

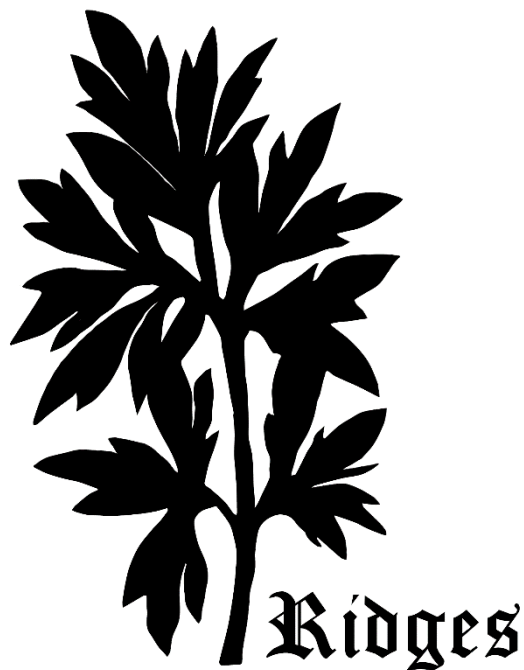
Dokumentation und Annotationsrichtlinien für das Korpus

Ridges Herbology Version 8.0 (ANNIS-und PAULA-Format*)

auf Grundlage des Metadatenframeworks nach LAUDATIO

Stand 06.04.2018

-deutsch-



Malte Belz
Carolin Odebrecht
Laura Perlitz
Gohar Schnelle
Vivian Voigt

*Diese setzen sich aus Annotationen zusammen, die in verschiedenen Formaten erstellt wurden. Alle (Zwischen-) Formate stehen auf korpling.org/ridges zum Download bereit, enthalten aber jeweils nicht alle hier aufgeführten Annotationsebenen, bzw. einige zusätzliche (irrelevante) Ebenen, die nicht nach ANNIS und PAULA konvertiert wurden. Eine ausführlichere Dokumentation der einzelnen Formate finden Sie im LAUDATIO-Repository unter laudatio-repository.org.

Inhalt

1. Korpus.....	7
2. Dokumente.....	11
2.1. Kraeuterbuch_1914_Losch.....	11
2.2. NatürlichePflanzenfamilien_1887_Engler	12
2.3. GemeinnuetzigesKraeuterbuch_1874_Siegmund	12
2.4. FloraDerPreussischenRheinlande_1870_Wirtgen.....	13
2.5. DeutschePflanzennamen_1870_Grassmann.....	13
2.6. VorlesungenUeberKraeuterkunde_1843_Link.....	14
2.7. GrossesIllustriertesKraeuterbuch_1860_Mueller	14
2.8. NochEinigeWorte-VR_1840_Meyen	15
2.9. NochEinigeWorte_1840_Meyen	15
2.10. EigenschaftenAllerHeilpflanzen_1828_Anonymous.....	16
2.11. EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous	16
2.12. Flora-7_1821_Sieber.....	17
2.13. Flora-6_1821_Wilbrand.....	17
2.14. GetreueDarstellungUndBeschreibung_1809_Hayne	18
2.15. Apothekerlexikon_1793_Hahnemann	18
2.16. GrundrissKraeuterkunde_1792_Willdenow	19
2.17. Unterricht_1774_Eisen	19
2.18. EinleitungZuDerKraeuterkenntnisz_1764_Oeder	20
2.19. BlackwellischesKraeuterbuch_1750_Blackwell.....	20
2.20. MysterivmSigillorvm_1735_Hiebner	21
2.21. FloraSaturnizans_1722_Henckel	21
2.22. ViridariumReformatum_1719_Valentini.....	22
2.23. TheatrumBotanicum_1696_Verzascha.....	22
2.24. SchweizerischerBotanicus_1687_vonRoll.....	23
2.25. SonderbaresKraeuterbuch-21-36_1675_Anonymous.....	24
2.26. SonderbaresKraeuterbuch-11-21_1675_Anonymous.....	24
2.27. SonderbaresKraeuterbuch-1-11_1675_Anonymous.....	25
2.28. ThesaurusSanitatis_304-321_1673_Nasser	25
2.29. ThesaurusSanitatis_1673_Nasser.....	26
2.30. Phythologia_1662_Becher	26
2.31. Wund-Artzney_1652_Greiff.....	27
2.32. PflantzGart-VR_1639_Rhagor	27

2.33.	PflantzGart-c4_1639_Rhagor	28
2.34.	PflantzGart_1639_Rhagor	29
2.35.	Kraeuterbuch_1609_Carrichter.....	29
2.36.	HortulusSanitatis_1609_Uffenbach.....	30
2.37.	AlchymistischePractic_1603_Libavius.....	30
2.38.	AlchymistischePractic-VR_1603_Libavius.....	31
2.39.	Paradeiszgaertlein_1588_Rosbach.....	31
2.40.	NewKreueterbuch_1563_Handsch	32
2.41.	WieSichMeniglich_1557_vonBodenstein	33
2.42.	WieSichMeniglich-VR_1557_vonBodenstein	33
2.43.	NewKreueterbuch_1543_Fuchs.....	34
2.44.	NewKreuterbuch-c19-25_1543_Fuchs	34
2.45.	NewKreuetterBuch_1539_Bock.....	35
2.46.	NewKreuetterbuch-CXXXIIIJ-CXXXVIIIJ_1539_Bock.....	36
2.47.	NewKreuetterBuch-VR_1539_Bock	36
2.48.	ContrafaytKreuterbuch_1532_Brunfels.....	37
2.49.	ContrafaytKreuterbuch- CCXXXVII-CCXLVIII_1532_Brunfels	37
2.50.	ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels	38
2.51.	ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels	39
2.52.	ContrafaytKreuterbuch-VR_1532_Brunfels.....	39
2.53.	ArtzneyBuchleinDerKreutter_1532_Tallat.....	40
2.54.	ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat	40
2.55.	ArtzneyBuchleinDerKreutter-Cretanus_1532_Tallat	41
2.56.	ArtzneyBuchleinDerKreutter-VR_1532_Tallat.....	41
2.57.	GartDerGesundheit_1487_vonCuba.....	42
2.58.	GartDerGesundheit-c15-23_1487_Cuba	43
2.59.	GartDerGesundheit-c23-33_1487_Cuba	43
2.60.	GartDerGesundheit-VR_1487_vonCuba.....	44
2.61.	BuchDerNatur_1482_vonMegenberg.....	44
3.	Annotationsebenen – Transkription/Normalisierung.....	45
3.1.	ocr	45
3.2.	dipl	47
3.3.	clean.....	50
3.4.	norm.....	52
4.	Annotationsebenen – Annotationen zu linguistischen Eigenschaften	55

4.1.	pos.....	55
4.2.	lemma	57
4.3.	cat.....	59
4.4.	cat-const.....	60
4.5.	func	63
4.6.	deprel	66
4.7.	lemma-deprel	69
4.8.	pos-deprel	71
4.9.	morph.....	73
4.10.	pos-const	75
4.11.	foreign.....	78
4.12.	foreign_trans	80
4.13.	lang.....	82
4.14.	sentence_end.....	84
4.15.	position_verb.....	86
4.16.	subclause_type	88
4.17.	KOUS_sem	90
4.18.	comp.....	92
4.19.	comp_orth	94
4.20.	prot	96
4.21.	comp_n.....	99
4.22.	comp_n_graph.....	101
4.23.	comp_n_mod.....	103
4.24.	comp_a	105
4.25.	comp_a_graph.....	107
4.26.	comp_amb.....	109
4.27.	infl_fuge	111
4.28.	comp_lex.....	113
4.29.	adja_uninfl	116
4.30.	attr_gen	117
4.31.	morph_ellipsis.....	119
4.32.	comment_lex.....	121
4.33.	comment.....	122
4.34.	reference.....	124
4.35.	form_disease	126

4.36.	problem	128
4.37.	herbname_norm	129
4.38.	herbprep	131
4.39.	form_prep.....	133
4.40.	noun_nom	135
4.41.	form_noun.....	136
4.42.	clause_type.....	139
4.43.	position_rel	141
4.44.	position_referent	143
4.45.	form_referent	145
4.46.	position_verb_rel	147
4.47.	form_relpron	149
4.48.	mod_referent	150
4.49.	ppk_e1	152
4.50.	ppk_e2.....	154
4.51.	ppk_e3.....	155
5.	Annotationsebenen – Strukturelle Annotation.....	157
5.1.	lb	157
5.2.	column	160
5.3.	pb	162
5.4.	pb_n	165
5.5.	pb_ana.....	167
5.6.	unclear	169
5.7.	atLeast.....	171
5.8.	atMost	173
5.9.	interpretation.....	175
5.10.	figure	177
5.11.	figure_p	186
5.12.	hi	188
5.13.	script.....	190
5.14.	hi_rend	191
5.15.	head	193
5.16.	note.....	195
5.17.	ref	197
5.18.	ref_target.....	199

5.19.	ref_type	201
5.20.	quote.....	203
6.	Annotationsebenen – Inhaltliche Annotation.....	205
6.1.	definition.....	205
6.2.	disease.....	207
6.3.	persname	209
6.4.	title	213
6.5.	author_ref.....	214
6.6.	reader_ref.....	216
6.7.	plant	218
6.8.	property.....	220
6.9.	name.....	222
6.10.	name_type	224
7.	Metadaten.....	226
8.	Referenzen	230
Anhang	231
	Zuordnung von Annotationen und Dokumenten.....	231
	Transkription und Normalisierung	248
	In Verbindung mit Excel benutzte Tools	272

1. Korpus

Typ: Corpus – Ridges Herbology

Name:	Ridges Herbology
Herausgeber:	Anke Lüdeling, Carolin Odebrecht, Laura Perlitz, Amir Zeldes, Korpuslinguistik, Humboldt-Universität zu Berlin
Annotatoren:	Ilham Abed-Ali, Silke Andresen, Henriette Ast, Valeska Beckert, Malte Belz, Katrin Benning, Thomas Berg, Cornelia Binnyus, Simon Blum, Doreen Christen, Janin Czilwik, Mascha Dayal, Antonia Dittberner, Cora Döhn, Imke Driemel, Christian Ebert, Olja Efremova, Gill-Maria Eichhorn, Stefanie Enneper, Judith Esser, Catharina Fischer, Felix Gehrke, Annegret Gerlach, Linda Giesel, Rahel Gajaneh Hartz, Yelyzaveta Hiebert, Hagen Hirschmann, Bodil Jessen, Keqin Ji, 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, Annika Mittelstädt, Lisa Mohs, Sandra Müller, Vinzent Müller, Andrew Murphy, Johannes Mursell, Carolin Odebrecht, Akiko Okuda, Mareike Otto, Maik Papenhagen, Angelo Papenhoff, Laura Perlitz, Viktoria Peselmann, Joachim Polte, Gina Reetz, Valerie Reichardt, Katharina Reinig, Ina Riesler, Lena Rosin, Romy Sachs, Franz-Josef Sachse, Anna Sapronova, Simon Sauer, Jan Christian Schaffert, Anna Greta Schmahl, Claudia Schmidt, Gohar Schnelle, Athina Sioupi, Sarah Sippach, Andrea Skotovic, Lucia Sohmen, Iryna Sorokovska, Madeleine Spitzer, Uwe Springmann, Kristina Stephan, Helen Stuke, Aleksandra Swiech, Juliane Tiemann, Anna Tóth, Réka Tóth, Phuong Thao Tran, Alexander Turtureanu, Hanna Varachkina, Magdalena Vock, Mirjam Vogler, Monika Walak, Juliana Wekel, Vita Rosalie Wijffels, Svenja Wolter, Isabell Ines Zander, Karolina Zuchewicz (nachfolgend Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen genannt)
Veröffentlichung:	06.04.2018
Anzahl der Textausschnitte:	61
Zeitraum:	1482-1914
dipl-Token:	257537
Sprache:	Deutsch
Register:	Kräutertexte

Zugang:	https://korpling.german.hu-berlin.de/annis3/
Projekt:	http://korpling.org/ridges
Projektbeschreibung:	<p>Das RIDGES-Projekt (Register in Diachronic German Science) 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.</p> <p>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.</p> <p>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.</p>
Forschungsfrage:	<p>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.</p>
Version:	8.0
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, Dependenzparses und Konstituentenbäumen versehen.

Art der Änderung:

Folgende Dokumente wurden in Version 8.0 hinzugefügt:

SonderbaresKraeuterbuch-21-36_1675_Anonymous

ThesaurusSanitatis_304-321_1673_Nasser

- da in ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat ein Teil des Textes bereits in

ArtzneyBuchleinDerKreutter_1532_Tallat vorhanden war, wurde dieser redundante Teil entfernt und die Datei umbenannt in ArtzneyBuchleinDerKreutter-Cretanus_1532_Tallat

- aus der Datei NochEinigeWorte_1840_Meyen wurde das Vorwort entnommen und in eine separate Datei NochEinigeWorte-VR_1840_Meyen eingefügt

- die Ebene lemma-dep wurde in lemma-deprel und die Ebene pos-dep in pos-deprel umbenannt

- die Dependenzparsed in der Ebene deprel und die dazugehörigen Annotationsebenen lemma-deprel und pos-deprel wurden in HortulusSanitatis_1609_Uffenbach und ContrafaytKreuterbuch_1532_Brunfels manuell und durchgängig korrigiert.

- durchgängige Korrektur der pb_n-Annotationen (Unterscheidung von lateinischen und römischen Zahlenangaben)

- durchgängige Korrektur der norm-Ebene in den neuen Dokumenten der Version 7:

ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat

ArtzneyBuchleinDerKreutter-Cretanus_1532_Tallat

ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels

ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels

GartDerGesundheit-c15-23_1487_Cuba

GartDerGesundheit-c23-33_1487_Cuba

NewKreutterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock

NewKreuterbuch-c19-25_1543_Fuchs

- die Ebenen figure_rend, item und nlp_morph wurden gelöscht

- in SonderbaresKraeuterbuch-1-11_1675_Anonymous und SonderbaresKraeuterbuch-11-21_1675_Anonymous wurden die Annotationsebenen „figure“ und „figure_p“ hinzugefügt
- der Wert „end“ in hi_rend wurde entfernt (Annotationsfehler)
- punktuelle Korrekturen in comp, comp_orth und prot
- in den Dokumenten, die nach 1652 erschienen sind, wurden alle „strD“-Werte in der Ebene „morph_ellipsis“ durch „morph_ellipsis“ ersetzt (dieser Schritt wurde in Version 6 nur für Dokumente durchgeführt, die bis einschließlich 1652 erschienen sind)
- in ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat wurden in der norm-Ebene 82 Leerzeichen entfernt sowie versteckte Werte aus einem Makro gelöscht, die in ANNIS als norm-Werte angezeigt wurden

Enthaltende Dokumente (Kürzel):

BuchDerNatur_1482_vonMegenberg
 GartDerGesundheit_1487_vonCuba
 GartDerGesundheit-c15-23_1487_Cuba
 GartDerGesundheit-c23-33_1487_Cuba
 GartDerGesundheit-VR_1487_vonCuba
 ArtzneyBuchleinDerKreutter_1532_Tallat
 ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat
 ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat
 ArtzneyBuchleinDerKreutter-VR_1532_Tallat
 ContrafaytKreuterbuch_1532_Brunfels
 ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels
 ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels
 ContrafaytKreuterbuch-VR_1532_Brunfels
 ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels
 NewKreuetterBuch_1539_Bock
 NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock
 NewKreuetterBuch-VR_1539_Bock
 NewKreuterbuch_1543_Fuchs
 NewKreuterbuch-c19-25_1543_Fuchs
 WieSichMeniglich_1557_vonBodenstein
 WieSichMeniglich-VR_1557_vonBodenstein
 NewKreuetterbuch_1563_Handsch
 Paradeiszgaertlein_1588_Rosbach
 AlchymistischePractic_1603_Libavius
 AlchymistischePractic-VR_1603_Libavius
 HortulusSanitatis_1609_Uffenbach
 Kraeutterbuch_1609_Carrichter
 PflantzGart-c4_1639_Rhagor
 PflantzGart_1639_Rhagor

PflantzGart-VR_1639_Rhagor
 Wund-Artzney_1652_Greiff
 Phythologia_1662_Becher
 ThesaurusSanitatis_1673_Nasser
 ThesaurusSanitatis_304-321_1673_Nasser
 SonderbaresKraeuterbuch-1-11_1675_Anonymous
 SonderbaresKraeuterbuch-11-21_1675_Anonymous
 SonderbaresKraeuterbuch-21-36_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
 NochEinigeWorte-VR_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, ppk_e2, prot, quote, reader_ref, ref, ref_target, script, subclause_type

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-VR_1840_Meyen

Titel: Noch einige Worte über den Befruchtungsakt und die Poljembryonie bei den höheren Pflanzen (Vorwort)
Kürzel: NochEinigeWorte-VR_1840_Meyen
Autor: Meyen, Franz Julius Ferdinand
Jahr: 1840
Ort: Berlin
Verlag: Haude und Spenersche Buchhandlung (S. J. Joseephy.)
Seitenangabe: III-IV
Publikationstyp: Monographie
dipl-Einheiten: 365
Beschreibung: Kräutertext
Annotierte Ebenen: attr_gen, author_ref, comp, comp_orth, head, hi, hi_rend, KOUS_sem, lang, lb, lemma, name, name_type, note, pb, pb_n, pos, position_verb, ppk_e1, prot, quote, ref, ref_target, ref_type, script, subclause_type

2.9. 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: 7619
Beschreibung: Kräutertext
Annotierte Ebenen: attr_gen, author_ref, comp, comp_orth, foreign, foreign_trans, 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

2.10. 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

2.11. 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

2.12. 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

2.13. 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

2.14. 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

2.15. 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

2.16. 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

2.17. 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, 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

2.18. 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

2.19. 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

2.20. 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

2.21. 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

2.22. 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

2.23. 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 vorgestellet sind. Allen Aerzten sonderlich auch denen auff dem Land wohnenden höchst nutzlich 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

2.24. SchweizerischerBotanicus_1687_vonRoll

Titel: Der Schweizerische Botanicus
Kürzel: SchweizerischerBotanicus_1687_vonRoll
Autor: Timotheus von Roll
Jahr: 1687
Ort: NA
Verlag: Franz Carl Rooß
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 in der Transkription zugrunde liegenden PDF doppelt vorhanden und wurden von den Studierende auch doppelt annotiert.

2.25. SonderbaresKraeuterbuch-21-36_1675_Anonymous

Titel: Curioser Botanicus oder sonderbares Kräuter Buch
Kürzel: SonderbaresKraeuterbuch-11-36_1675_Anonymous
Autor: NA
Jahr: 1675
Ort: NA
Verlag: NA
Seitenangabe: 21-36
Publikationstyp: Monographie
dipl-Einheiten: 2876
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, figure, figure_p, head, hi_rend, lang, lb, lemma, pb, pb_ana, pb_n, persname, pos, quote, script, unclear

2.26. 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, figure, figure_p, 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

2.27. 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, figure, figure_p, 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

2.28. ThesaurusSanitatis_304-321_1673_Nasser

Titel: Thesaurus Sanitatis
Kürzel: ThesaurusSanitatis_1673_Nasser
Autor: Adrian Nasser
Jahr: 1673
Ort: Nürnberg

Verlag: Johann Hoffmann
 Seitenangabe: 304-321
 Publikationstyp: Monographie
 dipl-Einheiten: 4493
 Beschreibung: Kräutertext
 Annotierte Ebenen: author_ref, comment, head, hi_rend, lang, lb, lemma, pb, pb_n, persname, pos, quote, reader_ref, script, title

2.29. 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

2.30. 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

2.31. 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, cat-const, comment , deprel, disease, foreign, func, head, hi, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, lemma-deprel, morph_ellipsis, morph, note, pb, pb_n, plant, pos, pos-dep, pos-const, position_verb, ppk_e1, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.32. 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, foreign, foreign_trans, head, hi, hi_rend, KOUS_sem, lang, lb, lemma, note, pb, pb_n, pos, position_verb, quote, reader_ref, reference, script, sentence_end, subclause_type, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.33. 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, 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, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.34. 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, figure, head, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, note, pb, pb_n, pos, position_verb, ppk_e1, reference, script, sentence_end, subclause_type, unclear, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.35. Kraeuterbuch_1609_Carrichter

Titel: Kräuterbuch des Edelen und hochgelehrten herren Bartholomei Carrichters
Kürzel: Kraeuterbuch_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, disease, head, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, plant, pos, position_verb, ppk_e1,

reader_ref, reference, script, sentence_end, subclause_type, unclear, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, lemma-const

2.36. HortulusSanitatis_1609_Uffenbach

Titel: Hortulus Sanitatis
Kürzel: HortulusSanitatis_1609_Uffenbach
Autor: Peter Uffenbach
Jahr: 1609
Ort: Frankfurt am Main
Verlag: Ioniae Rhodij
Seitenangabe: 1-21
Publikationstyp: Monographie
dipl-Einheiten: 6516
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, comment, 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, quote, reference, script, sentence_end, subclause_type, unclear, deprel, deprelGold, morph, cat-const, func, pos-deprel, posGold, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.37. 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, 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, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.38. AlchymistischePractic-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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.39. Paradeiszgaertlein_1588_Rosbach

Titel: Paradeiszgärtlein
Kürzel: Paradeiszgaertlein_1588_Rosbach

Autor: Conrad Rosbach
 Jahr: 1588
 Ort: Frankfurt am Main
 Verlag: Johann Spieß
 Seitenangabe: 1-43
 Publikationstyp: Monographie
 dipl-Einheiten: 5054
 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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.40. 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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.41. **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, disease, foreign, foreign_trans, head, herbname_norm, interpretation, KOUS_sem, lang, lb, lemma, morph_ellipsis, pb, pb_n, persname, plant, pos, position_verb, ppk_e1, quote, reader_ref, reference, script, sentence_end, subclause_type, unclear, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.42. **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:	707
Beschreibung:	Kräutertext

Annotierte Ebenen: author_ref, comment, head, hi_rend, lang, lb, lemma, pb, persname, pos, quote, reader_ref, script, sentence_end, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.43. 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, 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, quote, reader_ref, script, sentence_end, subclause_type, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.44. NewKreuterbuch-c19-25_1543_Fuchs

Titel: New Kreüterbuch
Kürzel: NewKreuterbuch-c19-25_1543_Fuchs

Autor: Fuchs, Leonhard
 Jahr: 1543
 Ort: Basel
 Verlag: Michael Isingrin
 Seitenangabe: Kapitel 19-25
 Publikationstyp: Monographie
 dipl-Einheiten: 2821
 Beschreibung: Kräutertext
 Annotierte Ebenen: author_ref, comment, figure, figure_p, head, hi_rend, lang, lb, lemma, note, ocr, pb, pos, quote, persname, reader_ref, script, title, comp_n, comp_n_graph, comp_n_mod, infl_fuge, attr_gen, adja_uninfl, morph_ellipsis, comp_a, comp_a_graph, comp_lex, comp_amb

2.45. **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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.46. NewKreuetterbuch-CXXXIIIJ-CXXXVIIIJ_1539_Bock

Titel: New Kreütter Buch: von Unterscheydt, 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-CXXXIIIJ-CXXXVIIIJ_1539_Bock

Autor: Hieronymus Bock

Jahr: 1539

Ort: Straßburg

Verlag: Rihel

Seitenangabe: CXXXIIIJ-CXXXVIIIJ

Publikationstyp: Monographie

dipl-Einheiten: 4055

Beschreibung: Kräutertext

Annotationsebenen: author_ref, comment, head, hi_rend, lang, lb, lemma, note, ocr, pb, pb_n, persname, pos, reader_ref, reference, script, sentence_end, title, comp_n, comp_n_graph, comp_n_mod, infl_fuge, attr_gen, adja_uninfl, morph_ellipsis, comp_a, comp_a_graph, comp_lex, comp_amb

2.47. NewKreuetterBuch-VR_1539_Bock

Titel: New Kreütter Buch: von Unterscheydt, 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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.48. 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, disease, figure, head, herbname_norm, hi_rend, interpretation, KOUS_sem, lang, lb, lemma, mod_referent, pb, pb_n, persname, plant, pos, position_verb, ppk_e1, reader_ref, script, sentence_end, subclause_type, unclear, deprel, deprelGold, morph, cat-const, func, pos-deprel, pos-const, posGold, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.49. ContrafaytKreuterbuch- CCXXXVII-CCXLVIII_1532_Brunfels

Titel: Contrafayt kreüterbuch
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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.50. **ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels**

Titel: Contrafayt kreüterbuch
 Kürzel: ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels
 Autor: Otto Brunfels
 Jahr: 1532
 Ort: Straßburg
 Verlag: Hans Schotten
 Seitenangabe: CCLVII-CCLXX
 Publikationstyp: Monographie
 dipl-Einheiten: 3975
 Beschreibung: Kräutertext
 Annotierte Ebenen: author_ref, comment, figure, figure_p, hi_rend, lang, lb, lemma, note, ocr, pb, pb_n, persname, pos, quote, reader_ref, reference, script, sentence_end, title, comp_n, comp_n_graph, comp_n_mod, infl_fuge, attr_gen, adja_uninfl, morph_ellipsis, comp_a, comp_a_graph, comp_lex, comp_amb

2.51. **ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels**

Titel: Contrafayt kreüterbuch
Kürzel: ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels
Autor: Otto Brunfels
Jahr: 1532
Ort: Straßburg
Verlag: Hans Schotten
Seitenangabe: CCLXXI-CCXC
Publikationstyp: Monographie
dipl-Einheiten: 4401
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, figure, figure_p, hi_rend, lang, lb, lemma, note, ocr, pb, pb_n, persname, pos, quote, reader_ref, reference, script, sentence_end, title, comp_n, comp_n_graph, comp_n_mod, infl_fuge, attr_gen, adja_uninfl, morph_ellipsis, comp_a, comp_a_graph, comp_lex, comp_amb

2.52. **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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel,

adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph,
comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.53. 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, 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, reader_ref, script,
sentence_end, subclause_type, title, unclear, deprel, morph, cat-
const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl,
comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod,
comp_a, comp_a_graph, infl_fuge

2.54. ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat

Titel: Artzney Buchlein der kreutter
Kürzel: ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat
Autor: Johannes Tallat
Jahr: 1532
Ort: Leipzig

Verlag: Michael Blum
 Seitenangabe: ab Abrotanum
 Publikationstyp: Monographie
 dipl-Einheiten: 3530
 Beschreibung: Kräutertext
 Annotierte Ebenen: author_ref, head, hi_rend, lang, lb, lemma, ocr, pb, persname, pos, quote, reader_ref, script, title, unclear, comp_n, comp_n_graph, comp_n_mod, infl_fuge, attr_gen, adja_uninfl, morph_ellipsis, comp_a, comp_a_graph, comp_lex, comp_amb

2.55. ArtzneyBuchleinDerKreutter-Cretanus_1532_Tallat

Titel: Artzney Buchlein der kreutter
 Kürzel: ArtzneyBuchleinDerKreutter-Cretanus_1532_Tallat
 Autor: Johannes Tallat
 Jahr: 1532
 Ort: Leipzig
 Verlag: Michael Blum
 Seitenangabe: Cretanus bis Dens leonis
 Publikationstyp: Monographie
 dipl-Einheiten: 1120
 Beschreibung: Kräutertext
 Annotierte Ebenen: ocr, dipl, clean, norm, infl_fuge, comp_amb, comp_n, comp_n_graph, comp_n_mod, comp_lex, comp_a, comp_a_graph, adja_uninfl, attr_gen, morph_ellipsis, lang, pb, pb_n, lb, head, hi_rend, script, author_ref, quote, comment

2.56. 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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.57. 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, disease, figure, form_disease, form_noun, form_prep, form_referent, form_relpron, head, herbyname_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, reader_ref, script, sentence_end, subclause_type, title, unclear, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.58. GartDerGesundheit-c15-23_1487_Cuba

Titel: Gart der Gesundheit
Kürzel: GartDerGesundheit_1487_vonCuba
Autor: Johannes von Cuba
Jahr: 1487
Ort: Ulm
Verlag: NA
Seitenangabe: Kapitel 15-23
Publikationstyp: Monographie
dipl-Einheiten: 3922
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, column, figure, head, hi_rend, lang, lb, lemma, ocr, pb, persname, pos, quote, reader_ref, script, title, unclear, comp_n, comp_n_graph, comp_n_mod, infl_fuge, attr_gen, adja_uninfl, morph_ellipsis, comp_a, comp_a_graph, comp_lex, comp_amb

2.59. GartDerGesundheit-c23-33_1487_Cuba

Titel: Gart der Gesundheit
Kürzel: GartDerGesundheit-c23-33_1487_Cuba
Autor: Johannes von Cuba
Jahr: 1487
Ort: Ulm
Verlag: NA
Seitenangabe: Kapitel 23-33
Publikationstyp: Monographie
dipl-Einheiten: 3843
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, comment, column,figure, head, hi_rend, lang, lb, lemma, ocr, pb, persname, pos, quote, reader_ref, script, title, comp_n, comp_n_graph, comp_n_mod, infl_fuge, attr_gen, adja_uninfl, comp_lex, comp_amb

2.60. 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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

2.61. 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, deprel, morph, cat-const, func, pos-deprel, pos-const, lemma-deprel, adja_uninfl, comp_amb, comp_lex, comp_n, comp_n_graph, comp_n_mod, comp_a, comp_a_graph, infl_fuge

3. Annotationsebenen – Transkription/Normalisierung

3.1. ocr

Typ: *Layer* – ocr

Diese Ebene wurde ausschließlich für die neuen Dokumente der Version 7 erstellt (insgesamt 8).

Name:	ocr
Dokumentation:	Siehe Ridges-Website https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/ridges-projekt/ocr
ocr-Einheiten:	30390
Beschreibung:	Diese Ebene enthält OCR-erkannten Text in den Dokumenten, die in der Version 7 hinzugekommen sind. Die OCR wurde mit eigens für diese Titel trainierten Modellen für die OCR-Engine OCRopus durchgeführt.

Typ: Preparationstep – ocr

Schritt:	1
Aufbereitung:	Die OCR wurde mit eigens für diese Titel trainierten Modelle für die OCR-Engine OCRopus durchgeführt.
Aufbereitungsart:	automatisch
Tool	OCRopus
Format:	Text-Datei, zeilenweise
Segmentierung der Annotation:	eigenständig
Qualitätsprüfer:	NA
Datum:	2017
Qualitätsprüfung:	NA
Editor:	Uwe Springmann
Schritt:	2
Aufbereitung:	Zusammenführung der OCR-Zeilen

Aufbereitungsart: automatisch
Tool Unix shell
Format: Text-Datei
Segmentierung der Annotation: eigenständig
Qualitätsprüfer: NA
Datum: 2017
Qualitätsprüfung: NA
Editor: Uwe Springmann

Schritt: 3
Aufbereitung: Tokenisierung
Aufbereitungsart: automatisch
Tool TreeTagger Version 3.2
Format: TreeTagger-Output
Segmentierung der Annotation: eigenständig
Qualitätsprüfer: NA
Datum: 2017
Qualitätsprüfung: NA
Editor: Uwe Springmann

Schritt: 3
Aufbereitung: Überführung nach Excel 2013
Aufbereitungsart: manuell
Tool Excel 2013
Format: Excel 2013
Segmentierung der Annotation: eigenständig
Qualitätsprüfer: NA
Datum: 2017
Qualitätsprüfung: NA
Editor: Uwe Springmann

Schritt:	4
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	Automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	eigenständig
Qualitätsprüfer:	Korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

3.2. dipl

Typ: *Layer* - dipl

Name:	dipl
Dokumentation:	Siehe Ridges-Website: https://hu.berlin/ridges-dokumentation-v7 Beispiele: Siehe Anhang
dipl-Einheiten:	252777
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!
	ſ wird nicht von Z unterschieden. ß wird in Frakturtexten als ſz transkribiert, da hier eine Unterscheidung zwischen der Ligatur und der beiden Einzelzeichen häufig schwierig ist.

Typ: Preparationstep – dipl

Die dipl-Ebene wird für gewöhnlich folgendermaßen erstellt (für die Aufbereitungsschritte in den hinzugekommenen Dokumenten der Version 7 siehe weiter unten):

Schritt: 1a
Aufbereitung: Transkript
Aufbereitungsart: manuell
Tool: NA
Format: Text-Datei
Segmentierung der Annotation: eigenständig
Qualitätsprüfer: Korpling
Datum: 2018
Qualitätsprüfung: manuell
Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt: 2a
Aufbereitung: Tokenisierung
Aufbereitungsart: Automatisch
Tool: Treetagger Version 3.2
Format: Treetagger-Output
Segmentierung der Annotation: eigenständig
Qualitätsprüfer: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling
Datum: 2018
Qualitätsprüfung: manuelle Konsistenzprüfung
Editor: Laura Perlitz

Schritt: 3a
Aufbereitung: Konvertierung von Excel 2010 und Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch
 Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
 Format: PAULA-XML, ANNIS
 Segmentierung der Annotation: eigenständig
 Qualitätsprüfer: korpling
 Datum: 2018
 Qualitätsprüfung: NA
 Editor: LAUDATIO, LangBank

In den hinzugekommenen Dokumenten der Version 7 wurde die dipl-Ebene folgendermaßen erstellt:

Schritt: 1b
 Aufbereitung: Auf Basis des automatisch erstellten und tokenisierten OCR (siehe ocr) entsteht eine (korrigierte) diplomatische Transkription, die den gleichen Regeln folgt, wie die dipl-Ebene in den vorigen Dokumenten.

Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2013
 Segmentierung der Annotation: eigenständig
 Qualitätsprüfer: Laura Perlitz
 Datum: 2017
 Qualitätsprüfung: manuell
 Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2b
 Aufbereitung: Konvertierung von Excel 2010 und Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	eigenständig
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

3.3. clean

Typ: *Layer* – clean

Name:	clean
Dokumentation:	Siehe Ridges-Website http://korpling.org/ridges Beispiele: Siehe Anhang
clean-Einheiten:	246204
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 (*Speifz=und Nahrungs=Saffts* 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	clean-skript_V3.py (https://hu.berlin/clean-script-v3)
Format:	Clean-Skript-Output
Segmentierung der Annotation:	eigenständig
Qualitätsprüfer:	Korpling
Datum:	2017
Qualitätsprüfung:	manuelle Konsistenzprüfung
Editor:	Korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	Automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter.

Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: Eigenständig
Qualitätsprüfer: Korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

3.4. norm

Typ: *Layer* – norm

Name: Norm
Dokumentation: Siehe Ridges-Website
<http://korpling.org/ridges>
Beispiele: Siehe Anhang
norm-Einheiten: 244532
Beschreibung: Die norm-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier Tokenisierung, Graphematik, (veraltete) 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:

Morphologisch wird dann normalisiert, wenn keine Flexion vorhanden ist oder wenn Flexion vorhanden ist, die heute so nicht mehr existiert. In diesem Fall verändern wir nicht den Kasus selbst, sondern nur die Form.

das kal Haupt → *das kahle Haupt*
funden → *gefunden*
in die Nasen → *in die Nase* (bleibt Akkusativ, nur die Form wird normalisiert)
dem Haupte → *dem Haupt* (bleibt Dativ, nur die Form wird normalisiert)

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

trinke des Wassers → *trinke des Wassers* (vs. nhd. Akkusativ *trinke das Wasser*, bleibt in der norm Genitiv)
das Milz → *das Milz* (vs. nhd. Femininum *die Milz*, bleibt in der norm Neutrum)

Lexikologie:

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

Vergefz → *Vergess* (comment: *das Vergessen*)

Wortbildung:

Ausgestorbene Wortbildungsmorpheme werden nach Möglichkeit durch nhd. Bildungen ersetzt. Es werden jedoch keine Affixe ergänzt, die nicht da sind, außer es betrifft die Bildung von Partizipialformen mit *ge-*.

halben → *halber*

stachelecht → *stachelig*

Anzündet vff den kolen / vñ gereüchet / vertreibet die schlangē → *Angezündet*

Es meret auch der frawen milch → *mehrt* (nicht *vermehrt*)

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.

Fußnoten innerhalb von Fließtext

Wenn „Anker“ von Fußnoten den Fließtext unterbrechen, dann werden diese in der norm nicht übernommen. Da die Information jedoch zumindest in dipl beibehalten werden soll, wird die entsprechende Zelle, die einen solchen Anker enthält, mit der folgenden Zelle verbunden und nur der Folgewert in der norm eingetragen, also das folgende Token. Sollte der Anker innerhalb einer größeren norm-Einheit stehen, dann werden alle Zellen, die zu dieser Einheit gehören und die Zelle mit dem Anker zu einer Zelle verbunden. Der Anker wird in einer Annotationsebene „ref“ mit dem Wert „ref“ markiert., vgl. folgendes Beispiel:

/	/	/	
g	g	Ellen	ref
Ehlen	Ehlen		
hoch	hoch	hoch	
.	.	.	
Die	Die	Die	
er-		ersten	
fte	erste		
fo	so	so	
heraufz	herausz	herauskommen	
kommen	kommen		
/	/	/	
find	sind	sind	
halb	halb	halbrund	
rund	rund		
/	/	/	
dem	dem	dem	
Mertzẽ	Mertzem Mertzen	Märzveilchenkraut	
h	h		ref
veyels	veyels		
kraut	kraut		
gleich	gleich	gleich	
/	/	/	

Typ: Preparationstep – norm

Schritt: 1

Aufbereitung: Normalisierung, Annotation

Aufbereitungsart: Manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: Eigenständig

Qualitätsprüfer: Korpling

Datum: 2018

Qualitätsprüfung: manuelle Korrektur

Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt: 2

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

Aufbereitungsart: Automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und

PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: Eigenständig
Qualitätsprüfer: Korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

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:	2017
Qualitätsprüfung:	NA
Editor:	Korpling
Schritt:	2
Aufbereitung:	Überführung des TreeTagger-Outputs nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	Korpling
Datum:	2017
Qualitätsprüfung:	NA
Editor:	Korpling
Schritt:	3
Aufbereitung:	Verbinden von leeren Zellen und der jeweils davor stehenden Zelle, die einen Wert enthält
Aufbereitungsart:	Automatisch
Tool:	SearchAndMerge.bas (https://hu.berlin/search-and-merge)
Format:	EXCEL
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	Korpling
Datum:	2017
Qualitätsprüfung:	Korpling
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/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: Korpling
Datum: 2018
Qualitätsprüfung: Korpling
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – pos

Annotationsart: Spannenannotation je norm-Einheit
Beschreibung: Wortarten-Annotation nach dem Stuttgart-Tübingen-Tagset.

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:	2017
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	Korpling
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – lemma

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

Typ: Annotationvalue – lemma

Wert: Wertbeschreibung:
Normalisiertes Lemma Lemmatisierung.

4.3. cat

Typ: *Layer* – cat

Name: cat
Dokumentation: Spannenannotation von Sätzen, die beim Konvertieren von CoNLL nach ANNIS ausgegeben wird, wenn „conll.SENTENCE“ im CoNLLImporter aktiviert ist. Mehr Informationen zum Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – cat

Schritt: 1
Aufbereitung: Konvertierung von CoNLL nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf ‘norm’
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer– cat

Annotationsart: Spannenannotation über mehrere norm-Einheiten
Beschreibung: Spannenannotation von Sätzen, die beim Konvertieren von CoNLL nach ANNIS ausgegeben wird, wenn „conll.SENTENCE“ im CoNLLImporter aktiviert ist.

Typ: Annotationvalue – cat

Wert:	Wertbeschreibung:
S	Satz

4.4. cat-const

Typ: *Layer* – cat-const

Name: cat-const
Dokumentation: Annotation von Konstituenten, die automatisch mit Hilfe des Berkeley Parsers (<https://github.com/slavpetrov/berkeleyparser>) aus dem Output des Mate Dependenzparsers abgeleitet wurden (<http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html>). In cat-const sind die Kategorien der Konstituenten annotiert. Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – cat-const

Schritt:	1
Aufbereitung:	Dependenzparsing
Aufbereitungsart:	Automatisch

Tool: Mate Dependenzparser (parser-ger-3.6.model)

Format: CoNLL 2009

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: Gohar Schnelle

Datum: 2017

Qualitätsprüfung: manuell (nur teilweise)

Editor: LAUDATIO, LangBank

Schritt: 2

Aufbereitung: Transformation der Abhängigkeiten in Konstituenten

Aufbereitungsart: Automatisch

Tool: Berkeley Parser 1.7

Format: PTB

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: NA

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Schritt: 3

Aufbereitung: Konvertierung nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer– cat-const

Annotationsart: Baumannotation (Konstituenten)
Beschreibung: Konstituentenannotation basierend auf dem TIGER-
Annotationsschema (https://www.linguistics.ruhr-uni-bochum.de/~dipper/pub/tiger_annot.pdf)

Typ: Annotationvalue – cat-const

Wert:	Wertbeschreibung:
AA	Superlativphrase mit “am”
AP	Adjektivphrase
AVP	Adverbialphrase
CAP	koordinierte Adjektivphrase
CAVP	koordinierte Adverbialphrase
CH	Chunk
CNP	koordinierte Nominalphrase
CO	Koordination
CPP	koordinierte Präpositionalphrase
CS	koordinierter Satz
CVP	koordinierte Verbalphrase (infinit)
CVZ	koordinierter Infinitiv mit “zu”
DL	Diskurslevelkonstituente
ISU	idiosynkratische Einheit
NP	Nominalphrase
PN	Personenname
PP	Präpositionalphrase
PSEUDO	NA
ROOT	Wurzel
S	Satz

TOP	NA
VP	Verbalphrase
VZ	Infinitiv mit “zu”

4.5. func

Typ: *Layer* – func

Name:	func
Dokumentation:	Kantenannotationen von Konstituenten, die automatisch mit Hilfe des Berkeley Parsers (https://github.com/slavpetrov/berkeleyparser) aus dem Output des Mate Dependenzparsers abgeleitet wurden (http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html). In func sind die Kanten annotiert. Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – func

Schritt:	1
Aufbereitung:	Dependenzparsing
Aufbereitungsart:	Automatisch
Tool:	Mate Dependenzparser (parser-ger-3.6.model)
Format:	CoNLL 2009
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	NA
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank
Schritt:	2

Aufbereitung: Transformation der Abhängigkeiten in Konstituenten
Aufbereitungsart: Automatisch
Tool: Berkeley Parser 1.7
Format: PTB
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Schritt: 3
Aufbereitung: Konvertierung nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer– func

Annotationsart: Baumannotation (Kanten)
Beschreibung: Konstituentenannotation mit Kategorie- und Kantenannotationen basierend auf dem TIGER-Annotationsschema (<https://www.linguistics.ruhr-uni->

Typ: Annotationvalue – func

Wert:	Wertbeschreibung:
AC	adpositioneller Kasusmarker
AG	Genitivattribut
AMS	Maßangabe, die Argument eines Adjektivs ist
AVC	Komponente einer Adverbialphrase
APP	Apposition
CC	komparatives Komplement
CD	koordinierende Konjunktion
CJ	Konjunkt
CM	komparaive Konjunktion
CP	Komplementierer
CVC	Funktionsverbgefüge
DA	Dativ
DH	Diskurslevelkopf
DM	Diskursmarker
EP	expletives “es”
FM	Fremdsprachliches Material
HD	Kopf
JU	Junktor
MNR	postnominaler Modifikator
MO	Modifikator
NG	Negation
NK	Element der Kern-NP
NMC	numerale Komponente
OA	Akkusativobjekt
OA2	zweites Akkusativobjekt
OC	Objektsatz
OG	Genitivobjekt
PAR	Parenthese
PD	Prädikat
PG	Phrasaler Genitiv

PH	Platzhalter
PM	morphologische Partikel
PNC	Teil eines Personennamens
RC	Relativsatz
RE	sich wiederholendes Element
RS	indirekte Rede
SB	Subjekt
SBP	passiviertes Subjekt
SVP	abgetrennte Vebrpartikel
UC	Einheitskomponente (für flache Anntoationne, z.B. in chunks)
VO	Vokativ

4.6. deprel

Typ: *Layer* – deprel

Name:	deprel
Dokumentation:	Manuell korrigierte Dependenzannotation, die zunächst mit den Mate Tools erstellt wurde (http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html). Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – deprel

Schritt:	1
Aufbereitung:	Dependenzparsing
Aufbereitungsart:	Automatisch
Tool:	Mate Dependenzparser (parser-ger-3.6.model)
Format:	CoNLL 2009
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	Gohar Schnelle

Datum: 2018
Qualitätsprüfung: manuell (nur teilweise)
Editor: LAUDATIO, LangBank

Schritt: 2
Aufbereitung: Korrektur der Mate Parses in zwei Dateien
(HortulusSanitatis_1609_Uffenbach und
ContrafaytKreuterbuch_1532_Brunfels)

Aufbereitungsart: manuell

Tool: TrEd

Format: CoNLL

Segmentierung der
Annotation: basiert auf 'norm'

Qualitätsprüfer: Gohar Schnelle

Datum: 2018

Qualitätsprüfung: manuell

Editor: Gohar Schnelle

Schritt: 3

Aufbereitung: Konvertierung nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013,
CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden
Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter.
Merging mithilfe des Merger-Moduls. Export ins ANNIS- und
PAULA-Format mit dem ANNISExporter und dem PAULAExporter
(https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – deprel

Annotationsart:	Dependenzannotation
Beschreibung:	Dependenzannotation basierend auf dem TIGER-Annotationsschema (https://www.linguistics.ruhr-uni-bochum.de/~dipper/pub/tiger_annot.pdf)

Typ: Annotationvalue – deprel

Wert:	Wertbeschreibung:
AC	adpositioneller Kasusmarker
AG	Genitivattribut
AMS	Maßangabe, die Argument eines Adjektivs ist
AVC	Komponente einer Adverbialphrase
APP	Apposition
CC	komparatives Komplement
CD	koordinierende Konjunktion
CJ	Konjunkt
CM	komparaive Konjunktion
CP	Komplementierer
CVC	Funktionsverbgefüge
DA	Dativ
DM	Diskursmarker
EP	expletives “es”
JU	Junktor
MNR	postnominaler Modifikator
MO	Modifikator
NG	Negation
NK	Element der Kern-NP
OA	Akkusativobjekt
OA2	zweites Akkusativobjekt
OC	Objektsatz
OG	Genitivobjekt
OP	Präpositionalobjekt

PAR	Parenthese
PD	Prädikat
PG	phrasaler Genitiv
PM	morphologische Partikel
PNC	Teil eines Personennamens
RC	Relativsatz
RE	sich wiederholendes Element
RS	indirekte Rede
SB	Subjekt
SBP	passiviertes Subjekt
SP	Subjekt oder Prädikat
SVP	abgetrennte Verbpartikel
UC	Einheitskomponente (für flache Annotationen, z.B. in chunks)
VO	Vokativ
--	NA

4.7. lemma-deprel

Typ: *Layer* – lemma-deprel

Name:	lemma-deprel
Dokumentation:	Lemmatisierung mithilfe der Mate Tools (http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html). Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – lemma-deprel

Schritt:	1
Aufbereitung:	Lemmatisierung
Aufbereitungsart:	Automatisch
Tool:	Mate Tools (lemma-ger-3.6.model)

Format: CoNLL 2009
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: Gohar Schnelle
Datum: 2018
Qualitätsprüfung: manuell, nur in ContrafaytKreuterbuch_1532_Brunfels und HortulusSanitatis_1609_Uffenbach
Editor: LAUDATIO, LangBank

Schritt: 2
Aufbereitung: Konvertierung nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – lemma-deprel

Annotationsart: Spannenannotation je norm-Einheit
Beschreibung: Lemmatisierung.

Typ: Annotationvalue – lemma-deprel

Wert: Wertbeschreibung:

Normalisiertes Lemma Lemmatisierung.

4.8. pos-deprel

Typ: *Layer* – pos-deprel

Name: pos-deprel

Dokumentation: Wortartenannotation mithilfe der Mate Tools. Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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

Typ: Preparationstep – pos-deprel

Schritt: 1

Aufbereitung: POS-Tagging

Aufbereitungsart: Automatisch

Tool: Mate Tools (tag-ger-3.6.model)

Format: CoNLL 2009

Segmentierung der Annotation: basiert auf ‘norm’

Qualitätsprüfer: NA

Datum: 2018

Qualitätsprüfung: manuell, nur in ContrafaytKreuterbuch_1532_Brunfels und HortulusSanitatis_1609_Uffenbach

Editor: LAUDATIO, LangBank

Schritt: 2

Aufbereitung: Konvertierung nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter

https://korpling.german.hu-berlin.de/saltpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – pos-deprel

Annotationsart: Spannenannotation je norm-Einheit
Beschreibung: Wortartenannotation.

Typ: Annotationvalue – pos-deprel

Wert:	Wertbeschreibung:
\$,	Part-of-Speech-Tagging.
\$.	
\$LRB	
ADJA	
ADJD	
ADV	
APPO	
APPR	
APPRART	
APZR	
ART	
CARD	
FM	
ITJ	
KOKOM	
KON	
KOUI	
KOUS	
NE	
NN	
PDAT	

PDS
PIAT
PIS
PPER
PPOSAT
PPOSS
PRELAT
PRELS
PRF
PROAV
PTKA
PTKANT
PTKNEG
PTKVZ
PTKZU
PWAT
PWAV
PWS
TRUNC
VAFIN
VAPP
VMFIN
VMINF
VVFIN
VVIMP
VVINF
VVIZU
VVPP
XY

4.9. morph

Typ: *Layer* – morph

Name:	morph
Dokumentation:	Morphologische Annotation. Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – morph

Schritt: 1
Aufbereitung: Morphologische Annotation
Aufbereitungsart: Automatisch
Tool: Mate Tools (morphology-ger-3.6.model)
Format: CoNLL 2009
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: Gohar Schnelle
Datum: 2018
Qualitätsprüfung: manuell, nur teilweise
Editor: LAUDATIO, LangBank

Schritt: 2
Aufbereitung: Konvertierung nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – morph

Annotationsart: Spannenannotation je norm-Einheit

Beschreibung: Morphologische Annotation mit additiven Werten bestehend aus Kasus|Numerus|Genus|Modus|Tempus|Komparationsstufe|Genus verbi (je nach Wortart).

Typ: Annotationvalue – morph

Wert:	Wertbeschreibung:
sg/pl	Singular/Plural
neut/masc/fem	Neutrum/Maskulinum/Femininum
nom/gen/dat/acc	Nominativ/Genitiv/Dativ/Akkusativ
1/2/3	1./2./3. Person
pres/past	Präsens/Präteritum
pos/comp/sup	Positiv/Komparativ/Superlativ
ind/subj	Indikativ/Subjunktiv
*	unterspezifiziert

4.10. pos-const

Typ: Layer – pos-deprel

Name: pos-const

Dokumentation: Wortartenannotation (<http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html>). Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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

Typ: Preparationstep – pos-const

Schritt: 1

Aufbereitung: POS-Tagging

Aufbereitungsart: Automatisch

Tool: Mate Tools (tag-ger-3.6.model)

Format: CoNLL 2009

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: NA

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Schritt: 2

Aufbereitung: Transfer nach PTB

Aufbereitungsart: Automatisch

Tool: Berkeley Parser 1.7

Format: PTB

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: NA

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Schritt: 3

Aufbereitung: Konvertierung nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – pos-const

Annotationsart: Spannenannotation je norm-Einheit

Beschreibung: Wortartenannotation.

Typ: Annotationvalue – pos-const

Wert:

\$,

\$.

*\$LRB

ADJA

ADV

APPO

APPR

APPRART

APZR

ART

CARD

FM

ITJ

KOKOM

KON

KOUI

KOUS

NE

NN

PDAT

PDS

PIAT

PIS

PPER

PPOSAT

PRELAT

PRELS

PRF

PROAV

PTKA

PTKNEG

Wertbeschreibung:

Part-of-Speech-Tagging.

PTKVZ
PTKZU
PWAT
PWAV
PWS
TRUNC
VAFIN
VAINF
VAPP
VMFIN
VMINF
VVFIN
VVIMP
VVINF
VVIZU
VVPP
XY

4.11. foreign

Typ: *Layer* – foreign

Name: foreign
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der
Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ:*Annotationlayer* – foreign

Annotationsart: Spannenannotation für einzelne dipl-Einheiten
Beschreibung: Markiert einen Text, der in einer Fremdsprache geschrieben ist.

Typ:*Annotationvalue* – foreign

Wert: Wertbeschreibung
foreign Allgemeiner Tag für fremdsprachliches Material.

4.12. foreign_trans

Typ: *Layer* – foreign_trans

Name: foreign_trans
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation:

basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – foreign_trans

Annotationsart: Spannenannotation für einzelne norm-Einheit
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.

4.13. lang

Typ: *Layer* – lang

Name: lang
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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	2018
Qualitätsprüfung:	NA

Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt: 2

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

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – lang

Annotationsart: Annotationswert wird je norm-Einhei zugewiesen

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)

deu: Deutsch (Default-Wert)

lat	Latein
grc	Altgriechisch
spa	Spanisch
ita	Italienisch
ara	Arabisch
fra	Französisch
pol	Polnisch
nld	Niederländisch
eng	Englisch
rus	Russisch
goh	Althochdeutsch
dan	Dänisch

Römische Zahlen und Interpunktionszeichen werden je nach Kontext (meistens “deu” oder “lat”) annotiert. „deu“ wird als default-Wert verwendet, wenn der Annotator/die Annotatorin keine Entscheidung treffen konnte.

4.14. 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: 2017
Qualitätsprüfung: manuelle Korrektur
Editor: Gohar Schnelle

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – Name: sentence_end

Annotationsart: Spannenannotation für die jeweils letzte dipl-Einheit einer "T-Unit" (s. Dokumentation)
Beschreibung: Annotation der Satzenden nach folgenden Guidelines zur frühneuhochdeutschen Satzsegmentierung: http://sfs.uni-tuebingen.de/langbank/src/enhg_sent_seg-german-v4.pdf

Typ: Annotationvalue – Name: sentence_end

Wert: Wertbeschreibung:
S Markiert das Ende eines Satzes.

4.15. position_verb

Typ: *Layer* – position_verb

Name: position_verb
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – Name: position_verb

Annotationsart: Spannenannotation für einzelne norm-Einheiten
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

4.16. subclause_type

Typ: *Layer* – subclause_type

Name:	subclause_type
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – subclause_type

Annotationsart: Spannenannotation für einzelne norm-Einheiten
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.

4.17. KOUS_sem

Typ: *Layer* – KOUS_sem

Name:	KOUS_sem
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – KOUS_sem

Annotationsart: Spannenannotation für einzelne norm-Einheiten
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

4.18. comp

Typ: Layer – komp

Name:	comp
Dokumentation:	Siehe Dokumentation von prot . Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markiert alle im Korpus enthaltenen zusammen, getrennt oder mit Bindestrich geschriebenen Komposita mit einem substantivischen Zweitglied.

Typ: Annotationvalue – comp

Wert: Wertbeschreibung:
k Kompositum mit substantivischem Zweitglied

4.19. 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. Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-

berlin.de/saltpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp_orth

Annotationsart: Spannenannotation über mehrere dipl-Einheiten

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)

4.20. 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. *ftabwurz*).

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. Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – prot

Annotationsart:	Spannenannotation über mehrere dipl-Einheiten
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

4.21. comp_n

Typ: *Layer* – comp_n

Name:	comp_n
Dokumentation:	Morphologische Annotation von Komposita mit nominalem Zweitglied. Hierin können auch getrenntgeschriebene Komposita vorkommen, die unter bestimmten Umständen auch als Konstruktionen mit pränominalem Genitiv interpretiert werden können. Da jedoch konkrete Anhaltspunkte vorliegen, diese als Komposita zu interpretieren (z.B. Modifikatoren, Flexions-/ Fugenelemente), wurden sie in dieser Ebene annotiert. Fälle, die weder als Kompositum noch als Syntagma eingeordnet werden konnten, wurden in comp_amb annotiert. Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comp_n

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	Laura Perlitz
Datum	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp_n

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markiert alle im Korpus enthaltenen zusammen, getrennt oder mit Bindestrich geschriebenen Komposita mit einem substantivischen Zweitglied.

Typ: Annotationvalue – comp_n

Wert: Wertbeschreibung:
Additive Werte zur Beschreibung der morphologischen Struktur der nominalen Komposita, z.B. N_N für Komposita, die aus zwei Nomina bestehen. Suffixe wurden nur berücksichtigt, wenn Suffigierung der letzte Wortbildungsprozess war ([[*Kindbett*]erin]). Zwischen den Kürzeln für die Kategorien der einzelnen morphologischen Elemente stehen Unterstriche. Ist die Kategorie eines Elements nicht ermittelbar, wird der Platzhalter „X“

	verwendet.
N	Nomen, z.B. <i>Ampfer</i> in <i>Ampffer kraut</i>
A	Adjektiv, z.B. <i>weiß</i> in <i>Weißs kol</i>
V	Verb, z.B. <i>destillieren</i> in <i>distillier Gläsern</i>
ADV	Adverb, z.B. <i>aber</i> in <i>Aberglauben</i>
APPR	Präposition, z.B. <i>bei</i> in <i>beynamen</i>
CARD	Kardinalzahl, z.B. <i>drei</i> in <i>treyfuß</i>
SUFF	Suffixe (bei mehreren hintereinander wird nur eins gezählt, z.B. <i>er-in</i> in <i>kindbetterin</i>)
CONV	Konversion, z.B. <i>bei das hertz zittern</i>
X	Platzhalter für intransparente Wortteile, z.B. <i>schel</i> in <i>Schelwurtz</i>

4.22. comp_n_graph

Typ: *Layer* – comp_n_graph

Name:	comp_n_graph
Dokumentation:	Beschreibung der Graphematik der Komposita mit nominalem Zweitglied. 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. Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comp_n_graph

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: Laura Perlitz
Datum: 2018
Qualitätsprüfung: manuelle Korrektur
Editor: Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp_n_graph

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Beschreibt die Graphematik der in der Ebene „comp_n“

annotierten Komposita.

Typ: Annotationvalue – comp_n_graph

Wert:	Wertbeschreibung:
nospace	zusammen geschrieben
sep	getrennt geschrieben
hyph	Bindestrichmarkierung (innerhalb einer Zeile)
lb1	unterbrochen durch Zeilenumbruch (ohne Bindestrichmarkierung)
lb2	unterbrochen durch Zeilenumbruch (mit Bindestrichmarkierung)
camel	Binnengroßschreibung

Komposita, die aus mehr als zwei lexikalischen Teilen entstehen, erhalten einen additiven Wert (verbunden durch Unterstriche), z.B. „nospace_sep“ bei *Saurampffer wasser*.

4.23. comp_n_mod

Typ: Layer – comp_n_mod

Name:	comp_n_mod
Dokumentation:	Gibt an, ob die in der Ebene <comp_n> annotierten potentiellen Komposita einen oder mehrere(flektierende) Modifikatoren erhalten haben und welche(n).
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comp_n_mod

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

Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: Laura Perlitz
Datum: 2018
Qualitätsprüfung: manuelle Korrektur
Editor: Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp_n_mod

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Gibt an, ob die in der Ebene <comp_n> annotierten potentiellen Komposita einen oder mehrere (flektierende) Modifikatoren erhalten haben und welche(n).

Typ: Annotationvalue – comp_n_mod

Wert:	Wertbeschreibung:
	Dem jeweiligen (potentiellen) Kompositum wurde ein Wert aus dem STTS (in Kleinbuchstaben) zugewiesen, der die Wortart des Modifikators beschreibt. Bei mehreren Modifikatoren wurden additive Werte vergeben, die durch einen Unterstrich separiert sind, z.B. 'art_adja'.
art	bestimmter oder unbestimmter Artikel
apprart	Präposition mit Artikel
adja	attributives Adjektiv
piat	attribuierendes Indefinitpronomen ohne Determiner (Determinierer wurden separat durch "art" oder "apprart" erfasst)
pposat	attribuierendes Possessivpronomen
pdat	attribuierendes Demonstrativpronomen
prelat	attribuierendes Relativpronomen
card	Kardinalzahl
NA	keine Angabe möglich
n	kein Modifikator vorhanden

4.24. comp_a

Typ: *Layer* – comp_a

Name:	comp_a
Dokumentation:	Beschreibung der Graphematik der Komposita mit nominalem Zweitglied. 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. Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comp_a

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	Laura Perlitz
Datum	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – comp_a

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Beschreibt die Orthographie der in der Ebene „komp“ annotierten Komposita.

Typ: Annotationvalue – comp_a

Wert: Wertbeschreibung:
Additive Werte zur Beschreibung der morphologischen Struktur der nominalen Komposita, z.B. A_A für Komposita, die aus zwei Adjektiven bestehen. Suffixe wurden nur berücksichtigt, wenn Suffigierung der letzte Wortbildungsprozess war. Zwischen den Kürzeln für die Kategorien der einzelnen morphologischen Elemente stehen Unterstriche.

N Nomen, z.B. *Asche* in *afchenfarben*
A Adjektiv, z.B. *wohl* in *wohrliechend*
ADV Adverb, z.B. *aber* in *Aberglauben*
CARD Kardinalzahl, z.B. *vier* in *viereckicht*
SUFF Suffixe (bei mehreren hintereinander wird nur eins gezählt, z.B. *-isch* in *Waldenburgifchen*
CONV Konversion, z.B. bei *die Hochgelehrten*
farb Das letzte Glied enthält die Wurzel „farb“, z.B. *braunfarb*.

4.25. comp_a_graph

Typ: Layer – comp_a_graph

Name: comp_a_graph
Dokumentation: Beschreibung der Graphematik der Komposita mit nominalem Zweitglied. 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.
Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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

Typ: Preparationstep – comp_a_graph

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	Laura Perlitz
Datum	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling

Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp_a_graph

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Beschreibt die Orthographie der in der Ebene „comp_a“
annotierten Komposita.

Typ: Annotationvalue – comp_a_graph

Wert:	Wertbeschreibung:
nospace	zusammen geschrieben
sep	getrennt geschrieben
hyph	Bindestrichmarkierung (innerhalb einer Zeile)
lb1	unterbrochen durch Zeilenumbruch (ohne Bindestrichmarkierung)
lb2	unterbrochen durch Zeilenumbruch (mit Bindestrichmarkierung)

4.26. comp_amb

Typ: Layer – comp_amb

Name: comp_amb
Dokumentation: Annotation von Wortfolgen, die sowohl als Komposita als auch
als Syntagmen interpretiert werden können und bei denen der
Kontext oder die Morphologie keine Hinweise für die
Einordnung in die eine oder andere Kategorie liefern.
Informationen zum gesamten Konvertierungsworkflow gibt es
hier: <https://github.com/langbank/conversion-pipeline>.

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

Typ: Preparationstep – comp_amb

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf ‘dipl’
Qualitätsprüfer: Laura Perlitz
Datum 2018
Qualitätsprüfung: manuelle Korrektur
Editor: Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf ‘dipl’
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp_amb

Annotationsart: Spannenannotation über mehrere dipl-Einheiten

Beschreibung: Annotation von Wortfolgen, die möglicherweise Komposita sein könnten.

Typ: Annotationvalue – comp_amb

Wert: Wertbeschreibung:

a Adjektiv-Nomen-Abfolgen, bei denen das Adjektiv unflektiert ist und bei denen man im modernen Deutschen eher ein Kompositum verwenden würde (z.B. das edel gefeine). Es gibt noch weitere AN-Abfolgen, die ebenfalls ein unflektiertes Adjektiv enthalten, die aber im heutigen Deutschen auch genauso (oder noch eher) als attributive Adjektive verwendet werden können. Diese wurden jedoch in ADJA_uninfl annotiert (z.B.).

gpre Nomen-Nomen-Abfolgen, die sowohl als Komposita als auch als Nomina mit pränominalem Genitivattribut interpretiert werden können und bei denen der Kontext oder die Flexion keinen Aufschluss auf den Status geben können.

4.27. infl_fuge

Typ: Layer – infl_fuge

Name: infl_fuge

Dokumentation: Annotation der (Nicht-) Existenz von Fugenelementen, bzw. Flexionselementen zwischen zwei oder mehr lexikalischen Einheiten, die potentielle Komposita darstellen (v.a. N-N-Abfolgen). Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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

Material zusammengefasst.

Typ: Preparationstep – infl_fuge

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	Laura Perlitz
Datum	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – infl_fuge

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Annotation von Flexion bzw. Fugenelementen bei Wortfolgen, die möglicherweise Komposita sein könnten.

Typ: Annotationvalue – infl_fuge

Wert:	Wertbeschreibung:
y	Es ist ein Fugenelement, bzw. ein Flexionselement vorhanden (,yes‘).
n	Es ist kein Fugenelement, bzw. Flexionselement vorhanden (,no‘).
yn	Gilt für potentielle Komposita mit mehr als zwei lexikalischen Wortteilen, die sowohl mit als auch ohne Fugenelement oder Flexionselement realisiert wurden (z.B. <i>Jungkfrauen har</i>)
NA	Es kann nicht beurteilt werden, ob ein Fugenelement, bzw. Flexionselement vorhanden ist (meistens weil das betroffene Wort/lexikalische Material nicht mehr transparent ist oder aus einer anderen Sprache (Latein) stammt).

4.28. comp_lex

Typ: Layer – comp_lex

Name: comp_lex
Dokumentation: Kategorisierung der (potentiellen) Komposita in lexikalisierte und nicht-lexikalisierte Komposita nach bestimmten Kriterien (s. unten). Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – comp_lex

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: Laura Perlitz
Datum: 2018
Qualitätsprüfung: manuelle Korrektur
Editor: Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comp_lex

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
 Beschreibung: Annotation der Lexikalisierung von Wortfolgen, die möglicherweise Komposita sein könnten.
Typ: Annotationvalue – comp_lex

Wert: Wertbeschreibung:
 lex lexikalisierte Komposita, die nicht (mehr) als Syntagma ausgedrückt werden können, weil ihre addierten Teilbedeutungen stark von der Kompositionsbedeutung abweichen (betroffene Themenbereiche: Pflanzennamen, geographische Namen, bestimmte Krankheiten, Körperteile, Tierarten, Sternzeichen), aber nicht wenn sie mit einem zusätzlichen Glied versehen sind, das nicht zu dem lexikalisierten Wort gehört, z.B. *Eisenkrautsaft* oder *Beifußblumen* (vs. *Johannisblumen*), *Blutwassersucht*. Verbindungen mit "Baum" werden wie Verbindungen mit "Kraut" behandelt - meistens sind sie ein zusätzliches Glied zur eigentlichen Pflanzenbezeichnung - *Kirsche* -> *Kirschbaum*, *Eiche* -> *Eichenbaum*; manchmal gehören sie aber zur lexikalisierten Pflanzenbezeichnung (z.B. *Schildkraut*, *Rutelkraut*, *Wunderbaum*)
 n nicht-lexikalisierte (potentielle) Komposita, die nicht den oben genannten Kriterien entsprechen

Einzelentscheidungen

Vorkommen	lex	n
Jungfrau	x	
Liebhaber	x	
Wohlfahrt	x	
Mehrteil		x
Vaterland	x	
Bauchfluss		x
Aberglaube	x	
Aderlass	x	
Brachmonat	x	
Brosam	x	
Erdapfel	x	
Gutdünken	x	
Maibutter		x
Notdurft	x	
Freiherr	x	
Gegenteil	x	
Nachteil	x	
Ohnmacht	x	

Gliedmaßen	x
Nachgeburt	x
Widerwille	x

4.29. adja_uninfl

Typ: *Layer* – adja_uninfl

Name:	adja_uninfl
Dokumentation:	Annotation von unflektierten Adjektiven, die direkt vor einem Nomen stehen. Stehen vor einem Nomen mehrere unflektierte Adjektive, wird nur dasjenige annotiert, das unmittelbar vor dem Nomen steht. Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – adja_uninfl

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	Laura Perlitz
Datum	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und

	ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – adja_uninfl

Annotationsart:	Spannenannotation für einzelne norm-Einheiten
Beschreibung:	Annotation von Wortfolgen, die möglicherweise Komposita sein könnten.

Typ: Annotationvalue – adja_uninfl

Wert:	Wertbeschreibung:
uninfl	unflektierte Adjektive, die direkt vor einem Nomen stehen; stehen vor einem Nomen mehrere unflektierte Adjektive, wird nur dasjenige annotiert, das unmittelbar vor dem Nomen steht

4.30. 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. <i>Beobachtungen der nach der Bestäubung in den vegetabilischen Eyern erfolgenden Veränderungen</i>). Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der	

Annotation: basiert auf ‘dipl’
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – attr_gen

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markiert alle im Korpus enthaltenen Nominalphrasen, die ein Genitivattribut enthalten.

Typ: Annotationvalue – attr_gen

Wert:	Wertbeschreibung:
gpre	pränominales Genitivattribut
gpost	postnominales Genitivattribut

4.31. morph_ellipsis

Typ: Layer – morph_ellipsis

Name: morph_ellipsis
Dokumentation: Annotation von Wortteillellipsen, z.B. *Anis- und Eppichsamen*. Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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: 2018
Qualitätsprüfung: manuelle Korrektur
Editor: Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – morph_ellipsis

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markiert alle im Korpus enthaltenen koordinierten Komposita („strukturelle Destruktionen“; z.B.: *gelb= und*

Waffenfucht).

Typ: Annotationvalue – morph_ellipsis

Wert:	Wertbeschreibung:
morph_ellipsis	Wortteilellipse (morphologische Ellipse)

4.32. comment_lex

Typ: Layer – comment_lex

Name:	comment_lex
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: NA
basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comment_lex

Annotationsart: Spannenannotation über einzelne norm-Einheiten

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.

4.33. comment

Typ: Layer – comment

Name: comment

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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: 2018

Qualitätsprüfung: NA

Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling

Schritt: 2

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

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – comment

Annotationsart: Spannenannotation über einzelne oder mehrere dipl-Einheiten

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 (*dicke* für *oft*) werden erläutert.

Typ: Annotationvalue – comment

Wert: Wertbeschreibung:
Zeichenkette Erläuterungen zur Lexik.

4.34. reference

Typ: *Layer* – reference

Name: reference

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	NA basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – reference

Annotationsart:	Spannenannotation über einzelne oder mehrere dipl-Einheiten
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.

4.35. form_disease

Typ: *Layer* – form_disease

Name:	form_disease
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

. **Typ:** Annotationlayer – form_ disease

Annotationsart: Spannenannotation über einzelne oder mehrere dipl-Einheiten

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

4.36. problem

Typ: *Layer* – werkname

Name:	problem
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
 Format: PAULA-XML, ANNIS
 Segmentierung der Annotation: basiert auf 'dipl'
 Qualitätsprüfer: korpling
 Datum: 2018
 Qualitätsprüfung: NA
 Editor: LAUDATIO, LangBank

. **Typ:** Annotationlayer – problem

Annotationsart: Spannenannotation über einzelne oder mehrere dipl-Einheiten
 Beschreibung: NA

Typ: Annotationvalue – problem

Wert: Wertbeschreibung:
 Zeichenkette NA

4.37. herbname_norm

Typ: Layer – herbname_norm

Name: herbname_norm
 Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
 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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – herbname_norm

Annotationsart: Spannenannotation über einzelne oder mehrere dipl-Einheiten

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.

4.38. herbprep

Typ: Layer – herbprep

Name: herbprep

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – herbprep

Annotationsart:	Spannenannotation über einzelne oder mehrere dipl-Einheiten
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>saft_des_wermuts</i> <i>des roten beifusz bletter</i> → <i>des_roten_beifuß_blätter</i> <i>safft von weremut</i> → <i>saft_von_wermut</i> <i>weremutsafft</i> → <i>wermutsaft</i> <i>weremut safft</i> → <i>wermutsaft</i>

4.39. form_prep

Typ: Layer – form_prep

Name:	form_prep
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

. **Typ:** Annotationlayer – form_prep

Annotationsart: Spannenannotation über einzelne oder mehrere dipl-Einheiten

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</i>

4.40. noun_nom

Typ: *Layer* – noun_nom

Name:	noun_nom
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-

[berlin.de/saltpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip](https://www.berlin.de/saltpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip))

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – noun_nom

Annotationsart: Spannenannotation über einzelne oder mehrere dipl-Einheiten
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.

4.41. form_noun

Typ: Layer – form_noun

Name: form_noun
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf ‘dipl’
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

. **Typ:** Annotationlayer – form_noun

Annotationsart:	Spannenannotation über einzelne oder mehrere dipl-Einheiten
Beschreibung:	In dieser Ebene werden alle Nomina morphologisch annotiert.

. **Typ:** Annotationvalue – form_noun

Wert:	Wertbeschreibung:
simplex	Mit <i>simplex</i> 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 <i>capitel</i> oder <i>quintin</i>) 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.

4.42. clause_type

Typ: *Layer* – clause_type

Name:	clause_type
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – clause_type

Annotationsart: Spannenannotation über mehrere dipl-Einheiten

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	<p>Klare Relativsätze, dabei sowohl w-Relativsätze als auch d-Relativsätze, z.B.</p> <p><i>im sch^z adet auch nít ob aín thier das nít sinnig wár ín beiffet</i></p> <p><i>Wer dē beyfusz bei ún tregt wen er wandert der wirt nít müde</i></p>
padv	<p>Sätze, die durch ein Präpositionaladverb eingeleitet werden, z.B.</p> <p><i>das stabwurz genüczet verz zeret vberflüßig feuchtikaít die ín den dármen find dauo aín kranck haít komet genant colíca passíó</i></p> <p>Hier ist oft unsicher, ob das überhaupt Relativsätze sind.</p>
rsx	<p>Relativsätze ohne Hauptsatz. Das kommt häufig in Überschriften vor, z.B.</p> <p><i>Wer eynen kurtzen atthem hat</i></p>
rsdem	<p>Ambige Fälle: Relativsatz oder Demonstrativsatz, z.B.</p> <p><i>vnd nimpt traurigkeit die kompt von Melancoley</i></p>
padvpart	<p>Sätze mit Präpositionaladverb und Partizip, z.B.</p> <p><i>dauon genomen auf aín halb lot</i></p> <p>Diese Sätze sind nicht ambig wie rsdem, sondern enthalten beides! Ob sie überhaupt Relativsätze sind, ist unklar.</p>
dem	<p>Demonstrativsätze. Hier sind alle Sätze gemeint, deren Subjekt ein Demonstrativpronomen ist, z.B.</p> <p><i>der hilfft on zweifel</i></p>
part	<p>Partizipien, die sich ähnlich verhalten wie Relativsätze, z.B.</p> <p><i>Jtem stabwurz mít oleí vn salcz zefamē gestoffen vnd auff den pulz hende vnd füßz gelegtift gút wider das fieber</i></p> <p>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.</p>

4.43. position_rel

Typ: Layer – position_rel

Name: position_rel

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – position_rel

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
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 nít hat die sol den beyfüsz fieden mit wein vn sol den trinckē der hilfft on zweifel</i> Wir annotieren auch ‚vor‘, wenn es ein Einleitungswort wie <i>item</i> 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 stáin habē in den lenden</i>
int	Eingebetteter Relativsatz, z.B. <i>Jtem stabwurz macht auch wachfzen den bart á langsam herfür komet domít bez strichē</i>

4.44. position_referent

Typ: Layer – position_referent

Name: position_referent
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – position_referent

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
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 kainen schaden zufügen</i>
adja-n	Adjazent nachgestellt, z.B. <i>im sch^z adet auch nit ob ain thier das nit sinnig wårin 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 hūften vn rainiget die brufft dē fy erkaltet ift</i>

4.45. form_referent

Typ: Layer – form_referent

Name:	form_referent
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – form_ referent

Annotationsart:	Spannenannotation über mehrere dipl-Einheiten
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 blasen wachzt</i>
d-pron	<i>der, die, das, dieser</i> etc., z.B.

*Wer dē beyfuß bei im tregt wen er wandert **der** wirt nit müde*

p-pron

Personalpronomina, z.B.

*Jtem wer € beyfuß wurczel an feinem halz tregt kain vergifftig tier mag **im** nit geschaden*

null

Bei freien und asyndetischen Relativsätzen, die kein overttes Korrelat im Hauptsatz haben, z.B.

vn was nach dem getranck vberbeleibt vo kreütern sol man warm legen auf die milcz

4.46. position_verb_rel

Typ: Layer – position_verb_rel

Name: position_verb_rel

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – position_verb_rel

Annotationsart:	Spannenannotation über mehrere dipl-Einheiten
Beschreibung:	Verbstellung im Relativsatz

Typ: Annotationvalue – position_verb_rel

Wert:	Wertbeschreibung:
v2	Verb-Zweit-Stellung, z.B. <i>Vn frawen die kinder find seügen solz 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 bekomme 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.

4.47. form_relpron

Typ: *Layer* – form_relpron

Name: form_relpron
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (<https://korpling.german.hu->

berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – form_relpron

Annotationsart: Spannenannotation über einzelne dipl-Einheiten
Beschreibung: Form der reltivsatzeinleitenden Kategorie

Typ: Annotationvalue – form_relpron

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

4.48. mod_referent

Typ: Layer – mod_referent

Name: mod_referent

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – mod_referent

Annotationsart: Spannenannotation über eine oder mehrere dipl-Einheiten

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

4.49. ppk_e1

Typ: Layer – ppk_e1

Name: ppk_e1

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

. **Typ:** Annotationlayer - ppk_e1

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
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.

4.50. ppk_e2

Typ: *Layer* – ppk_e1

Name:	ppk_e2
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

. **Typ:** Annotationlayer - ppk_e2

Annotationsart: Spannenannotation über mehrere dipl-Einheiten

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.

4.51. ppk_e3

Typ: Layer – ppk_e1

Name: ppk_e3
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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 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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf ‘dipl’
Qualitätsprüfer: korpling
Datum: 2018

Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

. **Typ:** Annotationlayer - ppk_e3

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
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: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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

Die lb-Ebene wurde bis zur Version 6 folgendermaßen erstellt (für die Aufbereitungsschritte in den neuen Dokumenten der Version 7 siehe weiter unten:

Schritt:	1a
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling
Schritt:	2a
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	LAUDATIO, LangBank

In den neuen Dokumenten der Version 7 wurde die lb-Ebene folgendermaßen erstellt:

Schritt:	1b
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	automatisch
Tool	ocrtagger.pl

Format: Text-Datei
Segmentierung der Annotation: basiert auf 'ocr'
Qualitätsprüfer: Uwe Springmann
Datum: 2017
Qualitätsprüfung: NA
Editor: Uwe Springmann

Schritt: 2b
Aufbereitung: Überführung nach Excel 2013
Aufbereitungsart: semi-automatisch
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'ocr'
Qualitätsprüfer: Uwe Springmann
Datum: 2017
Qualitätsprüfung: NA
Editor: Uwe Springmann

Schritt: 3b
Aufbereitung: Korrektur und Anpassung der Segmentierung an die diplomatische Transkription in dipl
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2017
Qualitätsprüfung: manuelle Konsistenzprüfung
Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt: 4b
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch
 Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
 Format: PAULA-XML, ANNIS
 Segmentierung der Annotation: basiert auf 'dipl'
 Qualitätsprüfer: korpling
 Datum: 2018
 Qualitätsprüfung: manuelle Korrektur
 Editor: LAUDATIO, LangBank

Typ: Annotationlayer – lb

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
 Beschreibung: Markierung einer ganzen Zeile.

Typ: Annotationvalue – lb

Wert: Wertbeschreibung:
 lb Markierung pro ganzer Zeile.

5.2. column

Typ: *Layer* – column

Name: column
 Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
 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 – column

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2017
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2018

Qualitätsprüfung: manuelle Korrektur
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – column

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markierung einer ganzen Spalte

Typ: Annotationvalue – lb

Wert:	Wertbeschreibung:
l	dipl-Einheiten, die zusammen in der linken Spalte stehen
r	dipl-Einheiten, die zusammen in der rechten Spalte stehen
c	bei dreispaltigem Text: dipl-Einheiten, die zusammen in der mittleren Spalte stehen (center)

5.3. pb

Typ: *Layer* – pb

Name: pb

Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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

Die pb-Ebene wurde bis zur Version 6 folgendermaßen erstellt (für die Aufbereitungsschritte in den neuen Dokumenten der Version 7 siehe weiter unten:

Schritt:	1a
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling
Schritt:	2a
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	LAUDATIO, LangBank

In den neuen Dokumenten der Version 7 wurde die pb-Ebene folgendermaßen erstellt:

Schritt:	1b
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	automatisch
Tool	ocrtagger.pl

Format: Text-Datei
Segmentierung der Annotation: basiert auf 'ocr'
Qualitätsprüfer: Uwe Springmann
Datum: 2017
Qualitätsprüfung: NA
Editor: Uwe Springmann

Schritt: 2b
Aufbereitung: Überführung nach Excel 2013
Aufbereitungsart: semi-automatisch
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'ocr'
Qualitätsprüfer: Uwe Springmann
Datum: 2017
Qualitätsprüfung: NA
Editor: Uwe Springmann

Schritt: 3b
Aufbereitung: Korrektur und Anpassung der Segmentierung an die diplomatische Transkription in dipl
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2017
Qualitätsprüfung: manuelle Konsistenzprüfung
Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt: 4b
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: manuelle Korrektur
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – pb

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markierung einer ganzen Seite.

Typ: Annotationvalue – pb

Wert: Wertbeschreibung:
pb Markierung pro ganzer Seite.

5.4. pb_n

Typ: Layer – pb_n

Name: pb_n
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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: 2018

Qualitätsprüfung: manuelle Korrektur

Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling

Schritt: 2

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

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf ‘dipl’

Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – pb_n

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markierung der Nummerierung einer Seite.

Typ: Annotationvalue – pb_n

Wert: Wertbeschreibung:
numerischer Wert Seitenzahl als lateinische oder römische Zahlenangabe (wird aus dem Faksimilé übernommen); ist keine Seitenangabe vorhanden, wird auch keine eingetragen

5.5. pb_ana

Typ: *Layer* – pb_ana

Name: pb_ana
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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: 2018
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – pb_ana

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Interpretation von vermutlich nicht korrekten Seitenzahlangaben.

Typ: Annotationvalue – pb_ana

Wert: Wertbeschreibung:
Int Entsprechende Angabe der vermutlich korrekten Seitenzahl.

5.6. unclear

Typ: *Layer* - unclear

Name: unclear
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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: 2018
Qualitätsprüfung: Manuelle Konsistenzprüfung. Der Layer reason, der eine Begründung für unleserliche Elemente beschreiben sollte, wurde entfernt.

Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: *Annotationlayer* – unclear

Annotationsart:	Annotationswerte werden je dipl-Einheit zugewiesen
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 dipl-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 unclear-Ebene dies allgemein angezeigt. In die Ebenen atLeast und atMost

wird angegeben, wieviele Zeichen „von“ „bis“ fehlen bzw. nicht gelesen werden können.

5.7. atLeast

Typ: *Layer* – atLeast

Name:	atLeast
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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:	2014
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ:*Annotationlayer* – atLeast

Annotationsart:	Spannenannotation für einzelne dipl-Einheiten
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.

5.8. atMost

Typ: *Layer* – atMost

Name: atMost
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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: 2014
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – atMost

Annotationsart: Spannenannotation für einzelne dipl-Einheiten

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.

5.9. interpretation

Typ: *Layer*– interpretation

Name:	interpretation
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: *Annotationlayer*– interpretation

Annotationsart: Spannenannotation für einzelne dipl-Einheiten
Beschreibung: Interpretationsvorschläge für unleserliche

Zeichen/Zeichenketten, die in der Spalte **unclear** als „unclear“ getagged sind.

Typ: *Annotationvalue*– interpretation

Wert: Wertbeschreibung:
Text Angabe der Interpretation.

5.10. figure

Typ: *Layer* – figure

Name: figure
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf ‘dipl’
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA

Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling

Schritt: 2

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

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: Annotationlayer – figure

Annotationsart: Spannenannotation über einzelne oder mehrere dipl-Einheiten

Beschreibung: Bisher wurden Abbildungen als Spanne über diejenigen dipl-Einheiten annotiert, mit denen sie auf einer Höhe sind, bzw. mit denen sie auf einer Seite stehen. Für die Dokumente der Version 7 wurde eine präzisere Annotation vorgenommen:

Markiert eine Graphik, die im Text eingebettet ist. Grundsätzlich wird unterschieden zwischen Abbildungen mit und ohne Bildbeschriftung.

Besitzt eine Abbildung eine Beschriftung, werden alle dipl-Einheiten der Bildbeschriftung in einer Spanne als ‚figure‘ annotiert. Beispiele finden sich in der Tabelle „Abbildungen

mit Beschriftung.

Besitzt eine Abbildung keine Beschriftung, werden Teile des Fließtextes als „Anker“ zur Annotation in der figure-Ebene verwendet. Diese sind in der Tabelle „Abbildungen ohne Beschriftung“ definiert.

Abbildungen mit Beschriftung

Beschreibung	Annotation
Bild mit Bildbeschriftung auf einer Extraseite	Die Bildbeschriftung wird wie der restliche Text auch transkribiert und bekommt eine eigene Seitenzahl zugewiesen. In ‚note‘ wird sie als <i>caption</i> annotiert und in ‚figure‘ als <i>figure</i> . In <i>pb_n</i> und in ‚figure_p‘ wird die entsprechende Seitenzahl angegeben. Sollte diese Extraseite einen Fließtext unterbrechen, d.h. nicht zwischen abgegrenzten Absätzen stehen, dann wird die Bildbeschriftung vor den entsprechenden Absatz gestellt und in <i>pb_n</i> ignoriert. In <i>figure_p</i> wird die Originalseitenzahl eingetragen. Bei dem hier aufgeführten Beispiel kann ein zweisepaltiger Text angenommen werden (<i>Wal-wurtz</i> bekäme also in ‚column‘ den Wert <i>l</i> und <i>männ-lin</i> den Wert <i>r</i>). Auch Zeilenumbrüche werden annotiert.

Screenshot



Zwei Bilder mit Bildbeschriftung auf einer Extraseite

Die Bildbeschriftung wird wie der restliche Text auch transkribiert und bekommt eine eigene Seitenzahl zugewiesen. In ‚note‘ wird jede Beschriftung für sich als *caption* annotiert und in ‚figure‘ als *figure*. Es gibt also zwei *caption*- und zwei *figure*-Annotationen. In ‚figure_p‘ wird jeweils die entsprechende Seitenzahl angegeben. Bei dem hier aufgeführten Beispiel kann ein zweisepaltiger Text angenommen werden (*Weiß Hornungsblüm* bekäme also in ‚column‘ den Wert *l* und *Geel Hornungsblüm* den Wert *r*).



Mehr als zwei Bilder mit Bildbeschriftung auf einer Extraseite

Bei mehreren Bildern wird von links nach rechts und von oben nach unten vorgegangen (links nach rechts geht vor). In diesem Beispiel werden die Figuren also in der folgenden Reihenfolge transkribiert und annotiert:

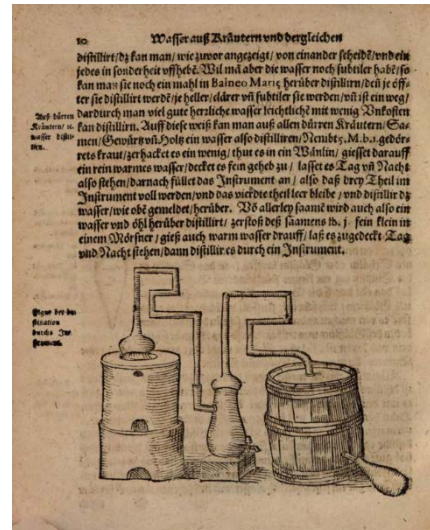
- oben links
- oben rechts
- unten links
- unten rechts

werden jeweils in figure als „figure“ annotiert und erhalten in pb_n eine gemeinsame und in figure_p jeweils eine Seitenzahl zugewiesen



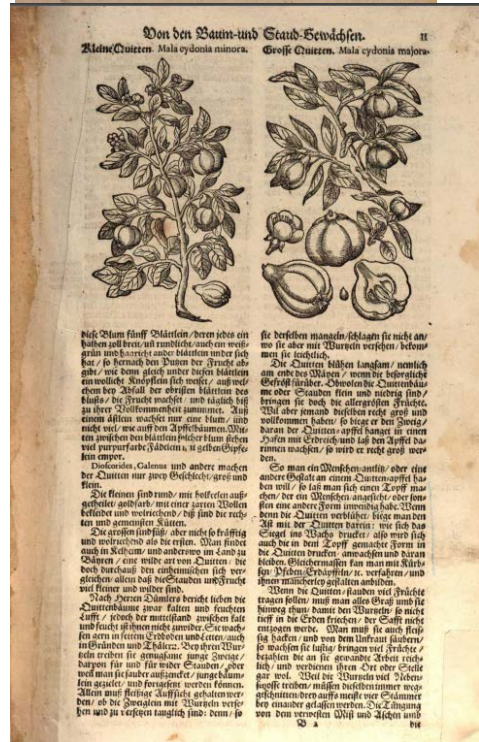
Bild mit Marginalie als Bildbeschriftung

Bezieht sich eine Marginalie offensichtlich auf eine Abbildung, die unter einem abgeschlossenen Absatz steht, wird diese Marginalie im Anschluss an den Absatz transkribiert und in ‚note‘ als *margin* sowie in ‚figure‘ als *figure* annotiert.



Zwei Bilder mit Beschriftung zwischen Fließtext

Zuerst wird die Beschriftung des linken Bildes (oder des oberen, wenn übereinander), dann die des rechten Bildes vor dem entsprechenden Absatz transkribiert. Die Transkriptionsabschnitte werden in ‚note‘ jeweils als *caption* und in ‚figure‘ als *figure* annotiert. In ‚figure_p‘ wird die originale Seitenzahl angegeben. Bei dem konkreten Beispiel kann man außerdem eine Annotation von Spalten vornehmen.



Abbildungen ohne Beschriftung

Beschreibung Annotation

Screenshot

Bild(er) allein
auf Seite

Die letzte dipl-Einheit der
vorangehenden sowie die erste
dipl-Einheit der folgenden Seite
werden als eine Figur annotiert.
In ‚figure_p‘ wird die Seitenzahl
angegeben.
Sind auf einer Seite mehrere
Abbildungen ohne Text werden
diese als eine Figur annotiert.



Bilder neben
einspaltigem
Text

Annotation von der ersten bis
zur letzten dipl-Einheit des
danebenstehenden Textes (wie
bisher auch).

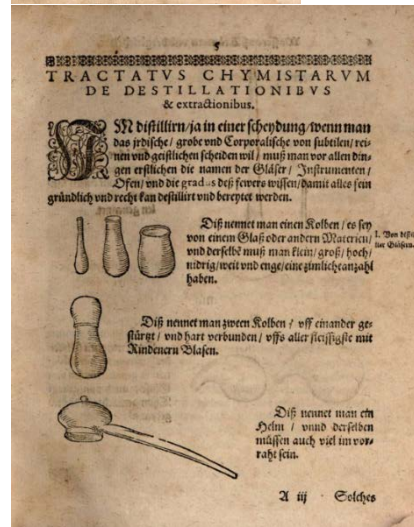


Bild über Text

Das Bild wird in ‚figure‘ für die
erste dipl-Einheit des folgenden
Textabschnittes annotiert (hier
Hafzelwurtz).

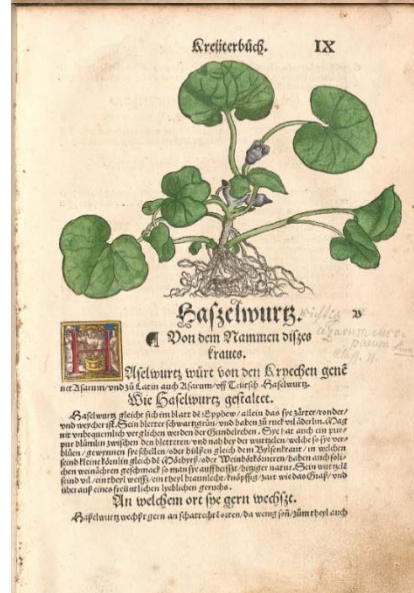


Bild unter
Text

Das Bild wird in ‚figure‘ für die
letzte dipl-Einheit des
vorangehenden Textabschnittes
annotiert.

Bild zwischen Text
Das Bild wird in ‚figure‘ für die letzte dipl-Einheit des vorangehenden Textabschnittes und die erste dipl-Einheit des folgenden Textabschnittes annotiert (hier also für den Punkt hinter *Kuchenschell* und *Kuchenschell* im unteren Textabschnitt).



Bild in
zweispaltigem
Text
oben in einer Spalte: Annotation
an der ersten nachfolgenden
dipl-Einheit
unten in einer Spalte:
Annotation an der letzten
vorangehenden dipl-Einheit
innerhalb einer Spalte:
Annotation an der letzten
vorangehenden und ersten
folgenden dipl-Einheit



die feuchtblaten darauff gelegt
 so auf feuchtblat den eck darnece
 so sol man dz mit ainet schreiffen
 ten blätend machē vñ darauff dz
 balza legen. Item beysuf ist se
 bestoffung der gelde außsin als
 mley vñ leuen die do von ainet
 falkē moterē komet also geniet
 Nym beysuf vñ hiefigen yeg
 liche ein hand vol vñ wenig wa
 müt in wein gefotē vñ mit zuck
 sif gemachet. Diste teurec ist
 auch gut vñ die gelstet so mā
 doep ist vñ wenig vñ taufet
 guldin genēt teurec vñ was
 nach dem getranck überbelet vñ
 teurec sol man warm legen auf
 die mley. Das selb ist auch gut so
 die mley außwendig gefetwelle
 ist. Item beysuf vñ walle genēt
 tarus barbatu gefoten mit wein
 vñ also warm darauff gefet
 ist gut dem der arsdarm ansetzt.



Stabwurtz
Das ander Capitel
 Botunum. Das wec ist
 in grecum vñ in latinis
 also genant. Hefam ist es
 genant in arabischen. Dae mai
 ste Escaplo in dem buch aggre
 gators in dē capitel holum id est
 Abrotanum spricht. Das dif teant
 sey zwoyet hand/ains ist feulich
 das and mällich vñ sind beide
 ainander gleich an der gestalt vñ
 an der tagend. allain die feulich
 hat waf bläm vñ die mällich
 gelb blümen vñ gleich dem so
 fram. Dae maistre Plinius in sei
 nem xij buch spricht. Dae stab
 wurtz hab ainet güten geant
 Dae teant bon fait blaine blet
 te vñ vil stengel vñ auch gar
 elain bläm vñ ain elain samē
 Dae maistre Aucena in seinem
 andn buch spricht. Dae stabwurtz
 sey haif an dem besten grade vñ
 trecken an dem andern. Dae wie
 dig maistre Aucema in seinē an
 dem buch in dē capitel Abrotanū
 spricht. Das stabwurtz gebreuet
 vñ gebilfet vñ das gemilchet
 mit si vñ mällich gemachet vñ das
 geschmiet auf die kal stat machet
 har wachsen. Item stabwurtz
 macht auch wachsen den bart d
 langelom herfar komet do mit bes
 treichē mit neme nym stabwurtz
 sifit mit byll die vñ wenig. Dae
 den bart oder die stat geschmiet
 do man hat beget. Stabwurtz

gesteffe vñ mit byptom wurtz
 gebilfet mit hönig vñ wenig
 ainet salben weise vñ auf wundē
 gelegt ist beulen vñ doonen auf
 der wunden ziehen.



Osterlucien
Das xi Capitel
 Estilogia longa. Die
 weidige maistre Aucena
 Galien. Dae scorde pla
 tearia vñ Plinius spricht. Dae
 die lang holwurtz seye warm an
 dem besten grad vñ trecken an
 dē andn grad. Die lang holwurtz
 hat ain langen sil vñ blette dar
 an die gelliche den bletten an der
 holwurtz allain das ostelucie
 es holwurtz blette wachsen sind
 an dem geiff. Dae wurtz ist lang
 gleich den petrefiligen wurtzeln.
 Plinius spricht. Das balde ostel

lucien die lang vñ die vñ gebul
 fet vñ darinder gemilchet müt
 yegliche gleich vil vñ dife ge
 nüt ain quintin mit warmē wei
 vñ get die miltē genant matrix
 von ier vñ schlägt ait vñ do treibt
 auf die tere gebret. Item das po
 do gram nym wegworte eybisch
 wurtzeln vñ lang holwurtzeln
 yegliche gleich vil vñ miltch dar
 under hönig vñ leg das auff den
 gebreuten ca senfeger vñ mindet
 den vortogen. Dae ostelucie ist
 fast güt an familie wunden des bul
 ftes darē gestret. Dae liden
 psard gewundet wiet vñ vil reitē
 oder tragen der steue des vulfae
 von dife ostelucie in die wundē/
 vñ schlope darauff p fardē milt
 es hailet eschond. Platearius
 nym ostelucie vñ alopatium
 yegliche gleich vil vñ miltch dar
 under vñ hönig vñ mach dar
 auf ain pflaster. Dae pflaster ist
 güt genüzet allen alten schäden/
 als d an ist der trech fistel d woff
 wie die sein mügen an den bamen
 vñ anderwo hailet es senftig.
 Elich daraf gelegt vñ die schaldē
 oder gebreue sollen wech in geseu
 heet werden mit weinlein d es
 mit woffe dariff milt gefotē ist.
 Dae also genüzet benymet die
 auffsetzt ait. Dae scorde ostel
 lucie gebilfet vñ gemilchet mit
 hönig woffe vñ das getwante
 komet osmo das ist das trecken/
 vñ raumet die busi. Dae ain
 psal in seinē leb het d neme baid

Bild mittig
zwischen
zweispaltigem
Text;
Lesefluss:
links rechts,
links rechts

Das Bild wird für die letzte dipl-
Einheit in der rechten Spalte
oben bis zur ersten dipl-Einheit
in der linken Spalte unten
annotiert.

Bild mittig
zwischen
zweispaltigem
Text;
Lesefluss:
links, rechts

Das Bild wird für die letzte dipl-
Einheit in der linken Spalte
oben und die erste dipl-Einheit
in der linken Spalte unten
annotiert.

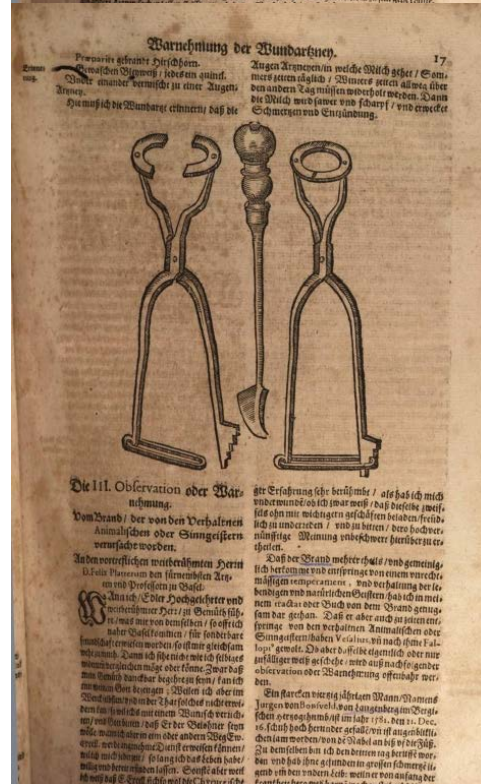
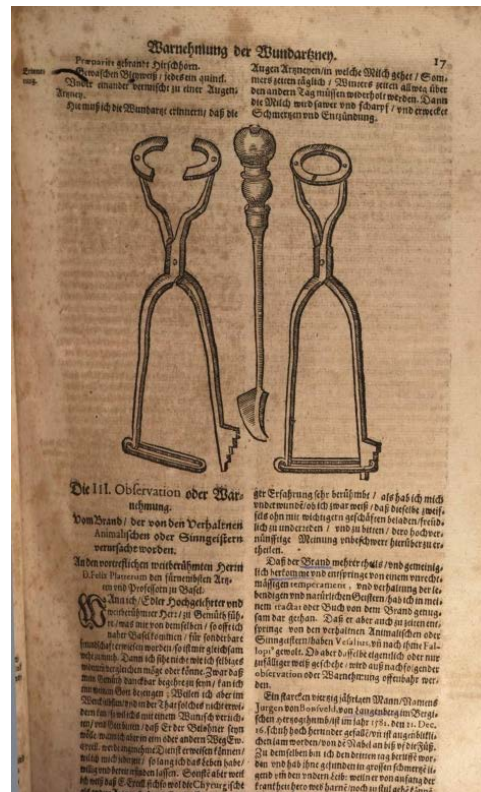


Bild mittig über zweispaltigem Text/ Bild unten in der linken Spalte

Das obere Bild wird an der ersten dipl-Einheit der linken Spalte annotiert, das untere Bild für die nebenstehenden dipl-Einheiten.



Bild von Text umgeben

Das Bild wird für den gesamten Text auf dieser Seite annotiert.

T1 T2 T3
T 4 **B** T5
T6 T7 T8

Typ: Annotationvalue – figure

Wert:
figure
table

Wertbeschreibung:
Allgemeine Kennzeichnung einer Grafik im Text.
Allgemeine Kennzeichnung einer Tabelle im Text

5.11. figure_p

Typ: Layer – figure_p

Name:

figure_p

Dokumentation:

Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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_p

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.humboldt-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’

Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: *Annotationlayer* – figure_p

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Angabe der Seite, auf der eine Abbildung im Original zu sehen ist.

5.12. hi

Typ: *Layer* – hi

Name: hi
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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: basiert auf ‘dipl’

Annotation:
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – hi

Annotationsart: Spannenannotation über einzelne dipl-Einheiten
Beschreibung: Markierung eines irgendwie hervorgehobenen Bereichs.

Typ: Annotationvalue – hi

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

5.13. script

Typ: *Layer* – 5.12. script

Name:	script
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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	2018
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter,

CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ:*Annotationlayer* – script

Annotationsart: Spannenannotation für einzelne dipl-Einheiten
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.

5.14. hi_rend

Typ: *Layer* – hi_rend

Name: hi_rend

Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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	2018
Qualitätsprüfung:	Manuelle Konsistenzprüfung: Korrektur der false positives
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ:*Annotationlayer* – hi_rend

Annotationsart:	Spannenannotation über einzelne dipl-Einheiten
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.

5.15. head

Typ: *Layer* – head

Name:	head
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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:	2018
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’

Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – head

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
Beschreibung: Markierung einer Überschrift.

Typ: Annotationvalue – head

Wert: Wertbeschreibung:
head Allgemeine Markierung einer Überschrift.

5.16. note

Typ: *Layer* – note

Name: note
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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: 2017
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – note

Annotationsart: Spannenannotation über mehrere dipl-Einheiten
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.
caption	In einer Spannenannotation wird die gesamte Bildbeschriftung mit dem Tag „caption“ markiert. Der Text der Bildbeschriftungen wird immer am Anfang des jeweiligen Absatzes, in dem sie realisiert sind, unabhängig davon, ob so mehrere Bildbeschriftungen hintereinander oder/und auf der vorherigen Seite realisiert werden müssen, in die dipl-Ebene/Transkription aufgenommen.
end	Eine Endnote.

5.17. ref

Typ: Layer – ref

Name:	ref
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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:	2017
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, gemeinsame Konvertierung der Formate Excel 2013,

CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – ref

Annotationsart: Spannenannotation über eine oder mehrere dipl-Einheiten
Beschreibung: Referenz einer Fußnote.

Typ: Annotationvalue – ref

Wert: Wertbeschreibung:
ref Referenz einer Fußnote.

5.18. ref_target

Typ: Layer – ref_target

Name: ref_target
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: *Annotationlayer* – ref_target

Annotationsart:	Spannenannotation über eine oder mehrere dipl-Einheiten
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.

5.19. ref_type

Typ: *Layer* – ref_type

Name:	ref_type
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: *Annotationlayer* – ref_type

Annotationsart:	Spannenannotation über eine oder mehrere dipl-Einheiten
Beschreibung:	Art der Referenz, z.B. TEI "noteAnchor".

Typ: *Annotationvalue* – ref_type

Wert:	Wertbeschreibung:
noteAnchor	Art der Referenz.

5.20. quote

Typ: *Layer* – quote

Name:	quote
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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:	2018
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling

Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – quote

Annotationsart: Spannenannotation für einzelne dipl-Einheiten
Beschreibung: Ordnet Fließtext als Zitat oder Nicht-Zitat ein.

Typ: Annotationvalue – quote

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

6. Annotationsebenen – Inhaltliche Annotation

6.1. definition

Typ: *Layer* – definition

Name: definition
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpapper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: *Annotationlayer* – definition

Annotationsart: Spannenannotation über mehrere norm-Einheiten
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: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2018
Qualitätsprüfung:	NA
Editor:	LAUDATIO, LangBank

Typ: Annotationlayer – disease

Annotationsart:	Spannenannotation über eine oder mehrere norm-Einheiten
Beschreibung:	Nennung einer Krankheit (Beschwerden, Symptome etc.) im Text. Es wird jeweils die komplette Phrase, wenn vorhanden auch mit restriktiven Relativsätzen, getagged, z.B. <i>der Bauch, der weh tut</i> oder <i>Bauchweh</i> .

Typ: Annotationvalue – disease

Wert:	Wertbeschreibung:
di	Benennung einer Krankheit

6.3. persname

Typ: Layer – persname

Name:	persname
Dokumentation:	Es werden alle Personennamen annotiert, auf die ein Autor referiert. Es wird der jeweilige Personennamen im Nominativ

angegeben. Eine Liste der vereinheitlichten Namen findet sich unten. Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.

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	2018
Qualitätsprüfung:	manuelle Korrektur
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen, korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2018

Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – persname

Annotationsart: Spannenannotation über eine oder mehrere norm-Einheiten
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. 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 [dieser Liste](#) (diese enthält offizielle Abkürzungen und die entsprechende Ausschreibung vieler BotanikerInnen) und wenn dort nicht aufgeführt nach ihrer Schreibung in der deutschsprachigen Wikipedia. Anstelle von Leerzeichen werden Unterstriche gesetzt. Im Folgenden sind alle bisher in RIDGES vorkommenden eindeutigen Personennamen aufgeführt. Diese werden weiterhin so übernommen. Hinweis: Viele Autoren haben eine Art Künstlernamen und es gibt auch viele verschiedene Schreibweisen. In dieser Ebene muss darauf geachtet werden, dass ein und dieselbe Person immer mit dem gleichen Namen annotiert wird.

Achilles, Adriaan_van_Royen, Aetios_von_Amida, Agrippa_von_Nettesheim, Aimé_Jacques_Alexandre_Bonpland, Albrecht_von_Haller, Alexander_von_Humboldt, Alexander_von_Tralleis, Alexander_Wilhelm_Martini, Alexandre_Henri_Gabriel_de_Cassini, André_Michaux, Andreas_Cleyer, Antoine-Laurent_de_Jussieu, Anton_Rehmann, Apollonius, Apuleius, Aristoteles, Artemis, Asa_Grey, Asklepiades_von_Bithynien, Attalos_I., August_Quirinus_Rivinus, Augustin_Pyramus_De_Candolle, Averroes, Avicenna, Bartholomaeus_Anglicus, Bartolomeo_Platina, Caelius_Apicius, Carl_Heinrich_Schultz_Bipontinus, Carl_Peter_Thunberg, Carl_Sigismund_Kunth, Carl_von_Linné, Carrichter, Caspar_Bauhin, Caspar_Hoffmann, Christian_Friedrich_Lessing, Christian_Gottlieb_Ludwig, Christoph_Wilhelm_Hufeland, Chrysippos_von_Soloi, Constantinus, David_Heinrich_Hoppe, Diodoros, Dominique_Villars, Eberhard_Gmelin,

Eduard_August_von_Regel, Eucharius_Rösslin_der_Jüngere, Fabio_Colonna,
Ferdinand_Jacob_Heinrich_Sir_Müller, François_Fulgis_Chevallier,
Franz_Xaver_von_Wulfen, Franziskus_Joel, Galenos, Gaza, Georg_Franz_Hoffmann,
George_Bentham, Giovanni_da_Vigo, Gregor_Horst, Gustav_Karl_Wilhelm_Karsten,
Guy_de_La_Brosse, Heinrich_Bernhard_Rupp, Heinrich_Kramer, Heinrich_Petreus,
Henri_Ernst_Baillon, Henry_Charles_Andrews, Herakleides_Pontikos, Herkules,
Herman_Boerhaave, Hermann_Hager, Hermann_von_Neuenahr_der_Ältere,
Hermolaus_Barbarus, Hesiod, Hieronymus_Bock, Hipólito_Ruiz_López,
Hippokrates_von_Kos, Homer, Hugh_Algernon_Weddell, Isaac_Newton, Isaac_Rand,
Isaak_ben_Salomon_Israeli, Isidor_von_Sevilla, Jacques_Peschier, Jean_Louis_Lassaigne,
Joachim_Camerarius, Joachim_Camerarius_der_Jüngere, João_de_Loureiro, Johann_Bauhin,
Johann_Georg_Volkamer, Johann_Heinrich_von_Heucher, Johann_Jacob_Dillen,
Johann_Schröder, Johann_Wonnecke_von_Kaub, Johannes_Bodaeus_van_Stapel,
Johannes_Jacobus_Manlius_de_Bosco, Johannes_Manardus, Johannes_Mesuë_der_Ältere,
Johannes_Theodor_Schmalhausen, John_Gerard, John_Hutton_Balfour, John_Kirk,
John_Parkinson, José_Antonio_Pavón_y_Jiménez, Joseph_Pitton_de_Tournefort,
Juan_José_Martínez_de_Lexarza, Karl_Gottfried_Hagen, König_Salomo,
Königin_von_Halikarnassos, Leonhart_Fuchs, Lorenz_Heister,
Lucius_Iunius_Moderatus_Columella, Marcellus_Vergilius,
Marcus_Porcius_Cato_der_Ältere, Marschall_von_Bieberstein, Marsilio_Ficino,
Martin_Schmuck, Matthaeus_Platearius, Mausolos_II, Michael_Ettmüller,
Michelangelo_Tilli, Mithridates, Nathaniel_Wallich, Niccolò_Leonceno, Nicolas_Lémery,
Nicolaus_Joseph_Jacquin, Nikandros_aus_Kolophon,
Nikolai_Stepanowitsch_Turtschaninow, Oreibasios, Orpheus, Otto_Brunfels,
Pablo_de_La_Llave, Pandecta, Paulos_von_Aigina, Paulus. Paulus_der_Arzt,
Pedanios_Dioskurides, Philip_Miller, Pietro_Andrea_Mattioli, Platon, Plinius_der_Ältere,
Pythagoras, Raius, Ramon_Llull, Rembert_Dodoens, Robert_Brown, Robert_Morison,
Rudolf_Goclenius, Rutilius_Taurus_Aemilianus_Palladius, Sebastian_Kneipp,
Serapion_von_Alexandria, Serenus, Simon_Pauli_der_Jüngere, Simon_von_Genua,
Sosimenes, Tabernaemontanus, Teukros_(König_von_Troas), Themison_von_Laodikeia,
Theophilus_Protopatharius, Theophrastos_von_Eresos, Thomas_Nuttall,
Wilhelm_Daniel_Joseph_Koch, Wilhelm_Ernst_Scheffer, William_Henry_Harvey,
William_Jackson_Hooker, William_Watson, Xenokrates, Xenophon, Zenon

6.4. title

Typ: *Layer* – title

Name:	title
Dokumentation:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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	2018
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen; korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: NA
basiert auf ‘dipl’
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – title

Annotationsart: Spannenannotation über mehrere norm-Einheiten
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. Dies gilt nicht für fremdsprachliches Material. Dieses wird aus dem Original übernommen.

6.5. author_ref

Typ: *Layer* – author_ref

Name: author_ref
Dokumentation: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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* – author_ref

Schritt: 1

Aufbereitung: Inhaltliche Annotation
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2018
Qualitätsprüfung: Laura Perlitz
Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt: 2
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: automatisch
Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: *Annotationlayer* – author_ref

Annotationsart: Spannenannotation über eine oder mehrere norm-Einheiten
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:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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:	2018
Qualitätsprüfung:	Laura Perlitz

Editor: Studierende der Humboldt-Universität zu Berlin und der Georg-August-Universität Göttingen

Schritt: 2

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

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

Typ: *Annotationlayer* – reader_ref

Annotationsart: Spannenannotation über eine oder mehrere norm-Einheiten

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 Grofzachtbahren 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:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2018

Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: *Annotationlayer* – plant

Annotationsart: Spannenannotation über eine oder mehrere norm-Einheiten
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: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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, gemeinsame Konvertierung der Formate Excel 2013,
CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden
Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter.
Merging mithilfe des Merger-Moduls. Export ins ANNIS- und
PAULA-Format mit dem ANNISExporter und dem PAULAExporter
(https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: *Annotationlayer* – property

Annotationsart:	Spannenannotation über mehrere norm-Einheiten
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:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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* – name

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der	basiert auf 'norm'

Annotation:
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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling
Datum: 2018
Qualitätsprüfung: NA
Editor: LAUDATIO, LangBank

Typ: Annotationlayer – name

Annotationsart: Spannenannotation über eine oder mehrere norm-Einheiten
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: Informationen zum gesamten Konvertierungsworkflow gibt es hier: <https://github.com/langbank/conversion-pipeline>.
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* – 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, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)
 Format: PAULA-XML, ANNIS
 Segmentierung der Annotation: basiert auf 'norm'
 Qualitätsprüfer: korpling
 Datum: 2018
 Qualitätsprüfung: NA
 Editor: LAUDATIO, LangBank

Typ: *Annotationlayer* – name_type

Annotationsart: Spannenannotation über eine oder mehrere norm-Einheiten
 Beschreibung: Spezifiziert den Typ der Eigennamen.

Typ: *Annotationvalue* – name_type

Wert:	Wertbeschreibung:
herb	Eigenname eines Krauts.
scholar	Eigenname 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:	Informationen zum gesamten Konvertierungsworkflow gibt es hier: https://github.com/langbank/conversion-pipeline .
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:	2018
Qualitätsprüfung:	manuell
Editor:	korpling

Schritt: 2

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

Aufbereitungsart: automatisch

Tool: SaltNPepper, gemeinsame Konvertierung der Formate Excel 2013, CoNLL und PTB nach ANNIS und PAULA mithilfe der folgenden Importer: SpreadsheetImporter, PTBImporter, CoNLLImporter. Merging mithilfe des Merger-Moduls. Export ins ANNIS- und PAULA-Format mit dem ANNISExporter und dem PAULAExporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/snapshot/Pepper_2018.03.27-SNAPSHOT.zip)

Format: PAULA-XML, ANNIS

Qualitätsprüfer: korpling

Datum: 2018

Qualitätsprüfung: NA

Editor: LAUDATIO, LangBank

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	1.0 2.0 3.0 4.0 5.0 6.0 7.0	Version, in der das jeweilige Dokument zum Korpus hinzugefügt wurde.
edition_first	yes	Wenn bekannt, wird hier angegeben, ob es sich

	no	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.
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.
deprelGold	yes	Gibt an, ob die Dependenzpares und die

no

korrespondierende Lemmatisierung sowie das pos-Tagging in dem entsprechenden Dokument manuell korrigiert wurden oder nicht.

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/>

Anhang

Zuordnung von Annotationen und Dokumenten

Ridges hat eine lange Versionsgeschichte. Deshalb sind nicht alle Annotationen in allen Dokumenten enthalten. Die meisten Annotationsebenen wurden von Studierenden der Humboldt-Universität zu Berlin erstellt. Viele wurden aber auch von einzelnen ForscherInnen angelegt, die selbst entschieden haben, welche Dokumente sie annotieren wollen.

In der folgenden Tabelle sind je Annotationsebene die annotierten Dokumente aufgeführt. Manche zusammengehörige Annotationsebenen wurden gruppiert, wenn sie in exakt den gleichen Dokumenten annotiert wurden.

Annotationsebene	annotierte Dokumente
adja_uninfl	AlchymistischePractic_1603_Libavius, AlchymistischePractic-VR_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-VR_1487_vonCuba, HortulusSanitatis_1609_Uffenbach, Krautterbuch_1609_Carrichter, NewKreuterbuch_1563_Handsch, NewKreutterBuch_1539_Bock, NewKreutterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff
atLeast, atMost	AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ContrafaytKreuterbuch_1532_Brunfels, DeutschePflanzennamen_1870_Grassmann, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-6_1821_Wilbrand, Flora-7_1821_Sieber, FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, Krautterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff
attr_gen	BuchDerNatur_1482_vonMegenberg, AlchymistischePractic_1603_Libavius, AlchymistischePractic-

VR_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat,
ArtzneyBuchleinDerKreutter-VR_1532_Tallat,
ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-
VR_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-
CCXLVIII_1532_Brunfels, DeutschePflanzennamen_1870_Grassmann,
EigenschaftenAllerHeilpflanzen_1828_Anonymous,
EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-
6_1821_Wilbrand, Flora-7_1821_Sieber,
FloraDerPreussischenRheinlande_1870_Wirtgen,
FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba,
GartDerGesundheit-VR_1487_vonCuba,
GrundrissKraeuterkunde_1792_Willdenow,
HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter,
MysterivmSigillorvm_1735_Hiebner, NewKreuetterBuch_1539_Bock,
NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs,
NewKreuterbuch_1563_Handsch, NochEinigeWorte_1840_Meyen,
NochEinigeWorte-VR_1840_Meyen,
Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor,
PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor,
SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1-
11_1675_Anonymous, SonderbaresKraeuterbuch-11-
21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser,
Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link,
WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-
VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff

author_ref

AlchymistischePractic_1603_Libavius, AlchymistischePractic-
VR_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat,
ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat,
ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat,
ArtzneyBuchleinDerKreutter-VR_1532_Tallat,
BlackwellischesKraeuterbuch_1750_Blackwell,
BuchDerNatur_1482_vonMegenberg,
ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-
CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-
CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-
CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels,
DeutschePflanzennamen_1870_Grassmann,
EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-
6_1821_Wilbrand, Flora-7_1821_Sieber,
FloraDerPreussischenRheinlande_1870_Wirtgen,
FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba,
GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-
33_1487_Cuba, GartDerGesundheit-VR_1487_vonCuba,
GemeinnuetzigesKraeuterbuch_1874_Siegmund,
GetreueDarstellungUndBeschreibung_1809_Hayne,
GrossesIllustriertesKraeuterbuch_1860_Mueller,
GrundrissKraeuterkunde_1792_Willdenow,

	HortulusSanitatis_1609_Uffenbach, Kraeuterbuch_1914_Losch, Kraeutterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, NatürlichePflanzenfamilien_1887_Engler, NewKreueiterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock, NewKreuetterbuch-CXXXIIIJ-CXXXVIIIJ_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-21-36_1675_Anonymous, TheatrumBotanicum_1696_Verzascha, ThesaurusSanitatis_1673_Nasser, ThesaurusSanitatis_304-321_1673_Nasser, Unterricht_1774_Eisen, ViridariumReformatum_1719_Valentini, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff
cat-const, const, deprel, func (edglabel), lemma-deprel, pos-deprel, pos-const, morph, sentence_end	AlchymistischePractic-VR_1603_Libavius, AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, ArtzneyBuchleinDerKreutter_1532_Tallat, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels, ContrafaytKreuterbuch_1532_Brunfels, GartDerGesundheit-VR_1487_vonCuba, GartDerGesundheit_1487_vonCuba, HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter, NewKreueiterbuch_1563_Handsch, NewKreuetterBuch-VR_1539_Bock, NewKreuetterBuch_1539_Bock, NewKreuterbuch_1543_Fuchs, Paradeiszgaertlein_1588_Rosbach, PflantzGart-VR_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart_1639_Rhagor, WieSichMeniglich-VR_1557_vonBodenstein, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff
clause_type	ArtzneyBuchleinDerKreutter_1532_Tallat, GartDerGesundheit_1487_vonCuba
clean, dipl, norm, lemma, script, lang, lb, pb, pos column	alle GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba
comment	AlchymistischePractic_1603_Libavius, AlchymistischePractic-VR_1603_Libavius, Apothekerlexikon_1793_Hahnemann, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat, BlackwellischesKraeuterbuch_1750_Blackwell, BuchDerNatur_1482_vonMegenberg,

ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-
 CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-
 CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-
 CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels,
 DeutschePflanzennamen_1870_Grassmann,
 EigenschaftenAllerHeilpflanzen_1828_Anonymous,
 EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous,
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder,
 FloraDerPreussischenRheinlande_1870_Wirtgen,
 GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-
 23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba,
 GartDerGesundheit-VR_1487_vonCuba,
 GemeinnuetzigesKraeuterbuch_1874_Siegmund,
 GetreueDarstellungUndBeschreibung_1809_Hayne,
 GrossesIllustriertesKraeuterbuch_1860_Mueller,
 HortulusSanitatis_1609_Uffenbach, Kraeuterbuch_1914_Losch,
 Kraeuterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner,
 NatürlichePflanzenfamilien_1887_Engler,
 NewKreuterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock,
 NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock,
 NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs,
 NewKreuterbuch-c19-25_1543_Fuchs,
 Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor,
 PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor,
 Phythologia_1662_Becher, SonderbaresKraeuterbuch-1-
 11_1675_Anonymous, SonderbaresKraeuterbuch-11-
 21_1675_Anonymous, SonderbaresKraeuterbuch-21-
 36_1675_Anonymous, TheatrumBotanicum_1696_Verzascha,
 ThesaurusSanitatis_1673_Nasser, ThesaurusSanitatis_304-
 321_1673_Nasser, Unterricht_1774_Eisen,
 ViridariumReformatum_1719_Valentini,
 VorlesungenUeberKraeuterkunde_1843_Link,
 WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-
 VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff

comment_lex ArtzneyBuchleinDerKreutter_1532_Tallat,
 GartDerGesundheit_1487_vonCuba

comp, comp_orth, DeutschePflanzennamen_1870_Grassmann,
 prot EigenschaftenAllerHeilpflanzen_1828_Anonymous,
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-
 6_1821_Wilbrand, Flora-7_1821_Sieber,
 FloraDerPreussischenRheinlande_1870_Wirtgen,
 FloraSaturnizans_1722_Henckel,
 GrundrissKraeuterkunde_1792_Willdenow,
 HortulusSanitatis_1609_Uffenbach, Kraeuterbuch_1609_Carrichter,
 MysterivmSigillorvm_1735_Hiebner, NochEinigeWorte_1840_Meyen,
 NochEinigeWorte-VR_1840_Meyen,
 SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1-

<p>comp_n, comp_n_graph, comp_n_mod, comp_lex, infl_fuge</p>	<p>11_1675_Anonymous, SonderbaresKraeuterbuch-11- 21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link AlchymistischePractic_1603_Libavius, AlchymistischePractic- VR_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, ArtzneyBuchleinDerKreutter-Cretanus_1532_Tallat, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch- CCXXXVII-CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch- CCLXXI-CCXC_1532_Brunfels , ContrafaytKreuterbuch-CCLVII- CCLXX_1532_Brunfels , ContrafaytKreuterbuch-VR_1532_Brunfels, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15- 23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba, GartDerGesundheit-VR_1487_vonCuba, HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter, NewKreueterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock, NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich- VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff</p>
<p>comp_amb</p>	<p>AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, ArtzneyBuchleinDerKreutter-Cretanus_1532_Tallat, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch- CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCLVII- CCLXX_1532_Brunfels, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23- 33_1487_Cuba, HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter, NewKreueterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff</p>
<p>comp_a, comp_a_graph</p>	<p>AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, ArtzneyBuchleinDerKreutter- Abrotanum_1532_Tallat,BuchDerNatur_1482_vonMegenberg,</p>

	<p>ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-VR_1487_vonCuba, HortulusSanitatis_1609_Uffenbach, Krautterbuch_1609_Carrichter, NewKreuterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock, NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff</p>
definition	<p>AlchymistischePractic_1603_Libavius, DeutschePflanzennamen_1870_Grassmann, Flora-6_1821_Wilbrand, FloraSaturnizans_1722_Henckel, NewKreuterbuch_1543_Fuchs, Unterricht_1774_Eisen</p>
disease	<p>ArtzneyBuchleinDerKreutter_1532_Tallat, ContrafaytKreuterbuch_1532_Brunfels, EigenschaftenAllerHeilpflanzen_1828_Anonymous, GartDerGesundheit_1487_vonCuba, HortulusSanitatis_1609_Uffenbach, Krautterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, NewKreuterbuch_1543_Fuchs, Paradeiszaertlein_1588_Rosbach, SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff</p>
figure	<p>AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels, FloraDerPreussischenRheinlande_1870_Wirtgen, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba, GemeinnuetzigesKraeuterbuch_1874_Siegmund, GetreueDarstellungUndBeschreibung_1809_Hayne, GrossesIllustriertesKraeuterbuch_1860_Mueller, Kraeuterbuch_1914_Losch, NatürlichePflanzenfamilien_1887_Engler, NewKreuterbuch_1563_Handsch, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, Phythologia_1662_Becher, SonderbaresKraeuterbuch-1-</p>

	11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, SonderbaresKraeuterbuch-21-36_1675_Anonymous, TheatrumBotanicum_1696_Verzascha, ViridariumReformatum_1719_Valentini
figure_rend	AlchymistischePractic_1603_Libavius, NewKreuterbuch_1543_Fuchs
figure_p	ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, SonderbaresKraeuterbuch-21-36_1675_Anonymous
foreign	AlchymistischePractic_1603_Libavius, AlchymistischePractic-VR_1603_Libavius, DeutschePflanzennamen_1870_Grassmann, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-6_1821_Wilbrand, Flora-7_1821_Sieber, FloraDerPreussischenRheinlande_1870_Wirtgen, FloraSaturnizans_1722_Henckel, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, MysterivmSigillorvm_1735_Hiebner, NewKreuterbuch_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff
foreign_trans	AlchymistischePractic_1603_Libavius, DeutschePflanzennamen_1870_Grassmann, Flora-6_1821_Wilbrand, Flora-7_1821_Sieber, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, MysterivmSigillorvm_1735_Hiebner, NewKreuterbuch_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, SonderbaresKraeuterbuch-1-11_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein
form_disease, form_noun, form_prep, form_referent, form_relpron, herbprep, noun_nom, position_referent, position_rel, position_verb_rel, problem	ArtzneyBuchleinDerKreutter_1532_Tallat, GartDerGesundheit_1487_vonCuba
head	AlchymistischePractic_1603_Libavius, AlchymistischePractic-VR_1603_Libavius, Apothekerlexikon_1793_Hahnemann, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, BlackwellischesKraeuterbuch_1750_Blackwell,

BuchDerNatur_1482_vonMegenberg,
 ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-
 CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-
 CCXC_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels,
 DeutschePflanzennamen_1870_Grassmann,
 EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous,
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-
 6_1821_Wilbrand, Flora-7_1821_Sieber,
 FloraDerPreussischenRheinlande_1870_Wirtgen,
 FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba,
 GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-
 33_1487_Cuba, GartDerGesundheit-VR_1487_vonCuba,
 GetreueDarstellungUndBeschreibung_1809_Hayne,
 GrossesIllustriertesKraeuterbuch_1860_Mueller,
 GrundrissKraeuterkunde_1792_Willdenow,
 HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter,
 MysterivmSigillorum_1735_Hiebner,
 NatürlichePflanzenfamilien_1887_Engler,
 NewKreuterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock,
 NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock,
 NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs,
 NewKreuterbuch-c19-25_1543_Fuchs, NochEinigeWorte-
 VR_1840_Meyen, Paradeiszgaertlein_1588_Rosbach,
 PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-
 VR_1639_Rhagor, Phythologia_1662_Becher,
 SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1-
 11_1675_Anonymous, SonderbaresKraeuterbuch-11-
 21_1675_Anonymous, SonderbaresKraeuterbuch-21-
 36_1675_Anonymous, TheatrumBotanicum_1696_Verzascha,
 ThesaurusSanitatis_1673_Nasser, ThesaurusSanitatis_304-
 321_1673_Nasser, Unterricht_1774_Eisen,
 ViridariumReformatum_1719_Valentini,
 VorlesungenUeberKraeuterkunde_1843_Link,
 WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-
 VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff

herbname_norm ArtzneyBuchleinDerKreutter_1532_Tallat,
 ContrafaytKreuterbuch_1532_Brunfels, NewKreuterbuch_1543_Fuchs,
 WieSichMeniglich_1557_vonBodenstein,
 GartDerGesundheit_1487_vonCuba

hi AlchymistischePractic_1603_Libavius, AlchymistischePractic-
 VR_1603_Libavius, DeutschePflanzennamen_1870_Grassmann,
 EigenschaftenAllerHeilpflanzen_1828_Anonymous,
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder,
 FloraDerPreussischenRheinlande_1870_Wirtgen,
 FloraSaturnizans_1722_Henckel,
 GrundrissKraeuterkunde_1792_Willdenow,
 HortulusSanitatis_1609_Uffenbach,

MysterivmSigillorvm_1735_Hiebner, NewKreuterbuch_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, SonderbaresKrauterbuch-1-11_1675_Anonymous, SonderbaresKrauterbuch-11-21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKrauterkunde_1843_Link, Wund-
Artzney_1652_Greiff

hi_rend

AlchymistischePractic_1603_Libavius, Apothekerlexikon_1793_Hahnemann, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, BlackwellischesKrauterbuch_1750_Blackwell, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, DeutschePflanzennamen_1870_Grassmann, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous, EinleitungZuDerKrauterkenntnisz_1764_Oeder, FloraDerPreussischenRheinlande_1870_Wirtgen, FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba, GemeinnuetzigesKrauterbuch_1874_Siegmund, GetreueDarstellungUndBeschreibung_1809_Hayne, GrossesIllustriertesKrauterbuch_1860_Mueller, GrundrissKrauterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, Krauterbuch_1914_Losch, Krautterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, NatürlichePflanzenfamilien_1887_Engler, NewKreuterbuch_1563_Handsch, NewKreutterBuch_1539_Bock, NewKreutterbuch-CXXXIII-CXXXVIII_1539_Bock, NewKreutterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, Phythologia_1662_Becher, SchweizerischerBotanicus_1687_vonRoll, SonderbaresKrauterbuch-1-11_1675_Anonymous, SonderbaresKrauterbuch-11-21_1675_Anonymous, SonderbaresKrauterbuch-21-36_1675_Anonymous, TheatrumBotanicum_1696_Verzascha, ThesaurusSanitatis_1673_Nasser, ThesaurusSanitatis_304-321_1673_Nasser, Unterricht_1774_Eisen,

	ViridariumReformatum_1719_Valentini, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich- VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff
interpretation	AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ContrafaytKreuterbuch_1532_Brunfels, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora- 6_1821_Wilbrand, Flora-7_1821_Sieber, FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1- 11_1675_Anonymous, SonderbaresKraeuterbuch-11- 21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff
KOUS_sem, subclause_type, position_verb	AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ContrafaytKreuterbuch_1532_Brunfels, DeutschePflanzennamen_1870_Grassmann, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora- 6_1821_Wilbrand, Flora-7_1821_Sieber, FloraDerPreussischenRheinlande_1870_Wirtgen, FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, NewKreuterbuch_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1- 11_1675_Anonymous, SonderbaresKraeuterbuch-11- 21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff,
mod_referent	ArtzneyBuchleinDerKreutter_1532_Tallat, ContrafaytKreuterbuch_1532_Brunfels, GartDerGesundheit_1487_vonCuba
morph_ellipsis	AlchymistischePractic_1603_Libavius, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch- CCXXXVII-CCXLVIII_1532_Brunfels,

	<p>DeutschePflanzennamen_1870_Grassmann, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-7_1821_Sieber, FloraDerPreussischenRheinlande_1870_Wirtgen, FloraSaturnizans_1722_Henckel, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, NewKreuetterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NochEinigeWorte_1840_Meyen, Paradeiszgaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff</p>
name, name_type	<p>DeutschePflanzennamen_1870_Grassmann, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen</p>
note	<p>AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels, Flora-6_1821_Wilbrand, FloraDerPreussischenRheinlande_1870_Wirtgen, FloraSaturnizans_1722_Henckel, HortulusSanitatis_1609_Uffenbach, Kraeuterbuch_1914_Losch, MysterivmSigillorvm_1735_Hiebner, NewKreuetterBuch_1539_Bock, NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch-c19-25_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, TheatrumBotanicum_1696_Verzascha, VorlesungenUeberKraeuterkunde_1843_Link, Wund-Artzney_1652_Greiff</p>
ocr	<p>ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels GartDerGesundheit-c15-23_1487_Cuba GartDerGesundheit-c23-33_1487_Cuba NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock NewKreuterbuch-c19-25_1543_Fuchs</p>
pb_ana	<p>SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch_21-36_1675_Anonymous</p>
pb_n	<p>AlchymistischePractic_1603_Libavius, AlchymistischePractic-</p>

VR_1603_Libavius, Apothekerlexikon_1793_Hahnemann,
 ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-
 CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-
 CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-
 CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels,
 DeutschePflanzennamen_1870_Grassmann,
 EigenschaftenAllerHeilpflanzen_1828_Anonymous,
 EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous,
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-
 6_1821_Wilbrand, Flora-7_1821_Sieber,
 FloraDerPreussischenRheinlande_1870_Wirtgen,
 FloraSaturnizans_1722_Henckel, GartDerGesundheit-
 VR_1487_vonCuba, GemeinnuetzigesKraeuterbuch_1874_Siegmund,
 GrossesIllustriertesKraeuterbuch_1860_Mueller,
 GrundrissKraeuterkunde_1792_Willdenow,
 HortulusSanitatis_1609_Uffenbach, Kraeuterbuch_1914_Losch,
 Kraeutterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner,
 NatürlichePflanzenfamilien_1887_Engler,
 NewKreuterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock,
 NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock,
 NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs,
 NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen,
 Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor,
 PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor,
 Phythologia_1662_Becher, SchweizerischerBotanicus_1687_vonRoll,
 SonderbaresKraeuterbuch-1-11_1675_Anonymous,
 SonderbaresKraeuterbuch-11-21_1675_Anonymous,
 SonderbaresKraeuterbuch-21-36_1675_Anonymous,
 TheatrumBotanicum_1696_Verzascha,
 ThesaurusSanitatis_1673_Nasser, ThesaurusSanitatis_304-
 321_1673_Nasser, Unterricht_1774_Eisen,
 ViridariumReformatum_1719_Valentini,
 VorlesungenUeberKraeuterkunde_1843_Link,
 WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff

persname

AlchymistischePractic-VR_1603_Libavius,
 ArtzneyBuchleinDerKreutter_1532_Tallat,
 ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat,
 ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat,
 ArtzneyBuchleinDerKreutter-VR_1532_Tallat,
 BlackwellischesKraeuterbuch_1750_Blackwell,
 BuchDerNatur_1482_vonMegenberg,
 ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-
 CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-
 CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-
 CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels,
 GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-
 23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba,

GartDerGesundheit-VR_1487_vonCuba,
 GemeinnuetzigesKraeuterbuch_1874_Siegmund,
 GetreueDarstellungUndBeschreibung_1809_Hayne,
 GrossesIllustriertesKraeuterbuch_1860_Mueller,
 Kraeuterbuch_1914_Losch, NatürlichePflanzenfamilien_1887_Engler,
 NewKreueiterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock,
 NewKreuetterbuch-CXXXIIIJ-CXXXVIIIJ_1539_Bock,
 NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs,
 NewKreuterbuch-c19-25_1543_Fuchs, SonderbaresKraeuterbuch-21-
 36_1675_Anonymous, TheatrumBotanicum_1696_Verzascha,
 ThesaurusSanitatis_304-321_1673_Nasser,
 ViridariumReformatum_1719_Valentini,
 WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-
 VR_1557_vonBodenstein

plant

ArtzneyBuchleinDerKreutter_1532_Tallat,
 ContrafaytKreuterbuch_1532_Brunfels,
 DeutschePflanzennamen_1870_Grassmann,
 EigenschaftenAllerHeilpflanzen_1828_Anonymous,
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-
 6_1821_Wilbrand, Flora-7_1821_Sieber,
 FloraDerPreussischenRheinlande_1870_Wirtgen,
 FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba,
 GrundrissKraeuterkunde_1792_Willdenow,
 HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter,
 MysterivmSigillorvm_1735_Hiebner, NewKreuterbuch_1543_Fuchs,
 NochEinigeWorte_1840_Meyen, Paradeiszaertlein_1588_Rosbach,
 PflantzGart-c4_1639_Rhagor, SonderbaresKraeuterbuch-1-
 11_1675_Anonymous, SonderbaresKraeuterbuch-11-
 21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser,
 Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link,
 WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff

ppk_e1

AlchymistischePractic_1603_Libavius, AlchymistischePractic-
 VR_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat,
 ContrafaytKreuterbuch_1532_Brunfels,
 EigenschaftenAllerHeilpflanzen_1828_Anonymous,
 EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-
 6_1821_Wilbrand, Flora-7_1821_Sieber,
 FloraDerPreussischenRheinlande_1870_Wirtgen,
 FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba,
 GrundrissKraeuterkunde_1792_Willdenow,
 HortulusSanitatis_1609_Uffenbach, Kraeutterbuch_1609_Carrichter,
 MysterivmSigillorvm_1735_Hiebner, NewKreuterbuch_1543_Fuchs,
 NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen,
 Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor,
 PflantzGart-c4_1639_Rhagor,
 SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1-
 11_1675_Anonymous, SonderbaresKraeuterbuch-11-

	21_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff
ppk_e2	AlchymistischePractic_1603_Libavius, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-7_1821_Sieber, FloraDerPreussischenRheinlande_1870_Wirtgen, GartDerGesundheit_1487_vonCuba, NochEinigeWorte_1840_Meyen, VorlesungenUeberKraeuterkunde_1843_Link
ppk_e3	EinleitungZuDerKraeuterkenntnisz_1764_Oeder, NochEinigeWorte_1840_Meyen
property	AlchymistischePractic_1603_Libavius, DeutschePflanzennamen_1870_Grassmann, Flora-6_1821_Wilbrand, Flora-7_1821_Sieber, FloraSaturnizans_1722_Henckel, GrundrissKraeuterkunde_1792_Willdenow, NewKreuterbuch_1543_Fuchs, NochEinigeWorte_1840_Meyen, PflantzGart-c4_1639_Rhagor, SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, Unterricht_1774_Eisen,
quote	AlchymistischePractic_1603_Libavius, AlchymistischePractic-VR_1603_Libavius, Apothekerlexikon_1793_Hahnemann, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, ArtzneyBuchleinDerKreutter-Centaurea_1532_Tallat, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, BlackwellischesKraeuterbuch_1750_Blackwell, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels, DeutschePflanzennamen_1870_Grassmann, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EigenschaftenAllerHeilpflanzen-149-165_1828_Anonymous, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-6_1821_Wilbrand, Flora-7_1821_Sieber, FloraDerPreussischenRheinlande_1870_Wirtgen, FloraSaturnizans_1722_Henckel, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba, GemeinnuetzigesKraeuterbuch_1874_Siegmund, GetreueDarstellungUndBeschreibung_1809_Hayne, GrossesIllustriertesKraeuterbuch_1860_Mueller, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, MysterivmSigillorvm_1735_Hiebner, NatürlichePflanzenfamilien_1887_Engler, NewKreutterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock, NewKreutterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, Paradeiszgaertlein_1588_Rosbach, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor,

	<p>SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, SonderbaresKraeuterbuch-21-36_1675_Anonymous, ThesaurusSanitatis_1673_Nasser, ThesaurusSanitatis_304-321_1673_Nasser, Unterricht_1774_Eisen, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff</p>
reader_ref	<p>AlchymistischePractic_1603_Libavius, AlchymistischePractic-VR_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, ArtzneyBuchleinDerKreutter-VR_1532_Tallat, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-6_1821_Wilbrand, FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba, GartDerGesundheit-VR_1487_vonCuba, GemeinnuetzigesKraeuterbuch_1874_Siegmund, GrossesIllustriertesKraeuterbuch_1860_Mueller, GrundrissKraeuterkunde_1792_Willdenow, Kraeutterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, NewKreuetterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock, NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch_1543_Fuchs, NewKreuterbuch-c19-25_1543_Fuchs, NochEinigeWorte_1840_Meyen, Paradeiszgaertlein_1588_Rosbach, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, SchweizerischerBotanicus_1687_vonRoll, TheatrumBotanicum_1696_Verzascha, ThesaurusSanitatis_1673_Nasser, ThesaurusSanitatis_304-321_1673_Nasser, Unterricht_1774_Eisen, ViridariumReformatum_1719_Valentini, VorlesungenUeberKraeuterkunde_1843_Link, WieSichMeniglich_1557_vonBodenstein, WieSichMeniglich-VR_1557_vonBodenstein, Wund-Artzney_1652_Greiff</p>
ref, ref_target	<p>Flora-6_1821_Wilbrand, FloraDerPreussischenRheinlande_1870_Wirtgen, FloraSaturnizans_1722_Henckel, MysterivmSigillorvm_1735_Hiebner, NochEinigeWorte_1840_Meyen, NochEinigeWorte-VR_1840_Meyen, VorlesungenUeberKraeuterkunde_1843_Link, Wund-Artzney_1652_Greiff</p>
ref_type	<p>Flora-6_1821_Wilbrand, FloraSaturnizans_1722_Henckel,</p>

	NochEinigeWorte_1840_Meyen
reference	AlchymistischePractic_1603_Libavius, AlchymistischePractic-VR_1603_Libavius, BuchDerNatur_1482_vonMegenberg, ContrafaytKreuterbuch-CCXXXVII-CCXLVIII_1532_Brunfels, HortulusSanitatis_1609_Uffenbach, Krautterbuch_1609_Carrichter, NewKreuetterbuch_1563_Handsch, NewKreuetterBuch_1539_Bock, Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, PflantzGart-VR_1639_Rhagor, WieSichMeniglich_1557_vonBodenstein, Wund-Artzney_1652_Greiff
title	AlchymistischePractic-VR_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, BlackwellischesKraeuterbuch_1750_Blackwell, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, ContrafaytKreuterbuch-CCLXXI-CCXC_1532_Brunfels, ContrafaytKreuterbuch-VR_1532_Brunfels, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-c23-33_1487_Cuba, GartDerGesundheit-VR_1487_vonCuba, GetreueDarstellungUndBeschreibung_1809_Hayne, Kraeuterbuch_1914_Losch, NewKreuetterBuch_1539_Bock, NewKreuetterbuch-CXXXIIIJ-CXXXVIIJ_1539_Bock, NewKreuetterBuch-VR_1539_Bock, NewKreuterbuch-c19-25_1543_Fuchs, Phythologia_1662_Becher, TheatrumBotanicum_1696_Verzascha, ThesaurusSanitatis_304-321_1673_Nasser, ViridariumReformatum_1719_Valentini
unclear	AlchymistischePractic_1603_Libavius, ArtzneyBuchleinDerKreutter_1532_Tallat, ArtzneyBuchleinDerKreutter-Abrotanum_1532_Tallat, BlackwellischesKraeuterbuch_1750_Blackwell, ContrafaytKreuterbuch_1532_Brunfels, ContrafaytKreuterbuch-CCLVII-CCLXX_1532_Brunfels, DeutschePflanzennamen_1870_Grassmann, EigenschaftenAllerHeilpflanzen_1828_Anonymous, EinleitungZuDerKraeuterkenntnisz_1764_Oeder, Flora-6_1821_Wilbrand, Flora-7_1821_Sieber, FloraSaturnizans_1722_Henckel, GartDerGesundheit_1487_vonCuba, GartDerGesundheit-c15-23_1487_Cuba, GartDerGesundheit-VR_1487_vonCuba, GetreueDarstellungUndBeschreibung_1809_Hayne, GrundrissKraeuterkunde_1792_Willdenow, HortulusSanitatis_1609_Uffenbach, Krautterbuch_1609_Carrichter, MysterivmSigillorvm_1735_Hiebner, NewKreuetterBuch_1539_Bock, NewKreuetterBuch-VR_1539_Bock, Paradeiszaertlein_1588_Rosbach, PflantzGart_1639_Rhagor, PflantzGart-c4_1639_Rhagor, SchweizerischerBotanicus_1687_vonRoll, SonderbaresKraeuterbuch-1-11_1675_Anonymous, SonderbaresKraeuterbuch-11-21_1675_Anonymous, SonderbaresKraeuterbuch-21-

36_1675_Anonymous, ThesaurusSanitatis_1673_Nasser,
Unterricht_1774_Eisen, WieSichMeniglich_1557_vonBodenstein,
Wund-Artzney_1652_Greiff

Transkription und Normalisierung

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.

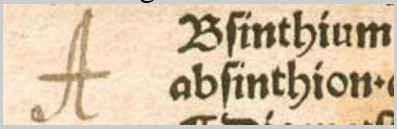
Die **norm**-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier die Tokenisierung, die Orthografie, Flexionsformen und Wortbildungen einheitlich dem nhd. Sprachstand angepasst werden.

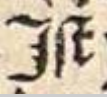

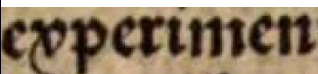
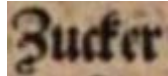
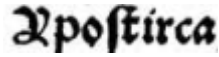
Durch die einzelnen Normalisierungsschritte und der daraus resultierenden multiplen Segmentierung ist es möglich, die uneinheitliche Orthografie in einzelnen Texten sowie im diachronen Verlauf zu vereinheitlichen und dennoch die spezifischen Besonderheiten des jeweiligen Textes/der jeweiligen Sprachstufe 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.

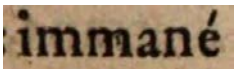
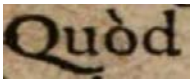
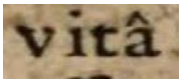

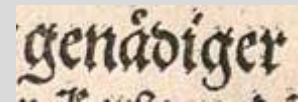
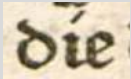
Nachfolgend werden die Richtlinien zur Transkription und Normalisierung im Einzelnen gegenübergestellt. Allgemein für jede Segmentierungsebene sowie Annotationsebene gilt, dass keine Leerzeilen oder Leerzeichen enthalten sein dürfen. Dazu wird für jeden Fall ein Beispiel aus dem Ridges Herbolgy Version 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.

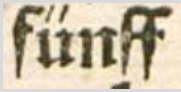
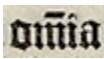



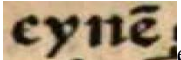


- 1 Transkribieren
 - 1.1 Lautzeichen
 - 1.2 Diakritika
 - 1.3 Satzzeichen
 - 1.4 Sonstige Zeichen
 - 1.5 Ligaturen
 - 1.6 Griechische Schrift
 - 1.7 Typographische Besonderheiten
- 2 Tokenisieren
 - 2.1 Wort
 - 2.2 Zeilenumbrüche
 - 2.3 Abkürzungen
 - 2.4 Zahlen
 - 2.5 Satzzeichen
 - 2.6 Fußnoten
- 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
	ABsinthium	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	Wie in der dipl-Ebene	Anpassung an moderne Orthographie.


	Zeichen individuell nach moderner Orthografie zu interpretieren. 		
	dipl	clean	norm
	Jft	Jst	Ist
1.2	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
	Das Zeichen '3' wird in der dipl-Ebene nicht von 'Z' unterschieden.	Wie in der dipl-Ebene	
	dipl	clean	norm
	 Zucker	Zucker	Zucker
	In manchen Drucken sieht das Zeichen „Y“ etwas anders aus; es wird aber als „normales Y“ transkribiert.		
	dipl	clean	norm
	 Ypostirca	Ypostirca	Ypostirca
1.3	Alle Akzente werden übernommen. Hierbei ist zu beachten, dass der jeweilige unäre	Alle Akzente werden übernommen.	Alle Akzente werden übernommen.

Hex-Code zu verwenden ist. Akut („ó“) - U 0300 Gravis („ò“)- U 0300 Zirkumflex („ô“) - U 0302 oder über sog. dead keys		
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
 'í' 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
 sciētiā meā	scientiam meam	scientiam meam
 eĩ	ein	ein
 samē	samen	Samen
 eynē	eynem	einem
 nēlich	nemlich	nämlich
- Verdopplung eines Buchstaben		
 hūmel	himmel	Himmel

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


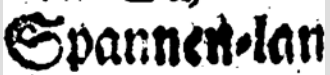
		inn	in
- Abkürzung frequenter Wörter und Endungen			
		unn (sic!)	und
		umm (sic!)	um
		darumm (sic!)	darum
	 complexiō	complexiom (sic!)	Komplexion
	Zwibel od Kn Das Dicit-Zeichen (Abkürzung für “der”) wird in der dipl-Ebene mit ð (AltGr oder als Sonderzeichen)	Es wird in der clean automatisch mit "der" 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 ⁹	deus	deus
	 Dieses Zeichen wird durch v+U+0309 realisiert.	Ersetzung durch <er>	Ersetzung durch <er>
	dipl	clean	norm
	v̇rzeren	vertzeren	verzehren
1.4	 Das Zeichen für morphologische	Die morphologische Trennung durch „-“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt.	Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum


	Worttrennung (oder Zeilenumbruch) wird mit U 2E17 ² übernommen.	Komposita, die in der dipl-Ebene mit einem „-“ realisiert 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.	wird analog zur modernen Orthographie zusammengeschieden. (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
	¶	¶	¶
1.5	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 übernommen.	Dieses Symbol wird mit dem U 2299 übernommen.	Dieses Symbol wird mit dem U 2299 übernommen.

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

	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
	\$	\$	\$
	 Dieses Zeichen wird als 7 (U+204A) transkribiert. Es handelt sich hierbei um das Tironische „et“.		
	7	et	et
1.6	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
 Dies ist eine Ligatur aus v̇ und v. Darstellung mit v+U+0308 Die Ligatur wird aufgelöst v̇v	v̇v	
dipl	clean	norm
 heüvschrecken	heüvschrecken	Heuschrecken
Wie auch u.a. in Beispiel 1.2 wird die Ligatur mit dem langen „f“ und „z“ übernommen.	Die Ligatur wird analog zu „sz“ aufgelöst.	Nach modernen Orthografieregeln wird für „sz“ das „ß“ verwendet.
dipl	clean	norm
fleißige	fleißige	fleißige
Griechische Ligaturen werden nicht abgebildet, sondern bereits in der dipl aufgelöst:  Ligatur aus o und ζ Als Hilfe: https://de.wikipedia.org/wiki/Griechisches_Alphabet#/media/File:Greek_alphabet_ligatures.jpg g		

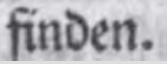
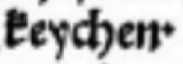

	dipl	clean	norm
	ἄκανθος	ἄκανθος	ἄκανθος
	 Dies bedeutet "etc."		
	dipl	clean	norm
	& c.	& c.	etc.
	Das kaufmännische Und „&“ wird mit U 0026 übernommen.	Das kaufmännische Und wird übernommen.	Das kaufmännische Und wird übernommen.
	dipl	clean	norm
	&	&	&
1.7	Sprachliches Material mit anderen Schriftsystemen (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
1.8	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		Wortbildung und Großschreibung, die nicht der modernen Orthographieregeln entsprechen, werden angeglichen.

	Orthographieregeln wird analog zur Textgrundlage tokenisiert.		
	dipl	clean	norm
	Spannen=lang	Spannen-lang	spannenlang
2.1		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 Orthografiereregeln 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.	

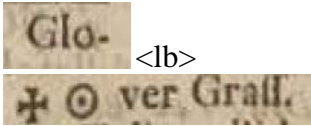
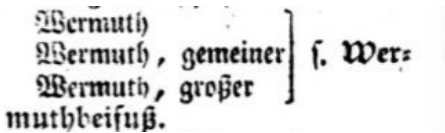
	dipl	clean	norm
	ge	ge	genannt
	nent	nent	
	Komposita, egal welcher Komplexität oder Bildungsweise, mit Gleichheitszeichen (U 2E17) werden als ein Token realisiert.	Die morphologische Trennung durch „≠“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt.	Gleichheitszeichen, die Kompositabildung 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	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.)

		Kompositum wird in einer Spanne zusammengefasst wiedergegeben.	
	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 satzbeendenden 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

	<p>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 auseinander, werden sie auch als verschiedene Tokens realisiert.</p>		<p>Norm: immer als mehrere Token</p>
2.5	<p></p> <p>Satzinterpunktion wie Kommata, Punkte, Virgeln oder Semikola werden als jeweils eigenständige Token realisiert.</p>		
	dipl	clean	norm
	finden	finden	finden
	.	.	.
	<p></p> <p>Die Höhe von Punkten wird nicht beachtet. Da wir nicht immer zweifelsfrei feststellen können, wann ein Punkt "hoch" ist und wann "tief" transkribieren wir das alles als normale Punkte.</p>		
	keychen	keychen	keuchen
	.	.	.
	<p></p> <p>Dieses Zeichen wird als Punkt transkribiert.</p>		

	werden	werden	werden
	.	.	.
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 (Postposition) stachelecht	halben stachelecht	halber (<i>allenthalben</i> ist ein Adverb und bleibt <i>allenthalben</i>) stachelig
3.5	Ausgestorbenes lexikalisches Material wird mit modernen Orthographieregeln übernommen, aber nicht lexikografisch übersetzt.		
	dipl	clean	norm
	Vergefz	Vergefz	Vergess
3.6	Flexionslose Adjektive, die im Neuhochdeutschen eine overte Flexion enthalten, werden in der norm flektiert, ohne dass der Kasus geändert wird.		
	das kal haupt	das kal haupt	das kahle Haupt
	bei schweinem fleisch	bei schweinem fleisch	bei schweinenem Fleisch
3.6	Intervenieren innerhalb eines Wortes Sonderzeichen, dann werden die Sonderzeichen in der norm-Ebene ignoriert und das unterbrochene Wort zusammengeschrieben.		
	dipl	clean	norm
		Glo ✖ ⊙ ver Grass	Glover Grass
	Text hinter (geschweiften) Klammern, wird nach dem umklammerten Textabschnitt eingefügt.		
	dipl		

	Wermuth Wermuth , gemeiner Wermuth , großer f. Werz muthbeifuß .
	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 zu ‚note‘.

dipl	ref	ref_target
.		
PAltinaten		
49)	ref	#f1
Wass		

dipl	not	xml_id
.		
49)		
PAltinaten		
find		
unterschiedener		
Arten		
,		
als		
PAltinaca		
Domestica		
,		
fylveltris		
,		
marina		
,		
Syriac_		
,		
man		
kan		
Hn.		
von		
Franckenau		
KräuterLexic.		
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“	ë	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
Tironisches „et“	7	204A

Tabelle SEQ Tabelle * ARABIC 2: Zeichentabelle

Tabelle: Normalisierungsbeispiele

	clean	norm
1	-icht/-echt/-acht	-(l)ich/-(l)ig
	ästicht	ästig
	knöpfflicht	knöpfflich
	haarecht	haarig
	holzecht	holzig
	schattecht	schattig
	langlecht	länglich
	laulecht	laulig
	weißlecht	weißlich
	rohtlecht	rötlich
	stachelecht	stachelig
2	deutlich erkennbare Ortsnamen werden als nhd. Form normalisiert	
	Franckfort	Frankfurt
	Muttetz	Muttenz
3	-lich-Adverbien werden durch nhd. lich-lose Formen ersetzt	
	kräftiglich	kräftig
	wunderbarlich	wunderbar
4	für- <i>fürnehmlich</i>	vor- <i>vornehmlich</i>
5	-für herfür	-vor hervor
6	fnhd. <i>etwan</i>	nhd. <i>etwa</i> (Adverb) nhd. <i>etwas</i> (Pronomen)
7	-erin	flektierte nhd. Form
	<i>in ein erhabnes Steinerin Grab gelegt</i>	steinernes
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)	
	wann wenn es einen Nebensatz einleitet → keine V2 Stellung, sondern VLetzt (auch VLetzt mit Nachfeldbesetzung: <i>wenn man Köl- und Haselbaum pflanzt zu Weinreben</i>)	wenn
	wann wenn es einen Hauptsatz einleitet → V2 und übersetzbar it nhd. ‚denn‘	wann
	wenn wenn es einen Hauptsatz einleitet → V2 und übersetzbar it nhd. ‚denn‘	wann
	wenn wenn es einen Nebensatz einleitet → keine V2 Stellung, sondern VLetzt (auch VLetzt mit Nachfeldbesetzung: <i>wenn man Köl- und Haselbaum pflanzt zu Weinreben</i>)	wenn
12	Eigennamen/Fremdwörter in Komposita werden getrennt geschrieben	

	St. Anthonius Feuer	Sankt Anthonius Feuer
	Cassia fistel	Cassia Fistel
13	Einzelentscheidungen	
	benommen	genommen
	Beweisung brauch (= Gebrauch)	Beweisung Brauch
	dannenher	daher
	dennocht	dennoch
	der selbe	der selbe
	derselbe	derselbe
	derowegen	deswegen
	entbeut	entbietet
	Epff	Epff
	fahen	fangen
	gel	gel
	gepreist	gepriesen
	gleich(e)sfalls	gleichfalls
	guldin	golden
	halb	halber
	harm	Harn
	harmen	harnen
	hauffecht	häufig
	leichtlich	leicht
	liebnusz	Liebnis
	Mannen	Männern
	mehrer teils	mehrenteils
	morgen (= morgens)	morgens
	namlichen	nämlich
	Nutz	Nutzen
	pestnachen	Pastinaken
	samlen	sammeln
	schlahen	schlagen
	sonders	sonders
	(wegen der) Viele	Viele

vierecket	viereckig
vor	vorher
wehtum/wehtumb	Wehtun
wofer	wofern
wurz	wurzel
yedickest	gedihest (Lexer: ‚oft‘)
yetliches	jegliches
zeuchen	ziehen

In Verbindung mit Excel benutzte Tools

Excel-Makro

SearchAndMerge.bas: <https://hu.berlin/search-and-merge>

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

Python-Skript

clean-skript_V3.py: <https://hu.berlin/clean-script-v3>

Die neuen Texte ab Version 5 stellten neue Anforderungen an die <clean>-Ebene, neue Zeichenersetzungen wurden nötig sowie ein komplett neuer Umgang mit Nasalstrichen. Die vorigen 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 ab 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.

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*.

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.