

Transkription im DMW

Kai-Uwe Carstensen (Siegen)

DMW-Workshop „Methoden der Transkription und Transliteration dialektaler Daten“

Münster, 9. August 2019

Ausgangspunkt: DMW-Previewkarten

dynamisch generierte Wortkarten mit IPA und POP(-Typen)

DMW Karten

Karteneinstellungen Ich interessiere mich für ... Ergebnisdarstellung

The map displays a region with numerous colored markers. A popup for 'Wenholthausen' is visible, showing IPA and POP information.

Wenholthausen
Kreis Hochsauerlandkreis

KEI-E ('kʰɛ̯.ə)

KII-E ('kʰi:ə)

Menü anzeigen plution:655.39

Map tiles by [Stamen Design](#), under [CC BY 3.0](#). © [OpenStreetMap](#) contributors.

"Kette.Substantiv.Akk.Sg.F"

- Typ KETT
- Typ KII-E
- Typ KIE
- Typ KERRE
- Typ KEDDE
- Typ KINGEN
- Typ GETT

Letzte Auswahl zurücknehmen

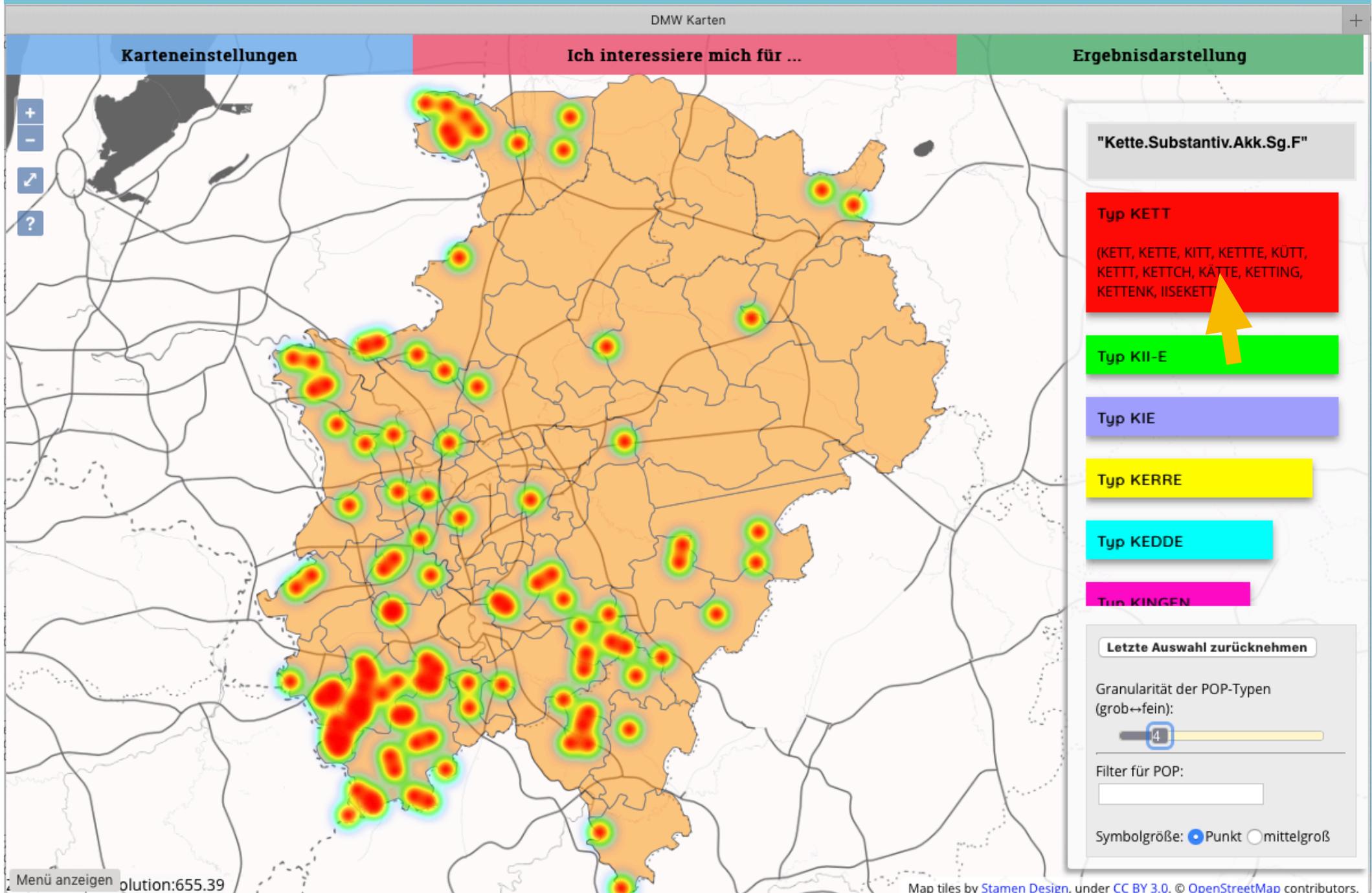
Granularität der POP-Typen (grob↔fein):

Filter für POP:

Symbolgröße: Punkt mittelgroß

Ausgangspunkt: DMW-Previewkarten mit POP-Typen

Heatmap-Karte des Vorkommens eines ausgewählten POP-Typs



Transkription im DMW: Desiderate

- **Schnelligkeit**

- Wie lässt sich eine effiziente Eingabe gewährleisten?
 - Transkription ist nur ein Teil der (Vor-)Analyse/Datenaufbereitung!
 - + Schneiden, Qualität der Wavs, metaling. Aspekte, Phänomenbearbeitung
 - ~800 Orte x 400 Items x 4 GPs = 1 280 000
 - z.Zt: 100 000 Transkriptionen überschritten

- **Qualität**

- Wie können vermeidbare Fehler in realistischen Kontexten (z.B.: moderate Einarbeitungszeit für studentische Mitarbeiterinnen) vermieden werden?

- **Lesbarkeit**

- Wie können Transkriptionen laienlesbar sein?

Schnelligkeit sichern

- durch komfortable Eingabe
 - ≠ händische Eingabe einzelner IPA-Symbole
 - stattdessen unterstützendes **Transkriptionstool**
- **POPs** (populäre/literarische Transkriptionen nach Hamburger Konventionen) werden **automatisch** erzeugt (kein Zeitaufwand)
- Bestimmte Informationen sind direkt im Analyseinterface verfügbar
 - Eingabekürzel als Infofenster
 - zum direkten Vergleich: wav und Transkription der anderen GP im Ort

Transkription

RWL: Weidekätzchen [L]

Referenzeinheit(en): Weidekätzchen

Observante:  GP-ID: 65 

IPA: 'vaɪ.də. kɛts.jə [bearbeiten](#)

RWL fehlt:

IPA-Notation: 'vaɪ.də. kɛts.jə

POP: Anmerkungen zur Transkription:

Konfidenzwert: sicher weder-noch unsicher

- **IPA-Notation:** Mix an graphematischer Eingabe, textueller SAMPA-Eingabe, IPA-Eingabe per Tastenkombination
 - erlaubt persönliche Präferenzen, ermöglicht Flexibilität und ist weniger monoton
- teilweise automatisch erzeugte Diakritika (Affrikate,(Nicht-)Silbizität)
- **automatische** Fehlererkennung (reduziert Überprüfungszeit)
- Autocompletion der Eingabe mit bisherigen Eingaben zu dem Item (reduziert Eingabezeit)

Transkriptionstool

RWS: **(ART) Eichhörnchen**
RWL(s): **Eichhörnchen** L:Eichhörnchen
Schneiden: **Ja** Audio: **Ja**

▸ wav-Bearbeitung

▾ Transkription

RWL: **Eichhörnchen [L]**

Referenzeinheit(en): **Eichhörnchen**

Observante:



loop

slow loop

RWL fehlt:

IPA-Notation: **k^h**

k^h

POP:

k^ha.p^he:ç!

k^ha.t^he: k^ha

k^ha.t^he: k^he

k^ha.t^he: k^hə

Konfidenzwert:

k^ho.ʎεt̪s̪iə̃ weder-noch

k^hə.t^hε: k^həl

GP-ID: 65

loop slow loop

IPA: ʔaɪç. hœɛn.çə

IPA-Eingabekürzel

Anmerkungen zur Transkription

Kürzel

- ɹ (Stimmhafter alveolarer Approximant): ɹ\
- ɹ# (Stimmhafter alveolarer Approximant): ɹ#
- ɹ̠ (Stimmhafter retrofleher Approximant): ɹ̠\
- ɹ̠# (Stimmhafter retrofleher Approximant): ɹ̠#
- ɹ̠̥ (Stimmhafter retrofleher Flap): ɹ̠̥
- ɹ̠̥# (Stimmhafter alveolarer Tap): ɹ̠̥#
- ɹ̠̥̥ (Stimmhafter uvularer Vibrant): ɹ̠̥̥\
- ɹ̠̥̥# (Stimmhafter uvularer Vibrant): ɹ̠̥̥#
- ɹ̠̥̥̥ (Stimmhafter uvularer Frikativ): ɹ̠̥̥̥
- ʃ sch
- ʃ S
- ʃ̥ (auch über Ctrl+Ü): }̥
- ʃ̥ U
- ʃ̥̥ (Stimmhafter labiodentaler Approximant): ʃ̥̥
- ʌ V
- ɹ Y
- ʒ (Stimmhafter alveolopalataler Frikativ): z\
- ʒ# (Stimmhafter alveolopalataler Frikativ): z#\
- ʒ Z
- ʒ̥ (Stimmhafter palataler Frikativ): ʒ̥\
- ʒ̥# (Stimmhafter palataler Frikativ): ʒ̥#\
- ʰ (aspiriert): _h
- ʲ (palatalisiert gesprochen[er Konsonant]): _j
- ʷ (labialisiert gesprochen[er Konsonant]): _w



Qualität sichern I

- Training der Transkription mit **Transkriptionsübungstool**
 - Verschiedene Notationen (IPA, SAMPA) und Tastaturkürzel erfordern intensive Einarbeitung
 - Silbengrenzen + -gelenke sowie Betonung werden auch notiert!
- Übungstool verwendet ausgesuchte reale wavs + Musterlösungen
- Automatische Fehlermeldungen und Warnungen (bei möglichen Fehlern)

Transkriptionsübungstool

AudioAnalysis – DMW-Projekt

Transkriptionstraining

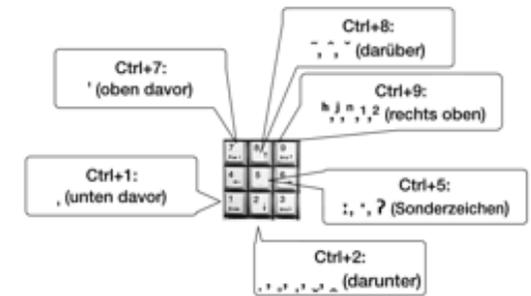


Akustische Übung der Transkription

Hilfe

Feder

Feder



,

Check

► Erläuterungen

Qualität sichern II

- Wavs lassen sich im (langsamen) **Loop** abspielen
- IPA der anderen GP ist hör-/sichtbar
- **POPs** werden bei der Eingabe *automatisch erzeugt* und angezeigt (direktes Feedback)
- obligatorische Eingabe eines **Konfidenzwertes** für die Transkription (Nachvollziehbarkeit und schnelle Identifikation notwendiger Verbesserungen/Korrekturen)
 - außerdem: Gründe für Transkriptionsprobleme müssen in Anmerkungsfeld angegeben werden (ggf. kann erst dann gespeichert werden)

▼ Transkription

RWL: Weidekätzchen [L]

Referenzeinheit(en): Weidekätzchen

Observante:  GP-ID: 65 

IPA: 'vaɪ.də. kɛts.jə [bearbeiten](#)

RWL fehlt:

IPA-Notation: 'vaɪ.də. kɛts.jə

POP: WAIDEKETZJE

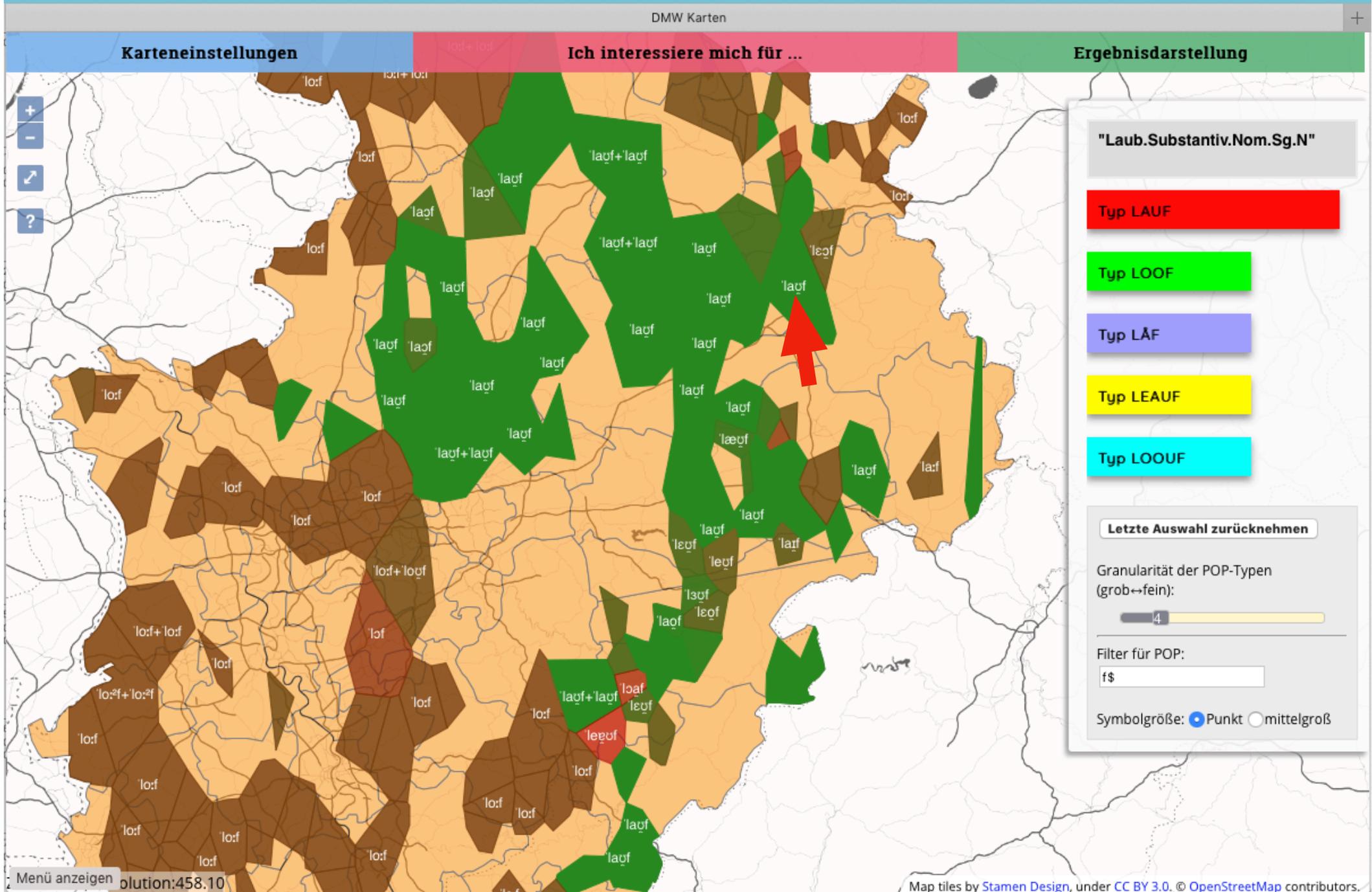
Anmerkungen zur Transkription:

Konfidenzwert: sicher weder-noch unsicher

Qualität sichern III

- Auf Karten für Items sind ***Ausreißer als potentielle Transkriptionsfehler*** sichtbar
- dialektometrische Karten mit Referenzpunktvergleichen:
 - feinere phonetische Unterschiede (\neq POPs) sichtbar

Dialektometrischer Referenzpunktvergleich: Phonetische Ähnlichkeit zu einer angeklickten Ortszelle



Qualität sichern IV

- **IPA-Repair-Tool:**
 - listet die Transkriptionen für auswählbare Items, Orte, IPA-/POP-Bestandteile, Konfidenzwerte auf
 - Ergebnisliste enthält jeweils Audioplayer und *direkten Link in das Analyseinterface* zur Bearbeitung/Korrektur der jeweiligen Transkription

IPA Qualitätscheck- und Reparaturtool

RWL:

IPA:

POP:

Ort:

Konfidenzwert: sicher weder-noch unsicher alle

71 Treffer

RWL	IPA	Audio	AI Link	Ort
Kaulquappe	'kaʊ̯l.kvap.pə		im AI öffnen	Silbach
Kaulquappe	'kaʊ̯l.kvap.pm̩		im AI öffnen	Barlo
Kaulquappe	'ku:l.kvap.pə		im AI öffnen	Schmidt
Kaulquappe	'kaʊ̯l.kvap.pə		im AI öffnen	Lienen
Kaulquappe	'ku:l.kvap.pə		im AI öffnen	Birgelen

Laienlesbarkeit der Transkriptionen

- POPs! (populäre Transkriptionen)
- nach den Hamburger Konventionen
 - minimalste Abweichungen in Theorie und Umsetzung
 - automatische POP-Erstellung aus IPA-Transkripten
 - Spezifischer Zweck der POPs im DMW lässt Raum für Diskussionen von Regelabweichungen (s. z.B. <er> in *Butter*-POPs)

'bʊtəɹ
'bʊtəɹ

BUTTER

'kɛn.vɔət
'kɛn.vɔət

KENNWOAT

Anzahl POPs (nach Hamburger Transkriptionskonventionen): > 80

KIIE, KII-E, KEA, KEEA, ISEKET, KET, KERRE, KEDDE, KETTE, KEEN, KEE-EN, KEI-EN, ISEKEA, KEE, KAI-E, KEU-E, KUI-E, KII-AN, KEAT, KIILN, KIILEN, KÖLLE, KII-EN, KIIRE, KIRRE, KEERE, ISEKEEA, KIIN, KEE-E, AISEKEAL, KII, KIT, KÄÄDE, KEEDE, KIIGE, KÄTTE, KEDDING, KEIDE, KETTING, KUU-E, KII-ER, KIIR, KEI-E, KIR, CHEEDE, KEIN, KIWWE, KHEE-E, KIEN, KIDDE, KÄÄT, KEEEN, KETE, KETCH, KETTTE, KIE, KETTENK, KEAN, KIETE, KUIE, KÖRRE, KETTT, AESEKETTE, KÄÄTE, KEEAN, KEDDN, KINGEN, KÜT, KAIN, KUIN, KIEN, KUIEN, KIATE, KEIE, KETT, GET, KÄIE, KIIJE, KIRE, KIIA, KÜ-E, KEATER

deshalb: Bildung von **POP-Typen** auf verschiedenen Granularitätsebenen, z.B.:

Typ KETT

(KETT, KETTE, KITT, KETTTE, KÜTT,
KETTT, KETTCH, KÄTTE, KETTING,
KETTENK, ISEKETT)

POP-Typen

- **automatisch** und **dynamisch** erzeugt
- POP-Typen werden angezeigt; per Mouseover auch die zugehörigen POPs
- mehrere Ebenen der **Körnigkeit** (Granularität)
 - einstellbar durch Regler (6 Ebenen)
 - grobe Sicht (POPs mit gemeinsamem Anfangsbuchstaben) zuerst
 - POPs als feinste Granularitätsebene
- **Reihenfolge** der POP-Typen in der Legende **nach Häufigkeit** der Typen
- **Farben** der POP-Typen: die zwölf häufigsten erhalten eine festgelegte Farbe, alle anderen grau

Vielen Dank
für die
Aufmerksamkeit!