

### 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

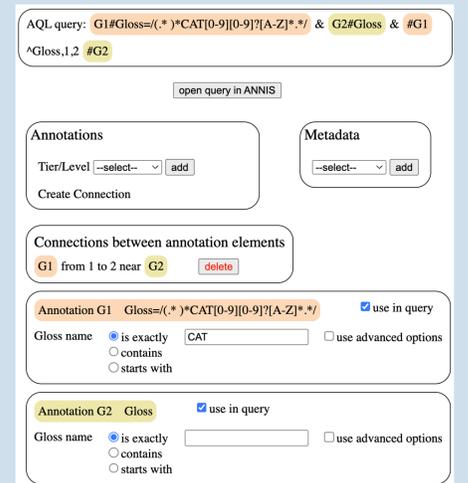


### ANNIS Query Language

- 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

### 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.



### 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]\*.\*

Gloss name  is exactly CAT  contains  starts with

AQL query: G1#Gloss=(.\*)CAT[0-9][0-9]?[A-Z]\*.\*

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

#### Wizard:

Gloss name  is exactly CAT  contains  starts with  use advanced options

all lexical variants  all phonological variants

enter lexical variant  enter phonological variant

allow supertypes  allow modified

only supertypes (^)  only modified (\*)

only subtypes  all

AQL query: G1#Gloss=(.\*)CAT[1A]\*1A\*(.\*)\*

### 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, 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.

#### Wizard:

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

Gloss name  is exactly CAT  contains  starts with  use in query

Annotation G2 Gloss  use in query

Gloss name  is exactly  contains  starts with

Create Connection Gloss G1 near 1 from 2 to Gloss G2 add

AQL query: G1#Gloss=(.\*)CAT[0-9][0-9]?[A-Z]\*.\* & G2#Gloss & #G1 ^Gloss,1,2 #G2

### 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 CAT  contains  starts with  use in query

metadata 1 Theme  use in query

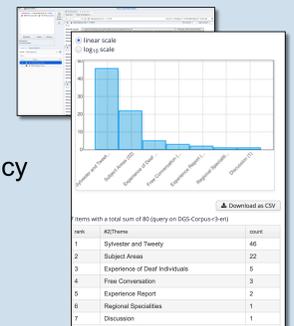
Theme any

metadata 2 TranscriptId  use in query

TranscriptId any

AQL query: G1#Gloss=(.\*)CAT[0-9][0-9]?[A-Z]\*.\* & #G1 @\* Theme & #G1 @\* TranscriptId

#### ANNIS query:



#### ANNIS frequency analysis:

rank	#2Theme	count
1	Sylvestre and Twenty	46
2	Subject Areas	22
3	Experiences of Deaf Individuals	6
4	Free Conversation	3
5	Experience Report	2
6	Regional Specialties	1
7	Discussion	1

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

#### Wizard:

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

Gloss name  is exactly CAT  contains  starts with

Annotation E1 English=/.cat.\*/

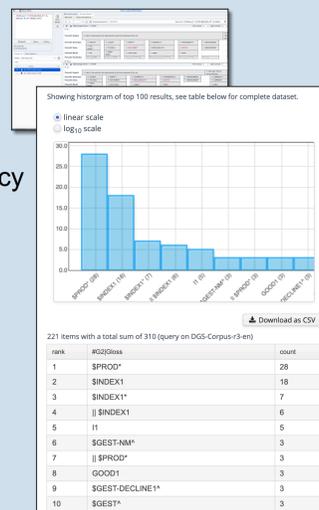
English  is exactly  contains  starts with

Create Connection Gloss G1 same time English E1 add

AQL query: G1#Gloss=(.\*)CAT[0-9][0-9]?[A-Z]\*.\* & E1#English & #G1 ->ident #E1

### MY DGS – ANNIS

- **ANNIS query:**
- **ANNIS frequency analysis:**



### Participant Metadata

### Participant Metadata

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

#### Wizard:

Annotation G1 Gloss=(.\*)CAT.\*

Gloss name  is exactly CAT  contains  starts with  use in query

metadata 1 AgeGroup  use in query

AgeGroup  any  18-30  31-45  46-60  61+

AQL query: ( G1#PersonA:Gloss=(.\*)CAT.\* @\* PersonA:AgeGroup="18-30" ) | ( G1#PersonB:Gloss=(.\*)CAT.\* @\* PersonB:AgeGroup="18-30" )