenstrukturierung und -analyse
4. Benutzerfunktionen
5. Möglichkeiten und Perspektiven

		Token-ID								Ontologie-ID		
<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		
		270625	1522-05	15	0	39	1	das	d			
		270626	1522-05	15	0	40	1	Korn	d	B.01	→	Ackerbau und Feldwirtschaft
		270627	1522-05	15	0	41	1	geschnitten	d			
		270628	1522-05	15	0	42	1	war	d			
		270629	1522c-07	1	0	1	1	###	d			
		270630	1522c-07	1	0	2	1	Kallesdorf	d	I, I.02, I.02.01	→	Geographie
		270631	1522c-07	1	0	3	1	fu#nfzig	d			
		270632	1522c-07	1	0	4	1	Jahre	d	A.04		
		270633	1522c-07	1	0	5	1	Raion	d			
		270634	1522c-07	1	0	6	1	Bistritz	d	I, I.02, I.02.01		
		270635	1522c-07	2	0	1	1	die	d			
		270636	1522c-07	2	0	2	1	Kallesdorfer	d	I, I.02, I.02.01		
		270637	1522c-07	2	0	3	1	Kollektivwirtschaft	d	H.03.01	→	Kollektivwirtschaft und Staatsgut
		270638	1522c-07	2	0	4	1	hat	d			
		270639	1522c-07	2	0	5	1	sich	d			
		270640	1522c-07	2	0	6	1	nun	d			
		270641	1522c-07	2	0	7	1	vereinigt	d			
		270642	1522c-07	2	0	8	1	vor	d			
		270643	1522c-07	2	0	9	1	drei	d			
		270644	1522c-07	2	0	10	1	Wochen	d	A.04	→	Tageszeiten und Jahreszeiten
		270645	1522c-07	2	0	11	1	mit	d			
		270646	1522c-07	2	1	12	1	Baierdorf	d	I, I.02, I.02.01		
		270647	1522c-07	2	2	13	1	Reus#en	d			
		270648	1522c-07	2	0	14	1	wir	d			
		270649	1522c-07	2	0	15	1	haben	d			
		270650	1522c-07	2	0	16	1	einen	d			
		270651	1522c-07	2	0	17	1	Plan	d			

	id_ont	schlagwoerter
<input type="checkbox"/>    	A.	Natur
<input type="checkbox"/>    	A.01	Flora
<input type="checkbox"/>    	A.02	Fauna
<input type="checkbox"/>    	A.03	Wetter und Naturgewalten
<input type="checkbox"/>    	A.04	Tageszeiten und Jahreszeiten
<input type="checkbox"/>    	A.04.01	Woche und Wochentage
<input type="checkbox"/>    	A.04.02	Monate
<input type="checkbox"/>    	B.	Landwirtschaft
<input type="checkbox"/>    	B.01	Ackerbau und Feldwirtschaft
<input type="checkbox"/>    	B.02	Viehzucht (auch Kleinvieh und Hirten)
<input type="checkbox"/>    	B.03	Imkerei
<input type="checkbox"/>    	B.04	Weinbau
<input type="checkbox"/>    	B.05	Garten- und Obstbau
<input type="checkbox"/>    	B.06	Ernte
<input type="checkbox"/>    	B.07	Landwirtschaftliche Gera#te
<input type="checkbox"/>    	C.	Haus, Haushalt, Hof (auch Geba#ude)
<input type="checkbox"/>    	C.01	Alltag(sarbeit)
<input type="checkbox"/>    	C.02	Handarbeiten
<input type="checkbox"/>    	C.03	Kulinarisches
<input type="checkbox"/>    	C.03.01	Kochen
<input type="checkbox"/>    	C.03.02	Backen

Datenstrukturierung (Beispiel)

```
text = "{WS 26} 'zi^azent^svint^siç"  
intervals [72]:  
  xmin = 207.69182984041652  
  xmax = 213.13286209442822  
text = "## ho^inder a:zæm 'ha^is fte^o:n dro^a 'o^ipælbo^im mæt 'rydn ,o^ipæln"  
intervals [73]:  
  xmin = 207.69182984041652  
  xmax = 213.13286209442822
```

nr	intervall	xmin	xmax	id_token	ws	token	ort	latitude	longitude
1449b-05	72	207.6918298404165	213.13286209442825	47253	26	ho^inder	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47254	26	a:zæm	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47255	26	'ha^is	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47256	26	fte^o:n	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47257	26	dro^a	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47258	26	'o^ipælbo^im	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47259	26	mæt	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47260	26	'rydn	Keisd	46.19482874	24.96494868
1449b-05	72	207.6918298404165	213.13286209442825	47261	26	'o^ipæln	Keisd	46.19482874	24.96494868

Georeferenzierung

Qualitative Datenanalyse – Beispiel WS

26

„Hinter unserem Haus stehen drei schöne Apfelbäume mit roten Äpfelchen“

ws	nr	token	kontext
26	90-02	● 'ry:t	'hæŋdər əzəm 'hōis' ʃto:n drʊə 'hi:f 'ɔ:pəlbi:m mæt 'ry:t əpeltʃər
26	95	● 'rəo:di	hɪntə ɪnzəm 'hɑ:əs ʃte:n traɪ 'ʃe:ne 'epfəlba:mə mit 'rəo:di 'epfələn
26	100	● 'ri:dn	'hæŋdər əzəm 'hōus ʃto:n dro: 'hi:f 'ɔ:pəlbi:m mæt 'ri:dn əpəl
26	118-03	● 'ri:dən	'hæŋdər əzəm 'hēous ʃto:n droi 'hi:əʃ 'apəlbi:m mæt 'ri:dən 'apeltʃəm
26	120-05	● 'ry:t	'hæŋdər əzəm 'hēous ʃto:n droə 'hi:əʃ 'apəlbi:m mæt 'ry:t əpeltʃər
26	127	● 'ry:t	'hæŋdər əzəm 'hēos ʃto:n dro: 'hi:əʃ 'apəlbi:m mæt 'ry:t 'apeltʃər
26	135-02	● 'ri:dn	'hæŋdər əzəm 'hōus ʃto:n dro: 'hi:f 'apəlbi:m mæt 'ri:dn 'æ:peltʃəm
26	144	● 'ri:dn	hɔŋdʒər əzərər əzər 'ʃtʊf ʃte:n drəʊ 'hi:əʃ 'æ:pləbi:m mæt 'ri:dn 'æ:peltʃəm
26	197	● 'rykt	hɔŋdʒər əzəm 'hāis ʃto:n dro: 'hi:f 'apəlbi:m mæt 'rykt əpəl
26	200	● 'ri:t	'hæŋdər əzəm 'hēos ʃto:n dro: 'hi:əʃ 'apəlbi:m mæt 'ri:t 'apeltʃər
26	210	● 'ri:t	hɔɪndər əzər 'ʃtʊf ʃte:n drʊ:ə 'apəlbaɪəm mæt 'hɔɪ 'ri:t əpəl

- roten
- rot
- rote



Qualitative Datenanalyse – Beispiel WS 26

Variante A: „roten“



- Umgangssprachlich: „Finde die Informanten, die im WS 26 ‚roten‘ gesagt haben“
- Technisch (SQL): „select nr from tokens_phon where ws=26 and token rlike '^r.*n\$';“
- Ergebnis:

anzahl	ergebnis
69	566b-03, 118-03, 135-02, 255-04, 261-05, 261-05, 486-04, 493-02, 502-11, 522-02, 59 ...

\$ = Ende des Wortes

Qualitative Datenanalyse – Beispiel WS 26

Variante B: „rot“

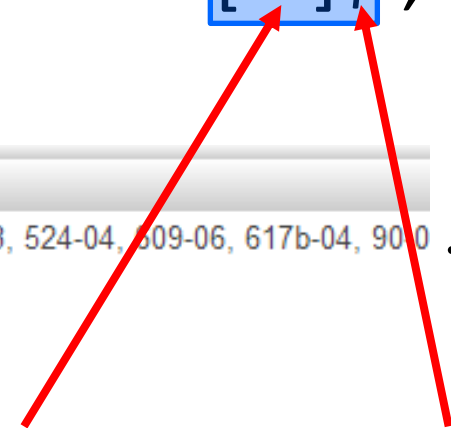


- Umgangssprachlich: „Finde die Informanten, die im WS 26 ‚rot‘ gesagt haben“
- Technisch (SQL): „select nr from tokens_phon where ws=26 and token rlike '^'r.*[td]\$'“;
- Ergebnis:

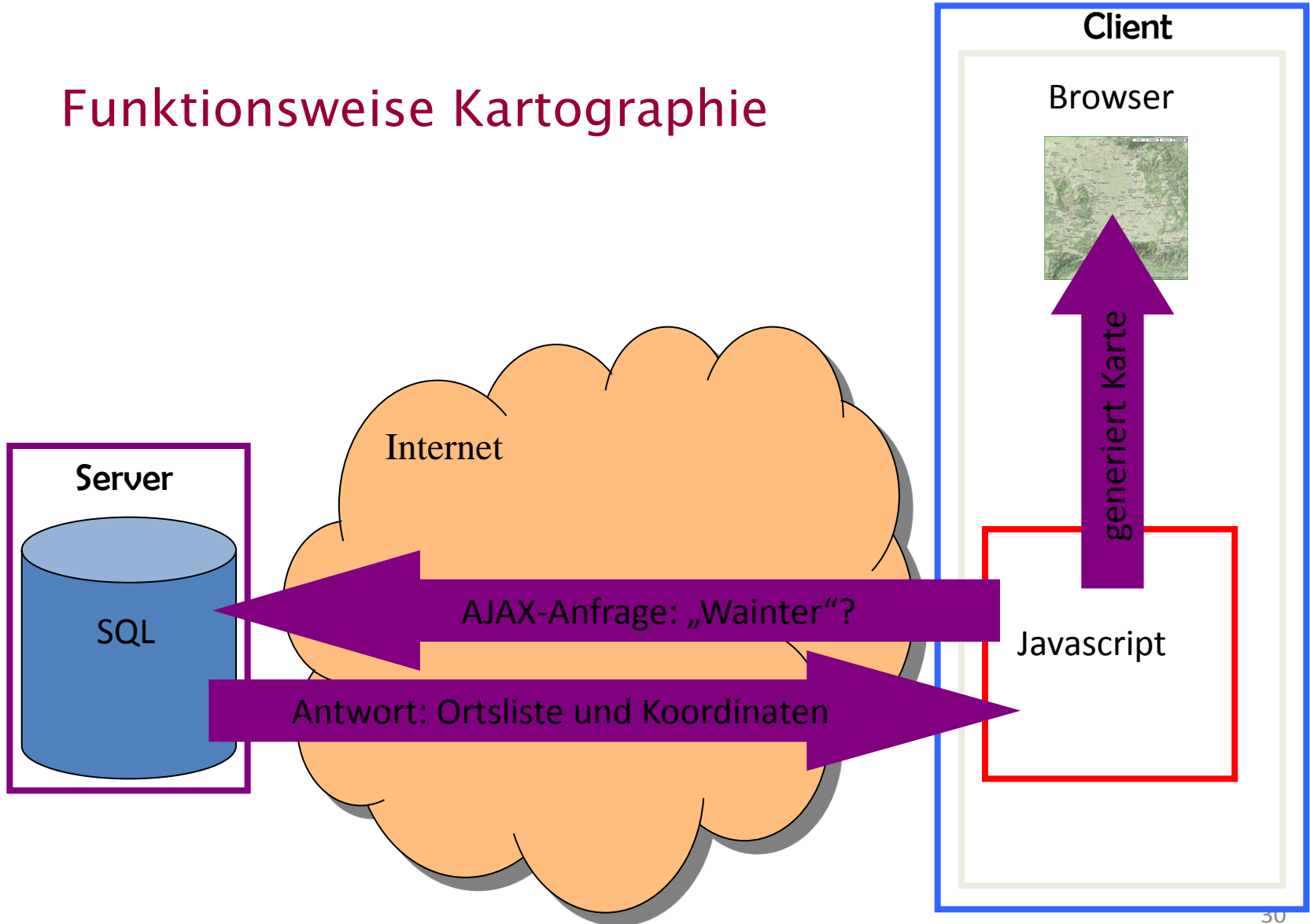
Anzahl	Ergebnis
50	394-06, 120-05, 285-09, 320-10, 347-02, 363-07, 364-08, 428-03, 524-04, 609-06, 617b-04, 900 ...

[td] = t oder d

\$ = Ende des Wortes



Funktionsweise Kartographie

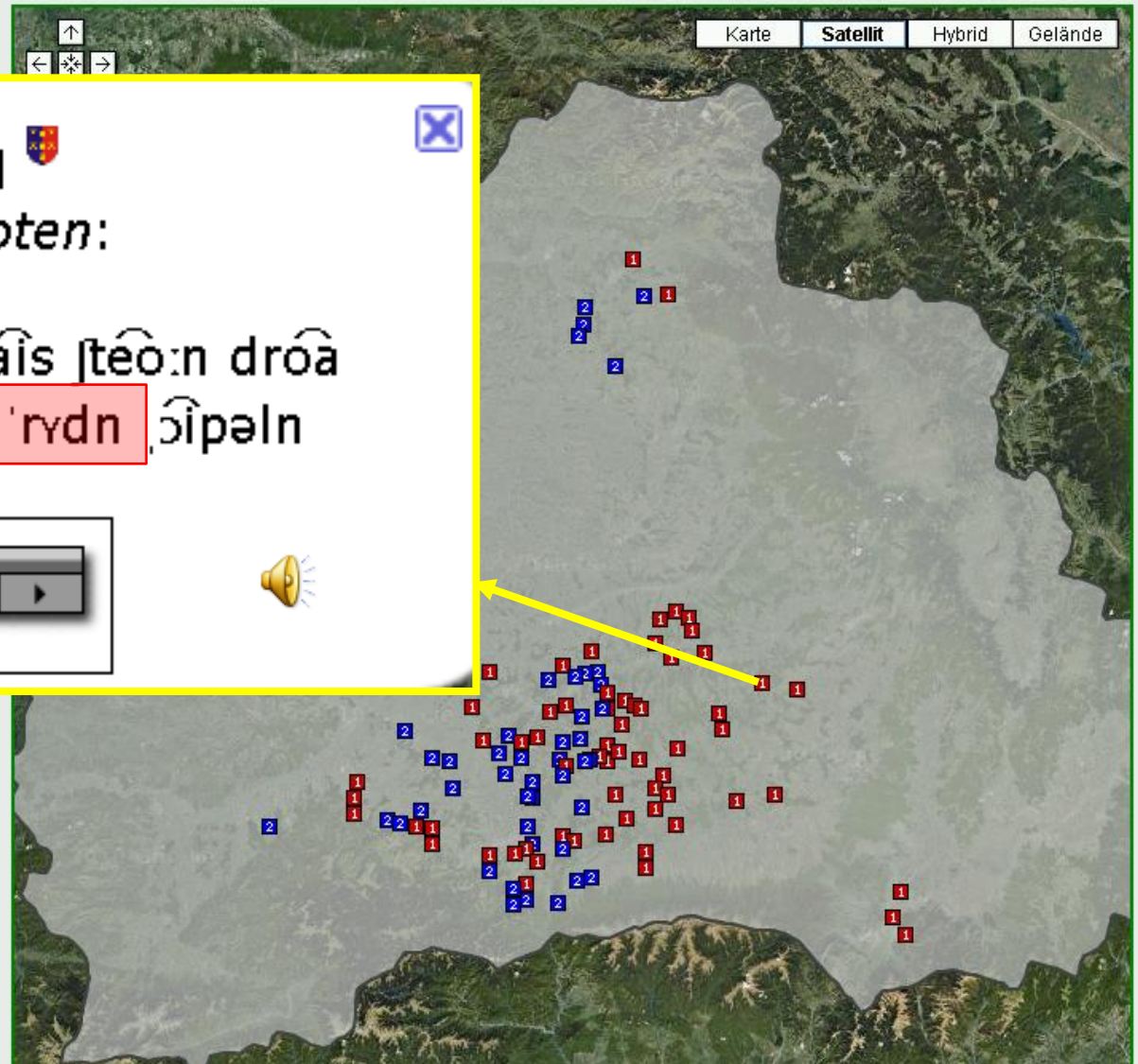


Wenkersatz 26:
Hinter unserem Haus stehen
drei schöne Apfelbäume mit
roten Äpfelchen

Qualitative Analyse: Kartierung

Wenkersatz 26, Morphosyntax: rot/roten

DIALEKTOMETRIE



Keisd 

1 => *roten*:

hōindər ɑːzəm 'hais [t̪eːn drōɑ
'ɔipəlɔim mæt 'rydn ɔipeln



A => *roten*

B => *rot*

Quantitative Datenanalyse

WS26: Hinter unserem Haus stehen drei schöne Apfelbäume mit roten Äpfelchen

WS33: Sein Bruder will sich zwei schöne neue Häuser in eurem Garten bauen

WS21: Wem hat er die neue Geschichte erzählt ?

nr	WS26	WS26	WS33	WS33	WS21	WS21
394-06	rot		sich		er	
617b-04	rot ●			sir		
1198a-04	rot ●			sir	er ●	
1016b	rot ●		sich ●		er ●	

Referenzinformant

1
2
3

- Übereinstimmung mit Referenzinformant

Ähnlichkeitsindex

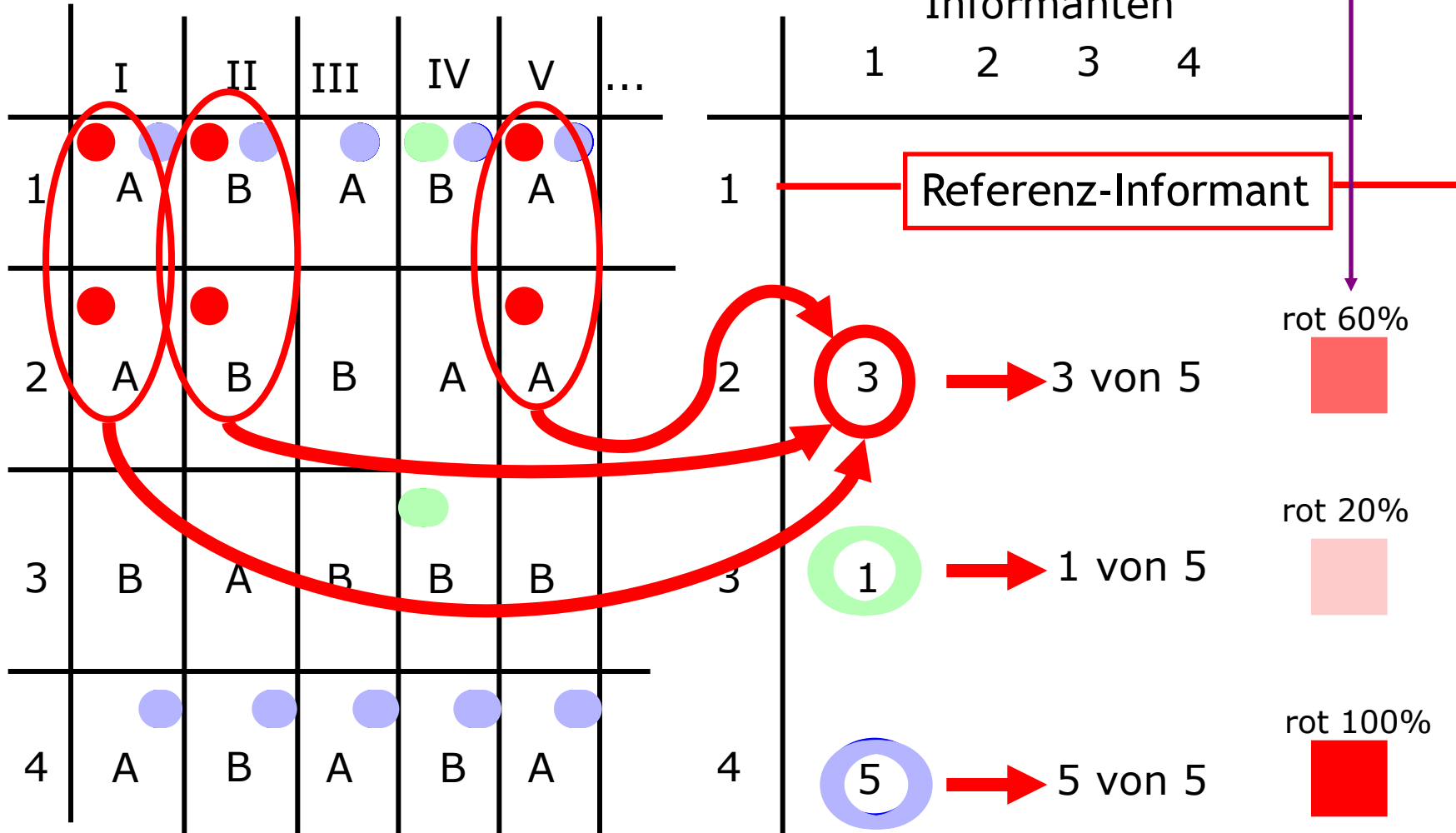
Quantitative Datenanalyse

Maß der Übereinstimmung

Merkmale

Informanten

Informanten



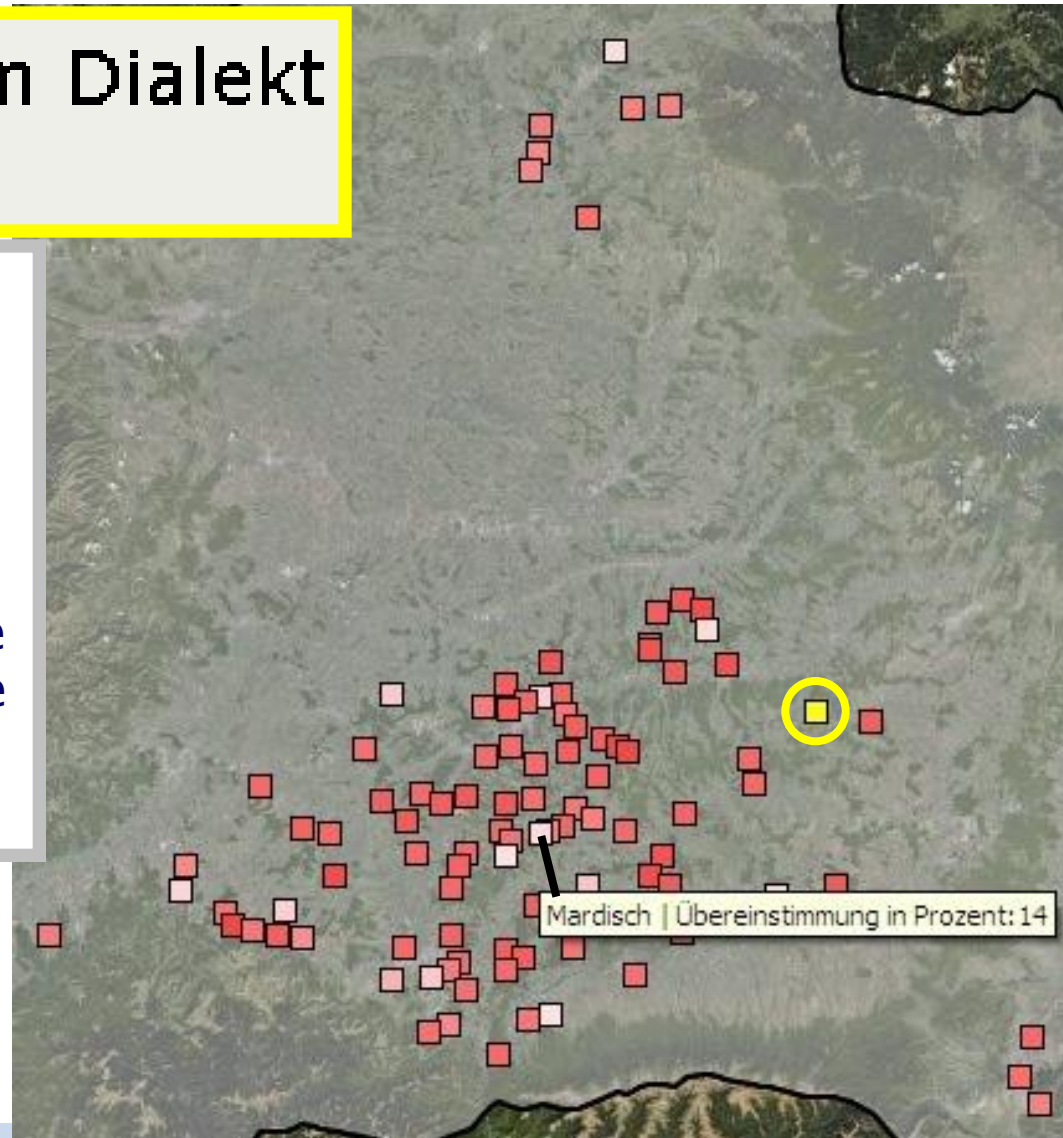
A, B: Varianten eines Merkmals

Quantitative Analyse: Kartierung

Ähnlichkeiten mit dem Dialekt von Keisd

je röter das Quadrat,
desto mehr Merkmale
hat ein Ort mit dem
Bezugspunkt gemeinsam

noch nicht sehr viele Merkmale
erfasst, dennoch auffällig geringe
Arealbildung (z.B. keine sehr klare
Korrelation zwischen Ähnlichkeit
und Entfernung)



Die informatische Seite des ASD

1. Projektgeschichte
2. Erschließung des Materials – Von der Audiodatei zur Tabelle
3. Datenstrukturierung und -analyse
4. Benutzerfunktionen
5. Möglichkeiten und Perspektiven

Gesamtbestandssuche

ÜBER DAS PROJEKT

GESAMTBESTAND

SUCHE

WENKERSATZANALYSE

DOKUMENTE

Datenbestand: **2274** | Anzahl gefundene Dateien: **0**

















[Alle Dateien](#) | [Nur transkribierte Dateien](#) | [Alle Filter entfernen](#) | [Karte bei Google Earth](#) | [i](#) | [Legende](#)





Audio	Ort	Jahr	Alter	Subcorpus	Inhalt
Filter entferne	Filter entferne	Filter entferne	Filter	Filter entferne	Filter entfernen

- Abspielen der Dateien
- orthographische und phonetische Transkriptionen
- Tagclouds aus ontologischen Begriffen

Datenbestand: **2274** | Anzahl gefundene Dateien: **67**

[Alle Dateien](#) | [Nur transkribierte Dateien](#) | [Alle Filter entfernen](#) | [Karte bei Google Earth](#) |  | [Legende](#)

Audio	Ort	Jahr	Alter	Subcorpus	Inhalt
Filter entf...	Filter entferner...	1971	Filt...	Filter entferner...	Filter entfernen
	Agnetheln (s)	1971	63	Bukarest	Schusterei; Fa Stiefelmachen
	Agnetheln (s)	1971		Bukarest	Wenkersätze
	Bistritz (s)	1971	11	Bukarest	Teppichwaschen im Bach; Winterfreuden; Heizholz spalten und schichten; Feuermachen; Bratkartoffeln kochen; zahme Eichhörnchen
	Bistritz (s)	1971	37	Bukarest	Sprache in Bistritz; Lage der Stadt; Ausflug; Bärengeschichten
	Bonnesdorf (s)	1971	43	Bukarest	Wenkersätze
	Deutsch Budak (s)	1971	61	Bukarest	Brauchtum; Streiche; Bespritzen; Hahnenschießen; Waldfest zu Pfingsten; Kirmes; Bartholomäjahrmart; Weihnachten; Neujahr; Spinnstuben; Bruderschaft; Kartoffelanbau
	Dobring (s)	1971	34	Bukarest	Wenkersätze
	Dürrbach (s)	1971	62	Bukarest	Arbeit im Weingarten; Weinbereitung
	Dürrbach (s)	1971	59	Bukarest	Backen
	Dürrbach (s)	1971	62	Bukarest	Feldwirtschaft; Bauernarbeit im Jahreslauf
	Großkopisch (s)	1971	35	Bukarest	Hopfenanbau auf dem Staatsgut
	Großkopisch (s)	1971	34	Bukarest	Kronenfest; Peter- und Paulstag; Arbeit im Baumgarten; Hausfrauenpflichten auf dem Dorf; Beschäftigung der Frauen im Winter
	Großkopisch (s)	1971	35	Bukarest	Wenkersätze
	Haschagen (s)	1971	29	Bukarest	Wenkersätze
	Haschagen (s)	1971	38	Bukarest	Nachbarschaften; Unterhaltungen im Dorf; Fastnacht; Arbeit als Schleifer in Salzburg; Arbeit als Transportarbeiter; Pendler; Militärdienst; Maurerlehre
	Heidendorf (s)	1971	75	Bukarest	Lebenslauf; 1914 Flucht nach Österreich; Rückkehr nach einem Jahr; Arbeit als Geschäftsdienner und als Kellermeister in Bistritz; Bruderschaft und Schwesternschaft; Unfug in der Spinnstube; Kindheitserinnerungen

Legende	
	phon. transkribiert
	orth. transkribiert
	orth. u. phon. transkribiert
	© Copyright Mag. Dr. Anton-Joseph Ilk

Datenbestand: 2274 | Anzahl gefundene Dateien: 1

[Alle Dateien](#) | [Nur transkribierte Dateien](#) | [Alle Filter entfernen](#) | [Karte bei Google Earth](#) | [i](#) | [Legende](#)

Audio	Ort	Jahr	Alter	Subcorpus	Inhalt
946-02	Filter entfernen	Filter entfe	Filt	Filter entfernen	Filter entfernen
	Bistritz (s)	1971	37	Bukarest	Sprache in Bistritz; Lage der Stadt; Ausflug; Bärengeschichten

Waldarbeit Haus, Haushalt, Hof **Orte und Städte** Literarische Traditionen Naturdenkmäler
Lebensabschnitte Ausflüge Sprache, Metasprachliches Kindheit Backen Mundart und Dialekt

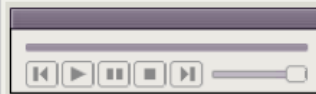
Woche und Wochentage **Geographie** Volksfrömmigkeit Flora **Orte und Städte**

Fauna Kulinarisches Bildungswesen Finanzwesen **Großfamilie und Verwandtschaft**

Sonntag Monate Jugend **Siebenbürgen**

Tageszeiten und Jahreszeiten Allgemein **Viehzucht**

Mensch und Gesellschaft



[Anhören mit Flashplayer](#)
[Anhören mit Quicktime-Player](#)

Phonetische Transkription Anmerkungen Textgrid-Datei		Orthographische Transkription Anmerkungen Textgrid-Datei	
	1 iç he:sn rudlf rø:zler		1 #1# ich heizse ### ich bin aus Bistritz
	2 iç bæ:n aõs bistrits zibmãtrãis:ç jo:r		37<simentraißig> Jahre geboren in Sankt
	3 'gabu:rãn æn 'sæn ke:n vo: mãi fõtãr 'li:rãr vo:r		Georgen<sentgergn> wo mein Vater Lehrer
	4 ont no: na'ty:rl:ç va iç hun fir fõf 'jo:r do: zãi mãr tsã:ræk æn dã 'ftãt ku		war und dann natürlich wie ich ungefähr 5
	5 unt sit 'dã:mo:Its le:bm mãr jo 'hã:		2 #1# ich will erzählen nicht wahr eine
	6 na vã iç 'vã:tãr tsiln		Geschichte was soll ich sage eine Geschichte
	7 næt vo:r æ gã'jriçt õdãr vat zõl iç 'zo:n gã'jriçt		ja na also sie kamen nun an einen
	8 çã na alzõ		3 #1# und der Vater kauft dem
	9 zã 'kun na un un 'shotãrt vo: m 'fã:gã fãr'ki:ft		Hãnschen<Hantsili> ein Paar Stiefel die
	10 ont dãr 'fõtãr ke:ft dãm 'hãntsili:pũãr 'ftãvãl		kommen ihm der Schaft bis über die Knie
	11 dã: 'kun dã 'fãft bæ:s ivãr dã 'knãrẽ		4 #1# und sie gehen weiter nun gehen sie bei
			einen Schattert wo man Kleider verkauft er
			kauft dem Hãnschen einen Rock er streift ihn
			hinauf so lang war er er kam ihm bis zu den

Suche

- Stichwortsuche im Transkriptionstext
- Schlagwortsuche
- Sprachauswahl

ÜBER DAS PROJEKT	Suche			
GESAMTBESTAND	<u>Stichwort</u>	<u>Schlagwort</u>	<u>Sprache</u>	
SUCHE	<input type="text"/>	<input type="text"/>	* <input type="text"/>	<input type="button" value="Suchen"/> 
WENKERSATZANALYSE				
DOKUMENTE				

Suche

Stichwort

fe

fe (s)
Feauser (s)
Feber (s)
Februar (d)
Feddelbiugen (s)
Fedemen (s)
Fedem (s)
Feder (d)
Federbetten (d)
Federchen (d)

Schlagwort

Sprache

Suchen



Natur

Natur

Flora

Fauna

Wetter und Naturgewalten

Tageszeiten und Jahreszeiten

Woche und Wochentage

Monate

Landwirtschaft

Landwirtschaft

Ackerbau und Feldwirtschaft

Viehzucht (auch Kleinvieh und Hirten)

Imkerei

Weinbau

Garten- und Obstbau

Ernte

Landwirtschaftliche Geräte

Haus, Haushalt, Hof (auch Gebäude)

Haus, Haushalt, Hof (auch Gebäude)

Suche

Stichwort <input style="width: 95%;" type="text"/>	Schlagwort <input style="width: 95%;" type="text"/>	Sprache * <input style="width: 20px;" type="text"/>	Suchen	
---	--	--	--------	--

Die Suche nach '**milch***' + '*****' + '*****' erzielte **184** Treffer.

- Milch
 Milchtopf
 Milchbrei
 Milchfrau
 Milchstern
 milchsauer
 Milchkühe
 Milchfabrik
 Milchkannen
 Milchkühen
 Milchpulver
 Milchkaffee
 Milchwirtschaft
 Milchproduktion
 Milchstehlerinnen

	Datei, Intervall	Kontext		Ort
	1092 , 20	na und so hat man immer müssen lernen immer besser bis man dann gut ein ist geübt na und dann an Heiligtagen wenn wir haben gebacken Stritzel und Hanklich wie wir es ja heißen haben wir fein geknetet wir haben hin Eier und Butter und Milch geschlagen		Minarken
	1094 , 31	na ja ja sie helfen schon wir haben eine Kuh wir haben sie melken sie helfen schon sie holen die Milch den Ferkelchen sie kochen Kartoffeln<krumpirn> sie helfen schon sehr viel vor allem die zwei großen Mädchen		Minarken
	1094 , 61	wir haben die geknetet mit Mehl und Eiern und Milch gaben dort einer Nachbarin entweder ein Kilo Zucker oder ein Kilo Reis und sie gab uns dann Milch dafür ich sagte ich kann sie mit Wasser nicht kneten wir haben mit Wasser nicht geknetet denn das klebt<pikt> dann so		Minarken
	1094 , 61	wir haben die geknetet mit Mehl und Eiern und Milch gaben dort einer Nachbarin entweder ein Kilo Zucker oder ein Kilo Reis und sie gab uns dann Milch dafür ich sagte ich kann sie mit Wasser nicht kneten wir haben mit Wasser nicht geknetet denn das klebt<pikt> dann so		Minarken

Suche

Stichwort <input style="width: 95%;" type="text"/>	Schlagwort <input style="width: 95%;" type="text"/>	Sprache *	Suchen	
---	--	--------------	--------	--

Die Suche nach '*' + '*' + 'u' erzielte **140** Treffer.

- a is en na de az tok van meg tej egy jai nem nagy fele fasz
 bika mint Csik volt kell Bike zseb Batja gyere raita Puika Bacs hozon
 bajam keves Janka Elesd fazek Annal leves ember kicsi lelke minden
 Bolyai Borviz ninden menyek Rabsag hogyne gulya z
 Kelemen Bethlen schovari magyarok Nagycsur Ves
 között Kuelvaros Morischka Vasarhely Udvarhely Ked
 kaporlasch Marosujvar Nagyengelt Karanszebes D
 Honved-Husaren Marosvasarhely Marosvasarehely Dicsoe

Legende

Treffer

- deutsche Passage
- siebenbürgisch-sächsische Passage
- mittelbayerische Passage
- rumänische Passage
- ungarische Passage
- russische Passage
- ukrainische Passage

	Datei, Intervall	Kontext	
	1169d-08 , 77	nagy fasz kicsi fasz minden fele fasz en van jai lelke nem az a bajam na hogyne en van fasz ninden fele nagy fasz kicsi fasz minden fele van	Pretai
	715-07 , 51	ein Kelch mit was wir das Heilige Abendmahl nehmen aus Gold das ist ein Geschenk von ihr nun würde ich das gerne veröffentlichen aber ich fürchte stets<ower ich grolen änj> die Ungarn sperren mich ein denn Annal ügyesebb ember mint Bolyai a magyarok között nem volt	Maldorf
	715-07 , 51	ein Kelch mit was wir das Heilige Abendmahl nehmen aus Gold das ist ein Geschenk von ihr nun würde ich das gerne veröffentlichen aber ich fürchte stets<ower ich grolen änj> die Ungarn sperren mich ein denn Annal ügyesebb ember mint Bolyai a magyarok között nem volt	Maldorf
	1169d-08 , 77	nagy fasz kicsi fasz minden fele fasz en van jai lelke nem az a bajam na hogyne en van fasz ninden fele nagy fasz kicsi fasz minden fele van	Pretai
	33-12 , 8	und hatten ja gefunden den alten Bacsi im Presshaus und er hatte auch ein paar Haare<Hooren> auf der Nase	Bartholomae

Morphosyntaktische Etikettierung

Spontansprachliches Material		Wenkersätze	
Datei	Bitte auswählen ▾	<	Keisd ▾ >
Intervall		<	1 ▾ >
			Ort Wenkersatz

	æm	'vanjtər	flæjən	də	gə'drɛçt	'bladər	an
Satz - +	11 ▾	11 ▾	11 ▾	11 ▾	11 ▾	11 ▾	11 ▾
Funktion	Atemp ▾	Atemp ▾	Vfin ▾	ENom ▾	ENom ▾	ENom ▾	Alok ▾
Wortart	praep ▾	subst ▾	verb ▾	art ▾	adj ▾	subst ▾	praep ▾
Spezif. ■	lok ▾ temp ▾ mod ▾ det ▾	normal ▾	voll ▾	▾	<i>Gebrauch</i> ▾ attr ▾ subst ▾ adv ▾	normal ▾	lok ▾ temp ▾ mod ▾ det ▾
Modus	▾	▾	ind ▾	▾	▾	▾	▾
Person	▾	▾	3 ▾	▾	▾	▾	▾
Numerus	▾	sg ▾	pl ▾	pl ▾	pl ▾	pl ▾	▾
Kasus	▾	dat ▾	▾	nom ▾	nom ▾	nom ▾	▾
Tempus	▾	▾	praesens ▾	▾	▾	▾	▾
Genus	▾	m ▾	▾	n ▾	n ▾	n ▾	▾
Lemma	im_1	Winter_2	fliegen_3	der_4	trocknen_5	Blatt_6	in_7
speichern							

Die informatische Seite des ASD

1. Projektgeschichte
2. Erschließung des Materials – Von der Audiodatei zur Tabelle
3. Datenstrukturierung und -analyse
4. Benutzerfunktionen
5. Möglichkeiten und Perspektiven

Möglichkeiten und Perspektiven

Regionale Erweiterung:

- Wassertal
- Banat
- Bukowina
- Zips
- ...

Thematische Erweiterung: Daten zur

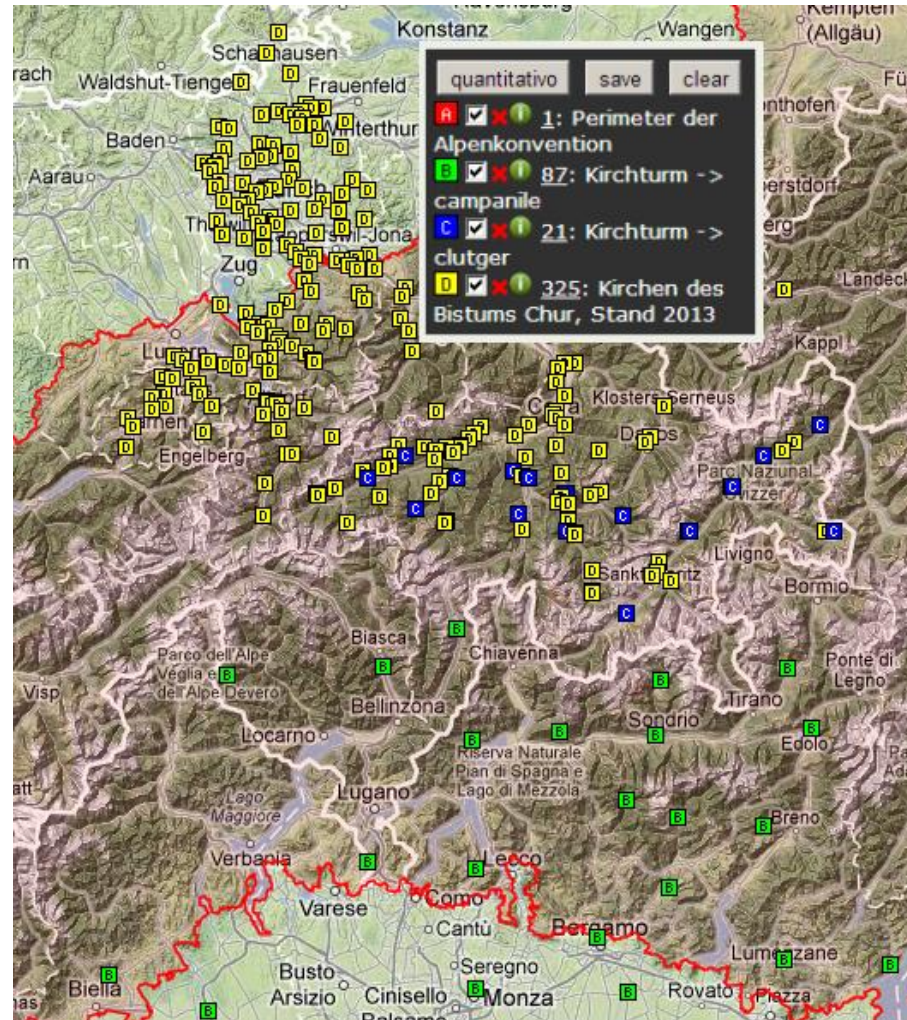
- Demographie
- Geschichte
- Infrastruktur
- ...

Speziell (Auswahl):

- Zeitungen samt Auflagenstärke
- Verlage samt Ort
- Ethnische Zusammensetzung
- Deutschsprachige Schulen

Einbindung von Laienwissen („crowdsourcing“)

Synopse: clutger, campanile, Bistum Chur

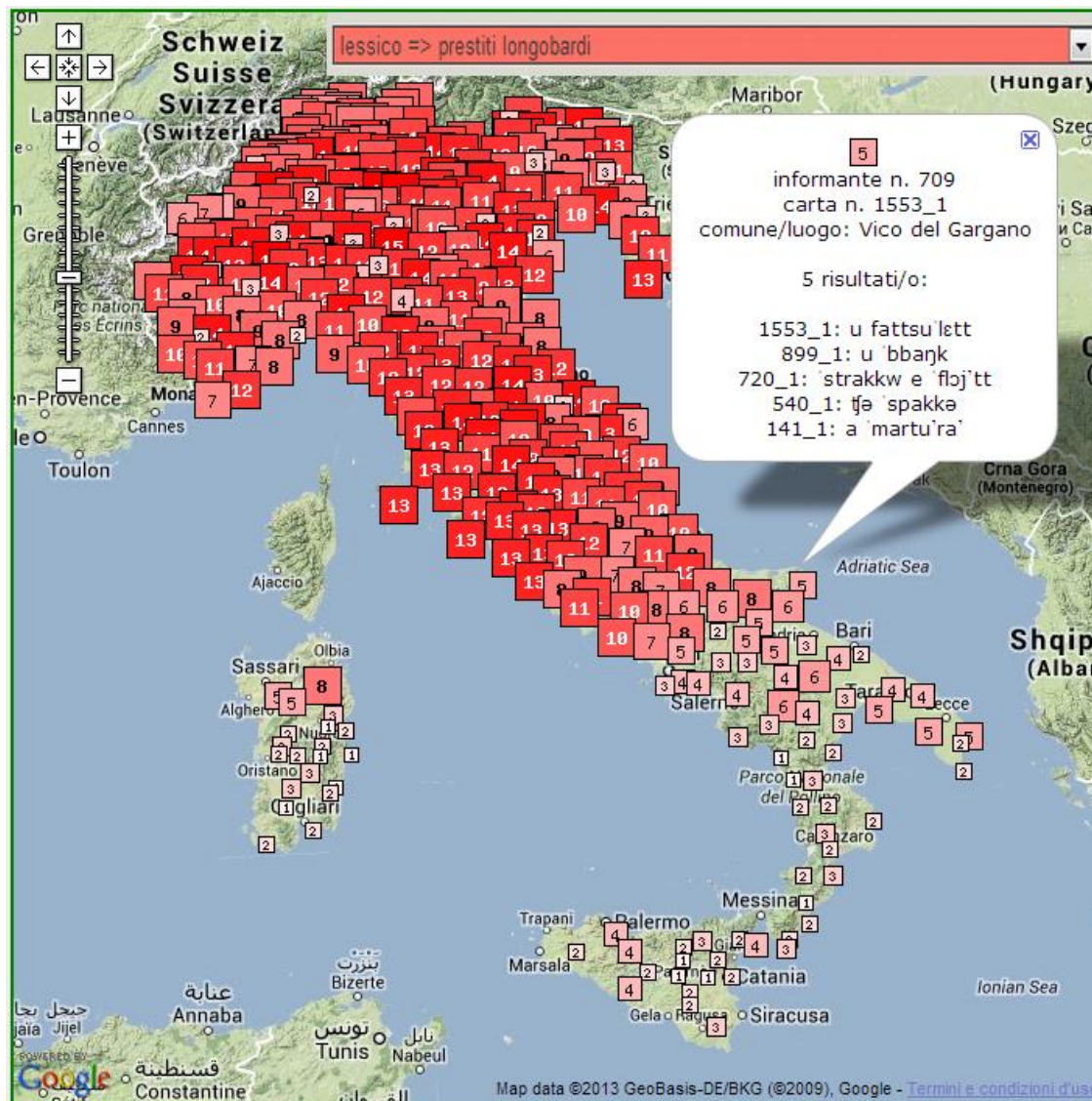


Phasen der Ausdehnung des Langobardenreichs



<http://de.wikipedia.org/wiki/Langobardenreich>

Quantitative Kartierung: Kumulierung und räumliche Verteilung der sprachlichen Relikte der Langobarden



Links & Kontakt

Vortrag abrufbar unter:

www.asd.gwi.uni-muenchen.de -> **DOKUMENTE**

emma.mages@itg.uni-muenchen.de

luecke@lmu.de