Das Öffentliche DGS-Korpus: Release 4

AKADEMIE DER WISSENSCHAFTEN IN HAMBURG

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Vortragsreihe am IDGS, WiSe 2024/2025

06.11.2024, Hamburg

Reiner Konrad, Amy Isard

Inhalte

- DGS-Korpus
  - 1. Erhebung
  - 2. Erhebung
  - Studio
- Öffentliches DGS-Korpus
  - Inhalte
  - MEINE DGS
  - MEINE DGS – annotiert
- MEINE DGS – ANNIS
- MEINE DGS – iLex
- Literatur

DGS-Korpus

1. Erhebung

- Korpusdesign
  - Projektziele
    - repräsentatives Korpus der DGS
    - korpusbasiertes Wörterbuch (DW-DGS)
    - öffentlich zugängliches annotiertes Korpus

DGS-Korpus

1. Erhebung

- Korpusdesign
  - ➔ Dokumentation der DGS: „more is more“ / „the bigger the better“
  - ➔ Ausgewogenheit: Alter, Geschlecht, Region (DGS-Erwerbsalter, soziale Stellung, prof. Umgang mit DGS)
  - ➔ Textsorten: Dialoge (u. Monologe), möglichst spontan, verschiedenste Themen
  - ➔ Einverständniserklärung (informed consent): Daten und Metadaten

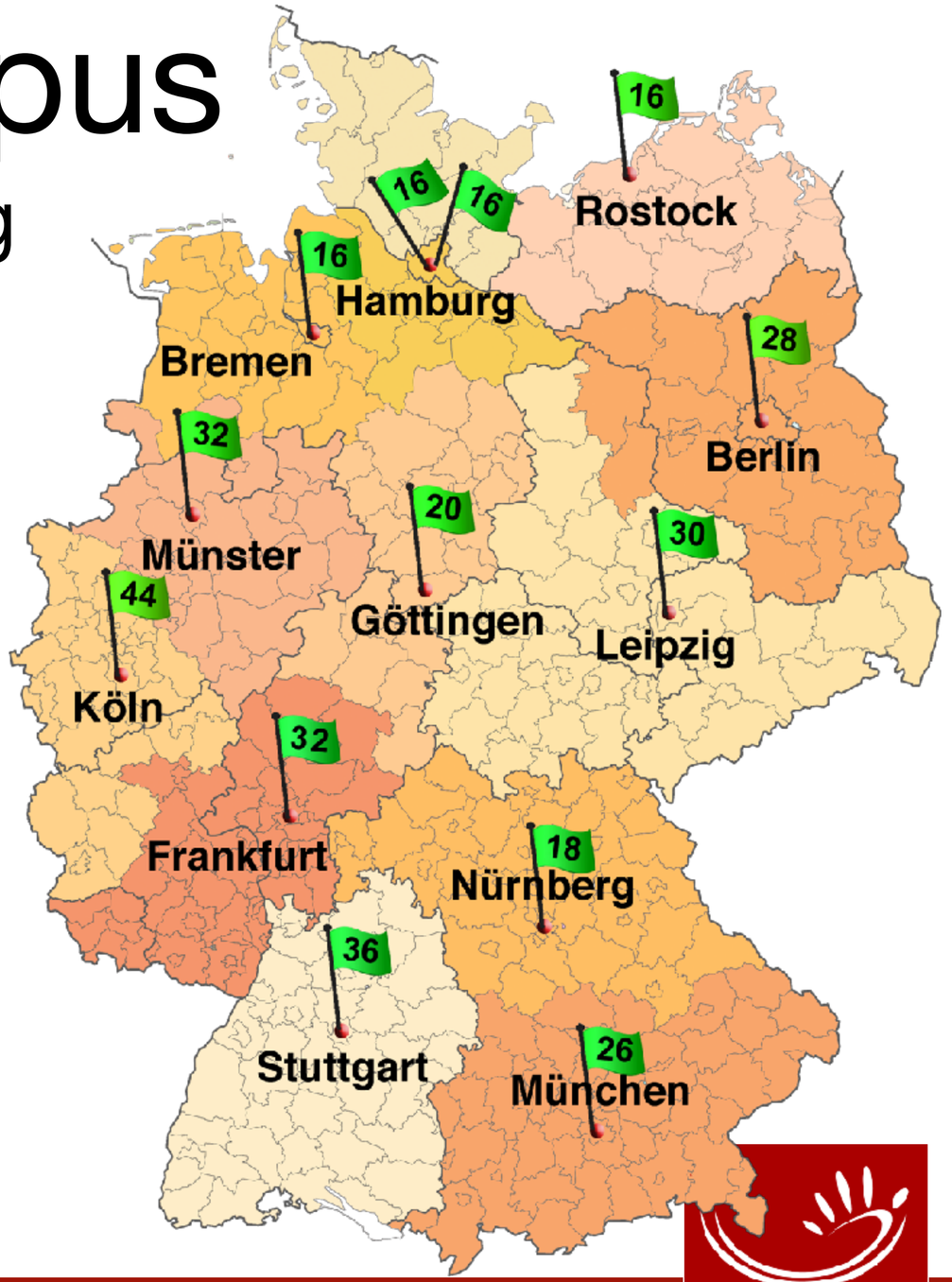
DGS-Korpus

1. Erhebung

- Informantensuche:
  regionale Kontaktpersonen (Schulung)
  Auswahlkriterien (Fragebögen)
- Datenerhebung (2010-2012)

➔ 13 Erhebungsregionen/
12 Erhebungsorte

➔ 165 Erhebungen (paarweise; 5-7 Std.)
330 Informanten
23 versch. Aufgaben (Tasks)



DGS-Korpus

2. Erhebung

- Projektantrag v. Jan. 2008
  - Jan. 2009–Dez. 2023
  - Erhebung: Jan. 2010 - März 2012:

    TN = 18–… Jahre (4 Altersgruppen)
- Neufestsetzung der Laufzeit („Verlängerung“)
  - Jan. 2024–Dez. 2027
  - Erhebung: Nov. 2024–2026: 1 Altersgruppe

    TN = 18-32 Jahre/Jahrgänge 1993–2007
    (5. Kohorte)

DGS-Korpus

2. Erhebung

- Ziel: Dokumentation des Sprachwandels in der DGS
  - Bildung: Inklusion und Einstellung zur DGS
  - Technologie: Smartphone-Nutzung, Social Media; CI
  - Demografie: Migration
- **78** TN angestrebt

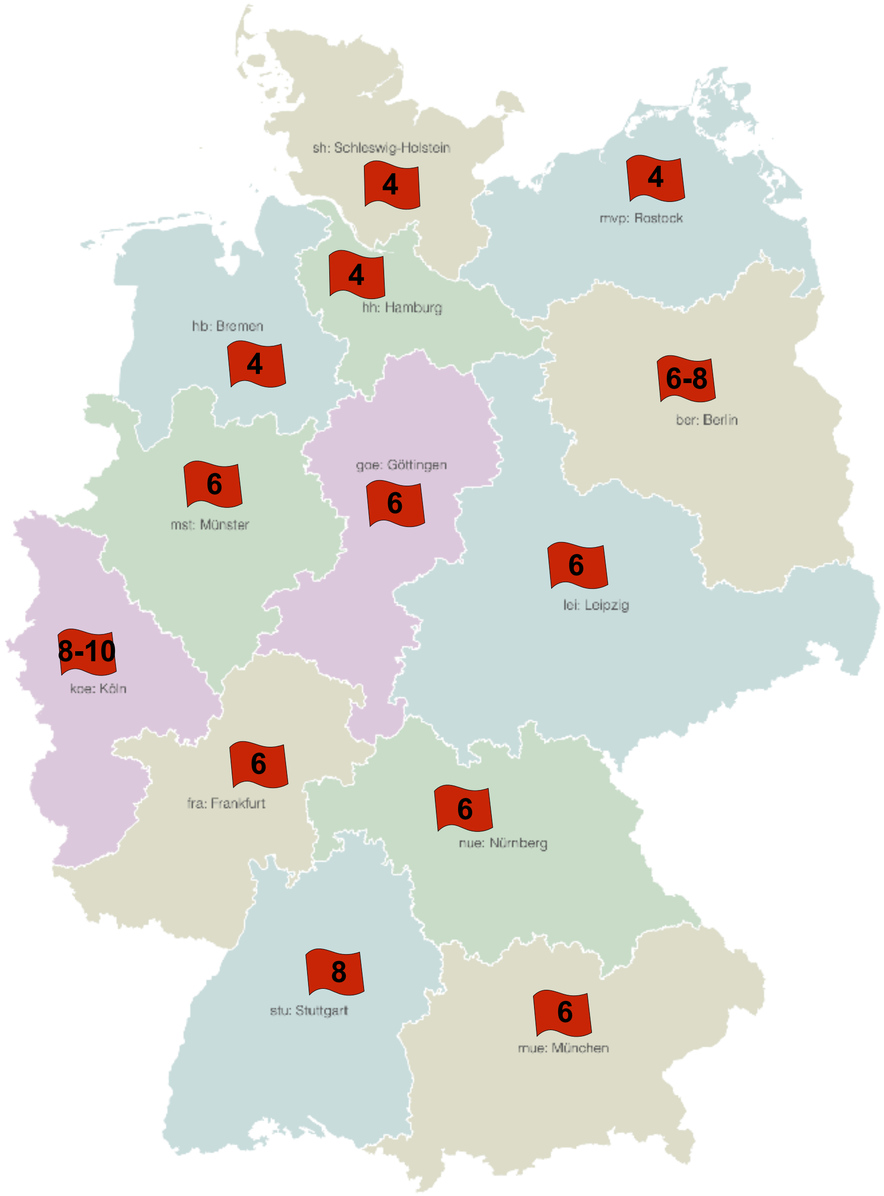
  zum Vergleich: 1. Erhebung: **82** TN pro Altergsruppe

➡ 9 junge Moderator:innen

➡ 13 Regionen (+ Subregionen)

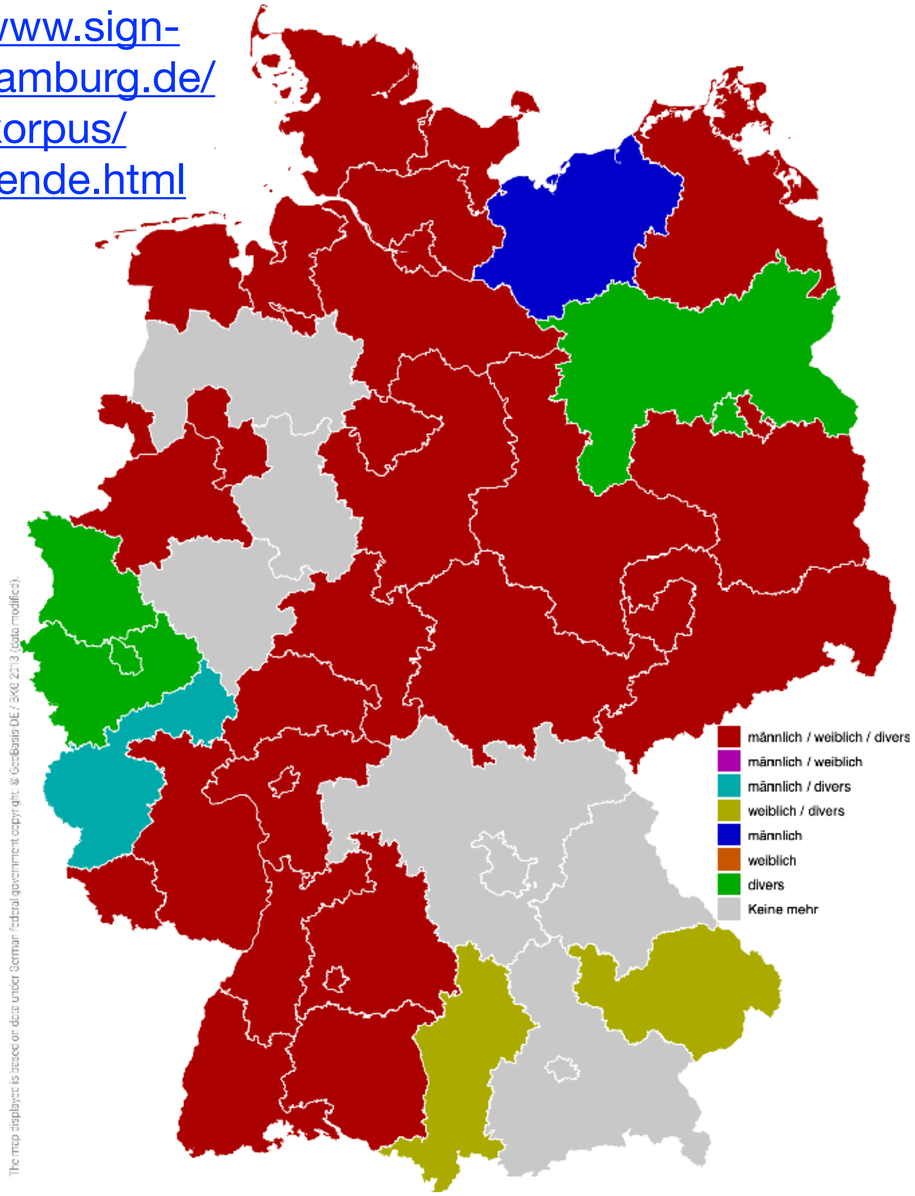
➡ 1 Erhebungsort: **Hamburg** (IDGS: Studio)

78 Teilnehmende



Teilnehmende der Jahrgänge 1993–2007 werden noch gesucht aus folgenden Regionen:

https://www.sign-lang.uni-hamburg.de/dgs-korpus/teilnehmende.html



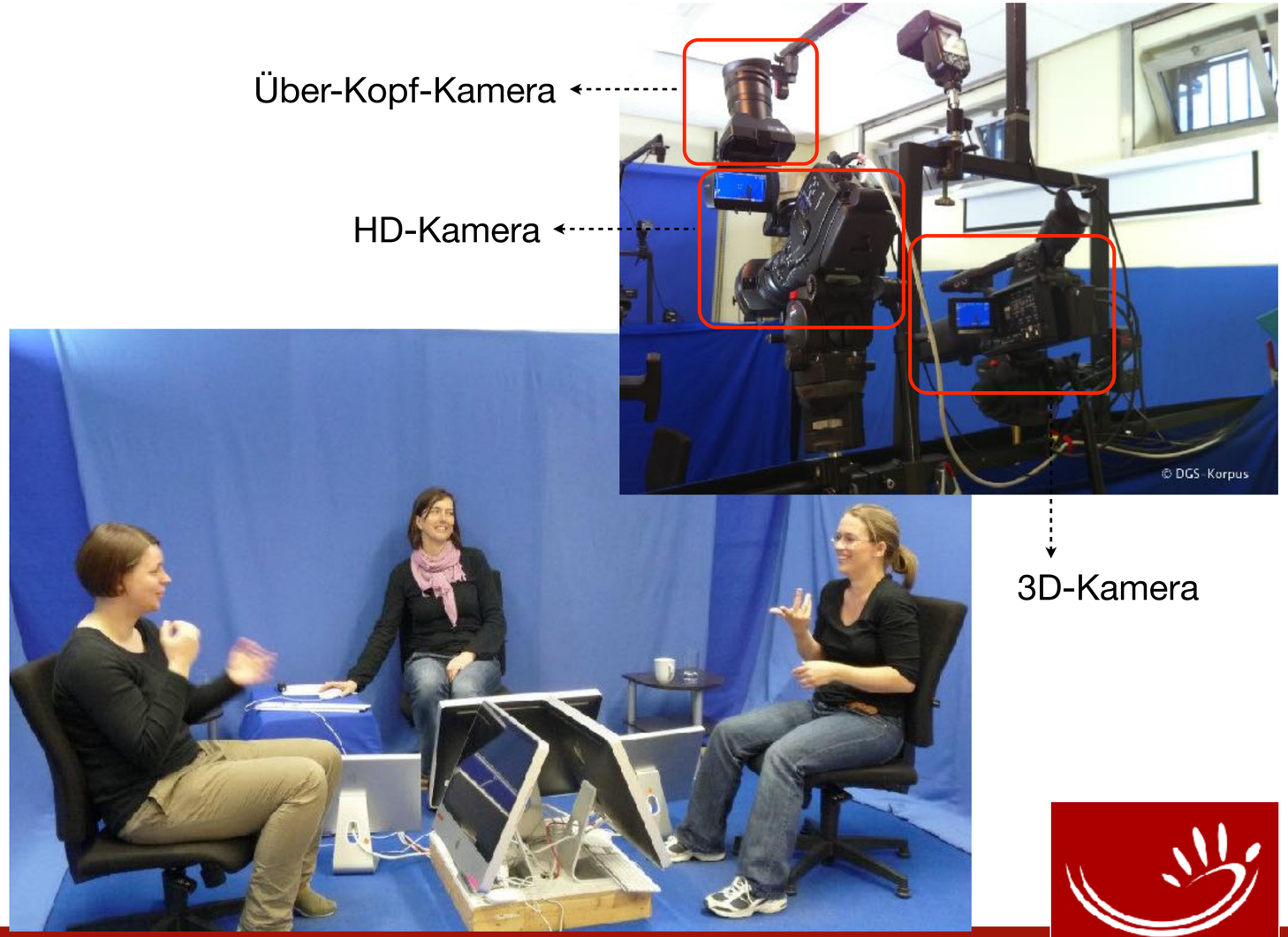
DGS-Korpus: Erstellung

Studio: 1. Erhebung

- Mobiles Studio:
  8 Kameras
  (3D und HD)
  5 Perspektiven

➔ ca. 1150 Std. Video
ca. 800 TB

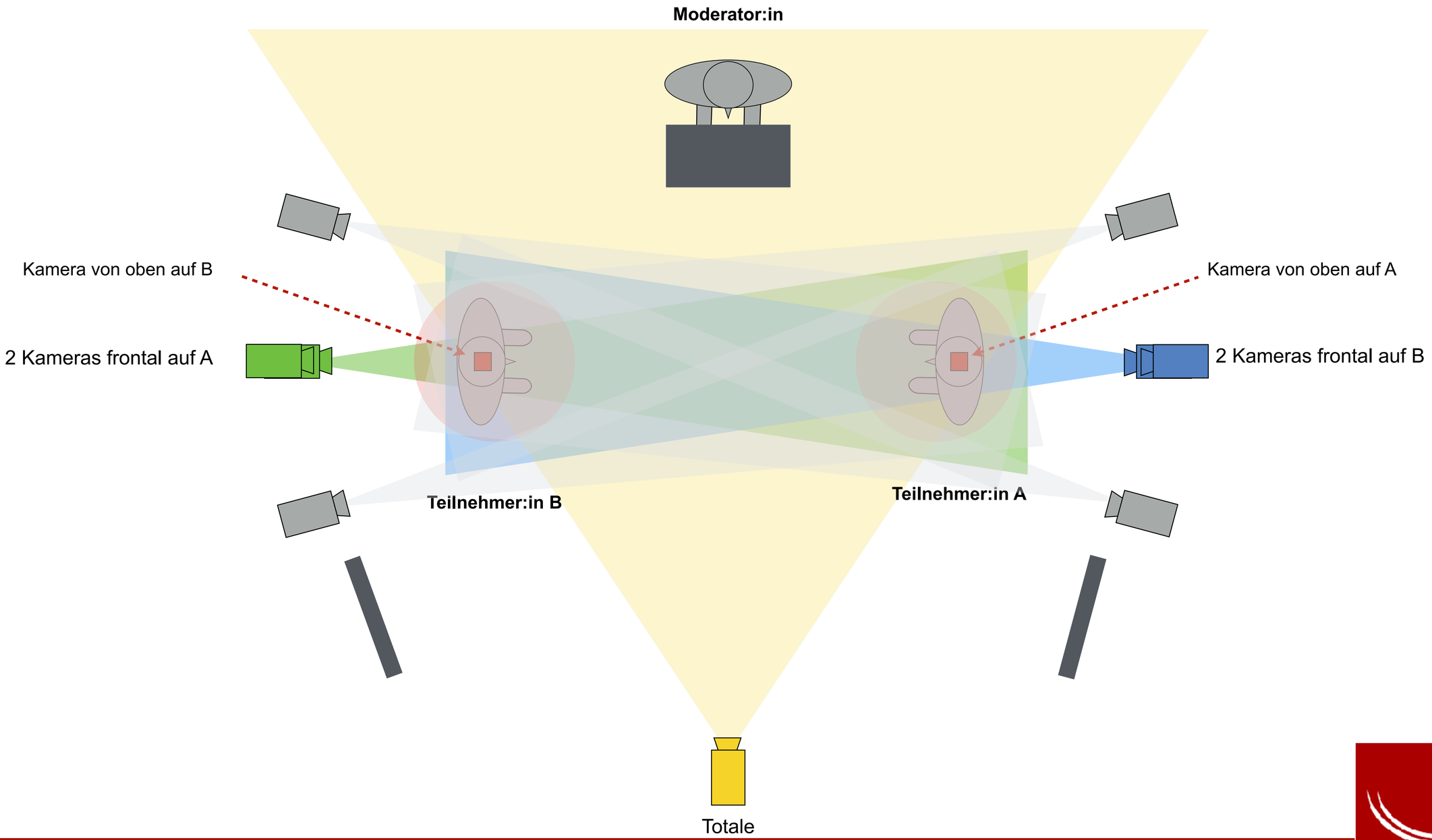
➔ ca. 560 Std. DGS
~ 3,5 Mio. Tokens;
(ø 115 Tokens/min)



DGS-Korpus: Erstellung

Studio: 2. Erhebung

- DGS-Korpus-Studio:
  11 Kameras (6K/4K)
  9 Perspektiven


Moderator:in
Kamera von oben auf B
Kamera von oben auf A
2 Kameras frontal auf A
2 Kameras frontal auf B
Teilnehmer:in B
Teilnehmer:in A
Totale

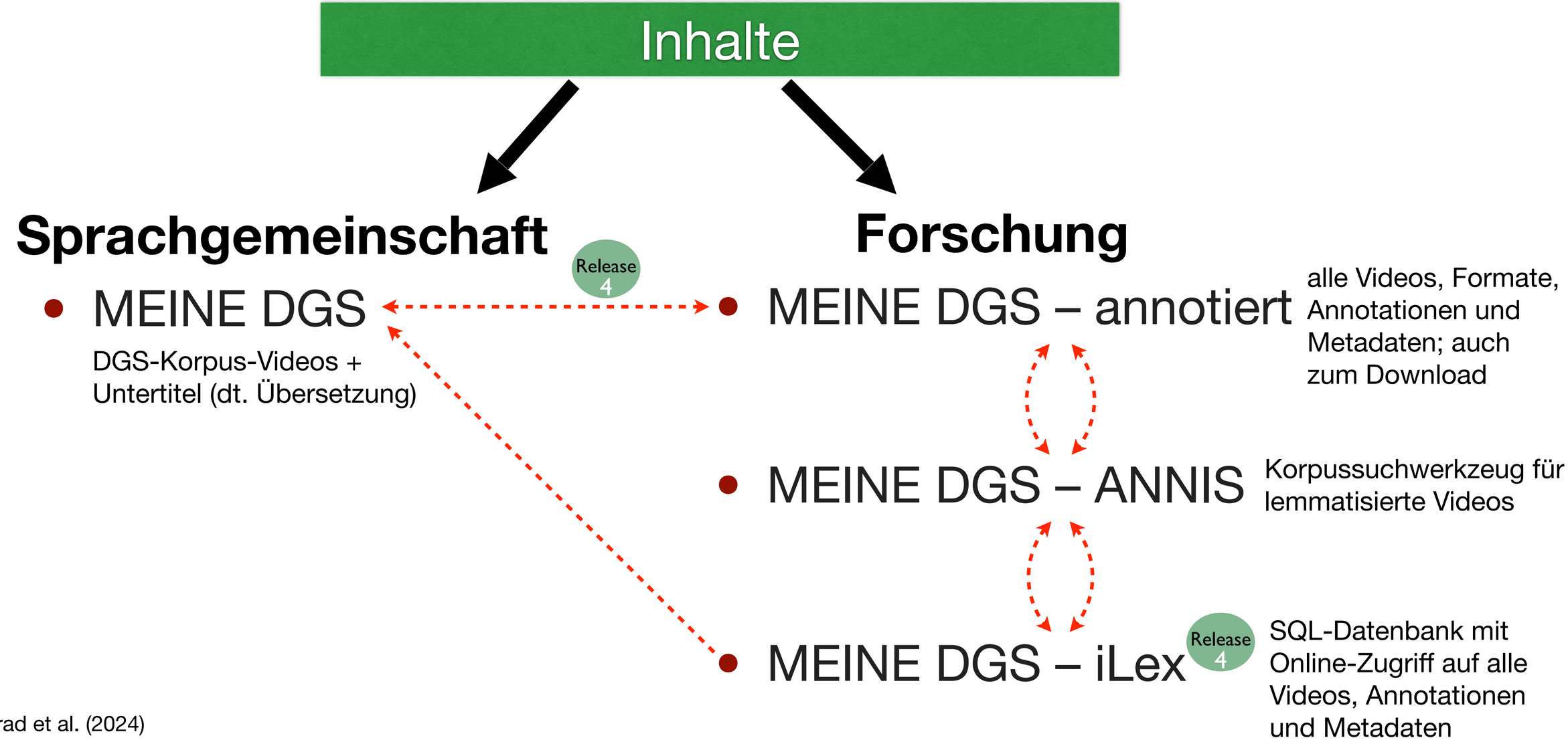
DGS-Korpus: Erstellung

Studio: 2. Erhebung

- DGS-Korpus-Studio:
  11 Kameras (6K/4K)
  9 Perspektiven

Monitor für Videos (Erklärung der Aufgaben) und Stimuli

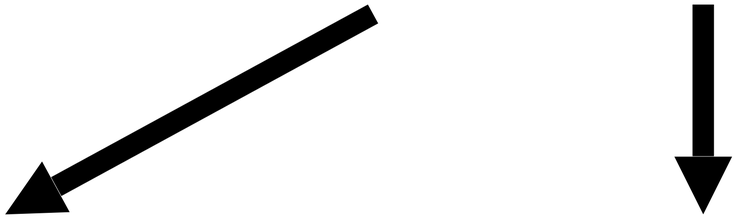
Öffentliches DGS-Korpus



Siehe Konrad et al. (2024)

Öffentliches DGS-Korpus

Inhalte



MEINE DGS

DGS-Videos +

- deutsche Übersetzung
  (als Untertitel, optional)
- Filter nach Region, Altersgruppe, Gesprächsformate, Themen
- Volltextsuche Release 4

MEINE DGS – annotiert

DGS-Videos +

- Online-Transkripte: Übersetzung, Glossen (Types), Mundbild/Mundgestik
  - Sortieren nach Transkript-ID Release 4, TN-Paar (Erhebung/Ort) Release 4, Altersgruppe+Geschlecht Release 4, Format, Annotationstiefe Release 4
- Types-Liste + Filter: Glossen oder Übersetzung (Bedeutungen) Release 4 + Suche Release 4 nach Glosse, Übersetzung, HamNoSys
- Sach-Index: gruppiert Release 4 nach Sachgruppe und Stichwörtern + Links Release 4 zu Transkripten
- Videos sowie verschiedene Inhalte und Formate zum Download; zusätzlich zu OpenPose auch MediaPipe Release 4 und Apple Vision Release 4
- Seiten auf Deutsch und Englisch (außer Mundbild)

Öffentliches DGS-Korpus

Inhalte

MEINE DGS – ANNIS

nur lemmatisierte DGS-Videos

- Alle Annotationen und Metadaten
- Kombinierte Suchabfragen zu Annotationen+/- Metadaten
- Frequenz-Analyse
  - Weiterverarbeitung der Daten (Download CSV-Dokument)
- Dokumentation und Austausch über URL
- Link auf Tokens und Types in MEINE DGS – annotiert Release 4
- Query Wizard Release 4

MEINE DGS – iLex Release 4

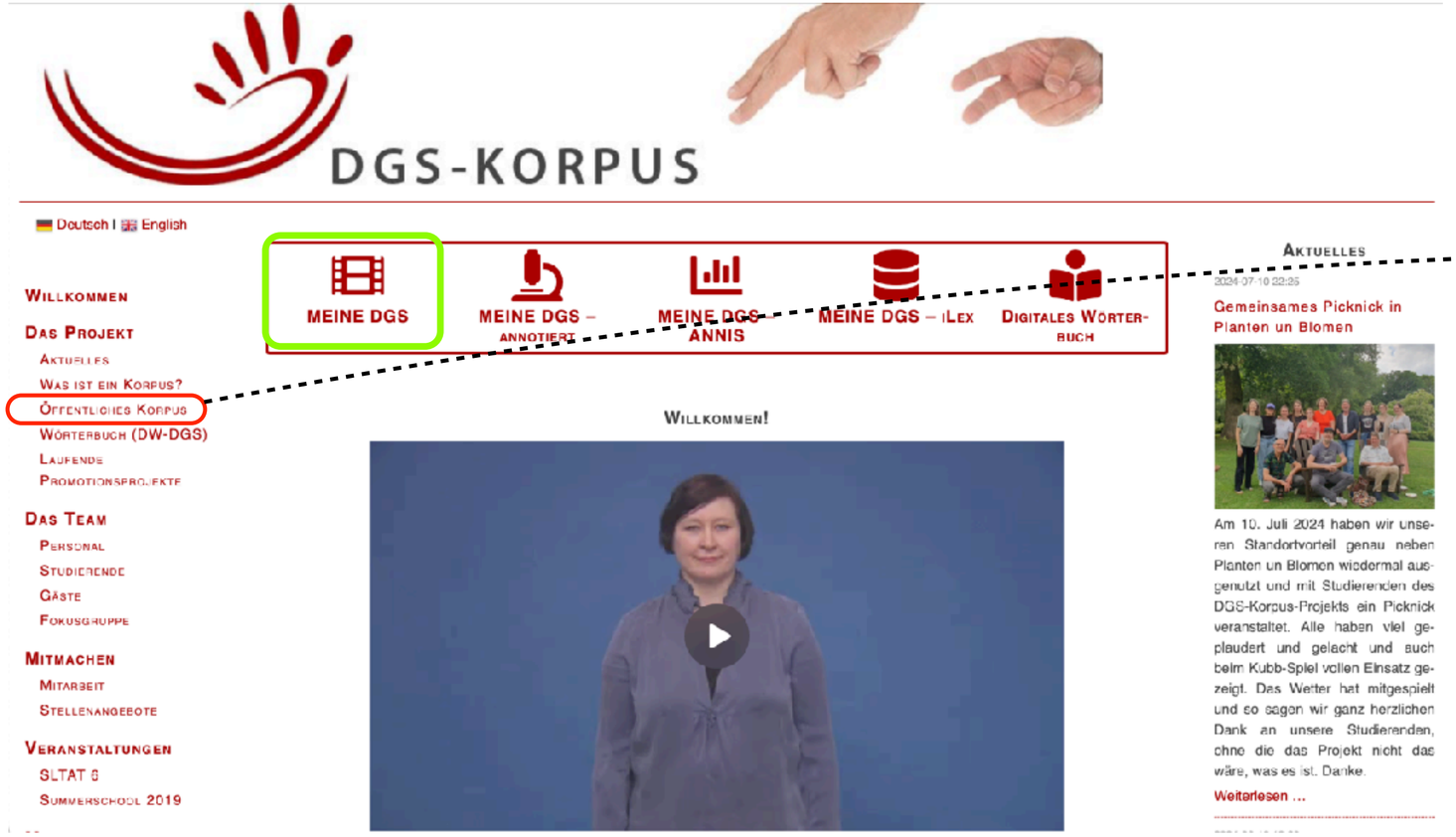
lexikalische Datenbank und Annotationswerkzeug (Transkripte)

- u.a. Type-Hierarchie (Types und Subtypes)
- benutzerdefinierte Sichten auf die Daten (Filter, Listen), SQL-Abfragen
- Alternativ zu ELAN für weitere Annotationen der Daten (nach Download der Videos und iLex-Daten von MEINE DGS – annotiert)

Öffentliches DGS-Korpus

Einführung: Projekt-Homepage: https://dgs-korpus.de

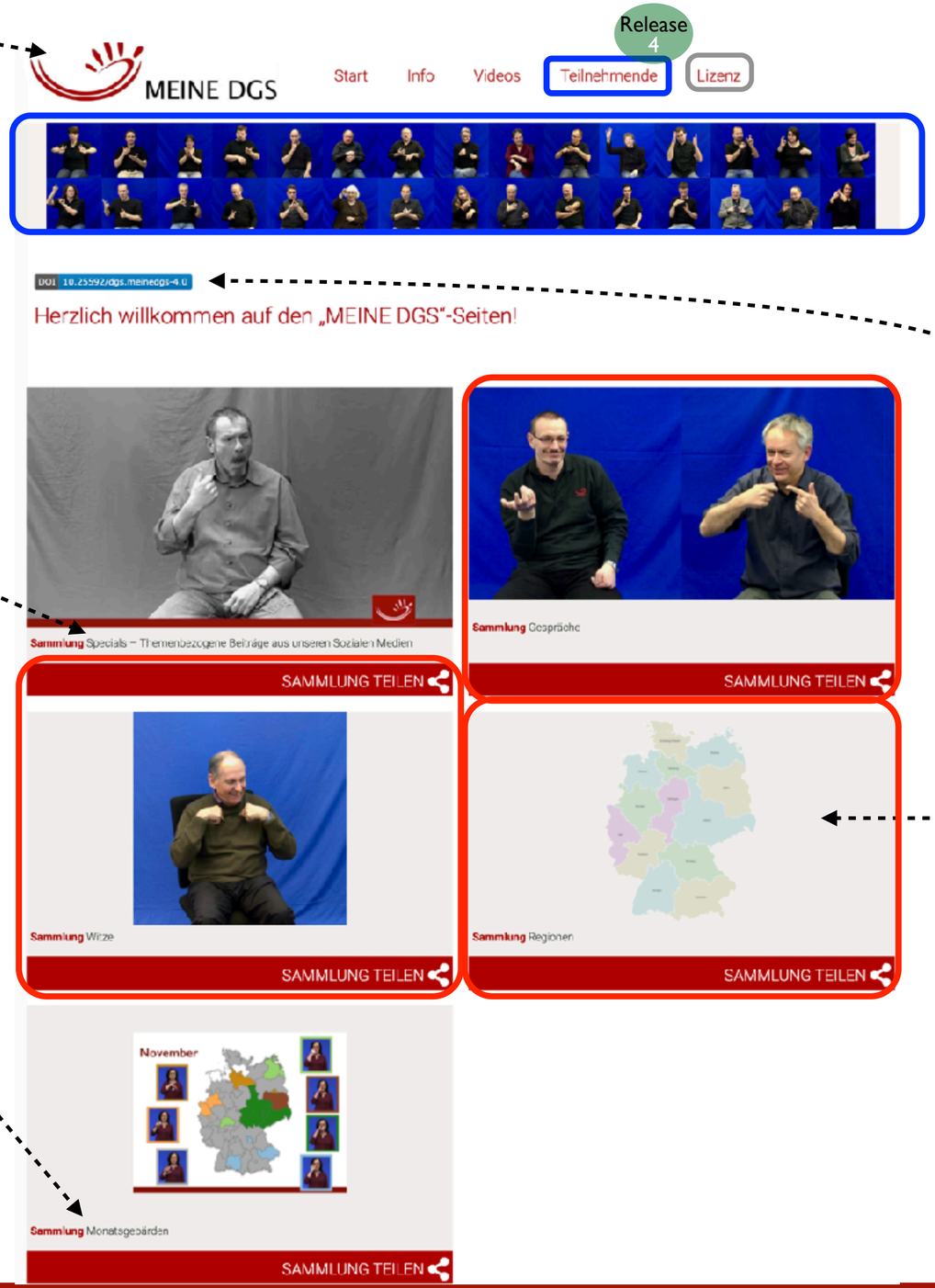
↳ Öffentliches Korpus



Link zur Homepage (DGS-Korpus)

MEINE DGS

https://meine-dgs.de



Startseite

(Kopf mit allen Teilnehmenden;
Klick öffnet neue Seite „Teilnehmende")

Lizenz:

Verwendung der Daten ist „beschränkt auf Unterrichtszwecke".

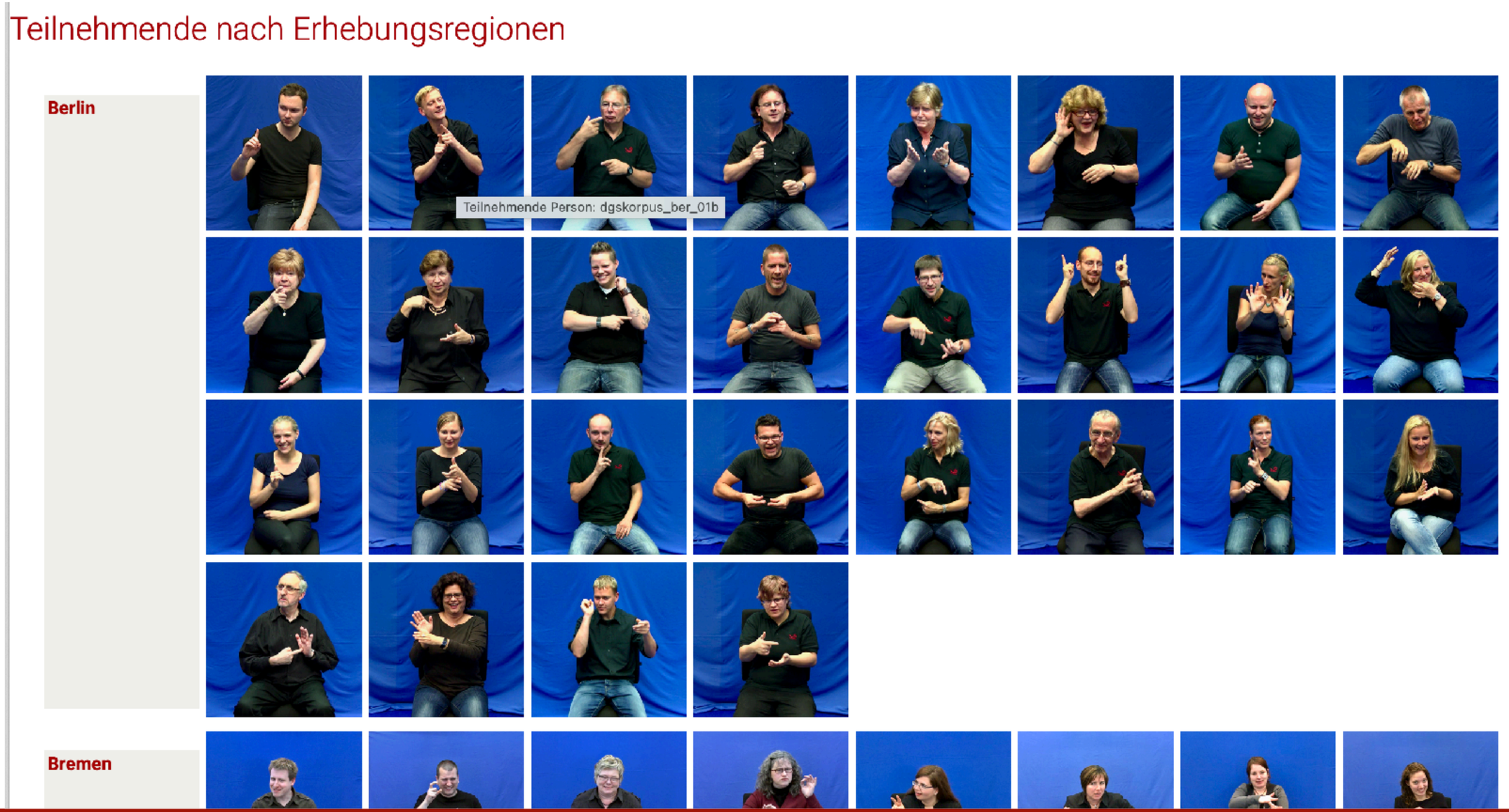
DOI: Link zur Übersichtsseite der verschiedenen Veröffentlichungen (Releases) von MEINE DGS + Zitierhinweis (s.u.)

Zusätzliche Inhalte:

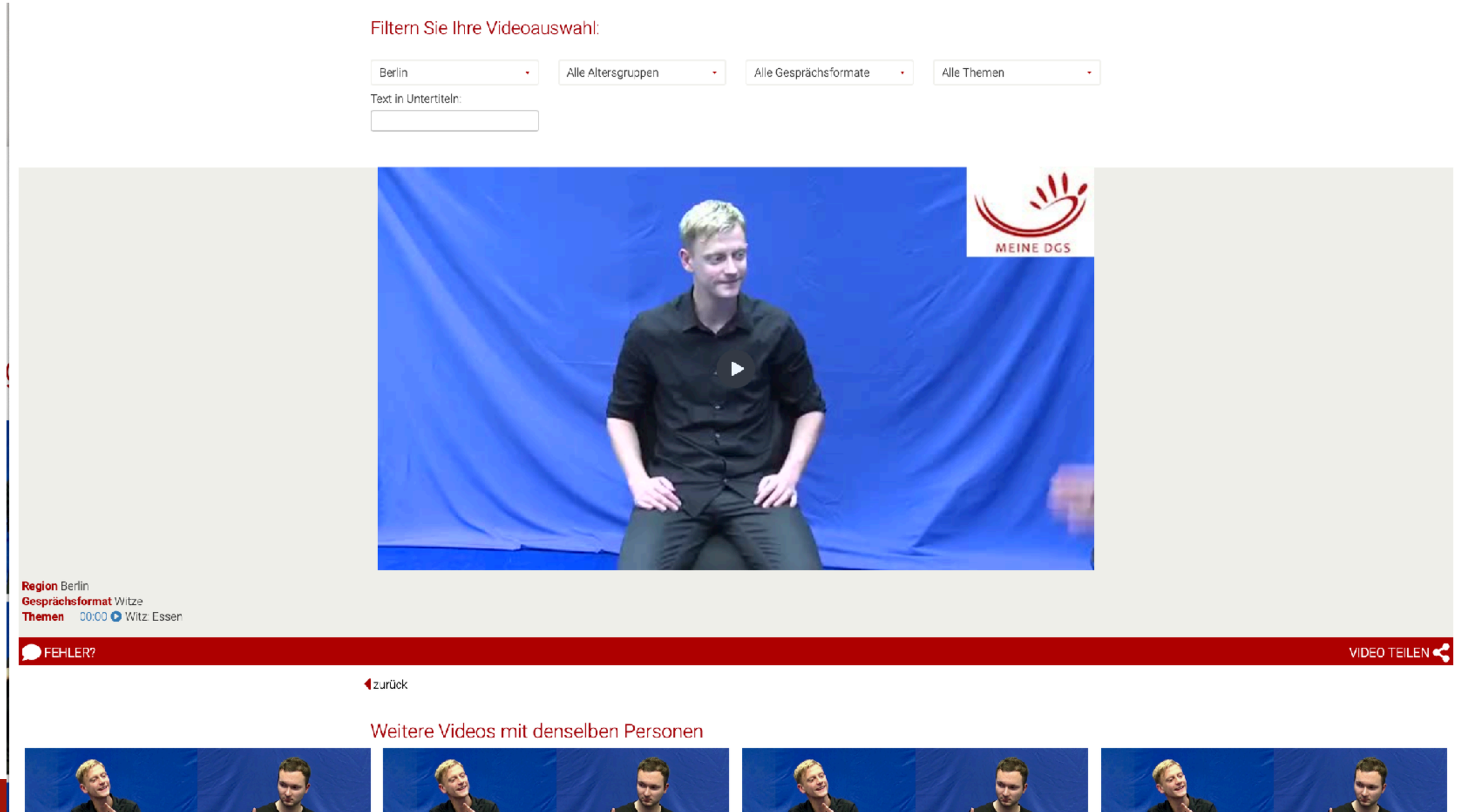
- Sammlung: Specials
- Sammlung: Monatsgebärden (Korpus- und Feedback-Daten)

Auswahl nach Erhebungsregion (Karte)

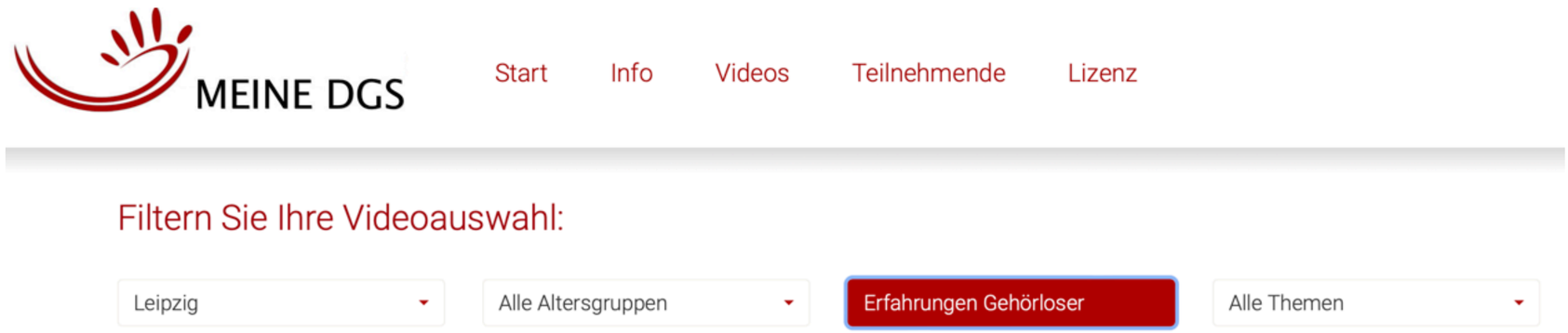
MEINE DGS


Teilnehmende nach Erhebungsregionen
Berlin
Teilnehmende Person: dgskorpus_ber_01b
Bremen

MEINE DGS

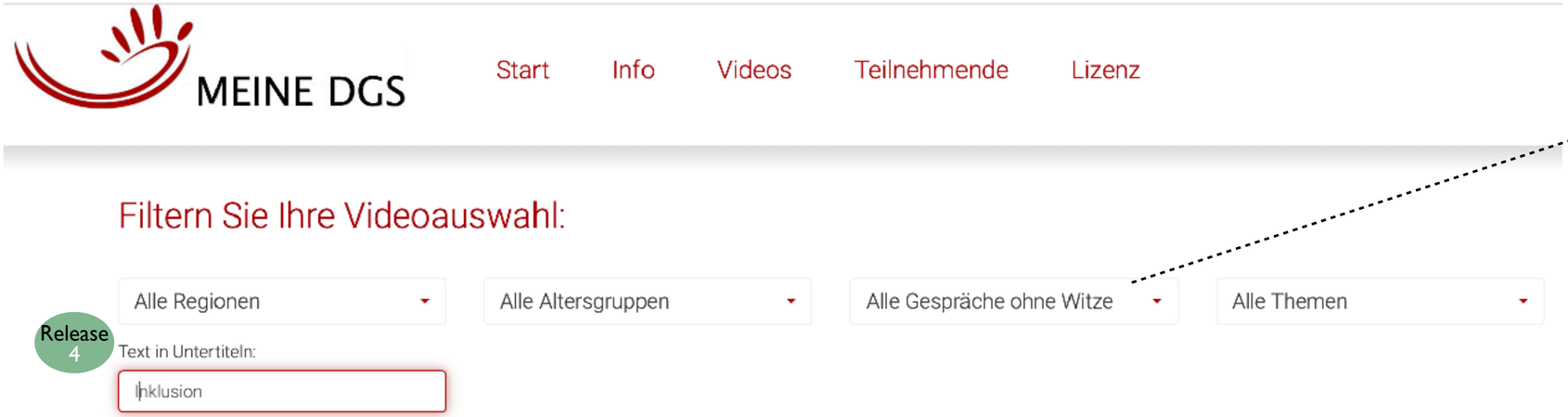

Filtern Sie Ihre Videoauswahl:
Berlin
Alle Altersgruppen
Alle Gesprächsformate
Alle Themen
Text in Untertiteln:
MEINE DGS
Region Berlin
Gesprächsformat Witze
Themen 00:00 Witz: Essen
FEHLER?
VIDEO TEILEN
zurück
Weitere Videos mit denselben Personen

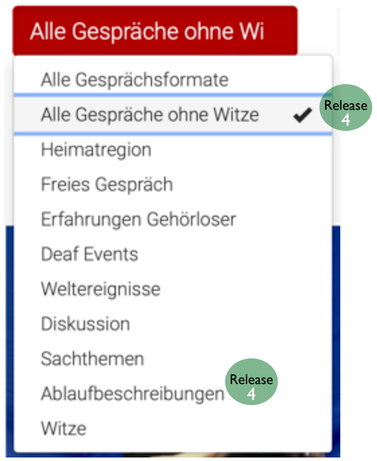
MEINE DGS



➔ 6 Gespräche aus einer Region

oder:



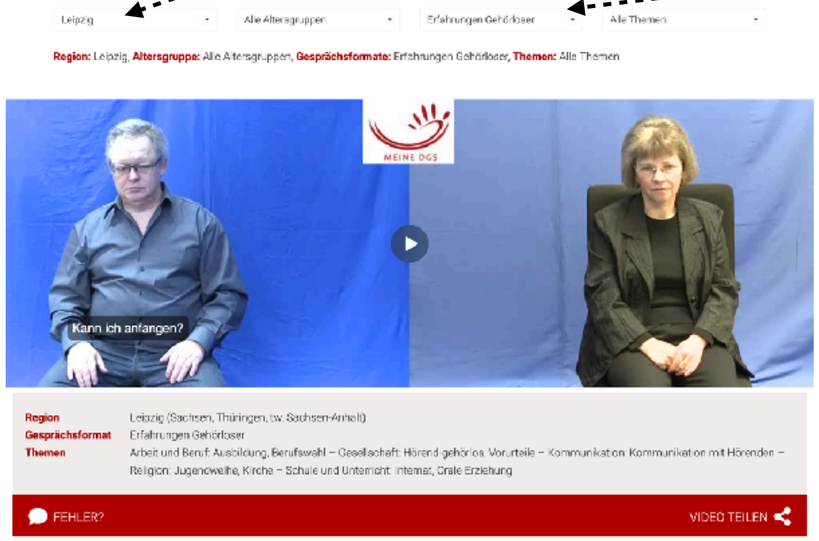


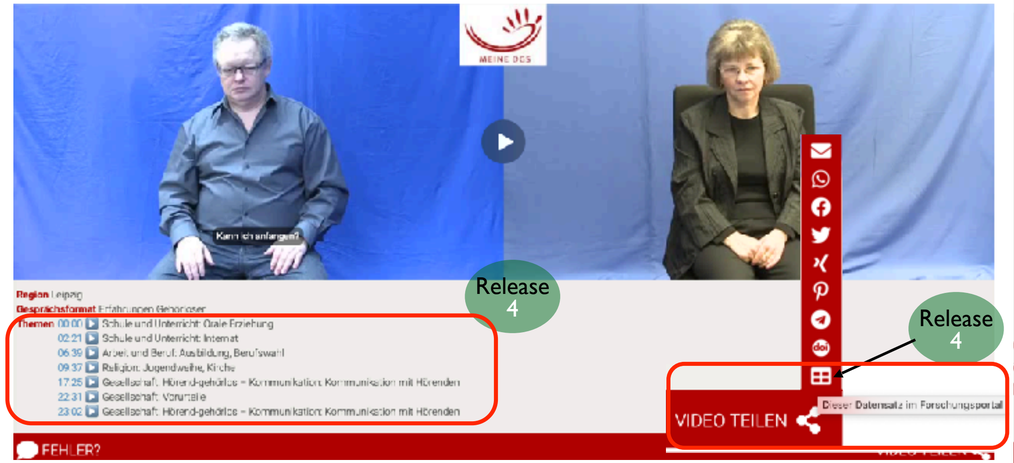
➔ 6 Gespräche aus versch. Regionen

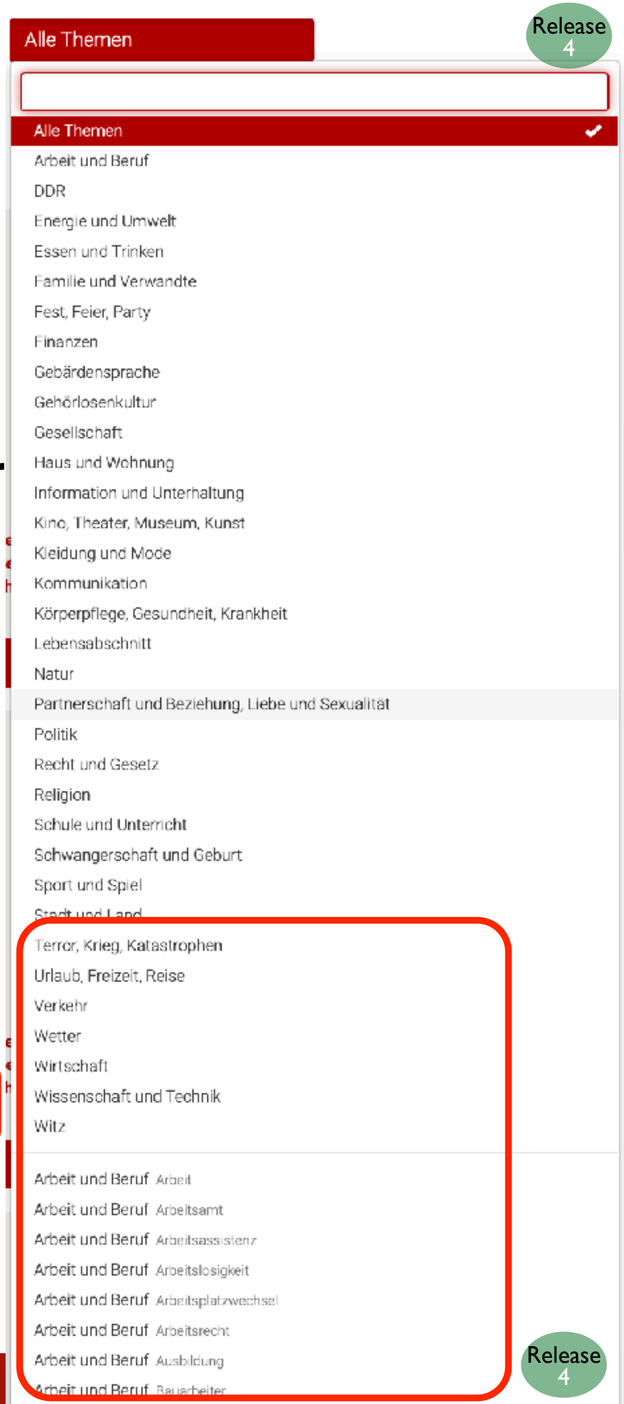
MEINE DGS

Filter:

Region: Leipzig, Gespr.format: Erfahrungen Gehörloser

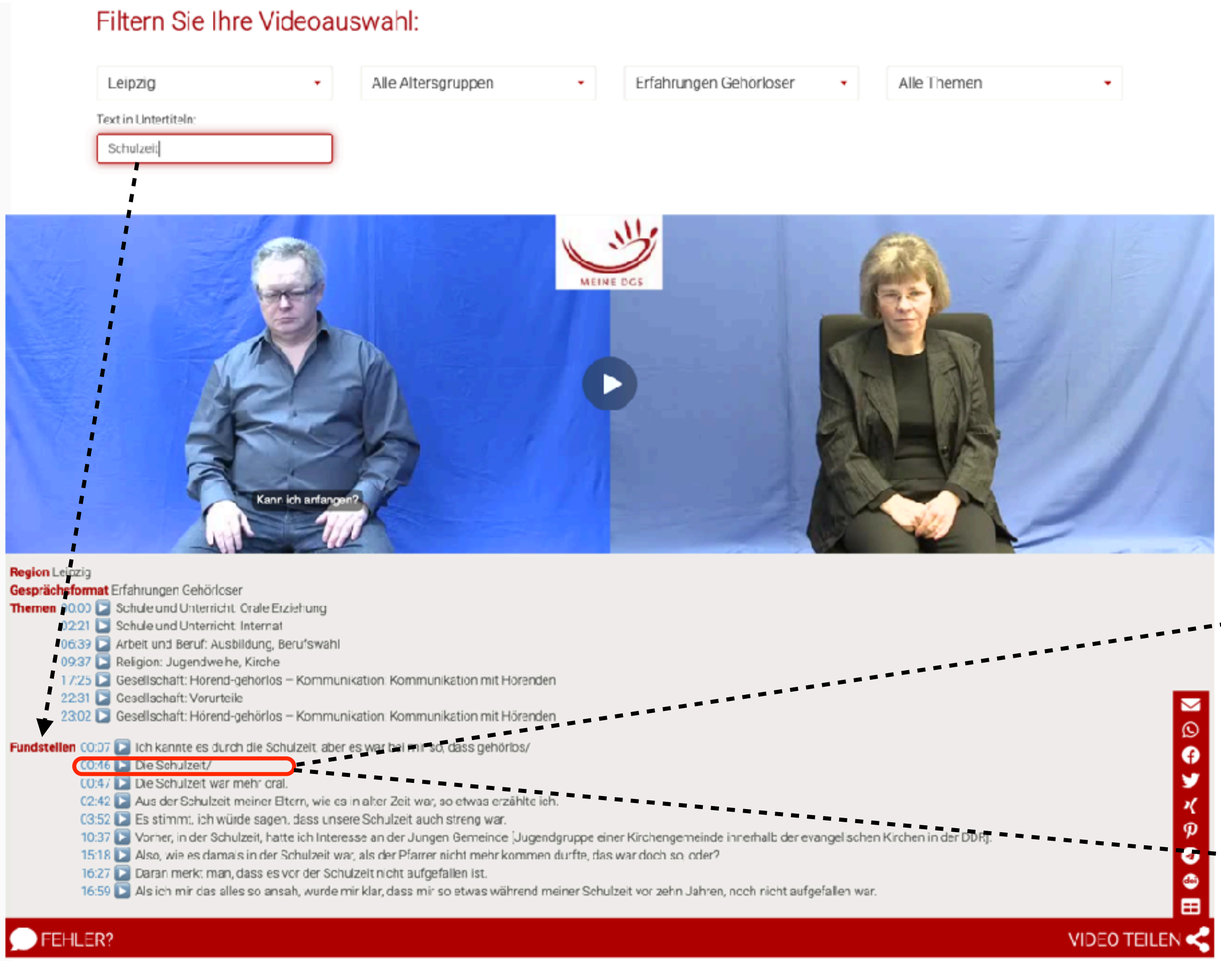






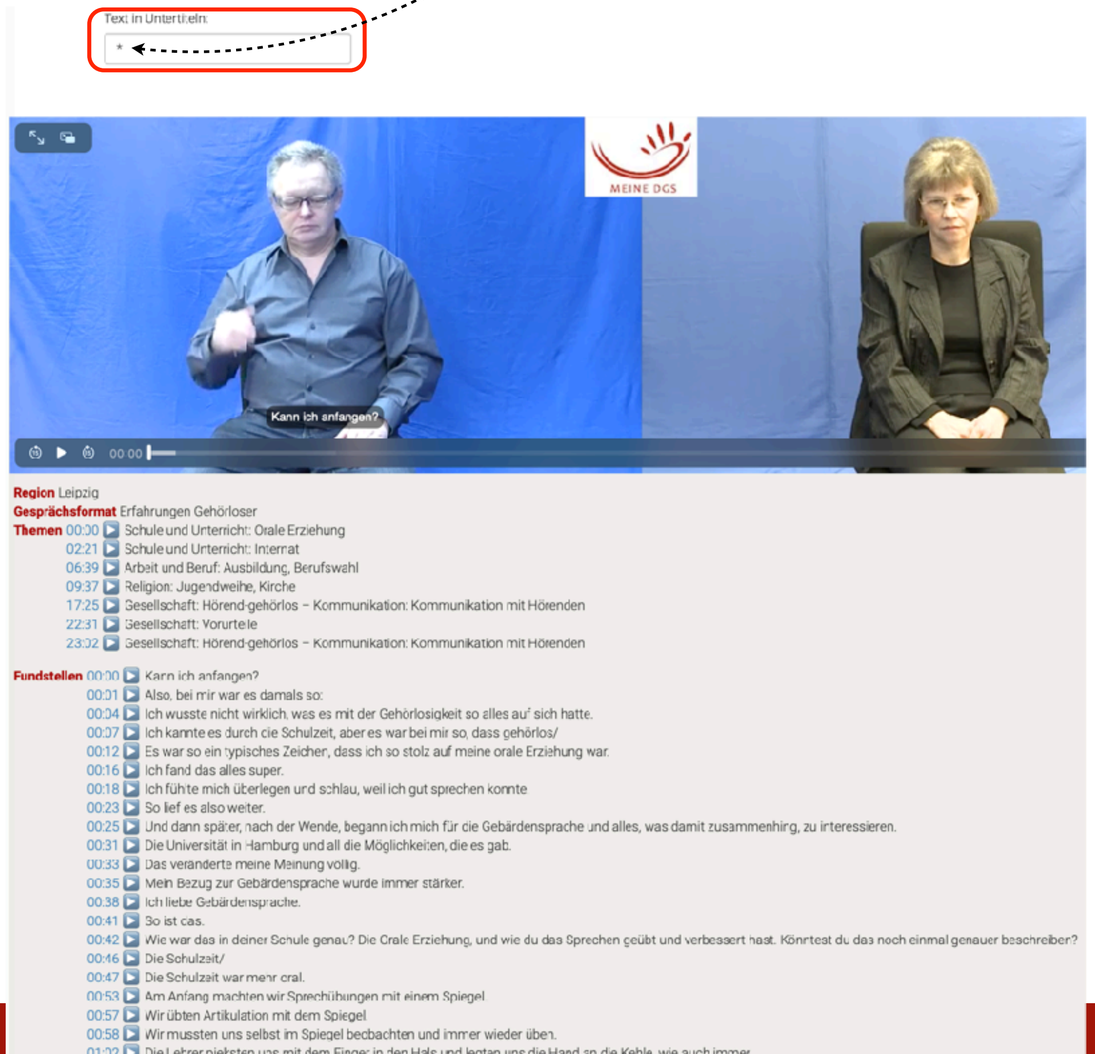
MEINE DGS

Suche in Untertiteln (~ dt. Übersetzung)



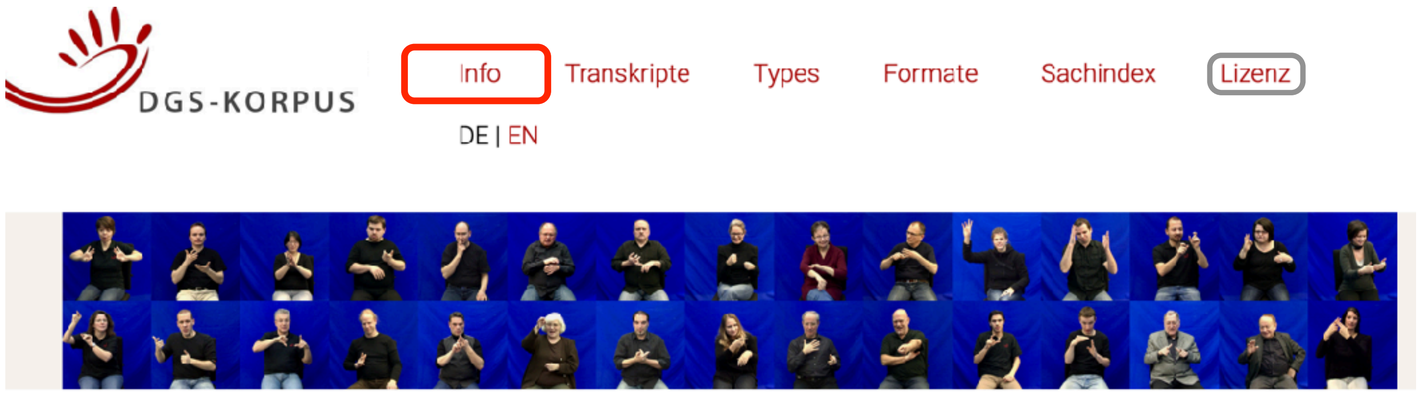
MEINE DGS

Suche in Untertiteln (= dt. Übersetzung): * = alles (Übersetzung von Anfang bis Ende)



MEINE DGS – annotiert

https://ling.meine-dgs.de



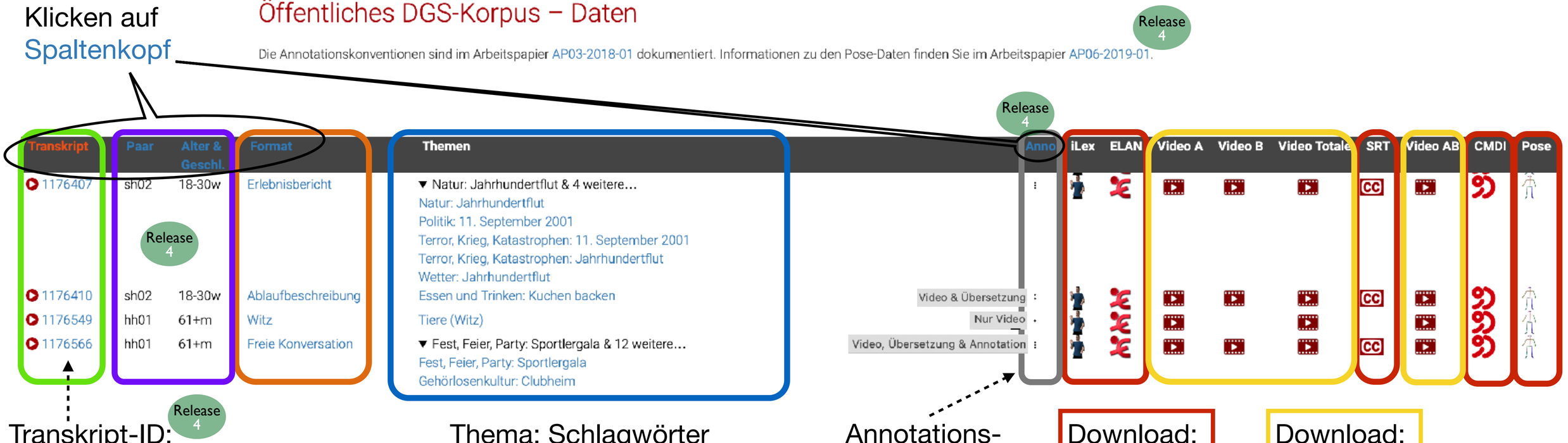
U.a. Annotationskonventionen!

Lizenz:

Verwendung der Daten nur „für sprachwissen-schaftliche Forschungs-zwecke“.

MEINE DGS – annotiert: Transkripte

Zum Sortieren:
Klicken auf
Spaltenkopf



Transkript-ID:
Link zur Online-Ansicht des Transkripts

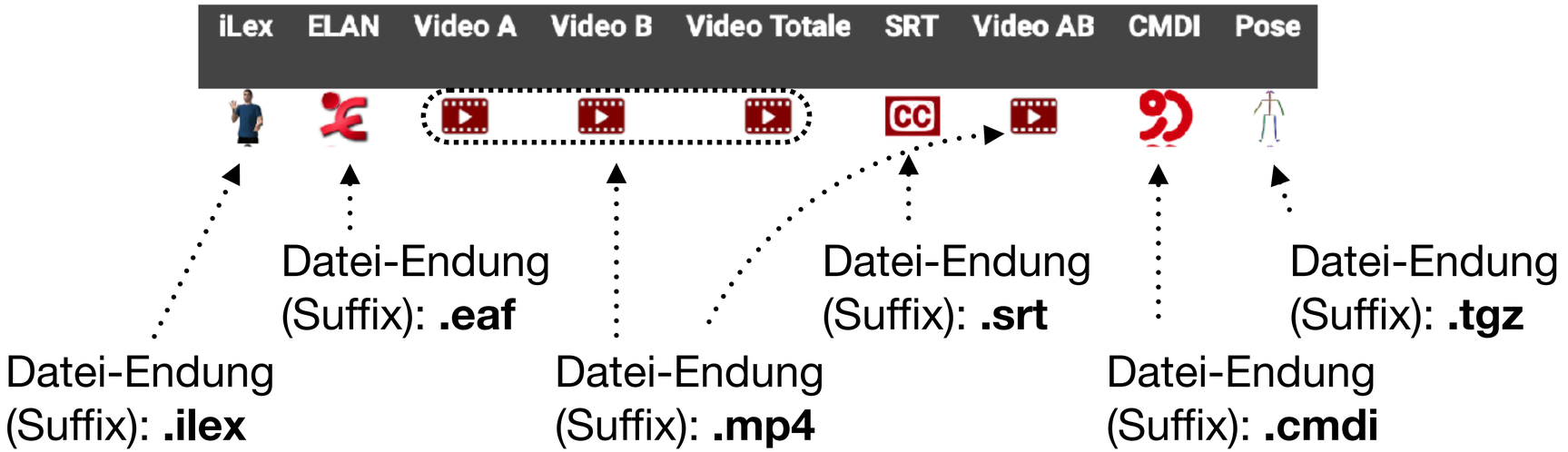
Thema: Schlagwörter

Annotations-tiefe

Download: Dateien

Download: Videos

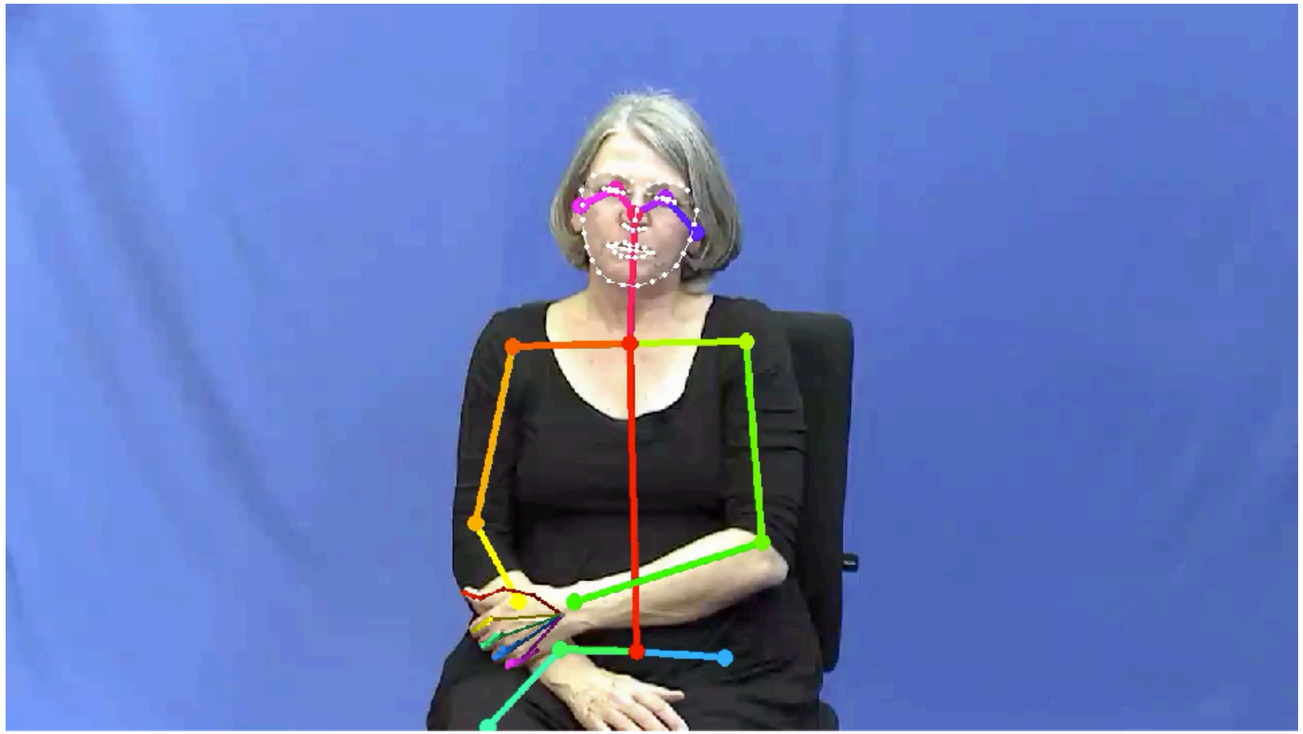
Download-Dateien



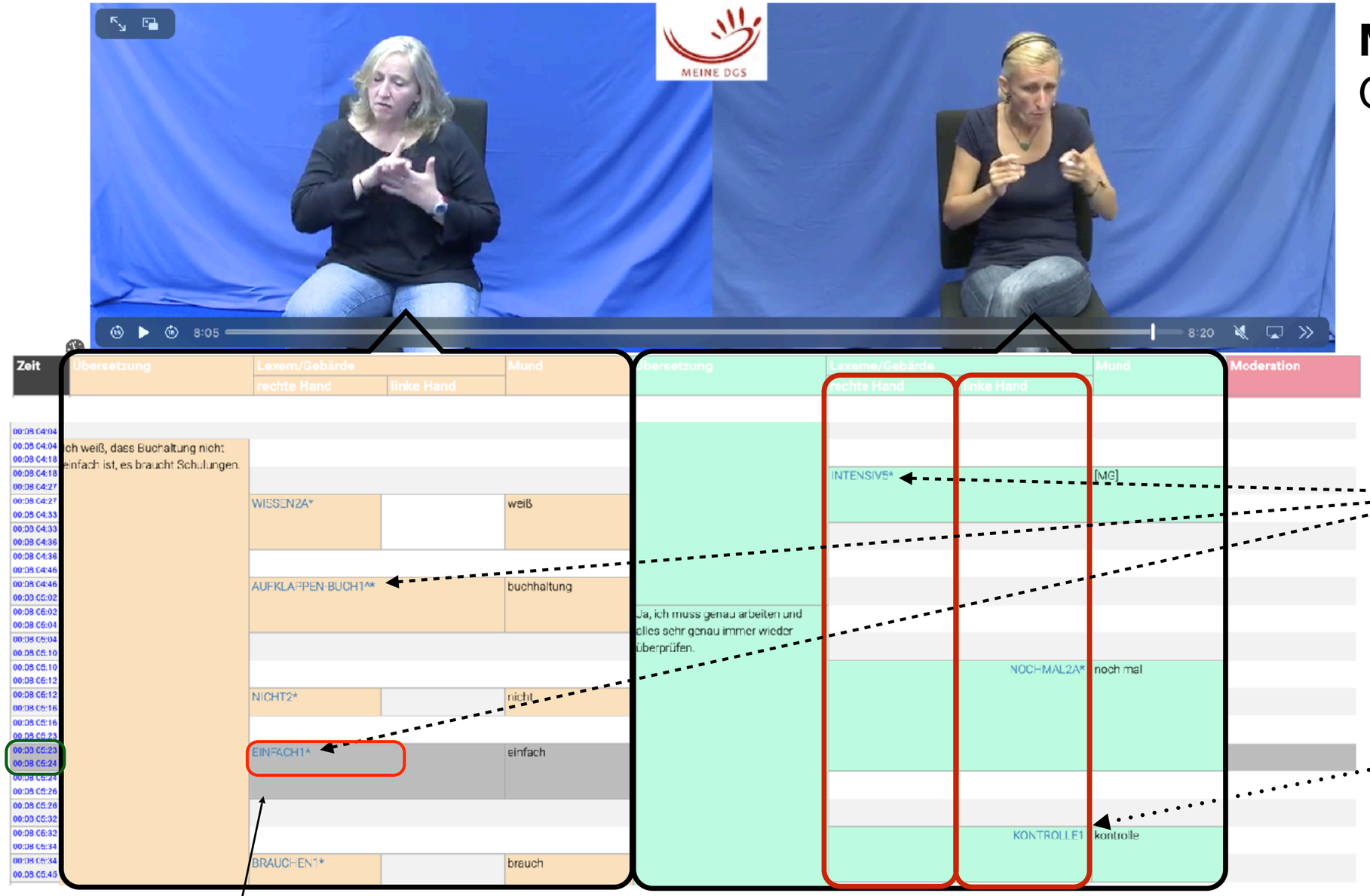
- **.ilex**: Importieren in iLex-DB (PostreSQL); enthält auch Metadaten (s. CMDI)
- **.eaf**: Öffnen mit ELAN
- **.mp4**: optional lokal gespeichert; (eingebunden in iLex oder ELAN-Dateien)
- **.srt**: Untertitel, gedacht für MaxQDA (in Verbindung mit Video AB)
- **.cmdi**: XML-Datei; Programm mit xml-Einstellung, z.B. Atom (farbl. Hervorhebung der Textteile; s. letzte Folie) Alternativ: mit **.xml**-Endung in Browser (z.B. Safari, Chrom, FireFox) öffnen
- **.tgz**: komprimierte/Zip-Datei, die zunächst dekomprimiert/entzippt werden muss; enthält OpenPose, MediaPipe und Apple Vision-Dateien zu den Videos

OpenPose

- Für jeden Punkt werden drei Werte ausgegeben:
  - X-Koordinate,
  - Y-Koordinate,
  - Zuverlässigkeit der Bestimmung (in %).
- Weitere Infos zu OpenPose-Daten des Öffentlichen DGS-Korpus: Schulder, Marc, & Hanke, Thomas. (2020, May 11). OpenPose in the Public DGS Corpus (Version 2). http://doi.org/10.25592/uhhfdm.1866.
- **Weitere Datenformate**: Release 4
  - MediaPipe
  - Apple Vision



MEINE DGS – annotiert:
Online-Transkript



zweihändige Gebärde
(re. Hand ist dominant/aktiv)

zweihändige Gebärde
(li. Hand ist dominant/aktiv)

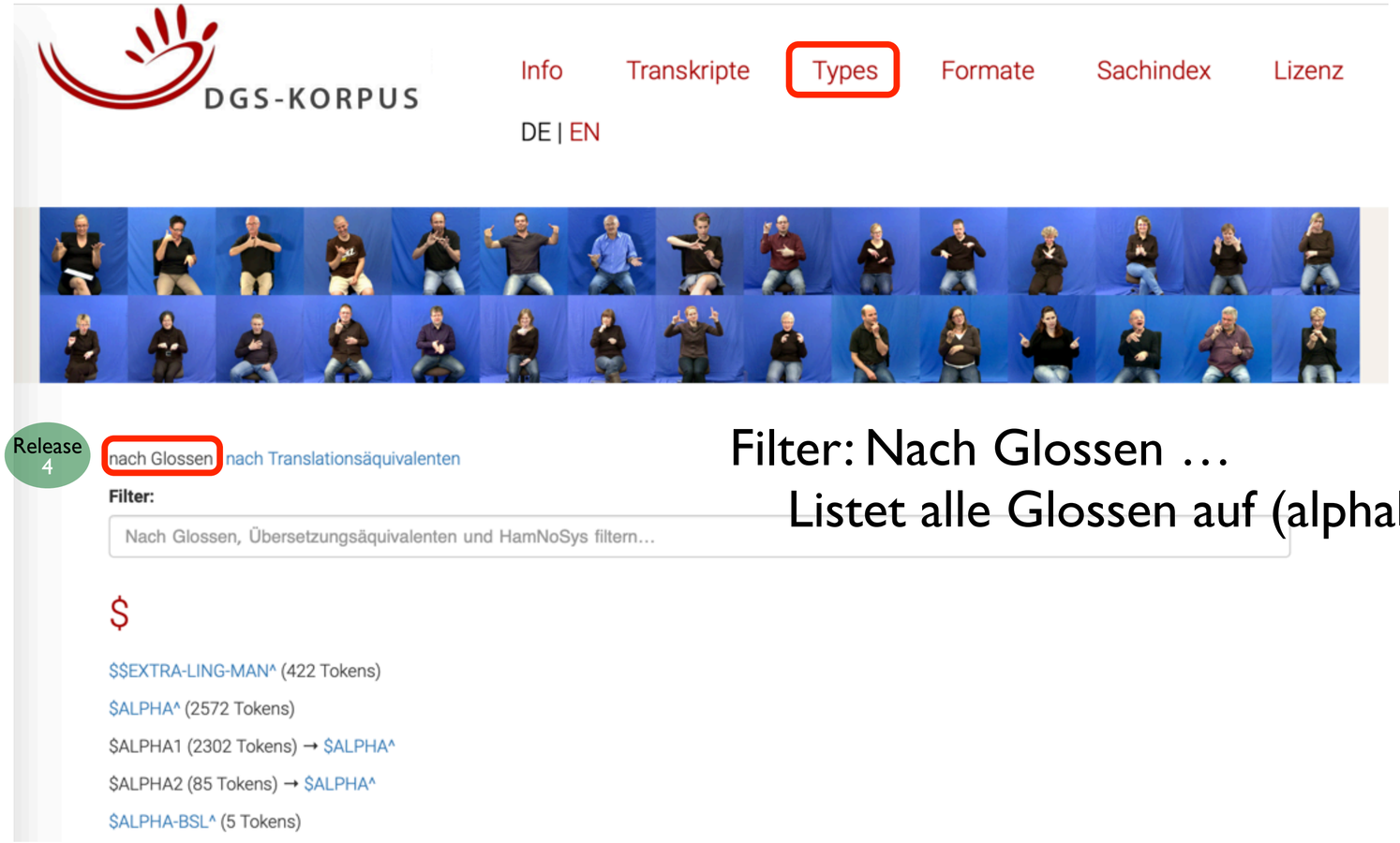
re. Hand li. Hand

Timecode 00:08:05:23 =
Anker #t00080523
auf Zeile im Transkript

Zur Types-Liste: Klicken auf Glosse

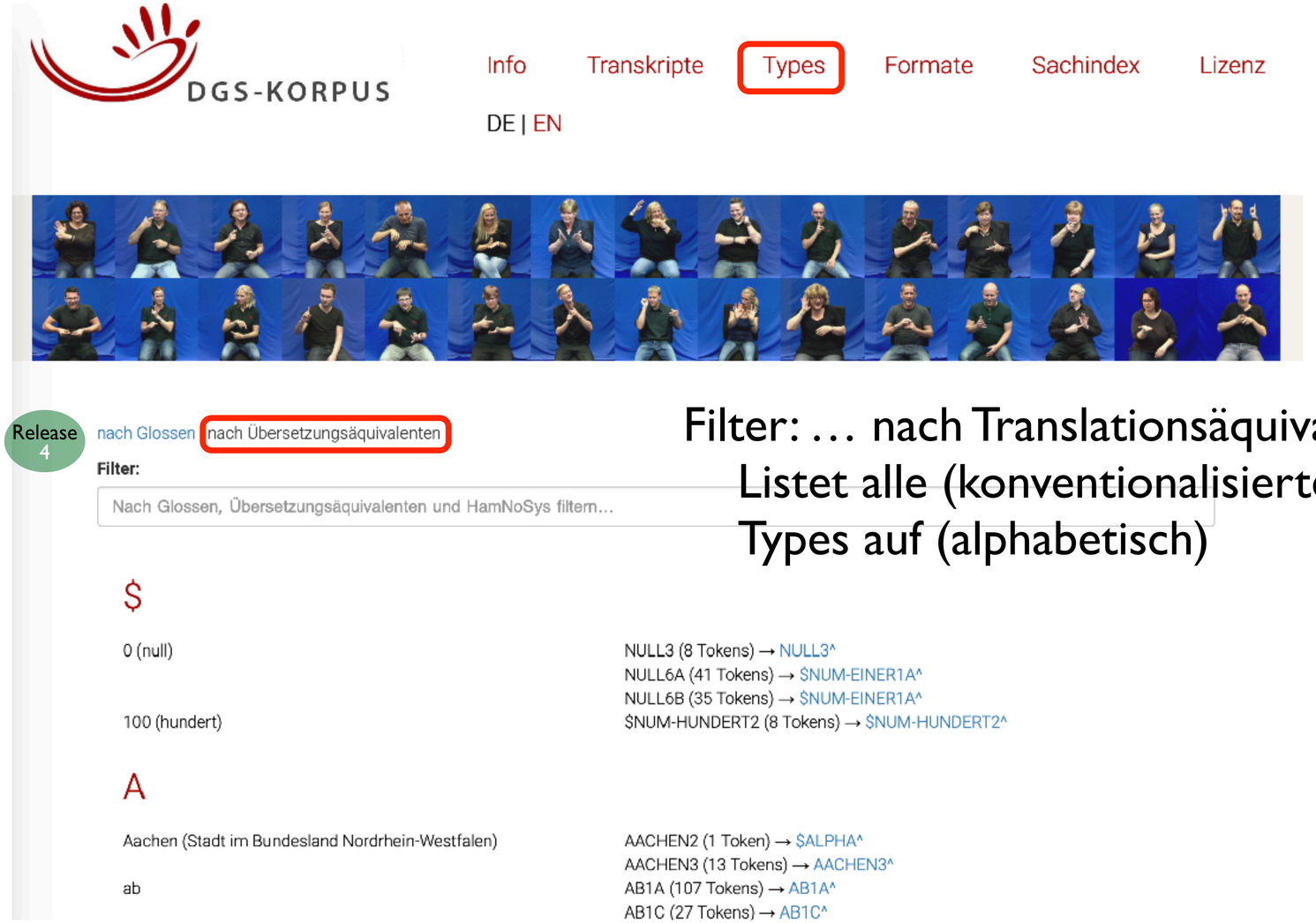
aus: dgskorpus_ber_08: Erfahrungen als Gehörloser Link 00:08:05:2300:08:05:24 = https://www.sign-lang.uni-hamburg.de/meinedgs/html/1418889_de.html#t00080523

MEINE DGS – annotiert: Types



Filter: Nach Glossen …
Listet alle Glossen auf (alphabetisch)

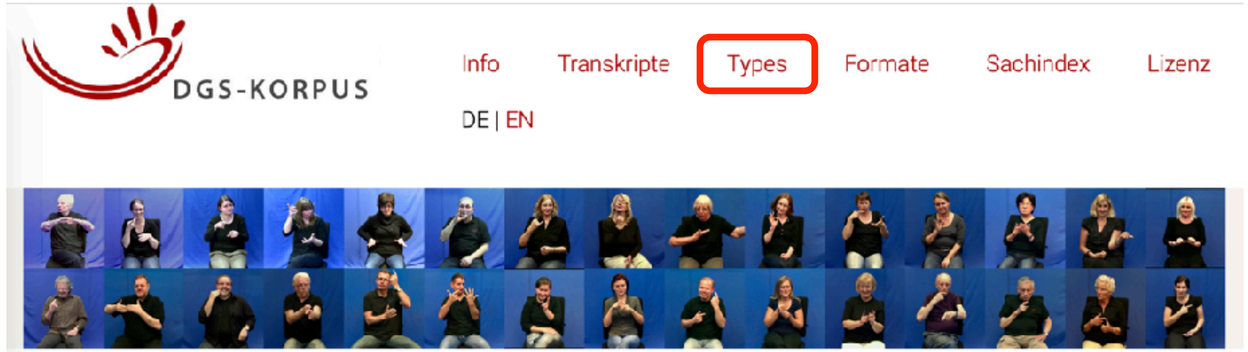
MEINE DGS – annotiert: Types



Filter: … nach Translationsäquivalenten

Listet alle (konventionalisierten) Bedeutungen der Types auf (alphabetisch)

MEINE DGS – annotiert: Types



Release 4

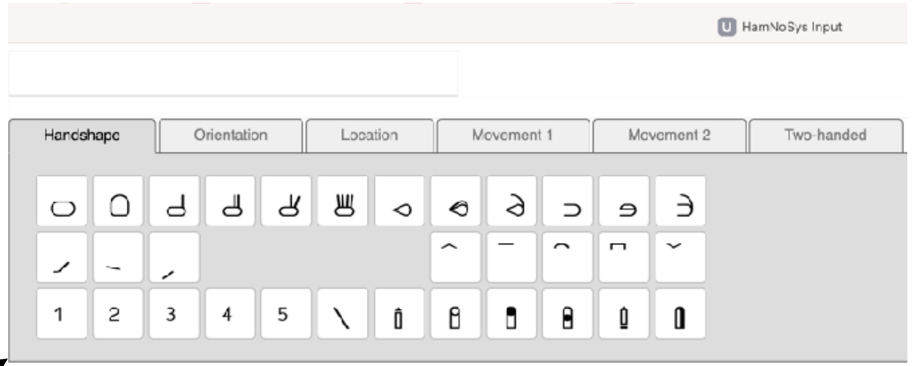
Filter: Suchtext

Sucht in Glossen, Bedeutungen und HamNoSys

HamNoSys

DGS-Korpus Homepage ↳ Tools ↳ HamNoSys

(https://www.sign-lang.uni-hamburg.de/dgs-korpus/hamnosys.html)



Dokumente

HamNoSys-Übersicht (englisch)

(Hanke, T. (2004), "HamNoSys - representing sign language data in language resources and language processing contexts." In: Streiter, Oliver, Vettori, Chiara (eds): LREC 2004, Workshop proceedings : Representation and processing of sign languages. Paris : ELRA, 2004, - pp. 1-6.)

HamNoSys-Dokumentation (englisch)

HamNoSys-Handformen (englisch)

Schriftsysteme

HamNoSys 4, plattformunabhängig

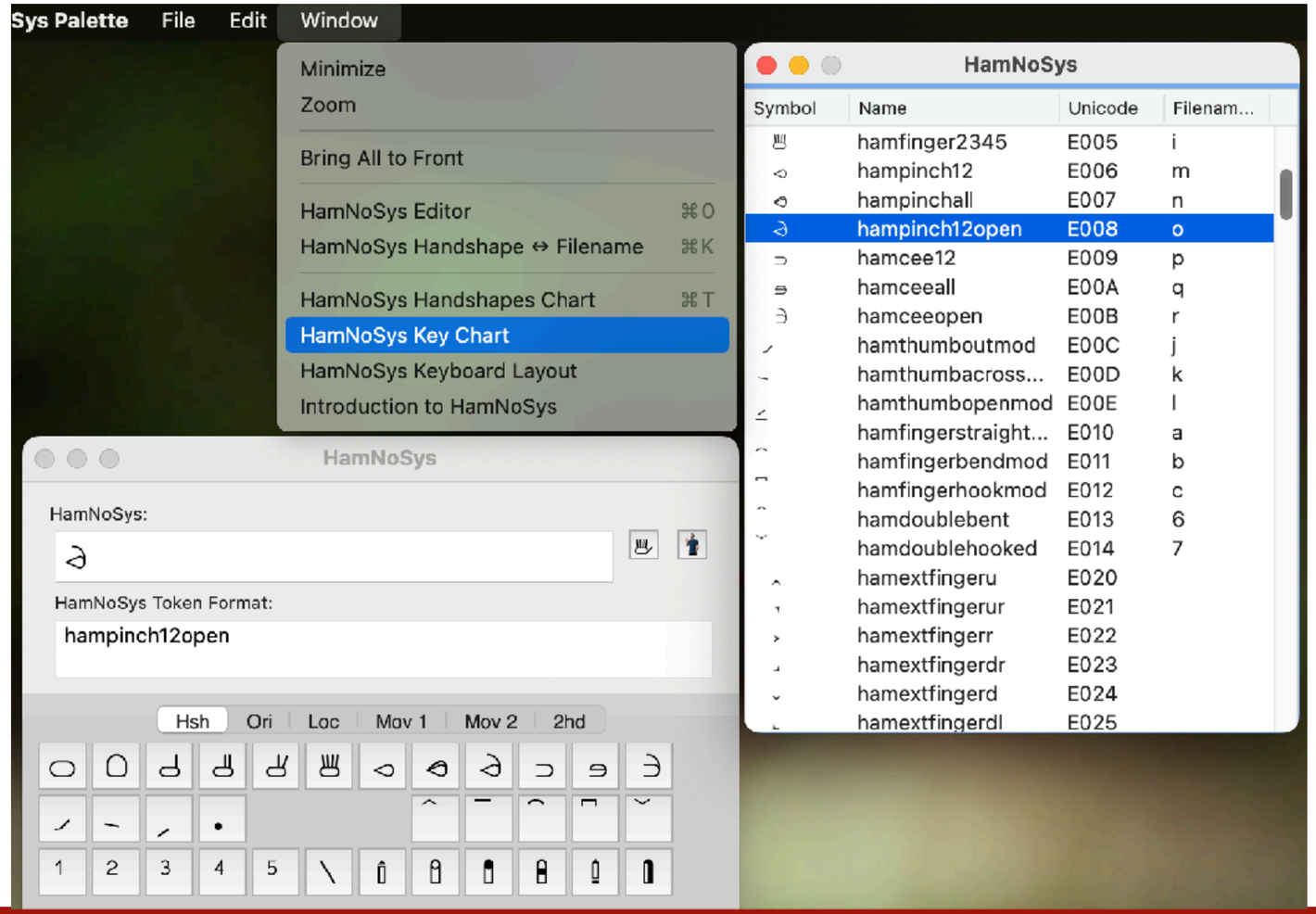
DOI 10.25592/uhhfdm.9724

Plattformunabhängige Eingabepalette (Browser-basiert, Zeichensatz muss nicht installiert sein)

Download:
hamnosys_4_1_7.zip

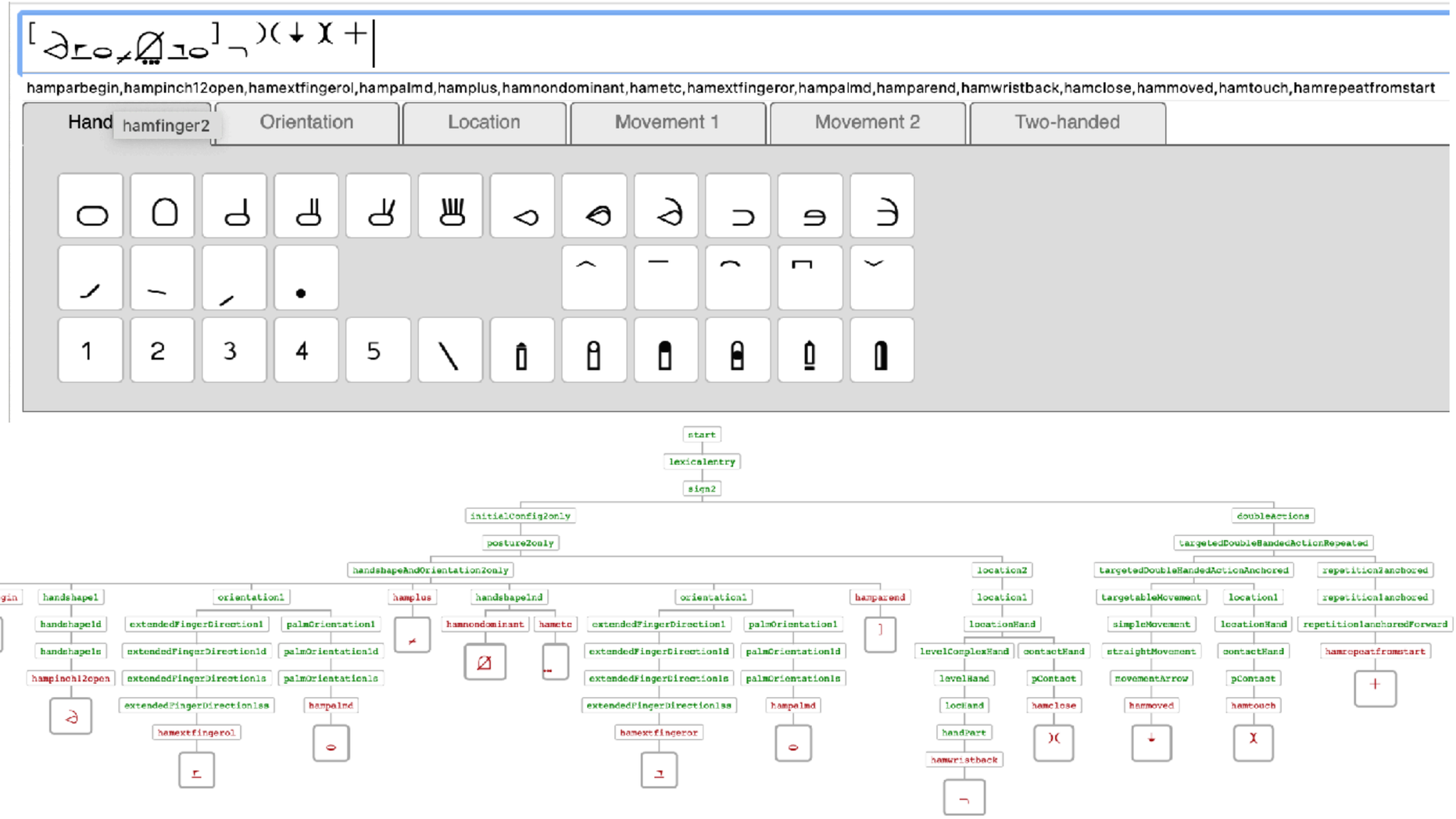
HamNoSys

HamNoSys Palette.app



HamNoSys

Plattformunabhängige Eingabepalette



MEINE DGS – annotiert: Types

00:08:04:04 – 00:08:04:18	Ich weiß, dass Buchaltung nicht einfach ist, es braucht Schulungen.		
00:08:04:18 – 00:08:04:27		WISSEN2A*	weiß
00:08:04:27 – 00:08:04:33			
00:08:04:33 – 00:08:04:36		AUFKLAPPEN-BUCH1^*	buchhaltung
00:08:04:36 – 00:08:04:46			
00:08:04:46 – 00:08:05:02			
00:08:05:02 – 00:08:05:04		NICHT2*	nicht
00:08:05:04 – 00:08:05:10			
00:08:05:10 – 00:08:05:12			
00:08:05:12 – 00:08:05:16		EINFACH1*	einfach
00:08:05:16 – 00:08:05:23			
00:08:05:23 – 00:08:05:24			
00:08:05:24 – 00:08:05:26			

Erwartbares Mundbild: „einfach“

(Konventionalisierte) Bedeutungen des Subtypes EINFACH1

1177704-… hh04 | 46-60m Ich nehme dann einfach Holzkohle und lege sie in den Grill.

	EINFACH1	HOLZ11	KOHLE2*	WERFEN1^*
r				
l				
m	einfach	holzkohle		[MG]

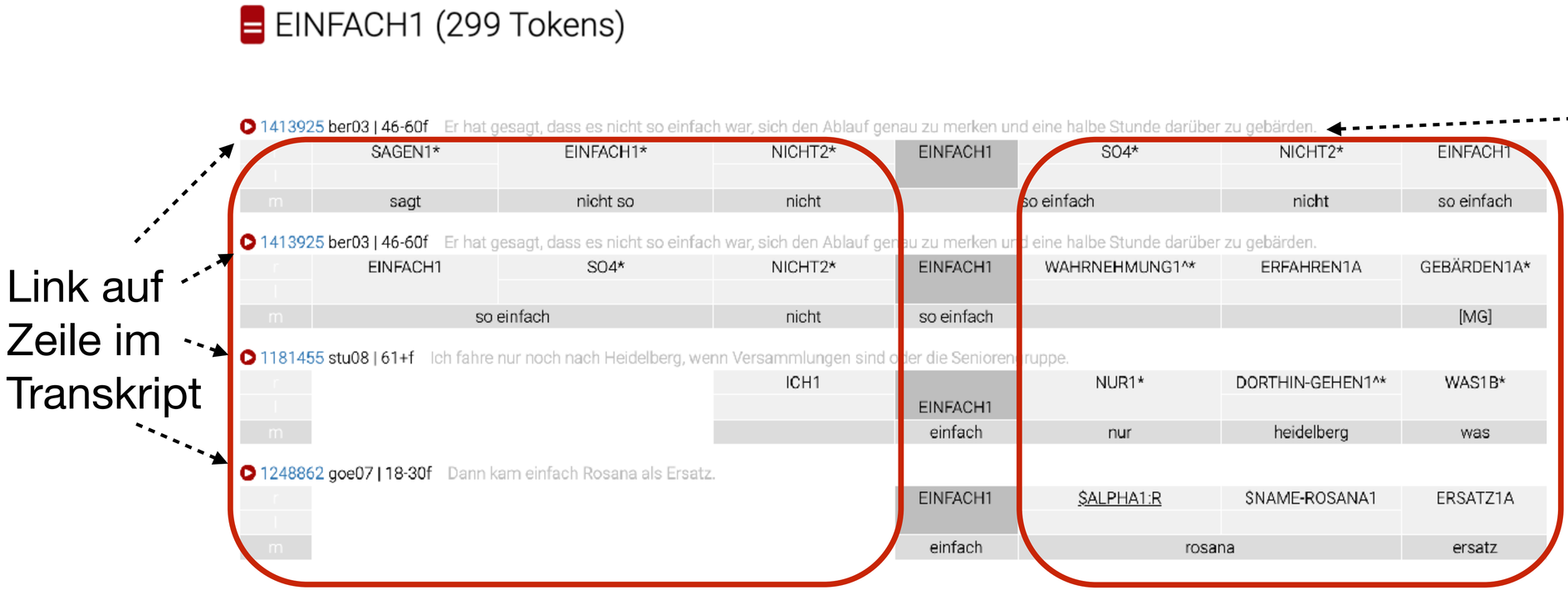
1200691 stu02 | 18-30f Ich hatte es als Kind nicht so einfach.

	ICH1	KIND2*	EINFACH1
r			
l			
m	ich	als kind	nicht so einfach

MEINE DGS – annotiert: Types

Zum Seitenanfang scrollen

Durch Klicken auf Glossenüberschrift Token-Zeilen ausblenden



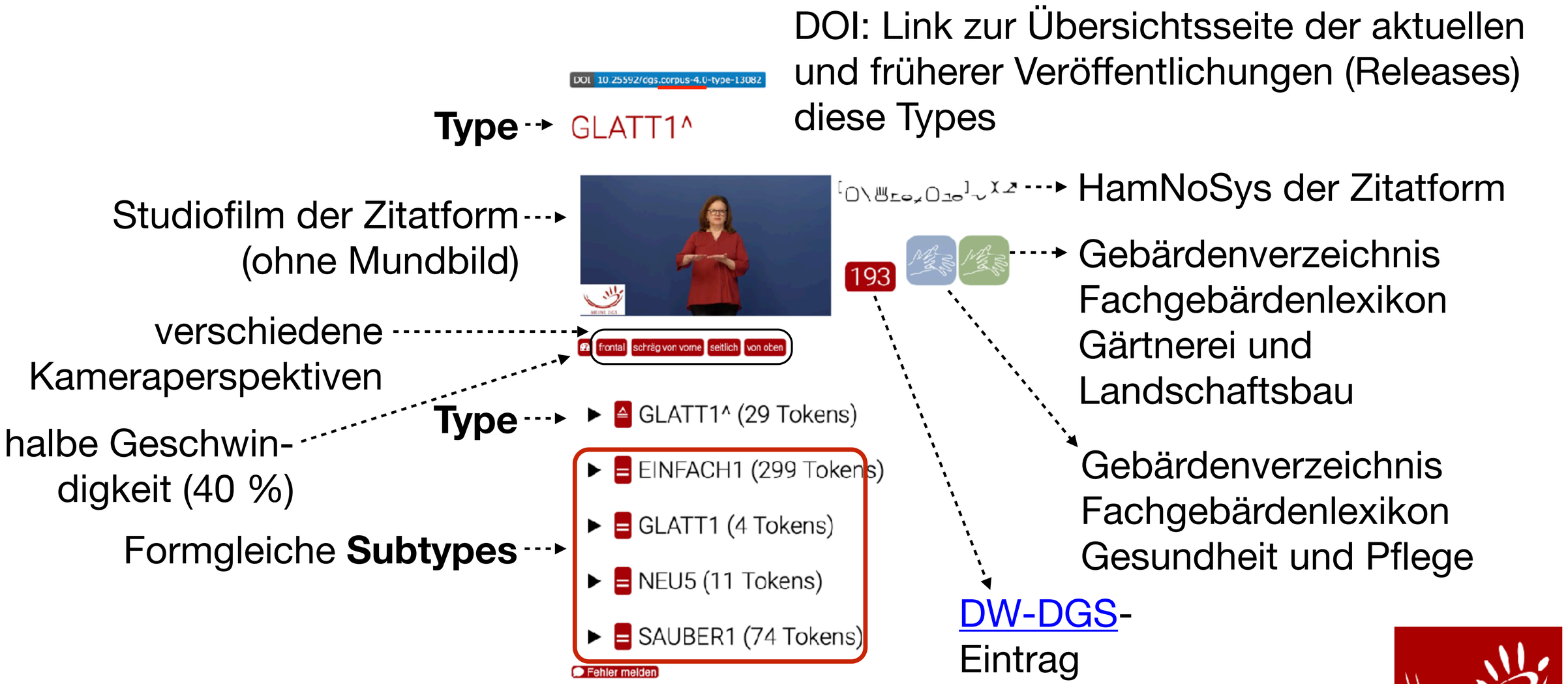
Übersetzung

Link auf Zeile im Transkript

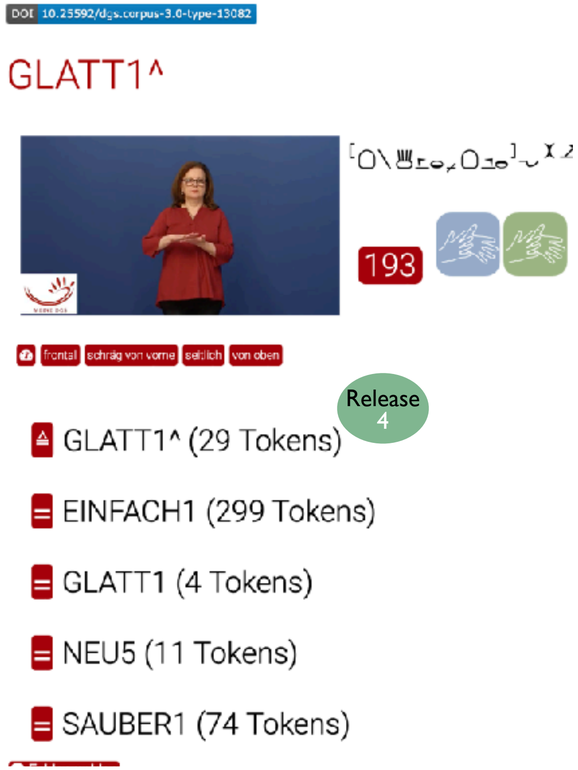
KWIC-Konkordanz
(**K**ey **W**ord **i**n **C**ontext)

Max. 3 linke und
3 rechte Nachbarn

MEINE DGS – annotiert: Types



MEINE DGS – annotiert: Types



Release 4

Download-Optionen

Types-ID	Glosse	HamNoSys	Mund	iLex	ELAN	SRT	Video frontal	Video schräg von vorne	Video seitlich	Video von oben	OpenPose
13082	GLATT1^										
8302	EINFACH1		einfach								
2541	GLATT1		glatt								
90653	NEU5		neu								
2886	SAUBER1		sauber								

MEINE DGS: Fehler melden

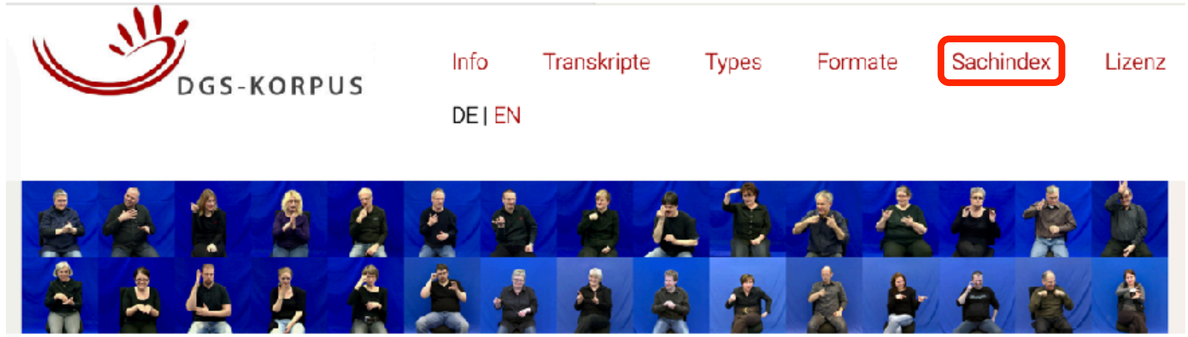
MEINE DGS:
am Fuß jeder Video-Kachel

MEINE DGS – annotiert: am Ende jedes

- Online-Transkripts
- Type-Eintrags aus der Types-Liste

„In addition, **annotating** should be seen as **open-ended** in two senses: first, it is never completed in the sense of **being immune from correction**; second, it is never completed in the sense that differing perspectives (theoretical or practical) can always be taken on the same piece of text, allowing for it **to be annotated in another way**. In short, we expect the annotations in the Auslan Corpus to be revised and augmented over time.“ (Johnston 2019: 6; Herv. R.K.)

MEINE DGS – annotiert: Sachindex



Sachindex

1963
1969
1974

1981
1986

→ Terror, Krieg, Katastrophen: Kennedy-Attentat (1963)
→ Wissenschaft und Technik: Mondlandung (1969)
→ DDR: Fußball-WM 1974 (BRD-DDR)
→ Sport und Spiel: Fußball-WM 1974 (BRD-DDR)
→ Sport und Spiel: Deaflympics 1981 in Köln
→ Energie und Umwelt: Tschernobyl (1986)
→ Terror, Krieg, Katastrophen: Tschernobyl (1986)

Im Online-Transkript:
Themen Sport und Spiel: Deaflympics 1981 in Köln

Sport und Spiel

35 Themen hier: Sport und Spiel

Bowling
Boxen
Deaflympics
Deaflympics 1981 in Köln
Deaflympics 1989 in Neuseeland

über 570 Stichwörter

Transkript-ID	Format	Paar	Alter & Geschl.	Anno
1413703	Freie Konversation	ber02	46-60m	
1248090	Freie Konversation	goe04	31-45m	
1428805	Sachgebiete	koe08	31-45m	
1429781-...	Freie Konversation	koe13	61+mw	
1582654	Freie Konversation	lei03	18-30m	
1289793	Heimatregion	mst02	18-30w	
1176566	Freie Konversation	hh01	61+m	
1583950	Erfahrungen Gehörloser	lei09	31-45w	
1585089	Freie Konversation	lei14	31-45m	
1413703	Freie Konversation	ber02	46-60m	
1414123	Erfahrungen Gehörloser	ber04	31-60m	
1414312	Gehörlosenveranstaltungen	ber04	31-60m	
1427158-...	Freie Konversation	koe01	18-30mw	
1582654	Freie Konversation	lei03	18-30m	
1290581	Freie Konversation	mst06	18-45m	
1246064	Gehörlosenveranstaltungen	fra10	61+w	
1431224	Gehörlosenveranstaltungen	koe19	31-45mw	

MEINE DGS – annotiert: Formate

DGS-KORPUS

Info | Transkripte | Types | **Formate** | Sachindex | Lizenz

DE | EN

Erhebungsformate

- Ablaufbeschreibung (31 Transkripte)
- Birnengeschichte (3 Transkripte)
- Diskussionsthemen (28 Transkripte)
- Erfahrungen Gehörloser (63 Transkripte)
- Erlebnisbericht (52 Transkripte)
- Feueralarmgeschichte (1 Transkript)
- Freie Konversation (34 Transkripte)
- Frosch und Junge (6 Transkripte)
- Gehörlosenveranstaltungen (30 Transkripte)
- Heimatregion (13 Transkripte)
- Junge vs. alte Gebärden (1 Transkript)
- Katze und Vogel (21 Transkripte)
- Lustige Geschichte (3 Transkripte)
- Realer Film (Signs) (1 Transkript)
- Reisegeschichte (7 Transkripte)
- Sachgebiete (26 Transkripte)
- Schilder (16 Transkripte)
- Stadtspaziergang (1 Transkript)
- Termine (1 Transkript)
- Vater und Sohn (2 Transkripte)
- Witz (88 Transkripte)

Datenerhebungsformat „Sachgebiete“

Schau dir die Bilder an und erzähle deinem Gesprächspartner alles, was dir dazu einfällt, oder unterhalte dich mit ihm über dieses Thema.

- Stimuli: verschiedene Fotos und Bilder zu einem Thema, um die Assoziationen der Informanten anzuregen
- Ziel: Gebärden zu 25 Sachthemen, um den Grundwortschatz der DGS abzudecken; Gespräch
- Verwendung in anderen Korpora: Verweise im Sign Language Dataset Compendium

Transkripte zum Erhebungsformat „Sachgebiete“

1414503	ber05	61+w	:
1419265	ber09	18-30w	:

MEINE DGS – annotiert: Formate

The Sign Language Dataset Compendium

Start | About | Corpora | Lexical Resources | Tasks | Languages | Credit

Data Collection Task

Subject areas

Different subject areas are presented to the signers with pictures, written subject titles or in signed instructions. Commonly used subject areas are family, home, work, hobbies, childhood, education and travel. The signers are asked to describe the topics or their personal family, work, home, etc. This task aims at collecting a solid basis of basic vocabulary additionally the Deaf culture is documented. Some projects collect discussions on this task afterwards and provide them bundled together across different tasks.

Stimulus	Written phrases, pictures and/or signed instruction
Target	Basic vocabulary, documentation of Deaf culture
Degree of Interaction	Low – middle
Duration	5–15 min
Source	See for example Nishio et al. (2010)

Task uses in corpora

Corpus	DGS Corpus
Corpus Language	German Sign Language
# recordings – open access	26 (25 subject areas)
# recordings – restricted access	349
Data available	https://meine-dgs.de/formats/format13_en.html

The Sign Language Dataset Compendium

Start | About | Corpora | Lexical Resources | Tasks | Languages | Credit

Corpus

DGS Corpus

The DGS Corpus is a collection of German Sign Language data from 330 signers from Germany. The 15-year long-term
the Institute of German Sign Language and Communication of the Deaf at the Universität Hamburg and started in 200
Thomas Hanke and Annika Herrmann. The DGS Corpus is used to build the DGS-German dictionary DW-DGS.

The signers were recorded in pairs in a mobile studio travelling to 13 spots in Germany. The signers were sitting opposi
front of a blue background. In total seven cameras were used for the recordings, five HD cameras and two Bumblebees
were later replaced by three HD stereo cameras. The cameras were set up in three different angles: one recording a to
the moderator, one filming the signers from the front and one from above. The original resolution is 1080i50 for the vid
720p50 for the videos from 2011 onwards. Public data is provided in 360p50. A Deaf moderator was leading through th

The DGS Corpus is available in different formats:

MY DGS is a community portal which offers an easy access to the data tailored for users interested in the content of th
Videos can be watched in an online viewer with subtitles.

MY DGS – annotated is a research portal which offers the annotated corpus data for linguistic research.

MY DGS – ANNIS is another research portal making the DGS Corpus available via the corpus tool ANNIS, a web browse
visualization architecture for complex multilayer linguistic corpora.

Language	German Sign Language
Size	560 hours recorded, 657000 tokens annotated
Participants	330 participants 4 age groups: 18–30, 31–45, 46–60, 61 years and older 165 female, 165 male From all over Germany
Metadata Format	CMDI
Translation	German and English, 375.8 hours (German), 113 hours (English)
Annotation	See Konrad et al. (2022) 90.9 hours annotated
Data Format	iLex
Licence	DGS Corpus License
Access	Public access via browsable homepage Open access to 50 hours of video, annotation and translation in iLex, ELAN and SRT format Restricted access for researchers to further data requires individual license agreement
Webpages	Project page: https://dgs-korpus.de Dataset: https://ling.meine-dgs.de Public access: https://meine-dgs.de ANNIS: https://annis.meine-dgs.de
Institution	Universität Hamburg
Publications	https://dgs-korpus.de/publications.html

Cite as

Konrad, R., Hanke, T., Langer, G., Blanck, D., Bleicken, J., Hofmann, I., Jeziorski, O., König, L., König, S., Nishio, R., Regen, A., Salden, U., Wagner, S., Worseck, S., Böse, O., Jahn, E., Schulder, M. 2020. *MEINE DGS – annotiert. Öffentliches Korpus der Deutschen*

Common tasks used in this corpus

Task	Calendar
# recordings – open access	1
# recordings – restricted access	167
Data available	https://meine-dgs.de/formats/format16_en.html

Task	Deaf life experiences
# recordings – open access	63
# recordings – restricted access	259
Data available	https://meine-dgs.de/formats/format3_en.html

Task	Describe process
# recordings – open access	13
# recordings – restricted access	153
Data available	https://meine-dgs.de/formats/format4_en.html

Task	Diachronic changes
# recordings – open access	1
# recordings – restricted access	157
Data available	https://meine-dgs.de/formats/format19_en.html

Task	Fire Alarm Story
# recordings – open access	1
# recordings – restricted access	67
Data available	https://meine-dgs.de/formats/format7_en.html

Task	Free conversation
# recordings – open access	34
# recordings – restricted access	131
Data available	https://meine-dgs.de/formats/format8_en.html

Task	Frog Story
# recordings – open access	1 (in 6 parts)
# recordings – restricted access	81
Data available	https://meine-dgs.de/formats/format9_en.html

Task	Jokes
# recordings – open access	66
# recordings – restricted access	49
Data available	https://meine-dgs.de/formats/format21_en.html

MEINE DGS – ANNIS

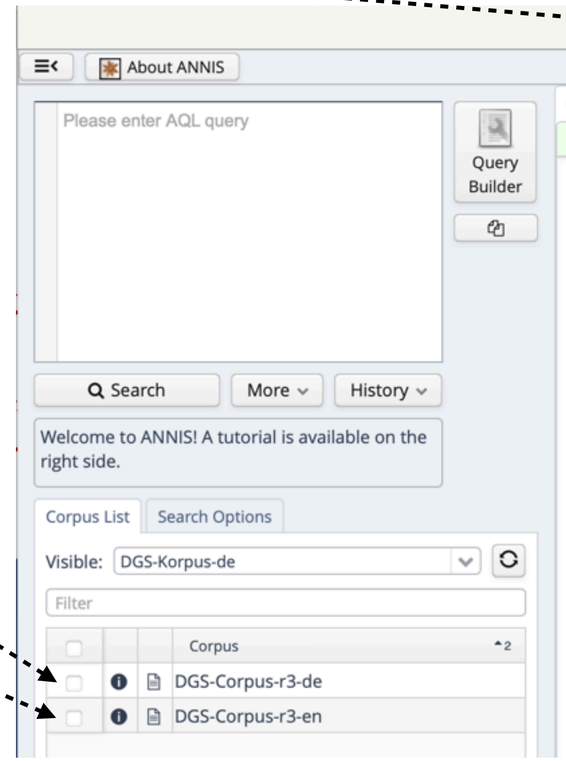
- MEINE DGS – annotiert:
  - Types-Liste ist kein Wörterbuch!
    - Trotz kombinierter Suche (Release 4): Webseite ersetzt kein Korpus-Suchwerkeug für Korpuslinguist:innen und Studierende
- Deshalb: **ANNIS** (**ANN**otation of **I**nformation **S**tructure)
  - Häufig verwendet für Korpora in vielen Formaten
  - Abfragesprache: **ANNIS Query Language** (AQL)
  - Freier Zugang über Webbrowser (ohne Anmeldung oder Download)
  - Ermöglicht komplexe Suchabfragen (inkl. Metadaten)

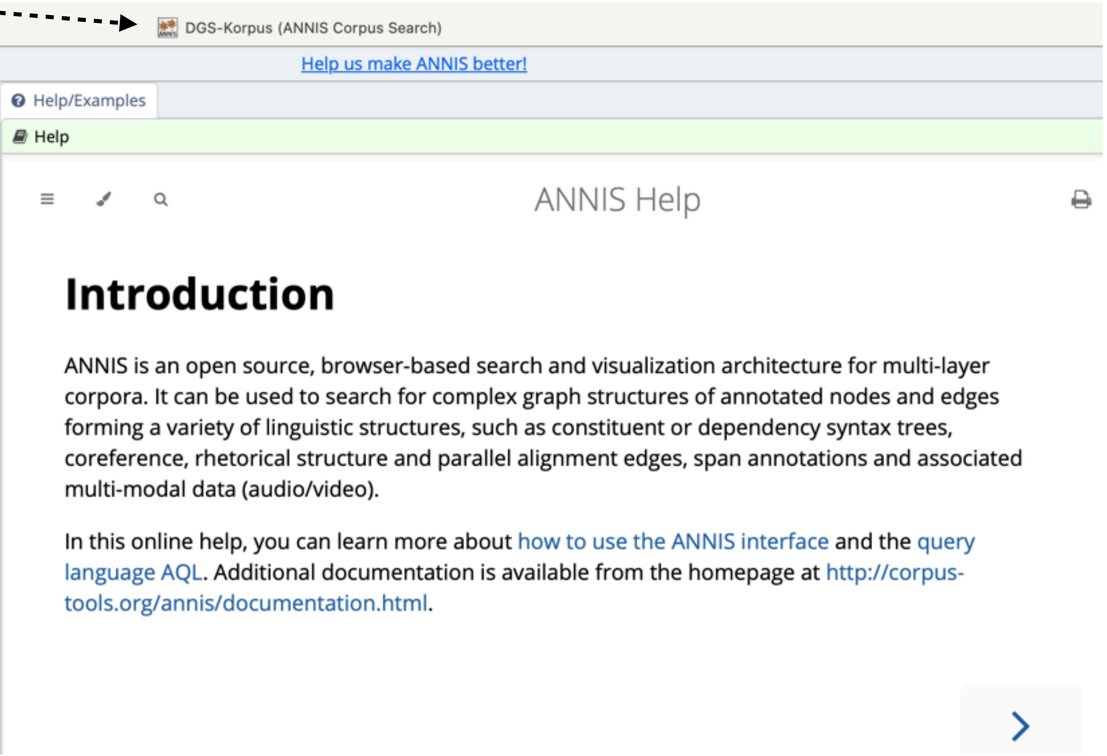
Siehe Krause/Zeldes (2016)

MEINE DGS – ANNIS

https://annis.meine-dgs.de

Annotationen auf Dt. oder Engl. (oder beides) durchsuchen





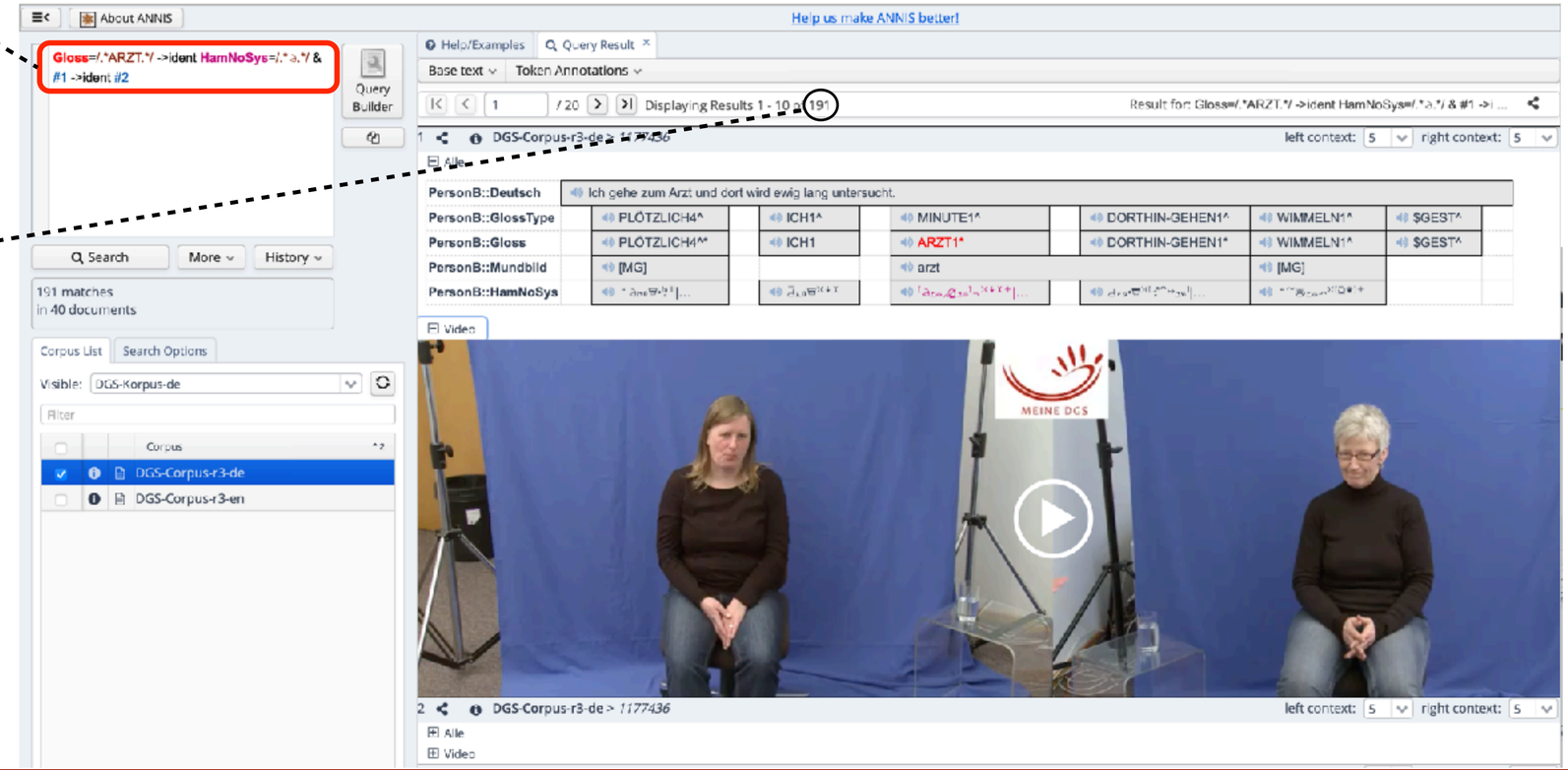
MEINE DGS – ANNIS (Release 3)

URL der Seite mit den Suchergebnissen:

https://dock.fdm.uni-hamburg.de/meinedgs/#_q=R2xvc3M9Ly4qQVJaVC4qLyAtPmlkZW50IEhhbU5vU3lzPS8uKu6AiC4qLyAmICMxIC0-aWRlbnQgIzI&ql=aql&_c=REdTLUNvcnB1cy1yMy1kZQ&cl=5&cr=5&s=0&l=10

AQL-Abfrage:
Gloss=/.*ARZT.*/ ->ident HamNoSys=/.*ɔ.*/ & #1 ->ident #2

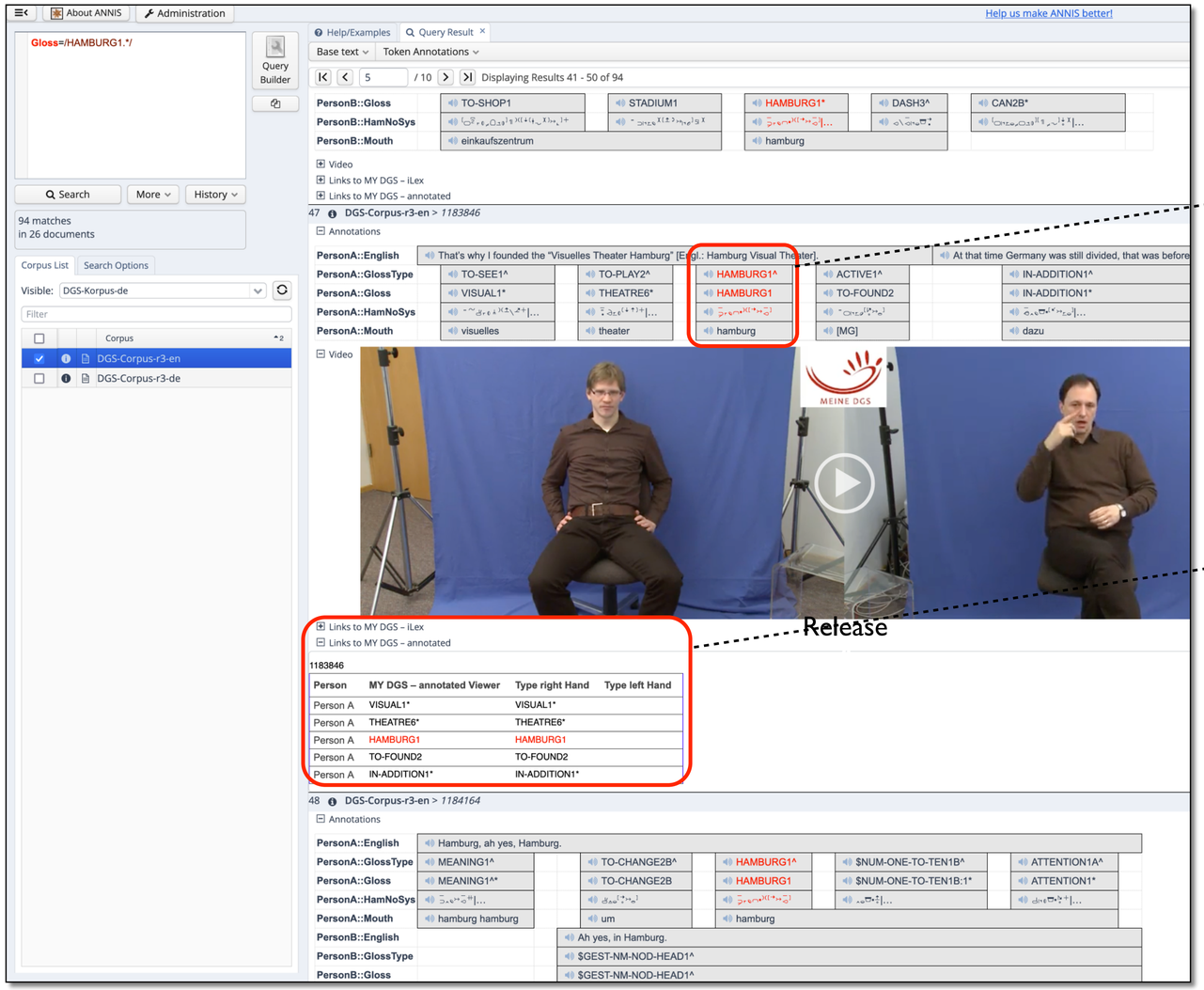
191 Treffer



ANNIS: Release 4

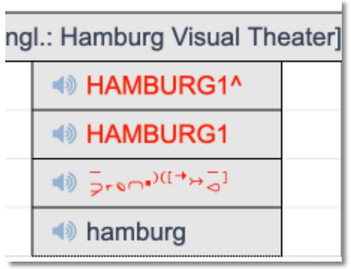
- Startseite mit Infos und Hinweisen
- AQL auch mit: ham2name=/.*hampinch12 open.*
- Links auf Types und Tokens in MEINE DGS – annotiert
- Query Wizard

MEINE DGS – ANNIS (Release 4)



Release

Treffer/Token für HAMBURG1



Link zu anderen Portalen

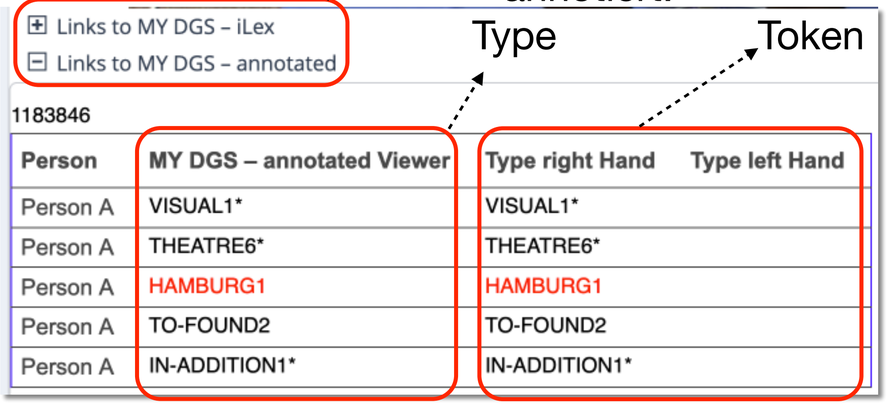
Release 4

Release 4

Direkte Links zu MEINE DGS – annotiert:

Type

Token

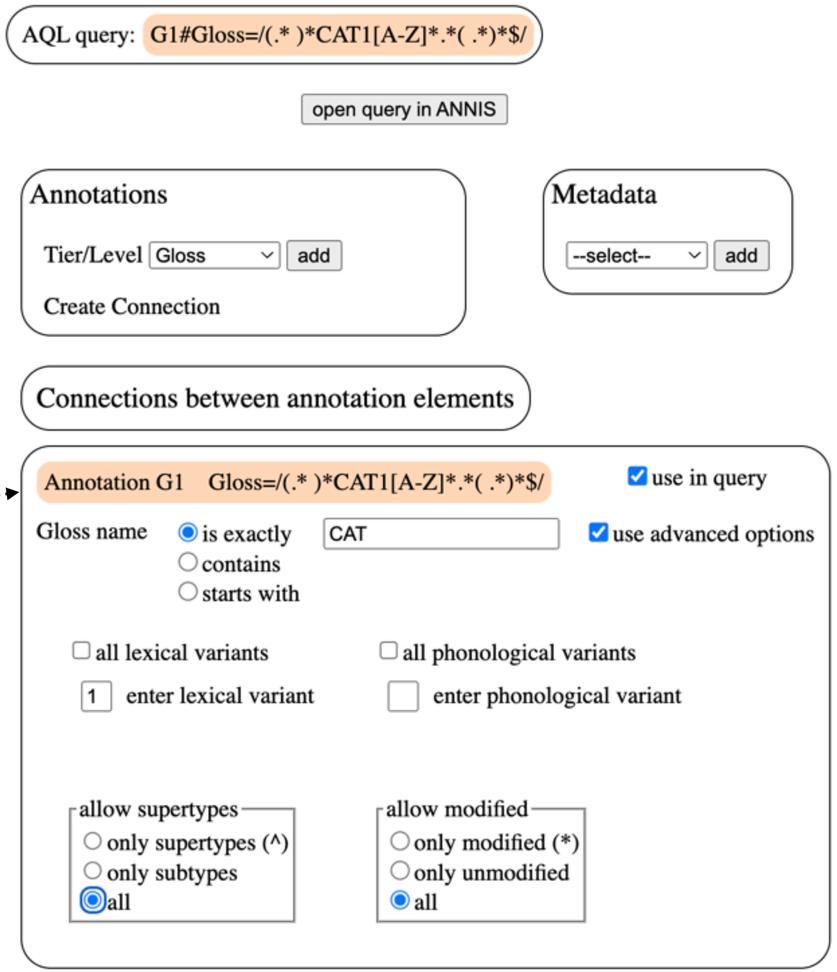


MEINE DGS – ANNIS (Release 4)

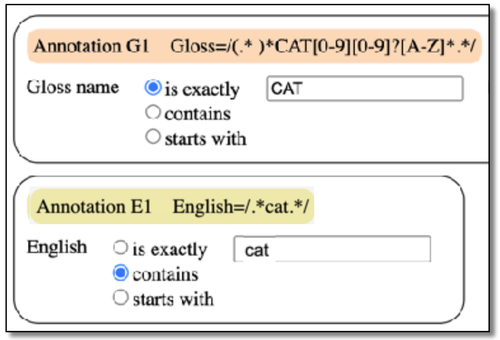
Query Wizard

Release 4

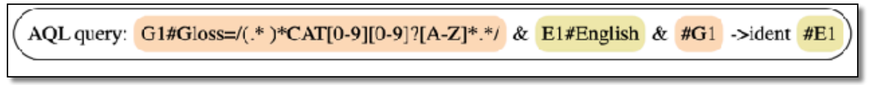
Eingabemasken als Hilfe zur Formulierung einer AQL-Abfrage



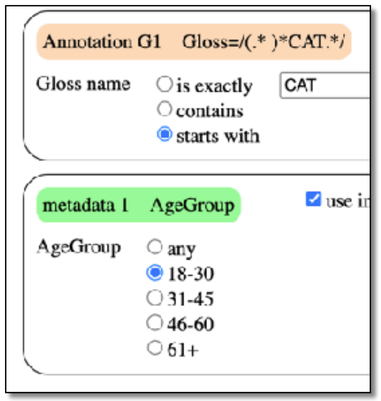
AQL-Abfrage sucht nach Tokens mit CAT1 im Glossennamen, entweder im Type oder Subtype, inkl. aller phonologischen Varianten (CAT1A/B/C …) und modifizierten Formen (Glossen mit *)



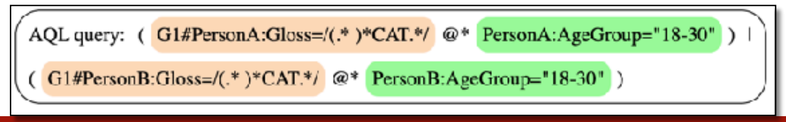
Glosse = CAT (inkl. Var.) und dazugehörige englische Übersetzung enthält cat



Kombinierte Abfrage: Glossenname enthält CAT und Teilnehmende:r ist in der Altersgruppe 18-30 Jahre

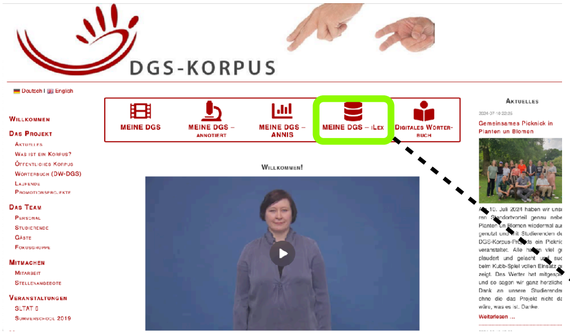


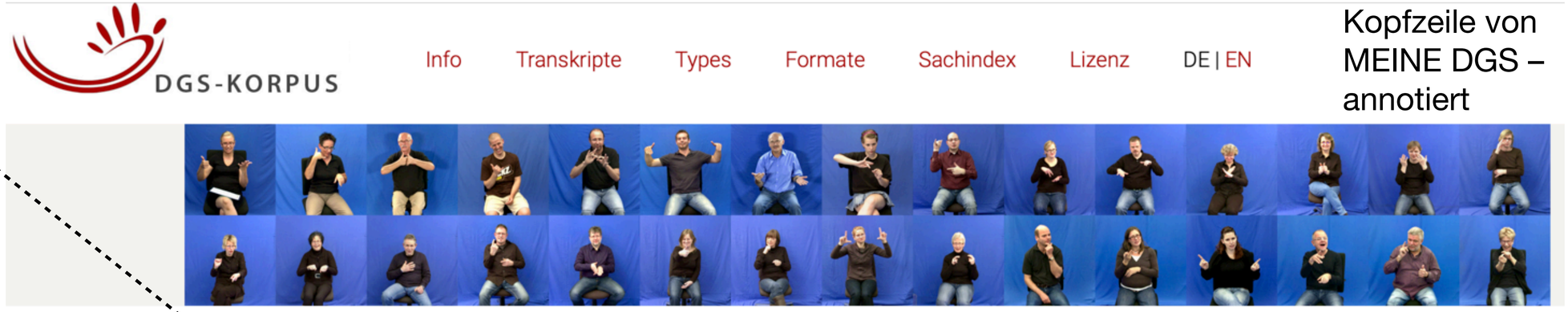
Damit beide TN berücksichtigt werden, wird die Query für PersonA und PersonB explizit formuliert



Weitere Infos s. Isard (2024)

MEINE DGS – iLex





Kopfzeile von MEINE DGS – annotiert

https://ilex.meine-dgs.de

Startseite mit Infos zum Download von:

- iLex-App
- Videos und iLex-Dateien des Öff. DGS-Korpus
- HamNoSys-Zeichensatz

iLex-Datenbank des Öffentlichen DGS-Korpus, Release 4

Wenn Sie an der Nutzung des Öffentlichen DGS-Korpus interessiert sind, haben Sie nun die Wahl zwischen vier verschiedenen Formaten:

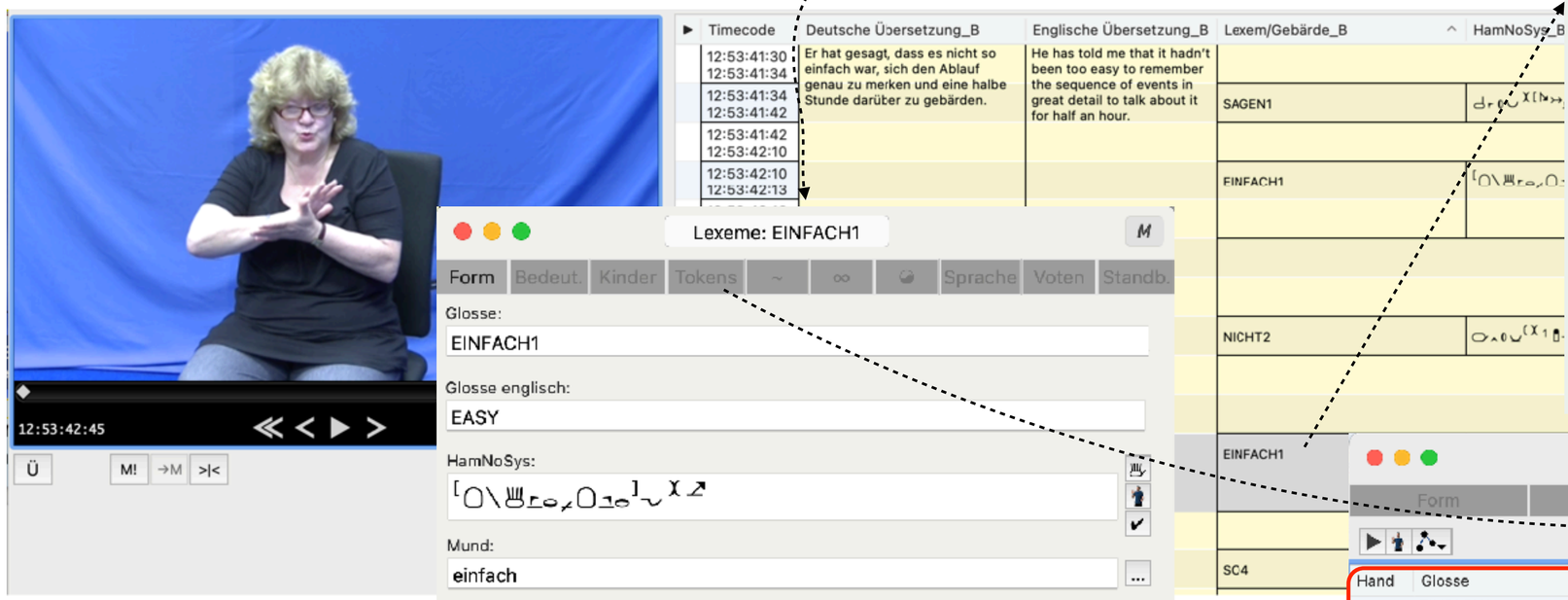
- Die einfachste Möglichkeit ohne jede Installation, ohne jeden Download, ist dieses Portal: Nutzen Sie einfach das Menü oben, um Transkripte anzuschauen, alle im Korpus verwendeten Gebärden-Types und mehr. Schauen Sie unbedingt in „Info" hinein, um mehr über die Daten zu erfahren!
- Eine weitere Möglichkeit des Zugangs ohne Download oder Installation ist die Nutzung des Korpusabfrage-Werkzeugs ANNIS mit vorinstalliertem Öffentlichen DGS-Korpus: https://annis.meine-dgs.de.
- Auf diesem Portal können Sie zu den Transkripten navigieren, an denen Sie Interesse haben, und sie entweder im ELAN-Format (.eaf) oder im iLex-Format herunterladen (ggf. auch die zugehörigen Videos), um die Daten in Ihrer lokalen ELAN- oder iLex-App zu nutzen.
- Neu mit Release 4 ist schließlich (zur Zeit nur für macOS-Nutzer:innen) die Möglichkeit, mit der iLex-App genau auf den Datenbestand des Öffentlichen DGS-Korpus zuzugreifen. Aus offensichtlichen Gründen haben Sie in iLex nur Lesezugriff auf die Daten.

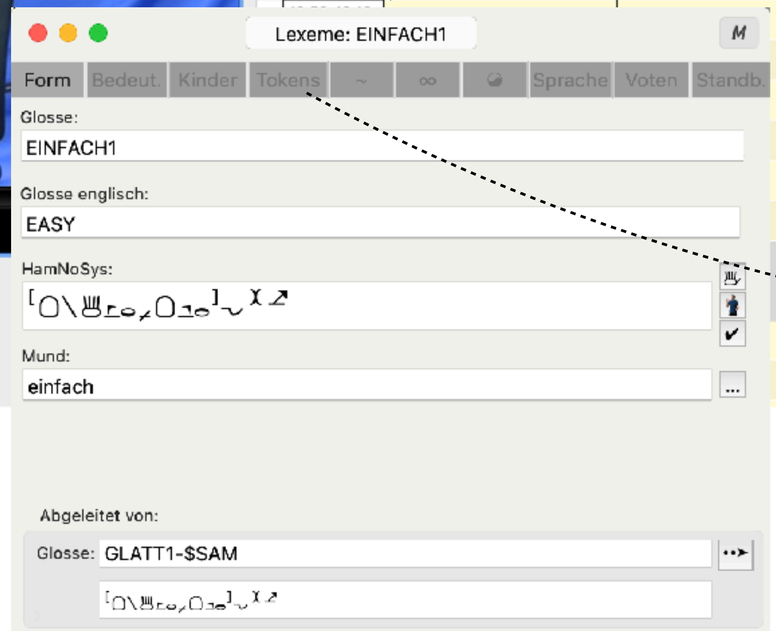
  Um diese Möglichkeit zu nutzen, gehen Sie wie folgt vor:

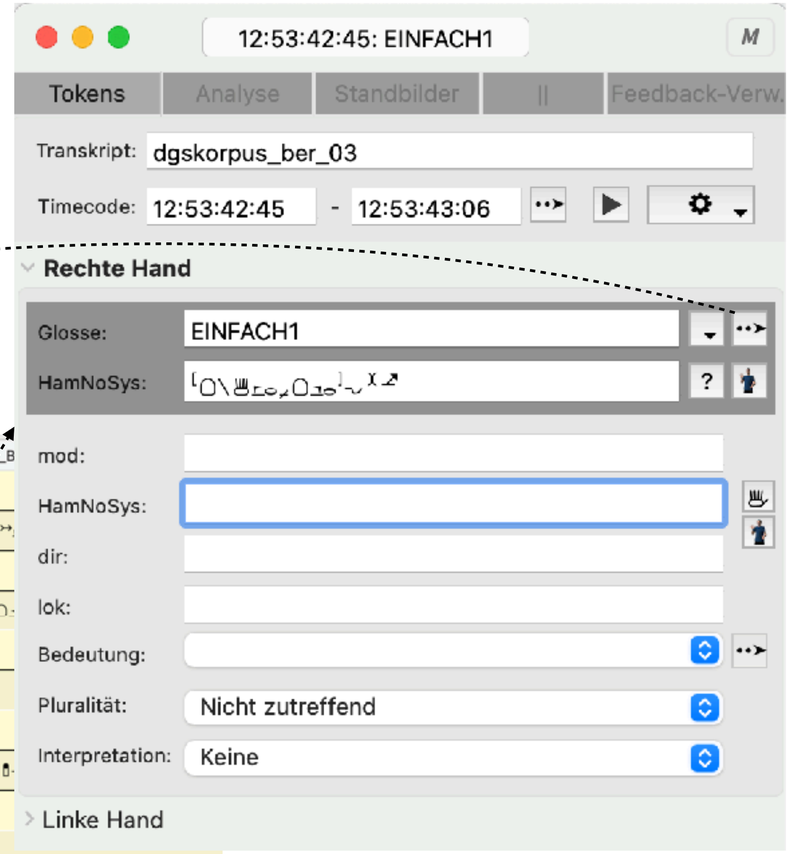
  - Sofern nicht bereits vorhanden, laden Sie die iLex-App von https://www.sign-lang.uni-hamburg.de/ilex/ilex.xml.
  - Für die Nutzung des Korpus ist es sinnvoll, auch den HamNoSys-Zeichensatz herunterladen: https://www.sign-lang.uni-hamburg.de/hamnosys.
  - Wenn Videodaten nicht lokale vorliegen, lädt iLex sie von unserem Web-Server. Sollte sich das in Ihrem Fall als zu langsam herausstellen, laden Sie die Videodateien einfach von Transkripte in den Filme-Ordner ihres macOS-Gerätes.
  - Starten Sie iLex und wählen das Öffentliche DGS-Korpus als Ihren Datenbestand. Sie müssen in diesem Fall keinen Benutzernamen oder Kennwort angeben.
  - Zum Hin- und Herspringen zwischen den Portalen MEINE DGS – annotiert, MEINE DGS – ANNIS und MEINE DGS – iLex klicken Sie auf die entsprechenden Querverweise. Wenn Sie in iLex hineinspringen wollen, fragt Ihr Browser ggf. um Erlaubnis, die iLex-App zu öffnen: Bestätigen Sie dies einfach.

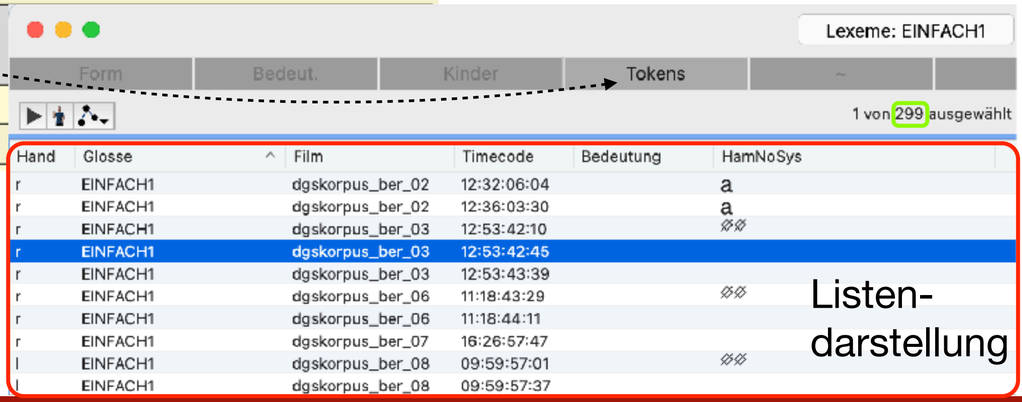
MEINE DGS – iLex

MEINE DGS – annotiert







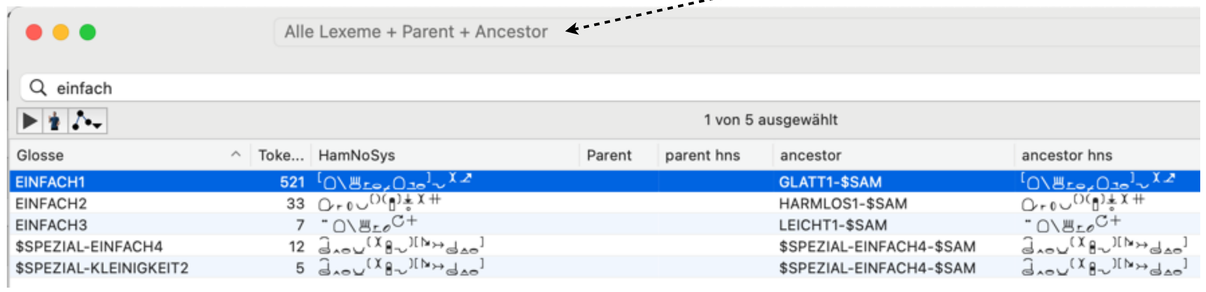


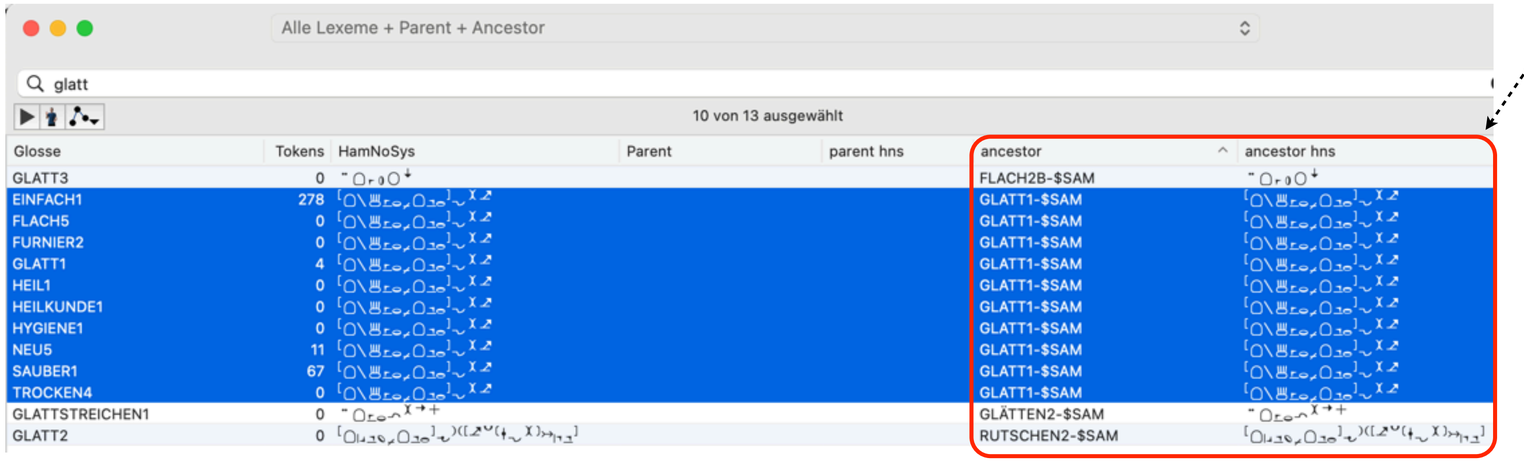
Listen-darstellung

MEINE DGS – iLex

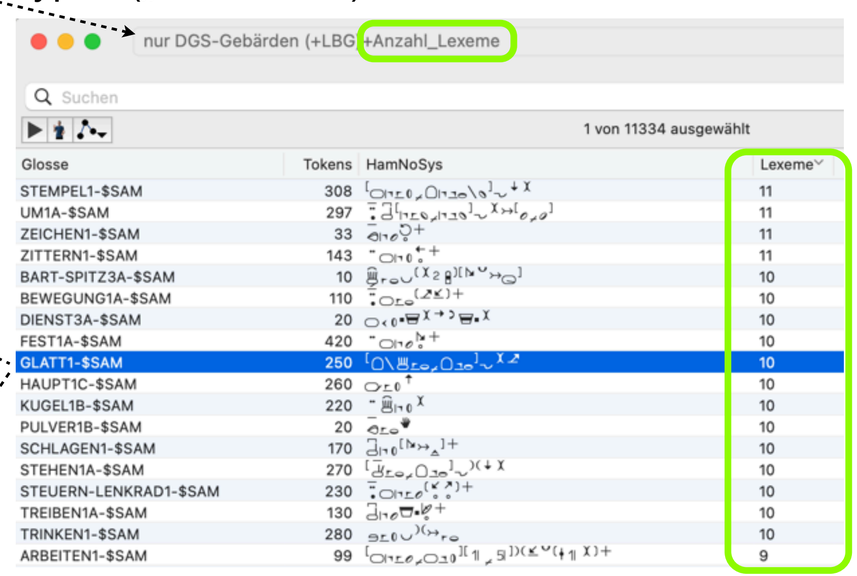
Filter (für Darstellung von Einträgen in Tabellen)

Subtypes („Lexeme“)





Types („Gebärden“)



MEINE DGS – iLex

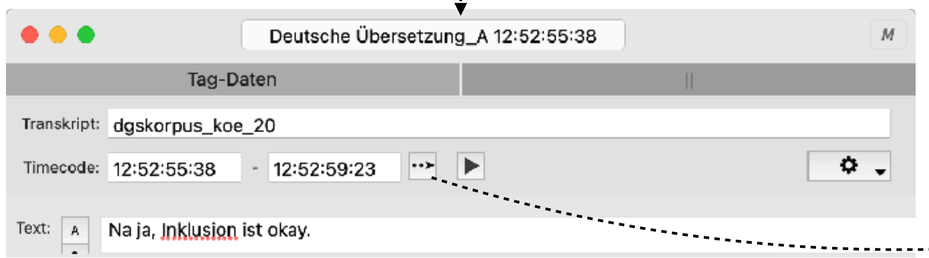
SQL-Abfrage: Suche in Übersetzungs-Tags von TN A oder B nach 'inklu' (aber nicht: 'alles inklusiv', 'inklusive alle')

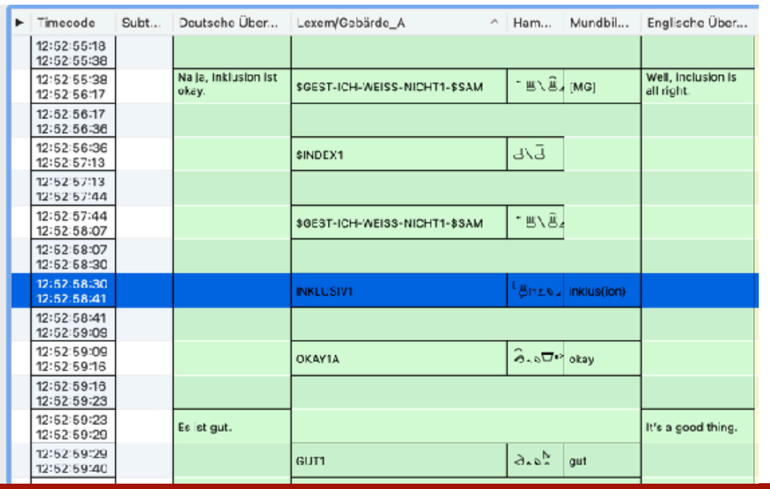
Abfrage 6

```
select * from (with pub as (select transcripts.id as trans,transcripts.name,pub.tagid,st.value as subtask,pub.timecode_start,pub.timecode_end from public_subcorpus.ling as pub join tags st on st.id=pub.tagid join tiers on st.tier=tiers.id join transcripts on tiers.transcript=transcripts.id)
select distinct pub.name as Transkript,pub.subtask,ttueb.timecode_start,ttueb.value as dt_Uebersetzung,tag_to_glossstring(ttueb.id) as glosses,ttueb.id as tagsid from pub join tiers on tiers.transcript=pub.trans and tiers.name~'^Deutsche .bersetzung_[AB]' join tags ttueb on
ttueb.tier=tiers.id and pub.timecode_start<=ttueb.timecode_start and pub.timecode_end>=ttueb.timecode_end where ttueb.value~*'inklu' and ttueb.value!~*'alles inklusiv|inklusiv[e]? (alle|F)?') as q order by 1,2,3
```

1 von 12 ausgewählt

transkript	subtask	timecode_start	dt_uebersetzung	glosses	tagsid
dgskorpus_ber_10	Diskussion	11:18:49:34	Inklusion geht nicht ohne Gebärdensprache, es sollte gebärdet werden.	INKLUSIV2.MUSS1.GEBÄRDENSPRACHE1B.DORT1.GEBÄRDEN1G.$$EXTRA-LING-MAN-$SAM	3323735
dgskorpus_hh_07	Diskussion	12:17:44:10	Jetzt ist das Wort „Inklusion" immer mehr ein Thema.	BEDEUTUNG1.JETZT1.MEHR3.AUCH3A.WORT3.INKLUSIV1.$ALPHA'1:I-N-K-L.INKLUSIV1	2974897
dgskorpus_hh_07	Diskussion	12:29:03:13	Es geht um das Thema Inklusion.	THEMA1.$SPEZIAL-GENAU-DAS1.INKLUSIV1.$ALPHA'1:I	3035656
dgskorpus_hh_07	Diskussion	12:29:04:44	Ach scheiße, ich vergesse das Wort immer. Es heißt „Inklu"/ Inklusion.	SCHEISSE1B.WORT2.IMMER4A'hd:1'phs:1.ICH1.VERGESSEN1'hd:2x.$ALPHA'1:I-N-K.INKLUSIV1	3035689
dgskorpus_koe_20	Freie Konversation ohne Moderation	12:51:47:12	Ich habe mich über den Unterschied zwischen der integrativen Klasse und der Inklusion informiert.	ICH2.WAS1A.UNTERSCHIED2.INTEGRATION1.KLASSE2.UND2B.INKLUSIV1.SCHULE1A'phs:0.$INDEX1.ERKLÄREN1.ICH1.ACH-SO1	3108034
dgskorpus_koe_20	Freie Konversation ohne Moderation	12:52:00:36	In der Inklusion wird ein Dolmetscher für die gesamte Unterrichtszeit, für bis zu acht Stunden, eingesetzt.	INKLUSIV1.BEDEUTUNG1.DOLMETSCHER1.VOLL2A.STUNDE1A'movdir:Kreis_gg_Uhr'phs:1.+STUNDE2A'q:8d	3108053
dgskorpus_koe_20	Freie Konversation ohne Moderation	12:52:04:17	Durch die UN-Konvention gibt es die Inklusion.	DURCH2B.UN1.UN-KONVENTION1.$INDEX1.INKLUSIV1	3108061
dgskorpus_koe_20	Freie Konversation ohne Moderation	12:52:18:00	Sie alle sollen nicht mehr ausgegrenzt, sondern in der Gemeinschaft inkludiert werden.	BEISEITE1B.INKLUSIV1	3108102
dgskorpus_koe_20	Freie Konversation ohne Moderation	12:52:34:44	Sie haben doch bei den Kulturtagen in Köln einen Vortrag gehalten über die Inklusion ihrer Kinder.	$INDEX1.KULTUR1A.KULTUR1A.TAG1A.IN1.KÖLN2.$INDEX1.VORTRAG1.MEIN1.KIND2.INKLUSIV1	3108903
dgskorpus_koe_20	Freie Konversation ohne Moderation	12:52:55:38	Na ja, Inklusion ist okay.	$GEST-ICH-WEISS-NICHT1-$SAM.$INDEX1.$GEST-ICH-WEISS-NICHT1-$SAM.INKLUSIV1.OKAY1A	3114030
dgskorpus_mst_01	Freie Konversation ohne Moderation	12:51:09:37	Auf der inklusiven Schule kann es natürlich sein, dass die Kinder schneller etwas lernen.	$INDEX-MONITOR1.+INTEGRATION1-$SAM.KIND2.SCHNELL3A.KOGNITION2B-$SAM.MÖGLICH1.$GEST-$SAM'h:Bezug	2395560





Literatur

- Isard, Amy. 2024. Building Your Query Step by Step: A Query Wizard for the MY DGS – ANNIS Portal of the DGS Corpus. In: Efthimiou, Eleni / Fotinea, Stavroula-Evita / Hanke, Thomas / Hochgesang, Julie A. / Mesch, Johanna / Schulder, Marc (eds): Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources. Torino, Italia: ELRA and ICCL,131–139. Online verfügbar, URL: https://aclanthology.org/2024.signlang-1.14/; letzter Zugriff: 2024-11-05.
- Johnston, Trevor. 2019: Auslan Corpus Annotation Guidelines. August 2019 version. Online verfügbar, URL: http://media.auslan.org.au/django-summernote/2020-09-04/69f180d0-04bf-4831-b59c-0b0b8f07cbb5.pdf; letzter Zugriff: 2024-11-05.
- Konrad, Reiner. 2024: DGS-Korpus und Öffentliches DGS-Korpus. Poster gezeigt auf: 7. Kulturtage der Gehörlosen 2024, Friedrichshafen, 19.-21.09.2024. [S. demnächst: DGS-Korpus-Homepage ↪ Veröffentlichungen]
- Konrad, Reiner / Hanke, Thomas / Isard, Amy / Schulder, Marc / König, Lutz / Bleicken, Julian / Böse, Oliver. 2024: Corpus à la carte – Improving Access to the Public DGS Corpus. In: Efthimiou, Eleni / Fotinea, Stavroula-Evita / Hanke, Thomas / Hochgesang, Julie A. / Mesch, Johanna / Schulder, Marc (eds): Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources. Torino, Italia: ELRA and ICCL,184–193. Online verfügbar, URL: https://aclanthology.org/2024.signlang-1.20/; letzter Zugriff: 2024-11-05.

Literatur

- Konrad, Reiner / Bleicken, Julian / Khan, Calvin / Isard, Amy / Langer, Gabriele / Müller, Anke / Schulder, Marc. 2022: Deutsche Gebärdensprache zugänglich und nutzbar machen. In: Kupietz, Marc / Schmidt, Thomas (Hrsg.): Neue Entwicklungen in der Korpuslandschaft der Germanistik. Beiträge zur IDS-Methodenmesse, S. 237-255. DOI: https://doi.org/10.24053/9783823396024
- Konrad, Reiner / Isard, Amy. 2022: MEINE DGS-ANNIS Das Öffentliche DGS-Korpus durchsuchen [Vortrag gehalten am 26.10.2022 (hybrid)]. Online verfügbar, URL: https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/Vortrag_ANNIS_korr.pdf; letzter Zugriff: 2024-11-05.
- Krause, Thomas & Zeldes, Amir (2016). ANNIS3: A new architecture for generic corpus query and visualization. In Digital Scholarship in the Humanities 2016 (31).

Links

- DGS-Korpus-Homepage: https://dgs-korpus.de
  ↳ Öffentliches Korpus:

  https://dgs-korpus.de/oeffentliches-korpus.html
  - MEINE DGS: https://meine-dgs.de
  - MEINE DGS – annotiert: https://ling.meine-dgs.de
  - MEINE DGS – ANNIS: https://annis.meine-dgs.de (noch Release 3!)
  - MEINE DGS – iLex: https://ilex.meine-dgs.de

Fragen, Anregungen, Kritik bitte an:

info@dgs-korpus.de