

ANNIS and the reference corpora of historical German

partly based on Perlitz & Odebrecht, ANNIS Einführung
RIDGES Herbology 4.1. https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/lehre/folienmaterial/2015_historische_korpora (CC-BY)

Historical reference corpora

REA (750-1050)

Reference Corpus of Old High (and Low) German
Referenzkorpus Altdeutsch
<https://www.deutschdiachrondigital.de/rea/>

REM (1050-1350)

Reference Corpus of Middle High German
Referenzkorpus Mittelhochdeutsch
<https://www.linguistics.rub.de/rem/>

REN (1200–1650)

Reference Corpus of Middle Low German / Low Rhenish - *Referenzkorpus Mittelniederdeutsch / Niederrheinisch*
<https://www.fdr.uni-hamburg.de/record/9195>

REF (1350-1650)

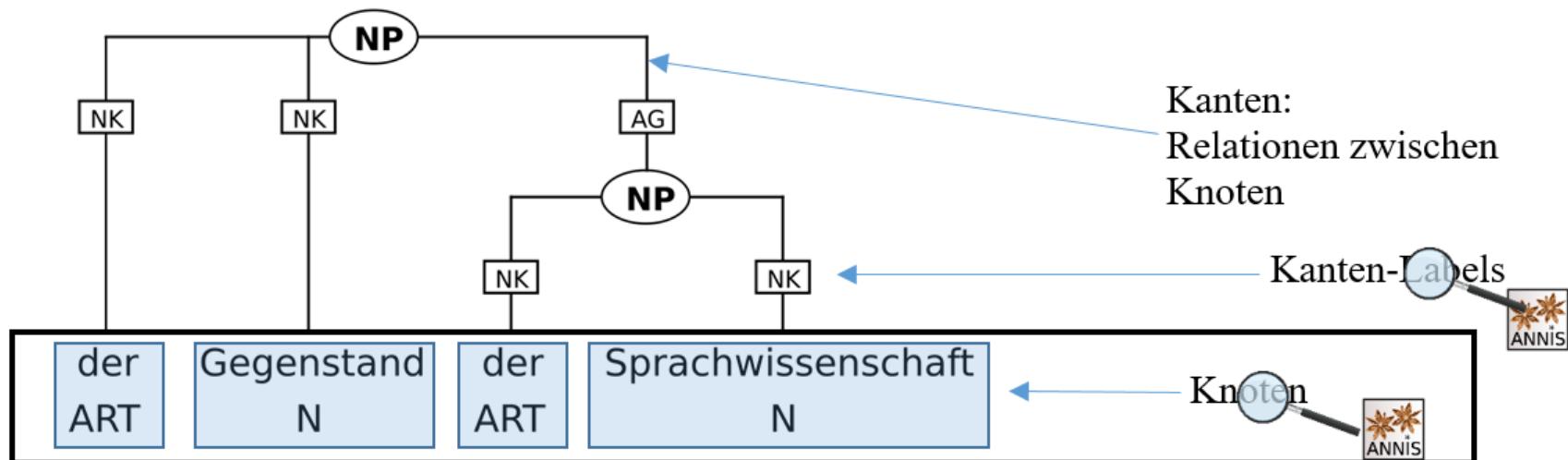
Reference Corpus of Early New High German
Referenzkorpus Frühneuhochdeutsch
<https://www.linguistics.rub.de/ref/>

ANNIS

- Another corpus query system...
- developed at HU Berlin
- for searching existing, annotated corpora (no inbuilt annotation function)
- some basic visualization possibilities

ANNIS Query Language

- can search for **nodes** and **edge labels**
- The latter are only available in **treebanks**, i.e. syntactically annotated corpora



ANNIS Query Language

- How do I tell ANNIS what I'm looking for?
- Basic principle: variable-value pairs

variable = value

e.g. **pos="NN"**

- and **relations** between variable-value pairs
 - pos="ART".pos="NN"** 'followed by'
 - lemma="Tanzen"=_ pos="NN"** 'congruent'



- web-based search tool for corpora
 - project website <http://annis-tools.org/>
 - Access to web-based ANNIS tool <https://korpling.german.hu-berlin.de/annis3>

➤ What can ANNIS do?

- finds annotations in a corpus
- finds multiple annotations in a specific relation to one another
- can visualize the annotations of the hits in multiple ways
- exports the hits
- etc.

ANNIS

Examples for corpora available via ANNIS:

- Referenzkorpus Altdeutsch (= Deutsch Diachron Digital; REA)
- Referenzkorpora Mittel- und Frühneuhochdeutsch (REM / REF)
- Referenzkorpus Mittelniederdeutsch /Niederrheinisch (REN)
- RIDGES (historical herbology texts)

ANNIS

Examples for synchronic corpora in ANNIS:

- Potsdamer Kommentarkorpus
(Zetungskommentare)
- KiezDeutschKorpus
- Corpus d'Études pour le Français Contemporain (CEFC)
- Georgetown University Multilayer Corpus (GUM)

Example corpus: RIDGES Herbology



- Herbology corpus
 - 15th-19th century
 - different dialects from the German-speaking area
 - annotated by ~50 students from three B.A./M.A. seminars
 - project website: http://korpling.german.hu-berlin.de/ridges/documentation_v4.1_en.html
- Annotation guidelines
 - http://korpling.german.hu-berlin.de/ridges/download/pubs/annotationGuidelines_v4.1.pdf

<https://korpling.german.hu-berlin.de/annis3-snapshot/>

The screenshot shows the ANNIS web interface. On the left, there is a search bar with placeholder text "Please enter AQL query" and a "Search" button. Below it is a "Corpus List" section with a table showing corpus details like name, texts, tokens, and edition. On the right, there is a "Help us to make ANNIS better!" sidebar with links to "Help/Examples", "Tutorial", and "Example Queries". The main content area displays a table of example queries with their descriptions and links to related resources.

Example Query	Description	
Q pos="NN"	Findet einen Nomen.	RIDGES Herbology Version4.1
Q lang="foreign"	search for foreign language (non-Hausa) segments within the Hausa text	a5.hausa.umarnin.uwa V2
Q pos="NPRO"		
Q norm="noy"		
Q pos="VSTA"		
Q norm="άπα"	search for the normalized word απα	apophthegmata.patrum.11
Q morph="μντ"	search for words containing the morpheme μντ	Besa.letters
Q pos="PPERI"	search for independent personal pronouns (e.g. ονο)	Besa.letters
Q norm="νκοτκ"	search for the normalized word νκοτκ	Besa.letters
Q edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-BenediktinerRegel
Q document	Access all texts in the corpus for reading purposes	DDD-BenediktinerRegel
Q edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-Genesis
Q document	Access all texts in the corpus for reading purposes	DDD-Genesis
Q edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-Heliand
Q document	Access all texts in the corpus for reading purposes	DDD-Heliand
Q edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-Isidor
Q document	Access all texts in the corpus for reading purposes	DDD-Isidor
Q edition="enti"	Search for the form "enti" in the transcription of the edition	DDD-IsidorLatein
Q document	Access all texts in the corpus for reading purposes	DDD-IsidorLatein



Interface

Start page, help, query window

Interface Start page



Please enter AQL query

Help us to make ANNIS better!

not logged in Login

About ANNIS Report Problem

Query Builder

Tutorial

Example Queries

Help/Examples

Example Query Description open corpus browser

[Q pos="NN"](#) Findet ein search for segments

[Q lang="foreign"](#) search for fo

[Q pos="NPROP"](#) search for fo

[Q norm="NOYR"](#) search for the normalized word NOYR

[Q pos="VSTAT"](#) search for stative verb forms

[Q norm="ANNA"](#) search for the normalized word ANNA

[Q morph="HNT"](#) search for words containing the morpheme HNT

[Q pos="PPERI"](#) search for independent personal pronouns (e.g. ANON)

[Q norm="NKOTK"](#) search for the normalized word NKOTK

[Q document](#) Search for the form "enti" in the transcription of the edition

[Q edition="enti"](#) Access all texts in the corpus for reading purposes

[Q document](#) Search for the form "enti" in the transcription of the edition

[Q edition="enti"](#) Access all texts in the corpus for reading purposes

[Q document](#) Search for the form "enti" in the transcription of the edition

[Q edition="enti"](#) Access all texts in the corpus for reading purposes

[Q document](#) Search for the form "enti" in the transcription of the edition

[Q edition="enti"](#) Access all texts in the corpus for reading purposes

[Q document](#) Search for the form "enti" in the transcription of the edition

[Q edition="enti"](#) Access all texts in the corpus for reading purposes

Welcome to ANNIS! A tutorial is available on the right side.

Corpus List Search Options

Visible: All

Filter

Name	Texts	Tokens	
a5.hausa.umarnin.uwa_\	47	10.194	i d
abraham.our.father	7	7.671	i d
apophthegmata.patrum.	11	1.500	i d
arabic.tree.test	1	11	i d
b2.hausa	50	6.991	i d
b4.tatian2.0	2.031	11.295	i d
b4.tatian2.1	2.030	11.295	i d
b7.wolof.web.V2	4	14.676	i d
b7.wolof.wiki.V4	14	12.737	i d
BeMaTaC I 1.2013-01	12	11.192	i d

Tutorial & example queries

Example queries for all corpora in ANNIS

List of all corpora in ANNIS

Interface Tutorial



Verschiedene Kapitel

Help us to make ANNIS better!

not logged in [Login](#)

Please enter AQL query

About ANNIS Report Problem

Help/Examples Tutorial Choose topic Print

ANNIS interface > ANNIS Query language >

Using the ANNIS interface and the results tab.

The Search Form

head _1_ pos=/V/ >dep[func="dobj"]

Search for Word Forms
Searching for Annotations
Searching using Regular Expressions
Searching for Trees
Searching for Pointing Relations
Exporting Results
Frequency Analysis
Complete List of Operators

Visible: All

Corpus List Search Options

Filter

Name	Texts	Tokens		
a5.hausa.umarnin.uwa_	47	10.194	i	d
abraham.our.father	7	7.671	i	d
apophthegmata.patrum	11	1.500	i	d
arabic.tree.test	1	11	i	d
b2.hausa	50	6.991	i	d
b4.tatian2.0	2.031	11.295	i	d
b4.tatian2.1	2.030	11.295	i	d
b7.wolof.web.V2	4	14.676	i	d
b7.wolof.wiki.V4	14	12.737	i	d
ReMaTaC_I1_2013-01	12	11.192	i	d

Nicht-aktiver Reiter

Example Queries

Interface

Query window



Please enter AQL query

Query can be built via the query builder

enter query, start with Cmd+Enter

Welcome to ANNIS! A tutorial is available on the right side.

Corpus List Search Options

Visible: All

Filter

Name	Texts	Tokens
a5.hausa.umarnin.uwa_	47	10.194
abraham.our.father	7	7.671
apophthegmata.patrum	11	1.500
arabic.tree.test	1	11
b2.hausa	50	6.991

corpus list
choose corpus by clicking – selected corpora highlighted in blue

Name	Texts	Tokens
Sta-D-Anselm	2	2.710
Sta-D-Kafka	2	10.388
Sta-D-TueBaDZ	2	10.832
Sta-D-Unicum	2	11.312
lamentsreden_Deuts	35	3.134.192
2	2	399
GES_Herbology_Vers	22	122.698
GES_Herbology_Vers	29	154.266
GES_Herbology_Vers	29	154.267
ges_Herbology_Versi	13	60.811



ANNIS Query Language

- **oder:** How do I tell ANNIS what I'm looking for?
- two basic principles:
 - Variable-value pairs
 - Relations between variable-value pairs

Tokens

- 1) Tokens are the **smallest (technical) unit** in a corpus.
- 2) A token usually, but not always, corresponds to one **orthographic word or punctuation mark!**
- 3) You can **search** for these items in ANNIS.

Token Token Token Token Token Token ...

dipI	wider	auff	ein	frîches	zerftoffenes	kraut	/	thue	es	wider	in
clean	wider	auff	ein	frisches	zerstossenes	kraut	/	thue	es	wider	in
norm	wieder	auf	ein	frisches	zersto�enes	Kraut	/	tue	es	wieder	in

Principle 1: Variable-value pair

- AQL
 - Name of annotation level → Variable dipl =
 - annotated category → value /kraut/
- We have to know which variables are available!
 - Corpus metadata
 - annotation guidelines of the corpus (e.g. for RIDGES
http://korpling.german.hu-berlin.de/ridges/download/pubs/annotationGuidelines_v4.1.pdf)

Principle I: Variable-value pairs

dipl= /kraut/



Variable

(Layer, Tier, Ebene ...)

Wert

(Wort, Lemma, Satz, Wortart ...)

- 1) Prerequisite: Existence of a level called „dipl“. (Metadata!)
- 2) Expected result: finding all attestations of this character string on the level „dipl“ in the selected corpus (or corpora)

dip l	wider	auff	ein	frîches	zerftoffenes	kraut	/	thue	es	wider	in
clean	wider	auff	ein	frisches	zerstossenes	kraut	/	thue	es	wider	in
norm	wieder	auf	ein	frisches	zersto�enes	Kraut	/	tue	es	wieder	in

Interface Query



The screenshot shows the ANNIS Query interface. On the left, a search bar contains the query `dip1=/kraut/`. Below it is a message: "Valid query, click on 'Search' to start searching." To the right of the search bar is a sidebar with links: Help/Example, Tutorial, Example Query, and a list of recent queries: `pos="NN"`, `norm="h"`, `pos="NN"`, and `pos="NN"`. At the bottom is a "Corpus List" table with the following data:

Name	Texts	Tokens	Actions
RIDGES_Herbology_Version 22	22	122.698	Info Edit
RIDGES_Herbology_Version 29	29	154.266	Info Edit
RIDGES_Herbology_Version 29	29	154.267	Info Edit
Ridges_Herbology_Version 13	13	60.811	Info Edit

Red arrows point from four callout boxes to specific elements:

- An arrow points from the text "query language AQL" to the search bar.
- An arrow points from the text "example queries for RIDGES corpus" to the sidebar's recent queries list.
- An arrow points from the text "shows if query is valid" to the validation message below the search bar.
- An arrow points from the text "for every query at least one corpus has to be selected" to the highlighted row in the corpus list table.

Interface Errors in the query



The screenshot shows the ANNIS interface with the following elements:

- Query Editor:** Displays the AQL query `dip1=/kraut`. A red arrow points from the text "AQL with errors!" to this line.
- Error Message:** Below the query editor, a message box contains the text "line 1:6 no viable alternative at input 'dip1=/kraut'". A red arrow points from the text "Feedback on the error" to this message.
- Toolbar:** Includes buttons for "About ANNIS", "Report Problem", "Help/Exam", "Tutorial", "Example Qu", "Search", "More", and "History".
- Corpus List:** Shows a list of available corpora:

Name	Texts	Tokens	Actions
RIDGES_Herbology_Version	22	122.698	[i] [e]
RIDGES_Herbology_Version	29	154.266	[i] [e]
RIDGES_Herbology_Version	29	154.267	[i] [e]
Ridges_Herbology_Version	13	60.811	[i] [e]

Variable-value pair dipl= /kraut/



Help us to make ANNIS better!

not logged in [Login](#)

dipl=/kraut/

Query Builder

Base text ▾

Displaying Results 1 - 10 of 145 Result for: dipl="kraut"

1 1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dip1 1905 left context: 5 right context: 5 - 1915)
Nembt 5,000000 M.b._ gedör̄s rets kraut / zerhacket es ein wenig

⊖ Transcription

dip1	Nembt	5,000000	M.b._	gedör̄s	rets	kraut	/	zerhacket	es	ein	wenig
clean	Nembt	5,000000	unknown	gedörretts		kraut	/	zerhacket	es	ein	wenig
norm	Nehmt	5,000000	unknown	gedörrettes		Kraut	/	zerhacket	es	ein	wenig

⊖ Lexical Annotation

⊖ Syntactical Annotation

⊖ Content Annotation

⊖ Graphical Annotation

⊖ All

2 1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dip1 2455 left context: 5 right context: 5 - 2465)
distillirte waffer vff ein frisches kraut / vnd distillir es wider

⊖ Transcription

⊖ Lexical Annotation

⊖ Syntactical Annotation

⊖ Content Annotation

⊖ Graphical Annotation

⊖ All

3 1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dip1 3858 left context: 5 right context: 5 - 3868)
wider auff ein frisches zerftoffenes kraut / thue es wider in

⊖ Transcription

⊖ Lexical Annotation

⊖ Syntactical Annotation

⊖ Content Annotation

Corpus List Search Options

Visible: All

Filter

Name	Texts	Tokens
HSJ-Baseler_Maesebuch	10	10.020	i	d
HSJ-Briefe	2	374	i	d
HSJ-Maese_Nissima	4	4.867	i	d
HSJ-Sheveth_Jehude	9	11.547	i	d
HSJ-Varia	11	22.918	i	d
KAJUK	8	119.420	i	d
kobaltL1v1.4	20	12.984	i	d
kobaltL2v1.4	51	33.368	i	d
Maerchenkorpus	211	295.880	i	d
Mercurius	2	187.423	i	d



Interface

Hits, visualization, metadata, search history

Interface hits



of hits

Key word in Context (KWIC)

Grouping of annotations

set context

Help us to make ANNOTATION and NAVIGATION better!

not logged in | Login

dipl=/kraut/

Query Builder

Base text

Displaying Results 1 - 10 of 145

Result for: dipl=/kraut/

Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 1905 left context: 5 right context: 5)
- 1915) Nembt 5,000000 M.b._ gedors rets kraut / zerhacket es ein wenig

Transcription

Lexical Annotation

Syntactical Annotation

Content Annotation

Graphical Annotation

All

Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 2455 left context: 5 right context: 5)
- 2465) distillirte waffer vff ein frisches kraut / vnd distillir es wider

Transcription

Lexical Annotation

Syntactical Annotation

Content Annotation

Graphical Annotation

All

Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 3858 left context: 5 right context: 5)
- 3868) wider auff ein frisches zerftoffenes kraut / thue es wider in

Transcription

Lexical Annotation

Syntactical Annotation

Content Annotation

Graphical Annotation

All

Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 3993 left context: 5 right context: 5)
- 4003)

Corpus List Search Options

Visible: All

ridges

Name	Texts	Tokens
RIDGES_Herbology_Version 22	122.698	
RIDGES_Herbology_Version 29	154.266	
RIDGES_Herbology_Version 29	154.267	
Ridges_Herbology_Version 13	60.811	

145 matches in 10 documents

Interface Visualization



Query # of hits

dipl=/Kraut/

Help/Examples Query Result

Base text ▾

Displaying Results 1 - 10 of 120

Result for: dipl=/Kraut/

Partitur

1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 2356 left context: 5 right context: 5)

ein Ich̄enes Saltz von dem Kraut / das du genommen haft

Transcription

dipl	ein	Ich̄enes	Saltz	von	dem	Kraut	/	das	du	genommen	haft
clean	ein	schönes	Saltz	von	dem	Kraut	/	das	du	genommen	hast
norm	ein	schönes	Salz	von	dem	Kraut	/	das	du	genommen	hast

Lexical Annotation

norm	ein	schönes	Salz	von	dem	Kraut	/	das	du	genommen	hast
pos	ART	ADJA	NN	APPR	ART	NN	\$(PRELS	PPER	VVPP	VAFIN
pos_klein	ART	ADJ	N	APPR	ART	N	ZEICHEN	REL	N	VINF	VFIN
lemma	ein	schön	Salz	von	d	Kraut	/	d	du	nehmen	haben

Syntactical Annotation
Content Annotation

Variables (Annotation levels)

Values (Annotation values)

dipl=/Vers/

1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 3690 left context: 5 right context: 5)

Results Per Page: 10

Order: normal

Left Context: 5

Right Context: 5

Show context in: dipl

Help us to make ANNIS better!

not logged in | Login

Query Builder

Search Options

Corpus List

120 matches in 17 documents

Transcription

Lexical Annotation

Syntactical Annotation

Content Annotation

Graphical Annotation

All

Interface Metadata



The screenshot shows the ANNIS interface with the following elements:

- Left Panel (Corpus Metadata):** A table titled "corpus metadata" showing statistics for different texts. The table has columns: Name, Texts, and Tokens.
- Right Panel (Document Metadata):** A list of search results for the query "dipl=/kraut/". Each result includes:
 - A path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 1905)
 - A snippet of text: Nembt 5000000 M.b._ gedors rets **kraut** / zerhacket es ein wenig
 - An annotation sidebar with options: Transcription, Lexical Annotation, Syntactical Annotation, Content Annotation, Graphical Annotation, and All.

Annotations in red:

- A red box highlights the "corpus metadata" table.
- A red arrow points from the "corpus metadata" table to the "document metadata + name of the document" sidebar.
- A red box highlights the "document metadata + name of the document" sidebar.

Name	Texts	Tokens
RIDGES_Herbology_Version 22	22	122.698
RIDGES_Herbology_Version 29	29	154.266
RIDGES_Herbology_Version 29	29	154.267
Ridges_Herbology_Version 13	13	60.811

Interface

Corpus metadata



Help us to make ANNIS better!

dip1=/kraut/

About ANNIS Report Problem

Query Builder

Search More History

145 matches in 10 documents

Corpus List

Visible: All

corpus metadata

Annotation layers (variables) of the corpus

Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dip1 1905) left context: 5 right context: 5

Corpus information for RIDGES_Herbology_Version4.1 (ID: 10158)

Metadata

Select corpus/document RIDGES_Herbology_Version4

Name Value

Homepage http://korpling.german.hu-berlin.de /ridges/index_en.html

Abed-Ali, Ilham; Andresen, Silke; Ast, Henriette; Belz, Malte; Christen, Doreen; Dayal, Mascha; Dittberner, Antonia; Döhn, Cora; Driemel, Ilmke; Efremova, Olja; Eichhorn, Gill-Maria; Esser, Judith; Gerlach, Annegret; Giese, Linda; Kiraga, Sebastian; Kolbik, Ewa Anna; Kovács, Kornél; Krüger, Daisy; Lehmann, Anna-Maria; Lober, Maria; Lueders, Laura; Lüdeling, Anke; Maniscalco, Samuele; Metzig, Manuel; Meyer, Alexander; Müller, Sandra; Müller, Vinzenz; Murphy, Andrew; Mursell, Johannes; Okuda, Akiko; Perlitz, Laura; Reinig, Katharina; Riesler, Ina; Rosin, Lena; Sachse, Franz-Josef; Sapronova, Anna; Sauer, Simon; Schmidt, Claudia; Sorokovska, Iryna; Springmann, Uwe; Stephan, Kristina; Tiemann, Julianne; Tóth, Anna; Tóth, Réka; Turtureanu, Alexander; Wekel, Juliana; Zuchewicz, Karolina et al.

RIDGES Project (Register In Diachronic German Science) funded by the

Available annotations

Node Annotations

Name Example (click to use query)

atLeast atLeast="1,000000"

atMost atMost="2,000000"

attr_gen attr_gen="gpost"

author_ref author_ref="pron1sg"

bemerkung bemerkung="grammatisch stellt es sich als verbalsubstantiv mit (idg.) ti-suffix zu got. siukan 'krank sein'."

brace brace="brRight"

brace_dir brace_dir="left"

clean clean="/"

Edge Annotations

Edge Types

Meta Annotations

Help/Examples Query Result

not logged in Login

/kraut/

List of all annotation layers (variables) of the corpus

general information about the corpus

is3-snapshot/# c=UkIER0VTX0hlcmJvbG9neV9WZXJzaW8=

Interface Corpus metadata II



clean="/" (arrow pointing to this field)

Help us to make ANNIS better!

not logged in Login

Query Result

Base text ▾

Displaying Results 1 - 10 of 145

Result for: dipl=/kraut

1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 1905 left context: 5 right context: 5)

Corpus information for RIDGES_Herbology_Version4.1 (ID: 10158) + x

Metadata

Select corpus/document: RIDGES_Herbology_Version4

Annotations

Name	Value
Homepage	http://korpling.german.hu-berlin.de/ridges/index_en.html
Abed-Ali, Ilham; Andresen, Silke; Ast, Henriette; Belz, Malte; Christen, Doreen; Dayal, Mascha; Dittberner, Antonia; Döhn, Cora; Driemel, Imke; Efremova, Olja; Eichhorn, Gill-Maria; Esse, Judith; Gerlach, Annegret; Giesel, Linda; Kiraga, Sebastian; Kolbik, Ewa Anna; Kovács, Kornél; Krüger, Daisy; Lehmann, Anna-Maria; Lober, Maria; Lueders, Laura; Lüdeling, Anke; Maniscalco, Samuel; Metzig, Manuel; Meyer, Alexander; Müller, Sandra; Müller, Vinzenz; Murphy, Andrew; Mursell, Johannes; Okuda, Akiko; Perlitz, Laura; Reining, Katharina; Riesler, Ina; Rosin, Lena; Sachse, Franz Josef; Sapronova, Anna; Sauer, Simon; Schmidt, Claudia; Sorokovska, Iryna; Springmann, Uwe; Stephan, Kristina; Tiemann, Julianne; Tóth, Anna; Tóth,	

Available annotations

Name	Example (click to use query)
atLeast	atLeast="1,000000"
atMost	atMost="2,000000"
attr_gen	attr_gen="gpost"
author_ref	author_ref="pron1sg"
bemerkung	bemerkung="grammatisch stellt es sich als verbalsubstantiv mit (idg.) ti-suffix zu got. siukan 'krank sein'. "
brace	brace="brRight"
brace_dir	brace_dir="left"
clean	clean="/" (highlighted)

2 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 1905 left context: 5 right context: 5)

3 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 3993 left context: 5 right context: 5)

4 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 3993 left context: 5 right context: 5)

Corpus List Search Options

Visible: All

ridges

Name	Texts	Tokens
RIDGES_Herbology_Version4.1	22	122.698
RIDGES_Herbology_Version4.1	29	154.266
RIDGES_Herbology_Version4.1	29	154.267
Ridges_Herbology_Version4.1	13	60.811

Click on an annotation level
-> is automatically pasted to the search field

Interface Document metadata



Help us to make ANNIS better!

not logged in [Login](#)

dip1=/Kraut/

Query Builder

Search More History

120 matches in 17 documents

Corpus List Search Options

Visible: All

Filter

Name	Texts	Tokens		
NoSta-D-Anselm	2	2.710	i	d
NoSta-D-Kafka	2	10.388	i	d
NoSta-D-TueBaDZ	2	10.832	i	d
NoSta-D-Unicum	2	11.312	i	d
Parlamentsreden_Deuts	35	3.134.192	i	d
pcc2	2	399	i	d
RIDGES_Herbology_Vers	22	122.698	i	d
RIDGES_Herbology_Vers	29	154.266	i	d
RIDGES_Herbology_Vers	29	154.267	i	d
Ridges_Herbology_Vers	13	60.811	i	d

Base text ▾

Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dip1 2356 left context: 5 right context: 5 - 2366)

1 ⓘ Metadata

document: AlchemistischePraktik_1603

Name Value

annis:doc AlchemistischePraktik_1603

default_ns:author Andreas Libavius

default_ns:bibl Libavius, Andreas (1603) Alchimistische Praktik. Frankfurt am Main. Johann Saurn. 4-26.

default_ns:date 1603

default_ns:pubPlace Frankfurt

default_ns:publisher Johann Saurn

default_ns:title Alchimistische Praktik

default_ns:version 4.1

2 ⓘ

3 ⓘ

4 ⓘ

Help/Examples Query Result

Displaying Results 1 - 10 of 120

Document metadata

Information on a document (text), e.g. author, year, title

603 (dip1 3498 left context: 5 right context: 5)

wider in

603 (dip1 3690 left context: 5 right context: 5)

603 (dip1 4123 left context: 5 right context: 5)

Interface

Search history



The screenshot shows the ANNOTATION and NAVIGATION interface. In the top right corner, there is a logo with two stylized flowers and the word "ANNIS". The main window has a toolbar at the top with "About ANNIS" and "Report Problem" buttons. Below the toolbar, a search bar contains the query "dipl=/kraut/". To the right of the search bar is a "Query Builder" button and a keyboard icon. On the far left, there is a "Corpus List" and "Search Options" section, and a "Visible" dropdown set to "All". The main content area displays a table of corpus entries:

Name	Text	Count	Details
RIDGES_Herbology_Version 22		154.266	View Edit
RIDGES_Herbology_Version 29		154.267	View Edit
Ridges_Herbology_Version 13		60.811	View Edit

To the right of the table, there is a "Help/Examples" section with a "Base text" dropdown set to "1 / 15". Below this, a "Search history" dropdown is open, showing a list of previous queries:

- dipl=/kraut/
- dipl="kraut"

A red arrow points from the "Search history" label to the dropdown. Another red arrow points from the "List of all queries in a session" label to the list of queries in the dropdown.

Interface

Query options



dipl=/Kraut/

120 matches
in 17 documents

Search Options

Left Context: 5

Right Context: 5

Show context in: dipl

Results Per Page: 10

Order: normal

set size of context

Which transcription level should
be shown in the list of hits?

Historical texts

- Variance in orthography and composition
- some things are quite "unpredictable"
- graphemic / (text)-structural information
 - line break, page break, page composition

Kräutern	Krautern	Alchimistische Praktik 1603
Kraut	Kraut	Alchimistische Praktik 1603
frau	kraut	Alchimistische Praktik 1603
Kreutern	Kreutern	Alchimistische Praktik 1603
Kreutter	Kreutter	Alchimistische Praktik
kreüter	kreüter	Fuchs New Kreüterbuch 1543
Kräuteren	Krauteren	Pflantz.Gart. 1639
Kreuter	Kreuter	Alchmistiche Praktik 1603
Kräuter	Kräuter	Deutsche Pflanzennamen 1870

Example

Graphemic variants



- Try to find all graphemic and inflectional variants of *Kraut* 'herb'!
 - use the fitting annotation layer (variable) for the query
 - most suitable variable would be **lemma**
 - and enter the value **/Kraut/**
- lemma=/Kraut/

Example Graphemic variants



Screenshot of the ANNOTATION and NLP INTEGRATION SYSTEM (ANNIS) interface. The search bar contains the query "lemma=/Kraut/". The results pane shows three examples of matches, each with a detailed view of its context and annotations.

Results 1 - 10 of 448

Result for: lemma=/Kraut/

1 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 889 - left context: 5 right context: 5)

weifz vnd weg / aufz krautern / bletern / wurz tzeln

(+ Transcription
+ Lexical Annotation
+ Syntactical Annotation
+ Content Annotation
+ Graphical Annotation
+ All)

2 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 922 - left context: 5 right context: 5)

weg / viel waffer aufz Krautern vff einmal zu distilliren .

(+ Transcription
+ Lexical Annotation
+ Syntactical Annotation
+ Content Annotation
+ Graphical Annotation
+ All)

3 Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 970 - left context: 5 right context: 5)

/ denn nimb was fur Krauter du wilt / zerftoz die

(+ Transcription
+ Lexical Annotation
+ Syntactical Annotation
+ Content Annotation
+ All)

Corpus List Search Options

Visible: All

Filter

Name	Texts	Tokens		
a5.hausa.umarnin.uwa_v	47	10.194		
abraham.our.father	7	7.671		
apophthegmata.patrum.	11	1.500		
arabic.tree.test	1	11		
b2.hausa	50	6.991		
b4.tatian2.0	2.031	11.295		
b4.tatian2.1	2.030	11.295		

Normalization

- Take a closer look at the hits for lemma=/Kraut/

weifz vnd weg / aufz **kräutern** / blettern / wurz tzeln

⊖ Transcription

dipI	weifz	vnd	weg	/	aufz	kräutern	/	blettern	/	wurz	tzeln
clean	weisz	vnd	weg	/	ausz	kräutern	/	blettern	/	wurtzeln	
norm	weiß	und	weg	/	aus	Kräutern	/	Blättern	/	Wurzeln	

- Our query finds all historical word forms of *Kraut* attested in the corpus.



Example: Normalization

- Search for the lemma of *zusammensetzen!*
 - We find very different historical spelling variants...
 - lemma=/zusammensetzen/
 - Hit 1
erften grade / vnd ley züfammen gefetzt aufz widerwertiger substantz . das
 - Hit 2
, die aus drei Worten zusammen gefetzt sind , wie die oben
 - Hit 3
muſz aus zwei einfachen Worten zuſammengeſetzt ſein , von denen der

Example: Normalization

- Stepwise normalization (dipl>clean>norm)
 - dipl: Unification/merging of historical spelling variants
 - POS and lemma annotation on the normalized level (norm)

3 ⓘ Path: RIDGES_Herbology_Version4.1 > DeutschePflanzennamen_1870 (dipl) left context: 5 right co
5444 - 5455

muſz aus zwei einfachen Worten zuſammengeſetzt fein , von denen der

⊖ Transcription

dipl	muſz	aus	zwei	einfachen	Worten	zuſammengeſetzt	fein	,	von	denen	der
clean	musz	aus	zwei	einfachen	Worten	zusammengesetzt	sein	,	von	denen	der
norm	muss	aus	zwei	einfachen	Worten	zusammengesetzt	sein	,	von	denen	der

⊖ Lexical Annotation

norm	muss	aus	zwei	einfachen	Worten	zusammengesetzt	sein	,		von	
pos	VMFIN	APPR	CARD	ADJA	NN	VVPP	VAINF	\$,		APPR	
pos_klein	VFIN	APPR	CARD	ADJ	N	VINF	VINF	ZEICHEN		APPR	
lemma	müssen	aus	zwei	einfach	Wort	zusammensetzen	sein	,		von	



Operators for pattern search

- any character
- ? 0 or 1 repetition (of preceding element)
- * 0-infinitely many rep. (of preceding element)
- + 1-infinitely many rep. (of preceding element)
- \ verbatim (following element)
- ! not
- (a|b) a or b (also: [ab])

Important operators for pattern search



- .
 - ?
 - *
 - +
 - \ \
 - !
 - (a|b)
 - [abc]
 - a{2,3}
- any character
0 or 1 repetition (of preceding element)
0-infinitely many rep. (of preceding element)
1-infinitely many rep. (of preceding element)
verbatim (following element)
not
a or b (also: [ab])
set (or [^abc]=all except abc)
a 2-3times

dip1=/ .*/
= pos=/PRELS/

important relations between VV pairs



- . direct precedence
- .^{*} indirect precedence
- _=_ identical coverage
- _i_ inclusion
- _o_ overlap
- _l_ left overlap
- _r_ right overlap

```
dip1=/.*/  
=/  
pos=/PRELS/
```



Example: Operators

- Which results do you expect for the following pattern searchers? norm=/g.b./
 - *gebe, gibt* (for RIDGES, others are theoretically possible)
 - dipl=/r(a|o)t/
 - *rot, rat* (for RIDGES)
 - dipl=/meint?/
 - *mein, meint* (for RIDGES)

Example

Part of speech



- How can we find all common nouns (not proper nouns) in RIDGES?
 - relevant variable is **pos**
 - pos=/NN/

Example

Part of speech



- How can we find all adjectives in RIDGES?
 - What do the guidelines say (STTS)?
 - relevant variable here **pos**
 - STTS: difference between ADJA and ADJD
 - ADJA: attributive adjective, *die schöne Katze* 'the beautiful cat'
 - ADJD: predicative or adverbial adjectives, e.g. *er fährt schnell* 'he drives fast', *er ist schnell* 'he is fast'
- pos=/ADJ./

Example: Querying annotation layers



- Find out if there is a dedicated layer for headers in RIDGES!
 - Can we build a query that searches for headers?
 - relevant variable is **head**
- `head=/head/`



Example: Combining searches

- Find the diplomatic word form *Wermut* occurring in headers!
 - relevant variables **head** and **dpl**
 - **head=/head/**
 - **dpl=/Wermut/**
- What's happening?**

Aufgabe

Kombinierte Suche



Help us to make ANNIS better!

not logged in [Login](#)

[About ANNIS](#) [Report Problem](#)

[Help/Examples](#) [Export](#) [Frequency Analysis](#)

[Tutorial](#)

[Example Queries](#)

error 😊

line 2:0-15 mismatched input 'dipl'
expecting {<EOF>, '&', '|'}

Corpus List Search Options
Visible: All
Filter

Name	Texts	Tokens		
a5.hausa.umarnin.uwa_\	47	10.194	i	d
abraham.our.father	7	7.671	i	d
apophthegmata.patrum.	11	1.500	i	d
arabic.tree.test	1	11	i	d
b2.hausa	50	6.991	i	d
b4.tatian2.0	2.031	11.295	i	d
b4.tatian2.1	2.030	11.295	i	d

Findet einen Nomen.

[Q norm="hat" & dipl="hats" & #2_o_#1](#) Findet eine bestimmte normierte Wortform "hat", die sich in einer klitischen Verbindung im diplomatischen Transkript befindet.

[Q pos="NN" & pos=/V.*/ & #1.#2](#) Findet einen Nomen und ein Verb mit den möglichen Annotationen /VV(FIN|INF)/, /VAFIN(FIN|INF)/ oder /VM(FIN|INF)/.

[Q pos="NN" & pos=/V.*/ & pb="pb" & #1.#2 &](#) Findet einen Nomen gefolgt von einem Verb. Nomen und Verb befinden sich auf derselben Seite des Faksimiles.

[open corpus browser](#)

[RIDGES Herbology Version4.1](#)

[RIDGES Herbology Version4.1](#)

[RIDGES Herbology Version4.1](#)

[RIDGES Herbology Version4.1](#)

Red arrows point from the error message box to the search bar and the history dropdown.

Principle II: Relations

head=/head/

VV pair1

O

Relation

dpl=/Wermut/

VV pair2

- 1) There are **several types of annotations!**
- 2) We have to know which **relation** these annotations can have to one another.
- 3) **Spans** (head) overlap with **token annotations** (dpl)!

XVIII. Saturninische Kräuter . Von ḥ Kräutern ihrer Natur

⊖ Transcription

dpl	XVIII.	Saturninische	Kräuter	.	Von	ḥ	Kräutern	ihrer	Natur
clean	XVIII.	Saturninische	Kräuter	.	Von	ḥ	Kräutern	ihrer	Natur
norm	XVIII.	Saturninische	Kräuter	.	Von	?	Kräutern	ihrer	Natur
head	head								

Aufgabe

Kombinierte Suche



Help us to make ANNIS better!

not logged in [Login](#)

Query Builder

head=/head/
o
dipl=/Wermut/

Search More History

2 matches in 1 document

Corpus List Search Options

Visible: All

Name	Texts	Tokens	Annotations
a5.hausa.umarnin.uwa_\	47	10.194	i d
abraham.our.father	7	7.671	i d
apophthegmata.patrum.	11	1.500	i d

Base text ▾

Displaying Results 1 - 2 of 2 Result for: head=/head/_o_dipl=/Wermut/

1 [i](#) Path: RIDGES_Herbology_Version4.1 > SonderbaresKraeuterbuch-1-11_1675 left context: 5 right context: 5
(dipl 1500 - 1514)

an ftatt des ZittwerSamens . **Abfinthium Vulgare** , Wermut . 1,000000 NAhmen . Wermuth /

[Transcription](#)
[Lexical Annotation](#)
[Syntactical Annotation](#)
[Content Annotation](#)
[Graphical Annotation](#)
[All](#)

2 [i](#) Path: RIDGES_Herbology_Version4.1 > SonderbaresKraeuterbuch-1-11_1675 left context: 5 right context: 5
(dipl 1927 - 1942)

Bruchen in böfen Fiebern . **Abfinthium Ponticum** , Welfcher Wermut . 1,000000 NAmen . Pontischer Wermuth

[Transcription](#)
[Lexical Annotation](#)
[Syntactical Annotation](#)
[Content Annotation](#)
[Graphical Annotation](#)
[All](#)



Syntax-Highlighting

- one color per VV pair - some examples:
 - Red for all values of the variable **head**
 - Purple for all values of the variable **dipI**
`head=/head/
dipI=/oWermut/`
- The hits are marked in the same colors:
 - multiple tokens are red, as they occur in a header together
 - one token is purple because it is the dipI value we were searching for: `Ablinthium Vulgare , Wermut .`



Example: Overlaps

- Search for a verb that occurs in a header.
 - relevant variables: **pos** and **head**
 - Operator _o_
- pos=/V.*/ _o_ head=/head/



Example: Identity

- Look for all diplomatic word forms tagged as substituting relative pronouns.
 - relevant variables: **dipI** and **pos**
 - Operator _=_
- dipI=/.*/ _=_ pos=/PRELS/



Example: Direct precedence

- Search for an article directly preceding a noun.
 - relevant variable: **pos**
 - Operator .
➤ pos=/ART/ . pos=/N./



Example: Direct precedence

Help us to make ANNIS better!

not logged in [Login](#)

Base text ▾ 1 / 916 Displaying Results 1 - 10 of 9158 Result for: pos=/ART/ . pos=/N./

1 ⓘ Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 1 - 8) left context: 5 right context: 5

An den Lefer . ES findet mir lieber

⊕ Transcription

⊖ Lexical Annotation

norm	An	den	Leser	.	Es	sind	mir	lieber
pos	APPR	ART	NN	\$.	PPER	VAFIN	PPER	ADJD
pos_klein	APPR	ART	N	ZEICHEN	N	VFIN	N	ADJ
lemma	an	d	Leser	.	es	sein	ich	lieb

⊕ Syntactical Annotation

⊕ Content Annotation

⊕ Graphical Annotation

⊖ All

2 ⓘ Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 68 - 79)

findet . Ich hab mir das judicium Athenienium von Hippocrate belieben lassen

⊕ Transcription

⊖ Lexical Annotation

⊕ Syntactical Annotation

⊕ Content Annotation

⊕ Graphical Annotation

⊖ All

3 ⓘ Path: RIDGES_Herbology_Version4.1 > AlchemistischePraktik_1603 (dipl 115 - 126)

zu Gottes ehr / und defz Nechsten sonderlich Deutcher Nation . nutz

Query Builder

Search More ▾ History ▾

9158 matches in 29 documents

Corpus List Search Options

Visible: All

Filter

Name	Texts	Tokens	
a5.hausa.umarnin.uwa_V	47	10.194	ⓘ Ⓜ
abraham.our.father	7	7.671	ⓘ Ⓜ
apophthegmata.patrum.	11	1.500	ⓘ Ⓜ
arabic.tree.test	1	11	ⓘ Ⓜ
b2.hausa	50	6.991	ⓘ Ⓜ
b4.tatian2.0	2.031	11.295	ⓘ Ⓜ
b4.tatian2.1	2.030	11.295	ⓘ Ⓜ



Example: indirect precedence

- Search for an article (**A**) that **indirectly precedes** a noun (**B**). Why indirectly? You also want to catch a potential prenominal modifier (**C**) in your results: *die schöne Katze.*
 - relevant variable: **pos**
 - Operator **.** and distance **1,2** (read: I search **A** and **B** either in direct precedence or one token **C** can be in-between)
- pos=/ART/ .**1,2** pos=/N./



Example: Sequences

- Search for a noun directly followed by a preposition which in turn is directly followed by an adjective.
 - relevant variable: **pos**
 - Operator .
➤ pos=/N./ . pos=/APPR/ . pos=/ADJ./

Example: Sequences

- Search for a noun directly followed by a preposition directly followed by an adjective.
 - relevant variable **pos**
 - Operator .
- **pos=/N./ . pos=/APPR/ . pos=/ADJ./**

/ zwey feine Alchy^s miftische Tractälein von guten freundten zukomēn / wiewol vnbewuft

⊕ Transcription

⊖ Lexical Annotation

norm	/	zwei	feine	Alchimistische	Traktälein	von	guten	Freunden	zukommen
pos	\$()	CARD	ADJA	ADJA	NN	APPR	ADJA	NN	VVINF

Example: Identity

- Search for particle verbs containing the particle *auf* (e.g. *aufwachen* 'wake up')
 - relevant variables: **lemma** and **pos**
 - Operator _=_
- pos=/VV.*/ _=_ lemma=/auf.*/

Example: Identity

- Search for particle verbs containing the particle *auf* (e.g. *aufwachen* 'wake up')
 - relevant variables: **lemma** and **pos**
 - Operator _=_
- pos=/VV.*/ _=_ lemma=/auf.*/
- What do we **not** find using this query?



Example: Metadata

- Find out if the normalized word form *Blume* can be found in a text from the year 1543.
- If yes, how often?
 - relevant variables **norm** and **meta::date**
 - Verknüpfungszeichen &
 - norm=/Blume/ & meta::date=/1543/
 - 2 matches



Hints

- The list of hits only shows the annotation layers (variables) that have been annotated in the span that is shown. In other words, there might be more annotation layers that are not displayed because there are no annotations for the span currently displayed.
- Take a look at corpus metadata and annotation guidelines to know which annotation layers are actually in the corpus.
- Not all documents in a corpus necessarily have the same (number and type of) annotation layers.



Summing up

- Querying ANNIS is based on
 - **Variables** (Annotation layers) and **values** (categories in the annotation layers), e.g.
 - Search for **exact values**, e.g. pos=/ADJA/
 - Search for **patterns**, e.g. pos=/ADJ./
 - variable-value pairs can be searched for in **relation** to one another, e.g.
 - Annotations can **overlap**
 - Annotations cover an **identical area**
 - Variable-value pairs can be searched for together with **metadata**, e.g.:
 - pos=/ADJA/ & meta::date=/1870/



Interface

Export, Frequency analysis

Interface Exporting hits



The screenshot shows the ANNIS interface with the following elements:

- Query Bar:** Shows the query `dip1=/kraut/`.
- Search Results:** Displays "145 matches in 10 documents".
- Export Dialog:** Shows settings for "WekaExporter" with "Left Context" and "Right Context" both set to 5.
- Annotations:** A red box highlights the "More" button in the search options and the "Export" section of the dialog, with the text "see 'More'" pointing to it.
- Help Box:** A red box contains the text "Choose type of export, format, context, annotation layers" with an arrow pointing to the "Parameters" section of the dialog.
- Table:** A table titled "Corpus List" shows document statistics:

Name	Texts	Tokens	Annotations
RIDGES_Herbology_Version3.0	22	122.698	1
RIDGES_Herbology_Version4.0	29	154.266	1
RIDGES_Herbology_Version4.1	29	154.267	1
Ridges_Herbology_Version_2.0	13	60.811	1

Interface Exporting hits



The screenshot shows the ANNOTATION NAVIGATOR application window. In the top right corner, there is a logo with two golden star-like shapes and the text "ANNIS". The main interface has several tabs at the top: "Help/Examples" (selected), "Query Result", and "Export". The "Export" tab contains fields for "Exporter" (set to "WekaExporter"), "Left Context" (set to 5), "Right Context" (set to 5), and "Annotation Keys" (empty). Below these are "Parameters" and "Perform Export" and "Download" buttons. A red arrow points from a "help" button in a tooltip to the "Annotation Keys" field. The left side of the window shows a search bar with the query "dip1=/kraut/", a search history, and a message indicating 145 matches in 10 documents. The bottom section displays a table titled "Corpus List" with rows for "ridges" and various corpus entries. A tooltip for "Annotation Keys" provides information about limiting exported data to specific annotation keys.

dip1=/kraut/

Help us to make ANNIS better!

About ANNIS Report Problem

Help/Examples Query Result Export

Exporter WekaExporter ?

Left Context 5

Right Context 5

Annotation Keys ?

Parameters ?

Perform Export Download

Search More History

145 matches in 10 documents

Corpus List Search Options

Visible: All

ridges

Name	Texts	Tokens	
RIDGES_Herbology_Version3.0	22	122.698	i
RIDGES_Herbology_Version4.0	29	154.266	i
RIDGES_Herbology_Version4.1	29	154.267	i
Ridges_Herbology_Version_2.0	13	60.811	i

Help for "Annotation Keys"
(Click here to close)

Some exporters will use this comma separated list of annotation keys to limit the exported data to these annotations.

help

Interface Exporting hits



Text-Exporter

```
0. of the International Brotherhood of [Magicians] Wednesday , October 9
1. Magic Month in the United [States] . Wikinews spoke with William
2. of the International Brotherhood of [Magicians] , about the current st
3. - " Scarne on Card [Tricks] " and " Scarne on
4. and " Scarne on Magic [Tricks] " . That started me
```

Grid-Exporter

0.	tok	of the International Brotherhood of Magicians Wednesday
	pos	IN[1-1] DT[2-2] NP[3-3] NP[4-4] IN[5-5] NPS[6-6] NP[7-7]
	cat	S[1-6] VP[1-6] NP[1-6] PP[1-6] NP[2-4] PP[5-6] NP[6-6] NP[7-12]

CSV-Exporter

```
'11318611','the current state','NP','11318616','current','AJ0','current','JJ'
'11318686','magic','NP','11318688','magic','AJ0','magic','JJ'
'11318757','some basic tricks','NP','11318760','basic','AJ0','basic','JJ'
```

Interface Frequency Analysis



Query has already been formulated

selected corpora:
RIDGES_Herbology_Version4.1

query to analyze:
dipl=.*/_=_pos=/PRELS/

Node number/name	Selected annotation of node	Comment
1	dipl	automatically created from dipl=.*/_=_
2	pos	automatically created from pos=/PRELS/_=_

Example: finding all word forms (dipl) annotated as PRELS (pos)

Corpus List Search Options

Visible: All

ridges

Name	Texts	Tokens
RIDGES_Herbology_Version3.0	22	122.698
RIDGES_Herbology_Version4.0	29	154.266
RIDGES_Herbology_Version4.1	29	154.267
Ridges_Herbology_Version_2.0	13	60.811

Metadata Select

Add Delete selected row(s)

Automatic mode

Perform frequency analysis

go to „More“

start of frequency analysis

The screenshot shows the ANNIS interface with a query for frequency analysis and a corpus list. Red annotations highlight the 'More' button, the start of frequency analysis, and the perform analysis button.

Annotations:

- Query has already been formulated
- selected corpora: RIDGES_Herbology_Version4.1
- query to analyze: dipl=.*/_=_pos=/PRELS/
- Table of selected annotations:

Node number/name	Selected annotation of node	Comment
1	dipl	automatically created from dipl=.*/_=_
2	pos	automatically created from pos=/PRELS/_=_

- Example: finding all word forms (dipl) annotated as PRELS (pos)
- Corpus List
- Visible: All
- ridges
- Table of corpora:

Name	Texts	Tokens
RIDGES_Herbology_Version3.0	22	122.698
RIDGES_Herbology_Version4.0	29	154.266
RIDGES_Herbology_Version4.1	29	154.267
Ridges_Herbology_Version_2.0	13	60.811

- Metadata
- Select
- Add
- Delete selected row(s)
- Automatic mode
- Perform frequency analysis
- go to „More“
- start of frequency analysis

REM

REM: Choosing a corpus

Corpus List		Search Options		
Visible:		<input type="text"/> ▼ ⟳		
Filter		All MiGraKo eReM		
11-12_1- obd -PV-X		55	20,457	📄
11-12_1-rhfrhess-PV-X	5	956	ℹ️ 📄	
11_2-12_1-obd-PV-G	7	44,986	ℹ️ 📄	
12-13_1-mdnd-PV-X	4	36,281	ℹ️ 📄	
12-13_1-mdnd-V-X	5	33,992	ℹ️ 📄	
12_2-alem-PV-G	4	30,309	ℹ️ 📄	

- MiGraKo: Corpus of MHD Grammar
- eReM: Additional texts

Find individual tokens

- very easy – but not always helpful, see

"so"

"mann"

"vrouw" – **keine Treffer**

Lemma search

- Lemmas in REM are MHG lemmas (as in MHD dictionary)
- in case of doubt, check mhdwb.de to search for the right lemma (or use the query builder)

Layers in REM

Path: 12_2-alem-PV-G > M065-G1 (tok_dipl 371 - 381) left context: 5 right context: 5

. der vbel hellevarte . ern vverde ir geverte . des

annotations

	6a,8	6a,9				6a,10				
reference		der	vbel	hellevar	te	ern	vverde	ir	geverte	.
tok_dipl	.	der	vbel	hellevar	te	er	n	vverde	ir	g
tok_anno	.	der	vbel	hellevar	te	er	ne	vverde	ir	g
norm		der	übel	hellewarte		er	ne	vverde	ire	g
tokenization						MS1	MS2			
pos	\$_	DDART	ADJA	NA	\$_	PPER	PTKNEG	VAFIN	DPOSA	N
posLemma	\$_	DD	ADJ	NA	\$_	PPER	PTK	VA	DPOS	N
lemma		dér	übel	helle-warte		ér	ne	wérden	ir(e)	g
lemmald		29817000	174642000	71433000		40380000	119841000	225138000	83982000	5
lemmaLemma		dér	übel	helle-warte		ér	ne	wérden	ir(e)	g
inflection		Masc.Nom.Sg	Pos.Masc.Nom.Sg.0	Nom.Sg		Masc.Nom.Sg.3	--	Subj.Pres.Sg.3	Masc.Nom.Sg.0	N
inflectionClass	--	--	*.Masc			--	--	st3b	--	w
inflectionClassLemma	--	--	*.Masc			--	--	st3b	--	w
punc				S*						D

Example: Negation in MHG

- We want to investigate the development of negation in MHG:
- "double" negation with *en* and *niht* vs. simple negation
- What do we have to look for?

Example: Conversion in MHD

- We want to find all nominalized infinitives in MHG (e.g. NHG *das Singen, das Tanzen*)
- We can use the pos and posLemma annotation for that purpose.

Finding individual tokens

- Pay attention for spelling variants (uppercase/lowercase!)
- *so* only finds lowercase attestations, *So* only uppercase ones
- For finding all variants, we use regular expressions

Finding individual tokens

"so" | "So" | "SO"

Finding individual lemmas

- Important: Attributes can have different names!
- Example: The layer that is called "lemma" in virtually all other corpora is called "hw" in the British National Corpus.

From result to export file

The screenshot shows the ANNIS interface with the following elements:

- Search Bar:** Contains the query `tok=/[jJ]e/ . pos=/A.*/ .2,10 tok=/[dD]esto|[jJ]e|[Uu]mso/`.
- Results Summary:** Shows 39 matches in 36 documents.
- Buttons:** Search, More, History, Export, Frequency Analysis.
- Visible Corpus List:** FalkoEssayL2WHIGv2.0 (195), FalkoGeorgetownL2v1.0 (92), FalkoSummaryL1v1.2 (57), FalkoSummaryL2v1.2 (106), FalkoWHIGL2v2.1 (196).
- Help/Example:** A tooltip for the 'More' button provides an example query: `tok=/[jJ]e/ . pos=/A.*/ .2,10 tok=/[dD]esto|[jJ]e|[Uu]mso/`.
- Base text:** Displays the first few matches from the search results.
- Export Options:** Buttons for ZHverb (grid), falko (grid), and ZH1 (grid) are visible.
- Contextual Information:** Left context: 5, right context: 5.
- Table:** A table at the bottom lists corpus details: FalkoEssayL2WHIGv2.0 (195), FalkoGeorgetownL2v1.0 (92), FalkoSummaryL1v1.2 (57), FalkoSummaryL2v1.2 (106), FalkoWHIGL2v2.1 (196). The last row shows FNHD context: 5, 2674.

From result to export file

The screenshot shows a software interface for managing and searching a corpus. On the left, a search results panel displays two captured query strings:

```
tok=/[jJ]e/ . pos=/A.*/ .2,10
tok=/[dD]esto|[jJ]e|[Uu]mso/
```

Below these are search controls: "Search", "More", and "History". A message indicates "39 matches in 36 documents". At the bottom, there are tabs for "Corpus List" and "Search Options", and a "Visible" dropdown set to "All".

On the right, a "Query Result" panel is open, showing export configuration options:

- Exporter: GridExporter
- Left Context: 20
- Right Context: 20
- Annotation Keys: tok, pos
- Parameters: (empty)

Buttons include "Perform Export" (disabled), "Cancel Export", and "Download". A status message above the "Perform Export" button says: "The Grid Export each annotation To display only". A blue circular progress indicator is visible below the "Perform Export" button.

Export

- not all exporters work for all ANNIS corpora 😞

The screenshot shows the ANNIS web interface. On the left, there is a search query builder with the query `n="sein" .* POS=/AD./*`. Below it, a results summary says "3078 matches in 154 documents". At the bottom, there's a corpus list and a filter table.

In the center, the main interface has tabs for "Help/Examples", "Export", and "Query Result". The "Export" tab is active. A red arrow points from the text below to the "Exporter" dropdown menu, which is set to "GridExporter".

The "GridExporter" documentation explains:

The Grid Exporter can export all annotations of a search result and its context. Each annotation layer is represented in a separate line, and the tokens covered by each annotation are given as number ranges after each annotation in brackets. To suppress token numbers, input numbers=false into the parameters box below. To display only a subset of annotations in any order use the "Annotation keys" text field, input e.g. "tok,pos,cat" to show tokens and the annotations pos and cat.

Below this, there are sections for "Parameters" and "Annotation Keys".

At the bottom of the export dialog, there are buttons for "Perform Export", "Cancel Export", and "Download".

At the very bottom, it says "exported 350 items in 35,08 s".

Name	Texts	Tokens		
mo	82	146,704		
mu	191	346,039		

Using conditional formatting in Excel

Screenshot of an Excel spreadsheet showing the 'Formatvorlage' dialog box for creating a new conditional format rule.

The dialog box is titled "Neue Formatvorlagentyp" (New Conditional Format Rule Type) and shows the following settings:

- Formatvorlage:** Klassisch (Classic)
- Formel für die Ermittlung der zu formatierenden Zellen verw...** (Formula to determine which cells to format): `=Istleer(A1)`
- Formatieren mit:** Hellrote Füllung mit dunkler...
- Color:** Red fill with dark red text.

The background of the spreadsheet shows several rows of text data, some of which are highlighted with red color due to the conditional formatting rule applied to column A.

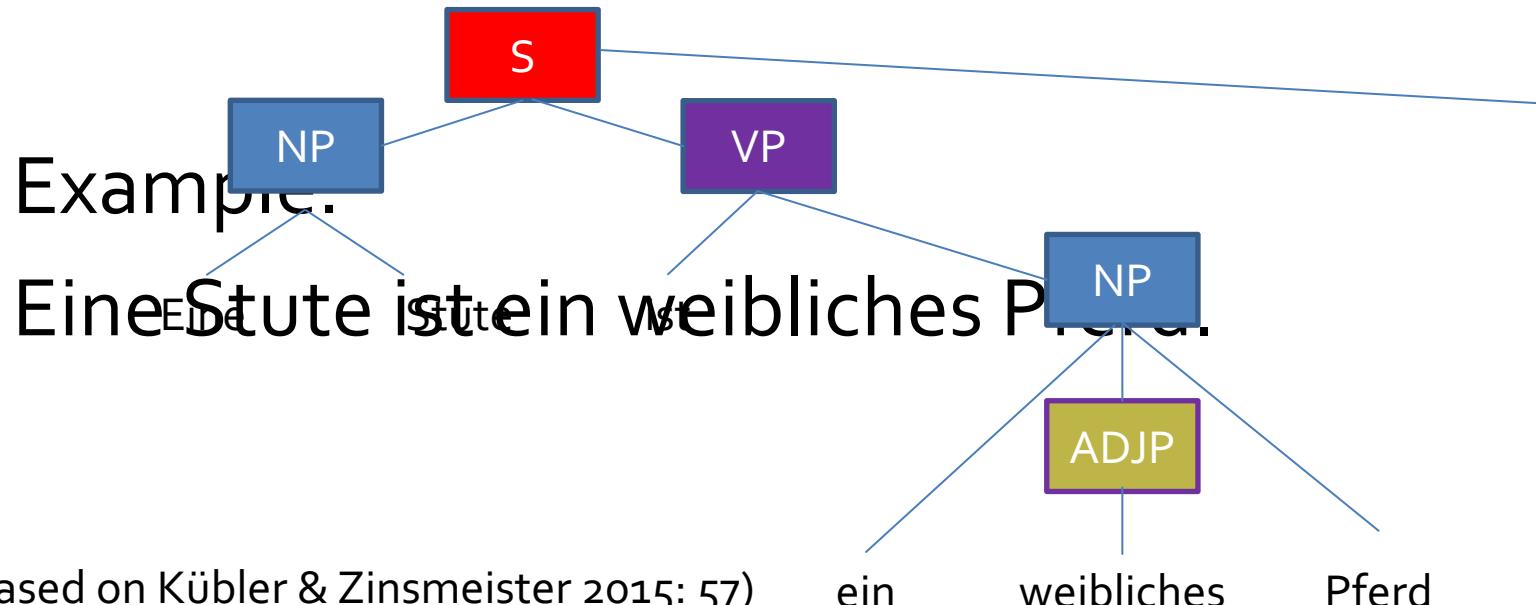
A	B	C	D
0 tok	von Kriminalität ist nur kurzfris...		
pos	APPR[1-1] NN[2-2] VAFIN[3-3]		
1 tok	ja keine Veränderungen , und		
pos	ADV[1-1] PIAT[2-2] NN[3-3] \$,)		
2 tok	eigenen Büchern und Lernque...		
pos	ADJA[1-1] NN[2-2] KON[3-3] N		
3 tok	die Gesellschaft entsprechen . Aber was heißt eigentlich " entsprechen " ? Man könnte zuerst denken , dass je mehr man arbeitet , desto mehr Geld bekommt man . :		
pos	ART[1-1] NN[2-2] VVINF[3-3] \$.[4-4] KON[5-5] PWS[6-6] VVFIN[7-7] ADV[8-8] \$.[9-9] VVINF[10-10] \$.[11-11] \$.[12-12] PIS[13-13] VMFIN[14-14] ADV[15-15] VVINF[16-16]		
4 tok	Beitrag für die Gesellschaft , wenn man daran denkt : " je nützlicher man ist , desto besser muss man bezahlt werden " . Das kommt man aber zu einer tieferen		
pos	ADV[1-1] NN[2-2] APPR[3-3] ART[4-4] NN[5-5] \$.[6-6] KOUS[7-7] PIS[8-8] PROAV[9-9] VVFIN[10-10] \$.[11-11] \$.[12-20] ADV[21-21] ADJD[22-22] PIS[23-23] VAFIN[24-24]		
5 tok	meisten Fällen) noch weniger verdienen als die Frauen . So , hier sehe ich eigentlich das Problem : je stärker die Frauen werden , je schwächer werden die Männer . De		
pos	PIAT[1-1] NN[2-2] \$.[3-3] ADV[4-4] ADV[5-5] VVFIN[6-6] KOKOM[7-7] ART[8-8] NN[9-9] \$.[10-10] ADV[11-11] \$.[12-12] ADV[13-13] VVFIN[14-14] PPER[15-15] ADV[16-16]		
6 tok	am meisten aus Geld zusammensetzt . Anders gesagt ist das Prinzip dieser Welt nichts anderes als die finanzielle Entlohnung . Je grösser ist das Lohn , desto wirklicher i		
pos	APPRART[1-1] PIS[2-2] APPR[3-3] NN[4-4] VVFIN[5-5] \$.[6-6] ADV[7-7] VVPP[8-8] VAFIN[9-9] ART[10-10] NN[11-11] PDAT[12-12] NN[13-13] PIS[14-14] PIS[15-15] KOKOM		

Treebanks

- syntactically parsed corpora
- syntax represented via syntactic trees
- a syntactic tree consists of **edges** and **nodes**
- widespread tagset for German: Tagset of the TIGER treebank (Albert et al. 2004)

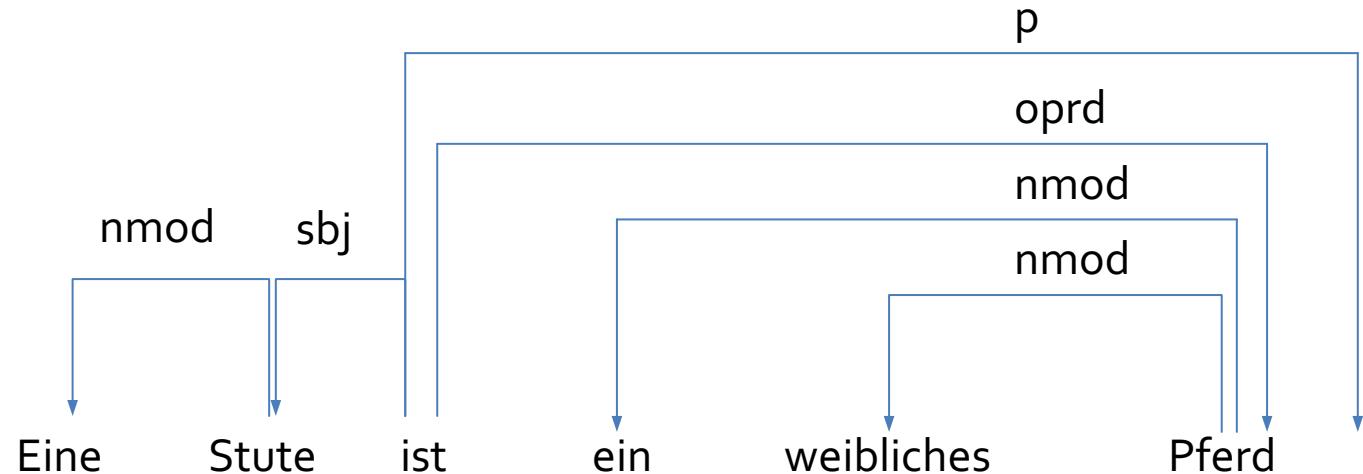
Constituent vs. dependency structure

- two key syntactic formalisms
- Constituent structures: Words are grouped into **phrases**, with one word functioning as **head** in each phrase.



Constituent vs. dependency structure

- Dependency structure: Relationships between word pairs (**head and dependent**)



Constituent vs. dependency structure

- both analyses assume hierarchical structuring of sentences
- but: in constituent structure analysis constituents (i.e. abstract entities) are ordered hierarchically
- dependency analysis, by contrast, is constrained to the words themselves.
- Constituent structure represents syntactic categories, dependency structure takes syntactic functions into account.

Hybrid models

- some projects use a mixture of both variants
- advantage: being able to take both syntactic categories and syntactic functions into account.

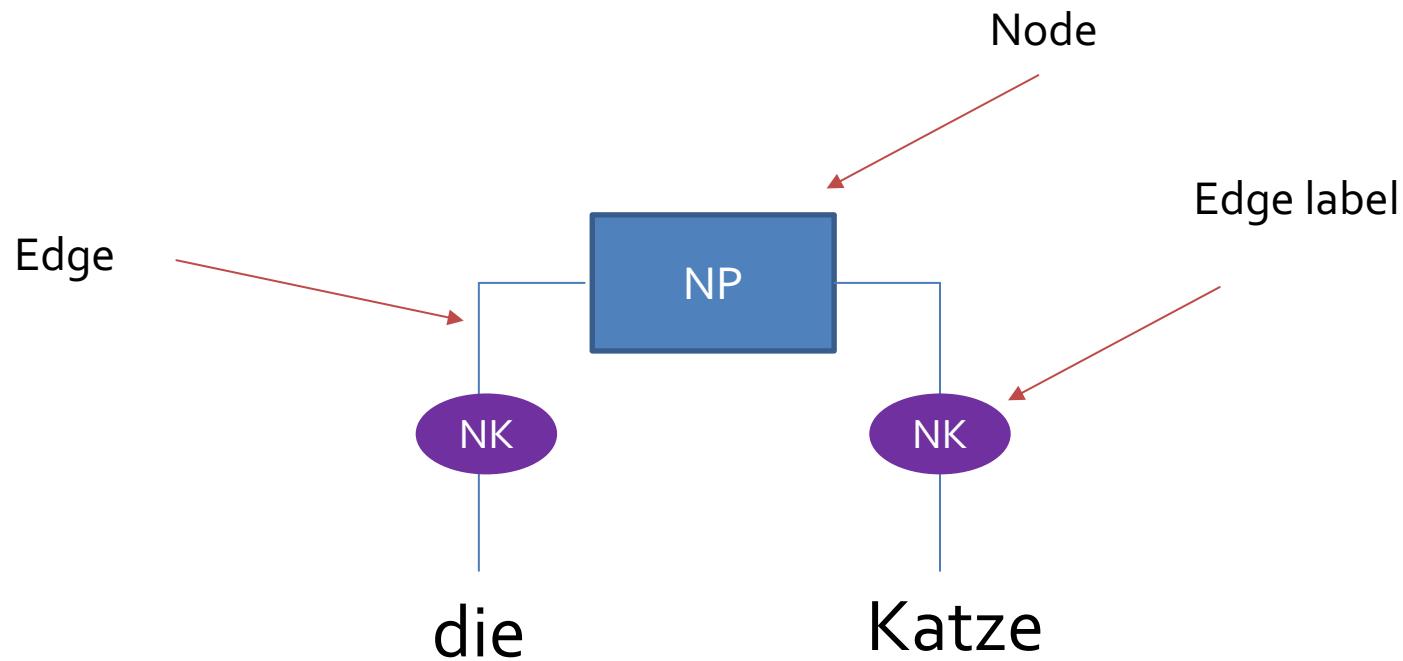
Terminology

- "Constituent": category-neutral description of a phrase
- A phrase is always assigned a specific category, e.g. VP, NP
- Apart from phrases, syntactic analysis has started to work increasingly with chunks
- Chunks are prosodic units of speech (recognizable by small pauses in spoken language) or multi-word units that belong together.

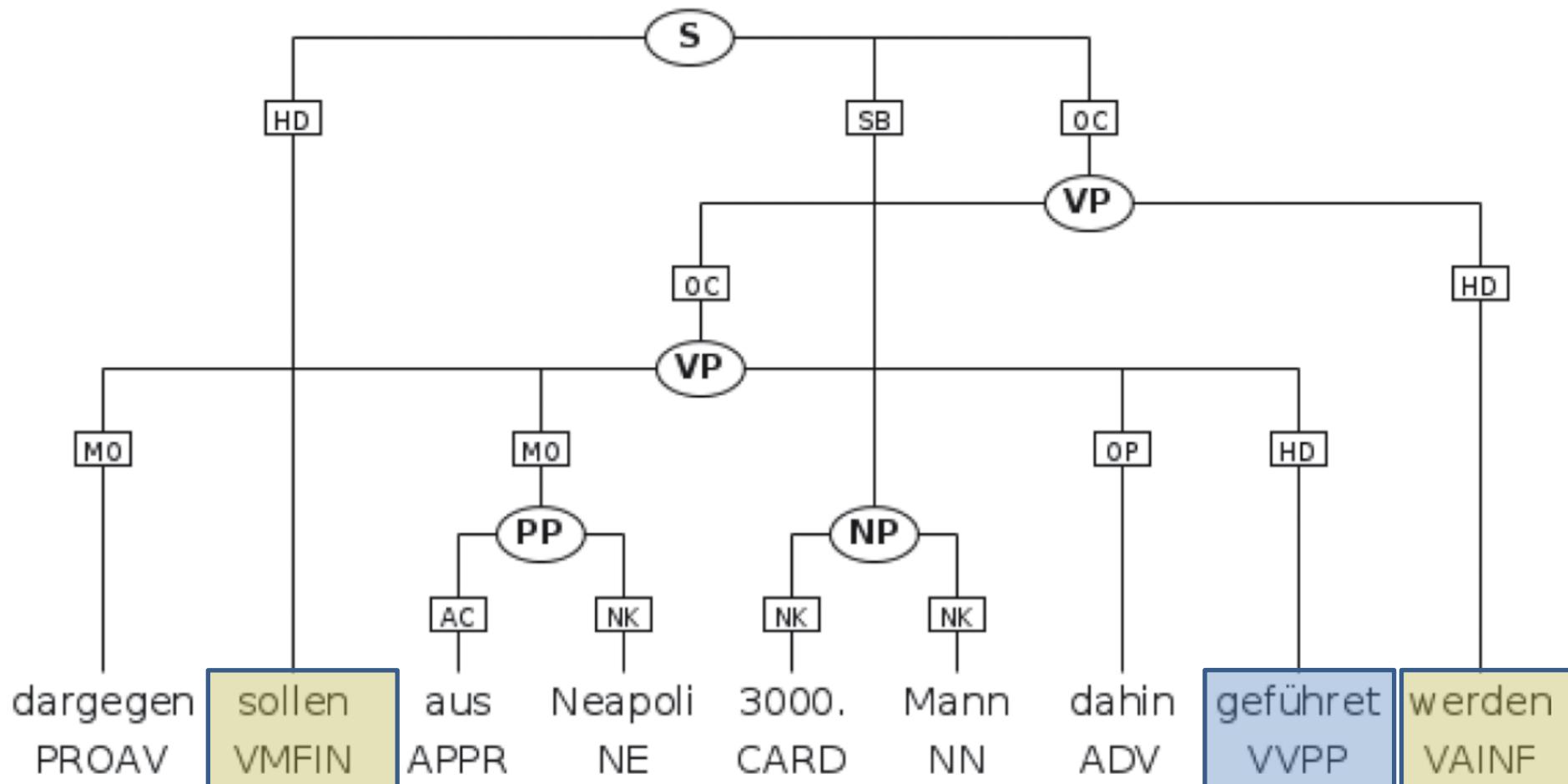
Historical treebanks of German

- Mercurius corpus
 - Newspaper texts from two different years: *Annus Christi 1597*, *Mercurius 1667*
- Deutsche Diachrone Baumbank
 - c. 2500 Tokens from two texts each for OHD, MHG, ENHG

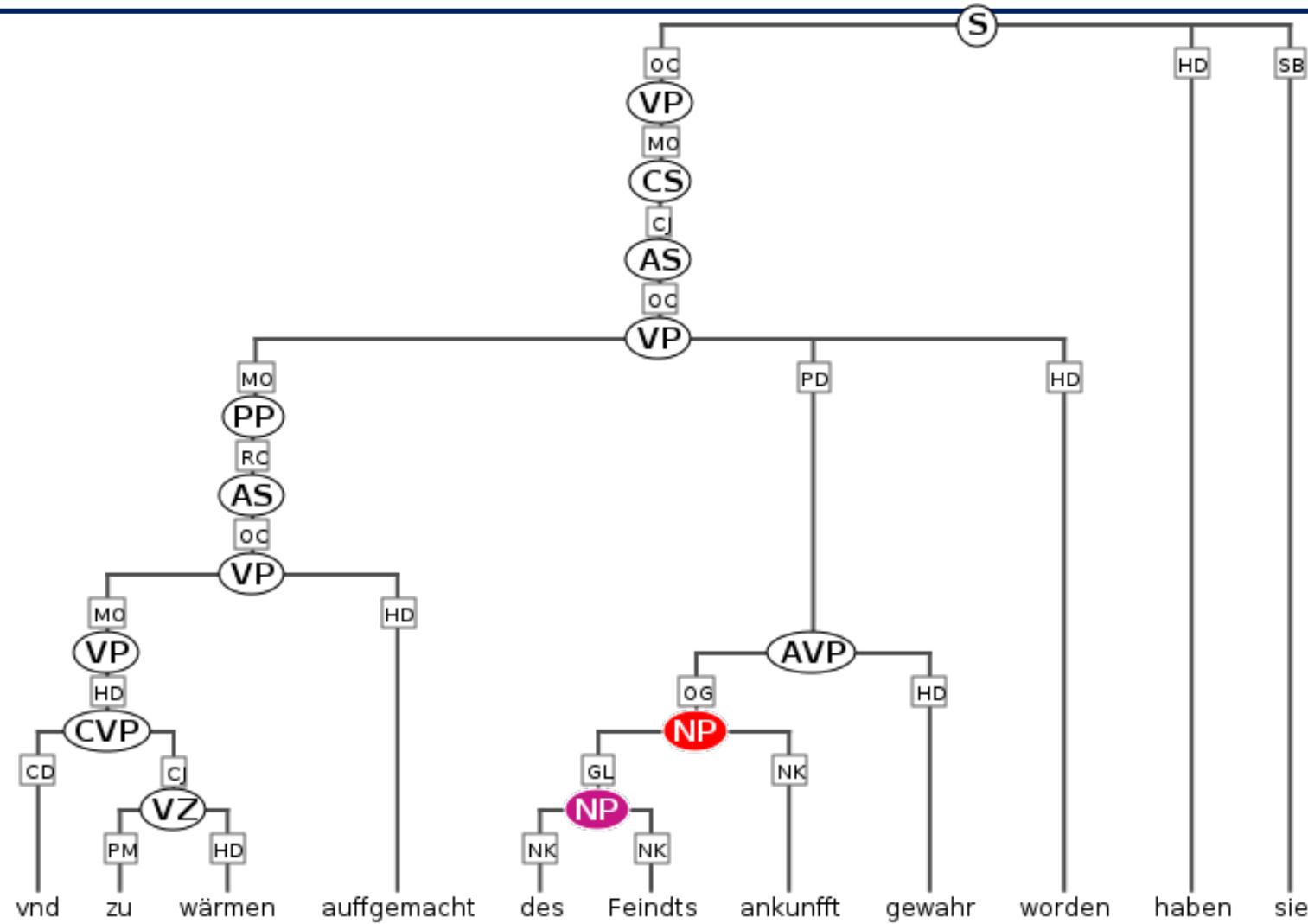
Syntax tree: Example



Crossing edges



Example: Mercurius



Example: Mercurius

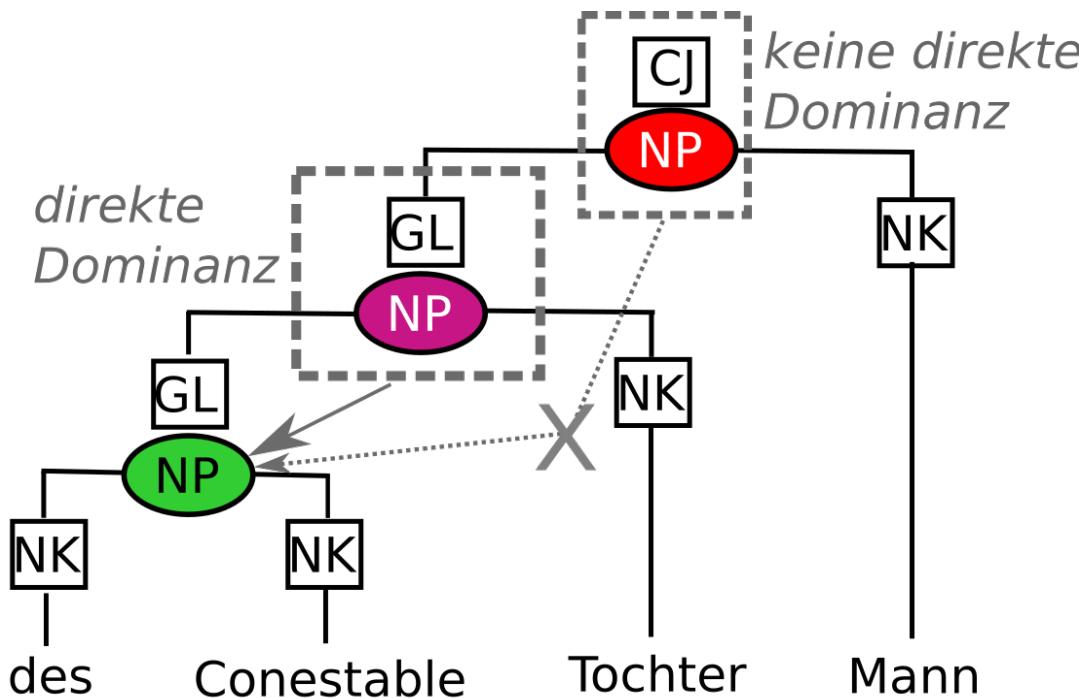
- node > [label="GL"] cat="NP"

any
node

direct
dominance

edge
label

noun phrase



Example: Mercuris

What can we find using the following queries?

- node > cat="NP",
- node >[label="GL"] cat="NP" &
meta::doc="Mercurius-1667"
- node >* cat="NP"