

# **Building Your Query Step by Step:** A Query Wizard for the MY DGS – ANNIS Portal of the DGS Corpus



Amy Isard

Institute of German Sign Language and Communication of the Deaf and House of Computing and Data Science,

University of Hamburg, Germany

# Motivation

- ANNIS is a browser-based query search and visualisation tool.
- MY DGS ANNIS allows searches of the Public DGS Corpus.
- The complexity of the DGS Corpus annotations can quickly result in complicated queries.
- Understanding how to phrase ANNIS queries correctly can be an access barrier to users.
- The Query Wizard helps users compose complex queries from smaller building blocks.

# **MY DGS – ANNIS**

ANNIS interface of the Public DGS Corpus

- Annotation tiers:
- Gloss, GlossType, HamNoSys, Mouth, Translation
- Metadata:
- for each transcript:
- TranscriptId, Date, Region, RegionCode, Theme, Keywords
- for each participant:
- ParticipantId, AgeGroup, Gender

ut ANNIS Administration		0.1111/00/1	and the second second	tota w									
/(.* )*CAT[0-9][0-9]?[A-Z]*.*/ & #G1	3	Base text y Tokon A	uery Result × Lat Frequency Ana	ilysis ^									
	Query	Lase text v   loven Annotations v											
	Builder	- K S 1 / 8 2 D Daplaying Results 1-10 of 80 Result for: G1#Gloss=0(1,*)CAT[0-9](0-9]7(A-2];*/4.#G1 •											
	2	1 0 Uss-Corput-y-ten-> 17/0640											
		PersonA::English											
		PersonA::English	Later in the evening, she really	Later in the evening, she really wanted to no home because of her cat				Vou, what time was that?				r 8.	
		PersonB::GlossType	40 AWAY1^	AWAY1^		MEASURE2B <sup>^</sup>	49    \$INDEX1	^ (I) ALONE5^	\$GEST-DEC	CLINE1^	Despite that, my morn s	ayed at my place for six hours.	
earch More ~ History ~		PersonB::Gloss	49 GO-HOME1*	BECAUSE-OF1*	40 CAT1A*	MEASURE2B <sup>^</sup>	1 \$100 - 100 -	II ALONE5	49 \$GEST-DEC	CLINE1^	<0 I2	49 FIGURE2A*	
str		PersonB::HamNoSys	48 \$\vec{-}_{**}O^{*}\)(*^{10(\pm_{2}*}\vec{-}_{2})	40 [OH26, OH26][~,8~]](\$X+]	40 * 8×+±%=2*++	40 ° OA8	-Б/ь II Ф	(1) □ - 1 + 10 <sup>-10</sup> ·	40 ° 0\84494		40 0, € B× <sup>+</sup> X	40 (dze,dz)]2)(\$X+	
10		PersonB::Mouth	nach hause	4 wegen	katze		<li>allein</li>	allein				trotzdem	
Search Options		Video  Video  Video	notated										
Korpus-de	v 0	2 O DGS-Corpus-r3-er	> 1176846									left context: 5	5 🗸 right context:
		Annotations		8 > > Displavin	g Results 1 - 10 of 80					Re	sult for: G1#Glos	s=/(.* )*CATI0-910-91?	?[A-Z]*.*/ & #G1
Corpus	<b>*</b> 2	PersonA::English			8								
DGS-Corpus-r3-en		PersonA::GlossType	Annotations										
BGS-Corpus-r3-de		PersonA::Gloss	PersonB::English	The cat got really	excited. "Grandma is ba	ack again."							
		PersonB::English	BaroonBuClosoTuno			A	O DET2A						
		PersonB::GlossType	PersonibGloss Type	WHISKERS IA"			O-PEI3"	NOT2.				W PRESENT-O	DR-HEREIT
		PersonB::Gloss	PersonB::Gloss	CAT1A*	HAPPY1	4) (	GRANDMA3*	NOT2		I ONC	E-MORE1A	PRESENT-O	DR-HERE1
		PersonB::HamNoSys	PersonB::HamNoSys	************************************	ا ت ۵٫۵۰۰ ♦	•×~ 40 c		(X) ت المحالية (X) المحالية (200)	<sup>(χ</sup> 10~ <sup>)(</sup> Ο <sup>Ψ)</sup>	dro <sup>[*</sup>	° → <u>r</u> ø]		
		FersonB::Mouth	PersonB::Mouth	<ul> <li>katze</li> </ul>	<ul> <li>froh</li> </ul>	م (1	ma	<ul> <li>nicht</li> </ul>		📣 wied	er	🐠 da	
		The second secon			2		ME	NE DGS					
		Bunks to MY DOG ar     Des Corpus Paer     PersonA::English     PersonA::Closs Type     PersonA::Closs     Demons +: Membre PersonA::Closs	Links to MY DCS − a	nnotated				NE DCS					
		Bunks to MYDGS     OSC-OTPUR-Ben     O	Links to MY DGS — a	notated		T		NE DGS					
		Bunks to MYDGS     OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R      OSC-COPUE-26-R         OSC-COPUE-26-R	Links to MY DGS – a I176846 Person MY DGS – MY DGS –	nnotated Viewer	Type rjaht Hand	Type left Hand		NE DGS					
		B Links to MP DGS — an     Constraints	Links to MY DGS – a Links to MY DGS – a Links to MY DGS – a Person MY DGS –	nnotated Viewer	Type right Hand	Type left Hand		NE DGS					
		E Links to MYOS—an     OS-Corper-2ie     Annotations PersonA::Gloss PersonA:	E Links to MY DGS – a It T684ê Person MY DGS – Person MY DGS – Person B LATAPY	nnotated Viewer	Type right Hand CATIA' HAEPY	Type left Hand		NE DOS					
		Ebist of MY OGS - are     Constantial     Constantial     PersonA: English     PersonA: Gloss Type     PersonA: Gloss     Decend 4: Missible B	E Links to MY DGS – a 1176846 Person B CAT1A* Person B HAPPY1 Person B HAPPY1	nnotated Viewer	Type right Hand CATIA* HAPPY1 CPDANEMORI	Type left Hand		NE DCS					
		<ul> <li>■ Enks to Mr065 – ar</li> <li>■ Disk to Mr065 – ar</li> <li>■ Disk Corpus Ber</li> <li>■ Arcostoins</li> <li>■ PersonA::Gloss Type</li> <li>PersonA::Gloss Type</li> <li>PersonA::Gloss</li> <li>DersonA::Gloss</li> </ul>	E Links to MY DGS – a 176846 Person MY DGS – Person B CAT1A* Person B CAT1A* Person B CAT1A*	nnotated Viewer 3'	Type right Hand CATIA* HAPPY1 GRANDMA3*	Type left Hand		NE DCS					
		Bunks to MY DGS	Links to MY DGS – a 176846 Person MY DGS – Person B CATIA* Person B HAPPY1 Person B GRANDMA Person B NOT2	nnotated Viewer	Type right Hand CATIA' HAPP'1 GRANDMA3' NOT2	Type left Hand		NE DGS					
		B Links for MOG—ar     Comparison     Comparis	E Links to MY DGS – a 1176846 Person B CAT14' Person B CAT14' Person B GRANDMA Person B GRANDMA Person B NOT2 Person B ONCE-MOF	nnotated Viewer are TA	Type right Hand CAT1A* HAPPY1 GRANDMA3* NOT2 ONCE-MORE1A	Type left Hand		NE DOS					
		E Inits to MYOS—an     Desconstructure     PersonA::Eligith     PersonA::ClossType	E Links to MY DGS – a Intræké Person B CATIA* Person B CATIA* Person B CATIA* Person B CATIA* Person B CATIA* Person B ONCE-MOF Person B ONCE-MOF Person B ONCE-MOF Person B ONCE-MOF	- annotated Viewer a* RE1A OR+HERE1	Type right Hand CATIA* HAPP11 GRANDMA3* NOT2 ONCE-MORE1A PRESENT-OR-HERE1	Type left Hand		NE DCS					
		Einis to MYDGS ar     Construction	Links to MY DGS – a TREAS Person MY DGS – Person B CAT1A* Person B HAPPY1 Person B GRANDMA Person B NOT2 Person B ONCE-MOF Person B ONCE-MOF Person B ONCE-MOF Person B ONCE-MOF	nnotated Viewer anotated Viewer as RE1A OR-HERE1 in > 1176846	Type right Hand         CATIA*         HAPPY1         GRANDMA3*         NOT2         ONCE-MORE1A         PRESENT-OR-HERE1	Type left Hand		NE DCS				left context: 5	right context
		Enrist to MYDGS	Links to MY DGS – a 176846 Person MY DGS – Person B CAT1A* Person B HAPPY1 Person B GRANDMA Person B NOT2 Person B ONCE-MOF Person B ONCE-MOF	nnotated Viewer anotated Viewer area ar	Type right Hand CATIA' HAPPY1 GRANDMA3' NOT2 ONCE-MORE1A PRESENT-OR-HERE1	Type left Hand		NE DCS				left context: 5	right context
		B Links to MP DGS — an     Construction     Constructin     Constructin     Construction     Construction     Constructi	Links to MY DGS – a 176846 Person MY DGS – Person B CAT1A* Person B HAPPY1 Person B HAPPY1 Person B GRANDMA Person B NOT2 Person B NOT2 Person B NOT2 Person B NOT2 Person B NOT2 Person B NOT2 Person B CAT1A* Person B CAT1A* Perso	Innotated annotated Viewer as as as annotated Viewer as	Type right Hand CATIA* HAPPY1 GRANDMA3* NOT2 ONCE-MORE1A PRESENT-OR-HERE1	Type left Hand						left context: 5	right context

# **ANNIS Query Language**

- Queries can include:
  - regular expression search of the text associated with an item
  - links between items in different tiers
  - collocation distances between items in the same tier
  - transcript or participant metadata

## **Gloss Annotations**

- Glosses consist of a name with optional characters and numbers for lexical and phonological variants and symbols for modification (\*) and type markers (^).
- Goal: find tokens with Gloss name CAT
- including CAT1 CAT1B CAT2A
- avoiding CATASTROPHE CATHEDRAL CATHOLIC COMPLICATED LOCATION

• Wizard:

Annotation G1 Gloss=/(.\* )\*CAT[0-9][0-9]?[A-Z]\*.\*;

#### **The Query Wizard**

- Build search queries from smaller building blocks.
- Each block designed for a specific annotation tier or metadata field.
- Users can see the AQL query change as they compose blocks, helping them learn how AQL queries for MY DGS – ANNIS work.
- Each query item is assigned a colour to aid understanding of query structure.
- Interface is available in German and English for the German or English variant of the annotations.
- Once a query has been created, it can be opened with one click in MY DGS – ANNIS.

		open query in ANN	S	
Annotation	S		Metad	ata
Tier/Level -	select > a	dd	sele	ct > add
Create Conr	nection			
Connection G1 from 1	ns between an to 2 near G2	notation eleme	nts	
Connection G1 from 1 Annotation	ns between an to 2 near G2 G1 Gloss=/(.*	notation eleme delete )*CAT[0-9][0-9]?	nts [A-Z]*.*/	✓ use in query
Connection G1 from 1 Annotation Gloss name	ns between an to 2 near G2 G1 Gloss=/(.* is exactly contains starts with	notation eleme delete )*CAT[0-9][0-9]? CAT	nts [A-Z]*.*/	✓ use in query use advanced optic
Connection G1 from 1 Annotation Gloss name Annotation	ns between an to 2 near G2 G1 Gloss=/(.* is exactly contains starts with G2 Gloss	notation eleme delete )*CAT[0-9][0-9]? CAT ✓ use in query	nts [A-Z]*.*/	✓ use in query use advanced optic

## **Collocation Distances**

- Collocation distances are distances between items on the same tier.
- In AQL, they are expressed using the dot (.) or caret (^) operators, followed by the tier name,

## **Transcript Metadata**

• **Goal:** find out in which discussion themes and transcripts CAT appears most often.

Wizard:	Annotation (	G1 Gloss=/(.*	)*CAT[0-9][0-9]?[A-Z]*.*/
	Gloss name	● is exactly ○ contains	CAT



• **Goal:** find tokens with Gloss name CAT including phonological variant 1A and modified forms.



## Links Between Tiers

- Queries can include links between tokens which occur at the same time on different tiers.
- **Goal:** find tokens with Gloss name CAT and

- then optionally by two numbers which specify the minimum and maximum distances.
- **Goal:** find tokens which appear most frequently with a distance of 1 to 2 from a token with Gloss name CAT.





## **Participant Metadata**

- **Goal:** find tokens starting with CAT signed by participants in the 18-30 age group.
- Query must be repeated to ensure metadata

translations containing the word "cat" which happen at the same time.



• ANNIS freque analysis:	ncy	25.0 20.0 15.0 10.0 5.0 0.0	20 <sup>5</sup> 2 <sup>80</sup> 3 <sup>40</sup> 2 <sup>51</sup> 1 <sup>80</sup> 1 <sup>18</sup> 1 <sup>51</sup> 1 <sup>51</sup> 1 <sup>60</sup> 1 <sup>16</sup> 1 <sup>16</sup> 1 <sup>18</sup> 1 <sup>18</sup>	PROD <sup>®</sup> BOOD <sup>®</sup> RCLIPER <sup>®</sup>
MY DGS – ANNIS		221 items rank 1 2 3 4 5 6 7 8 9 10	<pre>with a total sum of 310 (query on DGS-Corpus-r3 #G2[Gloss \$PROD* \$INDEX1 \$INDEX1* II \$INDEX1* II \$GEST-NM^ GOOD1 \$GEST-NM^ \$GEST-DECLINE1^ \$GEST-DECLINE1^ \$COD1</pre>	B-en) Count 28 18 7 6 5 3 3 3 3 3 3 3 3 3 3 3 3 3



**Poster presented** at the 11th Workshop on the Representation and Processing of Sign Languages (sign-lang@LREC-COLING 2024), Turin, Italy. 25 May 2024.



This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.