

LAUDATIO - Eine Infrastruktur zur linguistischen Analyse historischer Korpora

Carolin Odebrecht, Humboldt-Universität zu Berlin
Florian Zipser, Humboldt-Universität zu Berlin, INRIA

Linguistische Analyse historischer Korpora - Fragestellung

Fragestellung:

Was heißt es, (korpus-)linguistisch zu arbeiten?

Welche Benutzerszenarien ergeben sich aus der Datenaufbereitung und -analyse (historisch-)linguistischer Forschungsdaten?

Welche Anforderungen stellen sich dann für eine Infrastruktur, die diese Nutzer ansprechen will?

Linguistische Analyse historischer Korpora - Forschung

Forschungsfrage bedingt

- Datenauswahl
 - Annotationen
 - Art der Annotationen
 - Auswertung der Annotationen
- Folge: heterogenes Feld an Daten (Editoren + Formate) und Auswertungen (Analysetool + Formate)
- Kombination dieser verschiedenen Annotationen und Analysen
- Infrastrukturen müssen damit umgehen können

Linguistische Analyse historischer Korpora - Nutzerperspektive

Nutzerperspektive Linguistik:

- (Weiter-)Nutzung dieser heterogenen Forschungsdaten in Abhängigkeit von Forschungsfragen
- grundsätzliche Anwendungen

/1/ Daten annotieren

/2/ Daten speichern

/3/ Daten durchsuchen

/4/ auf Daten zugreifen

Linguistische Analyse historischer Korpora

Base text Token Annotations Show Citation URL

1 / 394 Displaying Results 1 - 10 of 3933

1 Path: DDB.AHD > PAULA_merge > Henchl

#	sagetun	al	ioh	fona	dhem	diubilsiuhhom	mannum	enti	see	saar	alle	dhea
#	sagen	al	io	fona	ther	tiuvalsioh	man	inti	sehan	sar	al	ther
-	3.Pl.Past.Ind	Acc.Sg.Neut	-	-	Dat.Pl.Masc	Pos.Dat.Pl.Masc.Wk	Dat.Pl.Masc	-	2.Sg	-	Nom.Pl.Masc	Nom.Pl.Masc
-	VVFIN	PIS	ADV	APPR	PDAT	ADJA	NN	KON	VVIMP	ADV	PIAT	PDAT

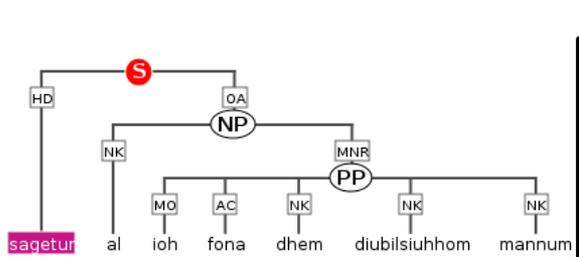
1 Path: DDD-AD_Tatian_171012_exmaralda > Tatian0

Prologus Bithiu uuanta manage zilotun ordinon

meta (grid)

seite	25					
sprache	lat	goh	goh	goh	goh	
txt	Prologus	Bithiu uuanta	manage	zilotun	ordinon	
tok	Prologus	Bithiu	uuanta	manage	zilotun	ordinon

base (grid)



Wortarten
 Lemmatisierung
 Morphologie
 Syntax

.....

Suche und Analyse der Daten in ANNIS

4 Path: Ridges_Herbology_Vers

nichts verhalten will , ob ichs

normalizations

clean	nichts	verhalten	will									
dipl	nichts	verhalten	will									
norm	nichts	verhalten	will	,	ob	ich	es	gleich	tun	könnte	,	weil

default_ns (grid)

st		
Pos_Masc_Pl_Nom_st	Ind_Past_Pl_3	
P	wk2	wk2
a,o	wk2	wk2
Masc		
om		
	Ind	
	Pl	
	3	
DJ	VV	VV
	Past	
anage	zilotun	ordinon

Linguistische Analyse historischer Korpora - Multikey-Annotationen

Tatian DDD-AD

Morphologie:

- Annotationskey und Werte als direkte Analysehilfe für den jeweils konkreten Untersuchungsgegenstand

- Token und/oder Spannen-annotationen erhalten auf unterschiedlichen Ebenen Werte:

Numerus

Kasus

Genus

Modus

Part of Speech

1 Path: DDD-AD_Tatian_171012_exmaralda > Tatian0

Prologus Bithiu uuanta manage zilotun ordinon
meta (grid)

Diplomatische Transkription

seite	25					
sprache	lat	goh	goh	goh	goh	
txt	Prologus	Bithiu uuanta	manage	zilotun	ordinon	
tok	Prologus	Bithiu	uuanta	manage	zilotun	ordinon

Annotations eines einzigen Tokens

dekl			st			
flexB	Sg_Nom		Pos_Masc_Pl_Nom_st	Ind_Past_Pl_3		
flexkB	o		P	wk2	wk2	
flexkL	o		a,o	wk2	wk2	
gen			Masc			
genB	Masc					
genL	Masc					
kasus	Nom		Nom			
modus					Ind	
numerus	Sg		Pl	Pl		
person					3	
posL	NA	KO	ADJ	VV	VV	
tempus					Past	
tok	Prologus	Bithiu	uuanta	manage	zilotun	ordinon

Linguistische Analyse historischer Korpora - Multikey-Annotationen

Tatian DDD-AD

Morphologie:

- Annotationskey und Werte als direkte Analysehilfe für den jeweils konkreten Untersuchungsgegenstand
- Token und/oder Spannenannotationen erhalten auf unterschiedlichen Ebenen

Numerus

Kasus

Genus

Modus

Part of Speech

1 Path: DDD-AD_Tatian_171012_exmaralda > Tatian0

Prologus Bithiu uuanta manage zilotun ordinon

meta (grid)

Diplomatische Transkription

seite	25					
sprache	lat	goh	goh	goh	goh	
txt	Prologus	Bithiu uuanta	manage	zilotun	ordinon	
tok	Prologus	Bithiu	uuanta	manage	zilotun	ordinon

Annotationskey und Werte können nicht antizipiert werden, da abhängig von den jeweiligen Forschungsfragen!

gerB	Masc					
gerL	Masc					
kasus	Nom		Nom			
modus				Ind		
numerus	Sg		PI	PI		
person				3		
posL	NA	KO	ADJ	VV	VV	
tempus				Past		
tok	Prologus	Bithiu	uuanta	manage	zilotun	ordinon

Linguistische Analyse historischer Korpora - Bäume und Konstituenten

Deutsche Diachrone Baubank

- Syntax-Annotationen ermöglichen eine Suche und Darstellung von komplexen Hierarchie-Beziehung
- Suche nach syntaktisch abhängigen Strukturen:

Base text Token Annotations Show Citation URL

< 1 / 394 > | Displaying Results 1 - 10 of 3933

1 Path: DDB.AHD > PAULA merge > Henchl

sagetun	al	ioh	fona	dhem	diubilsiuhhom	mannum	enti	see	saar	alle	dhea
3.Pl.Past.Ind	Acc.Sg.Neut	--	--	Dat.Pl.Masc	Pos.Sg.Pl.Masc.Wk	Dat.Pl.Masc	--	2.Sg	--	Nom.Pl.Masc	Nom.Pl.Masc
VVFIN	PIS	ADV	APPR	PDAT	ADJA	NN	KON	VVIMP	ADV	PIAT	PDAT

tree:

Diplomatische Transkription

Token in Konstituenten- & Hierarchie-Beziehungen

bspw. Nominalphrase, die mit einer bestimmten grammatischen Funktion, wie Subjekt, gelabelt sind

→ theoretische Schlüsse über die Syntax so möglich

Linguistische Analyse historischer Korpora - Bäume und Konstituenten

Deutsche Diachrone Baubank

- Syntax-Annotationen ermöglichen eine Suche und Darstellung von komplexen Hierarchie-Beziehung

- Suche nach syntaktisch abhängigen Strukturen:

bspw. Nominalphrasen, die nach ihrer grammatischen Funktion, wie Subjekt, getabelliert sind

→ theoretische Schlüsse über die Syntax so möglich

Base text Token Annotations Show Citation URL

1 Path: DDB.AHD > PAULA merge > Henchl

sagetun	al	ioh	fona	dhem	diubilsiuhhom	mannum	enti	see	saar	alle	dhea
3.Pl.Past.Ind	Acc.Sg.Neut	--	--	Dat.Pl.Masc	Pos.Pl.Pl.Masc.Wk	Dat.Pl.Masc	--	2.Sg	--	Nom.Pl.Masc	Nom.Pl.Masc
VVFIN	PIS	ADV	APPR	PDAT	ADJA	NN	KON	VVIMP	ADV	PIAT	PDAT

tree:

Annotationskey und Werte können nicht antizipiert werden, da abhängig von den jeweiligen Forschungsfragen!

Diplomatische Transkription

Konstituenten- & e-Beziehungen

grammatischen

Linguistische Analyse historischer Korpora - Multiple Tokenisierung

RIDGES Herbology

- Erhalt der diplomatischen Transkription
 - Normalisierungen ermöglichen eine verbesserte Mustersuche/Suche nach Typen sowie bessere Performanz von Tagging- bzw. Parsingtools
 - morphologisch wie syntaktisch interessante Phänomene so erst suchbar: bspw. Partikelverb-Konstruktionen, Klitika, Komposition
- Mehrebenen-Normalisierung durch multiple Tokenisierung
- linguistischer Vorteil: Varianz der Orthographie noch nachvollziehbar

4 ⓘ Path: Ridges_Herbology_Version_2.0 > Ridges_v2 > flora.saturnizans.1722

nichts verhalten will , ob **ichs** gleich thun könnte , weil

normalizations

clean	nichts	verhalten	will	,	ob	ichs		gleich	thun	könte	,	weil
dipl	nichts	verhalten	will	,	ob	ichs		gleich	thun	könte	,	weil
norm	nichts	verhalten	will	,	ob	ich	es	gleich	tun	könnte	,	weil

default_ns (grid)

Orthographische
Realisierung von Klitika +
Normalisierung

Linguistische Analyse historischer Korpora - Multiple Tokenisierung

RIDGES Herbology

- Erhalt der diplomatischen Transkription
 - Normalisierungen ermöglichen eine verbesserte Mustersuche/Suche nach Typen sowie bessere Performanz von Tagging- bzw. Parsingtools
 - morphologisch wie syntaktisch interessante Phänomene so erst suchbar:
bspw. Partikelverknüpfung, Partikelposition
- Mehrebenen-Normalisierung
- linguistischer Vorkontext nachvollziehbar

Annotationskey und Werte können nicht antizipiert werden, da abhängig von den jeweiligen Forschungsfragen!

4 ⓘ Path: Ridg... zans.1722

nichts verhalten will , ob **ichs** gleich thun könnte , weil

normalizations

clean	nichts	verhalten	will	,	ob	ichs	gleich	thun	könte	,	weil	
dipl	nichts	verhalten	will	,	ob	ichs	gleich	thun	könte	,	weil	
norm	nichts	verhalten	will	,	ob	ich	es	gleich	tun	könnte	,	weil

default_ns (grid)

Orthographische Realisierung von Klitika + Normalisierung

Infrastruktur für historische Korpuslinguistik - Annotationen

- Annotationen **als Grundlage** für Analyse
 - Forschungsfrage, immer theoretisch motiviert
 - Auswahl der Daten
 - unterschiedliche Tag-Sets
 - Datenaufbereitung
 - Bildung von Kategorien, Zuweisung dieser (Annotation), Suche/Analyse dieser über große Menge von Daten (Korpus)
 - viele Gemeinsamkeiten mit modernen Korpora (Aufbereitung und Analyse)
- Aufgabe: Dokumentation dieser Informationen

Infrastruktur für historische Korpuslinguistik - Metadaten

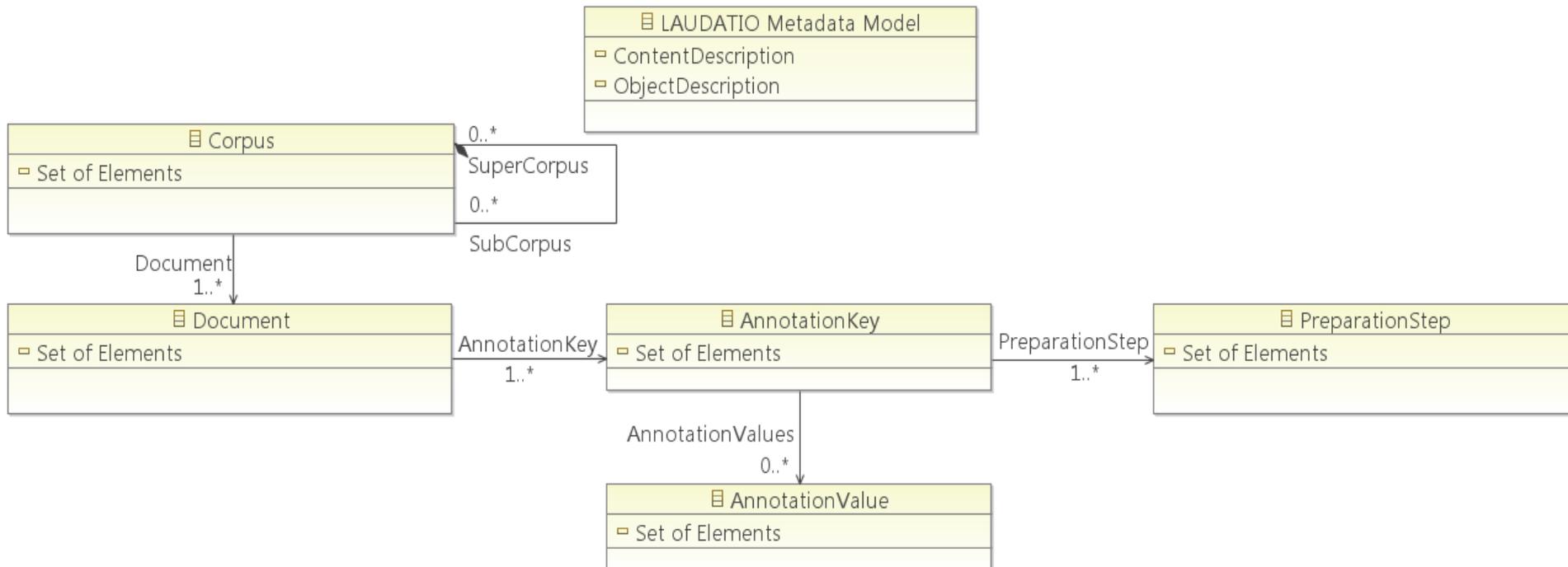
Neben klassischen Metadaten wollen wir (u.a.) wissen:

- Wie wurde transkribiert?
- Wie wurde annotiert?
- Mit welchem Tagset?
- Mit welchen Tools?
- Welche Ebenen beziehen sich auf einander?
- Wurde die Qualität überprüft?

→ (Weiter-)Verwendung dieser Daten durch Dokumentation der Verarbeitungsschritte eines Korpus

→ einheitlich strukturierte Erfassung (Dokumentation) der heterogenen Datengrundlage durch standardisierte Metadaten

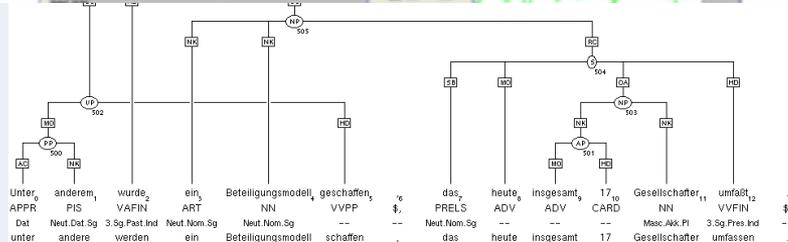
Infrastruktur für historische Korpuslinguistik - Metadaten



Modulares Metadatenmodell mit den Komponenten "Corpus", "Document", "AnnotationKey", "AnnotationValue" und "PreparationStep"

Infrastruktur für historische Korpuslinguistik - Daten annotieren

/1/ Daten annotieren



Viele spezifische Tools für unterschiedliche Annotationen:

- MMAX2, EXMARaLDA, TIGERSearch (Synpathie), RSTTool...

Infrastruktur für historische Korpuslinguistik - Daten speichern

/2/ Daten speichern

The image shows a screenshot of the MMAX2 software interface. The window title is "MMAX2 0.51 BETA c:\HTC\002_hfc_abn.mmax [dirty]". The menu bar includes "File", "Settings", "Display", and "Tools". The main text area contains a paragraph of German text with several words and phrases highlighted in blue. A callout box points to the highlighted phrase "[Das Stadtheater]". Below the text, there is a diagram of the EXMARaLDA interface, which includes a speech bubble icon, the text "EXMARaLDA Transkription, Annotation und Analyse gesprochener Sprache", and a screenshot of the software's main window. Below the diagram, there is a complex tree diagram representing a syntactic tree for the sentence "Unter anderem wurde ein Beteiligungsmodell geschaffen, das heute insgesamt 17 Gesellschafter umfasst." The tree diagram shows hierarchical structures for phrases like "Unter anderem", "wurde ein Beteiligungsmodell geschaffen", and "das heute insgesamt 17 Gesellschafter umfasst". Below the tree diagram, there is a table of morphological and syntactic information for each word in the sentence.

Word	Tag	Form	Case	Number	Gender	Person	Verb Form	Verb Class
Unter	APPFR	unter	Dat	Neut	Dat	Sg		
anderem	PIS	andere	Dat	Neut	Dat	Sg		
wurde	VAFIN	werden	3. Sg	Past	Ind			
ein	ART	ein	Neut	Nom	Sg			
Beteiligungsmodell	NN	Beteiligungsmodell	Neut	Nom	Sg			
geschaffen	VVPP	schaffen						
,	%	,						
das	PRELS	das	Neut	Nom	Sg			
heute	ADV	heute						
insgesamt	ADV	insgesamt						
17	CARD	17						
Gesellschafter	NN	Gesellschafter	Masc	Acc	Pl			
umfasst	VVFIN	umfassen	3. Sg	Pres	Ind			
.	%	.						

- Jedes Tool hat ein eigenes Format :-)
- wenig Interoperabilität :-)
- manche Tools werden nicht mehr gepflegt :-)

Infrastruktur für historische Korpuslinguistik - Daten speichern

/2/ Daten speichern

- Jedes Tool hat ein eigenes Format :-)
- geringe Interoperabilität :-)
- Manche Tools werden nicht mehr gepflegt :-)

The image shows a screenshot of the MMAX2 software interface. The window title is "MMAX2 0.51 BETA c:\HTC\002_htc_abn.mmax [dirty]". The menu bar includes "File", "Settings", "Display", and "Tools". The main text area contains a German sentence: "Im Gegensatz zu [anderen Städten] steht [das] nicht an [herausgehobener Stelle], sondern [e] Straßenfucht] ohne [Vorplatz] ein . [Der Haupt] Arkadenmotiv mit [den flachen Segmentbögen]". A green speech bubble is overlaid on the text. Below the screenshot is the TEI logo, which consists of the letters "TEI" in a bold, black font, enclosed in a yellow, 3D-style frame. Below the logo, the text "TEXT ENCODING INITIATIVE" is written in a black, italicized font.

Below the TEI logo, there is a complex tree diagram representing a text encoding structure. The diagram shows a hierarchy of nodes and branches, with various labels and numbers. Below the diagram, there is a table of text elements and their corresponding TEI tags and grammatical information.

Unter ₆	anderem ₆	wurde ₂	ein ₃	Beteiligungsmodell ₅	geschaffen ₅	%	das ₂	heute ₉	insgesamt ₃	17 ₁₀	Gesellschafter ₁₁	umfaßt ₂	\$
APPR	PIS	VAFIN	ART	NN	VVPP	--	PRELS	ADV	ADV	CARD	NN	VVFIN	--
Dat	Neut.Dat.Sg	3.Sg.Past.Ind	Neut.Nom.Sg	Neut.Nom.Sg	--	--	Neut.Nom.Sg	--	--	--	Misc.Akk.Pl	3.Sg.Pres.Ind	--
unter	andere	werden	ein	Beteiligungsmodell	schaffen	,	das	heute	Insgesamt	17	Gesellschafter	umfassen	.

Infrastruktur für historische Korpuslinguistik - TEI

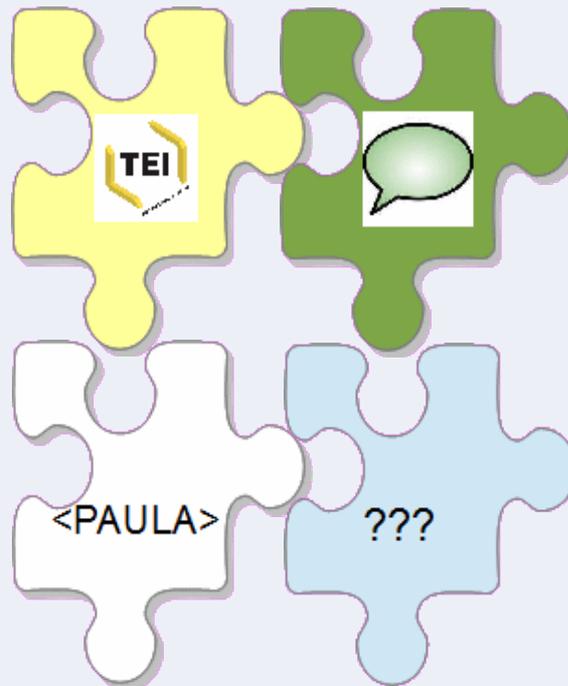
- Warum reicht uns TEI nicht?



- bereits existierende Korpora oft nicht in TEI
- viele bestehende Tools unterstützen kein TEI
- viele neue Kategorien pro Tag (individuelle Forschungsfrage)
 - Formate müssen flexibel/ generisch sein
- viele Annotationsarten nicht gut von TEI unterstützt
- Linguistik: nicht nur reine Digitalisierung des Originals

Infrastruktur für historische Korpuslinguistik - Was nun?

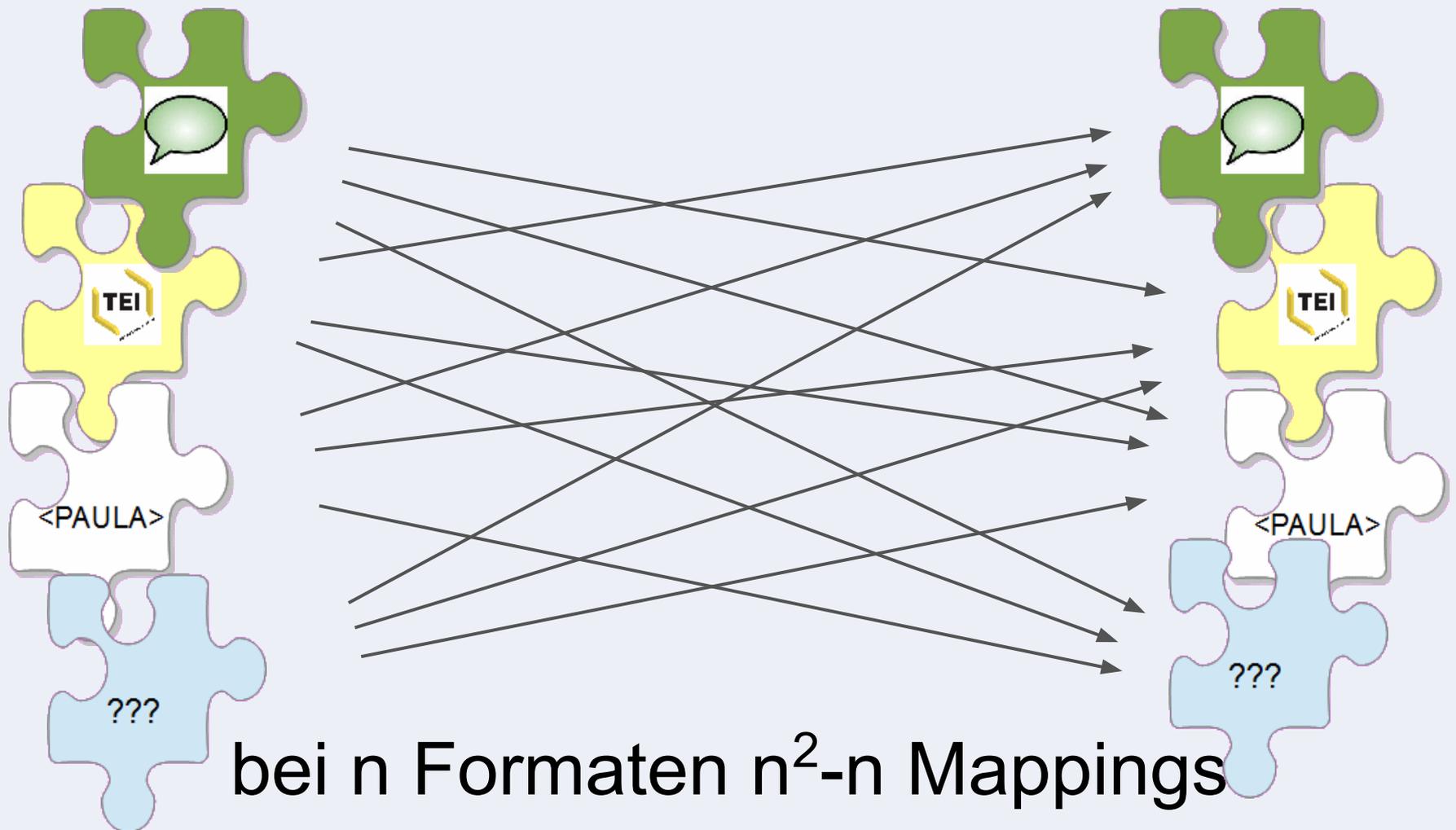
- erst Kombination der Formate ergibt möglichst umfangreiches Bild
→ Formatpluriversum



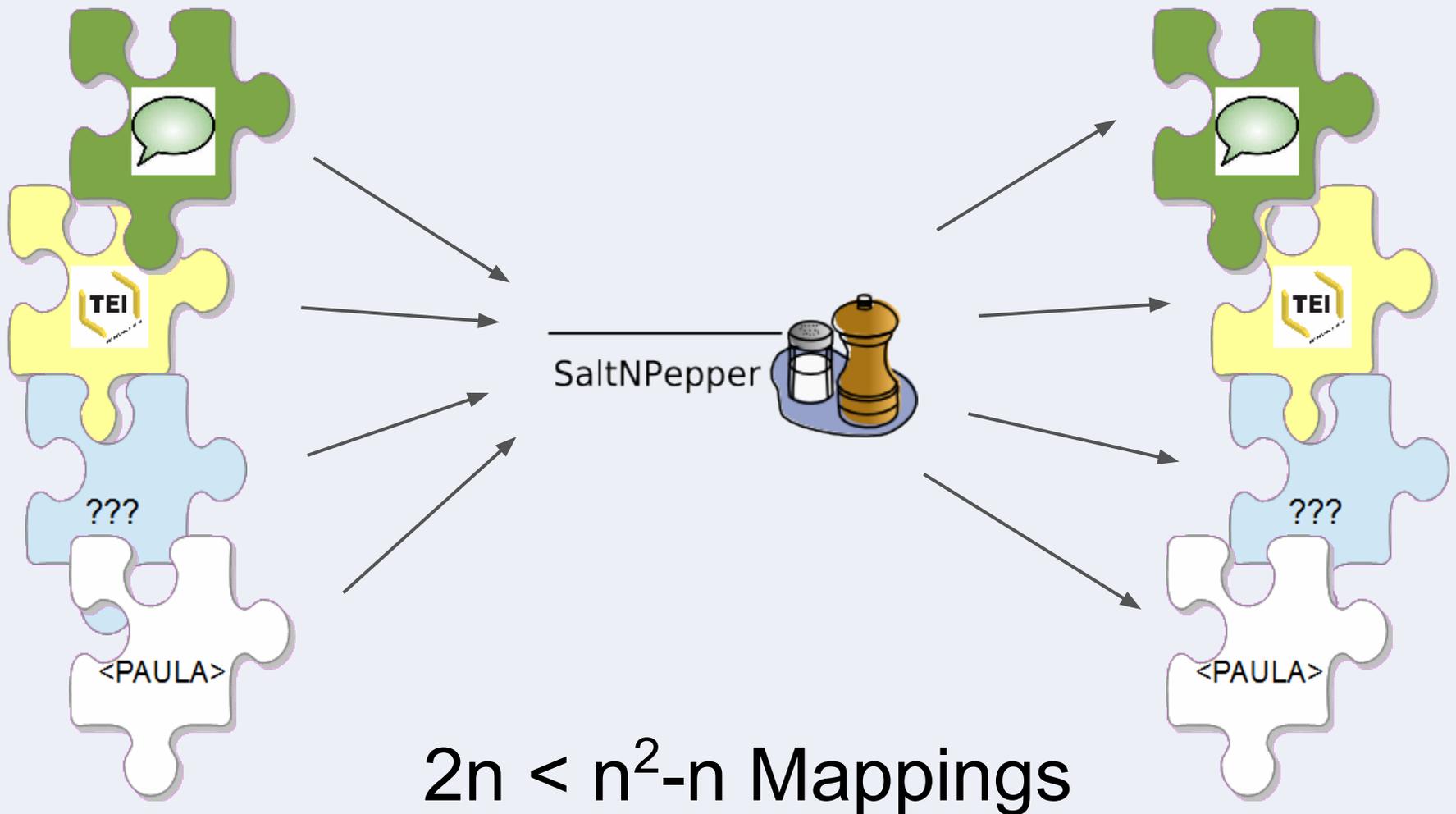
Infrastruktur für historische Korpuslinguistik - Formatpluriversum

- aber wir wollen Interoperabilität zwischen Formaten
soweit wie möglich
 - Digitaler Text,
 - Tokenisierung und
 - bestimmte Annotationin verschiedenen Tools nutzbar
 - Annotationen abbildbar auf Mehrebenenformate
- Überführen was zu überführen geht

Infrastruktur für historische Korpuslinguistik - Formatpluriversum

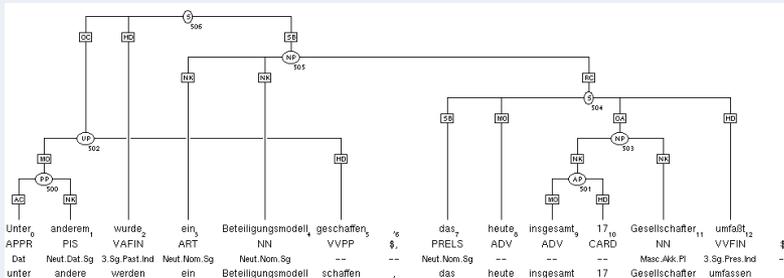


Infrastruktur für historische Korpuslinguistik - Formatpluriversum



Infrastruktur für historische Korpuslinguistik - Daten durchsuchen

/3/ Daten durchsuchen



- auf einzelner Ebene

Search Form

ANNSYS: Tutorial

Search Form

ANNSYSQL: [tok & tok & #1 ->dep [func="OA"] #2 & cat="S" & #2 _#1 & node & #3 >secede #4 | correction="correcting" | correct]

Query Builder:

Show >>>

Result: 43

History: Query History

More Corpora

Name	Texts	Tokens
FalkEisner2V2_0	248	131511
ONTONOTES_v1.5_small	4	6450
SMILTRON_Banana	2	2782
TueSWS_no_cvc	2187	770949
sgn_J	24	164
b4.miller2.0	2031	11295
pos-3	3	573
pos2	2	399
tiger1_dep	1	929
tiger2	1971	888578

Search Export

Context Left: 0

Context Right: 0

Results per page: 10

Show Result

Search Result - tok & tok & #1 ->dep[func="OA"] #2 & cat="S" & #3 _#1 & node & #3 >secede #4 (0, 0)

Page: 1 of 5

Token Annotations Show Citation URL

Displaying Results 1 - 10 of 43

während 78 Prozent sich für Bush und vier Prozent für Clinton aussprechen

während 78 Prozent sich für Bush und vier Prozent für Clinton aussprechen

KOLS_CARD NN PRF APPR NE KON_CARD NN APPR NE VVFIN

-- **Neut 3.Acc.Pl -- Acc.Sa* -- **Neut -- Acc.Sg* 3.PiPast.Ind

dependencies

constituents

Die Vase auf dem Tisch ist größer als die Vase

animacy (grd)

Select Displayed Annotation Levels

mmaxref_type Baum

mmaxref_type Baum

tok Die Vase auf dem Tisch ist größer als die Vase

conference (discourse)

Die Vase auf dem Tisch ist größer als die Vase auf der Fensterbank . ich fride . sie sieht nicht so gut aus . weil der Tisch zu klein ist

- parallel auf mehreren Ebenen

Infrastruktur für historische Korpuslinguistik - Datenzugriff

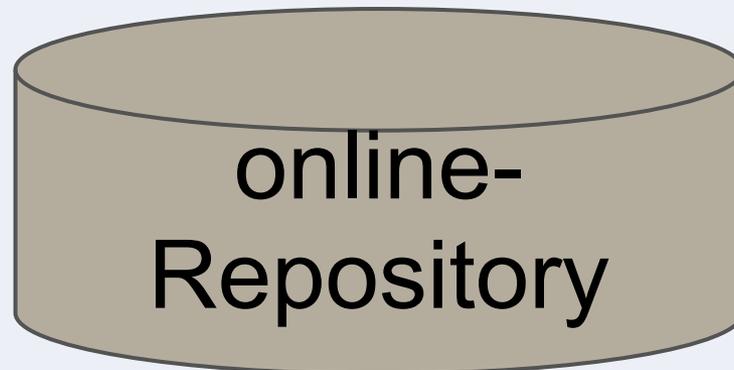
/4/ auf Daten zugreifen

- immer und überall
- Upload/ Download
 - in verschiedenen Formaten- Facetten
- Versionierungssystem
- einheitliche Metadatensuche

Infrastruktur für historische Korpuslinguistik - Datenzugriff

/4/ auf Daten zugreifen

- immer und überall
- Upload/ Download
 - in verschiedenen Formaten- Facetten
- Versionierungssystem
- einheitliche Metadatensuche

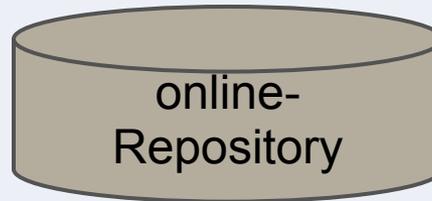


LAUDATIO - Long-term Access and Usage of Deeply Annotated Information

Filter by

- Corpora
 - ridges (2)
 - herology (2)
 - changed (1)
- Projects
 - ridges (2)
 - herology (2)
 - changed (1)
- Formats
 - heron (2)
 - relanis (2)
 - exmaralda (2)

Title: Ridges Herbiology, 2011-MM-DD
 Status: draft Change: 1 Type: NA
 63754 Seiten
[http://dipol1-7.oms.hu-berlin.de/corpus/](#)
[2.2.3.3/MAX2/objekte/halbschriftlich/akroponridges1_compact](#)
 Homepage: [http://korling.german.hu-berlin.de/ridges/index_de.html](#)
 Project Description: Das RIDGES-Projekt (Register in Diachronic German Science) untersucht die Entstehung und Entwicklung der deutschen Wissenschaftspraxis ab Mitte des 16. Jahrhunderts bis ins späte 19. Jahrhundert. Bis zum 16. Jahrhundert war die Sprache der Wissenschaft in Europa Latein, alle wissenschaftlichen Texte wurden auf Latein verfasst. Erst im Laufe des 16. Jahrhunderts ... (more)
 Documents of the Corpus: Deutsche Pflanzennamen Noch einige Worte über den Dichtungsart und die Polyembryonie bei den höheren Pflanzen? Flora oder Botanische Zeitung? Flora oder Botanische Zeitung? Grundriss der Kräuterkunde zu Vorkesslingen entworfen Unterricht von der allgemeinen Kräuter- und Wurzeltröckung Flora satumiana, Die Verwandschaft des Pflanzen mit dem Mineral Reich Curator Botanische oder sonderbare Kräuter Buch Curator Botanische oder sonderbare Kräuter Buch Pflanz-Gart Pflanz-Gart Alchimistische Praktik New Krieditbuch Flora francisca rediva



EXMARaLDA
 Transkription, Annotation und Analyse gesprochener Sprache

Im Gegensatz zu [anderen Städten] steht das [Heidelberger Stadttheater] nicht an [herausgehobener Stelle], sondern [es] fügt sich in [die Straßenführung] ohne [Auszeichnung] ein. [Der Hauptgang] zeigt noch [das alte Ankerdenkmal] [sein flachen Segmentbögen]. Erst [1674] übernimmt [die ...]

Search Results

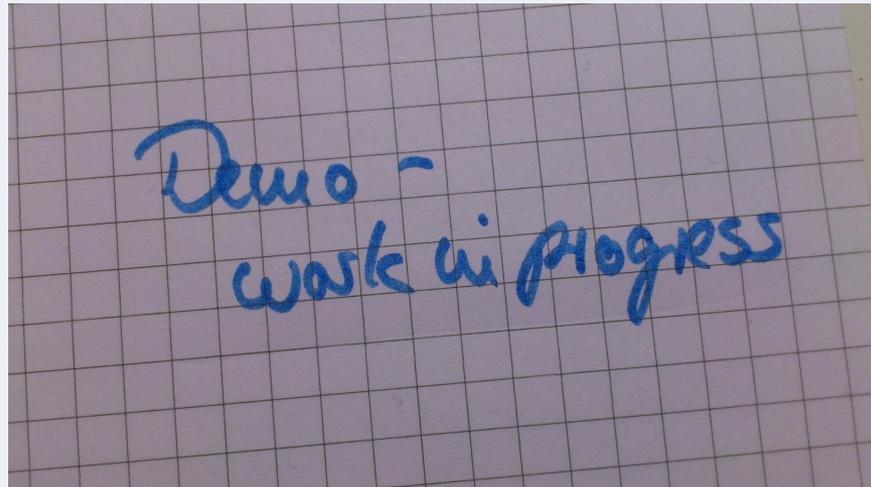
Query: `ridges & ridges & #1 -> ridges`

Result: 43

Item	Texts	Items
Palustris2V2_0	249	131511
ORTONORDE_us_s_journal	4	6460
SALZMINERALE	2	2302
SalzMinerale	2107	770949
pp1	24	164
Artenbuch	2020	11296
erick	2	672
erick	2	399
Sp11.de	1	509
Sp2	1071	88078

Vielen Dank für Ihre Aufmerksamkeit!

Demo von LAUDATIO
auf der DGfS-Tagung 2013 in Potsdam!



Referenzen

ANNIS	Zeldes, Amir, Ritz, Julia, Lüdeling, Anke & Christian Chiarcos (2009) ANNIS: A Search Tool for Multi-Layer Annotated Corpora . In: <i>Proceedings of Corpus Linguistics 2009</i> , Liverpool, July 20-23, 2009.
Deutsche Diachrone Baumbank	Hirschmann, Hagen & Sonja Linde (erscheint): Annotationsbeschreibung für syntaktische Annotationen unterschiedlicher Sprachstufen des Deutschen . Technical Report. Humboldt-Universität zu Berlin. https://korpling.german.hu-berlin.de/annis3/
EXMARaLDA	Schmidt, Thomas (2002) EXMARaLDA - ein System zur Diskurstranskription auf dem Computer . <i>Arbeiten zur Mehrsprachigkeit</i> , Folge B 34:1 ff. http://www.exmaralda.org/files/AZM.pdf .
LAUDATIO	Krause, Thomas, Odebrecht, Carolin & Dennis Zielke (erscheint) Langfristiger Zugang und Nutzung von tief annotierten Korpora: LAUDATIO . <i>32. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft</i> . http://www.laudatio-repository.org/
MMAX2	Müller, Christoph & Michael, Strube (2006): Multi-Level Annotation of Linguistic Data with MMAX2 . In: Sabine Braun, Kurt Kohn & Joybrato Mukherjee (Eds.): <i>Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods</i> . Frankfurt: Peter Lang, pp. 197-214. (English Corpus Linguistics, Vol.3).
PAULA	Dipper, Stefanie (2005) XML-based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation . In: Eckstein R, Tolksdorf R (Eds.) <i>Berliner XML Tage</i> .

Referenzen

RIDGES Herbology	Krause, Thomas, Lüdeling, Anke, Odebrecht, Carolin & Amir Zeldes (2012) Multiple Tokenization in a Diachronic Corpus . <i>Exploring Ancient Languages through Corpora Conference (EALC), 14.-16.Juni 2012</i> . http://korpling.german.hu-berlin.de/ridges/documentation_en.html https://korpling.german.hu-berlin.de/annis3/
RSTTool	O'Donnell, Michael (2000) "RSTTool 2.4 -- A Markup Tool for Rhetorical Structure Theory". <i>Proceedings of the International Natural Language Generation Conference (INLG'2000)</i> , 13-16 June 2000, Mitzpe Ramon, Israel. 253 - 256.
SaltNPepper	Zipser, Florian & Laurent Romary (2010) A model oriented approach to the mapping of annotation formats using standards In: <i>Proceedings of the Workshop on Language Resource and Language Technology Standards, LREC 2010. Malta</i> . URL: http://hal.archives-ouvertes.fr/inria-00527799/en/
Tatian DDD	Linda, Sonja, Unverzagt, Silke & Karin Donhauser (erscheint) Old German Reference Corpus . <i>32. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft</i> . https://korpling.german.hu-berlin.de/annis3/
TEI	Burnard, Lou & Syd Bauman (Eds.) (2008). TEI P5: Guidelines for Electronic Text Encoding and Interchange . Oxford. http://www.tei-c.org/Guidelines/P5/ .
TIGERSearch	Lezius, Wolfgang (2002) Ein Suchwerkzeug für syntaktisch annotierte Textkorpora . http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/
Treetagger	Schmid, Helmut (1994) Probabilistic Part-of-Speech Tagging Using Decision Trees . In: <i>Proceedings of International Conference on New Methods in Language Processing</i> .

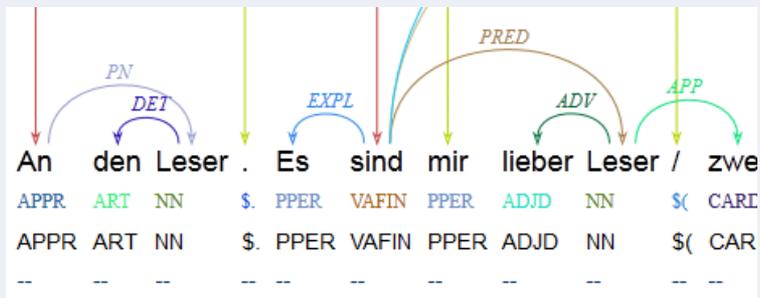
DTA Basis Format

default_ns (grid)												
Select Displayed Annotation Levels ▾												
bibl									bibl			
body	body											
cit									cit			
div	div					div						
figure										figure		
fw												
head						head						
hi						hi		hi				
hi								hi				
lg												
n	1					1						
p	p									p		
place												
quote									quote			
rendition						#b		#et				
rendition								#aq				
text	text											
type												
tok	klärt, bei einigen ist Erläuterung überflüssig. So möge denn das Ge-	wand bunt genug feyn, um manche Fehler und Schwächen der Sache	fehlt zu verdecken oder doch minder fühlbar zu machen, kurz möchten	diese in der That anspruchslosen Betrachtungen, nachrichtige und	freundliche Lefer finden.	Erste Vorlesung.	Das Auge und das Microscop	Oculus ad vitam nihil facit, ad vitam beatam nihil magis. —	Seneca	Kein Organ ist so unwichtig für das Leben als das Auge,	Keins so wichtig für die Schönheit des Lebens. —	Die Vignette giebt einen idealen Durchschnitt durch die kleine Camera ob-

Linguistische Analyse historischer Korpora - Bäume und Konstituenten

- weitere Möglichkeit der syntaktischen Annotation:

Dependenzen



Pointing Relation

und	verzagt	wie	die	deutsche	Abwehrreihe	der	Fußballkicker	.	Und	dann
und	verzagt	wie	der	deutsch	Abwehrreihe	der	Fußballkicker	.	und	dann
--	Pos	--	Nom.Sg.Fem	Pos.Nom.Sg.Fem	Nom.Sg.Fem	Gen.Pl.Masc	Gen.Pl.Masc	--	--	--
KON	ADJD	KOKOM	ART	ADJA	NN	ART	NN	\$.	KON	ADV

mmax (discourse)

Stellpass Wunder gibt es immer wieder ! Erst spielen die Dallgower Gemeindevertreter so statisch und Tiefe solch ein fulminanter Stellpass , von dem man hofft , dass die Seeburger oder Groß-Glienicker , es vorerst keine Gefahr fürs Dallgower Tor gab . Die Seeburger und einige Groß-Glienicker haben de sie zeigen , wie sie die Chance verwerten . Eine Diskussion , wo künftig die Trainerkabine stehen si deutschen Grotten-Kickern gibt es immer noch . Auch wenn die Spieler aus den verschiedenen Verei einer Mannschaft " Döberitzer Heide " spielen . Und das heißt geme **Component: 5, Type: anaphor_a** kleinsten Schubser gegenseitig zu zerfleischen , sind normalerweise überflüssig . Vorerst allerdings hi