

Dokumentation des Korpus
Ridges Herbology Version 6.0

auf Grundlage des Metadatenframeworks nach LAUDATIO

Stand 14.03.2017

-deutsch-



Malte Belz
Carolin Odebrecht
Laura Perlitz
Gohar Schnelle
Vivian Voigt

Inhalt

1. Korpus.....	6
2. Dokumente.....	11
2.1. Kraeuterbuch_1914_Losch.....	12
2.2. NatürlichePflanzenfamilien_1887_Engler	12
2.3. GemeinnuetzigesKraeuterbuch_1874_Siegmund	13
2.4. FloraDerPreussischenRheinlande_1870_Wirtgen.....	13
2.5. DeutschePflanzennamen_1870_Grassmann.....	14
2.6. VorlesungenUeberKraeuterkunde_1843_Link.....	14
2.7. GrossesIllustriertesKraeuterbuch_1860_Mueller	15
2.8. NochEinigeWorte_1840_Meyen	15
2.9. EigenschaftenAllerHeilpflanzen_1828_Anonymous	16
2.10. EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous.....	16
2.11. Flora-7_1821_Sieber.....	17
2.12. Flora-6_1821_Wilbrand.....	17
2.13. GetreueDarstellungUndBeschreibung_1809_Hayne	18
2.14. Apothekerlexikon_1793_Hahnemann	18
2.15. GrundrissKraeuterkunde_1792_Willdenow	19
2.16. Unterricht_1774_Eisen	19
2.17. EinleitungZuDerKraeuterkenntnisz_1764_Oeder	20
2.18. BlackwellischesKraeuterbuch_1750_Blackwell.....	20
2.19. MysterivmSigillorvm_1735_Hiebner	21
2.20. FloraSaturnizans_1722_Henckel	21
2.21. ViridariumReformatum_1719_Valentini.....	22
2.22. TheatrumBotanicum_1696_Verzascha.....	22
2.23. SchweizerischerBotanicus_1687_vonRoll.....	23
2.24. SonderbaresKraeuterbuch-11-21_1675_Anonymous.....	24
2.25. SonderbaresKraeuterbuch-1-11_1675_Anonymous.....	24
2.26. ThesaurusSanitatis_1673_Nasser.....	25
2.27. Phythologia_1662_Becher	25
2.28. Wund-Artzney_1652_Greiff.....	26
2.29. PflantzGart-VR_1639_Rhagor	26
2.30. PflantzGart-c4_1639_Rhagor	27
2.31. PflantzGart_1639_Rhagor	27
2.32. Kraeuterbuch_1609_Carrichter.....	28

2.33.	HortulusSanitatis_1609_Uffenbach.....	29
2.34.	AlchymistischePractic_1603_Libavius.....	29
2.35.	AlchymistischePractic-VR_1603_Libavius.....	30
2.36.	Paradeiszgaertlein_1588_Rosbach.....	30
2.37.	NewKreueterbuch_1563_Handsch.....	31
2.38.	WieSichMeniglich_1557_vonBodenstein.....	31
2.39.	WieSichMeniglich-VR_1557_vonBodenstein.....	32
2.40.	NewKreueterbuch_1543_Fuchs.....	32
2.41.	NewKreuetterBuch_1539_Bock.....	33
2.42.	NewKreuetterBuch-VR_1539_Bock.....	34
2.43.	ContrafaytKreuterbuch_1532_Brunfels.....	34
2.44.	ContrafaytKreuterbuch- CCXXXVII-CCXLVIII_1532_Brunfels.....	35
2.45.	ContrafaytKreuterbuch-VR_1532_Brunfels.....	35
2.46.	ArtzneyBuchleinDerKreutter_1532_Tallat.....	36
2.47.	ArtzneyBuchleinDerKreutter-VR_1532_Tallat.....	36
2.48.	GartDerGesundheit_1487_vonCuba.....	37
2.49.	GartDerGesundheit-VR_1487_vonCuba.....	38
2.50.	BuchDerNatur_1482_vonMegenberg.....	38
3.	Annotationsebenen – Transkription/Normalisierung.....	39
3.1.	dipl.....	39
3.2.	clean.....	40
3.3.	norm.....	42
4.	Annotationsebenen – Annotationen zu linguistischen Eigenschaften.....	44
4.1.	pos.....	44
4.2.	lemma.....	46
4.3.	foreign.....	48
4.4.	foreign_trans.....	50
4.5.	lang.....	52
4.6.	sentence_end.....	54
4.7.	position_verb.....	55
4.8.	subclause_type.....	57
4.9.	KOUS_sem.....	59
4.10.	comp.....	62
4.11.	comp_orth.....	64
4.12.	prot.....	66

4.13.	attr_gen	69
4.14.	morph_ellipsis.....	71
4.15.	comment_lex.....	73
4.16.	comment.....	74
4.17.	reference.....	76
4.18.	form_disease	77
4.19.	problem	79
4.20.	herbname_norm	81
4.21.	herbprep	82
4.22.	form_prep.....	84
4.23.	noun_nom	86
4.24.	form_noun.....	87
4.25.	clause_type.....	89
4.26.	position_rel	92
4.27.	position_referent	93
4.28.	form_referent	95
4.29.	position_verb_rel	97
4.30.	form_relpron	99
4.31.	mod_referent.....	100
4.32.	ppk_e1.....	102
4.33.	ppk_e2.....	103
4.34.	ppk_e3.....	105
5.	Annotationsebenen – Strukturelle Annotation.....	107
5.1.	lb.....	107
5.2.	pb.....	108
5.3.	pb_n	110
5.4.	pb_ana.....	111
5.5.	unclear	113
5.6.	atLeast.....	115
5.7.	atMost.....	117
5.8.	interpretation.....	119
5.9.	figure.....	121
5.10.	figure_rend.....	122
5.11.	hi	125
5.12.	script.....	127

5.13.	hi_rend	129
5.14.	head	130
5.15.	note.....	132
5.16.	ref	134
5.17.	ref_target	136
5.18.	ref_type	138
5.19.	quote.....	140
5.20.	item	142
6.	Annotationsebenen – Inhaltliche Annotation.....	144
6.1.	definition.....	144
6.2.	disease.....	146
6.3.	persname	148
6.4.	title	150
6.5.	author_ref.....	151
6.6.	reader_ref.....	153
6.7.	plant	155
6.8.	property.....	157
6.9.	name.....	159
6.10.	name_type	161
7.	Metadaten.....	163
8.	Referenzen	166
1.	Anhang	167

1. Korpus

Typ: Corpus – Ridges Herbology

Name:	Ridges Herbology
Herausgeber:	Anke Lüdeling, Carolin Odebrecht, Amir Zeldes, Korpuslinguistik, Humboldt-Universität zu Berlin
Annotatoren:	Ilham Abed-Ali, Silke Andresen, Henriette Ast, Malte Belz, Cornelia Binnyus, Simon Blum, Doreen Christen, Janin Czilwik, Mascha Dayal, Antonia Dittberner, Cora Döhn, Imke Driemel, Christian Ebert, Olja Efremova, Gill-Maria Eichhorn, Judith Esser, Annegret Gerlach, Linda Giesel, Yelyzaveta Hiebert, Hagen Hirschmann, Nikolaos Kartalis, Sebastian Kiraga, Ewa Anna Kolbik, Anna Konik, Kornél Kovács, Marco Krämer, Daisy Krüger, Anna-Maria Lehmann, Eliese-Sophia Lincke, Maria Lober, Anke Lüdeling, Laura Lueders, Tatjana Malon, Samuele Maniscalco, Maria Martynova, Kim Kristin Maser, Laura McKee, Manuel Metzsig, Alexander Meyer, Lisa Mohs, Sandra Müller, Vinzent Müller, Andrew Murphy, Johannes Mursell, Carolin Odebrecht, Akiko Okuda, Mareike Otto, Angelo Papenhoff, Laura Perlitz, Joachim Polte, Gina Reetz, Valerie Reichardt, Katharina Reinig, Ina Riesler, Lena Rosin, Romy Sachs, Franz-Josef Sachse, Anna Sapronova, Simon Sauer, Anna Greta Schmahl, Claudia Schmidt, Gohar Schnelle, Athina Sioupi, Lucia Sohmen, Iryna Sorokovska, Uwe Springmann, Kristina Stephan, Aleksandra Swiech, Juliane Tiemann, Anna Tóth, Réka Tóth, Alexander Turtureanu, Magdalena Vock, Monika Walak, Juliana Wekel, Vita Rosalie Wijffels, Svenja Wolter, Isabell Ines Zander, Karolina Zuchewicz (nachfolgend Studierende der Humboldt-Universität zu Berlin genannt)
Veröffentlichung:	14.03.2017
Anzahl der Texte:	50
Zeitraum:	1482-1914
dipl-Token:	222500
Sprache:	Deutsch
Register:	Kräutertexte
Zugang:	https://korpling.german.hu-berlin.de/annis3/
Projekt:	http://korpling.german.hu-Berlin.de/ridges/index_de.html
Projektbeschreibung:	Das RIDGES -Projekt (R egister i n D iachronic G erman S cience) untersucht die Entstehung und Entwicklung der deutschen Wissenschaftssprache ab Mitte des 15. Jahrhunderts bis ins späte 19. Jahrhundert. Bis zum 16. Jahrhundert war die Sprache der Wissenschaft

in Europa vorwiegend Latein und erst im Laufe des 15. Jahrhunderts beginnen deutsche Wissenschaftler, wissenschaftliche Texte (oder zumindest Texte mit wissenschaftlichen Inhalten für Laien) erstmalig auch auf Deutsch zu formulieren. Dazu mussten sie ein wissenschaftliches Register ‚erfinden‘ – sie mussten die Terminologie genauso wie angemessene Textstrukturen entwickeln und erproben. Wissenschaftliche Texte haben sich über die Jahrhunderte natürlich verändert. Dabei sind die meisten Veränderungen quantitativ und nicht qualitativ (kategorial). Man kann die Entwicklungen nur nachzeichnen, wenn man Texte oder Textsammlungen von vielen Zeitpunkten genau miteinander vergleicht. In unserem Projekt sollen wissenschaftliche Texte daher auf allen sprachlichen Ebenen (Syntax, Wortbildung, Lexik, Phraseologie, Textstruktur etc.) analysiert werden, um Entwicklungen und Tendenzen identifizieren und beschreiben zu können. In einem variationistischen Ansatz annotieren wir wissenschaftliche Texte mit korpuslinguistischen Methoden und nutzen quantitative Verfahren, um Veränderungen in den Merkmalen zu identifizieren.

Wichtig ist uns dabei die Mitwirkung unserer Studierenden, die im Rahmen ihres Studiums bei der Erstellung und der Annotation der Korpus-Ressourcen mit einbezogen werden. Die so entstandenen Ressourcen werden in mehreren Formaten unter einer Creative-Commons-Lizenz frei zur Verfügung gestellt.

Das RIDGES-Projekt wird von einem Google Digital Humanities Research Award unterstützt. Die originalen Faksimilia der Texte wurden in den meisten Fällen durch Google Books erhoben.

Forschungsfrage: In unserem Projekt sollen wissenschaftliche Texte auf allen sprachlichen Ebenen (Syntax, Wortbildung, Lexik, Phraseologie, Textstruktur etc.) analysiert werden, um Entwicklungen und Tendenzen identifizieren und beschreiben zu können. In einem variationistischen Ansatz annotieren wir wissenschaftliche Texte mit korpuslinguistischen Methoden und nutzen quantitative Verfahren, um Veränderungen in den Merkmalen zu identifizieren.

Version: 6.0

Korpusarchitektur: Das diachrone Korpus Ridges Herbology besitzt eine multiple Segmentierung, die es erlaubt, konfligierende Tokenisierung von multiplen Ebenen zu verarbeiten (Krause et al. 2012). Das Korpus ist mit Spannenannotationen versehen.

Art der Änderung: Folgende Dokumente wurden hinzugefügt:
GartDerGesundheit-VR_1487_vonCuba
ArtzneyBuchleinDerKreutter-VR_1532_Tallat
ContrafaytKreuterbuch-VR_1532_Brunfels

ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels
NewKreuetterBuch-VR_1539_Bock
WieSichMeniglich-VR_1557_vonBodenstein
AlchymistischePractic-VR_1603_Libavius
BlackwellischesKraeuterbuch_1750_Blackwell
Apothekerlexikon_1793_Hahnemann
GetreueDarstellungUndBeschreibung_1809_Hayne
EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous
GrossesIllustriertesKraeuterbuch_1860_Mueller
GemeinnuetzigesKraeuterbuch_1874_Siegmund
NatürlichePflanzenfamilien_1887_Engler

Folgende Annotationsebenen wurden hinzugefügt:

ppk_e1, ppk_e2, ppk_e3, sentence_end

- Änderung der Ebene „quote“
- basiert auf dipl-Token
- es gibt keinen Wert „quote“ mehr für die Markierung der Zitate, sondern den Wert „yes“, Nicht-Zitate bzw. unkenntliche werden mit „no“ annotiert

- Dokumente, die vorher diese Ebene besaßen, haben dafür die Ebenen „plant“ und „disease“ erhalten; der ehemalige term-Wert „h“ wurde als Wert „pl“ nach „plant“ überführt; der ehemalige term-Wert „d“ wurde als Wert „di“ nach „disease“ überführt

- folgende Ebenen wurden vollständig gelöscht:

p, p_rend, p_n, brace, brace_dir, pb_rend, div1-5, die dazugehörigen div_n/div_type, xml:id, list, list_type, pos_klein, head_n, head_rend, citation, dialekt, diachronie

- Umbenennung der Ebene typeface in „script“, der Wert „gothic“ heißt jetzt „blackletter“, der Wert „antiqua“ heißt jetzt „roman“

- Umbenennung aller Metadaten:

alter Name	neuer Name
autor	author
uebersetzer	translator
uebersetztAus	trans_from
datum	date
ort	place
verlag	publisher

bibl	bibl
version	version
herausgeber	editor
auflage	edition_first
band	issue
bereich	maintopic
register	register
thema	topic
einMehrspr	lingualism
originaldatum	orig_date
originalort	orig_place
repositorium	repository
sprachtyp	lang_type
sprachgebiet	lang_area
textgestaltung	text_type
gestaltungselemente	lyric_type
vorredeVorh	preface
wermutVorh	wormwood
kraeutermonographiesammlung	herb_sorting

Umbenennung aller Annotationsebenen mit deutschem Namen:

alter Name	neuer Name
erlaeuterung	comment
referenz	reference
nomen_nominativ	noun_nom
bemerkung_lexik	comment_lex
bemerkung_syntax	comment_syn
form_bezugskategorie	form_referent
form_des_relativpronomens	form_relpron
form_krankheit	form_disease
form_nomen	form_noun
form_zubereitung	form_prep
komp	comp
komp_orth	comp_orth
KOUS_Semantik	KOUS_sem
kraeutername_normiert	herbname_norm
kraeuterzubereitung	herbprep
modifikation_bezugskategorie	mod_referent
Nebensatztyp	subclause_type
personenname	persname
position_im_satz	position_rel

position_zur_bezugskategorie	position_referent
satztyp	clause_type
strD	morph_ellipsis
Verbposition	position_verb
verbstellung	position_verb_rel
werkname	title

- leere Zellen in der Ebene „lang“ wurden mit dem Wert „deu“ aufgefüllt

- Die Werte „include“ und „self“ in author_ref existieren nicht mehr. Die meisten wurden ersetzt durch andere mögliche Werte, z.B. „pron1sg“ oder „pron1pl“.

-Der Wert „address“ in reader_ref existiert nicht mehr und wurde durch andere mögliche Werte ersetzt, z.B. „reader“.

- grobe manuelle Korrektur der Ebenen „author_ref“ und „reader_ref“, hi, hi_rend, head und script in allen Dateien

- Ersetzung des Unicodes 0292 durch „z“ in AlchymistischePractic_1603_Libavius

- manuelle Korrektur kleinerer Fehler

- systematische manuelle Korrektur der lb-Annotationen in ContrafaytKreuterbuch_1532_Brunfels

Enthaltende Dokumente
(Kürzel):

BuchDerNatur_1482_vonMegenberg
 GartDerGesundheit_1487_vonCuba
 GartDerGesundheit-VR_1487_vonCuba
 ArtzneyBuchleinDerKreutter_1532_Tallat
 ArtzneyBuchleinDerKreutter-VR_1532_Tallat
 ContrafaytKreuterbuch_1532_Brunfels
 ContrafaytKreuterbuch-VR_1532_Brunfels
 ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels
 NewKreuetterBuch_1539_Bock
 NewKreuetterBuch-VR_1539_Bock
 NewKreuterbuch_1543_Fuchs

WieSichMeniglich_1557_vonBodenstein
 WieSichMeniglich-VR_1557_vonBodenstein
 NewKreuterbuch_1563_Handsch
 Paradeiszgaertlein_1588_Rosbach
 AlchymistischePractic_1603_Libavius
 AlchymistischePractic-VR_1603_Libavius
 HortulusSanitatis_1609_Uffenbach
 Krautterbuch_1609_Carrichter
 PflantzGart-c4_1639_Rhagor
 PflantzGart_1639_Rhagor
 PflantzGart-VR_1639_Rhagor
 Wund-Artzney_1652_Greiff
 Phythologia_1662_Becher
 ThesaurusSanitatis_1673_Nasser
 SonderbaresKraeuterbuch-1-11_1675_Anonymous
 SonderbaresKraeuterbuch-11-21_1675_Anonymous
 SchweizerischerBotanicus_1687_vonRoll
 TheatrumBotanicum_1696_Verzascha
 ViridariumReformatum_1719_Valentini
 FloraSaturnizans_1722_Henckel
 MysterivmSigillorvm_1735_Hiebner
 BlackwellischesKraeuterbuch_1750_Blackwell
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder
 Unterricht_1774_Eisen
 GrundrissKraeuterkunde_1792_Willdenow
 Apothekerlexikon_1793_Hahnemann
 GetreueDarstellungUndBeschreibung_1809_Hayne
 Flora-6_1821_Wilbrand
 Flora-7_1821_Sieber
 EigenschaftenAllerHeilpflanzen_1828_Anonymous
 EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous
 NochEinigeWorte_1840_Meyen
 VorlesungenUeberKraeuterkunde_1843_Link
 GrossesIllustriertesKraeuterbuch_1860_Mueller
 DeutschePflanzennamen_1870_Grassmann
 FloraDerPreussischenRheinlande_1870_Wirtgen
 GemeinnuetzigesKraeuterbuch_1874_Siegmund
 NatürlichePflanzenfamilien_1887_Engler
 Kraeuterbuch_1914_Losch

2. Dokumente

Alle Kurztitel werden, den Standards der ‚clean‘-Normalisierung gemäß, nach Originalschreibweise realisiert, wobei Sonderzeichen, die der modernen Orthografie nicht entsprechen an den modernen

Zeichensatz angepasst werden. Umlaute werden durch „ae“, „ue“ und „oe“ ersetzt.

Typ: Documents

2.1. Kraeuterbuch_1914_Losch

Titel: Kräuterbuch: Unsere Heilpflanzen in Wort und Bild
Kürzel: Kraeuterbuch_1914_Losch
Autor: Friedrich Losch
Jahr: 1914
Ort: Eßlingen, München
Verlag: Schreiber
Seitenangabe: 163-168
Publikationstyp: Monographie
dipl-Einheiten: 3961
Beschreibung: Kräutertext
Annotationsebenen: author_ref, comment, figure, hi_rend, lang, lb, lemma, note, pb, pb_n, persname, pos, script, title

2.2. NatürlichePflanzenfamilien_1887_Engler

Titel: Die natürlichen Pflanzenfamilien
Kürzel: NatürlichePflanzenfamilien_1887_Engler
Autor: Adolf Engler
Jahr: 1887
Ort: Leipzig
Verlag: Wilhelm Engelmann
Seitenangabe: 280-286
Publikationstyp: Monographie
dipl-Einheiten: 3828
Beschreibung: Kräutertext
Annotationsebenen: author_ref, comment, figure, head, hi_rend, lang, lb, lemma, pb, pb_n, persname, pos, quote, script

2.3. GemeinnuetzigesKraeuterbuch_1874_Siegmund

Titel:	Gemeinnütziges Kräuterbuch
Kürzel:	GemeinnuetzigesKraeuterbuch_1874_Siegmund
Autor:	Ferdinand Siegmund
Jahr:	1874
Ort:	Wien und Pest
Verlag:	Hartleben
Seitenangabe:	316-325
Publikationstyp:	Monographie
dipl-Einheiten:	4527
Beschreibung:	Kräutertext
Annotationsebenen:	author_ref, comment, figure, hi_rend, lang, lb, lemma, pb, pb_n, persname, pos, quote, reader_ref, script

2.4. FloraDerPreussischenRheinlande_1870_Wirtgen

Titel:	Flora Der Preussischen Rheinlande
Kürzel:	FloraDerPreussischenRheinlande_1870_Wirtgen
Autor:	Wirtgen, Philipp Wilhelm
Jahr:	1870
Ort:	Bonn
Verlag:	NA
Seitenangabe:	1-22
Publikationstyp:	Monographie
dipl-Einheiten:	7519
Beschreibung:	Kräutertext
Annotationsebenen:	attr_gen, author_ref, comment, comp, comp_orth, figure, foreign, head, hi, hi_rend, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, plant, pos, position_verb, ppk_e1, ppk_e2, prot, quote, ref, ref_target, script, subclause_type

2.5. DeutschePflanzennamen_1870_Grassmann

Titel:	Deutsche Pflanzennamen
Kürzel:	DeutschePflanzennamen_1870_Grassmann
Autor:	Grassmann, Hermann
Jahr:	1870
Ort:	Stettin
Verlag:	NA
Seitenangabe:	1-23
Publikationstyp:	Monographie
dipl-Einheiten:	10282
Beschreibung:	Kräutertext
Annotationsebenen:	atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, definition, foreign, foreign_trans, head, hi, hi_rend, KOUS_sem, lang, lb, lemma, morph_ellipsis, name, name_type, pb, pb_n, plant, pos, position_verb, property, prot, quote, script, subclause_type, unclear

2.6. VorlesungenUeberKraeuterkunde_1843_Link

Titel:	Vorlesungen Über Kräuterkunde
Kürzel:	VorlesungenUeberKraeuterkunde_1843_Link
Autor:	Link , Heinrich Friedrich
Jahr:	1843
Ort:	Berlin
Verlag:	Verlag von C. G. Lüderitz
Seitenangabe:	28-48
dipl-Einheiten:	7995
Beschreibung:	Kräutertext
Annotationsebenen:	attr_gen, author_ref, comment, comp, comp_orth, foreign, foreign_trans, head, hi, hi_rend, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, plant, pos, position_verb, ppk_e1,

2.7. GrossesIllustriertesKraeuterbuch_1860_Mueller

Titel:	Das große illustrierte Kräuter-Buch
Kürzel:	GrossesIllustriertesKraeuterbuch_1860_Mueller
Autor:	Ferdinand Müller
Jahr:	1860
Ort:	Ulm
Verlag:	NA
Seitenangabe:	617-622
Publikationstyp:	Monographie
dipl-Einheiten:	3852
Beschreibung:	Kräutertext
Annotationsebenen:	author_ref, comment, figure, head, hi_rend, lang, lb, lemma, pb, pb_n, persname, pos, quote, reader_ref, script

2.8. NochEinigeWorte_1840_Meyen

Titel:	Noch einige Worte über den Befruchtungsakt und die Poljembryonie bei den höheren Pflanzen
Kürzel:	NochEinigeWorte_1840_Meyen
Autor:	Meyen, Franz Julius Ferdinand
Jahr:	1840
Ort:	Berlin
Verlag:	Haude und Spenersche Buchhandlung (S. J. Joseephy.)
Seitenangabe:	24-50
Publikationstyp:	Monographie
dipl-Einheiten:	7984
Beschreibung:	Kräutertext
Annotierte Ebenen:	attr_gen, author_ref, comp, comp_orth, foreign, foreign_trans, head, hi, hi_rend, KOUS_sem, lang, lb, lemma, morph_ellipsis, name, name_type, note, pb, pb_n, plant, pos, position_verb, ppk_e1, ppk_e2, ppk_e3, property, prot, quote, reader_ref, ref, ref_target, ref_type, script,

subclause_type

2.9. EigenschaftenAllerHeilpflanzen_1828_Anonymous

Titel: Die Eigenschaften Aller Heilpflanzen
Kürzel: EigenschaftenAllerHeilpflanzen_1828_Anonymous
Autor: NA
Jahr: 1828
Ort: München
Verlag: Jos. Lindauer'sche Buchhandlung
Seitenangabe: 3-21
dipl-Einheiten: 6849
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, comment, comp, comp_orth, disease, foreign, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, prot, quote, script, subclause_type, unclear

2.10. EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous

Titel: Die Eigenschaften Aller Heilpflanzen
Kürzel: EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous
Autor: NA
Jahr: 1828
Ort: München
Verlag: Jos. Lindauer'sche Buchhandlung
Seitenangabe: 149-165
Publikationstyp: Monographie
dipl-Einheiten: 5561
Beschreibung: Kräutertext
Annotationsebenen: comment, head, hi_rend, lang, lb, lemma, pb, pb_n, pos, quote, script

2.11. Flora-7_1821_Sieber

Titel:	Flora oder Botanische Zeitung 7
Kürzel:	Flora-7_1821_Sieber
Autor:	Sieber
Jahr:	1821
Ort:	Regensburg
Verlag:	Königl. botanische Gesellschaft
Seitenangabe:	97-112
Publikationstyp:	Zeitschrift
dipl-Einheiten:	3636
Beschreibung:	Kräutertext
Annotierte Ebenen:	atLeast, atMost, attr_gen, author_ref, comp, comp_orth, foreign, foreign_trans, head, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, ppk_e2, property, prot, quote, script, subclause_type, unclear

2.12. Flora-6_1821_Wilbrand

Titel:	Flora oder Botanische Zeitung 6
Kürzel:	Flora-6_1821_Wilbrand
Autor:	B. Wilbrand
Jahr:	1821
Ort:	Regensburg
Verlag:	Königl. botanische Gesellschaft
Seitenangabe:	81-96
Publikationstyp:	Zeitschrift
dipl-Einheiten:	3312
Beschreibung:	Kräutertext

Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comp, comp_orth, definition, foreign, foreign_trans, head, interpretation, KOUS_sem, lang, lb, lemma, note, pb, pb_n, plant, pos, position_verb, ppk_e1, property, prot, quote, reader_ref, ref, ref_target, ref_type, script, subclause_type, unclear

2.13. GetreueDarstellungUndBeschreibung_1809_Hayne

Titel: Getreue Darstellung und Beschreibung der in der Arzneykunde gebräuchlichen Gewächse, wie auch solcher, welche mit ihnen verwechselt werden können

Kürzel: GetreueDarstellungUndBeschreibung_1809_Hayne

Autor: Friedrich Gottlob Hayne

Jahr: 1809

Ort: Berlin

Verlag: NA

Seitenangabe: NA

Publikationstyp: Monographie

dipl-Einheiten: 3177

Beschreibung: Kräutertext

Annotationsebenen: author_ref, comment, figure, head, hi_rend, lang, lb, lemma, pb, persname, pos, quote, script, title, unclear

2.14. Apothekerlexikon_1793_Hahnemann

Titel: Apothekerlexikon

Kürzel: Apothekerlexikon_1793_Hahnemann

Autor: Samuel Hahnemann

Jahr: 1793

Ort: Leipzig

Verlag: NA

Seitenangabe: 437-446

Publikationstyp: Monographie

dipl-Einheiten: 4244

Beschreibung: Kräutertext
Annotationsebenen: comment, head, hi_rend, lang, lb, lemma, pb, pb_n, pos, quote, script

2.15. GrundrissKraeuterkunde_1792_Willdenow

Titel: Grundriss der Kräuterkunde zu Vorlesungen entworfen
Kürzel: GrundrissKraeuterkunde_1792_Willdenow
Autor: Willdenow, Carl Ludwig
Jahr: 1792
Ort: Berlin
Verlag: Haude und Spener
Seitenangabe: 282-305
Publikationstyp: Monographie
dipl-Einheiten: 5094
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comp, comp_orth, foreign, foreign_trans, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, property, prot, quote, reader_ref, script, subclause_type, unclear

2.16. Unterricht_1774_Eisen

Titel: Unterricht von der allgemeinen Kräuter- und Wurzeltrocknung
Kürzel: Unterricht_1774_Eisen
Autor: Eisen, Johann Georg
Jahr: 1774
Ort: Riga
Verlag: NA
Seitenangabe: 44-72
Publikationstyp: Monographie

dipl-Einheiten: 3982
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, definition, disease, foreign, foreign_trans, head, hi, hi_rend, interpretation, item, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, property, prot, quote, reader_ref, script, subclause_type, unclear

2.17. EinleitungZuDerKraeuterkenntnisz_1764_Oeder

Titel: Einleitung Zu Der Kräuterkenntnisz
Kürzel: EinleitungZuDerKraeuterkenntnisz_1764_Oeder
Autor: Oeder , George Christian
Jahr: 1764
Ort: Kopenhagen
Seitenangabe: 304-339
dipl-Einheiten: 7115
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, foreign, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, ppk_e2, ppk_e3, prot, quote, reader_ref, script, subclause_type, unclear

2.18. BlackwellischesKraeuterbuch_1750_Blackwell

Titel: Vermehrtes und verbessertes Blackwellisches Kräuter-Buch, das ist Elisabeth Blackwell Sammlung der Gewächse die zum Arzney-Gebrauch in den Apothecken aufbehalten werden, deren Beschreibung und Kräfften
Kürzel: BlackwellischesKraeuterbuch_1750_Blackwell
Autor: Elisaneth Blackwell
Jahr: ca. 1750
Ort: Nürnberg
Verlag: Nicolaus Friedrich Eisenberger

Seitenangabe: NA
Publikationstyp: Monographie
dipl-Einheiten: 5609
Beschreibung: Kräutertext
Annotationsebenen: author_ref, comment, head, hi_rend, lang, lb, lemma, pb, persname, pos, quote, script, title, unclear

2.19. MysterivmSigillorvm_1735_Hiebner

Titel: Mysterivm Sigillorvm
Kürzel: MysterivmSigillorvm_1735_Hiebner
Autor: Hiebner von Schneebergk, Israel
Jahr: 1735
Ort: Erfurt
Seitenangabe: 65-89
dipl-Einheiten: 7864
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, disease, foreign, foreign_trans, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, plant, pos, position_verb, ppk_e1, prot, quote, reader_ref, ref, ref_target, script, subclause_type, unclear

2.20. FloraSaturnizans_1722_Henckel

Titel: Flora saturnizans, Die Verwandschafft des Pflanzen mit dem Mineral Reich
Kürzel: FloraSaturnizans_1722_Henckel
Autor: Henckel, Johann Friedrich
Jahr: 1722

Ort: Leipzig
 Verlag: Johann Christian Martini
 Seitenangabe: 647-671
 Publikationstyp: Monographie
 dipl-Einheiten: 6219
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comp, comp_orth, definition, foreign, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, plant, pos, position_verb, ppk_e1, property, prot, quote, reader_ref, ref, ref_target, ref_type, script, subclause_type, unclear

2.21. ViridariumReformatum_1719_Valentini

Titel: Viridarium Reformatum, Seu Regnum Vegetabile, Das ist: Neu-eingerichtetes und Vollständiges Kräuter-Buch
 Kürzel: ViridariumReformatum_1719_Valentini
 Autor: Michael Bernhard Valentini
 Jahr: 1719
 Ort: Frankfurt am Main
 Verlag: Heinscheidt
 Seitenangabe: 459-465
 Publikationstyp: Monographie
 dipl-Einheiten: 3596
 Beschreibung: Kräutertext
 Annotationsebenen: author_ref, comment, figure, head, hi_rend, lang, lb, lemma, pb, pb_n, persname, pos, reader_ref, script, title

2.22. TheatrumBotanicum_1696_Verzascha

Titel: Theatrum Botanicum. Das ist: Neu Vollkommenes Kräuter-Buch: Worinnen

Allerhand Erdgewächse der Bäumen, Stauden und Kräutern, welche in allen vier Theilen der Welt, sonderlich aber in Europa herfür kommen, neben ihren Fürtrefflichen Würckungen, und deren Gebrauch, wider allerley Kranckheiten an Menschen und Vieh, Mit sonderbahrem Fleiß auff eine ganz neue Art und Weise beschrieben, Auch mit schönen, theils neuen Figuren geziert, und neben denen ordenlichen, so wohl Kräuter- als Kranckheit-Registern, mit nutzlichen Marginalien vorgestellt sind. Allen Aerzten sonderlich auch denen auff dem Land wohnenden höchst nützlich und ergetzlich

Kürzel: TheatrumBotanicum_1696_Verzascha
 Autor: Bernhard Verzascha
 Jahr: 1696
 Ort: Basel
 Verlag: Bertsch [u.a.]
 Seitenangabe: 659-664
 Publikationstyp: Monographie
 dipl-Einheiten: 3893
 Beschreibung: Kräutertext
 Annotationsebenen: author_ref, comment, figure, head, hi_rend, lang, lb, lemma, note, pb, pb_n, persname, pos, reader_ref, script, title

2.23. SchweizerischerBotanicus_1687_vonRoll

Titel: Der Schweizerische Botanicus
 Kürzel: SchweizerischerBotanicus_1687_vonRoll
 Autor: Timotheus von Roll
 Jahr: 1687
 Ort: NA
 Verlag: Franz Carl Roß
 Seitenangabe: 54-79
 Publikationstyp: Monographie
 dipl-Einheiten: 3510
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comp, comp_orth, head, hi_rend,

interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, pos, position_verb, ppk_e1, prot, reader_ref, script, subclause_type, unclear

Anmerkung: Aus der Originaltranskription der Studierende wurden 2 Seiten gelöscht, diese waren im der Transkription zugrunde liegenden PDF doppelt vorhanden und wurden von den Studierende auch doppelt annotiert.

2.24. SonderbaresKraeuterbuch-11-21_1675_Anonymous

Titel: Curioser Botanicus oder sonderbares Kräuter Buch
Kürzel: SonderbaresKraeuterbuch-11-21_1675_Anonymous
Autor: NA
Jahr: 1675
Ort: NA
Verlag: NA
Seitenangabe: 11-21
Publikationstyp: Monographie
dipl-Einheiten: 1907
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, comment, comp, comp_orth, disease, foreign, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, property, prot, quote, script, subclause_type, unclear

2.25. SonderbaresKraeuterbuch-1-11_1675_Anonymous

Titel: Curioser Botanicus oder sonderbares Kräuter Buch
Kürzel: SonderbaresKraeuterbuch-1-11_1675_Anonymous
Autor: NA
Jahr: 1675
Ort: NA
Verlag: NA

Seitenangabe: 1-11
 Publikationstyp: Monographie
 dipl-Einheiten: 2245
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, comment, comp, comp_orth, disease, foreign, foreign_trans, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_ana, pb_n, plant, pos, position_verb, ppk_e1, property, prot, quote, script, subclause_type, unclear

2.26. ThesaurusSanitatis_1673_Nasser

Titel: Thesaurus Sanitatis
 Kürzel: ThesaurusSanitatis_1673_Nasser
 Autor: Adrian Nasser
 Jahr: 1673
 Ort: Nürnberg
 Verlag: Johann Hoffmann
 Seitenangabe: 497-524
 Publikationstyp: Monographie
 dipl-Einheiten: 7041
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, disease, foreign, foreign_trans, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, prot, quote, reader_ref, script, subclause_type, unclear

2.27. Phythologia_1662_Becher

Titel: Phythologia; Das ist: Deß erläuterten Medicinalischen Parnassi Ander Theil, Nemlich das Kräuter-Buch. Band 2.
 Kürzel: Phythologia_1662_Becher
 Autor: Johann Joachim Becher

Jahr: 1662
 Ort: Ulm
 Verlag: Johann Görlin
 Seitenangabe: 305-319
 Publikationstyp: Monographie
 dipl-Einheiten: 4321
 Beschreibung: Kräutertext
 Annotationsebenen: comment, figure, head, hi_rend, lang, lb, lemma, pb, pb_n, pos, script, title

2.28. Wund-Artzney_1652_Greiff

Titel: Wund-Artzney
 Kürzel: Wund-Artzney_1652_Greiff
 Autor: Guilelmus Fabricius Hildanus, Friedrich Greiff
 Jahr: 1652
 Ort: Frankfurt am Main
 Verlag: Johann Beyer
 Seitenangabe: 218-223
 Publikationstyp: Monographie
 dipl-Einheiten: 5253
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, disease, foreign, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, plant, pos, position_verb, ppk_e1, prot, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear

2.29. PflantzGart-VR_1639_Rhagor

Titel: Pflantz-Gart (VR)
 Kürzel: PflantzGart-Vorrede_1639_Rhagor
 Autor: Daniel Rhagor

Jahr: 1639
 Ort: Bern
 Verlag: Ben Stephan Schmid
 Seitenangabe: 1-10
 Publikationstyp: Monographie
 dipl-Einheiten: 2318
 Beschreibung: Kräutertext
 Annotierte Ebenen: attr_gen, author_ref, comment, comp, comp_orth, foreign, foreign_trans, head, hi, hi_rend, KOUS_sem, lang, lb, lemma, note, pb, pb_n, pos, position_verb, prot, quote, reader_ref, reference, script, sentence_end, subclause_type

2.30. PflantzGart-c4_1639_Rhagor

Titel: Pflantz-Gart (Capitel 4)
 Kürzel: PflantzGart-c4_1639_Rhagor
 Autor: Daniel Rhagor
 Jahr: 1639
 Ort: Bern
 Verlag: Ben Stephan Schmid
 Seitenangabe: 33-45
 Publikationstyp: Monographie
 dipl-Einheiten: 2846
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, foreign, foreign_trans, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, plant, pos, position_verb, ppk_e1, property, prot, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear

2.31. PflantzGart_1639_Rhagor

Titel: Pflantz-Gart

Kürzel: PflantzGart_1639_Rhagor
 Autor: Daniel Rhagor
 Jahr: 1639
 Ort: Bern
 Verlag: Ben Stephan Schmid
 Seitenangabe: 92-110
 Publikationstyp: Monographie
 dipl-Einheiten: 4067
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, figure, head, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, pos, position_verb, ppk_e1, prot, reference, script, sentence_end, subclause_type, unclear

2.32. Krauterbuch_1609_Carrichter

Titel: Kräuterbuch des Edelen und hochgelehrten herren Bartholomei Carrichters
 Kürzel: Krauterbuch_1609_Carrichter
 Autor: Bartholomeus Carrichter
 Jahr: 1609
 Ort: Straßburg
 Verlag: Antonium Bertram
 Seitenangabe: 47-75
 Publikationstyp: Monographie
 dipl-Einheiten: 4992
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, disease, head, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, prot, reader_ref, reference, script, sentence_end, subclause_type, unclear

2.33. HortulusSanitatis_1609_Uffenbach

Titel:	Hortulus Sanitatis
Kürzel:	HortulusSanitatis_1609_Uffenbach
Autor:	Peter Uffenbach
Jahr:	1609
Ort:	Frankfurt am Main
Verlag:	Ionae Rhodij
Seitenangabe:	1-21
Publikationstyp:	Monographie
dipl-Einheiten:	6516
Beschreibung:	Kräutertext
Annotierte Ebenen:	atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, disease, foreign, foreign_trans, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, plant, pos, position_verb, ppk_e1, prot, quote, reference, script, sentence_end, subclause_type, unclear

2.34. AlchymistischePractic_1603_Libavius

Titel:	Alchymistische Practic
Kürzel:	AlchymistischePractic_1603_Libavius
Autor:	Andreas Libavius
Jahr:	1603
Ort:	Frankfurt
Verlag:	Johann Saur
Seitenangabe:	5-26
Publikationstyp:	Monographie
dipl-Einheiten:	5063
Beschreibung:	Kräutertext
Annotierte Ebenen:	atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, definition, figure, figure_rend, foreign, foreign_trans, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb,

pb_n, pos, position_verb, ppk_e1, ppk_e2, property, prot, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear

2.35. Alchymistische Practic-VR_1603_Libavius

Titel: Alchymistische Practic (Vorrede)
Kürzel: AlchymistischePractic-VR_1603_Libavius
Autor: Andreas Libavius
Jahr: 1603
Ort: Frankfurt
Verlag: Johann Saur
Seitenangabe: 3-4
Publikationstyp: Monographie
dipl-Einheiten: 441
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, foreign, head, hi, lang, lb, lemma, pb, pb_n, persname, ppk_e1, pos, quote, reader_ref, reference, script, sentence_end, title

2.36. Paradeiszaertlein_1588_Rosbach

Titel: Paradeiszgärtlein
Kürzel: Paradeiszaertlein_1588_Rosbach
Autor: Conrad Rosbach
Jahr: 1588
Ort: Frankfurt am Main
Verlag: Johann Spieß
Seitenangabe: 1-43
Publikationstyp: Monographie
dipl-Einheiten: 5055
Beschreibung: Kräutertext

Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, disease, figure, head, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1, prot, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear

2.37. NewKreuterbuch_1563_Handsch

Titel: New Kreüterbuch: Mit den allerschönsten vnd artlichsten Figuren aller Gewechß, dergleichen vormals in keiner sprach nie an tag kommen

Kürzel: NewKreuterbuch_1563_Handsch

Autor: Georg Handsch

Jahr: 1563

Ort: Prag

Verlag: Melantrich von Auentin und Valgriß

Seitenangabe: 277-283

Publikationstyp: Monographie

dipl-Einheiten: 3785

Beschreibung: Kräutertext

Annotationsebenen: author_ref, comment, figure, head, hi_rend, lang, lb, lemma, pb, pb_n, persname, pos, reader_ref, reference, script, sentence_end

2.38. WieSichMeniglich_1557_vonBodenstein

Titel: Wie sich meniglich von dem Cyperlin, so Podagra genennet, hüten soll

Kürzel: WieSichMeniglich_1557_vonBodenstein

Autor: von Bodenstein, Adam

Jahr: 1557

Ort: Basel

Verlag: Bartholome Stähälin

Seitenangabe: 28-47

Publikationstyp: Monographie
dipl-Einheiten: 5508
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth, disease, foreign, foreign_trans, head, herbyname_norm, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, persname, plant, pos, position_verb, ppk_e1, prot, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear

2.39. WieSichMeniglich-VR_1557_vonBodenstein

Titel: Wie sich meniglich von dem Cyperlin, so Podagra genennet, hüten soll (Vorrede)
Kürzel: WieSichMeniglich-VR_1557_vonBodenstein
Autor: von Bodenstein, Adam
Jahr: 1557
Ort: Basel
Verlag: Bartholome Stähälin
Seitenangabe: NA
Publikationstyp: Monographie
dipl-Einheiten: 706
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, head, hi_rend, lang, lb, lemma, pb, persname, pos, quote, reader_ref, script, sentence_end

2.40. NewKreuterbuch_1543_Fuchs

Titel: New Kreüterbuch
Kürzel: NewKreuterbuch_1543_Fuchs
Autor: Fuchs, Leonhard

Jahr: 1543
 Ort: Basel
 Verlag: Michael Isingrin
 Seitenangabe: 2-e4
 Publikationstyp: Monographie
 dipl-Einheiten: 5221
 Beschreibung: Kräutertext
 Annotierte Ebenen: attr_gen, author_ref, comment, comp, comp_orth, definition, disease, figure, figure_rend, foreign, foreign_trans, head, herbname_norm, hi, hi_rend, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, persname, plant, pos, position_verb, ppk_e1, property, prot, quote, reader_ref, script, sentence_end, subclause_type

2.41. NewKreuetterBuch_1539_Bock

Titel: New Kreütter Buch: von Underscheydt, Würckung und Namen der Kreütter so in teütschen Landen wachsen; auch der selbigen eygentlichem und wolgegründtem Gebrauch in der Artzney, zu behalten und zu fürdern Leibs Gesuntheyt fast nutz und tröstlichen, vorab gemeynem Verstand
 Kürzel: NewKreuetterBuch_1539_Bock
 Autor: Hieronymus Bock
 Jahr: 1539
 Ort: Straßburg
 Verlag: Rihel
 Seitenangabe: XCVI-C
 Publikationstyp: Monographie
 dipl-Einheiten: 4408
 Beschreibung: Kräutertext
 Annotationsebenen: author_ref, comment, head, hi_rend, lang, lb, lemma, note, pb, pb_n, persname, pos, reader_ref, reference, script, sentence_end, title, unclear

2.42. NewKreuetterBuch-VR_1539_Bock

Titel: New Kreütter Buch: von Underscheydt, Würckung und Namen der Kreütter so in teütschen Landen wachsen; auch der selbigen eygentlichem und wolgegründtem Gebrauch in der Artzney, zu behalten und zu fürdern Leibs Gesuntheyt fast nutz und tröstlichen, vorab gemeynem Verstand (Vorrede)

Kürzel: NewKreuetterBuch_1539_Bock

Autor: Hieronymus Bock

Jahr: 1539

Ort: Straßburg

Verlag: Rihel

Seitenangabe: VII-X

Publikationstyp: Monographie

dipl-Einheiten: 2021

Beschreibung: Kräutertext

Annotationsebenen: author_ref, comment, figure, head, hi_rend, lang, lb, lemma, note, pb, pb_n, persname, pos, quote, reader_ref, script, sentence_end, title, unclear

2.43. ContrafaytKreuterbuch_1532_Brunfels

Titel: Contrafayt kreüterbuch

Kürzel: ContrafaytKreuterbuch_1532_Brunfels

Autor: Otto Brunfels

Jahr: 1532

Ort: Straßburg

Verlag: Hans Schotten

Seitenangabe: 312-323

Publikationstyp: Monographie

dipl-Einheiten: 4359

Beschreibung: Kräutertext

Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, comp, comp_orth,

disease, figure, head, herbyname_norm, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, mod_referent, pb, pb_n, persname, plant, pos, position_verb, ppk_e1, prot, reader_ref, script, sentence_end, subclause_type, unclear

2.44. ContrafaytKreuterbuch- CCXXXVII-CCXLVIII_1532_Brunfels

Titel: Contrafayt kreüterbuch (S. CCXXXVII-CCXLVIII)
Kürzel: ContrafaytKreuterbuch- CCXXXVII-CCXLVIII_1532_Brunfels
Autor: Otto Brunfels
Jahr: 1532
Ort: Straßburg
Verlag: Hans Schotten
Seitenangabe: CCXXXVII-CCXLVIII
Publikationstyp: Monographie
dipl-Einheiten: 3988
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, figure, hi_rend, lang, lb, lemma, pb, pb_n, persname, pos, quote, reader_ref, reference, script, sentence_end

2.45. ContrafaytKreuterbuch-VR_1532_Brunfels

Titel: Contrafayt kreüterbuch (Vorrede)
Kürzel: ContrafaytKreuterbuch-VR_1532_Brunfels
Autor: Otto Brunfels
Jahr: 1532
Ort: Straßburg
Verlag: Hans Schotten
Seitenangabe: VIII-IX
Publikationstyp: Monographie

dipl-Einheiten: 664
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, figure, head, lang, lb, lemma, note, pb, pb_n, persname, pos, quote, script, sentence_end, title

2.46. ArtzneyBuchleinDerKreutter_1532_Tallat

Titel: Artzney Buchlein der kreutter
Kürzel: ArtzneyBuchleinDerKreutter_1532_Tallat
Autor: Johannes Tallat
Jahr: 1532
Ort: Leipzig
Verlag: Michael Blum
Seitenangabe: NA
Publikationstyp: Monographie
dipl-Einheiten: 4630
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, clause_type, comment, comment_lex, comp, comp_orth, disease, form_disease, form_noun, form_prep, form_referent, form_relpron, head, herbname_norm, herbprep, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, mod_referent, note, noun_nom, pb, persname, plant, pos, position_referent, position_rel, position_verb, position_verb_rel, ppk_e1, problem, prot, reader_ref, script, sentence_end, subclause_type, title, unclear

2.47. ArtzneyBuchleinDerKreutter-VR_1532_Tallat

Titel: Artzney Buchlein der kreutter (Vorrede)
Kürzel: ArtzneyBuchleinDerKreutter-VR_1532_Tallat
Autor: Johannes Tallat

Jahr: 1532
 Ort: Leipzig
 Verlag: Michael Blum
 Seitenangabe: NA
 Publikationstyp: Monographie
 dipl-Einheiten: 270
 Beschreibung: Kräutertext
 Annotierte Ebenen: author_ref, figure, head, hi_rend, lang, lb, lemma, pb, persname, pos, quote, reader_ref, script, sentence_end

2.48. GartDerGesundheit_1487_vonCuba

Titel: Gart der Gesundheit
 Kürzel: GartDerGesundheit_1487_vonCuba
 Autor: Johannes von Cuba
 Jahr: 1487
 Ort: Ulm
 Verlag: NA
 Seitenangabe: NA
 Publikationstyp: Monographie
 dipl-Einheiten: 4700
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, clause_type, comment, comment_lex, comp, comp_orth, disease, figure, form_disease, form_noun, form_prep, form_referent, form_relpron, head, herbname_norm, herbprep, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, mod_referent, noun_nom, pb, persname, plant, pos, position_referent, position_rel, position_verb, position_verb_rel, ppk_e1, ppk_e2, problem, prot, reader_ref, script, sentence_end, subclause_type, title, unclear

2.49. GartDerGesundheit-VR_1487_vonCuba

Titel: Gart der Gesundheit (Vorrede)
Kürzel: GartDerGesundheit-VR_1487_vonCuba
Autor: Johannes von Cuba
Jahr: 1487
Ort: Ulm
Verlag: NA
Seitenangabe: aij-aiij
Publikationstyp: Monographie
dipl-Einheiten: 1352
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, head, lang, lb, lemma, pb, pb_n, persname, pos, reader_ref, script, sentence_end, title, unclear

2.50. BuchDerNatur_1482_vonMegenberg

Titel: Das Buch der Natur
Kürzel: BuchDerNatur_1482_vonMegenberg
Autor: Conradus von Megenberg
Jahr: 1482
Ort: Augsburg
Verlag: NA
Seitenangabe: NA
Publikationstyp: Monographie
dipl-Einheiten: 5215
Beschreibung: Kräutertext
Annotationsebenen: author_ref, comment, head, hi_rend, lang, lb, lemma, pb, persname, pos, reader_ref, reference, script, sentence_end

3. Annotationsebenen – Transkription/Normalisierung

3.1. dipl

Typ: *Layer* - dipl

Name:	dipl
Dokumentation:	Siehe Ridges-Website: http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: Siehe Anhang
dipl-Einheiten:	182738
Beschreibung:	Die Transkription von Faksimiles stellt für die korpuslinguistische Aufbereitung zumeist die grundlegende, diplomatische Ebene (dipl). Diese Ebene soll sich graphisch hinsichtlich Orthographie, Getrennt- und Zusammenschreibung und Sonderzeichen möglichst nah am zugrunde liegenden Faksimile orientieren. Grundsätzlich wird so auch entgegen modernen Orthographieregeln segmentiert oder transkribiert!

Typ: Preparationstep – dipl

Schritt:	1
Aufbereitung:	OCR/Transkript
Aufbereitungsart:	semi-automatisch/manuell
Tool:	NA
Format:	Text-Datei
Segmentierung der Annotation:	eigenständig
Qualitätsprüfer:	Korpling
Datum:	2015
Qualitätsprüfung:	manuell
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Tokenisierung
Aufbereitungsart:	Automatisch
Tool:	Treetagger Version 3.2

Format:	Treetagger-Output
Segmentierung der Annotation:	eigenständig
Qualitätsprüfer:	Studierende der Humboldt-Universität zu Berlin, korpling
Datum:	2015
Qualitätsprüfung:	manuelle Konsistenzprüfung
Editor:	Laura Perlitz
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2010 und Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	eigenständig
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

3.2. clean

Typ: *Layer* – clean

Name:	clean
Dokumentation:	Siehe Ridges-Website http://korpling.german.hu-berlin.de/ridges/index_de.html Beispiele: Siehe Anhang
clean-Einheiten:	177795

Beschreibung: Die clean-Ebene enthält erste vollautomatisch erstellte Normalisierungen hinsichtlich Sonderzeichen und graphischer Strukturierungen. So werden u.a. Ligaturen, die nach moderner Rechtschreibung nicht mehr verwendet werden, normalisiert. Graphische Markierungen der einzelnen Texte wie Zeilenumbrüche werden aufgelöst und Sonderzeichen einiger Fonts wie Fraktur auf die heutigen Schriftsätze übertragen. Die Dokumente, die seit der Version 4.0 dazugekommen sind, stellen neue Anforderungen an die automatisch erstellte clean-Ebene. Mit Tilden versehene Vokale können nicht länger kontextunabhängig normalisiert werden. Deshalb werden ab Version 4.0 für Token, die Vokale mit Tilden enthalten, alle möglichen Formen dieser Token in der clean angegeben. Die verschiedenen Formen werden durch ‚|‘ getrennt (zum Beispiel: *aufzwēdig* wird zu *auszwemdig|auszwendig*).

In der clean-Ebene werden außerdem Wörter, die durch einen Zeilenumbruch getrennt und mit einem Bindeelement versehen sind, zusammengezogen. Beginnt das zweite Wort mit einem Großbuchstaben, wird dieser in der clean-Ebene in Kleinschreibung realisiert. Ist der komplette zweite Bestandteil in Großbuchstaben geschrieben, bleibt dies so bestehen (*Gelb-Sucht* wird zu *Gelbsucht*; *MON- <lb> TANUM* wird zu *MONTANUM*).

Trunkierte Elemente, die am Zeilenende stehen, werden bisher nicht als solche erkannt und daher automatisch mit dem ersten Element der folgenden Zeile zusammengezogen (*Speiſzund Nahrungsſaffts* wird zu *Speiszund (sic!)Nahrungssaffts*).

Für eine komplette Auflistung aller Ersetzungen, die für die Erstellung der clean-Ebene gemacht wurden, siehe die Readme zum Skript, das zur Erstellung benutzt wurde.

Typ: Preparationstep – clean

Schritt:	1
Aufbereitung:	Normalisierung, Annotation
Aufbereitungsart:	Automatisch
Tool	cleanV22.py
Format:	Clean-Skript-Output
Segmentierung der Annotation:	Eigenständig
Qualitätsprüfer:	Laura Perlitz

Datum:	2016
Qualitätsprüfung:	manuelle Konsistenzprüfung
Editor:	Carolin Odebrecht
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	Automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	Eigenständig
Qualitätsprüfer:	Korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

3.3. norm

Typ: *Layer* – norm

Name:	Norm
Dokumentation:	Siehe Ridges-Website http://korpling.german.hu-berlin.de/ridges/index_de.html Beispiele: Siehe Anhang
norm-Einheiten:	176855
Beschreibung:	Die norm-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier Tokenisierung, Graphematik, Flexionsformen und Lexeme an den nhd. Sprachstand angepasst werden.
	Graphematik: orthographische Anpassung nach Duden kreutter → Kräuter

Phonologie:

zu beachten sind die fnhd. Lautwandelprozesse (Diphthongierung, Monophthongierung, Synkope, Apokope, Nukleussenkung etc.)

wänlîn → Wännlein

lehret → lehrt

Morphologie:

in die Nasen → in die Nase

das kal Haupt → das kahle Haupt

funden → gefunden

Lexikologie:

Ausgestorbenes lexikalisches Material wird mit modernen Orthographieregeln übernommen und ggf. in der Ebene Erläuterung erläutert..

Vergefz → Vergess ‚das Vergessen‘

Wortbildung:

Ausgestorbene Wortbildungsmorpheme werden nach Möglichkeit durch nhd. Bildungen ersetzt.

halben → halber

stachelecht → stachelig

Es erfolgt keine morphosyntaktische Normalisierung (z.B. innerhalb der Nominalphrase), keine Normalisierung des Genus und keine Normalisierung von syntaktischem Kasus.

(es heilt die) versehrte Därmelein → versehrte Därmelein (vs. nhd. die versehrten Därmelein)

das Milz → das Milz (vs. nhd. die Milz)

Satzanfänge werden großgeschrieben, wenn sie im Original auch großgeschrieben wurden. Satzanfänge bleiben kleingeschrieben, wenn sie im Original auch kleingeschrieben sind.

Beispiele und Einzelentscheidungen siehe Anhang.

Typ: Preparationstep – norm

Schritt: 1

Aufbereitung: Normalisierung, Annotation

Aufbereitungsart:	Manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	Eigenständig
Qualitätsprüfer:	Korpling
Datum:	2015
Qualitätsprüfung:	manuelle Korrektur
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	Automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	Eigenständig
Qualitätsprüfer:	Korpling
Datum:	2016
Qualitätsprüfung:	Gohar Schnelle, Laura Perlitz
Editor:	Carolin Odebrecht

4. Annotationsebenen – Annotationen zu linguistischen Eigenschaften

4.1. pos

Typ: *Layer* – pos

Name:	Pos
Dokumentation:	http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html

Beschreibung: Unter linguistischen Annotationen werden u. a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – pos

Schritt: 1
Aufbereitung: Part-of-Speech-Tagging
Aufbereitungsart: Automatisch
Tool: Treetagger Version 3.2
Format: Treetagger-Output
Segmentierung der Annotation: basiert auf ‘norm’
Qualitätsprüfer: Korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 2
Aufbereitung: unclear-POS-Tagging
Aufbereitungsart: Automatisch
Tool: ReplacePosOfUnclear.bas
Format: Excel 2013
Segmentierung der Annotation: basiert auf ‘norm’
Qualitätsprüfer: Carolin Odebrecht
Datum: 2016
Qualitätsprüfung: semi-automatisch
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: Automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS
Segmentierung der
Annotation: basiert auf 'norm'
Qualitätsprüfer: Korpling
Datum: 2016
Qualitätsprüfung: Korpling
Editor: Carolin Odebrecht

Typ: Annotationlayer – pos

Annotationsart: Spannenannotation
Beschreibung: Wortarten-Annotation nach dem Stuttgart-Tübingen-Tagset. Alle in norm als „unknown“ getaggten Token wurden in pos nachträglich mithilfe des ReplacePosOfUnclear.bas Skripts mit „XY“ getaggt.

Typ: Annotationvalue – pos

Wert: Wertbeschreibung:
STTS Wortarten-Annotation nach dem Stuttgart-Tübingen-Tagset.

4.2. lemma

Typ: Layer – lemma

Name: lemma
Dokumentation: <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – lemma

Schritt: 1
Aufbereitung: Lemmatisierung
Aufbereitungsart: Automatisch
Tool: Treetagger Version 3.2
Format: Treetagge-Output
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: Korpling
Editor: Carolin Odebrecht

Typ: Annotationlayer– lemma

Annotationsart: Spannenannotation
Beschreibung: Lemmatisierung: Normalisierung der flektierten Wortformen, Zuordnung der entsprechenden Lexikoneinträge zu jeder Wortform.

Typ: Annotationvalue – lemma

Wert: Normalisiertes Lemma
Wertbeschreibung: Lemmatisierung.

4.3. foreign

Typ: *Layer* – foreign

Name: foreign
Dokumentation: Siehe Anhang
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: *Preparationstep* – foreign

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2013
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ:*Annotationlayer* – foreign

Annotationsart: Spannenannotation

Beschreibung: Markiert einen Text, der in einer Fremdsprache geschrieben ist.

Typ:*Annotationvalue* – foreign

Wert: Wertbeschreibung

foreign: Allgemeiner Tag für fremdsprachliches Material.

4.4. foreign_trans

Typ: *Layer* – foreign_trans

Name: foreign_trans
Dokumentation: Siehe Anhang
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: *Preparationstep* – foreign_trans

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: LAUDATIO
Datum: 2013
Qualitätsprüfung: Manuelle Konsistenzprüfung: Die Tags aus der Annotationsebene **definition**, die eine Markierung von Übersetzungen darstellen, wurden unter dieser Annotationsebene zusammengefasst.
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – foreign_trans

Annotationsart: Spannenannotation
Beschreibung: Beschreibung der Zielsprache und der Ausgangssprache einer Übersetzung.

Typ: *Annotationvalue* – foreign_trans

Wert:	Wertbeschreibung:
trans_to_german	Übersetzung eines Wortes in Deutsch.
trans_from_german	Übersetzung eines Wortes vom Deutschen.
trans_to_german_extended	Übersetzung einer Phrase in Deutsch.
trans_from_german_extended	Übersetzung einer Phrase vom Deutschen.

4.5. lang

Typ: Layer – lang

Name: lang
Dokumentation: Siehe Anhang
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – lang

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2016
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Leere Zellen wurden mit dem Wert „deu“ aufgefüllt.
Aufbereitungsart: semi-automatisch
Tool: Excel 2013
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2016
Qualitätsprüfung: NA

Editor: Laura Perlitz
 Schritt: 3
 Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
 Aufbereitungsart: automatisch
 Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
 Format: PAULA-XML, ANNIS
 Segmentierung der Annotation: basiert auf 'dipl'
 Qualitätsprüfer: korpling
 Datum: 2016
 Qualitätsprüfung: NA
 Editor: Carolin Odebrecht

Typ: Annotationlayer – lang

Annotationsart: Spannenannotation
 Beschreibung: Kodiert die konkrete Sprache des im Layer **foreign** markierten Textes nach ISO Drei-Buchstaben-Code (z. B. lat).

Typ: Annotationvalue – lang

Wert:	Wertbeschreibung:
Kürzel der entsprechenden Sprache	Getagged nach ISO Drei-Buchstaben-Code ISO 639-3 (http://www.sil.org/iso639-3/codes.asp?order=639_3&letter=%25)
lat	Latein
grc	Altgriechisch
spa	Spanisch
ita	Italienisch
ara	Arabisch
fra	Französisch

pol	Polnisch
nld	Niederländisch
eng	Englisch
rus	Russisch
goh	Althochdeutsch
dan	Dänisch

4.6. sentence_end

Typ: *Layer* – sentence_end

Name:	sentence_end
Dokumentation:	http://sfs.uni-tuebingen.de/langbank/src/enhg_sent_segm-german-v4.pdf
Beschreibung:	Unter linguistischen Annotationen werden u. a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – sentence_end

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	Gohar Schnelle
Datum:	2016
Qualitätsprüfung:	manuelle Korrektur
Editor:	Gohar Schnelle
Schritt:	2

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – Name: sentence_end

Annotationsart: Spannenannotation
Beschreibung: Annotation der Satzenden nach folgenden Guidelines zur frühneuhochdeutschen Satzsegmentierung: http://sfs.uni-tuebingen.de/langbank/src/enhg_sent_segm-german-v4.pdf

Typ: Annotationvalue – Name: sentence_end

Wert: Wertbeschreibung:
S Markiert das Ende eines Satzes.

4.7. position_verb

Typ: Layer – position_verb

Name: position_verb
Dokumentation: NA
Beschreibung: Unter linguistischen Annotationen werden u. a. pos-Tagging,

Lemmatisierung und Markierung von fremdsprachlichem Material
zusammengefasst.

Typ: Preparationstep – position_verb

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	semi-automatisch und manuell (mit VBA-Skripten und simplen Ersetzungs- und Konsistenzprüfungsheuristiken)
Tool:	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	Hagen Hirschmann
Datum:	2014
Qualitätsprüfung:	manuelle Korrektur
Editor:	Hagen Hirschmann
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3

Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – Name: position_verb

Annotationsart:	Spannenannotation
Beschreibung:	Annotation der Verbstellung in untergeordneten Sätzen.

Typ: Annotationvalue – Name: position_verb

Wert:	Wertbeschreibung:
V2	Verbzweitstellung in einem gegebenen untergeordneten Satz mit subordinierender Konjunktion; analysiert als Token-Feature auf der Basis von pos=KOUS
Vletzt	Verbletzstellung in einem gegebenen untergeordneten Satz mit subordinierender Konjunktion; analysiert als Token-Feature auf der Basis von pos=KOUS
V?	Unklare Stellung des Verbs; in einem gegebenen untergeordneten Satz mit subordinierender Konjunktion; analysiert als Token-Feature auf der Basis von pos=KOUS
V1	Verberststellung in einem gegebenen untergeordneten Satz mit subordinierender Konjunktion; analysiert als Token-Feature auf der Basis von pos=KOUS

4.8. subclause_type

Typ: Layer – subclause_type

Name: subclause_type
Dokumentation: NA
Beschreibung: Unter linguistischen Annotationen werden u. a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – subclause_type

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: semi-automatisch und manuell (mit VBA-Skripten und simplen Ersetzungs- und Konsistenzprüfungsheuristiken)
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: Hagen Hirschmann
Datum: 2014
Qualitätsprüfung: manuelle Korrektur
Editor: Hagen Hirschmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016

Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – subclause_type

Annotationsart:	Spannenannotation
Beschreibung:	Annotation der Nebensatztypen.

Typ: Annotationvalue – subclause_type

Wert:	Wertbeschreibung:
Adverbial	Adverbiale Funktion eines untergeordneten Satzes.
Attribut	Attributive Funktion eines untergeordneten Satzes.
Komplement	Komplementfunktion eines untergeordneten Satzes.

4.9. KOUS_sem

Typ: Layer – KOUS_sem

Name: KOUS_sem

Dokumentation: NA

Beschreibung: Unter linguistischen Annotationen werden u. a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – KOUS_sem

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: semi-automatisch und manuell (mit VBA-Skripten und simplen Ersetzungs- und Konsistenzprüfungsheuristiken)

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: Hagen Hirschmann

Datum: 2014

Qualitätsprüfung: manuelle Korrektur

Editor: Hagen Hirschmann

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – KOUS_sem

Annotationsart:	Spannenannotation
Beschreibung:	Annotation der semantischen Funktion von subordinierenden Konjunktionen.

Typ: Annotationvalue – KOUS_sem

Wert:	Wertbeschreibung:
additiv	Additive Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
final	Finale Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
k.a.	Nicht-analysierbare Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
kausal	Kausale Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS

konditional	Konditionale Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
konsekutiv	Konsekutive Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
konzessiv	Konzessive Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
modal	Modale Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
temporal	Temporale Semantik einer subordinierenden Konjunktion; analysiert auf der Basis von Vorkommen mit pos=KOUS
0	NA

4.10.comp

Typ: *Layer* – komp

Name:	comp
Dokumentation:	siehe Dokumentation von ‘prot’
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comp

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	Laura Perlitz
Datum	2014
Qualitätsprüfung:	manuelle Korrektur

Editor: Laura Perlitz

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: Annotationlayer – comp

Annotationsart: Spannenannotation

Beschreibung: Markiert alle im Korpus enthaltenen zusammen, getrennt oder mit Bindestrich geschriebenen Komposita mit einem substantivischen Zweitglied.

Typ: Annotationvalue – comp

Wert: k
Wertbeschreibung: Kompositum mit substantivischem Zweitglied

4.11.comp_orth

Typ: Layer – comp_orth

Name: comp_orth

Dokumentation: Ist ein Kompositum an einer Morphemgrenze durch einen Zeilenumbruch unterbrochen, wurde es entweder mit „lb1“ oder „lb2“ markiert, je nachdem, ob ein Bindestrich vorhanden ist oder nicht. Ist dies nicht der Fall und das Kompositum wurde innerhalb einer Morphemgrenze unterbrochen, wird es als zusammen geschrieben markiert, da davon ausgegangen wird, dass der Autor das Kompositum nicht getrennt geschrieben hätte, wenn an dieser Stelle kein Zeilenumbruch vorkäme. Mit „bs“ wurden nur diejenigen Komposita markiert, die innerhalb einer Zeile mit einem Bindestrich versehen sind.

Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comp_orth

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell

Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: Laura Perlitz
Datum: 2014
Qualitätsprüfung: manuelle Korrektur
Editor: Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: Annotationlayer – comp_orth

Annotationsart: Spannenannotation

Beschreibung: Beschreibt die Orthographie der in der Ebene „komp“ annotierten Komposita.

Typ: Annotationvalue – comp_orth

Wert:	Wertbeschreibung:
zs	zusammen geschrieben
gtr	getrennt geschrieben
bs	Bindestrichmarkierung (innerhalb einer Zeile)
lb1	unterbrochen durch Zeilenumbruch (ohne Bindestrichmarkierung)
lb2	unterbrochen durch Zeilenumbruch (mit Bindestrichmarkierung)

4.12.prot

Typ: Layer – prot

Name: prot

Dokumentation: Der Prototyp 1 (prot1) markiert alle Token, die eindeutig als Komposita mit substantivischem Zweitglied zu identifizieren sind. Folgende Kriterien müssen dabei erfüllt sein: Es liegen mindestens zwei freie Morpheme vor, die entweder zusammen oder mit Bindestrich geschrieben wurden. Es darf wortintern keine Flexion auftreten und das Kompositum kann syntaktisch nicht erweitert werden, beispielsweise durch Adjektiv- oder Präpositionalphrasen (z.B. *stabwurz*).
Der Prototyp 2 (prot2) fasst alle Tokens zusammen, die sehr wahrscheinlich Komposita sind. Diese bestehen ebenfalls aus zwei freien Morphemen, werden aber getrennt geschrieben oder wurden

durch einen Zeilenumbruch unterbrochen und enthalten keine Bindestrichmarkierung. Sie dürfen wie Komposita des Prototyp 1 wortintern nicht flektierbar und syntaktisch nicht erweiterbar sein (z.B. *wermut faft*).

Der Prototyp 3 (prot3) markiert alle Zweifelsfälle, also Fälle, die weder eindeutig als Phrase noch als Kompositum zu identifizieren sind. Wieder liegen zwei freie Morpheme sowie Getrennschreibung vor, jedoch kann man keine genaue Aussage darüber treffen, ob eine wortinterne Flexion vorgenommen wurde oder möglich wäre und ob das jeweilige Vorkommen syntaktisch erweiterbar ist (z.B. *widers Teuffels Biffz*). Die Zweifelsfälle wurden nicht in der Ebene "komp" aufgenommen.

Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – prot

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	Laura Perlitz
Datum	2014
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013

Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – prot

Annotationsart:	Spannenannotation
Beschreibung:	Weist jedem Wert in der Ebene Kompositum einen Prototypen zu. Der Prototyp gibt an, wie wahrscheinlich ein Vorkommen ein Substantivkompositum ist.

Typ: Annotationvalue – prot

Wert:	Wertbeschreibung:
prot1	sicher als Kompositum zu identifizieren
prot2	sehr wahrscheinlich Kompositum

prot3

Zweifelsfall; nicht in der Ebene „komp“ markiert

4.13.attr_gen

Typ: Layer – attr_gen

Name: attr_gen

Dokumentation: Das Bezugsnomen und das Genitivattribut wurden in ein und derselben Spanne annotiert. Das Gleiche gilt für verschachtelte sowie koordinierte und durch weitere Phrasen erweiterte Genitivkonstruktionen (z.B. *Beobachtungen der nach der Bestäubung in den vegetabilischen Eyern erfolgenden Veränderungen*).

Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – attr_gen

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: Laura Perlitz

Datum: 2014

Qualitätsprüfung: manuelle Korrektur

Editor: Laura Perlitz

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – attr_gen

Annotationsart:	Spannenannotation
Beschreibung:	Markiert alle im Korpus enthaltenen Nominalphrasen, die ein Genitivattribut enthalten.

Typ: Annotationvalue – attr_gen

Wert:	Wertbeschreibung:
gprä	pränominales Genitivattribut
gpost	postnominales Genitivattribut

4.14.morph_ellipsis

Typ: *Layer* – morph_ellipsis

Name:	morph_ellipsis
Dokumentation:	„strD“ kann strukturelle Destruktionen (Koordinationen von Kompositionserst- und Zweitglied) mit und ohne Bindestrichmarkierung enthalten.
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – morph_ellipsis

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	Laura Perlitz
Datum	2014
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der	

Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – morph_ellipsis

Annotationsart:	Spannenannotation
Beschreibung:	Markiert alle im Korpus enthaltenen koordinierten Komposita („strukturelle Destruktionen“; z.B.: <i>gelb= und Wafferfucht</i>).

Typ: Annotationvalue – morph_ellipsis

Wert:	Wertbeschreibung:
strD	strukturelle Destruktion

4.15.comment_lex

Typ: Layer – comment_lex

Name: comment_lex
Dokumentation: NA
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comment_lex

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2015
Qualitätsprüfung: NA
Editor: Anke Lüdeling

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: NA
Qualitätsprüfer: korpling
Datum: 2016

Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – comment_lex

Annotationsart: Spannenannotation
Beschreibung: Dies ist eine unsystematische Ebene für Bemerkungen und Fragen zur Lexik.

Typ: Annotationvalue – comment_lex

Wert: Wertbeschreibung:
Zeichenkette Bemerkung und/oder Frage zur Lexik.

4.16.comment

Typ: Layer – comment

Name: comment
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comment

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer:	NA
Datum	2015
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin, korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – comment

Annotationsart:	Spannenannotation
Beschreibung:	Dies ist keine systematische Ebene. Sie dient lediglich der besseren Lesbarkeit. Wörter, deren Bedeutung sich aus dem Neuhochdeutschen nicht direkt erschließt (z.B. Krankheitsnamen oder Wörter, die keine neuhochdeutsche Entsprechung haben), werden hier erklärt. Auch Wörter, deren Form es heute noch gibt, die aber ihre Bedeutung geändert haben (<i>dicke</i> für <i>oft</i>) werden erläutert.

Typ: Annotationvalue – comment

Wert:	Wertbeschreibung:
Zeichenkette	Erläuterungen zur Lexik.

4.17.reference

Typ: *Layer* – reference

Name:	reference
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – reference

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	Uwe Springmann
Datum	2016
Qualitätsprüfung:	manuelle Korrektur
Editor:	Gohar Schnelle, Uwe Springmann
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	NA basiert auf ‘dipl’

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – reference

Annotationsart: Spannenannotation
Beschreibung: Diese unsystematische Ebene dient der Referenz von Interpretationen, die zum annotierten Sprachmaterial gefällt wurden.

Typ: Annotationvalue – reference

Wert: Wertbeschreibung:
Zeichenkette: Jede Art Angabe, die ein Nachvollziehen einer Interpretation erleichtert. Bei einschlägigen Wörterbüchern, wie Grimm, Lexer, Adelung, genügt der Nachname. Möglich sind z.B. bibliographische Angaben oder Weblinks.

4.18.form_disease

Typ: Layer – form_disease

Name: form_disease
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

. **Typ:** Preparationstep – form_disease

Schritt: 1

Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum:	2014
Qualitätsprüfung:	NA
Editor:	Anke Lüdeling, Uwe Springmann
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

. **Typ:** Annotationlayer – form_ disease

Annotationsart:	Spannenannotation
Beschreibung:	NA

. **Typ:** Annotationvalue – form_ disease

Wert:	Wertbeschreibung:
deriv	NA

derivat	NA
kompNN	NA
kompNNgetrennt	NA
lat	NA
phrase	NA
Phrase	NA
phraseDasIst	NA
phraseGen	NA
phraseGEN	NA
phraseGenannt	NA
phraseHS	NA
phraseRS	NA
phraseSubj	NA
phraseV1	NA
phraseVP	NA
simplex	NA
wort	NA

4.19.problem

Typ: *Layer* – werkname

Name:	problem
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: *Preparationstep* – problem

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell

Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum	2014
Qualitätsprüfung:	NA
Editor:	Anke Lüdeling, Uwe Springmann
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – problem

Annotationsart:	Spannenannotation
Beschreibung:	NA

Typ: Annotationvalue – problem

Wert:	Wertbeschreibung:
Zeichenkette	NA

4.20.herbname_norm

Typ: Layer – herbname_norm

Name: herbname_norm
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – herbname_norm

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016

Qualitätsprüfung: NA
Editor: Carolin Odebrecht

. **Typ:** Annotationlayer – herbname_norm

Annotationsart: Spannenannotation
Beschreibung: In dieser Ebene wird ein systematischer Kräuternamen gegeben. Manchmal ist der nicht eindeutig, dann wird in den Annotationsebenen „erläuterung“ oder „bemerkung_lexik“ darauf hingewiesen. Es gibt keine Leerzeichen, die beiden Bestandteile von zweiteiligen Bezeichnungen werden durch einen Unterstrich verbunden.

. **Typ:** Annotationvalue – herbname_norm

Wert: Wertbeschreibung:
Zeichenkette Normierter, systematischer Kräuternamen.

4.21. herbprep

Typ: Layer – herbprep

Name: herbprep
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – herbprep

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell

Tool: Excel 2013
 Format: Excel 2013
 Segmentierung der Annotation: basiert auf 'dipl'
 Qualitätsprüfer: NA
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
 Aufbereitungsart: automatisch
 Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
 Format: PAULA-XML, ANNIS
 Segmentierung der Annotation: basiert auf 'dipl'
 Qualitätsprüfer: korpling
 Datum: 2016
 Qualitätsprüfung: NA
 Editor: Carolin Odebrecht

Typ: Annotationlayer – herbprep

Annotationsart: Spannenannotation
 Beschreibung: In dieser Ebene geht es nur um die Zubereitungen oder Darreichungsformen von Kräutern. Dabei werden nur NPs betrachtet, bei denen das Kraut der Kopf oder der Kopf der Modifikation ist (die syntaktischen Modifizierungen werden in anderen Annotationsebenen annotiert). Es wird normiert auf Neuhochdeutsch die Form im Nominativ Singular ausgeschrieben. Dabei werden Leerzeichen bei PPs und Genitiven durch Unterstriche ersetzt. Komposita werden immer

zusammengeschrieben, auch wenn sie im Original nicht zusammengeschrieben sind. Alles wird in Kleinbuchstaben geschrieben.

Typ: Annotationvalue – herbprep

Wert:	Wertbeschreibung:
Zeichenkette	Kräuterzubereitung, z.B. <i>safft des weremuts</i> → <i>safft_des_wermuts</i> <i>des roten beyfusz bletter</i> → <i>des_roten_beifuß_blätter</i> <i>safft von weremut</i> → <i>safft_von_wermut</i> <i>weremutsafft</i> → <i>wermutsaft</i> <i>weremut safft</i> → <i>wermutsaft</i>

4.22.form_prep

Typ: Layer – form_prep

Name:	form_prep
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – form_prep

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum	2014

Qualitätsprüfung:	NA
Editor:	Anke Lüdeling, Uwe Springmann
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – form_prep

Annotationsart:	Spannenannotation
Beschreibung:	In dieser Ebene werden die Kräuterzubereitungen morphologisch oder syntaktisch beschrieben.

Typ: Annotationvalue – form_prep

Wert:	Wertbeschreibung:
kompNN	NN-Komposita annotiert, die zusammengeschrieben oder mit einem Bindestrich verbunden sind
kompNNgetrennt	NN-Folgen annotiert, die ein Kompositum sein könnten, aber mit Spatium geschrieben wurden
phraseVon	Kräuterzubereitungen mit von-PP, z.B. <i>safft von weremut</i>
phraseGen	Kräuterzubereitungen mit Genitiv-Attribut, z.B. <i>safft des weremuts</i>

4.23.noun_nom

Typ: *Layer* – noun_nom

Name: noun_nom
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – noun_nom

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – noun_nom

Annotationsart: Spannenannotation
Beschreibung: In dieser Ebene werden alle Nomina aufgenommen, die im Text vorkommen und zwar immer in der ersten Schreibung des Nominativ Singular, in der sie vorkommen.

Typ: Annotationvalue – noun_nom

Wert: Wertbeschreibung:
Zeichenkette Wenn die erste Form von ‚Saft‘ safft ist, werden alle Vorkommen von ‚Saft‘ mit safft annotiert, auch wenn es später Formen von saft geben sollte. Alle Pluralformen werden durch die Singularform annotiert. Alles wird in Kleinbuchstaben geschrieben. Diese Ebene hilft bei der Feststellung der Variation innerhalb eines Textes.

4.24.form_noun

Typ: Layer – form_noun

Name: form_noun
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – form_noun

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013

Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum	2014
Qualitätsprüfung:	NA
Editor:	Anke Lüdeling, Uwe Springmann
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

. **Typ:** Annotationlayer – form_noun

Annotationsart: Spannenannotation
 Beschreibung: In dieser Ebene werden alle Nomina morphologisch annotiert.

. **Typ:** Annotationvalue – form_noun

Wert: Wertbeschreibung:
 simplex Mit *simplex* werden alle Nomina annotiert, die entweder morphologisch wirklich nicht komplex sind oder die als Simplex wahrgenommen werden. Das ist natürlich eine nicht immer einfache Entscheidung. Manchmal ist sie in der Annotationsebene „bemerkung_lexik“ kommentiert. Auch

lateinische und griechische Lehnwörter, die integriert sind (wie *capitel* oder *quintin*) werden hier annotiert. Die Komplexität in der Gebersprache spielt keine Rolle.

kompNN	NN-Komposita, die zusammengeschrieben oder mit einem Bindestrich verbunden sind
kompNNgetrennt	alle NN Folgen, die ein Kompositum sein könnten, aber mit Spatium geschrieben wurden
kompNEN	NE-N-Komposita, die zusammengeschrieben oder mit einem Bindestrich verbunden sind
kompNENgetrennt	alle NE-N Folgen, die ein Kompositum sein könnten, aber mit Spatium geschrieben wurden
kompNNNgetrennt	alle NNN Folgen, die ein Kompositum sein könnten, aber mit Spatium geschrieben wurden
kompAN	AN-Komposita
kompVN	VN-Komposita
derivat	Derivate
nom	implizite Nominalisierung (Konversion, Ablaut, syntaktische Nominalisierung)
gri	Klar griechische Nomina. Bereits integrierte Wörter griechischen Ursprungs werden nicht gekennzeichnet, sondern wie native Wörter behandelt.
lat	Klar lateinische Nomina. Bereits integrierte Wörter lateinischen Ursprungs werden nicht gekennzeichnet, sondern wie native Wörter behandelt.
ara	Klar arabische Nomina. Bereits integrierte Wörter arabischen Ursprungs werden nicht gekennzeichnet, sondern wie native Wörter behandelt.
lex	kraut: Mit lex werden bestimmte lexikalisierte Kräuternamen annotiert. Diese sind zwar ursprünglich morphologisch komplex, aber nicht mehr offen und der Schreiber hat keine Wahl (ob diese Bildungen noch transparent sind, ist unklar). Beispiele sind Beifuß, Wermut, Stabwurz, aber auch tausend guldin für Tausendguldenkraut.

4.25. clause_type

Typ: Layer – clause_type

Name: clause_type
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – clause_type

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: Annotationlayer – clause_type

Annotationsart: Spannenannotation
Beschreibung: Annotation von Satztypen. Da wir nur flach annotieren und nicht beliebig Ebenen vervielfachen wollen, haben wir ein Problem bei geschachtelten Strukturen (also Relativsätzen, die Relativsätze enthalten oder Relativsätzen, die Demonstrativsätze enthalten). Wir haben jeweils den äußersten (größten) Satz annotiert. In der Annotationsebene „bemerkungen_syntax“ steht ein Hinweis auf die Verschachtelung.

Typ: Annotationvalue – clause_type

Wert:	Wertbeschreibung:
rs	Klare Relativsätze, dabei sowohl w-Relativsätze als auch d-Relativsätze, z.B. <i>im schz adet auch nit ob ain thier das nit finnig wår in beiffet</i> Wer dē beyfusz bei im tregt wen er wandert der wirt nit müde
padv	Sätze, die durch ein Präpositionaladverb eingeleitet werden, z.B. <i>das stabwurz genüczet verz zeret vberflüffig feuchtikaít die ín den dårmen find dauo ain kranck haít komet genant colíca paffío</i> Hier ist oft unsicher, ob das überhaupt Relativsätze sind.
rsx	Relativsätze ohne Hauptsatz. Das kommt häufig in Überschriften vor, z.B. Wer eynen kurtzen atthem hat
rsdem	Ambige Fälle: Relativsatz oder Demonstrativsatz, z.B. <i>vnd nimpt traurigkeit die kompt von Melancoley</i>
padvpart	Sätze mit Präpositionaladverb und Partizip, z.B. dauon genomen auf ain halb lot Diese Sätze sind nicht ambig wie rsdem, sondern enthalten beides! Ob sie überhaupt Relativsätze sind, ist unklar.
dem	Demonstrativsätze. Hier sind alle Sätze gemeint, deren Subjekt ein Demonstrativpronomen ist, z.B.

der hilfft on zweifel

part

Partizipien, die sich ähnlich verhalten wie Relativsätze, z.B.

Jtem stabwurz mit olei vn salcz zefamē gestoffen vnd auff den pulz hende vnd fūsz gelegtiſt gūt wider das fieber

Oft sind mehrere Partizipien verkettet (zB durch und). Die ganze Kette wird als eine Spanne annotiert. Die Form der Partizipstrukturen ist unterschiedlich, manchmal ist das Partizip der Kopf, manchmal eine Präposition.

4.26.position_rel

Typ: Layer – position_rel

Name: position_rel

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html

Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – position_rel

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: NA

Datum: 2014

Qualitätsprüfung: NA

Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2

Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – position_rel

Annotationsart:	Spannenannotation
Beschreibung:	Position des Relativsatzes im Hauptsatz

Typ: Annotationvalue – position_rel

Wert:	Wertbeschreibung:
vor	Vorangestellter Relativsatz, z.B. <i>Vnd welche fraw dz zů rechter zeit nit hat die sol den beyfüß fieden mit wein vn sol den trinckē der hilfft on zweifel</i> Wir annotieren auch ‚vor‘, wenn es ein Einleitungswort wie item oder und oder ein XY spricht gibt (würde man wahrscheinlich in ein Vorvorfeld einsortieren).
nach	Nachgestellter Relativsatz, z.B. <i>Vn auch fast wol bekome den die den stain habē in den lenden</i>
int	Eingebetteter Relativsatz, z.B. <i>Jtem stabwurz macht auch wachfzen den bart d langkfam herfür komet domit bez strichē</i>

4.27. position_referent

Typ: Layer – position_referent

Name: position_referent
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – position_referent

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – position_referent

Annotationsart: Spannenannotation
Beschreibung: Position des Relativsatzes zur Bezugskategorie

Typ: Annotationvalue – position_referent

Wert:	Wertbeschreibung:
adja-v	Adjazent vorangestellt, z.B. <i>Item wer beyfufz in feim haufz hatdem mag der teufel kaínen fchaden zufügen</i>
adja-n	Adjazent nachgestellt, z.B. <i>im fchz adet auch nít ob aín thier das nít fínníg wárin beiffet</i>
dist	Distant, z.B. <i>dauo den bart oder die stat gefchmieret do man har begert</i> Interpunktionszeichen werden nicht berücksichtigt (d.h., wenn zwischen Relativsatz und Bezugskomponente ein Interpunktionszeichen steht (zum Beispiel eine Virgel), wird als adjazent annotiert.
na	Nicht anwendbar, z.B. <i>vnd benimbt den alten hűften vn rainíget die bruft dē fy erkaltet íft</i>

4.28.form_referent

Typ: Layer – form_referent

Name: form_referent
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – form_referent

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – form_ referent

Annotationsart: Spannenannotation
Beschreibung: Form der Bezugskategorie des Relativsatzes

Typ: Annotationvalue – form_ referent

Wert:	Wertbeschreibung:
np	Nicht pronominale NP, z.B. <i>vn treibt aufz den stain der in der blafen wachftz</i>
d-pron	<i>der, die, das, dieser</i> etc., z.B. <i>Wer dē beyfufz bei im tregt wen er wandert der wirt nit müde</i>
p-pron	Personalpronomina, z.B. <i>Jtem wer € beyfufz wurczel an feinem halz tregt kain vergifftig tier mag im nit geschaden</i>
null	Bei freien und asyndetischen Relativsätzen, die kein overt es Korrelat im Hauptsatz haben, z.B. <i>vn was nach dem getranck vberbeleibt vo kreütern fol man warm legen auf die milcz</i>

4.29. position_verb_rel

Typ: Layer – position_verb_rel

Name:	position_verb_rel
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – position_verb_rel

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA

Datum	2014
Qualitätsprüfung:	NA
Editor:	Anke Lüdeling, Uwe Springmann
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – position_verb_rel

Annotationsart:	Spannenannotation
Beschreibung:	Verbstellung im Relativsatz

Typ: Annotationvalue – position_verb_rel

Wert:	Wertbeschreibung:
v2	Verb-Zweit-Stellung, z.B. <i>Vn frawen die kinder find feügen folz len nít nützen eppich</i>
ve	Verb-End-Stellung, z.B. <i>vmb tuz gent willen die diße künigin an di fem kraut befand</i>
venf	Verb-End-Stellung mit besetztem Nachfeld, z.B. <i>Vn auch fast wol bekome den die den stain habē in den lenden</i> Ob das wirklich als Nachfeld analysiert werden soll, ist

unerheblich, das hier meint einfach, dass es eine klare Verb-End-Stellung ist, aber noch etwas danach kommt.

4.30.form_relpron

Typ: *Layer* – form_relpron

Name: form_relpron
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – form_relpron

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS

Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – form_relpron

Annotationsart:	Spannenannotation
Beschreibung:	Form der reltivsateinleitenden Kategorie

Typ: Annotationvalue – form_relpron

Wert:	Wertbeschreibung:
d-pron	Alle d-Pronomen, auch <i>do/dâ</i> , z.B. <i>Difes wechft gern bei dē faulen waffern do die frōsch wonent der da</i> wird als Relativsatzeinleiter wird als Spanne annotiert.
w-pron	<i>wer, welch-</i> , z.B. <i>wer beyfuß bei im hab so er vber feld gat€</i> wird nít müd
w-phras	welch frau etc., z.B. <i>Weliche fraw aín tods kind ín írem leíb het die trinck vo beyfuß</i>

4.31.mod_referent

Typ: Layer – mod_referent

Name:	mod_referent
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – mod_referent

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – mod_referent

Annotationsart: Spannenannotation
Beschreibung: Modifikation der Bezugskategorie

Typ: Annotationvalue – mod_referent

Wert:	Wertbeschreibung:
relsatz	An Pronomen, NP oder Satz annotiert, wenn von einem Relativsatz modifiziert. Bei freien Relativsätzen nicht anwendbar. Die ganze Bezugskategorie wird als Spanne annotiert.
d-pron	NA
m-padv	NA
m-part	NA
np	NA

4.32.ppk_e1

Typ: Layer – ppk_e1

Name:	ppk_e1
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – ppk_e1

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	Laura Perlitz
Datum	2016
Qualitätsprüfung:	manuell
Editor:	Marco Krämer

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer - ppk_e1

Annotationsart:	Spannenannotation
Beschreibung:	Die ppk_e1-Ebene gibt an, ob es sich bei einer Textsequenz um eine Präpositionalkonstruktion (Präpositionalattributskonstruktion bzw. Attributive Adverbiale Bestimmung) handelt.

Typ: Annotationvalue – ppk_e1

Wert:	Wertbeschreibung:
ppk	normale Präpositionalkonstruktion
zwf	Zweifelsfall
ppk_rek	rekursive (verschachtelte) Präpositionalkonstruktion
attr_X	Attribute, die sich auf ein Element einer PPK beziehen, ohne innerhalb einer syntaktischen Sequenz direkt an dieses angeschlossen zu sein. X ist ein Platzhalter für das jeweilige Bezugswort.

4.33.ppk_e2

Typ: Layer – ppk_e1

Name: ppk_e2
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – ppk_e2

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: Laura Perlitz

Datum: 2016

Qualitätsprüfung: manuell

Editor: Marco Krämer

Schritt: 2

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: Annotationlayer - ppk_e2

Annotationsart: Spannenannotation
Beschreibung: Die ppk_e2-Ebene zeigt alle Präpositionalkonstruktionen innerhalb der Struktur der in ppk_e1 erfassten Präpositionalkonstruktionen an.

Typ: Annotationvalue – ppk_e2

Wert:	Wertbeschreibung:
ppk_e2	normale Präpositionalkonstruktion
zwf	Zweifelsfall
BSP	Annotation besonderer Beispiele. Markierung für den Annotator.
BSP+	Annotation besonderer Beispiele. Markierung für den Annotator.
BSPBuchtitel	Annotation besonderer Beispiele, die einen Buchtitel betreffen. Markierung für den Annotator.

4.34.ppk_e3

Typ: Layer – ppk_e1

Name: ppk_e3
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – ppk_e3

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der: basiert auf 'dipl'

Annotation:

Qualitätsprüfer: Laura Perlitz

Datum: 2016

Qualitätsprüfung: manuell

Editor: Marco Krämer

Schritt: 2

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

. **Typ:** Annotationlayer - ppk_e3

Annotationsart: Spannenannotation

Beschreibung: Die ppk_e3-Ebene zeigt alle Präpositionalkonstruktionen innerhalb der Struktur der in ppk_e2 erfassten Präpositionalkonstruktionen an.

. **Typ:** Annotationvalue – ppk_e3

Wert:	Wertbeschreibung:
ppk_e3	normale Präpositionalkonstruktion

5. Annotationsebenen – Strukturelle Annotation

5.1. lb

Typ: *Layer* – lb

Name:	lb
Dokumentation:	Siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – lb

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2015
Qualitätsprüfung:	manuelle Korrektur
Editor:	Studierende der Humboldt-Universität zu Berlin, korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	manuelle Korrektur
Editor:	Carolin Odebrecht

Typ: Annotationlayer – lb

Annotationsart:	Spannenannotation
Beschreibung:	Markierung einer ganzen Zeile.

Typ: Annotationvalue – lb

Wert:	Wertbeschreibung:
lb	Markierung pro ganzer Zeile.

5.2. pb

Typ: Layer – pb

Name:	pb
Dokumentation:	Siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – pb

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2015
Qualitätsprüfung: manuelle Korrektur
Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – pb

Annotationsart: Spannenannotation
Beschreibung: Markierung einer ganzen Seite.

Typ: Annotationvalue – pb

Wert: Wertbeschreibung:
pb Markierung pro ganzer Seite.

5.3. pb_n

Typ: *Layer* – pb_n

Name: pb_n
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang
Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – pb_n

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2015
Qualitätsprüfung: manuelle Korrektur
Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – pb_n

Annotationsart:	Spannenannotation
Beschreibung:	Markierung der Nummerierung einer Seite.

Typ: Annotationvalue – pb_n

Wert:	Wertbeschreibung:
numerischer Wert	Seitenzahl

5.4. pb_ana

Typ: Layer – pb_ana

Name:	pb_ana
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: Siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist

die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – pb_ana

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: LAUDATIO
Datum: 2013
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (<https://korpling.german.hu->

berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: Annotationlayer – pb_ana

Annotationsart: Spannenannotation

Beschreibung: Interpretation von vermutlich nicht korrekten Seitenzahlangaben.

Typ: Annotationvalue – pb_ana

Wert: Wertbeschreibung:

Int Entsprechende Angabe der vermutlich korrekten Seitenzahl.

5.5. unclear

Typ: Layer - unclear

Name: unclear

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: Siehe Anhang

Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl-Ebene). So ist die Tokenisierung in der dipl-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – unclear

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: LAUDATIO
Datum: 2015
Qualitätsprüfung: Manuelle Konsistenzprüfung. Der Layer reason, der eine Begründung für unleserliche Elemente beschreiben sollte, wurde entfernt.
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpapper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: *Annotationlayer* – unclear

Annotationsart: Spannenannotation
Beschreibung: Markierung, dass Zeichen eines Tokens oder ein Token bzw. ein Text nicht lesbar sind.

Typ: *Annotationvalue* – unclear

Wert:	Wertbeschreibung:
unclear	Nicht leserliche Zeichen oder Zeichenketten werden unabhängig von der Anzahl unleserlicher Zeichen, in der <i>dipl</i> -Ebene mit nur <u>einem</u> Unterstrich („_“) gekennzeichnet (siehe Transkription und Normalisierung). Wenn in der Transkription ein solcher Fall eintritt, so wird mit dem Tag „unclear“ in der gleichnamigen <i>unclear</i> -Ebene dies allgemein angezeigt. In die Ebenen atLeast und atMost wird angegeben, wieviele Zeichen „von“ „bis“ fehlen bzw. nicht gelesen werden können.

5.6. atLeast

Typ: *Layer* – atLeast

Name:	atLeast
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – atLeast

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	Manuell
Tool:	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘dipl’

Qualitätsprüfer: LAUDATIO
Datum: 2012
Qualitätsprüfung: Manuelle Konsistenzprüfung
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach
Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013

Segmentierung der Annotation:
basiert auf 'dipl'

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: *Annotationlayer* – atLeast

Annotationsart: Spannenannotation

Beschreibung: Geschätzte, minimale Anzahl an nicht lesbaren Zeichen eines Token, die in der Spalte **unclear** als „unclear“ getagged sind.

Typ: *Annotationvalue* – atLeast

Wert: Wertbeschreibung:
numerischer Wert Dieser Tag gibt die geschätzte Mindestanzahl an unleserlichen Zeichen an.

5.7. atMost

Typ: *Layer* – atMost

Name: atMost
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: Siehe Anhang
Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die grafische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – atMost

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Datum: 2013
Qualitätsprüfer: LADUATIO
Qualitätsprüfung: Manuelle Konsistenzprüfung

Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach
Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation:
basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: Annotationlayer – atMost

Annotationsart: Spannenannotation

Beschreibung: Geschätzte, maximale Anzahl an nicht lesbaren Zeichen eines
Token, die in der Spalte **unclear** als „unclear“ getagged sind.

Typ: Annotationvalue – atMost

Wert:	Wertbeschreibung:
numerischer Wert	Dieser Tag gibt die geschätzte Höchstanzahl an unleserlichen Zeichen an.

5.8. interpretation

Typ: *Layer*– interpretation

Name:	interpretation
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – interpretation

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	LAUDATIO
Datum:	2013
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: *Annotationlayer*– interpretation

Annotationsart: Spannenannotation

Beschreibung: Interpretationsvorschläge für unleserliche Zeichen/Zeichenketten, die in der Spalte **unclear** als „unclear“ getagged sind.

Typ: *Annotationvalue*– interpretation

Wert: Wertbeschreibung:

Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2015
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – figure

Annotationsart:	Spannenannotation
Beschreibung:	Markiert eine Graphik, die im Text eingebettet ist.

Typ: Annotationvalue – figure

Wert:	Wertbeschreibung:
Figure	Allgemeine Kennzeichnung einer Grafik im Text.
Table	Allgemeine Kennzeichnung einer Tabelle im Text

5.10.figure_rend

Typ: Layer – figure_rend

Name:	figure_rend
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren,

strukturellen Annotationen.

Typ: *Preparationstep* – figure_rend

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2013
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS
 Segmentierung der Annotation: basiert auf 'dipl'
 Qualitätsprüfer: korpling
 Datum: 2016
 Qualitätsprüfung: NA
 Editor: Carolin Odebrecht

Typ: *Annotationlayer* – figure_rend

Annotationsart: Spannenannotation
 Beschreibung: Beschreibung, was auf dem Bild zu sehen ist (z. B. Darstellung einer Pflanze oder eines Destilliergeräts).

Typ: *Annotationvalue* – figure_rend

Wert:	Wertbeschreibung:
drawingOfTwoJars	Abbildung von zwei Gläsern.
drawingOfThreeJars	Abbildung von drei Gläsern.
drawingOfTwoGlasses	Abbildung von zwei Brillen.
drawingOfThreeGlasses	Abbildung von drei Brillen.
drawingOfTwoAlembics	Abbildung von zwei Destillierkolben.
drawingOfAnInstrument	Abbildung von einem Instrument.
drawingOfAnEibisch	Abbildung eines Eibischs.
drawingOfAStaubwurtz	Abbildung eines Staubwurtz.
drawingOfAKamille	Abbildung einer Kamille.
drawingOfAHühnerdarm	Abbildung eines Hühnerdarms.
drawingOfAFilter	Abbildung eines Filters.
drawingOfAHelmet	Abbildung eines Helms.
drawingOfADill	Abbildung von Dill.
drawingOfAHauswurtz	Abbildung von Hauswurtz.
drawingOfAWaldenburgischerKolben	Abbildung eines Waldenburgischen Kolbens.
drawingOfADrachenwurtz	Abbildung von Drachenwurtz.

drawingOfAGauchhey1

Abbildung von Gauchhey1.

drawingOfAHaselwurtz

Abbildung von Haselwurtz.

5.11.hi

Typ: *Layer* – hi

Name: hi

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang

Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – hi

Schritt: 1

Aufbereitung: Strukturelle Annotation

Aufbereitungsart: manuell

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: LAUDATIO

Datum: 2016

Qualitätsprüfung: NA

Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – hi

Annotationsart:	Spannen
Beschreibung:	Markierung eines irgendwie hervorgehobenen Bereichs.

Typ: Annotationvalue – hi

Wert:	Wertbeschreibung:
hi	Allgemeine Kennzeichnung eines hervorgehobenen Bereichs (highlighted area).

5.12.script

Typ: *Layer* – 5.12. script

Name: script

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang

Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die grafische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – script

Schritt: 1

Aufbereitung: Strukturelle Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt: 2

Aufbereitung: Umbenennung des Wertes „antiqua“ in „roman“ und des Wertes „gothic“ in „blackletter“. Die Spezifizierung von Schwabacher und Fraktur in NewKreuerbuch_1563_Handsch wurde aufgehoben.

Aufbereitungsart: automatisch

Tool: Excel 2013

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf ‘dipl’

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Laura Perlitz

Schritt: 3

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf ‘dipl’

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ:*Annotationlayer* – script

Annotationsart: Spannenannotation

Beschreibung: Weist jedem Token eine Schriftart zu. In manchen Texten kommt es vor, dass ein und dasselbe Token zum Teil in einem und zum Teil in einem anderen Font gedruckt wurde. Dies betrifft v.a. Wörter mit deutscher Flexionsendung und fremdsprachiger Wurzel. Für diese Token wurde innerhalb von „script“ der Wert „mixed“ annotiert.

Typ:*Annotationvalue* – script

Wert: Wertbeschreibung:

roman Text, der in Antiqua gedruckt ist.

blackletter	Text, der in Fraktur gedruckt ist.
mixed	Text, der in verschiedenen Schriftsätzen gedruckt ist.

5.13.hi_rend

Typ: *Layer* – hi_rend

Name:	hi_rend
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – hi_rend

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum	2016
Qualitätsprüfung:	Manuelle Konsistenzprüfung: Korrektur der false positives
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: *Annotationlayer* – hi_rend

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibung der Darstellung des hervorgehobenen Bereichs, z. B. graphische Hervorhebungen, Farben etc.

Typ: *Annotationvalue* – hi_rend

Wert:	Wertbeschreibung:
italics	Text, der kursiv gedruckt ist.
bold	Text, der fett gedruckt ist.
underlined	Text, der unterstrichen gedruckt ist.
red	Text, der rot gedruckt ist.
iniCap	Ein Text, der mit einem zusätzlichen Großbuchstaben gedruckt ist.
letter-spacing: 1em	Ein größerer Abstand zwischen Buchstaben, Sperrdruck.

5.14.head

Typ: *Layer* – head

Name:	head
-------	------

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang

Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – head

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: LAUDATIO
Datum: 2016
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016

Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – head

Annotationsart: Spannenannotation
Beschreibung: Markierung einer Überschrift.

Typ: Annotationvalue – head

Wert: Wertbeschreibung:
head Allgemeine Markierung einer Überschrift.

5.15. note

Typ: Layer – note

Name: note
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang (8.2)
Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – note

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2013

Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2015
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – note

Annotationsart: Spannenannotation
Beschreibung: In einer Spannenannotation wird die gesamte Fußnote mit dem Tag „note“ markiert. Der Text der Fußnoten wird immer am Ende des Absatzes, in dem sie eingepflegt sind, unabhängig davon, ob sich so die Fußnoten sammeln oder erst auf der nachfolgenden Seite realisiert werden müssten, in die dipl-Ebene/Transkription aufgenommen.

Typ: Annotationvalue – note

Wert: Wertbeschreibung:

note	In einer Spannenannotation wird die gesamte Fußnote mit dem Tag „note“ markiert. Der Text der Fußnoten wird immer am Ende des Absatzes, in dem sie eingepflegt sind, unabhängig davon, ob sich so die Fußnoten sammeln oder erst auf der nachfolgenden Seite realisiert werden müssten, in die dipl-Ebene/Transkription aufgenommen.
margin	In einer Spannenannotation wird die gesamte Marginalie mit dem Tag „margin“ markiert. Der Text der Marginalien wird immer am Anfang des jeweiligen Absatzes, in dem/ neben dem sie realisiert sind, unabhängig davon, ob so mehrere Marginalien hintereinander oder/und auf der vorherigen Seite realisiert werden müssen, in die dipl-Ebene/Transkription aufgenommen.
end	Eine Endnote.

5.16.ref

Typ: Layer – ref

Name:	ref
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – ref

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘dipl’

Qualitätsprüfer: NA
Datum: 2013
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach
Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013

Segmentierung der Annotation:
basiert auf 'dipl'

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – ref

Annotationsart: Spannenannotation

Beschreibung: Referenz einer Fußnote.

Typ: Annotationvalue – ref

Wert: Wertbeschreibung:
ref Referenz einer Fußnote.

5.17.ref_target

Typ: Layer – ref_target

Name: ref_target

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang

Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – ref_target

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2013
Qualitätsprüfung: NA

Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach
Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation:
basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: *Annotationlayer* – ref_target

Annotationsart: Spannenannotation

Beschreibung: ID der Fußnote, auf die referiert wird. Es wird fortlaufend mit #fZ nummeriert.

Typ: *Annotationvalue* – ref_target

Wert:	Wertbeschreibung:
#fINT	ID der Fußnote, auf die referiert wird. Es wird fortlaufend mit #fINT nummeriert.

5.18.ref_type

Typ: *Layer* – ref_type

Name:	ref_type
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – ref_type

Schritte:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2

Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ:*Annotationlayer* – ref_type

Annotationsart:	Spannenannotation
Beschreibung:	Art der Referenz, z.B. TEI "noteAnchor".

Typ:*Annotationvalue* – ref_type

Wert:	Wertbeschreibung:
noteAnchor	Art der Referenz.

5.19.quote

Typ: *Layer* – quote

Name:	quote
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – quote

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	LAUDATIO
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

Aufbereitung: Ersetzung des Wertes „quote“ durch den Wert „yes“, Ersetzung der leeren Zellen durch den Wert „no“

Aufbereitungsart: semi-automatisch

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Laura Perlitz

Schritt: 4

Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: Annotationlayer – quote

Annotationsart: Spannenannotation
Beschreibung: Ordnet Fließtext als Zitat oder Nicht-Zitat ein.

Typ: Annotationvalue – quote

Wert: Wertbeschreibung:
yes Ein Zitat im Text.
no Kein Zitat.

5.20.item

Typ: Layer – item

Name: item
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang
Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – item

Schritt: 1
Aufbereitung: Strukturelle Annotation
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA

Datum: NA
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – item

Annotationsart: Spannenannotation
Beschreibung: Element des Textes, der als Liste in der Ebene **list** markiert ist.

Typ: Annotationvalue – item

Wert: Wertbeschreibung:
Item Jedes Element der Aufzählung wird einzeln getagged.

6. Annotationsebenen – Inhaltliche Annotation

6.1. definition

Typ: *Layer* – definition

Name: definition
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang
Beschreibung: Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: *Preparationstep* – definition

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf ‘norm’
Qualitätsprüfer: LAUDATIO
Datum: 2013
Qualitätsprüfung: Manuelle Konsistenzprüfung. Eingrenzung der in dieser Ebene beschriebenen Informationen, Markierung von Übersetzung in eine eigene Ebene (**foreign_trans**) ausgelagert.
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: *Annotationlayer* – definition

Annotationsart: Spannenannotation
Beschreibung: Beschreibung oder Erklärung eines auf eine Abbildung bezogenen oder im Text genannten Begriffs.

Typ: *Annotationvalue* – definition

Wert:	Wertbeschreibung:
fig	Bildbeschreibung.
expl	Erklärung eines Begriffs.

6.2. disease

Typ: *Layer* – disease

Name:	disease
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: Preparationstep – disease

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	NA
Datum:	2012
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'norm'

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Überführung des Wertes „d“ aus der ehemaligen Ebene „term“ in diese Ebene (Wert: di)
Aufbereitungsart: automatisch
Tool: Excel 2013
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2016
Qualitätsprüfung: NA
Editor: Laura Perlitz

Schritt: 4
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – disease

Annotationsart: Spannenannotation

Beschreibung: Nennung einer Krankheit (Beschwerden, Symptome etc.) im Text. Es wird jeweils die komplette Phrase, wenn vorhanden auch mit restriktiven Relativsätzen, getagged, z.B. *der Bauch, der weh tut* oder *Bauchweh*.

Typ: Annotationvalue – disease

Wert: Wertbeschreibung:
di Benennung einer Krankheit

6.3. persname

Typ: Layer – persname

Name: persname
Dokumentation: NA
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – persname

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2015
Qualitätsprüfung: manuelle Korrektur
Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: Annotationlayer – persname

Annotationsart:	Spannenannotation
Beschreibung:	Es werden alle Personennamen annotiert, auf die ein Autor referiert.

Typ: Annotationvalue – persname

Wert:	Wertbeschreibung:
Zeichenkette	Diese Ebene dient der eindeutigen und schnellen Verknüpfung der im Korpus auftretenden Namen mit den dazugehörigen historischen Personen. Die Ebene erhebt keinen Anspruch auf Vollständigkeit der Namen, vielmehr wird nur so viel Namenmaterial geliefert, wie zur Identifikation der Person mit Hilfe digitaler Nachschlagewerke erforderlich ist. z.B. "Dioskurides" aber "Arthemisia II" (vs. "Arthemisia I") "Mausolos I" (vs. "Mausolos II") Ist eine Person nicht eindeutig identifizierbar wird der in der "norm" gegebene Wert im Nominativ Singular in die Ebene "persname" übertragen. Die Schreibung der Namen richtet sich nach ihrer Schreibung in der deutschsprachigen Wikipedia.

6.4. title

Typ: *Layer* – title

Name: title
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – title

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf ‘dipl’
Qualitätsprüfer: NA
Datum: 2015
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin; korpling

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: NA
Qualitätsprüfer: basiert auf ‘dipl’
Qualitätsprüfer: korpling

Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – title

Annotationsart: Spannenannotation
Beschreibung: Es werden alle Werknamen annotiert, auf die ein Autor referiert.

Typ: Annotationvalue – title

Wert: Wertbeschreibung:
Zeichenkette Es wird der jeweilige Werkname im Nominativ angegeben.

6.5. author_ref

Typ: Layer – author_ref

Name: author_ref
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang
Beschreibung: Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u.a. Annotationen von im Korpus enthaltenen Begriffen, Definitionen oder Referenzen.

Typ: Preparationstep – author_ref

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA

Datum: 2015
Qualitätsprüfung: Laura Perlitz
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Ersetzung der Werte "include" und „self“ durch andere mögliche Werte, z.B. „pron1sg“ und „pron1pl“

Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: NA
Datum: 2015
Qualitätsprüfung: NA
Editor: Laura Perlitz

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: *Annotationlayer* – author_ref

Annotationsart: Spannenannotation
Beschreibung: Markiert Referenzen des Autors auf sich oder andere Autoren bzw. auf beide.

Typ: *Annotationvalue* – author_ref

Wert:	Wertbeschreibung:
pron1pl	Referenz auf den Autor (und auf andere Autoren oder den Leser) mit der ersten Person Plural, z.B. „wir“, „unser“.
pron1sg	Referenz auf den Autor mit der ersten Person Singular, z.B. „ich“, „mein“.
pron3sg	Referenz auf einen Autoren mit der dritten Person Singular, z.B. „er“.
pron2pl	Referenz auf mehrere Autoren mit der zweiten Person Plural, z.B. „ihnen“.
author	Explizite Referenz auf einen benannten Autor, z.B. „Plinius“, „Dioscorides“, „ein Mönch“ (wenn dieser als Autor/Quelle zitiert wird)

6.6. reader_ref

Typ: *Layer* – reader_ref

Name:	reader_ref
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: *Preparationstep* – reader_ref

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	NA
Datum:	2015

Qualitätsprüfung: Laura Perlitz
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Ersetzung des Wertes “address” in “reader” oder andere mögliche Werte
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf ‘norm’
Qualitätsprüfer: NA
Datum: 2015
Qualitätsprüfung: NA
Editor: Laura Perlitz

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf ‘norm’
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: *Annotationlayer* – reader_ref

Annotationsart: Spannenannotation
Beschreibung: Markiert Referenzen des Autors auf den Leser.

Typ: *Annotationvalue* – reader_ref

Wert:	Wertbeschreibung:
pron1pl	Referenz mit der ersten Person Plural, also z.B. „wir“, „uns“, „unser“
pron2sg	Referenz mit der zweiten Person Singular, z. B. „du“, „dein“
pron3sg	Referenz mit der dritten Person Singular, z.B. „er“, „sie“
pron2pl	Referenz mit der zweiten Person Plural, z.B. „ihr“, „euer“
reader	Adressierung an den Leser in nicht-pronominaler Form, z.B. „an den Leser“, „lieber Leser“
author	Anrede eines anderen Autoren mit Namen, z.B. <i>An den Großzachtbahren vnd Hochgelehrten Herrn Johan Jacob Krafft</i>

Referenzen auf ein generisches „man“ oder Imperative werden in dieser Annotationsebene nicht annotiert.

6.7. plant

Typ: *Layer* – plant

Name:	plant
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: *Preparationstep* – plant

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	LAUDATIO

Datum: 2012
Qualitätsprüfung: Manuelle Konsistenzprüfung.
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Überführung des Wertes "h" aus der ehemaligen Annotationsebene "term" in diese Ebene (Wert: pl)
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2016
Qualitätsprüfung: NA
Editor: Laura Perlitz

Schritt: 4
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (<https://korpling.german.hu->

berlin.de/saltpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: *Annotationlayer* – plant

Annotationsart: Spannenannotation

Beschreibung: Nennung einer Pflanze (Kräuter wie Bäume wie Blumen etc.) im Text. Es wird jeweils die komplette Phrase, wenn vorhanden auch mit restriktiven Relativsätzen, getagged, z.B. *das Kraut, das hinter dem Bahnhof wächst* oder *Kamille*.

Typ: *Annotationvalue* – plant

Wert: Wertbeschreibung:

pl Benennung einer Pflanze.

6.8. property

Typ: *Layer* – property

Name: property

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang

Beschreibung: Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: *Preparationstep* – property

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: LAUDATIO
Datum: 2012
Qualitätsprüfung: Manuelle Konsistenzprüfung.
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Typ: *Annotationlayer* – property

Annotationsart: Spannenannotation
Beschreibung: Beschreibt den Bezug von Eigenschaften wie Geruch, Aussehen etc. auf eine Referenz (z.B. Kraut).

Typ: *Annotationvalue* – property

Wert:	Wertbeschreibung:
appearance	Beschreibt das Aussehen eines Krauts.
effect	Beschreibt den Effekt, die Wirkung eines Krauts.
smell	Beschreibt den Geruch eines Krauts.
preparation	Beschreibt die Verarbeitungs-/Bearbeitungsschritte eines Krauts.
taste	Beschreibt den Geschmack eines Krauts.
cultivation	Beschreibt den Abbau eines Krauts.

6.9. name

Typ: *Layer* – name

Name: name
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang
Beschreibung: Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: *Preparationstep* – name

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Tool: Excel 2010

Format: Excel 2010
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2012
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013

Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Typ: Annotationlayer – name

Annotationsart: Spannenannotation
Beschreibung: Markiert Eigennamen.

Typ: Annotationvalue – name

Wert: Wertbeschreibung:
Name: Allgemeine Markierung von Eigennamen.

6.10.name_type

Typ: Layer – name_type

Name: name_type
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang
Beschreibung: Inhaltliche Annotationen sind speziell auf das Register des Korpus bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: Preparationstep – name_type

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2012
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach
Excel 2013

Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

Typ: *Annotationlayer* – name_type

Annotationsart:	Spannenannotation
Beschreibung:	Spezifiziert den Typ der Eigennamen.

Typ: *Annotationvalue* – name_type

Wert:	Wertbeschreibung:
herb	Eigename eines Krauts.
scholar	Eigename einer Schule/Lehre.

plant	Eigenname einer Pflanze.
person	Eigenname einer Person.
flower	Eigenname einer Blume.
tree	Eigenname eines Baums.
gardener	Eigenname eines Gärtners.
publisher	Eigenname eines Verlegers.

7. Metadaten

Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele: siehe Anhang

Beschreibung: Die Metadaten enthalten Informationen in Bezug auf jeden historischen Text im Korpus. Zum Beispiel wird für jeden historischen Text der Titel, der Autor, das Erscheinungsjahr, der Sprachraum und die Textgestaltung angegeben.

Typ: *Preparationstep* - Metadaten

Schritt:	1
Aufbereitung:	Metadatum
Tool	Excel 2013
Format:	Excel 2013
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	manuell
Editor:	korpling

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-

berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

Metadatum	Annotationswerte	Beschreibung
author	Zeichenkette	Name des Autors des Textes.
bibl	Zeichenkette	Voller bibliographischer Eintrag für die Quelle inklusive der im Korpus annotierten Seitenzahlen.
date	JJJJ (Int)	Datum der Veröffentlichung, normalerweise das Jahr.
publisher	Zeichenkette	Verlag/Verleger.
place	Zeichenkette	Verlagsort.
title	Zeichenkette	Titel der Arbeit, aus dem das Dokument extrahiert wurde.
translator	Zeichenkette	Übersetzer des Textes, wenn vorhanden.
trans_from	it lat	Sprache, aus der der Text übersetzt wurde.
editor	Zeichenkette	Herausgeber des Textes (wenn zusätzlich zum Verlag/Verleger angegeben).
version	5.0	Versionierung des Korpus.
edition_first	yes no	Wenn bekannt, wird hier angegeben, ob es sich bei dem Faksimile um eine Erstauflage (yes) handelt oder nicht (no).
issue	Int	Nummer des Bandes, wenn vorhanden.
maintopic	science non-science	Angabe darüber, ob der Text eher Alltags- oder Wissenschaftsthemen behandelt.
topic	additiv aus: Al As B G K L M R	Ordnet einem Text einen oder mehrere Themenbereiche zu. Additiver Wert aus den verschiedenen Kürzeln, alphabetisch geordnet. Al: Alchemie, As: Astronomie, B: Botanik, G: Gärtnerei, K: Küche, L: Linguistik, M: Medizin, R: Religion,. Beispielwerte sind "B", "BM" oder "BKM".
register	herbology	Register des Textes.
lingualism	monoling multiling	Gibt an, ob größere Textstellen vorhanden sind, die nicht auf Deutsch verfasst wurden. Dies gilt nicht für einzelne Übersetzungen von Pflanzennamen.

orig_date	JJJJ (Int)	Ist ein Text eine Nichterstaufgabe, dann kann hier das Jahr der Erstauflage stehen, wenn bekannt.
repository	URL	URL zum Repositorium des Faksimiles.
lang_type	enhg nhg	Angabe des Sprachtyps. enhg: frühneuhochdeutsch (early new high german), nhg: neuhochdeutsch (new high german)
lang_area	md obd NA	Das Sprachgebiet wird angegeben. md: mitteldeutsch, obd: oberdeutsch. Bei jüngeren Texten, die schon weitestgehend standardisiert sind und denen kein Sprachgebiet zugeordnet werden konnte, wird als Wert "NA" angegeben.
text_type	prose lyric mixed	Angaben über die allgemeine Textgestaltung. Prosa: der Text ist komplett in Prosaform verfasst, Poesie: der Text ist komplett in Reimform verfasst; gemischt: der Text ist teilweise prosaisch, teilweise gereimt verfasst.
lyric_type	end_rhyme meter rhyme_meter	Wenn im Metadatum "text_type" die Werte "lyric" oder "mixed" angegeben sind, dann wird hier eingetragen, welche Gestaltungselemente dafür verwendet wurden.
preface	yes no	Gibt an, ob eine Vorrede mit transkribiert wurde.
wormwood	yes no	Gibt an, ob ein Textabschnitt zum Thema "Wermut" transkribiert wurde.
herb_sorting	yes no	Gibt an, ob es sich bei dem Text um eine Kräutermonographiesammlung handelt, d.h. dass nacheinander verschiedene Kräuter abgehandelt werden.

8. Referenzen

- Bollmann, Marcel; Petran, Florian & Stefanie Dipper (2011): Rule-Based Normalization of Historical Texts. In: Proceedings of the RANLP Workshop on Language Technologies for Digital Humanities and Cultural Heritage, Hissar, Bulgaria, 34-42.
- Dickinson, Markus & Detmar Meurers (2003): Detecting Errors in Part-of-Speech Annotation. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03). Budapest, Hungary.
- Krause, Thomas; Lüdeling, Anke; Odebrecht, Carolin & Amir Zeldes (2012): Multiple Tokenizations in a Diachronic Corpus. Exploring Ancient Languages through Corpora Conference (EALC), 14.-16. Juni 2012. Oslo, Norway.
- Reichmann, Oskar & Klaus-Peter Wegera (Hrsg.) (1993): Frühneuhochdeutsche Grammatik. Tübingen: Niemeyer Verlag.
- Schmid, Helmut (1994): *Probabilistic Part-of-Speech Tagging Using Decision Trees*. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK.

Empfohlene Literatur:

- Georges, Karl Ernst (2010) : Ausführliches deutsch-lateinisches Handwörterbuch. Unveränderter Nachdruck der achten verbesserten und vermehrten Auflage von Heinrich Georges. Darmstadt: WBG.
- Hartweg Frederic & Klaus-Peter Wegera (2005): Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit. 2., neu bearbeitete Auflage, Tübingen: Niemeyer Verlag.
- Niermeyer, J.F. & C. van de Kieft (2002): *Mediae Latinitatis Lexicon Minus*. Überarbeitet von J.W.J. Burgers. Darmstadt: WBG.

<http://www.cis.uni-muenchen.de/ocrworkshop/program.html>

Adelung: <http://woerterbuchnetz.de/Adelung/>

Bibel: <http://www.die-bibel.de/online-bibeln>

Grimm: <http://woerterbuchnetz.de/DWB/>

Lexer: <http://woerterbuchnetz.de/Lexer/>

1. Anhang

Transkription und Normalisierung

dipl-Ebene	clean-Ebene	norm-Ebene	
Grundlage für/	Grundlage für/		Normalisierungsschritte
grafische Annotationen		pos- Ebene	weitere Annotationen...

Die Transkription von Faksimiles stellt für die korpuslinguistische Aufbereitung zumeist die grundlegende, diplomatische Ebene (**dipl**). Diese Ebene soll sich grafisch hinsichtlich Orthografie, Getrennt- und Zusammenschreibung und Sonderzeichen möglichst nah am zugrunde liegenden Faksimile orientieren.

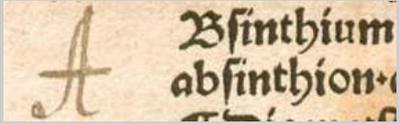
Die **clean**-Ebene enthält erste Normalisierungen hinsichtlich Sonderzeichen und grafischer Strukturierungen. So werden u. a. Ligaturen, die nach moderner Rechtschreibung nicht mehr verwendet werden, normalisiert. Graphische Markierungen der einzelnen Texte wie Zeilenumbrüche werden aufgelöst und Sonderzeichen einiger Fonts wie Fraktur auf die heutigen Schriftsätze übertragen.

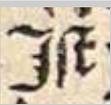
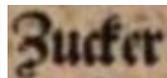
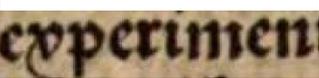
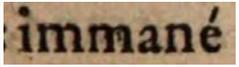
Die **norm**-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier die Tokenisierung, die Orthografie, Flexionsformen und Wortbildungen einheitlich dem nhd. Sprachstand angepasst werden. Durch die einzelnen Normalisierungsschritte und der daraus resultierenden multiplen Segmentierung ist es möglich, die uneinheitliche Orthografie in einzelnen Texten sowie im diachronen Verlauf zu vereinheitlichen und dennoch die spezifischen Besonderheiten des jeweiligen Textes/der jeweiligen Sprachstufe im Korpus abzubilden. Weiterhin können automatische Tagging- und ParsingTools, die in vielen Fällen auf modernem Deutsch trainiert sind, auf der norm-Ebene angewandt werden. Durch die Alignierung der einzelnen Ebenen können Annotationen und Suchanfragen, die auf Grundlage der norm-Ebene getätigt wurden, auf die dipl-Ebene zurückgeführt werden.

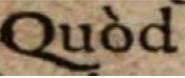
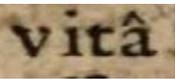
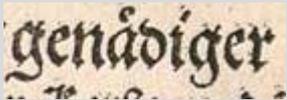
Nachfolgend werden die Richtlinien zur Transkription und Normalisierung im Einzelnen gegenübergestellt. Allgemein für jede Segmentierungsebene sowie Annotationsebene gilt, dass keine Leerzeilen oder Leerzeichen enthalten sein dürfen. Dazu wird für jeden Fall ein Beispiel aus dem Ridges Herbology Version 4.0 Korpus gegeben, die konkrete Regel beschrieben und die weiteren Normalisierungsschritte gegenübergestellt (dipl – clean – norm). Die Richtlinien sind nach Schwerpunkten gruppiert: (1) Zeichensetzung/Sonderzeichen (2) Segmentierung/Tokenisierung (3) Interpunktion.

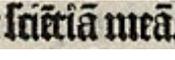
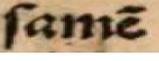
- 1 Transskribieren
 - 1.1 Lautzeichen
 - 1.2 Diakritika
 - 1.3 Satzzeichen
 - 1.4 Sonstige Zeichen
 - 1.5 Ligaturen
 - 1.6 Griechische Schrift
 - 1.7 Typographische Besonderheiten
- 2 Tokenisieren

- 2.1 Wort
- 2.2 Zeilenumbrüche
- 2.3 Abkürzungen
- 2.4 Zahlen
- 2.5 Satzzeichen
- 2.6 Fußnoten
- 3 Normalisieren
 - 3.1 Phonologie
 - 3.2 Morphologie
 - 3.3 Morphosyntax
 - 3.4 Wortbildung
 - 3.5 Lexik
 - 3.6 Graphematik

Nr.	dipl-Ebene	clean-Ebene	norm-Ebene
	Transkript des Faksimiles.	Wird automatisch durch ein Python-Skript (clean-skript.py) erstellt. Dieses ersetzt alle heute unüblichen Sonderzeichen durch heute verwendete Entsprechungen.	Erfolgt manuell im .xlsx Format nach der modernen neuen Rechtschreibung. Tipp: Kopieren Sie sich die clean-Ebene und verändern Sie dann die entsprechenden Stellen!
1.0	Nicht mehr lesbare Zeichen oder Zeichenketten werden mit <u>einem</u> Unterstrich markiert, unabhängig davon, wieviele Zeichen (in etwa) nicht mehr interpretiert werden können.	Durch die fehlende Transkription ist eine Normalisierung nicht mehr möglich, Dies wird mit dem Platzhalter-Tag „unknown“ angezeigt.	Durch die fehlende Transkription ist eine Normalisierung nicht mehr möglich, Dies wird mit dem Platzhalter-Tag „unknown“ angezeigt.
	dipl	clean	norm
	–	unknown	unknown
	Handschriftliche Versalien die offensichtlich zum Text gehören (KEINE Anmerkungen oder Kommentare von Lesern), werden mit annotiert. 		
	dipl	clean	norm
	ABfynthium	ABSinthium	Absinth
1.1	Die Zeichen für 'I' und 'J' sind in Fraktur-Texten graphisch in der Regel nicht voneinander unterscheidbar. Konvention in Ridges ist daher, das Zeichen in der dipl als 'J' darzustellen und erst in der norm-Ebene das Zeichen individuell nach moderner Orthografie zu interpretieren.	Wie in der dipl-Ebene	Anpassung an moderne Orthographie.

			
	dipl	clean	norm
	Jst	Jst	Ist
	Das Zeichen 'J' wird in der dipl-Ebene nicht von 'Z' unterschieden.	Wie in der dipl-Ebene	
	dipl	clean	norm
	 Zucker	Zucker	Zucker
	Das „f“ wird mit U 017F übernommen.	Das lange „f“ wird durch ein reguläres „s“ ersetzt.	
	Dipl	clean	norm
	Lefer	Leser	Leser
	 Dies ist ein <x>	x	x
	dipl	clean	norm
	 Experiment	Experiment	experiment
1.2	Alle Akzente werden übernommen. Hierbei ist zu beachten, dass der jeweilige unäre Hex-Code zu verwenden ist. Akut („ó“) - U 0300 Gravis („ò“) - U 0300 Zirkumflex („ô“) - U 0302 oder über sog. dead keys	Alle Akzente werden übernommen.	Alle Akzente werden übernommen.
	Dipl	clean	norm
		immané	immané

immané		
 Quòd	Quòd	Quòd
 vitâ	vitâ	vitâ
 Dieses Zeichen wird mit U 016F in der dipl-Ebene übernommen.	Dieses diakritische Zeichen wird in der clean-Ebene nicht mehr realisiert und durch den zugrundeliegenden Vokal ersetzt.	
Dipl	clean	norm
zû	zu	zu
 Umlaute mit superskribiertem „e“ werden mit Vokal + U 0364 übernommen.	Umlaute mit superskribiertem „e“ werden wie moderne Umlaute des Deutschen realisiert.	
dipl	clean	norm
genâdiger	genädiger	gnädiger
 'i' mit schrägem Strich wird nicht vom 'i' mit Punkt unterschieden.		
dipl	clean	norm
die	die	die
 Das ũ mit schrägen Strichen wird nicht vom ü mit Punkten unterschieden.	ü mit schrägen Strichen wird in der clean-Ebene durch 'ü' ersetzt.	ü mit schrägen Strichen wird in der norm-Ebene durch 'ü' ersetzt.

dipl	clean	norm
fünff	fünff	fünf
Alle horizontalen Striche über einem Zeichen werden als Tilde (U 0303) interpretiert Eine Tilde kann stehen für: ausgelassene Nasale, sog. Nasalstrich	Tilden werden entsprechend ihrer Funktion umgesetzt. Die kann u. a. aus Reichmann & Wegera (1993) entnommen werden. ¹ Achtung: In der clean-Ebene wird der Strich automatisch als Nasalstrich interpretiert. Gegebenenfalls in der norm ausgleichen.	Gegebenenfalls in der clean-Ebene eingefügte Nasale anders umsetzen.
dipl	clean	norm
 om̃ia	omnia	omnia
 sciēt̃iā meā	scientiam meam	scientiam meam
 ei	ein	ein
 samē	samen	Samen
 eynē	eynem	einem
 nēlich	nemlich	nämlich
- Verdopplung eines Buchstaben		
 hūmel	himmel	Himmel
 iñ	inn	in
- Abkürzung frequenter Wörter und Endungen		

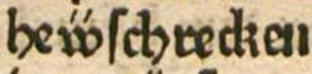
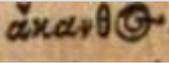
1 Nachweis für das Beispiel: Reichmann & Wegera 1993: 31.

	unn (sic!)	und	
	umm (sic!)	umb	
	darumm (sic!)	darumb	
 complexiō	complexiom (sic!)	complexione	
Zwibel od Kn Das Dicit-Zeichen (Abkürzung für "der") wird in der dipl-Ebene mit ð (AltGr oder als Sonderzeichen)	Es wird in der clean automatisch mit "der" ersetzt	In der norm auch.	
dipl	clean	norm	
oð	oder	oder	
 Dieses Zeichen wird mit U+A770 repräsentiert.	Ersetzung durch <us>	Ersetzung durch <us>	
dipl	clean	norm	
de□	deus	deus	
 Dieses Zeichen wird durch v+U+0309 realisiert.	Ersetzung durch <er>	Ersetzung durch <er>	
dipl	clean	norm	
vrzeren	vertzeren	verzehren	
1.3	 Das Zeichen für morphologische Worttrennung (oder Zeilenumbuch) wird mit	Die morphologische Trennung durch „≡“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt. Komposita, die in der dipl- Ebene mit einem „≡“ realisiert	Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthographie

	U 2E17 ² übernommen.	werden, behalten dieses Zeichen bei, um die allgemeine Suche nach Komposita in dieser Ebene zu ermöglichen. Für Erläuterungen zur rein graphischen Trennung, wie Zeilenumbrüchen, siehe Nummer 2.1.	zusammengeschrieben. (Grundlage hierfür ist der aktuelle Stand des Dudens)
	Dipl	clean	norm
	Dañ=reifz	Dann-reisz	Tannenreis
	Das "modifizierende Pluszeichen" wird mit U+02D6 übernommen.	In der clean wird das Zeichen zu ":" normalisiert.	In der norm wird das Zeichen zu ":" normalisiert.
	dipl	clean	norm
	.	:	:
	Das Absatzzeichen wird in dipl-, clean- und norm-Ebene mit U+00B6 übernommen.		
	dipl	clean	norm
	¶	¶	¶
1.4	Bruchangaben werden mit „/“ übernommen. Dabei wird getrennt segmentiert, um gemischte Brüche transparent zu halten.	Bruchangaben werden mit „/“ übernommen. Dabei wird getrennt segmentiert, um gemischte Brüche transparent zu halten.	Bruchangaben werden mit „/“ übernommen. Dabei wird getrennt segmentiert, um gemischte Brüche transparent zu halten.
	Dipl	clean	norm
	1	1	1
	1/2	1/2	1/2
	 Dieses Symbol wird mit dem U 2299	Dieses Symbol wird mit dem U 2299 übernommen.	Dieses Symbol wird mit dem U 2299 übernommen.

2 Sonderzeichen wie zum Beispiel das „/“ werden mit Hilfe von Unicodes in der dipl-Ebene realisiert. Eine Liste der zu verwendenden Unicodes ist angefügt.

	übernommen.		
	Dipl	clean	norm
	⊙	⊙	⊙
	diefe: haben Senkrechtstriche werden mit „ “ übernommen.	Senkrechtstriche werden übernommen.	Senkrechtstriche werden übernommen.
	dipl	clean	norm
	diefe	diese	diese
	 Dieses Symbol wird als “\$” transkribiert. Es bedeutet wahrscheinlich „Zwitter“ bzw. „zwitterig“.	Die als Platzhalter dienenden Dollarzeichen werden übernommen.	Dollarzeichen werden übernommen. In der Ebene „comment“ kann ergänzt werden, dass das Zeichen vermutlich „Zwitter“ oder „zwitterig“ bedeutet.
	dipl	clean	norm
	\$	\$	\$
1.5	Die Ligatur „æ“ wird mit U 00E6, für „Æ“ mit U 00C6 übernommen. Das gleiche gilt ebenfalls für „Œ“ U 0152 und „œ“ U 0153.	Die Ligatur wird aufgelöst.	
	dipl	clean	norm
	hæc	haec	haec
	 Die ct-Ligatur wird nicht übernommen.	Die ct-Ligatur wird nicht übernommen.	Die ct-Ligatur wird nicht übernommen.
	dipl	clean	norm
	Lactucis	Lactucis	Lactucis
		vv	

<p>Dies ist eine Ligatur aus \ddot{v} und v. Darstellung mit v+U+0308 Die Ligatur wird aufgelöst $\ddot{v}v$</p>		
dipl	clean	norm
 <p>heüvſchrecken</p>	heüvſchrecken	Heuschrecken
<p>Wie auch u.a. in Beispiel 1.2 wird die Ligatur mit dem langen „f“ und „z“ übernommen.</p>	Die Ligatur wird analog zu „sz“ aufgelöst.	Nach modernen Orthografieregeln wird für „sz“ das „ß“ verwendet.
dipl	clean	norm
fleifzige	fleiszige	fleißige
<p>Griechische Ligaturen werden nicht abgebildet, sondern bereits in der dipl aufgelöst:</p>  <p>Ligatur aus ο und ς</p> <p>Als Hilfe: https://de.wikipedia.org/wiki/Griechisches_Alphabet#/media/File:Greek_alphabet_ligatures.jpg</p>		
dipl	clean	norm
ἄκανθος	ἄκανθος	ἄκανθος
 <p>Dies bedeutet "etc."</p>		

	dipl	clean	norm
	& c.	& c.	etc.
	Das kaufmännische Und „&“ wird mit U 0026 übernommen.	Das kaufmännische Und wird übernommen.	Das kaufmännische Und wird übernommen.
	dipl	clean	norm
	&	&	&
1.6	Sprachliches Material mit anderen Schriftsystemen (z.B. Griechisch) wird in clean und norm nicht geändert http://www.unicode.org/charts/PDF/U0370.pdf sowie http://www.unicode.org/charts/PDF/U1F00.pdf		
	dipl	clean	norm
	ἄκανθος	ἄκανθος	ἄκανθος
1.7	Oft werden die ersten Zeichen bzw. ersten Wörter (im Bsp. Von Wegz) einer neuen Seite bei einem Seitenumbruch in der vorhergehenden Seite in einem eigenen Absatz/in einer eigenen Zeile doppelt realisiert (sog. Kustoden). Diese Zeichen bzw. das Wort wird nicht mit in das Transkript aufgenommen.		
2.0	 Auch entgegen moderner Orthographieregeln wird analog zur Textgrundlage tokenisiert.		Wortbildung und Großschreibung, die nicht der modernen Orthographieregeln entsprechen, werden angeglichen.
	dipl	clean	norm
	Spannenlang	Spannen-lang	spannenlang

2.1	 Klitika werden als ein Token realisiert.	Klitika werden als ein Token realisiert.	Klitika werden in dieser Ebene aufgelöst. Dazu wird die Tokenisierung verändert, in dem aus einem Token in dipl (oder clean) zwei Token gemacht werden!
	dipl	clean	norm
	Ichs	ichs	ich es
	Auch entgegen moderner Orthographieregeln wird analog zur Textgrundlage tokenisiert.		Getrennt- und Zusammenschreibung wird nach modernen Orthografieregeln angepasst. Dazu wird die Tokenisierung verändert, indem aus einem Token in der dipl-Ebene (oder clean) zwei Token in der norm-Ebene gemacht werden!
	dipl	clean	norm
	zufetzen	zusetzen	zu setzen
	Im Falle einer Worttrennung aufgrund von Zeilenumbrüchen, die keine overte, graphische Markierung wie „=-“ beinhalten, werden formal die getrennten Elemente als jeweils ein Token in der dipl-Ebene betrachtet.	Die Worttrennung ohne graphische Markierung ist nicht in jedem Fall transparent. Graphische Normalisierungen werden nicht vorgenommen.	Diese Art der Worttrennung, wenn sie transparent ist, wird nach modernen Orthographieregeln normalisiert und die graphische Trennung aufgehoben.
	dipl	clean	norm
	ge	ge	genannt
	nent	nent	

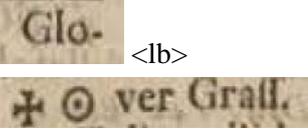
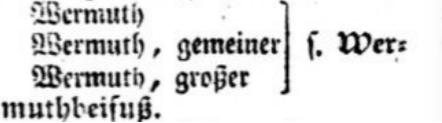
	Komposita, egal welcher Komplexität oder Bildungsweise, mit Gleichheitszeichen (U 2E17) werden als ein Token realisiert.	Die morphologische Trennung durch „≠“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt.	Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschrieben. (Grundlage hierfür bildet die aktuelle Ausgabe des Dudens.)
	dipl	clean	norm
	Artzney≠Kunft	Artzney-Kunst	Arzneikunst
2.2	Von Zeilenumbrüchen betroffene Elemente werden analog zur Primärquelle getrennt tokenisiert. Das Trennungszeichen „≠“ wird mit U 2E17 übernommen.	Wörter, die von Zeilenumbrüchen betroffen sind, werden in der clean-Ebene ohne den (Doppel-) Bindestrich, der diesen anzeigt, sowie in einer Spanne zusammengefasst realisiert. Die grafische Worttrennung wird somit aufgehoben.	
	dipl	clean	norm
	allge≠	allgemeinen	allgemeinen
	meinen		
	Von Zeilenumbrüchen betroffene Komposita werden analog zur Primärquelle getrennt tokenisiert, dabei bleibt die Kennzeichnung der morphologischen Worttrennung unberührt.	Wenn ein Kompositum, das durch Gleichheitszeichen grundsätzlich getrennt wird, von einem Zeilenumbruch betroffen ist, so wird dieses Gleichheitszeichen in der clean-Ebene entfernt, das andere Gleichheitszeichen für die morphologische Trennung wird analog zu Nummer 1.1 mit einem Minus ersetzt. Das Kompositum wird in einer Spanne zusammengefasst wiedergegeben.	Die Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschrieben. (Grundlage hierfür bildet die aktuelle Ausgabe des Dudens.)
	dipl	clean	norm

	Stab z	Stabwurtz-Oel	Stabwurzöl
	wurtz z Oel		
2.3	Abkürzungen werden analog zur Textgrundlage tokenisiert. Das heißt, auch Satzfehler oder Spatien, die nicht modernen Orthografieregeln entsprechen, werden realisiert.		Abkürzungen werden nach Dudenrichtlinien umgesetzt. Dies hat zur Folge, dass Abkürzungen unterschiedlich tokenisiert werden können (bspw. Abkürzungen als zwei oder mehr Tokens wie u. a. m. vs. solche, die als ein Token realisiert werden, wie usw.). vgl. dazu Nummer 2/3.1 und 2/3.3. In Fällen, in denen keine Dudenrichtlinie zugeordnet werden kann, wird immer nach einem Punkt segmentiert.
	dipl	clean	norm
	u.	u.	usw.
	f.	s.	
	w.	w.	
	Abkürzungen werden analog zur Textgrundlage tokenisiert. Das heißt, auch Satzfehler oder Spatien, die nicht modernen Orthografieregeln entsprechen, werden realisiert.		Abkürzungen werden nach Dudenrichtlinien umgesetzt. Dies hat zur Folge, dass Abkürzungen unterschiedlich tokenisiert werden können (bspw. Abkürzungen als zwei oder mehr Tokens wie u. a. m. vs. solche, die als ein Token realisiert werden, wie usw.). vgl. dazu Nummer 2/3.1 und 2/3.2. In Fällen, in denen keine Dudenrichtlinie zugeordnet werden kann, wird immer nach einem Punkt segmentiert.
	dipl	clean	norm
	u	u	u.
			d.

	d.g.	d.g.	g.
2.4	Punktsetzungen bei Ordinalzahlen werden mit der Ziffer als ein Token realisiert.		Nach modernen Orthographieregeln werden Ordinalzahlen mit Punkt realisiert.
	dipl	clean	norm
	I.	I.	I.
	Die Faksimilia weisen oft Kardinalzahlen mit Interpunktion auf. Diese Punktsetzungen werden mit der jeweiligen Ziffer als Token realisiert. Ausnahmen bilden Zahlen, die mit einem satzbeenden Punkt auftreten. Hier wird wie gewohnt die Satzinterpunktion getrennt von der Kardinalzahl tokenisiert.		Um die Kardinalzahlen von Ordinalzahlen in der norm-Ebene getrennt betrachten zu können, wird die Interpunktion bei Kardinalzahlen weggelassen.
			
	dipl	clean	norm
	I.	I.	1
	Loth	Loth	Lot
	.IX.	.IX.	IX
	Zeichenketten wie "5 %", "5-12", "800" werden auf der dipl und clean als ein Token und in der norm als mehrere Token betrachtet, wenn sie im Scan visuell zusammen stehen. Stehen sie visuell		Norm: immer als mehrere Token

	auseinander, werden sie auch als verschiedene Tokens realisiert.		
2.5			
	Satzinterpunktion wie Kommata, Punkte, Virgeln oder Semikola werden als jeweils eigenständige Token realisiert.		
	dipl	clean	norm
	finden	finden	finden
	.	.	.
2.6	Fußnotenreferenzierungen werden in dipl- und clean-Ebene als mehrere Token und in der norm-Ebene als ein Token realisiert		Norm: ein Token
	dipl	clean	norm
	(*) * *)	(*) * *)	(* **)
3.1	Formen die in phonologischer Hinsicht fnhd. oder dialektalen Lautstand aufweisen werden in der norm den (standard)nhd Entsprechungen angepasst		
	dipl	clean	norm
	Hellenpein	Hellenpein	Höllenpein

	wänlin	wänlin	Wännlein
3.2	Fnhd. Flexionsformen werden durch die nhd. Entsprechungen ersetzt.		
	dipl	clean	norm
	in die Nasen das kal Haupt	in die Nasen das kal Haupt	in die Nase das kahle Haupt
	Es erfolgt keine Anpassung des Genus		
	dipl	clean	norm
	das Milz	das Milz	das Milz
3.3	Es erfolgt keine morphosyntaktische Anpassung (z.B. innerhalb der Nominalphrase)		
	dipl	clean	norm
	(es heilt) die versehrte Därmelein	die versehrte Därmelein	die versehrte Därmelein
3.4	Ausgestorbene Wortbildungsmorpheme werden nach Möglichkeit durch entsprechende nhd. Bildungen ersetzt. (siehe Beispieltabelle am Schluss)		
	dipl	clean	norm
	halben stachelecht	halben stachelecht	halber stachelig
3.5	Ausgestorbenes lexikalisches Material wird mit modernen Orthographieregeln übernommen.		
	dipl	clean	norm
	Vergefz	Vergefz	Vergess
3.6	Intervenieren innerhalb eines Wortes Sonderzeichen, dann werden die Sonderzeichen in der norm-Ebene ignoriert und das unterbrochene Wort zusammengesrieben.		
	dipl	clean	norm

		Glo⚡ ⊙ ver Grass	Glover Grass
Text hinter (geschweiften) Klammern, wird nach dem umklammerten Textabschnitt eingefügt.			
dipl			
 <p>Wermuth Wermuth , gemeiner Wermuth , grofzer f. Werz muthbeifuß .</p>			
Graphematisch wird grundsätzlich nach dem Duden normalisiert.			

Fußnoten und Marginalien

Der Text der **Marginalie** wird immer am Anfang des jeweiligen Absatzes, in dem/neben dem sie realisiert ist, unabhängig davon, ob so mehrere Marginalien hintereinander oder/und auf der nachfolgenden Seite realisiert werden müssen, in die dipl-Ebene/Transkription aufgenommen. Der Text der **Fußnote** wird immer am Ende des Absatzes, in dem sie eingepflegt ist, unabhängig davon, ob sich so die Fußnoten sammeln oder erst auf der nachfolgenden Seite realisiert werden müssten, in die dipl-Ebene/Transkription aufgenommen.

Beispiele für die Referenzierung von Fußnoten

Siehe auch Abschnitt 5.34 (xml_id), 5.28 (ref_target), 5.27 (ref) und 5.26 (note).

dipl	ref	ref_target
.		
Paltinaten		
49)	ref	#f1
Wapp		

dipl	not	xml_id
.		
49)		
Paltinaten		
find		
unterschiedener		
Arten		
,		
als		
Paltinaca		
Domestica		
,		
fylveltris		
,		
marina		
,		
Syriac_		
,		
man		
kan		
Hn.		
von		
Franckenu		
KrauterLexic.		
voc.		
Paltinaca		
nachschlagen		
,		
wofelbiten		
der		
Ufus		
pharmacaticus		
weiter		
verfolget		
wird		
.	note	f1

Tabelle SEQ Tabelle * ARABIC 1: Fußnotenreferenzierung

Zeichentabelle

Beschreibung	Zeichen	Eingabe über die Tastatur
Schräger Doppelbindestrich	≈	2E17
Langes kleines „s“	ſ	017F Alt s
Tilde	˜	0303
Ligatur ae	æ	00E6
Ligatur AE	Æ	00C6
Ligatur oe	œ	0153
Ligatur OE	Œ	0152
Akut	´	dead key + <Vokal>
Gravis	`	shift+dead key+ <Vokal>
Cedille klein	ç	00E7
Cedille groß	Ç	00C7
Superskribiertes „e“	◌ ^e	0364
Kreis	å	030A
y mit Punkten	ÿ	00FF
Absatzzeichen	¶	00B6
Abkürzung „der“	ð	AltGr+d
Häkchen über Zeichen	ˇ	0309
Zeichen für lat. „recipe“	℞	211E
Zeichen für lat. „libra“	℔	2114
Zeichen für lat. „uncia“	℥	2125
Zeichen für lat. „scrupel“	℥	2108
3 Kreise	∴	2234
Kreis mit Punkt „einjährig“	⊙	2299
Latin Small Letter Rum Rotunda	ꝛ	A75D

Tabelle SEQ Tabelle * ARABIC 2: Zeichentabelle

Tabelle: Normalisierungsbeispiele

	clean	norm
1	-icht/-echt/-acht	-(I)ich/-(I)ig
	ästicht	ästig
	knöpfflicht	knöpfflich
	haarecht	haarig
	holzecht	holzig
	schattecht	schattig
	langlecht	länglich
	laulecht	laulig
	weißlecht	weißlich
	rohtlecht	rötlich
	stachelecht	stachelig
2	für-	vor-
	<i>fürnehmlich</i>	<i>vornehmlich</i>
3	-für	-vor
	herfür	hervor
4	fnhd. etwan	nhd. <i>etwa</i> (Adverb) nhd. <i>etwas</i> (Pronomen)
5	-en	-Ø
	<i>endlichen</i>	<i>endlich</i>
	<i>sonsten</i>	<i>sonst</i>
	<i>selbsten</i>	<i>selbst</i>
6	-Ø	-en
	<i>ob</i>	<i>oben</i>
	<i>beizeit</i>	<i>beizeiten</i>
7	(Nicht)Normalisierung von Kasus	
	<i>mit Normalisierung: Man trinke des wassers</i>	mit Normalisierung von Kasus: <i>Man trinke das Wasser</i>
	<i>Man trinke des wassers</i>	ohne Normalisierung von Kasus: <i>Man trinke des Wassers</i>
8	Vom Nhd. abweichende starke oder schwache Verbalflexion bzw. Rückumlaut wird dem nhd. Stand angepasst	

	<i>gennent</i>	<i>genannt</i>
	<i>gebauen</i>	<i>gebaut</i>
9	ggf. ge-Präfix in Partizipien ergänzen	
	funden	gefunden
	kommen	gekommen
	worden	geworden
10	Flexionsformen der Pronomina <i>der, die, das</i> an nhd. Formen anpassen	
	der	derer
	dero	derer
	des	dessen
11	fnhd. <i>wann/wenn</i>	nhd. wann (Interrogativpron) nhd. wenn (Subjunktion)
12	Einzelentscheidungen	
	benommen	genommen
	Beweisung	Beweisung
	derowegen	deswegen
	entbeut	entbietet
	gel	gel
	gleich(e)sfalls	gleichfalls
	halb	halber
	harm	Harn
	harmen	harnen
	leichtlich	leicht
	mehrer	mehrenteils
	teils	
	samlen	sammeln
	schlahen	schlagen
	(wegen der) Viele	Viele
	vierecket	viereckig
	vor	vorher
	wofer	wofern

zeuchen	ziehen
---------	--------

Dokumentation der Probleme bei der Normalisierung eines historischen Korpus – Ridges v.5

Normalisierung

Die Annotationsebene <norm> wurde händisch von Studierende erledigt (im Rahmen des Kurses „Historische Korpora“) und im Anschluss manuell korrigiert.

Erstellung von „clean“

Die Annotationsebene <clean> wird automatisch erstellt durch das Skript *cleanV22.py*, welches automatisch Sonderzeichen in den frühneuhochdeutschen Drucken durch äquivalente neuhochdeutsche Zeichen ersetzt. Zudem werden auf der <dipl> durch Zeilenumbruch getrennte Tokens in der <clean> zusammengezogen.

Bei der weiteren Bearbeitung in Excel kann es nach Anwendung des Clean-Skripts zu ungewollten automatischen Formatierungen kommen, ÜBERPRÜFEN!!!

1/2 wurde durch Excel zu *01.Februar*

Das normalisierte *falsch* wurde durch Excel zum logischen Operator *FALSCH*

Benutzte Tools:

Makros

SearchAndMerge.bas

Das Makro SearchAndMerge.bas sucht in einem markierten Bereich nach leeren Zellen (kleiner gleich 300 leere Zellen nacheinander) und vereinigt diese mit der letzten Vorgängerezelle, die einen Wert enthält.

ReplacePosOfUnclear.bas

Das Makro ReplacePosOfUnclear.bas ersetzt für alle in der norm-Ebene vorkommenden „unknown“-Werte, die bei unleserlichen Textstellen auftreten (vgl. Ebene „unclear“), in der pos-Ebene den ursprünglich vorhandenen Wert durch den Wert „XY“.

Python-Skript

cleanV22.py

Die neuen Texte stellten auch neue Anforderungen an die <clean>-Ebene, neue Zeichenersetzungen wurden nötig (siehe hierzu die cleanV2-README) sowie ein komplett neuer Umgang mit Nasalstrichen. Die bisherigen Texte ließen Zeichenersetzungen der Nasalstriche zu den jeweiligen Nasalen durch Kontextbetrachtung zu, während die neuen Texte viele von diesen alten Regularitäten nicht mehr erfüllen. Deshalb wird in der zweiten Version des clean-Skriptes anders mit solchen Nasalstrichen verfahren. Statt den Kontext zu betrachten und eine eindeutige Entscheidung zugunsten eines Nasals zu treffen, werden nun alle in Frage kommenden Zeichenersetzungen berücksichtigt und die möglichen Token werden durch „|“ getrennt dargestellt.

Shell-Skript

normalizeDipl.sh

Das Skript normalizeDipl.sh ersetzt eine Reihe an in der diplomatischen Ebene (fälschlicherweise) verwendeten Sonderzeichen durch andere festgelegte Sonderzeichen. Beispielsweise haben wir uns entschieden, keine Unterscheidung mehr zwischen Makra und Tilden vorzunehmen, da sich die beiden

Zeichen sehr ähneln und nicht immer auseinandergehalten werden können.