

MEINE DGS-ANNIS

Das Öffentliche DGS-Korpus durchsuchen



Vortrag von
Reiner Konrad / Amy Isard
am 26.10.2022
Universität Hamburg, IDGS
(und online)



Übersicht

1. Warum MEINE DGS – ANNIS?
MEINE DGS – annotiert vs. Korpussearchwerkzeuge
2. MEINE DGS – ANNIS
Suchmaske
Frequenzanalyse
Was geht schon? (Beispiele)
Was geht noch nicht? (Abfrage-Wünsche)
3. Literatur
4. Links



MEINE DGS – annotiert

<https://ling.meine-dgs.de>

- gesamtes Videomaterial: **52,4 Std.** Release 4
- übersetzt und lemmatisiert:
 - 7 Kernaufgaben (47,3 Std.)
 - exempl. ausgewählte Aufgaben (1,7 Std.)
 - zusätzl. Nacherzählungen (0,1 Std.) Release 4
- nur übersetzt: Ablaufbeschreibungen (0,9 Std.) Release 4
- nicht annotiert: Witze (2,4 Std.)



MEINE DGS – annotiert

<https://ling.meine-dgs.de>

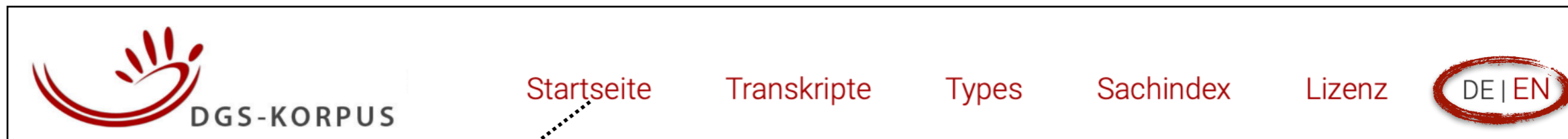
- Fragen:
 - Wer hat sich MEINE DGS und MEINE DGS – annotiert schon einmal angesehen?
 - Mit einer bestimmten Fragestellung?
 - Haben Sie gefunden, was sie gesucht haben?
 - Wie einfach/schwierig war die Suche?
 - Was haben Sie vermisst? Was hätte die Suche einfacher gemacht?
 - Wofür haben Sie die Ergebnisse benutzt?
 - Wie haben Sie die Ergebnisse weiterverarbeitet?
 - Haben Sie Fehler gefunden oder Unstimmigkeiten?
 - Wenn ja, was haben Sie gemacht?



MEINE DGS – annotiert

<https://ling.meine-dgs.de>

- Webseiten auf Deutsch und Englisch (inkl. Annotationskonventionen)



DGS-Korpus: Annotationskonventionen für das Öffentliche Korpus

Die Annotationskonventionen sind im Arbeitspapier [AP03-2018-01](#) dokumentiert.

Arbeitspapiere / Project Notes

[AP03-2018-01 Öffentliches DGS-Korpus: Annotationskonventionen / Public DGS Corpus: Annotation Conventions](#) DOI [10.25592/uhhfdm.822](https://doi.org/10.25592/uhhfdm.822)

[Current Version 4.1 \(released 2022-06-14\)](#) DOI [10.25592/uhhfdm.10251](https://doi.org/10.25592/uhhfdm.10251)

de: [DGS-Korpus_AP03-2018-01v04.1_de.pdf](#)

en: [DGS-Korpus_AP03-2018-01v04.1_en.pdf](#)

DGS-Korpus-Arbeitspapier AP03-2018-01

Öffentliches DGS-Korpus: Annotationskonventionen

Autoren: Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, Anja Regen

DOI: <https://doi.org/10.25592/uhhfdm.822>

Version 0: Dezember 2018

Version 1: Februar 2019: Abschnitt zu Komposita überarbeitet sowie Hyperlinks in das öffentliche DGS-Korpus ergänzt

Version 2: September 2019: Änderung der Zuordnung lexikalischer Formen (Subtypes) von Fingeralphabet-, initialisierten oder PMS-Gebärden zu Types

Version 3: September 2020: Anpassung an veränderte Darstellungen und neue Inhalte in Release 3 (s.a. Hanke et al. 2020) und Ergänzungen (Doppel-Token-Tags, Glossierungskonventionen, Komposita)

Version 4: Februar 2022: Differenzierung von Namensgebärden in Personennamen (\$NAME, \$NAME-...) und Organisationsnamen (\$ORG-...); Auflistung der Glossierung (Suffixe) für Tokens von Fremdsprachensprachen

Version 4.1: Juni 2022: Projektlinks werden jetzt mit https:// statt http:// angegeben, div. bibliographische Angaben aktualisiert.



MEINE DGS – annotiert

<https://ling.meine-dgs.de>

- Filme und Daten zum Download

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

Öffentliches DGS-Korpus – Daten

Die Annotationskonventionen sind im Arbeitspapier [AP03-2018-01](#) dokumentiert. Informationen zu den OpenPose-Daten finden Sie im Arbeitspapier [AP06-2019-01](#).

Release 4

Transkript	Paar	Alters- gruppe	Format	Themen
1176340	sh02	18-30	Erfahrungen Gehörloser	Arbeit und Beruf Arbeit und Beruf: Arbeit Arbeit und Beruf: Arbeitsplatzwechsel

iLex ELAN Video A Video B Video Totale SRT Video AB CMDI OpenPose



MEINE DGS – annotiert

<https://ling.meine-dgs.de>

- Online-Transkripte

The screenshot shows the 'MEINE DGS' website interface. At the top, there is a navigation menu with links: 'Startseite', 'Transkripte', 'Types', 'Sachindex', 'Lizenz', and 'DE | EN'. The logo 'DGS-KORPUS' is on the left. The main content area features a video player with two signers in a studio setting. Below the video is a transcription table with time stamps, original text, and annotated DGS labels.

00:02:53:45	Ich weiß es wirklich			
00:02:54:02	nicht mehr.			
00:02:54:02		\$GEST^		
00:02:54:13				
00:02:54:13				
00:02:54:27				
00:02:54:27				
00:02:55:05				
00:02:55:05				
00:02:55:12				
00:02:55:12				
00:02:55:20				
00:02:55:20				
00:02:55:25				
00:02:55:25				
00:02:55:36		\$GEST^	[MG]	
00:02:55:36				
00:02:55:41				
00:02:55:41				
00:02:55:49				
00:02:55:49				
00:02:56:10				
00:02:56:10				
00:02:56:20				
00:02:56:20				
00:02:56:20				
00:02:56:22				

Erinnerst du dich noch daran, wie wir damals zum Fußballspiel von Bayern gefahren sind und große Fans waren?

WISSEN2B^

\$GEST-AUFMERKSAMKEIT1^

BEISPIEL1* beispiel

FUSSBALL2 fußball

MEINE DGS – annotiert

<https://ling.meine-dgs.de>

- Online-Transkripte
→ Suchfunktion des Browsers

1 von 12 Treffern Enthält Fertig

DGS-KORPUS Startseite Transkripte Types Sachindex Lizenz DE | EN

00:02:54:13
00:02:54:13
00:02:54:27
00:02:54:27
00:02:55:05
00:02:55:05
00:02:55:12
00:02:55:12
00:02:55:20
00:02:55:20
00:02:55:25
00:02:55:25
00:02:55:36
00:02:55:36
00:02:55:41
00:02:55:41
00:02:55:49
00:02:55:49
00:02:56:10
00:02:56:10
00:02:56:20
00:02:56:20
00:02:56:22
00:02:56:22
00:02:56:43
00:02:56:43
00:02:57:03
00:02:57:03
00:02:57:12
00:02:57:12
00:02:57:21
00:02:57:21
00:02:57:22
00:02:57:22
00:02:57:30
00:02:57:30
00:02:57:38
00:02:57:38
00:02:57:40
00:02:57:40
00:02:57:47
00:02:57:47

Erinnerst du dich noch daran, wie wir damals zum Fußballspiel von Bayern gefahren sind und große Fans waren?

WASSEN2B^
\$GEST-AUFMERKSAMKEIT1^
BEISPIEL1*
FUSSBALL
VERGANGENHEIT1^*
ZUSAMMEN3A*
BAYERN|MÜNCHEN1A*

beispiel
fußball
zusammen
bayern

Was? WAS1B* was

02:55 11:58

Klick auf Glosse



MEINE DGS – annotiert

<https://ling.meine-dgs.de>

- Types-Liste mit Tokens im KWIC-Format (3 rechte und linke Nachbarn)
- Links zu DW-DGS und lexikalischen Ressourcen

BEISPIEL1

Berlin | dgskorpus_ber_01 | 18-30m Aber was für mich auch noch traurig ist, ist zum Beispiel, dass meine Eltern überhaupt nicht gebärden können.

r	ICH1	NUR2A*	TRAURIG4*	BEISPIEL1	\$INDEX1*	MEIN1	ELTERN7
l							
m		nur	traurig	beispiel		mein	eltern

Berlin | dgskorpus_ber_01 | 18-30m Meistens fahre ich weiter weg, um Freunde zu treffen, zum Beispiel nach Halle, Hannover, nach Hamburg oder Berlin.

r	ICH1	WEIT1*	TREFFEN1	BEISPIEL1	\$INDEX1*	HALLE-STADT1*	ODER4B*
l							
m	...	weit weit	treff	beispiel		halle	

Schleswig-Holstein | dgskorpus_sh_02 | 18-30f Erinnerst du dich noch daran, wie wir damals zum Fußballspiel von Bayern gefahren sind und große Fans waren?

r		WISSEN2B^	\$GEST-AUFMERKSAMKEIT1^	BEISPIEL1*	FUSSBALL2	VERGANGENHEIT1^*	ZUSAMMEN3A*
l							
m				beispiel	fußball		zusammen

Beginn der Tokens-Auflistung für Subtype BEISPIEL1, sortiert nach Transkript, innerhalb des Types BEISPIEL1^

Kopf derselben Seite (Types-Eintrag: BEISPIEL1^)

Links zum Eintrag im Gebärdenverzeichnis der Fachgebärdenlexika

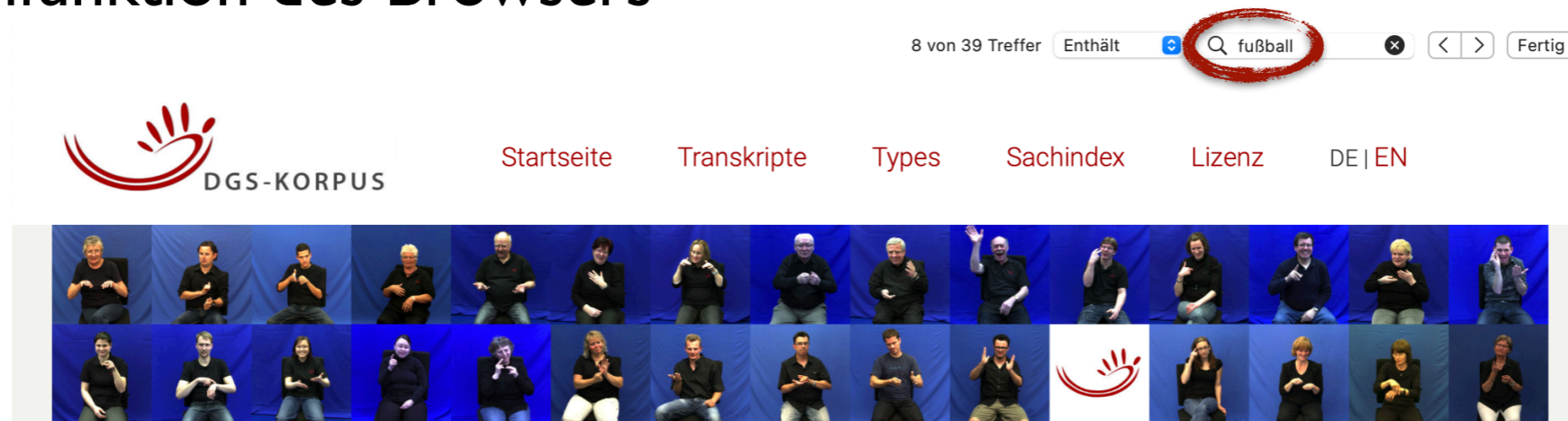
Link zum DW-DGS (Vorab-Eintrag)



MEINE DGS – annotiert

<https://ling.meine-dgs.de>

- Sachindex
35 Sachgruppen, 560 Stichwörter
→ Suchfunktion des Browsers



Sachindex

1

11. September 2001 → Politik
11. September 2001 → Terror, Krieg, Katastrophen

Fukushima (2011) → Terror, Krieg, Katastrophen

Fußball → Sport und Spiel

Fußball-WM → Sport und Spiel

Fußball-WM 1974 (BRD-DDR) → DDR

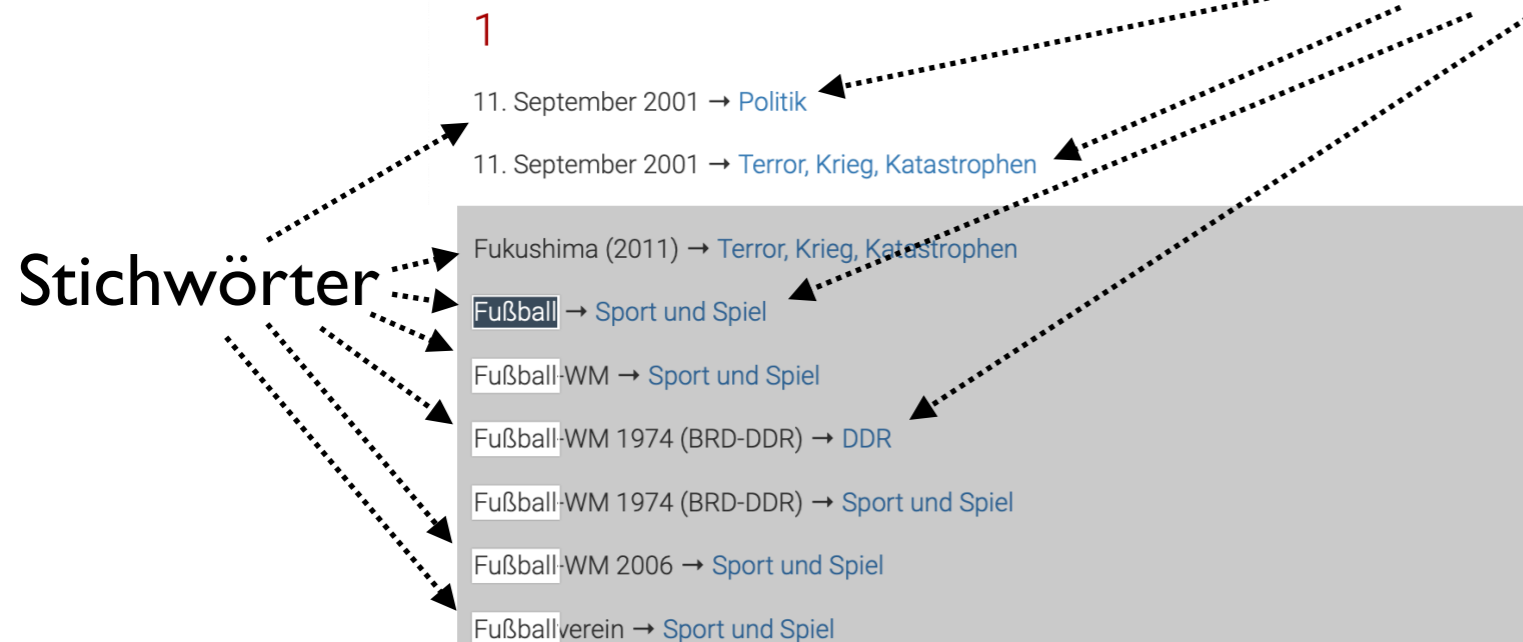
Fußball-WM 1974 (BRD-DDR) → Sport und Spiel

Fußball-WM 2006 → Sport und Spiel

Fußballverein → Sport und Spiel

Sachgruppen

Stichwörter



Warum ANNIS?

Textsuche,
Tokensuche,
kombinierte Suche
(Übersetzung, Glossen, Mundbild, Metadaten),
Token + Nachbar-Token
(in bestimmtem Abstand),
Suche nach Form (HamNoSys),
Frequenzanalyse ...



<https://debeste.de/183603/Was-hat-er,-was-ich-nicht-habe>



Fragen

- Wer kennt andere Korpora?
- Wer hat schon in einem Korpus nach Daten gesucht?
- Wer kennt Abfrage-Sprachen für Korpussuche?



Korpussuchwerkzeuge (corpus search tools)

- MEINE DGS – annotiert ist **kein** Korpussuchwerkzeug (corpus search tool)!
- Datenbank für Gesprochenes Deutsch (DGD) (https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome)
 - Volltextsuche
 - Token-Suche

↑
Registrierung erforderlich

The screenshot shows the DGD (Germanistik im Netz - COSMAS II) corpus search interface. The search parameters are: Transkribiert: z.B. 'kannst', Normalisiert: z.B. 'kannst', Lemma: z.B. 'können', POS: z.B. 'VMFIN'. The search results are displayed in a table with columns for 'Sprechereignis', 'Sprecher', and 'Treffer'. The results show various instances of the word 'kannst' in different contexts, with some instances marked as 'nie' (not found) and others as 'gefunden' (found).

Ergebnis	Sprechereignis	Sprecher	Treffer
1	FOLK_00002_01	CJ	es dauerte nie lange
2	FOLK_00002_01	CJ	in den kindergarten würde sesemi ohne marmelad nie gehen
3	FOLK_00002_01	CJ	dann seh ich marmelad nie wieder
4	FOLK_00004_01	TH	mit dem mit jemandem arbeitet der eigentlich nie was sacht und ach information ach für sich behalt in
5	FOLK_00004_01	GS	während der redet kann er ja auch nie zuhöre oder da kommt ma ja selbscht nie zu wort
6	FOLK_00004_01	GS	zuhöre oder da kommt ma ja selbscht nie zu wort oder
7	FOLK_00004_01	MS	kenn ich gar nich mehr oder no nie gehört muss ich mi jetzt mit beschäftigen
8	FOLK_00004_01	GS	schaffen se nie wieder des geht natürlich nich ja bitte
9	FOLK_00005_01	LB	grundregel mache se nie vorher s käschtsche hin
10	FOLK_00007_01	GS	grad ein hier als ergänzung comenius no nie gehört
11	FOLK_00008_01	LB	diese schaltung aber irgendwie hab ich s nie gschaft
12	FOLK_00009_01	LB	abt ihr des herr werner nie gemacht
13	FOLK_00011_01	VK	s schafft ma ja nie
14	FOLK_00011_01	SK	ach des hätt ich bei dir nie gedacht nina
15	FOLK_00011_01	VK	auf die schlossallee nie
16	FOLK_00011_01	SK	so n schwein da kommt ma fast nie hin
17	FOLK_00011_01	SK	dann musst du drauf sein un nie zahlen gar nich
18	FOLK_00012_01	VK	nein mach ma doch nie
19	FOLK_00012_01	VK	verraten ne frau sieht ma ja fast nie ne
20	FOLK_00014_01	CJ	ja noch nicht der war ja noch nie im kindergarten des is des erste mal

Korpussuchwerkzeuge (corpus search tools)

- COSMAS II

(<https://www2.ids-mannheim.de/cosmas2/>)

COSMAS II

Corpus Search, Management and Analysis System



COSMAS II - Übersicht

COSMAS II_{projekt}

COSMAS II_{web}

COSMAS II_{win}

COSMAS II_{script}

COSMAS II_{server}

Das Portal für die Korpusrecherche in den Textkorpora des *Instituts für Deutsche Sprache*



Korpussuchwerkzeuge (corpus search tools)

- COSMAS II

(<https://www2.ids-mannheim.de/cosmas2/>; Registrierung erforderlich)

COSMAS II für Einsteiger

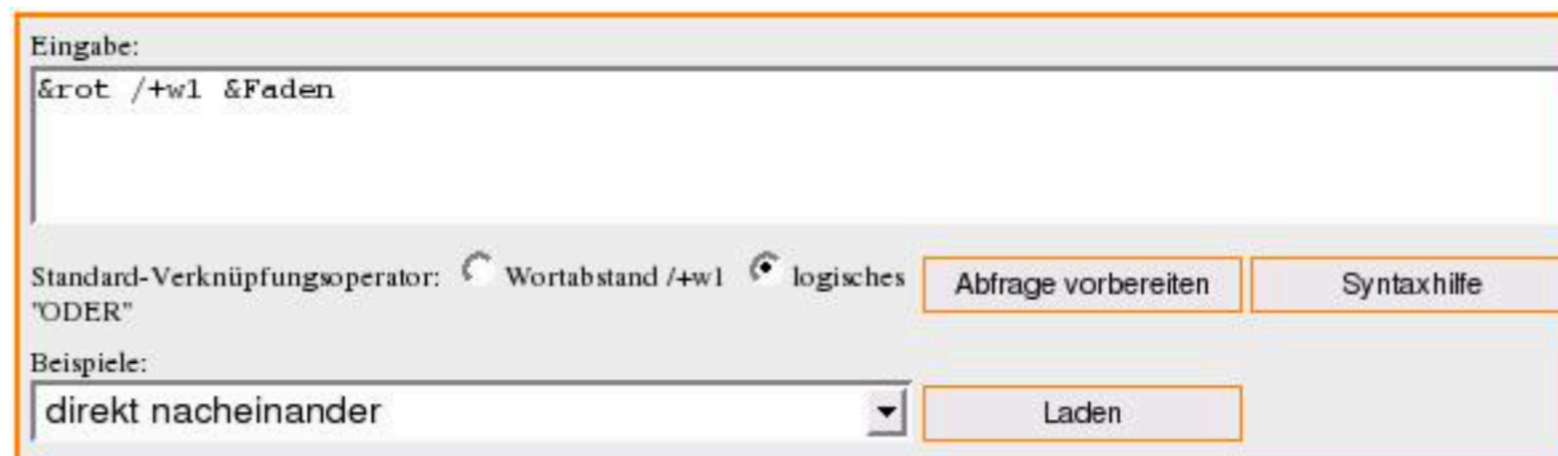
Recherchebeispiel - Suchanfrage

Die Suchanfrage lautet umschrieben:

"Suche alle Vorkommen des Ausdrucks *roter Faden* in allen Flexionsformen."

Übersetzt in die Suchanfragesprache:

"Alle Flexionsformen des Wortes *rot* gefolgt von allen Flexionsformen des Wortes *Faden* mit dem Wortabstand 1."



The screenshot shows the COSMAS II search interface. At the top, there is a text input field labeled 'Eingabe:' containing the query '&rot /+w1 &Faden'. Below the input field, there are two radio buttons for the 'Standard-Verknüpfungoperator': 'Wortabstand /+w1' (selected) and 'logisches "ODER"'. To the right of these buttons are two buttons: 'Abfrage vorbereiten' and 'Syntaxhilfe'. Below the operator selection, there is a 'Beispiele:' section with a dropdown menu showing 'direkt nacheinander' and a 'Laden' button.

Bildschirmausschnitt des Suchanfragefensters aus COSMAS II_{web}

<https://www2.ids-mannheim.de/cosmas2/projekt/einsteiger/wie.html>



Korpussuchwerkzeuge (corpus search tools)

- Suchabfragen über Eingabemaske oder direkt mithilfe einer Abfrage-Sprache (query language)
 - bei ANNIS:AQL (**ANNIS Query Language**)
Suchmaske, Korpora und Beispiele: <https://hu-berlin.de/annis>
Video-Tutorials: <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/corpus-tools/annis-tutorials/gui-tutorial>
- kombinierte Suche in mehreren Annotationszeilen (tiers)
- Kombination von Text-/Token-Suche mit Metadaten
- Token im Kontext: gezielte Suche nach Token-Umgebung:
Token + li./re. Nachbar in (un-)bestimmtem Abstand
- Verwendung von regulären Ausdrücken (regular expressions)



ANNIS (**ANN**otation of **I**nformation **S**tructure)



A web browser-based search and visualization architecture for complex multilayer linguistic corpora with diverse types of annotation.

(<https://corpus-tools.org/annis/corpora.html>)

- web browser-based → plattformunabhängig
- search tool → Werkzeug für Suchabfragen (auch search engine, z.B. Sketch Engine (<https://www.sketchengine.eu>))
- multi-layer → verschiedene Annotationsebenen, die sich überlappen können oder hierarchisch aufeinander aufbauen, z.B. syntaktische Analyse (Strukturbaum)



Öffentliche ANNIS Korpora

- Mehrere öffentliche Korpora von [Georgetown University Corpus Linguistics](#) und [Humboldt-Universität zu Berlin, Corpus Linguistics and Morphology](#) u.a.
 - BeMaTac-Korpus: gesprochene Dialoge mit MapTask-Design ([ANNIS-Link](#), [Korpus-Link](#))
 - Potsdam Commentary Corpus Sampler: eine Sammlung von Zeitungstexten, die dem Genre „Kommentar“ zuzuordnen sind ([ANNIS-Link](#), [Korpus-Link](#))
 - Georgetown University Multilayer (GUM) corpus ([ANNIS-Link](#), [Korpus-Link](#))
- Seit Juli 2022: das Öffentliche DGS-Korpus (<https://annis.meine-dgs.de>)



MEINE DGS – ANNIS

<https://annis.meine-dgs.de>

- nur lemmatisierte Videos
- übersetzt und lemmatisiert: **49,1** Std.
7 Kernaufgaben (47,3 Std.)
exempl. ausgewählte Aufgaben (1,7 Std.)
zusätzl. Nacherzählungen (0,1 Std.) Release 4
- ~~● nur übersetzt: Ablaufbeschreibungen (0,9 Std.) Release 4~~
- ~~● nicht annotiert: Witze (2,4 Std.)~~



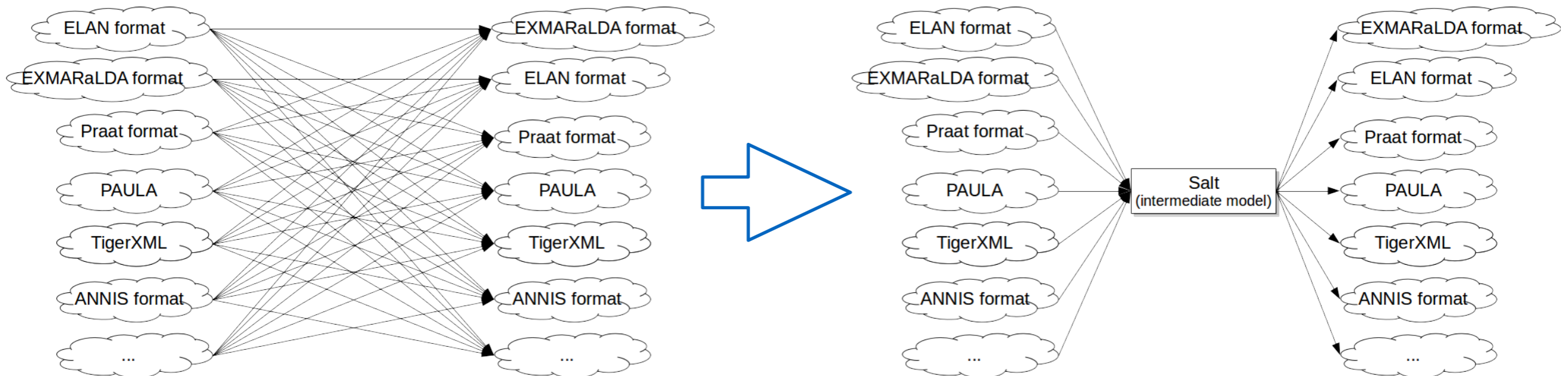
MEINE DGS – ANNIS

- Vorbereitung der Daten
- Interface
 - Suchmaske
 - Frequenzanalyse
- ANNIS Query Language AQL
 - Einführung
 - Was geht schon, mit Beispielen
 - Was geht noch nicht



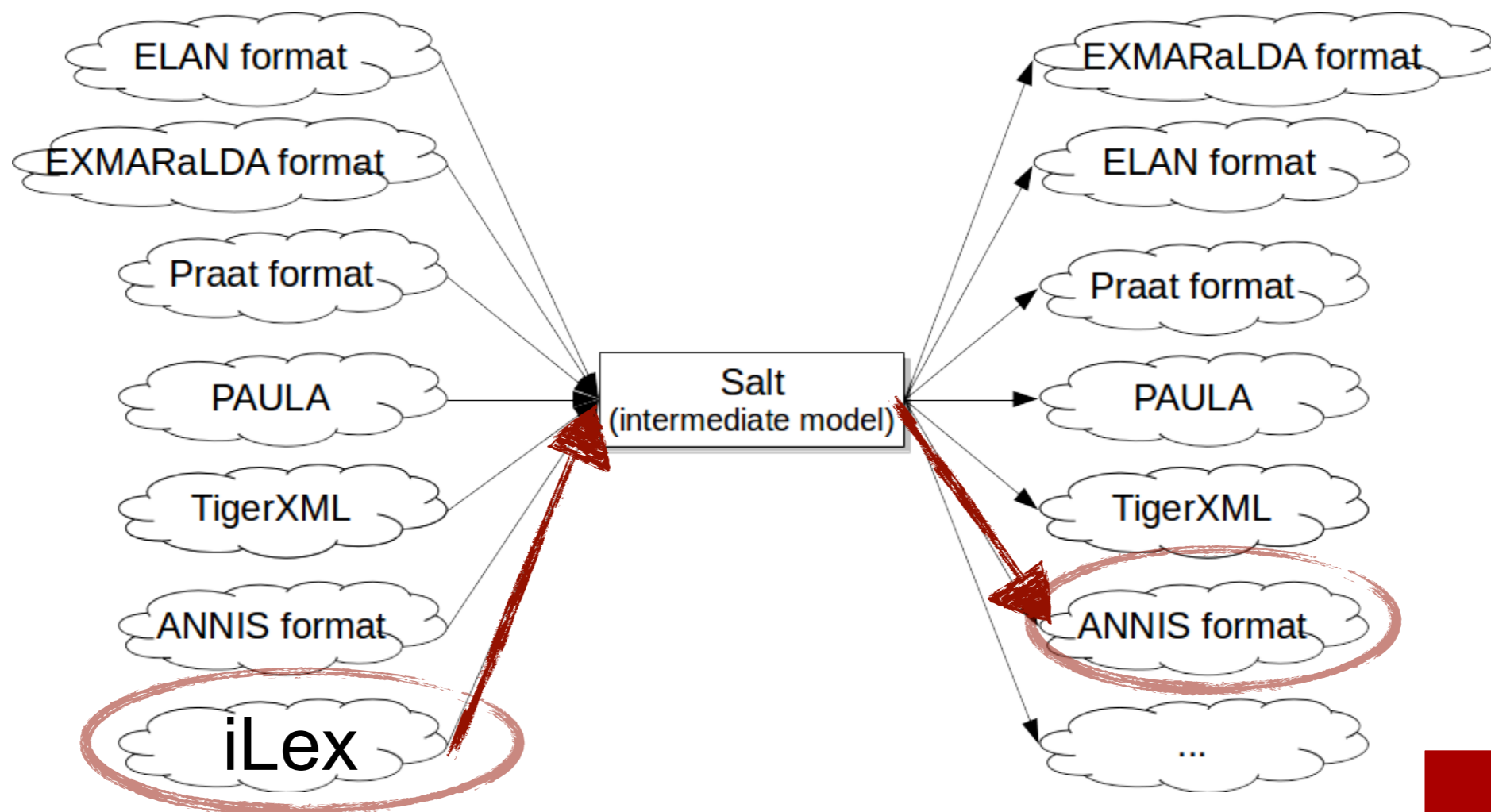
MEINE DGS – ANNIS: Vorverarbeitung der Daten

- Pepper Werkzeug <https://corpus-tools.org/pepper/>
- Pepper konvertiert Dokumente von und zu einer Vielzahl von linguistischen Formaten durch ein Zwischenmodell Salt



MEINE DGS – ANNIS: Vorverarbeitung der Daten

- Neues Peppermodul entwickelt für Öffentliches DGS-Korpus iLex Daten
- Konvertiert von iLex nach Salt und verwendet das vorhandene Modul um ANNIS-Format auszugeben



MEINE DGS – ANNIS: Interface

Eingabefeld für
Abfrage (AQL)

Ausführliche Einführung zu ANNIS
(**nicht spezifisch** zu MEINE DGS – ANNIS)

The screenshot shows the ANNIS interface. On the left, there is a search input field labeled 'Please enter AQL query' with a 'Query Builder' button below it. Below the input field are 'Search', 'More', and 'History' buttons. A welcome message states: 'Welcome to ANNIS! A tutorial is available on the right side.' Below this is a 'Corpus List' section with 'Search Options' tabs. The 'Visible' dropdown is set to 'DGS-Korpus-de'. A table lists corpora:

<input type="checkbox"/>	Corpus	↕
<input type="checkbox"/>	DGS-Corpus-r3-de	
<input type="checkbox"/>	DGS-Corpus-r3-en	

On the right, the 'ANNIS Help' page is displayed, featuring a table of contents with sections: 1. Introduction, 2. Using the ANNIS interface (sub-sections 2.1. Search Form, 2.2. Result Window, 2.3. Query Builder), and 3. ANNIS Query Language (AQL) (sub-sections 3.1. Searching for Word Forms, 3.2. Searching for Annotations, 3.3. Searching using Regular Expressions). A link for 'Example Queries' is at the bottom.

Deutsche und englische
Annotationen aufgeteilt
in zwei Korpora

Link zu vorbereitete Beispiel-Abfragen
(**spezifisch** zu MEINE DGS – ANNIS)

[Hier klicken zum Ausprobieren](#)



MEINE DGS – ANNIS: Interface

Beispiel-Abfrage wird durch Klicken in Eingabe-Feld kopiert (und sofort ausgeführt)

Gloss tokens from any participant which have the exact string KANN2B (exact string match with "TEXT"). All string and regular expression search is case sensitive, so you must use CAPITAL LETTERS to search Gloss and GlossType tokens.

The screenshot shows the ANNIS interface with a search query 'Gloss="KANN2B"' entered in the input field. Below the input field are buttons for 'Search', 'More', and 'History'. A message below the buttons says 'Valid query, click on "Search" to start searching.' On the left, there is a 'Corpus List' section with 'DGS-Korpus-de' selected. On the right, there is a 'Query Result' section with a list of example queries and their descriptions. The first query is 'Gloss="KANN2B"' with the description 'Gloss tokens from any participant which have the exact string KANN2B'. Other queries include 'Gloss=/. *KANN.*/', 'Gloss=/KANN.*/', 'PersonA:Gloss=/. *KANN.*/', 'Gloss!=/KANN.*/', 'Gloss=/KANN.*\|', 'Gloss=/MUSS.*\|', 'Gloss=/KANN.* & Deutsch!=/. *[Kk]ann.* / & #1 ->ident #2', 'Gloss=/KANN.* ->ident Deutsch!=/. *[Kk]ann.* /', 'Gloss ->ident GlossType=/KANN.* /', 'Gloss ->ident HamNoSys=/. *~.* / & Deutsch!=/. *denken.* / & #1 ->ident #3', 'Gloss=/KEIN1.* / .Gloss Gloss', 'Gloss=/KEIN1.* / ^Gloss Gloss', 'GlossType=/GLEICH.* / .GlossType,2,5 GlossType', and 'Gloss=/KANN.* / ->ident Mundbild!=/. *nicht.* /'. The 'Example Queries' section is circled in blue, and a blue arrow points from the 'Gloss="KANN2B"' query to the input field. A dotted arrow points from the text 'Gloss tokens from any participant which have the exact string KANN2B' to the first query in the list.

Deutsche Annotationen ausgewählt

Vorbereitete Beispiel-Abfragen mit Beschreibung

[Hier klicken zum Ausprobieren](#)



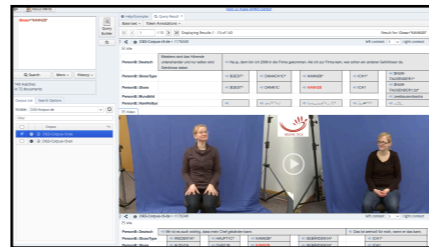
MEINE DGS – ANNIS: Interface

Übersetzung, Type, Gloss, Mundbild und HamNoSys (Type-Zitatform)

Suchergebnisse (insg. 143 Tokens mit KANN2B lemmatisiert)

The screenshot shows the ANNIS interface with the search term "KANN2B". The search results are displayed in a table format, showing the German text, the gloss, and the HamNoSys symbols. The interface includes a search bar, a corpus list, and a detailed view of search results. A red circle highlights the search bar and the "Search Options" button. A blue circle highlights the "Query Result" tab. A red circle highlights the search results table. A blue circle highlights the "Video" button. A red circle highlights the context settings (left and right context: 5).

Auf + klicken um Video zu öffnen



Voreingestellt: 5 rechte & linke Nachbar-Tokens

[Hier klicken zum Ausprobieren](#)



MEINE DGS – ANNIS: Interface

Noch nicht angepasst an
Gebärdensprachdaten!

Klicken auf Glosse, Übersetzung usw.
öffnet Video und spielt deren Abschnitt
(auf ganzen Sekunden gerundet)

Transkript-ID

Displaying Results 1 - 10 of 143

Result for: Gloss="KANN2B"

left context: 5 right context: 5

DGS-Corpus-r3-de > 1176340

Alle

PersonB::Deutsch	Meistens sind das Hörende untereinander und nur selten sind Gehörlose dabei.				
PersonB::GlossType	SGEST^	DANAC1C^	KANN2B^	ICH1^	\$NUM-TAUSENDER1^
PersonB::Gloss	SGEST^	DANN1C	KANN2B	ICH1	\$NUM-TAUSENDER1:2d*
PersonB::Mundbild					zweitausendsechs
PersonB::HamNoSys		d +v[±^]u	[O+Lp,O10]I[1~]±X	[d]±X	±X ...

Video

Play-Button spielt Video vom Beginn des Transkripts oder von der
zuletzt angewählten Stelle

Abspielgeschwindigkeit (noch) nicht variierbar.

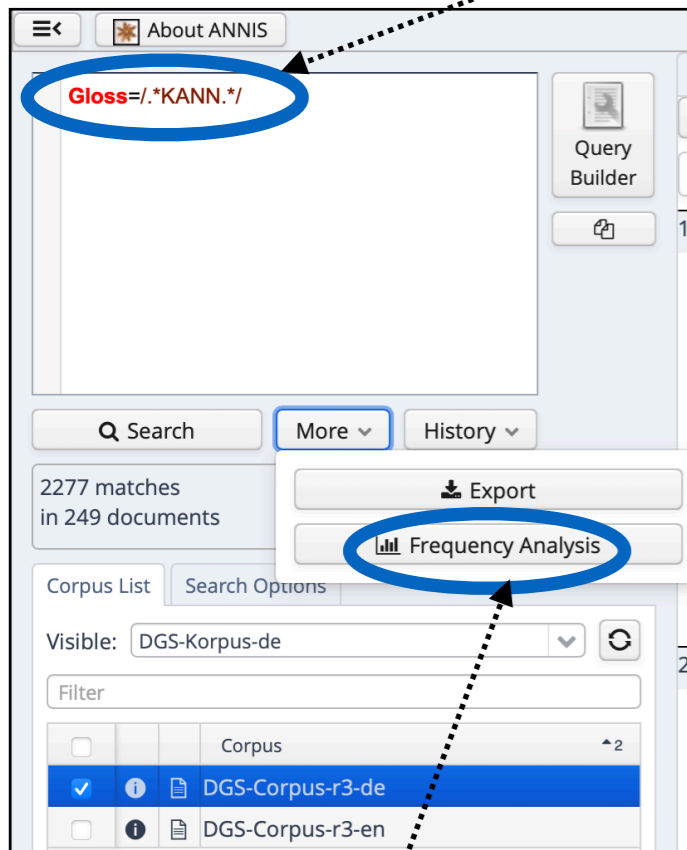
[Hier klicken zum Ausprobieren](#)



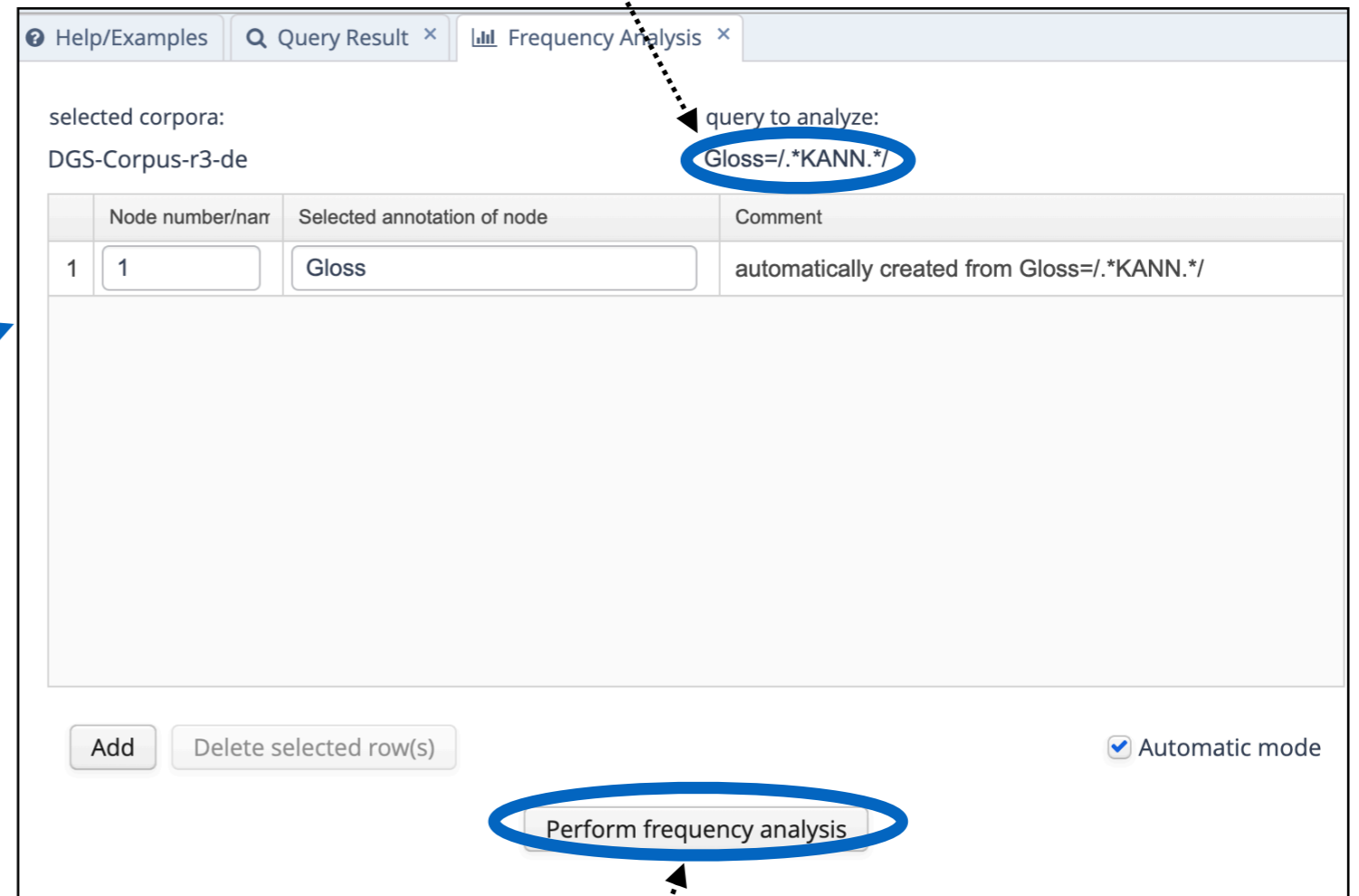
MEINE DGS – ANNIS: Frequenzanalyse

Abfrage in
Eingabe-Feld
eintragen

Hier zweite abgearbeitete Beispiel-Abfrage
Gloss=/.*KANN.* / d.H. alle Glossen die
“KANN” enthalten



Auf “More” Menu klicken
und “Frequency Analysis”
auswählen



Auf “Perform frequency analysis” klicken

[Hier klicken zum Ausprobieren](#)

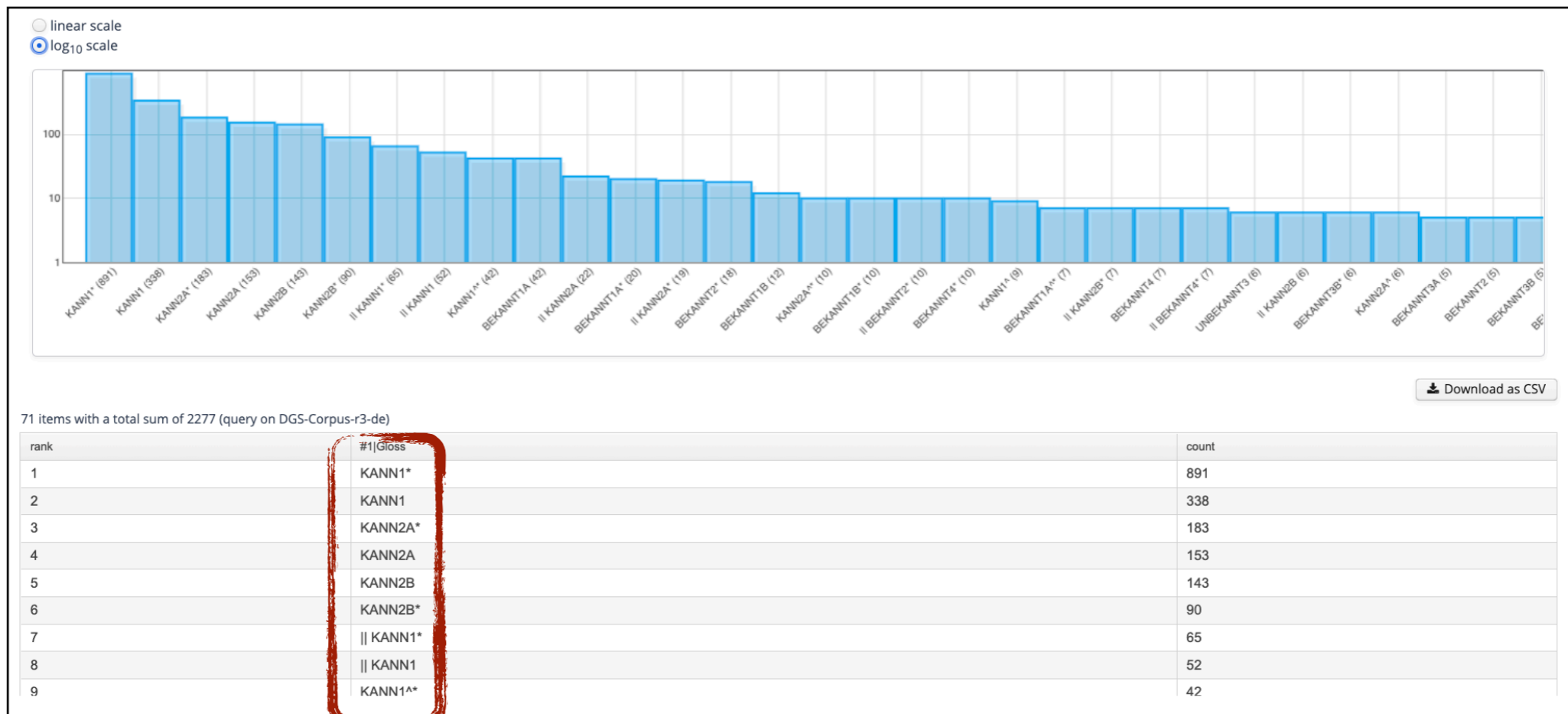


MEINE DGS – ANNIS: Frequenzanalyse

Abfrage: alle Glossen, die „KANN“ enthalten

Ergebnis sortiert nach Häufigkeit; unterschiedliche Glossen (s. Annotationskonventionen):

- Varianten: KANN1 KANN2A KANN2B
- Token mit Formabweichung: KANN1*
- produktive Verwendung einer Gebärde: KANN1^
- Token mit linkem Hand gebärdet: ||KANN...



[Hier klicken zum Ausprobieren](#)



MEINE DGS – ANNIS: Frequenzanalyse

Exportieren der Ergebnisliste, um sie z.B. in Excel weiter zu bearbeiten (z.B. Tokens von KANN1, KANN1*, || KANN1 ... zusammenführen)

Auf Spalte klicken, um nach ihr zu sortieren

Tipp: wiederholter Klick auf Spaltenkopf dreht Sortierreihenfolge um

rank	#1 Gloss	count
67	FERTIG1A* BEKANNT2	1
69	GIESSKANNE1A*	1
2	KANN1	338
20	KANN1^	9
9	KANN1^*	42
1	KANN1*	891
65	KANN1* \$GEST-OFF^*	1
33	KANN1* \$INDEX1	4
68	KANN1* \$INDEX1*	1
59	KANN1* \$PROD*	1
46	KANN1* AUFZÄHLEN1C	2
61	KANN1* AUFZÄHLEN1C*	1
51	KANN1* DU1*	1
56	KANN1 \$INDEX1	1
60	KANN1 \$INDEX2	1
4	KANN2A	153

71 items with a total sum of 2277 (query on DGS-Corpus-r3-de)

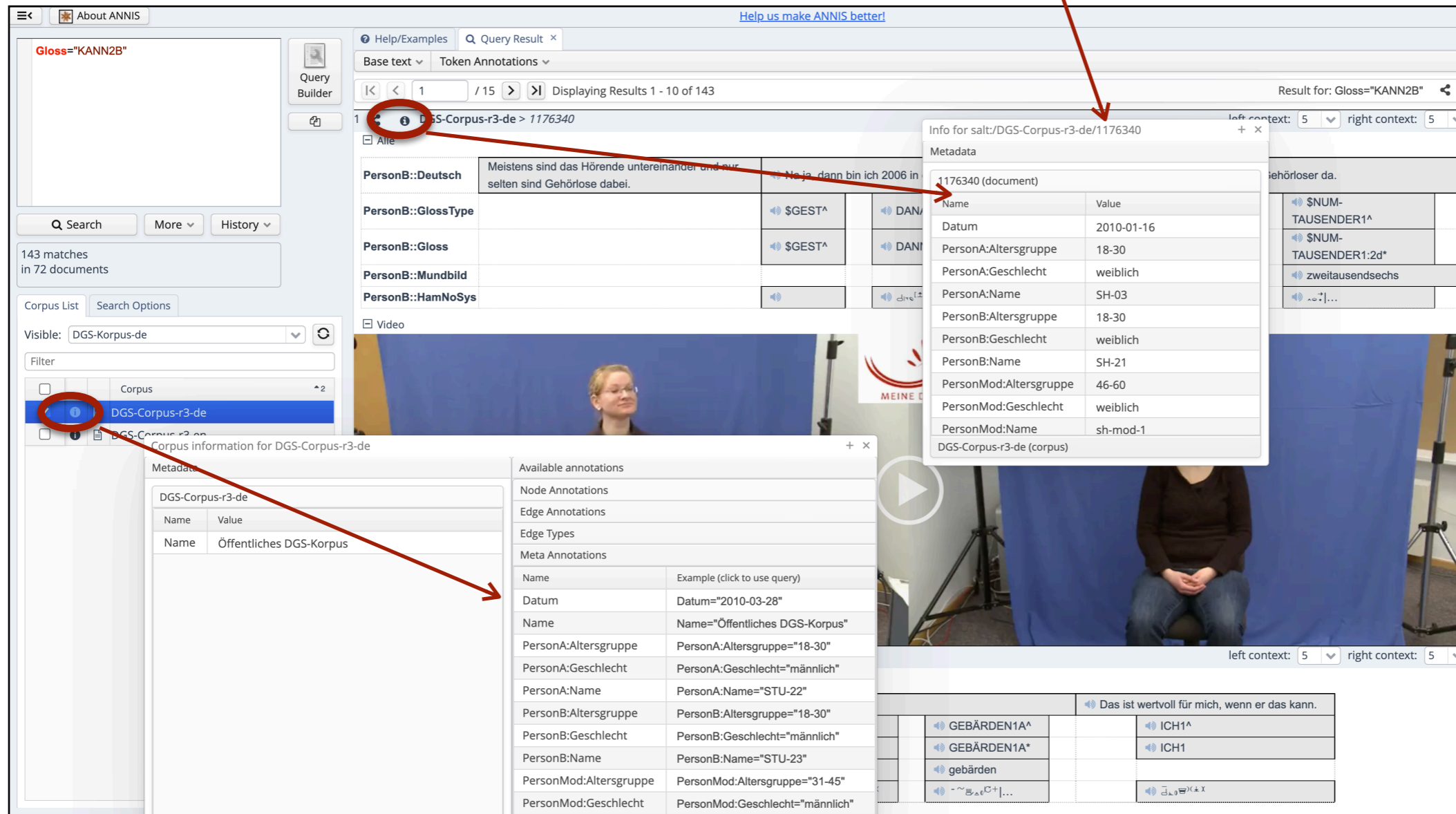
Download as CSV

[Hier klicken zum Ausprobieren](#)



MEINE DGS – ANNIS: Metadaten

Informationen über Metadaten für ein spezifisches Transkript: ID = 1176340



The screenshot shows the ANNIS interface with a search for 'Gloss="KANN2B"'. The results list 'DGS-Corpus-r3-de' with ID 1176340. A red circle highlights the information icon for this document. A red arrow points from this icon to a 'Metadata' popup window. Another red arrow points from the 'DGS-Corpus-r3-de' entry in the corpus list to a 'Corpus information for DGS-Corpus-r3-de' popup window. A third red arrow points from the 'Available annotations' table in the corpus information popup to the 'Info for salt:/DGS-Corpus-r3-de/1176340' popup window.

Info for salt:/DGS-Corpus-r3-de/1176340

Name	Value
1176340 (document)	
Datum	2010-01-16
PersonA:Altersgruppe	18-30
PersonA:Geschlecht	weiblich
PersonA:Name	SH-03
PersonB:Altersgruppe	18-30
PersonB:Geschlecht	weiblich
PersonB:Name	SH-21
PersonMod:Altersgruppe	46-60
PersonMod:Geschlecht	weiblich
PersonMod:Name	sh-mod-1

Corpus information for DGS-Corpus-r3-de

Name	Value
Öffentliches DGS-Korpus	

Available annotations

Name	Example (click to use query)
Datum	Datum="2010-03-28"
Name	Name="Öffentliches DGS-Korpus"
PersonA:Altersgruppe	PersonA:Altersgruppe="18-30"
PersonA:Geschlecht	PersonA:Geschlecht="männlich"
PersonA:Name	PersonA:Name="STU-22"
PersonB:Altersgruppe	PersonB:Altersgruppe="18-30"
PersonB:Geschlecht	PersonB:Geschlecht="männlich"
PersonB:Name	PersonB:Name="STU-23"
PersonMod:Altersgruppe	PersonMod:Altersgruppe="31-45"
PersonMod:Geschlecht	PersonMod:Geschlecht="männlich"

Informationen über Annotationen und Metadaten im ganzen Korpus; hier: Metadaten

[Hier klicken zum Ausprobieren](#)



MEINE DGS – ANNIS: Suchmaske

Suchergebnis teilen und dokumentieren

Funktionen stehen zurzeit nicht zur Verfügung, da Proxy-Server den direkten Zugriff verhindert.

Match reference link

Share your match: 1. Choose the visualization to share. 2. Copy the generated link or code. 3. Share this link with your peers or include the code in your website.

Link for publications
https://dock.fdm.uni-hamburg.de/meinedgs/?id=d980d8f7-8d40-46d9-8cf0-5d23f8655c7a#_q=R2xvc3M9IktBTk4yQiI&q1=aq1&c=REdTLUNvcnB1cylyMy1kZQ&c1=5&cr=5&s=0&l=10&m=0

Code for embedding visualization in your page

```
<iframe height="300px" width="1000px" src="https://dock.fdm.uni-hamburg.de/meinedgs/?id=d980d8f7-8d40-46d9-8cf0-5d23f8655c7a#_q=R2xvc3M9IktBTk4yQiI&q1=aq1&c=REdTLUNvcnB1cylyMy1kZQ&c1=5&cr=5&s=0&l=10&m=0"></iframe>
```

Preview

[Show in ANNIS search interface](#)

PersonB::Deutsch	Meistens sind das Hörende untereinander und nur selten sind Gehörlose dabei.	Na ja, dann bin ich 2006 in die Firma gekommen. Als ich zur Firma kam, war schon ein anderer Gehörloser da.
PersonB::GlossType		\$GEST^ DANACH1C^ KANN2B^ ICH1^ \$NUM-TAUSENDER1^
PersonB::Gloss		\$GEST^ DANN1C KANN2B ICH1 \$NUM-TAUSENDER1:2d*
PersonB::Mundbild		zweitausendsechs
PersonB::HamNoSys		dire[±^] [O1E0000I1~] X [O1E0000I1~] X [O1E0000I1~] X [O1E0000I1~] X [O1E0000I1~] X

Close

Share your query: 1. Copy the generated link
 2. Share this link with your peers.

https://dock.fdm.uni-hamburg.de/meinedgs/?id=8b79792f-5222-4e02-a2ed-0babc6295fdb#_q=R2xvc3M9IktBTk4yQiI&q1=aq1&c=REdTLUNvcnB1cylyMy1kZQ&c1=5&cr=5&s=0&l=10&m=0

1 / 15 > | Displaying Results 1 - 10 of 143

Result for: Gloss="KANN2B"

DGS-Corpus-r3-de > 1176340 left context: 5 right context: 5

PersonB::Deutsch	Meistens sind das Hörende untereinander und nur selten sind Gehörlose dabei.	Na ja, dann bin ich 2006 in die Firma gekommen. Als ich zur Firma kam, war schon ein anderer Gehörloser da.
PersonB::GlossType		\$GEST^ DANACH1C^ KANN2B^ ICH1^ \$NUM-TAUSENDER1^
PersonB::Gloss		\$GEST^ DANN1C KANN2B ICH1 \$NUM-TAUSENDER1:2d*
PersonB::Mundbild		zweitausendsechs
PersonB::HamNoSys		dire[±^] [O1E0000I1~] X [O1E0000I1~] X [O1E0000I1~] X [O1E0000I1~] X [O1E0000I1~] X



MEINE DGS – ANNIS: Suchmaske

Suchergebnis teilen und dokumentieren

Alternative:
URL der Seite kopieren und in
Dokument, E-Mail oder Textfeld
einsetzen

2022_W..._Programme DGS-Korpus-M...beiter-Wiki ▾ Transkriptionsmethodik ▾ DGS-Korpus_Wiki Meine DGS MEINE DGS-Release 4 ling DGS-Korpus-ANNIS DW-DGS IDGS LaT

DGS-Korpus (ANNIS Corpus Search)

[Help us make ANNIS better!](#)

Help/Examples Query Result ×

Base text ▾ Token Annotations ▾

1 1 / 15 > >| Displaying Results 1 - 10 of 143

1 DGS-Corpus-r3-de > 1176340

☐ Alle

PersonB::Deutsch	Meistens sind das Hörende untereinander und nur selten sind Gehörlose dabei.	Na ja, dann bin ich 2006 in die Firma gekommen. Als ich zur Firma kam, v		
PersonB::GlossType		\$GEST^	DANACH1C^	KANN2B^
PersonB::Gloss		\$GEST^	DANN1C	KANN2B
PersonB::Mundbild				
PersonB::HamNoSys			dʰo [ɪ̯ ʰ ɔ̯ ɪ]	[Oɦɪɔ̯ɔ̯ɪ] [ɪ̯ ʰ ɔ̯ ɪ]



ANNIS Query Language (AQL)

- Struktur ähnlich wie SQL (Structured Query Language)
- Suche nach Annotationsattributen und Beziehungen zwischen diesen
- Mehrere Suchoperatoren, die zu komplexen Abfragen kombiniert werden können
- Ausführliche Anleitung (auf Englisch) <http://korpling.github.io/ANNIS/4.9/user-guide/aql/>



ANNIS Query Language (AQL)

- Suchen nach:
 - Text (+/- regulärer Ausdruck)
 - Verlinkung zwischen Tags in verschiedenen Annotationszeilen (tiers)
 - Abstand zwischen Tags in derselben Zeile (tier)
 - Annotation + Metadaten



MEINE DGS – ANNIS: Annotationszeilen (tiers)

Gloss	mit Glossen lemmatisierte Tokens
GlossType	(Parent-)Type einer Glosse
HamNoSys	Type-Zitatform
Mundbild (Mouth)	Mundbild oder Mundgestik
Deutsch (English)	Übersetzung
Rechts, Links (Right, Left)	mit welcher aktiven Hand gebärdet wird



ANNIS Query Language (AQL)

- Genauer Begriff via Anführungszeichen
Gloss="KANN2B"
- Regulärer Ausdruck via Schrägstrich
Gloss=/KANN.*/
Gloss=/ICH.**/
- Negation via Ausrufezeichen
Deutsch!=/*.kann.*/*



ANNIS Query Language (AQL)

- Abstand via .
Gloss=/KANN.*/.Gloss Gloss
Gloss=/KANN.*/.Gloss,3,6 Gloss
- Kombinieren via &
(**Achtung:** kombinierte Teile sind vielleicht nicht von derselben Person)
Gloss=/KANN.*/ & Mundbild=/.*nicht.*/
- Verlinkung zwischen zwei Annotationszeilen von derselben Person via ->ident
Gloss=/KANN.*/ ->ident Mundbild=/.*nicht.*/
- Kombination mit Metadaten via @*
Gloss="KANN2B" @* Regionskennung="FRA"



MEINE DGS – ANNIS: Was geht schon?

- Suche Glosse (+/- regulärer Ausdruck)
 - Tokens mit KANN als Glossenname und Formabweichung:
[Gloss=/KANN.**/](#)
- Suche Glosse + Mundbild/Übersetzung
 - Tokens mit KANN, die mit einem Mundbild artikuliert werden, das „nicht“ enthält:
[Gloss=/KANN.* / ->ident Mundbild=/. *nicht.* /](#)
 - Tokens mit KANN innerhalb eines Übersetzungs-Tags, das weder „kann“ noch „Kann“ enthält:
[Gloss=/KANN.* / ->ident Deutsch!=/. *\[Kk\]ann.* /](#)
- Suche Glosse + Teilnehmende:r (Person A oder B)
 - Tokens, deren Glosse KANN enthält: nur von PersonA des Transkripts:
[PersonA:Gloss=/. *KANN.* /](#)



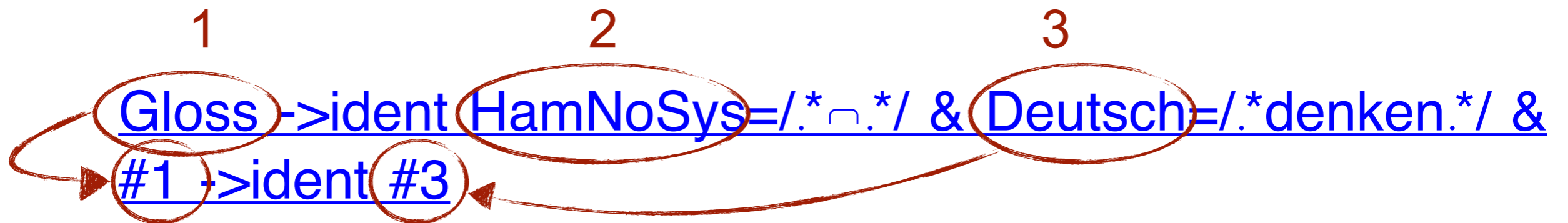
MEINE DGS – ANNIS: Was geht schon?

- Suche Glosse + Erhebungsregion/Altersgruppe (Metadaten)
- Tokens mit Glosse KANN2B von Teilnehmenden aus der Erhebungsregion FRA (Frankfurt a.M.):
[Gloss="KANN2B" @* Regionskennung="FRA"](#)
- Tokens von allen Teilnehmenden aus der Altersgruppe 31-45 Jahre:
[\(PersonA:Gloss @* PersonA:Altersgruppe="31-45"\) | \(PersonB:Gloss @* PersonB:AltersGruppe="31-45"\)](#)
- Suche Tokens mit Glossen, die dem GlossType KANN1 zugeordnet sind (Type-Hierarchie, s. Annotationskonv.):
[Gloss ->ident GlossType=/KANNI.* /](#)



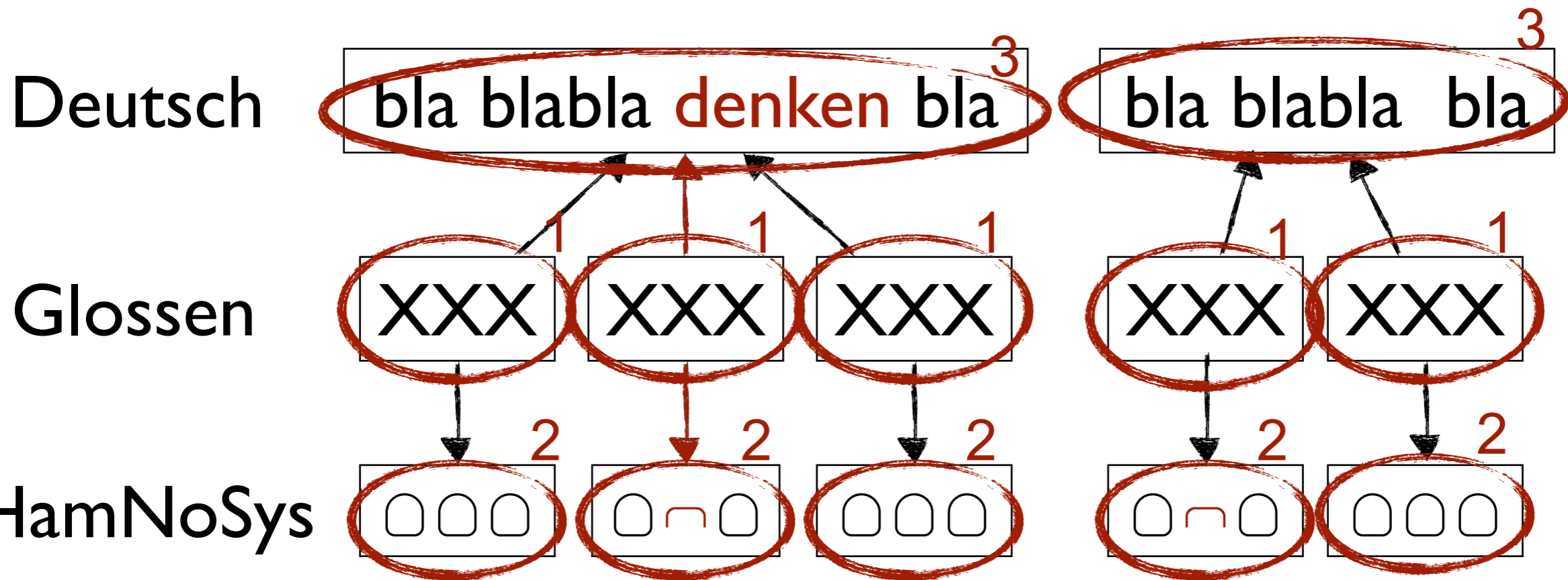
MEINE DGS – ANNIS: Was geht schon?

- Suche Glosse + HamNoSys + Übersetzung
- Tokens von Gebärden, deren Grundform an der Stirn ausgeführt werden (HamNoSys des GlossTypes enthält das Symbol für Stirn \cap) und die innerhalb eines Übersetzungs-Tags liegen, das „denken“ enthält:



MEINE DGS – ANNIS: Was geht schon?

1
2
3
1 Gloss ->ident 2 HamNoSys=/.*∩.* / & 3 Deutsch=/.*denken.* / & 1
#1 ->ident #3



Gloss ->ident HamNoSys=/.*∩.* / & Deutsch=/.*denken.* / & #1 ->ident #3



MEINE DGS – ANNIS: Was geht schon?

1

2

3

Gloss ->ident HamNoSys=/.***□***/ & Deutsch=/.***denken***/ & #1 ->ident #3

6 DGS-Corpus-r3-de > 1178364							
Alle							
PersonA::Deutsch	Meine Rückfahrt mit dem Zug dauerte zwei Stunden.		Ich musste daran denken.			Aber es war schön.	
PersonA::GlossType	ZUG2B^	\$PROD	WISSEN2B^	PASSEN1^	\$GEST^	\$GEST-NM-KOPFNICKEN1^	GUT^
PersonA::Gloss	ZUG2B*	\$PROD*	WISSEN2B*	PASSEN1*	\$GEST*	\$GEST-NM-KOPFNICKEN1*	SCHÖN^
PersonA::Mundbild	zug					aber	schön
PersonA::HamNoSys	0z\ao[+?]+...		dren[+?]+...	-d[+?]+...		(o[+?]+...	ao[+?]+...
Video							
7 DGS-Corpus-r3-de > 1178939							
Alle							
PersonB::Deutsch	Dann gibt es eben auch Hörende, bei denen es nicht klappt.		Die Gehörlosen denken, dass bei den Hörenden immer alles klappt, aber das stimmt nicht.				
PersonB::GlossType	NEIN3A^	\$GEST^	\$INDEX1^	WISSEN2A^	HÖREN21A^	ALLE1A^	KLA...
PersonB::Gloss	NICHT3A	\$GEST^	\$INDEX1	GLAUBEN2A*	HÖREND1A	ALLE1A	KLA...
PersonB::Mundbild				glaub	hö[rend}	alle	klap
PersonB::HamNoSys	dne[+?]+...		d\3	drene[+?]+...	dre[+?](k\<x+ ...)	ao[+?](k\<x+ ...)	[O[+?]+...
Video							
8 DGS-Corpus-r3-de > 1179389							
Alle							
PersonB::Deutsch	Wenn die Wirtschaft schlecht war, bekommen sie halt keine Prämie, aber das ist ihnen auch egal.		Entweder denken sie nur an den Gewinn und daran, durch die Prämien noch mehr Geld zu bekommen.				
PersonB::GlossType	ICH1^	WIE4A^	\$GEST^	WIE4A^	ERFOLG2^	SICH-DIE-HÄNDE-REIBEN1^	WISSEN2B^
PersonB::Gloss	ICH1	EGAL3	\$GEST^	WIE4A*	GEWINN1	SICH-DIE-HÄNDE-REIBEN1	DENKEN1B
PersonB::Mundbild	mir	egal		wie	gewinn	[MG]	denk
PersonB::HamNoSys	3ao[+?]+...	-d[+?]+...		-d[+?]+...	oro[+?]+...	o[+?](k\<x+ ...)	dren[+?]+...
Video							

Gloss ->ident HamNoSys=/.***□***/ & Deutsch=/.***denken***/ & #1 ->ident #3



MEINE DGS – ANNIS: Was geht schon?

- Suche Glosse + rechter Nachbar in bestimmtem Abstand
- Tokens mit Glossen, die einem GlossType mit dem Glossennamen GLEICH zugeordnet sind und das rechte Nachbar-Token, das im Abstand von 2-5 Tokens folgt:
[GlossType=/GLEICH.*/.GlossType,2,5 GlossType](#)
- Glosse + rechter Nachbar mit demselben Glossennamen im Abstand von 2-50 Tokens innerhalb desselben Übersetzungs-Tags (→ Reduplikation bzw. Sandwich-Konstruktion?):
[Gloss=/ICH.*/.Gloss,2,50 Gloss=/ICH.*/ & Deutsch & #1->ident #3 & #2->ident #3](#)



MEINE DGS – ANNIS: Was geht noch nicht?

- Suche Glosse zu Beginn/am Ende eines Übersetzungs-Tags
- z.B. \$GEST-OFF^ zu Beginn eines Übersetzungs-Tags: Alignierungsfehler?
- Problem: Übersetzungs-Tags beginnen i.d.R. früher als das erste Token / enden i.d.R. später als das letzte Token innerhalb einer Übersetzung, die in ANNIS vorhandenen Funktionen `left_aligned/right_aligned` sind daher nicht nutzbar.





- Try, search and repeat
- and give feedback!



Literatur

- Isard, Amy and Konrad, Reiner (2022). [MY DGS – ANNIS:ANNIS and the Public DGS Corpus](#). In Proceedings of the LREC 2022 10th Workshop on the Representation and Processing of Sign Languages.
- Krause, Thomas & Zeldes, Amir (2016). [ANNIS3: A new architecture for generic corpus query and visualization](#). In Digital Scholarship in the Humanities 2016 (31).
- Hanke, T., Schulder, M., Konrad, R., and Jahn, E. (2020). [Extending the Public DGS Corpus in size and depth](#). In Proceedings of the LREC 2020 9th Workshop on the Representation and Processing of Sign Languages: Marseille, France.



Links

- MEINE DGS – ANNIS <https://annis.meine-dgs.de>
- ANNIS Homepage <https://corpus-tools.org/annis>
- MEINE DGS – annotiert <https://ling.meine-dgs.de>



Kontakt

reiner.konrad@uni-hamburg.de

amy.isard@uni-hamburg.de

info@dgs-korpus.de

<https://dgs-korpus.de>

Gefördert durch



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

AKADEMIE DER
WISSENSCHAFTEN
IN HAMBURG

