

ANNIS-Schulung für historische Korpora: Referenzkorpus Altdeutsch

Carolin Odebrecht, Karin Donhauser, Anke Lüdeling, Thomas Krause,
Laura Perlitz, Gohar Schnelle

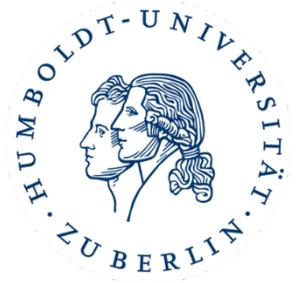
Sprachgeschichte und Korpuslinguistik

Humboldt-Universität zu Berlin

ANNIS-Schulung

- Organisiert von den Arbeitsgruppen Sprachgeschichte und Korpuslinguistik von
 - Prof. Dr. Karin Donhauser und Prof. Dr. Anke Lüdeling
 - mit verantwortlich für die Erstellung des Referenzkorpus Altdeutsch und Entwicklung des Suchtools ANNIS
- Schulung für Fortgeschrittene
 - neues Format, um die Funktionsweise von ANNIS vorzustellen
 - neues Format, um das Referenzkorpus Altdeutsch vorzustellen
- Ziel der Schulung
 - Zitation und Referenzierung von Suchanfragen und Treffern in ANNIS
 - Generierung von (komplexen) Suchanfragen
 - Verstehen von Annotationskonzepten und deren Visualisierungen
 - Exportfunktionen
 - best practice: Wie erarbeite ich mir Suchanfragen für die eigene Forschung?

ANNIS-Schulung



Wie arbeiten Sie bislang mit ANNIS?

Plan

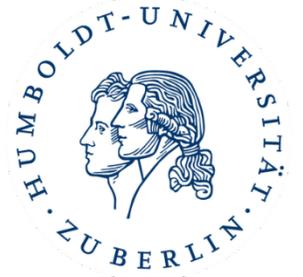
- I. ANNIS
- II. Referenzkorpus Altdeutsch
- III. Hands-On-Suche
 - Referenzieren mit ANNIS
 - AQL Prinzipien
 - Best Practice Beispiele 1+2
- IV. Exkurs
 - Datenaufbereitung für ANNIS
- V. Offene Diskussion
 - Ihre Beispiele
 - Ihre Fragen
 - Auswertung und Anregungen

I. ANNIS

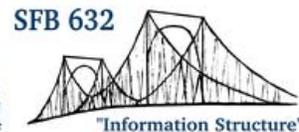




ANNIS



- ANNIS ist ein Suchwerkzeug, mit dem man in annotierten Korpora über alle **vorhandenen** Ebenen hinweg suchen kann
 - unterstützt verschiedene Annotationskonzepte wie Tokenannotation, Spannenannotation, Pointer und Bäume
- ANNIS hat eine eigene Suchabfragesprache AQL (ANNIS Query Language), die vielen anderen Korpusabfragesprachen ähnlich ist
 - basierend auf Variable-Wert-Paaren und Regulären Ausdrücken
- Kooperation:



Referenz:

- **Krause, Thomas; Zeldes, Amir** (2016): ANNIS3: A new architecture for generic corpus query and visualization. in: *Digital Scholarship in the Humanities* 2016 (31).
<http://dsh.oxfordjournals.org/content/31/1/118>
- Projekt: <http://corpus-tools.org/annis/>

II. Referenzkorpus Altdeutsch



Referenzkorpus Altdeutsch

- Historisches Korpus des Altdeutschen (Donhauser 2015)
 - gesamte ahd. Überlieferung bis auf den Glossenwortschatz des Deutschen 750-1050
 - z.B. Isidor, Tatian, Otfrid, Notker, Heliand und weitere kleine Texte
 - ca. 300 000 Token
 - Annotationen: Markup, Wortarten, Sätze, Flexionsmorphologie, Lemmata
- Kooperation:



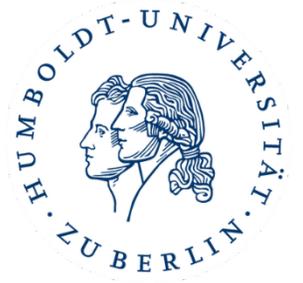
seit 1558



Referenz:

- Donhauser, Karin; Gippert, Jost; Lühr, Rosemarie; ddd-ad (Version 0.1), Humboldt-Universität zu Berlin.
<https://referenzkorpusaltdeutsch.wordpress.com/>
<http://hdl.handle.net/11022/0000-0000-7FC2-7> , CC-BY Lizenz
- Projekt: <http://www.deutschdiachrondigital.de/home/>

III. Hands-On-Suche mit ANNIS



AQL Prinzipien

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

AQL – Variable-Wert-Paare

- lemma=**/got/**

Variable (Layer, Tier, Ebene ...)

Wert (Wortart, Satz, Lemma...)

Setzt eine Ebene namens „lemma“ voraus. Wert in Schrägstrichen findet exakt diese Zeichenkette!

edition	si	kehlaufen	kepoto	cotes
ling	si	kehlaufen	kepoto	cotes
lemma	sîn	giloufan	gibot	got
posLemma	VA	VV	NA	NA
pos	VAFIN	VVPP	NA	NE

AQL - Mustersuche

- gegeben die Variable *pos*
 - enthält u.a. Werte VMFIN, VAFIN, VVFIN
- Welche Werte erwarten Sie bei folgenden Anfragen zu finden?
- $pos=/V.FIN/$
 - Punkt-Platzhalter
 - Findet alle Verben, die finit sind: VAFIN, VMFIN, VVFIN ...
- $pos=/N(A|E)/$
 - Alternatives Zeichen (a|b)
- $pos=/((VM)|(VA))FIN/$
 - Alternative Zeichenketten ((ab)|(ac))
 - Findet alle Modal- und Hilfsverben, die finit sind: VMFIN und VAFIN
- $pos=/V.+/$
 - Findet alle Verben.
 - VMFIN, VAFIN, VVFIN, ...

AQL – Relationen

edition	dera	suazzi	si	kehlaufen	kepoto	cotes
ling	dera	suazzi	si	kehlaufen	kepoto	cotes
lemma	dër	su03ī	sīn	giloufan	gibot	got
posLemma	DD	NA	VA	VV	NA	NA
pos	DDA	NA	VAFIN	VVPP	NA	NE
inflectionClassLemma		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflectionClass		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflection	FEM_SG_DAT	SG_DAT	SUBJ_PRESENT_SG_3		PL_GEN	SG_GEN
lang	goh	goh	goh	goh	goh	goh
clause	CF					

Trefferreferenzlink: <https://korpling.org/annis3/?id=406b5ce3-be8a-4a0e-92d7-0921329bbd27>

AQL – Relationen Präzedenz „.“



1	2	3	4	5	6
---	---	---	---	---	---

	1	2	3	4	5	6
edition	dera	suazzi	si	kehlaufen	kepoto	cotes
ling	dera	suazzi	si	kehlaufen	kepoto	cotes
lemma	dēr	suozī	sīn	giloufan	gibot	got
posLemma	DD	NA	VA	VV	NA	NA
pos	DDA	NA	VAFIN	VVPP	NA	NE
inflectionClassLemma		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflectionClass		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflection	FEM_SG_DAT	SG_DAT	SUBJ_PRES_SG_3		PL_GEN	SG_GEN
lang	goh	goh	goh	goh	goh	goh
clause	CF					

AQL – Relationen Identische Textabdeckung =

	1	2	3	4	5	6
edition	dera	suazzi	si	kehlaufen	kepoto	cotes
ling	dera	suazzi	si	kehlaufen	kepoto	cotes
lemma	dër	suozī	sīn	giloufan	gibot	got
posLemma	DD	NA	VA	VV	NA	NA
pos	DDA	NA	VAFIN	VVPP	NA	NE
inflectionClassLemma		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflectionClass		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflection	FEM_SG_DAT	SG_DAT	SUBJ_PRES_SG_3		PL_GEN	SG_GEN
lang	goh	goh	goh	goh	goh	goh
clause	CF					

AQL – Relationen Überlappung _o_

	1	2	3	4	5	6
edition	dera	suazzi	si	kehlaufen	kepoto	cotes
ling	dera	suazzi	si	kehlaufen	kepoto	cotes
lemma	dër	suozī	sīn	giloufan	gibot	got
posLemma	DD	NA	VA	VV	NA	NA
pos	DDA	NA	VAFIN	VVPP	NA	NE
inflectionClassLemma		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflectionClass		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflection	FEM_SG_DAT	SG_DAT	SUBJ PRES_SG_3		PL_GEN	SG_GEN
lang	goh	goh	goh	goh	goh	goh
clause	CF					

AQL – Relationen Überlappung + links aligniert _|_

	1	2	3	4	5	6
edition	dera	suazzi	si	kehlaufen	kepoto	cotes
ling	dera	suazzi	si	kehlaufen	kepoto	cotes
lemma	dēr	suozī	sīn	giloufan	gibot	got
posLemma	DD	NA	VA	VV	NA	NA
pos	DDA	NA	VAFIN	VVPP	NA	NE
inflectionClassLemma		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflectionClass		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflection	FEM_SG_DAT	SG_DAT	SUBJ_PRESENT_SG_3		PL_GEN	SG_GEN
lang	goh	goh	goh	goh	goh	goh
clause	CF					

AQL – Relationen Überlappung + rechts aligniert _r_

	1	2	3	4	5	6
edition	dera	suazzi	si	kehlaufen	kepoto	cotes
ling	dera	suazzi	si	kehlaufen	kepoto	cotes
lemma	dër	suoʒī	sīn	giloufan	gibot	got
posLemma	DD	NA	VA	VV	NA	NA
pos	DDA	NA	VAFIN	VVPP	NA	NE
inflectionClassLemma		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflectionClass		IN_FEM	IRR	RED2	A_NEUT	A_MASC
inflection	FEM_SG_DAT	SG_DAT	SUBJ_PRESENT_SG_3		PL_GEN	SG_GEN
lang	goh	goh	goh	goh	goh	goh
clause	CF					

Hands-On-Suche

- ANNIS-Instanz der Korpuslinguistik der Humboldt-Universität zu Berlin
 - <https://korpling.org/annis3/>
- Wir zeigen an ein paar Beispielen einen Best-Practice-Vorschlag, wie man sich in ANNIS Suchanfragen für die eigenen Forschung erstellen kann.
- Wir nutzen hier die Klauselschreibweise und die verkürzte Schreibweise von AQL!

Referenzierung von Suchanfragen und Treffern in ANNIS

- wichtig für eigene Publikationen, Homepage oder Lehre
- Korpusreferenzlink:
 - Referenzierung eines bestimmten Korpus oder mehrerer Korpora in einer ANNIS-Instanz
- Suchreferenzlink:
 - Referenzierung der konkreten Suchanfrage in einem bestimmten Korpus oder in mehreren Korpora in einer ANNIS-Instanz
- Trefferreferenzlink:
 - Referenzierung eines Treffers einer konkreten Suchanfrage in einem bestimmten Korpus oder in mehreren Korpora in einer ANNIS-Instanz

Korpus- referenzlink

The screenshot displays the ANNIS interface. On the left, there is a search area with a text input field containing 'Please enter AQL query', a 'Query Builder' button, and a 'Search' button. Below this is a 'Corpus List' section with a 'Search Options' button and a 'Visible: All' dropdown. The corpus list table is as follows:

Name	Texts	Tokens		
DDD				
DDD-AD:Kl.Ahd.Denkmaele	42	9.258	ⓘ	📄
DDD-AD:Tatian	245	54.690	ⓘ	📄
DDD-AD:Tatian_Latein	245	47.997	ⓘ	📄
DDD-BenediktinerRegel	65	16.964	ⓘ	📄
DDD-Genesis	3	4.042	ⓘ	📄
DDD-Heliand	70	68.655	ⓘ	📄
DDD-Isidor	9	6.564	ⓘ	📄
DDD-IsidorLatein	10	5.403	ⓘ	📄
DDD-KIAhdDm	83	32.330	ⓘ	📄
DDD-Monsee	41	14.923	ⓘ	📄
DDD-MurbacherHymnen	27	3.138	ⓘ	📄
DDD-MurbacherHymnenL	27	4.319	ⓘ	📄
DDD-Orfrid	150	81.303	ⓘ	📄

The main area shows a detailed view for 'DDD-Heliand' (ID: 5967). It includes a 'Select corpus/document:' dropdown set to 'DDD-Heliand'. The central text area displays '(no metadata)'. On the right, there is a 'Available annotations' table:

Name	Example (click to use query)	URL
chapter	chapter="XI"	🔗
clause	clause="CF_U_M"	🔗
document	document="Hel_44_Heliand.XLIV.komplett"	🔗
edition	edition=","	🔗
inflection	inflection="SG_ACC"	🔗
inflectionClass	inflectionClass="A_MASC"	🔗
inflectionClassLen	inflectionClassLemma="A,O"	🔗
lang	lang="osx"	🔗
lemma	lemma="hē"	🔗
ling	ling=","	🔗
markup	markup="K"	🔗
pos	pos="NA"	🔗
posLemma	posLemma="NA"	🔗

At the bottom of the detailed view, a red circle highlights the 'Link to corpus: https://korpling.org/annis3/#_c=RERELUhibGihbmQ'.

Such- referenzlink

The screenshot shows the ANNIS interface with a search query `pos="VVFIN"` and 3689 matches in 70 documents. A table lists the corpus items, with **DDD-Heliand** selected. A search results view is open for the first result, showing the text `* Manega uuâron , the sia iro môd`. A red arrow points to a share icon in the top right of the results area. A dialog box titled "Query reference link" is open, displaying the URL `https://korpling.org/annis3/?id=6199a22e-7afa-4468-86cc-74f5842eb43e`, which is circled in red.

Name	Texts	Tokens		
DDD-AD:Kl.Ahd.Denkma	42	9.258		
DDD-AD:Tatian	245	54.690		
DDD-AD:Tatian_Latein	245	47.997		
DDD-BenediktinerRegel	65	16.964		
DDD-Genesis	3	4.042		
DDD-Heliand	70	68.655		
DDD-Isidor	9	6.564		
DDD-IsidorLatein	10	5.403		
DDD-KIAhdDm	83	32.330		

Treffer- referenzlink

The screenshot shows the ANNIS interface with the search query `pos="VVFIN"` entered in the search box. The search results show 3689 matches in 70 documents. The corpus list table is visible, with the following data:

Name	Texts	Tokens		
DDD-AD:Kl.Ahd.Denkma	42	9.258		
DDD-AD:Tatian	245	54.690		
DDD-AD:Tatian_Latein	245	47.997		
DDD-BenediktinerRegel	65	16.964		
DDD-Genesis	3	4.042		
DDD-Heliand	70	68.655		
DDD-Isidor	9	6.564		
DDD-IsidorLatein	10	5.403		
DDD-KlAhdDm	83	32.330		

The search results are displayed in a list view, showing the path, edition, and context for each match. A red arrow points to the reference link icon (a document with a double arrow) next to the first result, which is:

1 Path: DDD-Heliand > Hel_01_Heliand.I.komplett left context: 5 right context: 5
(edition 1 - 8)
* Māriā **uuâron** , the sia iro môd
⊞ edit
⊞ annotations

Treffer- referenzlink

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.

Select visualization

Display Name
kwic
edition
annotations

Link for publications

<https://korpling.org/annis3/?id=8d72e34e-5a8b-4de0-af1a-ab60c3b8f0a3>

Code for embedding visualization into web page

```
<iframe height="300px" width="100%" src="https://korpling.org/annis3/?id=8d72e34e-5a8b-4de0-af1a-ab60c3b8f0a3"></iframe>
```

Preview

Show in ANNIS search interface

edition	*	Manega	uuâron
ling	*	Manega	uuâron
lemma		manag	wesan

Close

Treffer- referenzlink

[Show in ANNIS search interface](#) 

edition	*	Manega	uuâron	,	the	sia	iro	môd
ling	*	Manega	uuâron	,	the	sia	iro	môd
lemma		manag	wesan		the	hě	hě	môd
posLemma		DI	VA	,\$	DD	PPER	PPER	NA
pos		DIS	VVFIN	,\$	DDSREL	PPER	DPOS	NA
inflectionClassLemma		A,O	ST5					A_MASC
inflectionClass		P	ST5					A_MASC
inflection		MASC_PL_NOM_ST	IND_PAST_PL_3		MASC_PL_ACC	MASC_PL_ACC_3	MASC_PL_GEN_3	SG_NOM
lang		osx	osx		osx	osx	osx	osx
clause		CF_U_M			CF_I_Rel			
document	Hel_01_Heliand.I.komplett							
chapter	I							
verse	1							
rhyme		S1						S2
translation		mancher	sein; dasein, existieren		der, die, das, welcher, welche, welches	er, sie, es	er, sie, es	Sinn, Inneres, Herz

Best Practice Arbeitsschritte

- Wie erarbeite ich mir Suchanfragen für die eigene Forschung?
 1. Forschungsfrage so konkret wie möglich stellen
 2. Variablen der Forschungsfrage identifizieren
 3. Korpus auswählen und Dokumentation anschauen
 4. Abbildung der benötigten Variablen auf die Annotationen im Korpus
 5. Formulierung einer Suchanfrage in ANNIS und Suche nach diesen Annotationen
 6. Treffer, Visualisierungen und Exporte prüfen



Beispiel 1

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- „Ich beschäftige mich z.Zt. mit Verschmelzungsformen aus Präposition und Artikel und mich würde daher konkret interessieren, wie solche Formen aufgefunden werden können, die eine Wortform darstellen, aber zwei Lemmata zuzuordnen sind.“

Beispiel 1

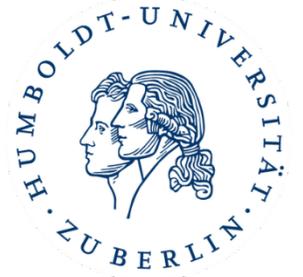
- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- „Ich beschäftige mich z.Zt. mit **Verschmelzungsformen** aus **Präposition** und **Artikel** und mich würde daher konkret interessieren, wie solche Formen aufgefunden werden können, die eine **Wortform** darstellen, aber zwei **Lemmata** zuzuordnen sind.“

Beispiel 1

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- Suche nach einem Korpus
 - in unserem Fall: Referenzkorpus Altdeutsch
 - (in der offenen Runde gern auch Fragen zu anderen Korpora!)
- Dokumentation des Referenzkorpus suchen
 - Gibt es eine Homepage, die Informationen über das Korpus auflistet?
 - <https://referenzkorpusaltdeutsch.wordpress.com>
 - <http://www.deutschdiachrondigital.de/home/projekt/>
 - Gibt es eine Publikation – Artikel/Guideline/Technischer Bericht?
 - z.B. Donhauser (2015), Dipper et al. (2013)
 - Ist das Korpus in einem Repository archiviert und dokumentiert?
 - LAUDATIO-Repository: Donhauser, Karin; Gippert, Jost; Lühr, Rosemarie; ddd-ad (Version 0.1), Humboldt-Universität zu Berlin. <https://referenzkorpusaltdeutsch.wordpress.com/>.
<http://hdl.handle.net/11022/0000-0000-7FC2-7>



DD Determinativ, definit/demonstrativ,
DDA Determinativ, definit, artikelartig
DDN Determinativ, definit/demonstrativ, nachgestellt
DDS Determinativ, definit/demonstrativ, substituierend
DDREL Determinativ, definit/ demonstrativ/relativ
DDSREL Determinativ, definit/demonstrativ, substituierend, relativ
DI Determinativ, indefinit, (mit und ohne Determiner)
DIA Determinativ, indefinit, artikelartig
DIN Determinativ, indefinit, nachgestellt
DW Determinativ, interrogativ
DWREL Determinativ, interrogativ, relativ
DWS Determinativ, interrogativ, substituierend
DWSREL Determinativ, interrogativ, substituierend, relativ
DWG Determinativ, (interrogativ,) generalisierend, attributiv
DWGREL Determinativ, (interrogativ,) generalisierend, relativ

DINEG Determinativ, indefinit, (mit und ohne Determiner), negativ
DINEGN Determinativ, indefinit, negativ, nachgestellt
DINEGS Determinativ, indefinit, negativ, substituierend
DIS Determinativ, indefinit, substituierend
DPOS Determinativ, possessiv
DPOSD Determinativ, possessiv, prädikativ oder adverbial
DPOSN Determinativ, possessiv, nachgestellt
DPOSS Determinativ, possessiv, substituierend
DPOS Sonderfall: Genitiv des PPER als DPOS

APPR Präposition; Zirkumposition links
APPO Postposition
APZR Zirkumposition rechts

Annotationen in der Ebene „pos“ für Wortarten
→ Tags für Determinierer und Präpositionen

Beispiel 1

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- „Ich beschäftige mich z.Zt. mit **Verschmelzungsformen** aus **Präposition** und **Artikel** und mich würde daher konkret interessieren, wie solche Formen aufgefunden werden können, die eine **Wortform** darstellen, aber zwei **Lemmata** zuzuordnen sind. “
- Annotationsebene für Wortarten vorhanden
 - „pos“
- Annotationsebene für Lemmata vorhanden
 - „lemma“
- Annotationsebene für die historischen Wortformen vorhanden
 - „edition“

Beispiel 1

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen

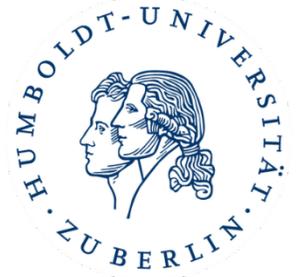


- Man möchte also etwas finden wie:

<i>zen</i>		<i>éwarton</i>
z	en	éwarton
Präposition	Artikel	Nomen
zu	d	Hohepriester

Beispiel 1

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



edition	<i>zen</i>		<i>éwarton</i>
(hypothese)	z	en	éwarton
pos	APPR #1	D... #2	...
lemma	zu	d	Hohepriester

pos=/APPR/ &

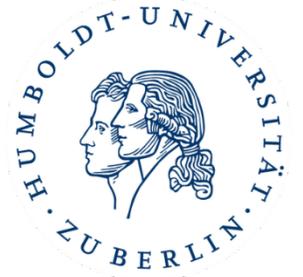
pos=/D.+/ &

#1 . #2

Suchreferenzlink: <https://korpling.org/annis3/?id=c1aad5e-5167-488a-9b76-61fad28b1b3e>

Beispiel 1

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



edition	<i>zen</i> #3		<i>éwarton</i>
(hypothese)	z	en	<i>éwarton</i>
pos	APPR #1	D... #2	...
lemma	zu	d	Hohepriester

pos=/APPR/ &

pos=/D.+/ &

edition &

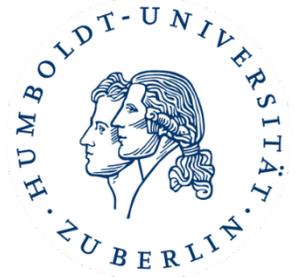
#1 . **#2** &

#3 _o_ **#1** & **#3** _o_ **#2**

Suchreferenzlink: <https://korpling.org/annis3/?id=f0de9053-8f3a-477f-b620-2704d27e6f1b>

Beispiel 1

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



edition	<i>zen</i> #3		<i>éwarton</i>
(hypothese)	z	en	<i>éwarton</i>
pos	APPR #1	D... #2	...
lemma	zu #4	d #5	Hohepriester

pos=/APPR/ &

pos=/D.+/ &

edition &

lemma & **lemma** &

#1 . **#2** &

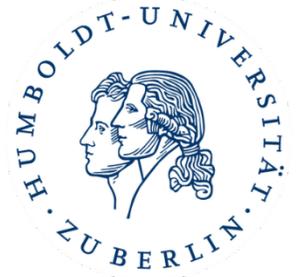
#3 _o_ **#1** & **#3** _o_ **#2** &

#4 _=_ **#1** & **#5** _=_ **#2**

Suchreferenzlink: <https://korpling.org/annis3/?id=a914d8d5-cd1c-4adb-8ccc-03902d658669>

Beispiel 1

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



[* Show in ANNIS search interface](#)

edition	Er	sprah	zen		ewarton	sélben	thesen	wórton	;
ling	Er	sprah	z	en	ewarton	sélben	thesen	wórton	;
lemma	ër	spréhhan	zi	dër	ēwawart,ēwawarto	sëlb	dëse	wort	
posLemma	PPER	VV	AP	DD	NA	DD	DD	NA	\$.
pos	PPER	VVFIN	APPR	DDA	NA	DD	DD	NA	\$.
inflectionClassLemma		ST4			A,N_MASC	A,O		A_NEUT	
inflectionClass		ST4			A,N_MASC	P		A_NEUT	
inflection	MASC_SG_NOM_3	IND_PAST_SG_3		MASC_PL_DAT	PL_DAT	NEUT_PL_DAT_ST	NEUT_PL_DAT	PL_DAT	
lang	goh	goh	goh	goh	goh	goh	goh	goh	
clause	CF_U_M								
document	O_Otfr.Ev.1.17								
chapter	17								
verse	35								
rhyme					E			E	
translation	er, sie, es	sprechen	zu	der, die, das	(Hohe)priester	selber, derselbe, selbst	dieser, diese, dieses	Wort; Rede	

Trefferreferenzlink: <https://korpling.org/annis3/?id=e6e5bb9b-d827-4525-b217-c66949bd43e2>

Beispiel 1

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- Export: csv (Auszug)

1_id	1_span	1_anno_elan:lang	1_anno_elan:lemma	1_anno_elan:ling	1_anno_elan:pos	1_anno_elan:posLemma	1_anno_elan:translation			
3532892	z	goh	zi	z	APPR	AP	zu			
2_id	2_span	2_anno_elan:inflection	2_anno_elan:inflectionClass	2_anno_elan:inflectionClassLemma	2_anno_elan:lang	2_anno_elan:lemma	2_anno_elan:ling	2_anno_elan:pos	2_anno_elan:posLemma	2_anno_elan:translation
3532894	en	MASC_PL_DAT	'NULL'	'NULL'	goh	dër	en	DDA	DD	der, die, das
3_id	3_span	3_anno_elan:edition								
3532891	zen	zen								

Beispiel 1

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



• Frequenzanalyse + csv-Export

#1 pos	#2 pos	#3 edition	#4 lemma	#5 lemma	count
APPR	DDA	zen	zi	dër	32
APPR	DDA	zemo	zi	dër	11
APPR	DPOS	iniro	in	ër	1
APPR	DIA	zeînéro	zi	ein	1
APPR	DD	indesemo	in	dëse	1
APPR	DIA	inei	in	ein	1
APPR	DDA	mitten	mit	dër	1
APPR	DPOS	zadinemo	zi	dīn	1
APPR	DI	zinoh	zi	noheinīg	1
APPR	DPOS	bira	bī	ër	1
APPR	DDA	zeru	zi	dër	1

Beispiel 1

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- Trefferkritik bzw. Kritik an der Suche

- Filterung nach Metadaten notwendig?
 - z.B. nur bestimmte Textdenkmäler

- Menge enthält teilweise Nomen in den Einheiten der Ebene *edition*, teilweise nicht.

<i>zererde</i>		
z	en	Erde
Präposition	Artikel	Nomen
zu	d	Erde

- Welche Art Determinierer ist gewünscht?
 - Eingrenzung der Tags

Best Practice Arbeitsschritte

- Wie erarbeite ich mir Suchanfragen für die eigene Forschung?
 1. Forschungsfrage so konkret wie möglich stellen
 2. Variablen der Forschungsfrage identifizieren
 3. Korpus auswählen und Dokumentation anschauen
 4. Abbildung der benötigten Variablen auf die Annotationen im Korpus
 5. Formulierung einer Suchanfrage in ANNIS und Suche nach diesen Annotationen
 6. Treffer, Visualisierungen und Exporte prüfen



Beispiel 2

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- Schwankungen in Deklinationsverhalten von Substantiven (st/sw), zwischen unterschiedlichen Dekl.-Klassen der st.Dekl., Distribution der Formenvarianten bei ein und derselben Wortform (z.B. Akk.Sg. man und mannan vom man 'mensch')

Beispiel 2

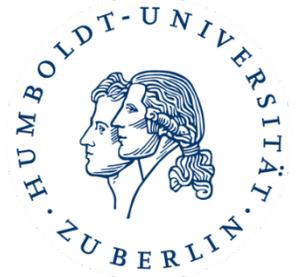
- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- Schwankungen in **Deklinationsverhalten** von **Substantiven (st/sw)**, zwischen unterschiedlichen **Dekl.-Klassen** der st.Dekl., **Distribution** der **Formenvarianten** bei ein und **derselben Wortform** (z.B. Akk.Sg. man und mannan vom man 'mensch')

Beispiel 2

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- Suche nach einem Korpus
 - in unserem Fall: Referenzkorpus Altdeutsch
 - (in der offenen Runde gern auch Fragen zu anderen Korpora!)
- Dokumentation des Referenzkorpus suchen
 - Gibt es eine Homepage, die Informationen über das Korpus auflistet?
 - <https://referenzkorpusaltdeutsch.wordpress.com>
 - <http://www.deutschdiachrondigital.de/home/projekt/>
 - Gibt es eine Publikation – Artikel/Guideline/Technischer Bericht?
 - z.B. Donhauser (2015), Dipper et al. (2013)
 - Ist das Korpus in einem Repository archiviert und dokumentiert?
 - LAUDATIO-Repository: Donhauser, Karin; Gippert, Jost; Lühr, Rosemarie; ddd-ad (Version 0.1), Humboldt-Universität zu Berlin. <https://referenzkorpusaltdeutsch.wordpress.com/>.
<http://hdl.handle.net/11022/0000-0000-7FC2-7>

Beispiel 2

- | | |
|-----------------------------|-----------------------------------|
| 1. Forschungsfrage | 4. Annotationen |
| 2. Variablen identifizieren | 5. Formulierung einer Suchanfrage |
| 3. Korpus + Dokumentation | 6. Ergebnis prüfen |



- Ebene für Wortarten, Flexionsklassen, Lemma und Edition sind vorhanden
- Man möchte eine Liste aller Substantive mit ihren Lemmata sowie mit den Informationen über ihre Flexionsklassen
 - Unterschiede in den Annotationen derselben Lemmata beobachten

Wort	herzun
Lemma	hërza
Wortart	Nomen
Flexionsklasse	N_NEUT

Beispiel 2

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



edition	herzun	#1
lemma	hërza	#2
pos	NA	#3
inflectionClass	N_NEUT	#4

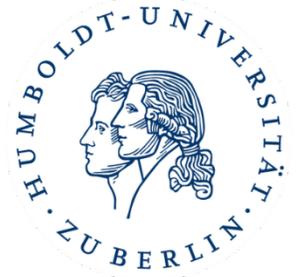
pos=/N./ &
lemma &
edition &
inflectionClass &
#1_=_**#2** & **#2**_=_**#3** & **#3**_=_**#4**

Es gibt zwei AQL-Schreibweisen:
Hier Klauselschreibweise

Suchreferenzlink: <https://korpling.org/annis3/?id=c689fe11-2348-4037-92e0-dd0a55052406>

Beispiel 2

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



edition	herzun	#1
lemma	hërza	#2
pos	NA	#3
inflectionClass	N_NEUT	#4

Es gibt zwei AQL-Schreibweisen:
Hier verkürzte Schreibweise

pos=/N./ _=_

lemma _=_

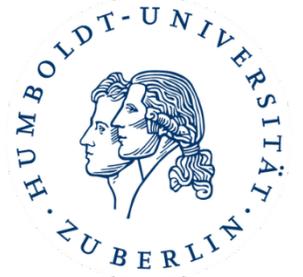
edition _=_

inflectionClass

Suchreferenzlink: <https://korpling.org/annis3/?id=129fca2f-9899-43aa-9b24-0ef8ab2ab1ca>

Beispiel 2

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



edition	herzun	#1
lemma	hërza	#2
pos	NA	#3
inflectionClass	N_NEUT	#4

pos=/N./ _=_

lemma _=_

edition _=_

inflectionClass

pos=/N./ &

lemma &

edition &

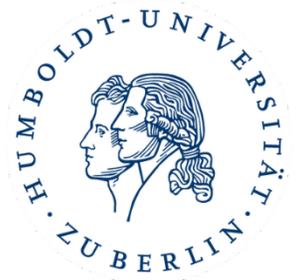
inflectionClass &

#1_=_#2 & #2_=_#3 & #3_=_#4

Beide liefern dieselben
Ergebnisse!
Beide können mit einander
kombiniert werden!

Beispiel 2

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



- Frequenzanalyse und csv-Export (Auszug)

#1 pos	#2 lemma	#3 edition	#4 inflectionClass	count
NA	annuzzi	annuzi	JA_NEUT	13
NA	annuzzi	annuzzi	JA_NEUT	3
NA	annuzzi	ánnuzze	A_NEUT	2
NA	annuzzi	annuze	A_NEUT	2
NA	annuzzi	ánnuzzi	JA_NEUT	2
NA	annuzzi	annuci	JA_NEUT	2
NA	annuzzi	ánluzzes	A_NEUT	1

Beispiel 2

1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



- Erweiterte Suchanfrage

edition &

pos=/N.+/ &

inflectionClassLemma=/.*,.* / &

lemma &

ling &

#1 _ _ #2 &

#1 _ _ #3 &

#1 _ _ #4 &

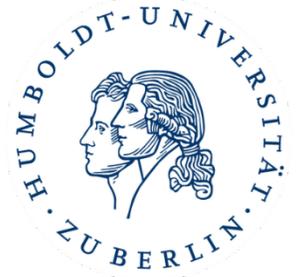
#1 _ _ #5

edition	manno	#1
lemma	man	#4
pos	NA	#2
inflectionClassLemma	C,A_MASC	#3
ling	manno	#5

Suchreferenzlink: <https://korpling.org/annis3/?id=355c5bb0-95a2-4e1d-8df4-4aef312c4d5c>

Beispiel 2

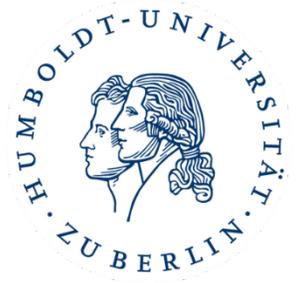
1. Forschungsfrage
2. Variablen identifizieren
3. Korpus + Dokumentation
4. Annotationen
5. Formulierung einer Suchanfrage
6. Ergebnis prüfen



- Schwankungen in **Deklinationsverhalten** von **Substantiven (st/sw)**, zwischen unterschiedlichen **Dekl.-Klassen** der st.Dekl., **Distribution** der **Formenvarianten** bei ein und **derselben Wortform** (z.B. Akk.Sg. man und mannan vom man 'mensch')
- Trefferkritik, bzw. Kritik an der Suche
 - Was genau meint Distribution? Welche Kontexte ggf. berücksichtigen?
 - Fehler im Korpus:

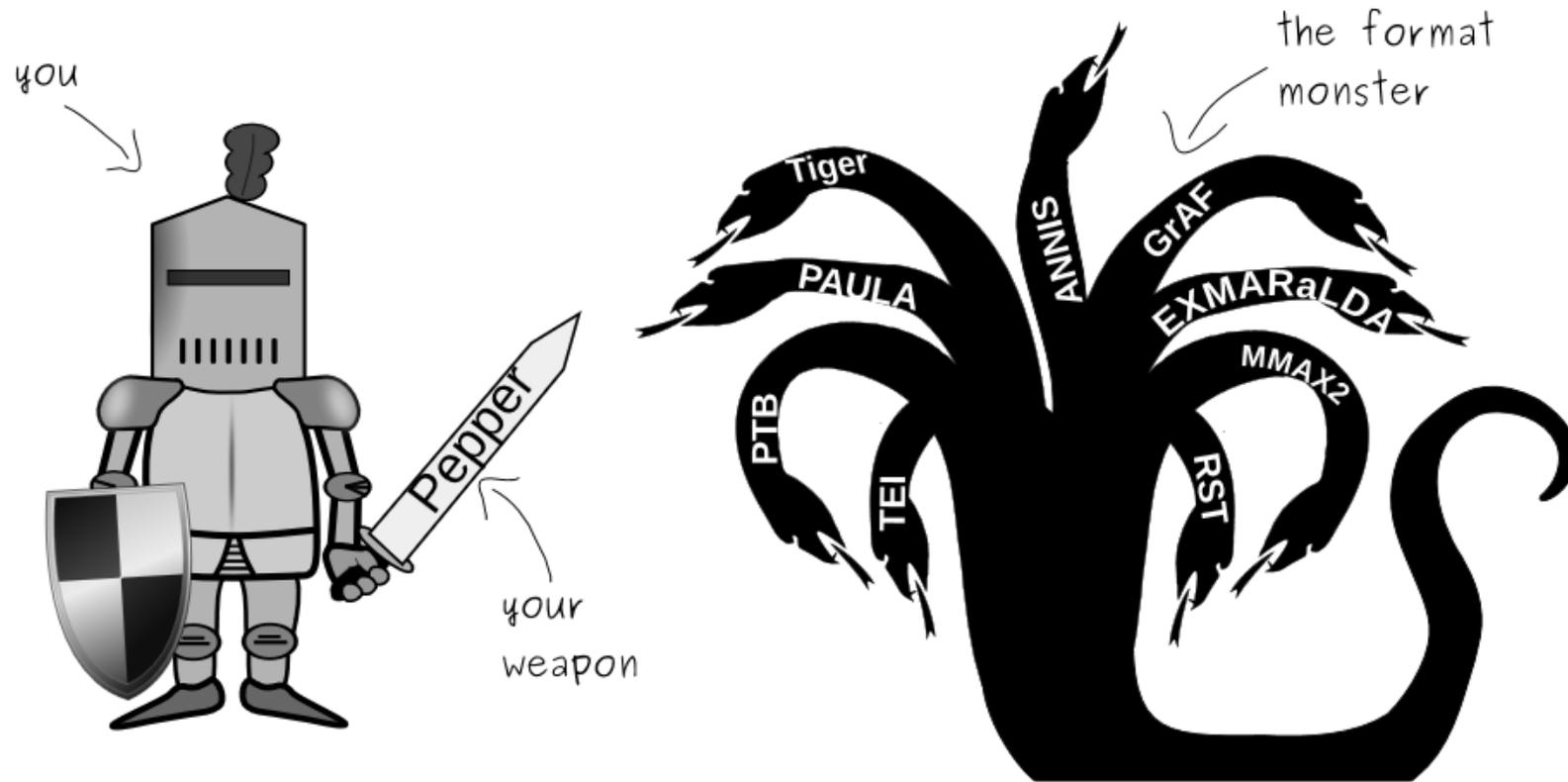
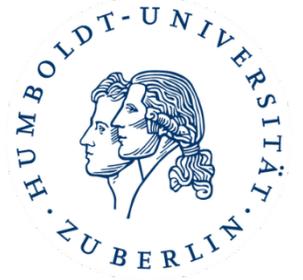
#1 pos	#2 lemma	#3 edition	#4 inflectionClass	count
NA	ābandmuos	abandmuase	A_NEUT	2
NA	ābandmuos	abandmuos	A_NEUT	1
NA	ābandmuos	abandmuos	I_NEUT	1
NA	ābandmuos	abandmúose	I_NEUT	1

IV. Exkurs Datenaufbereitung



Exkurs

Welche Formate nutzen Sie?



Exkurs

Wie bekomme ich mein Korpus nach ANNIS



- Wir nutzen das Konverterframework Pepper (Zipser & Romary 2010), um Korpora in ANNIS zu importieren
1. Überprüfen, ob mein Annotationsformat von Pepper unterstützt wird.
 - <http://corpus-tools.org/pepper/knownModules.html>
 - EXMARaLDA; ELAN, EXCEL etc.
 2. Konvertierung von eigenem Format in das ANNIS-Format durch Pepper
 - Download: <http://corpus-tools.org/pepper/#download>
 - Dokumentation: <http://corpus-tools.org/pepper/userGuide.html>
 3. Korpus im ANNIS-Format importieren
 - Kickstarter auf eigenem Rechner: <http://corpus-tools.org/annis/download.html>
 - Öffentliche ANNIS-Instanz: <http://corpus-tools.org/annis/#access>

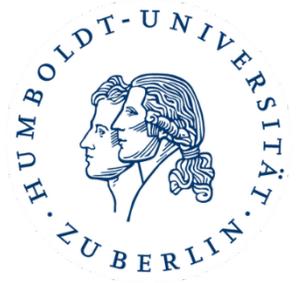
Exkurs

Wie bekomme ich mein Korpus nach ANNIS



- Haben Sie Interesse an einem Workshop speziell zur Datenaufbereitung?

V. Offene Diskussionsrunde



V. Offene Diskussionsrunde

- Themen in Vorbereitung
 - Verberstsätze
 - Satzmodus
 - Adjektivdeklinaton
 - Fragepartikel
 - Nebensätze ohne Konjunktion
 - Modus im Nebensatz
 - Verschmelzungsformen Präposition und Artikel
 - Univerbierung (obschon)
- Weitere Fragen?

Beispiel Verberstsätze I

- Suche nach Verberstsätzen, die mehr als 15 Wörter haben.

edition	<i>verb</i> (15)
pos	V.+
clause	Satz		

pos=/V.+/_|_ clause

- Spezifizieren, nur finite Verben? Welche Konstruktion ist konkret gesucht? Nur bei Matrixsätzen? Genaue Anzahl?

Suchreferenzlink: <https://korpling.org/annis3/?id=9d1c9d80-5574-41c9-bc98-1e867e1c48e3>

Beispiel Verberstsätze II

- Suche nach Verberstsätzen, die mehr als 15 Wörter haben.

edition	<i>verb</i> (15)
pos	V.+
clause	Satz		

pos=/V.+/_l_ clause

&

edition

&

#3_r_#2

&

#1 .15,20 #3

Suchreferenzlink: <https://korpling.org/annis3/?id=431fe7a6-a5b4-4710-b8be-1cd194bba5e6>

Beispiel Satzmodus I

Ist Satzmodus
annotiert?

▼ clause	CP participle clause, Partizipalsatz	Loc local clause, Lokalsatz
Ad adversative clause, Adversativsatz	CS suspended clause, Unterbrechung	M main clause, Hauptsatz
Adv adverbial clause, Adverbialsatz	Ex exclamation, Ausrufesatz	Mod modal clause, Modalsatz
Att attributive clause, Attributsatz	Fin final clause, Finalsatz	O object clause, Objektsatz
Caus causal clause, Kausalsatz	I introduction, eingeleiteter Satz	P predicative clause, Prädikativsatz
CC continued clause, Wiederaufnahme	ID dependent interrogative clause, abhängiger Fragesatz	S subject clause, Subjektsatz
CE elliptical clause, elliptischer Satz	Imp imperative clause, Imperativsatz	Temp temporal clause, Temporalsatz
CF finite clause, finiter Satz	Int interrogative clause, Fragesatz	U clause unintroduced, uneingeleiteter Satz
CI infinite clause, infiniter Satz		
Cnc concessive clause, Konzessivsatz		
Cnd conditional clause, Konditionalsatz		
Cns consecutive clause, Folgesatz		

Gewünschte Information ist in
anderen Informationen
eingebettet!

Beispiel Satzmodus II

Ist Satzmodus
annotiert?

z.B.

clause=/.+(Ex)|(Int)/

CP participle clause, Partizipalsatz
CS suspended clause, Unterbrechung
Ex exclamation, Ausrufesatz
Fin final clause, Finalsatz
I introduction, eingeleiteter Satz
ID dependent interrogative clause, abhängiger Fragesatz
Imp imperative clause, Imperativsatz
Int interrogative clause, Fragesatz

Suchreferenzlink:

<https://korpling.org/annis3/?id=c4f28008-0637-4f66-b640-85e9942ea118>

Gewünschte Information ist in
anderen Informationen
eingebettet!

Beispiel Adjektivdeklination I

- Verteilung der **starken** und **schwachen Adjektivdeklination**, mit und ohne **Artikel** und je nach Art des Artikels, auch in **Nominalisierungen**

des	guoten
(Artikel)	starkes oder schwaches Adjektiv, nominalisiert
der	guot

des	guoten	mannes
(Artikel)	starkes/ schwaches Adjektiv	(Nomen)
der	guot	man

- Ermittlung der Annotationswerte durch Suche eines Lieblingsworts
- gerade bei offenen Klassen wie Lemmata oder bei variierenden Werten sinnvoll
- z.B. **lemma**="guot"

Suchreferenzlink: <https://korpling.org/annis3/?id=52b7998d-f797-4970-99d8-3c6df8d4a69d>

Beispiel Adjektivdeklination II

- Verteilung der **starken** und **schwachen Adjektivdeklination**, mit und ohne **Artikel** und je nach Art des Artikels, auch in **Nominalisierungen**

des	guoten
(Artikel)	starkes oder schwaches Adjektiv, nominalisiert
der	guot

des	guoten	mannes
(Artikel)	starkes/ schwaches Adjektiv	(Nomen)
der	guot	man

Annotationsebene: inflection (variierende Werte!)

Tags: **ST|st** starke

WK|wk schwache

inflection = / . + ((ST) | (st) | (WK) | (wk)) /

Suchreferenzlink:

<https://korpling.org/annis3/?id=065992b0-a35b-4f6f-ad3f-115cb50798cc>

Beispiel Adjektivdeklination III

- Fall: starkes oder schwaches Adjektiv, das mit einem Artikel in einer prädikativen Struktur vorkommt

```

pos=/D.+/. pos=/ADJ.*/. pos=/N./
_=_ lemma &
inflection=/.+((WK)|(wk))/ &
lemma &
lemma &
#2 _=#6 &
#1 _=#7 &
#2 _=#5

```

edition	des	guoten	mannes
inflection	(Artikel)	starkes/ schwaches Adjektiv	(Nomen)
lemma	der	guot	man
pos	D...	ADJ	N.

Suchreferenzlink: <https://korpling.org/annis3/?id=a5dec780-2968-464d-9921-d7a36d3e5c99>

Beispiel Fragepartikel

Man kann auch nach **Lexemen** suchen, z.B. alle **Fragepartikeln** *inu* und *ia*, und die **Wortstellung in den Sätzen** untersuchen, die sie enthalten.

Inu	ne	larut	ir
<i>inu</i>	nī	lësan	ir
Partikel	Partikel	Verb	Pronomen
Satz			

Annotationsebenen
lemma, pos und clause

pos="PTKINT" _=_
lemma="inu" _|_
clause

- Präzision, wie man die Wortstellung untersuchen möchte
 - bestimmte Varianten als Hypothese überprüfen?
 - Soll Position von *inu* ermittelt werden oder die gesamte Wortreihenfolge im Satz?

Suchreferenzlink: <https://korpling.org/annis3/?id=fc301079-9db7-40ac-acee-a053a580249b>

Beispiel Nebensätze ohne Konjunktion

Nebensätze ohne Konjunktion finden.

- Annotationsebene clause enthält Informationen über Einleitung und Art der Sätze, additives Tagset
 - **U** clause unintroduced, uneingeleiteter Satz
 - **S** subject clause, Subjektsatz
 - **O** object clause, Objektsatz
 - **Adv** adverbial clause, Adverbialsatz

clause=/.+U.+(S|O|(ADV)|(Adv)).*/

- keine Satzsegmentierungsrichtlinien
- Annotation von unterbrochenen (CS) und fergesetzten Sätzen (CC)

Suchreferenzlink: <https://korpling.org/annis3/?id=7f2c7762-8292-404b-9ad3-23d43932561c>

Beispiel Modus im Nebensatz

- Suche nach Adverbialsätzen mit Indikativ
- Annotation von Sätzen und Flexion

edition	daz	kearneem	inan
inflection		SUBJ.+	...
clause	Satz		

clause=/.+((Adv)|(ADV)).*/ _o_ **inflection**=/((IND)|(Ind)).+/

- Suche nach Adverbialsätzen mit Konjunktiv I und II

clause=/.+((Adv)|(ADV)).*/ _o_ **inflection**=/SUBJ.+((PRES)|(Pres)|(PAST)|(Past)).+/

Suchreferenzlink: <https://korpling.org/annis3/?id=8c44f0ae-3ead-4dd3-9bc4-ad652263be93>

Suchreferenzlink: <https://korpling.org/annis3/?id=683b24ff-ece5-4e1e-8b08-d8ebe42e5a07>

Beispiel Entwicklung von Konzessivsätzen

- nhd. *ob* ≠ ahd. *ob*
- mögliche Varianten für Konzessivnebensatzeinleitungen:
 - ahd. *oba*, *ob* ‚oben, über... hinaus‘ **ahd. *ibu* ‚wenn‘**
- Suche nach Lemma *ibu* im **ahd**

lemma=/((ob)|(ibu))/

= lang="goh"

Suchreferenzlink: <https://korpling.org/annis3/?id=861904b2-f894-432e-ac78-ce11306679c9>

Beispiel Entwicklung von Konzessivsätzen

- nhd. *ob* ≠ ahd. *ob*
- mögliche Varianten für Konzessivnebensatzeinleitungen:
 - ahd. *oba*, *ob* ,oben, über... hinaus **ahd. *ibu* ,wenn'**
- Suche nach Lemma *ibu* als Nebensatzeinleitung

lemma=/((ob)|(ibu))/

= lang="goh,,

&

clause=/.+((Cnc)|(CNC))/ &

#1 _|_ #3

Suchreferenzlink: <https://korpling.org/annis3/?id=c86380bc-df49-42fe-b6d9-8c6c72046161>

Best Practice Arbeitsschritte

- Wie erarbeite ich mir Suchanfragen für die eigene Forschung?
 1. Forschungsfrage so konkret wie möglich stellen
 2. Variablen der Forschungsfrage identifizieren
 3. Korpus auswählen und Dokumentation anschauen
 4. Abbildung der benötigten Variablen auf die Annotationen im Korpus
 5. Formulierung einer Suchanfrage in ANNIS und Suche nach diesen Annotationen
 6. Treffer, Visualisierungen und Exporte prüfen



Hinweise

- Referenzkorpus:
 - Dokumentation der *inflectionClass* oder auch *clause* ist ohne weitere Kenntnisse über das Korpus unzureichend, die Projekthomepage gibt nicht genügend Informationen, in der Korpusdokumentation in LAUDATIO wurde nicht die Kombination der Tags beschrieben!
- Unterschied zwischen einer Annotationsguideline und einer Annotationsdokumentation
 - Was soll/darf/muss annotiert werden? vs. Was wurde annotiert? (Was kam tatsächlich im Korpus vor)
- Metadaten:
 - Metadatum: muss man nach Gruppen oder einzelnen Daten filtern;
 - Geht über die Anfrage selbst oder den Export
 - Beispielanfrage: <https://korpling.org/annis3/?id=50baf42d-7236-4d1d-a27d-3f5cd80a638f>

Weitere Hilfestellungen

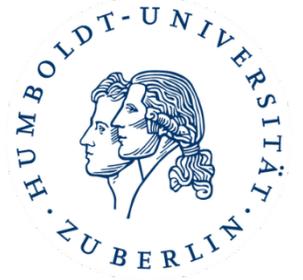
- ANNIS Demonstrationen und AQL-Einführungen
 - <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/lehre/folien-material>
 - mit verschiedenen Korpora und in unterschiedlichen Kontexten (Lehre, Vorträge, Workshops)
- ANNIS Video-Tutorials
 - <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/corpus-tools/annis-tutorials>
- ANNIS User Guide
 - [http://corpus-tools.org/annis/resources/ANNIS User Guide 3.4.3.pdf](http://corpus-tools.org/annis/resources/ANNIS%20User%20Guide%203.4.3.pdf)
- Beispielanfragen für RIDGES-Herbology
 - <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/ridges-projekt/access-de>

Fragen, Auswertung, Feedback



Wir freuen uns über Ihr Feedback, weitere Anregungen zur Schulung,
zum Referenzkorpus Altdeutsch und zu ANNIS!

Vielen Dank! 😊



Literatur

- Dipper, Stefanie/Donhauser, Karin/Klein, Thomas/Linde, Sonja/Müller, Stefan/Wegera, Klaus-Peter (2013). HiTS: Ein Tagset für historische Sprachstufen des Deutschen; in: Zinsmeister, Heike/Heid, Ulrich/Beck, Kathrin (Hg.): Das Stuttgart-Tübingen Tagset – Stand und Perspektiven, JLCL 28(1), 85-137.
- Donhauser, Karin (2015). Das Referenzkorpus Altdeutsch: Das Konzept, die Realisierung und die neuen Möglichkeiten; in: Gippert, Jost/Gehrke, Ralf (Hg.): Historical Corpora. Challenges and Perspectives (=CLIP 5). Tübingen.
- Krause, Thomas; Zeldes, Amir (2016): ANNIS3. A new architecture for generic corpus query and visualization. In: *Digital Scholarship in the Humanities* (31(1)), S. 118–139. Online verfügbar unter <http://dx.doi.org/10.1093/lc/fqu057>.
- Zipser, Florian; Romary, Laurent (2010). A model oriented approach to the mapping of annotation formats using standards. In: *Proceedings of the Workshop on Language Resource and Language Technology Standards, LREC 2010. Malta*. URL: <http://hal.archives-ouvertes.fr/inria-00527799/en/>