

Suche und Visualisierung von Annotationen mit ANNIS

Hands On Workshop

Korpuslinguistik und Morphologie
Humboldt-Universität zu Berlin



TraCES Projekt, Universität Hamburg | 26.4.2018

Werkauftrag für den Importer

- Quotations implementieren, sobald ich dafür Testdaten habe
- Kosmetik in ANNIS: Grids neu anordnen wenn/wie gewünscht
+ Links sich in neuem Fenster öffnen lassen
- Einbettung des Importers in die Pepper Grinder GUI

Bilddokumentation – Entwurf Anordnung der Grids in ANNIS

Gruppen Summation	
Flat Morphology	F10EO, TOKL, lex, NT,
Comment	TOKL, comment, F10EO, Comment
Deep Morphology	F10EO, TOKL, lex, NT (num, person, gender), number, case, state, genderxxx, numberxxx, NE (logogram,)
Div	way, LE, NR, NET, CR
NR	F10EO, TOKL, T, NET, R
Layout	F10, F10EO, F10EO, F10GT, F10LTD, TOKL, Tok, a, b, c, ^{play} , fl
All	X

Ziele für den Workshop

I. Teil Hands On Einführung

- ANNIS als Such- und Visualisierungstool
- Suche in ANNIS – Wiederholung
 - Query Language
 - Verschiedene Annotationskonzepte – Fokus auf Token- und Spannenannotationen
 - Suche nach Werten, nach Mustern, nach Kombinationen von Annotationen, Frequenzanalyse
- Aufgabengeleitet, flexibel im Ablauf, Fragen jederzeit

Ziele für den Workshop

II. Teil Gemeinsame offene Diskussionsrunde

- Ihre Fragen!
- Reihenfolge / Anordnung der Annotationen in ANNIS

① ANNIS Such- und Visualisierungstool

② Hands On – Einführung
Variable-Wert-Paare
Relationen zwischen Annotationen

③ Frequenzanalyse

Plan

① ANNIS Such- und Visualisierungstool

② Hands On – Einführung
Variable-Wert-Paare
Relationen zwischen Annotationen

③ Frequenzanalyse

ANNIS – Suchwerkzeug für tief annotierte Korpora

- Entwicklung Thomas Krause, Arbeitsgruppe Korpuslinguistik
- SFB 632 Informationsstruktur (Universität Potsdam und Humboldt-Universität zu Berlin)
- Amir Zeldes, Georgetown University
- Open Source Entwicklung
 - Offen für Mitarbeit anderer Arbeitsgruppen
 - Kann auf eigenem Server oder Desktop/Notebook installiert werden, kein Zwang zu zentralem Service

ANNIS – Suchwerkzeug für tief annotierte Korpora

- Generelle Funktionen
 - Finden von Annotationen und Kombinationen von Annotationen
 - Filterung nach Metadaten
 - Frequenzanalyse
 - Key Word In Context

ANNIS – Suchwerkzeug für tief annotierte Korpora

- Besonderheiten
 - Verschiedene Visualisierungen
 - Konzepte wie Tokenannotation, Spannenannotation, Pointing Annotationen oder Bäume
 - Agnostisch gegenüber korpuspezifischen Annotationswerten → Unterstützung möglichst vieler (auch externer) Korpora
 - Eigene Abfragesprache – ANNIS Query Language (AQL)
 - Referenzierung von Suchanfragen und Treffern mit einem permanenten Link (Server-Anwendung)
 - Spezialisiert für tief annotierte Mehrebenenkorpora und Nichtstandardvarietäten

Plan

① ANNIS Such- und Visualisierungstool

② Hands On – Einführung
Variable-Wert-Paare
Relationen zwischen Annotationen

③ Frequenzanalyse

Prinzipien von AQL

- I Variable-Wert-Paare (Attribut-Wert-Paare, Layer-Wert-Paare)
- II Relationen zwischen Variable-Wert-Paaren

Prinzipien von AQL

- I Variable-Wert-Paare (Attribut-Wert-Paare, Layer-Wert-Paare)
- II Relationen zwischen Variable-Wert-Paaren

Variable-Wert-Paare

FID	નહાં	નહાં	ફોર્માન				અફોર્માન							
FIDED	નહાંઃ	નહાંઃ	ફોર્માનઃ				અફોર્માનઃ							
FIDEKh	નહાં	નહાં	ફોર્માન				અફોર્માન							
FIDLET	ગ્રંથઃ	ગ્રંથ	નામઃ	નામ	સ્ત્રી	સ્ત્રી	જીવિ	જીવિ	માત્રા	માત્રા				
FIDLETED	ગ્રંથ	ગ્રંથ	નામ	નામ	સ્ત્રી	સ્ત્રી	જીવિ	જીવિ	માત્રા	માત્રા				
Id	1De0c19c47-0094-4a91-97bc-4e0f761ba632													
LE	1													
NA	sentence													
NEId							N5fcc4a71-2de6-455c-970f-aeb1989a6927		N5fcc4a71-2de6-455c-970f-aeb1989a6927					
NR	3													
NRI	3													
NT	Conjunction	Verb	Proper Name				Common Noun							
Name			James											
Occupation			apostle											
R			http://betamasaheft.eu/api/PRS5662James/tei											
Sex			1											
T			person											
TOKL	kama	yabe	yā'qob				hawāreŷā							
TR	ba-kama	yabe	yā'qob				hawāreŷā							
TRFID	ma	ya	be	yā	'	qo	b	ha	wā	ra	yā			

- **Variable:** Annotationsebene
 - z. B. FID, FIDED, NT, TR, Name, NA
- **Wert:** zugewiesene Kategorie
 - Offene und geschlossene Wertelisten, Links und IDs

Wissen über Korpora

- Annotationsrichtlinie
- Beispiel für Variable-Wert-Paare in den Metadaten
- Zusammenspiel von im Korpus enthaltenen Annotationen und der Suche nach diesen Annotationen

Aufgaben | Beispielanfragen

Hinweise zu den Aufgaben | Beispielanfragen:

- Die Auswahl der Anfragen und ihre Lösungen sind nach methodischen Gesichtspunkten (z.B. Prinzipien AQL) zusammengestellt.
- Die Ergebnisse der Anfragen und der Frequenzanalyse sind zum Teil auf einzelnen Texten des Korpus ausgeführt.
- Die Darstellungen der Annotationen und Frequenzanalysen sind Ausschnitte.

Metadaten

The screenshot shows the ANNIS GUI interface. On the left, there's a sidebar with a search bar, a 'More' dropdown, and a 'History' dropdown. Below these are sections for 'Corpus List' and 'Search Options'. The 'Visible' dropdown is set to 'All'. A 'Filter' section is present. The main area displays a list of corpora with columns for 'Name', 'Texts', and 'Tokens'. The first item in the list is 'ExportAmidaSeyonVer18_C_1' with 74.143 texts and 10 tokens. To the right of the list is a 'Base text' panel showing the path 'Path: ExportAmidaSeyonVer18_04_19_01_19_09 > AmidaSeyon18_04_19_01_19_09_EA (tokens 113 - 124)'. Below this is a 'No metadata available' message. A dropdown menu is open over the corpus name, showing options: 'ExportAmidaSeyonVer18_04_19_01_19_09' (selected) and 'AmidaSeyon18_04_19_01_19_09_EA'. To the right of the base text panel is a large 'Available annotations' section. This section contains a table with columns for 'Name', 'Example (click to use query)', and 'URL'. It includes rows for 'case' (example: 'case="Nominitive"', URL: '#'), 'Comm' (example: 'Comm="#<div>According to Huntingford, the author of the text miscalculated the Year of Mercy, which should be 514 instead of 516 (=1329 A.D.). For more datalists about this issue, cp. Huntingford 1965, p. 53, n. 2.'), 'Comment' (example: 'Comment="<p></p>"'), 'comment' (example: 'comment="#<div>Typing mistake in Marrassini's edition, where the word is written as ቀዱስቀ see Marrassini 1993, p. 866. </div>"'), 'CR' (example: 'CR="Annotator"'), 'distance' (example: 'distance="Near"'), 'FID' (example: 'FID="a"'), 'RIDED' (example: 'RIDED="a"'), 'RIDEKh' (example: 'RIDEKh="a"'), 'Edge Annotations' (example: 'Edge Annotations'), 'Edge Types' (example: 'Edge Types'), and 'Meta Annotations' (example: 'Meta Annotations'). At the bottom of the main panel, there's a link to the corpus: 'Link to corpus: http://localhost:8080/annis-gui/#.c=RXhwbdj0QW1kYVNLew3uVmVMyMThfMDRfMTfMDfMTfMDk'.

Metadaten

The screenshot shows the Annis GUI interface with the following components:

- Left Panel:** Contains a search bar with "FIDLET=/*/" and a "Search" button, a "More" dropdown, and a "History" dropdown. Below these are sections for "Visible: All" and "Corpus List". The "Corpus List" table has columns for Name, Texts, and Tokens, listing various corpora like "4Kaufleute_TEI", "ExportAkalKrestosGeTa%", "ExportAmdaSeyonVer18_1", etc.
- Top Center:** A navigation bar with "Base text" dropdown, page numbers (1 / 76), and a "Displaying Results 1 - 10 of 758" message.
- Middle Column:** A table titled "Select corpus/document:" showing the selected item "AmdaSeyon18_04_19_01_19_09_EA". It lists various metadata fields with their values:
 - GeTa:SCR: 0
 - GeTa:TR: 0
 - GeTa_META:ANNOT: Vitagrazia Pisani
 - GeTa_META:Comm
 - GeTa_META:DATE: 18_04_18_15_52_20
 - GeTa_META:EDITION
 - GeTa_META:LANG: Târika 'Āmda Ṣeyon/Chronicle of 'Amida Seyon
 - GeTa_META:NAME: Târika 'Āmda Ṣeyon/Chronicle of 'Amida Seyon
 - GeTa_META:PARTS: 1, 2
 - GeTa_META:SCR: 0
 - GeTa_META:SOFT: Ver4/ CreatorCristina Vertan
 - GeTa_META:TR: 0
 - annis:doc: AmdaSeyon18_04_19_01_19_09_EA
- Right Column:** A panel titled "Available annotations" containing a table of node annotations with columns for Name, Example (click to use query), and URL.
 - case: case="Nominative"
 - Comm: Comm=<div>> According to Huntingford, the author of the text miscalculated the Year of Mercy, which should be 514 instead of 516 (= 1329 A.D.). For more details about this issue, cp. Huntingford 1965, p. 53, n. 2. </div>"
 - Comment: Comment=<p> </p>"
 - comment: comment=<div>> Typing mistake in Marrassini's edition, where the word is written as ቀዱስስ see Marrassini 1993, p. 86:6, </div>"
 - CR: CR="Annotator"
 - distance: distance="Near"
 - FID: FID="#"
 - FIDED: FIDED="#"
 - FIDEDh: FIDEDh="#"
 - FOLEY: FOLEY="#"
- Bottom:** A link to the corpus: <http://localhost:8080/annis-gui/#.c=R0hwB3j0QW1kyVNleW9uvmVymThfMDRMfTfMDFfMTfMDk>

Plan

① ANNIS Such- und Visualisierungstool

② Hands On – Einführung
Variable-Wert-Paare
Relationen zwischen Annotationen

③ Frequenzanalyse

Variable-Wert-Paare – 1

- Suchen Sie nach Verben!

CR	Annotator							
FID	፩፻፻፻፻፻							
FIDED	፩፻፻፻፻፻፡፡							
FIDEDh	፩፻፻፻፻፡፡፡፡							
FIDLET	አ	ም	አ	ከ፡፡	ን	ኝ	ከ	ኩ፡፡
FIDLETED	አ	ም	አ	ከ	ን	ኝ	ከ	ኩ
Id	1D48d8b56d-1e74-499c-981f-1a3ba6d7c39a							
LE	1							
NA	sentence							
NR	1							
NRI	1							
NT	Common Noun				Verb			
TOKL	'amläk				naṣəḥef			
TR	1-'amläk				naṣəḥef			
TRFID	'a	m	lä	k	nə	ṣə	ḥə	f

Variable-Wert-Paare – 1

- Suchen Sie nach Verben!

CR	Annotator							
FID	፩፻፻፻፻፻							
FIDED	፩፻፻፻፻፻፻፻							
FIDEDh	፩፻፻፻፻፻፡፡							
FIDLET	አ	ም	አ	ከ፡	ን	ኝ	ከ	ኋ፡
FIDLETED	አ	ም	አ	ከ	ን	ኝ	ከ	ኋ
Id	1D48d8b56d-1e74-499c-981f-1a3ba6d7c39a							
LE	1							
NA	sentence							
NR	1							
NRI	1							
NT	Common Noun				Verb			
TOKL	'amlák				naṣəḥəf			
TR	1-'amlák				naṣəḥəf			
TRFID	'a	m	lä	k	nə	ṣə	ḥə	f

Variable: NT, Wert: Verb

- NT="Verb"

Variable-Wert-Paare – 2

- Suchen Sie nach dem Tokenlabel 'nəşəhəf'!

CR	Annotator							
FID	દ્વારા લખાયાની વિધાનનું							
FIDED	દ્વારા લખાયાની વિધાનનું:							
FIDEDh	દ્વારા લખાયાની :							
FIDLET	હ	ગૂ	અ	હઃ	ન	ક	હન	ફઃ
FIDLETED	હ	ગૂ	અ	હ	ન	ક	હન	ફ
Id	1D48d8b56d-1e74-499c-981f-1a3ba6d7c39a							
LE	1							
NA	sentence							
NR	1							
NRI	1							
NT	Common Noun				Verb			
TOKL	'amlák				nəşəhəf			
TR	1-'amlák				nəşəhəf			
TRFID	'a	m	lā	k	nə	şə	hə	f

Variable-Wert-Paare – 2

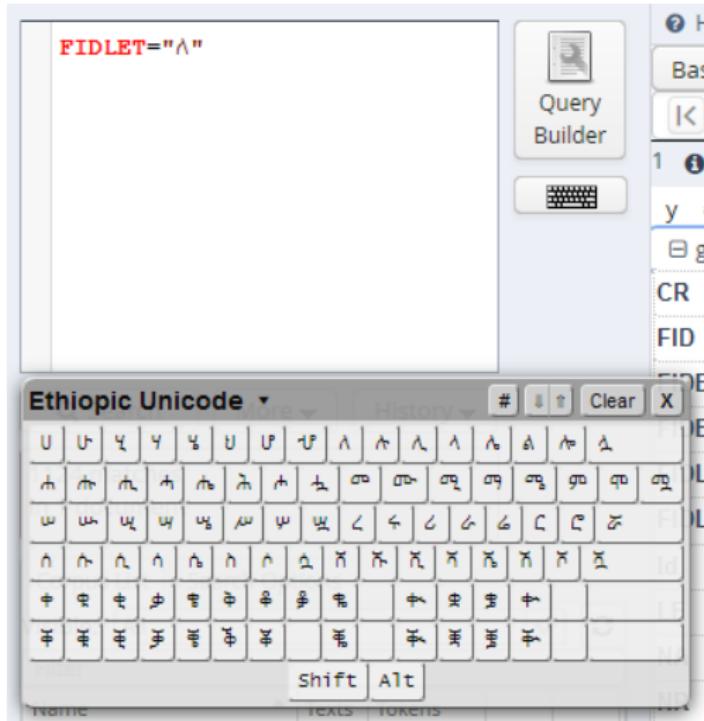
- Suchen Sie nach dem Tokenlabel 'nəʃəhəf'!

CR	Annotator							
FID	፩፻፭፻፪ሺ							
FIDED	፩፻፭፻፪ሺ:							
FIDEDh	፩፻፭፻፪ሺ:							
FIDLET	አ	ጥ	ለ	ከ	ን	ኋ	ከ	ቁ
FIDLETED	አ	ጥ	ለ	ከ	ን	ኋ	ከ	ቁ
Id	1D48d8b56d-1e74-499c-981f-1a3ba6d7c39a							
LE	1							
NA	sentence							
NR	1							
NRI	1							
NT	Common Noun				Verb			
TOKL	'amlák				nəşəhəf			
TR	1-'amlák				nəşəhəf			
TRFID	'a	m	lā	k	nə	şə	hə	f

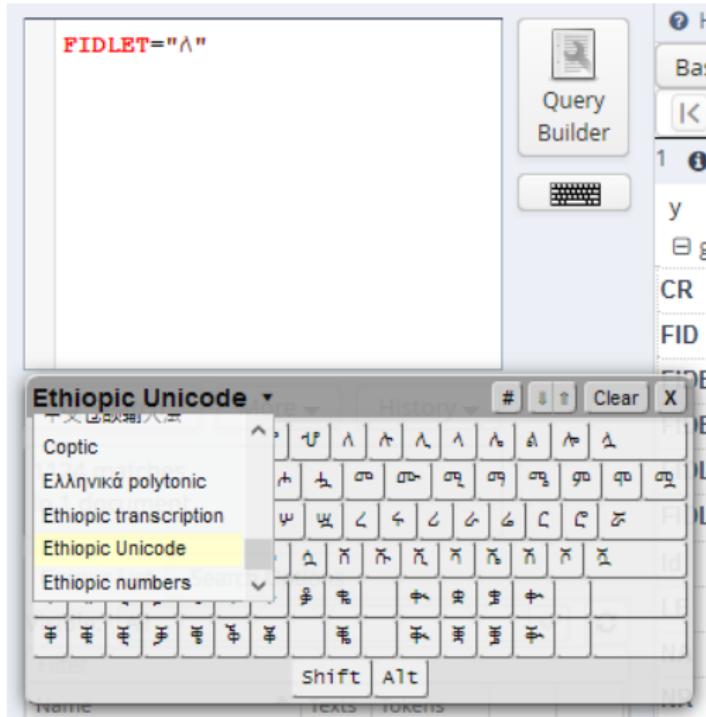
Variable: TOKL, Wert: 'nəʂəhɛf'

- TOKL**="nəşəhəf"

Virtuelle Tastatur



Virtuelle Tastatur



Variable-Wert-Paare – Mustersuche

Reguläre Ausdrücke können im **Wert** einer Variable angewendet werden:

regEx Funktion

- . ein beliebiges Zeichen
- ? 0 oder 1 Zeichen des vorherigen Zeichens im Wert
- * 0 bis unendlich viele Zeichen (des vorher. Zeichens)
- + 1 bis unendlich viele Zeichen (des vorher. Zeichens)
- (a|b) a oder b (auch a und b)

Variable-Wert-Paare – 3

- Suchen Sie nach Wörtern mit Ge'ez Buchstaben, die ል mit anfangen!

FID	ለማሙያ
FIDED	ለማሙያ:

Variable-Wert-Paare – 3

- Suchen Sie nach Wörtern mit Ge'ez Buchstaben, die  mit anfangen!

FID	
FIDED	

Variable: FID, 

- FID**=/./

Variable-Wert-Paare – Anfrage im Workshop konzipiert

- Mustersuche nach bestimmten Flexionsformen eines Verbs.
- Variable: TOKL, NT

```
TOKL=/ .@.@.@./ &
```

```
NT="Verb" &
```

```
#1_ _ _ #2
```

Variable-Wert-Paare – Anfrage im Workshop konzipiert

- Mustersuche nach bestimmten Flexionsformen einer Wortform, die kein Verb ist.
- Variable: TOKL, NT

TOKL=/*nə.ə.ə./* &

NT!="Verb" &

#1 _=_ #2

Variable-Wert-Paare – 4

- Suchen Sie nach allen Arten von Partikeln!

Variable-Wert-Paare – 4

- Suchen Sie nach allen Arten von Partikeln!

Variable: NT, **Wert:** am Ende *Particle*

- $\text{NT} = /.+\text{Particle}/$

Variable-Wert-Paare – 4

- Suchen Sie nach allen Arten von Partikeln!

Variable: NT, **Wert:** am Ende *Particle*

- $\text{NT} = /.+\text{Particle}/$

Variable-Wert-Paare – 5

- Suchen Sie nach Wörtern im Akkusativ oder Nominativ!

Variable: NT, Wert: *Accusative, Nominative*

Variable-Wert-Paare – 5

- Suchen Sie nach Wörtern im Akkusativ oder Nominativ!

Variable: NT, **Wert:** *Accusative, Nominative*

Variable-Wert-Paare – 5

- Suchen Sie nach Wörtern im Akkusativ oder Nominativ!

Variable: NT, **Wert:** *Accusative, Nominative*

- `case=/(Accusative|Nominative)/`

Plan

① ANNIS Such- und Visualisierungstool

② Hands On – Einführung
Variable-Wert-Paare
Relationen zwischen Annotationen

③ Frequenzanalyse

Prinzipien von AQL

- I Variable-Wert-Paare (Attribut-Wert-Paare, Layer-Wert-Paare)
- II Relationen zwischen Variable-Wert-Paaren

Relationen zwischen Variable-Wert-Paaren

Wir stellen uns häufiger komplexere Fragen:

- Suche nach einer Variable oder einem Wert nicht immer ausreichend
- Kombinationen von verschiedenen Variablen und Werten notwendig
- Mehrere Arten von Beziehungen zwischen Annotationen möglich

Variable-Wert-Paare

FID	નહાં	લબ્ધ	ફોર્માન				અન્યાં							
FIDED	નહાંઃ	લબ્ધઃ	ફોર્માનઃ				અન્યાંઃ							
FIDEKh	નહાં	લબ્ધ	ફોર્માન				અન્યાં							
FIDLET	દ્વાઃ	દ્વા	લબ્ધઃ	ફ	દ્વા	સ્ત્રી	જાઃ	શ્વર	મા	લ				
FIDLETED	દ્વા	દ્વા	લબ્ધ	ફ	દ્વા	સ્ત્રી	જા	શ્વર	મા	લ				
Id	1De0c19c47-0094-4a91-97bc-4e0f761ba632													
LE	1													
NA	sentence													
NEId							N5fcc4a71-2de6-455c-970f-aeb1989a6927		N5fcc4a71-2de6-455c-970f-aeb1989a6927					
NR	3													
NRI	3													
NT	Conjunction	Verb	Proper Name				Common Noun							
Name			James											
Occupation			apostle											
R			http://betamasaheft.eu/api/PRS5662James/tei											
Sex			1											
T			person											
TOKL	kama	yabe	yā'qob				hawāreyā							
TR	ba-kama	yabe	yā'qob				hawāreyā							
TRFID	ma	ya	be	yā	'	qo	b	ha	wā	ra				
										yā				

- **Variable:** Annotationsebene
 - z. B. FID, FIDED, NT, TR, Name, NA
- Token- und Spannenannotationen
 - Verschiedene Relationen möglich!

Relationen zwischen Annotationen – Operatoren

- Zwei Annotationen können über den Textbereich, den sie überlappen/abdecken, verglichen werden, z. B.

- Präzedenz (“kommt danach”) .

- Identische Textabdeckung _=_

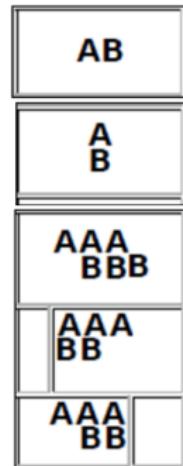
- Überlappung _o_

- Überlappung + links aligniert _l_

- Überlappung + rechts aligniert _r_

→ Setzen immer zwei Variablen mit einander in Verbindung.

→ Können beliebig oft und in verschiedenen Konstellationen eingesetzt werden.



Relationen zwischen Annotationen – 6

- Suchen Sie ein Verb gefolgt von einem Eigennamen!

FID	ନିର୍ମାଣ	କରି	ପୋଷିନ୍ତି				ଅପରିପ୍ରକାଶ				
FIDED	ନିର୍ମାଣ:	କରି:	ପୋଷିନ୍ତି:				ଅପରିପ୍ରକାଶ:				
FIDEDh	ନିର୍ମାଣ	କରି	ପୋଷିନ୍ତି				ଅପରିପ୍ରକାଶ				
FIDLET	ଦେଇ	ଦେ	କରି	ପୋ	ବୁଝି	ବୁଝି	ଜାଣି	ଜାଣି	ପାଇଲା	ପାଇଲା	
FIDLETED	ଦେଇ	ଦେ	କରି	ପୋ	ବୁଝି	ବୁଝି	ଜାଣି	ଜାଣି	ପାଇଲା	ପାଇଲା	
Id	1De0c19c47-0094-4a91-97bc-4e0f761ba632										
LE	1										
NA	sentence										
NEId							N5fcc4a71-2de6-455c-970f-aeb1989a6927	N5fcc4a71-2de6-455c-970f-aeb1989a6927			
NR	3										
NRI	3										
NT	Conjunction	Verb	Proper Name				Common Noun				
Name			James								
Occupation			apostle								
R			http://betamasaheft.eu/api/PRS5662James/tei								
Sex			1								
T			person								
TOKL	kama	yabe	yā'qob				hawāreyā				
TR	ba-kama	yabe	yā'qob				hawāreyā				
TRFID	ma	ya	be	ya	'	qo	b	ha	wā	rā	yā

Relationen zwischen Annotationen – 6

- Suchen Sie nach einem Verb gefolgt von einem Eigennamen!

Relationen zwischen Annotationen – 6

- Suchen Sie nach einem Verb gefolgt von einem Eigennamen!
- **Variablen:** NT, **Werte:** *Verb* und *Proper Name*

Relationen zwischen Annotationen – 7

- Suchen Sie nach einem Verb gefolgt von einem Eigennamen!
- Variablen: NT, Werte: *Verb* und *Proper Name*

```
NT="Verb" &  
NT="Proper Name" &  
#1 . #2
```

Relationen zwischen Annotationen – 8

- Suchen Sie nach Verben im Imperfekt!

Relationen zwischen Annotationen – 8

- Suchen Sie nach Verben im Imperfekt!
- **Variablen:** NT, tam, **Werte:** *Verb, Imperfect*

Relationen zwischen Annotationen – 8

- Suchen Sie nach Verben im Imperfekt!
- Variablen: NT, tam, Werte: *Verb, Imperfect*

```
NT="Verb" &  
tam="Imperfect" &  
#1=_#2
```

Relationen zwischen Annotationen – 9

- Negative Particle + Verb (tam: Imperfect; person: First; gender: number: Singular)
- Variablen: NT, tam, person, number, Werte: *Verb, Imperfect, First, Singular*

Relationen zwischen Annotationen – 9

- Negative Particle + Verb (tam: Imperfect; person: First; gender: number: Singular)
- **Variablen:** NT, tam, person, number, **Werte:** *Verb, Imperfect, First, Singular*

Relationen zwischen Annotationen – 9

- Negative Particle + Verb (tam: Imperfect; person: First; gender: number: Singular)
- Variablen: NT, tam, person, number, Werte: Verb, Imperfect, First, Singular

```
NT="Negative Particle" &
NT="Verb" &
#1. #2 &
tam="Imperfect" &
#2=_#3 &
person="First" &
#2=_#4 &
number="Singular" &
#2=_#5
```

Plan

① ANNIS Such- und Visualisierungstool

② Hands On – Einführung
Variable-Wert-Paare
Relationen zwischen Annotationen

③ Frequenzanalyse

Frequenzanalyse

Wir stellen uns ebenfalls Fragen nach Vorkommen:

- Frequenzen von Treffern
- Filterung durch Metadaten
- Export in Tools zur statistischen Auswertung

Frequenzanalyse – 14

- Gesucht werden als *Existentials* (negativ oder affirmativ) annotierte Token gefolgt von Nomen im Akkusativ oder gefolgt von Nomen im Nominativ.
- Variable: NT, case

Frequenzanalyse – 14

- Gesucht werden als *Existentials* (negativ oder affirmativ) annotierte Token gefolgt von Nomen im Akkusativ oder gefolgt von Nomen im Nominativ.
- **Variable:** NT, case

Frequenzanalyse – 14

- Gesucht werden als *Existentials* (negativ oder affirmativ) annotierte Token gefolgt von Nomen im Akkusativ oder gefolgt von Nomen im Nominativ.
- Variable: NT, case

```
NT=/Existential.*/ .*
NT="Common Noun" &
case=/ (Nominative|Accusative)
/ &
#2 _=_#3
```
- Frequenzanalyse

Frequenzanalyse – 14

rank	#1 NT	#2 NT
1	Existential Affirmative Base	Common Noun
2	Existential Negative Base	Common Noun
3	Existential Affirmative Base	Common Noun
4	Existential Negative Base	Common Noun

Anfrage im Workshop konzipiert

- Gesucht werden die ersten beiden FID-Einheiten am Anfang einer Sentence-Einheit. Paraphrasiert: die ersten beiden Wörter eines Satzes (wobei die Sentence-Annotation hier vereinfacht als Satz angenommen wird).
- **Variable:** FID, NA

```
FID &  
NA="sentence" &  
#1 _1_ #2  
& FID & #1 . #3
```

Anfrage im Workshop konzipiert

- Gesucht werden Nomen oder Verben im Construct State, die entweder im Akkusativ oder Dativ stehen.
- Variable: NT, state, case

```
case=/ (Accusative|Nominative)
/ &
NT=/ (Common Noun|Verb) / &
#1 _=_ #2 &
state="Construct state" &
#2 _=_ #3
```

Vielen Dank für Ihre
Aufmerksamkeit!



corpus-tools.org/annis

Dokumentationen zu ANNIS

- Zum Nachlesen
 - Dokumentation <http://corpus-tools.org/annis/documentation.html>
 - Nutzungshandbuch
http://corpus-tools.org/annis/resources/ANNIS_User_Guide_3.4.3.pdf
 - Krause, Thomas and Zeldes, Amir (2016) "ANNIS3: A New Architecture for Generic Corpus Query and Visualization". *Digital Scholarship in the Humanities* 31(1), 118-139.
<https://academic.oup.com/dsh/article/31/1/118/2605493/>
ANNIS3-A-new-architecture-for-generic-corpus-query
- Zum Nachschauen
 - Videotutorials <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/corpus-tools/annis-tutorials>

Anhang – Beispielanfrage A

- Suche nach Nomen mit allen Genera.
- Variablen: NT, Gender TÖKL

Anhang – Beispielanfrage A

- Suche nach Nomen mit allen Genera.
- **Variablen:** NT, Gender TOKL

Anhang – Beispielanfrage A

- Suche nach Nomen mit allen Genera.
- **Variablen:** NT, Gender TOKL
`NT="Common Noun" &`
`gender &`
`TOKL &`
`#1 _=_ #2 &`
`#1 _=_ #3`
- Frequenzanalyse

Anhang – Beispielanfrage A

#1 NT	#2 gender	#3 TOKL	count
Common Noun	Unmarked	ɖəmāwe	1
Common Noun	Unmarked	mā'bala	1
Common Noun	Unmarked	mā'leta	1
Common Noun	Masculine	səfəħa	1
Common Noun	Unmarked	bāħatitə	1
Common Noun	Unmarked	kʷərbatāta	1
Common Noun	Unmarked	śərnāy	1
Common Noun	Unmarked	tərgʷāme	1
Common Noun	Unmarked	mazāgəbti	1
Common Noun	Feminine	lelita	1
Common Noun	Unmarked	'azə'ebət	1
Common Noun	Unmarked	baħāwərtə	1
Common Noun	Unmarked	gəb'ata	1
Common Noun	Unmarked	qat̪tantə	1
Common Noun	Unmarked	ħəywət	1
Common Noun	Masculine	sab'a	1

Anhang – Beispielanfrage B

- Wie häufig wurde die Wortform *haba* als Präposition, Konjunktion oder Adverb annotiert?
- **Variable:** NT und TR

Anhang – Beispielanfrage B

- Wie häufig wurde die Wortform *haba* als Präposition, Konjunktion oder Adverb annotiert?
- Variable: NT und TR
TR=/*haba*/ &
NT &
#1=_#2
- Frequenzanalyse

Anhang – Beispielanfrage B

rank	#1 TR	#2 NT
1	ḥaba	Preposition
2	ḥaba	Conjunction