

Dokumentation des Korpus
Ridges Herbiology Version 5.0

auf Grundlage des Metadatenframeworks nach LAUDATIO

Stand 07.06.2016

-deutsch-

Malte Belz
Carolin Odebrecht
Laura Perlitz
Gohar Schnelle
Vivian Voigt

Inhalt

1. Korpus.....	6
2. Dokumente.....	10
2.1. Kraeuterbuch_1914_Losch.....	10
2.2. FloraDerPreussischenRheinlande_1870_Wirtgen.....	10
2.3. DeutschePflanzennamen_1870_Grassmann.....	11
2.4. VorlesungenUeberKraeuterkunde_1843_Link.....	11
2.5. NochEinigeWorte_1840_Meyen.....	12
2.6. EigenschaftenAllerHeilpflanzen_1828_Anonymous.....	13
2.7. Flora-7_1821_Sieber.....	13
2.8. Flora-6_1821_Wilbrand.....	14
2.9. GrundrissKraeuterkunde_1792_Willdenow.....	14
2.10. Unterricht_1774_Eisen.....	15
2.11. EinleitungZuDerKraeuterkenntnisz_1764_Oeder.....	16
2.12. MysterivmSigillorvm_1735_Hiebner.....	16
2.13. FloraSaturnizans_1722_Henckel.....	17
2.14. ViridariumReformatum_1719_Valentini.....	17
2.15. TheatrumBotanicum_1696_Verzascha.....	18
2.16. SchweizerischerBotanicus_1687_vonRoll.....	18
2.17. SonderbaresKraeuterbuch-11-21_1675_Anonymous.....	19
2.18. SonderbaresKraeuterbuch-1-11_1675_Anonymous.....	20
2.19. ThesaurusSanitatis_1673_Nasser.....	20
2.20. Phythologia_1662_Becher.....	21
2.21. Wund-Artzney_1652_Greiff.....	21
2.22. PflantzGart-Vorrede_1639_Rhagor.....	22
2.23. PflantzGart-c4_1639_Rhagor.....	22
2.24. PflantzGart_1639_Rhagor.....	23
2.25. Kraeuterbuch_1609_Carrichter.....	24
2.26. HortulusSanitatis_1609_Uffenbach.....	24
2.27. AlchymistischePractic_1603_Libavius.....	25
2.28. Paradeiszgaertlein_1588_Rosbach.....	25
2.29. NewKreueterbuch_1563_Handsch.....	26
2.30. WieSichMeniglich_1557_vonBodenstein.....	27
2.31. NewKreueterbuch_1543_Fuchs.....	27
2.32. NewKreuetterBuch_1539_Bock.....	28

2.33.	ContrafaytKreuterbuch_1532_Brunfels.....	28
2.34.	ArzneiDerKreutter_1532_Tallat.....	29
2.35.	GartDerGesundheit_1487_vonCuba.....	30
2.36.	BuchDerNatur_1482_vonMegenberg.....	30
3.	Annotationsebenen – Transkription/Normalisierung.....	31
3.1.	dipl.....	31
3.2.	clean.....	33
3.3.	norm.....	34
4.	Annotationsebenen – Annotationen zu linguistischen Eigenschaften	37
4.1.	pos.....	37
4.2.	pos_klein.....	39
4.3.	lemma	42
4.4.	foreign.....	44
4.5.	foreign_trans	46
4.6.	lang	48
4.7.	Verbposition	50
4.8.	Nebensatztyp	52
4.9.	KOUS_Semantik	54
4.10.	komp	56
4.11.	komp_orth.....	58
4.12.	prot.....	60
4.13.	attr_gen	63
4.14.	strD.....	65
4.15.	bemerkung_lexik.....	67
4.16.	erlaeuterung.....	68
4.17.	referenz.....	70
4.18.	form_krankheit.....	71
4.19.	problem	73
4.20.	kraeutername_normiert	74
4.21.	kraeuterzubereitung.....	76
4.22.	form_zubereitung	78
4.23.	nomen_nominativ	79
4.24.	form_nomen.....	81
4.25.	satztyp	83
4.26.	position_im_satz	86

4.27.	position_zur_bezugskategorie.....	87
4.28.	form_bezugskategorie.....	89
4.29.	verbstellung.....	91
4.30.	form_des_relativpronomens.....	92
4.31.	modifikation_bezugskategorie.....	94
4.32.	diachronie.....	96
4.33.	dialekt.....	97
5.	Annotationsebenen – Strukturelle Annotation.....	99
5.1.	lb.....	99
5.2.	brace.....	101
5.3.	brace_dir.....	103
5.4.	p.....	105
5.5.	p_n.....	107
5.6.	p_rend.....	108
5.7.	pb.....	110
5.8.	pb_n.....	112
5.9.	pb_rend.....	113
5.10.	pb_ana.....	116
5.11.	div1 – div5.....	118
5.12.	div1_type – div5_type.....	120
5.13.	div1_n – div5_n.....	123
5.14.	unclear.....	125
5.15.	atLeast.....	126
5.16.	atMost.....	128
5.17.	interpretation.....	130
5.18.	figure.....	132
5.19.	figure_rend.....	134
5.20.	hi.....	136
5.21.	typeface.....	138
5.22.	hi_rend.....	140
5.23.	head.....	142
5.24.	head_n.....	143
5.25.	head_rend.....	145
5.26.	note.....	147
5.27.	ref.....	149

5.28.	ref_target	151
5.29.	ref_type	153
5.30.	quote.....	155
5.31.	list.....	157
5.32.	list_type.....	159
5.33.	item	160
5.34.	xml_id	162
6.	Annotationsebenen – Inhaltliche Annotation.....	164
6.1.	definition.....	164
6.2.	disease.....	166
6.3.	term.....	168
6.4.	personenname	170
6.5.	werkname.....	171
6.6.	author_ref.....	173
6.7.	reader_ref.....	174
6.8.	plant	176
6.9.	property.....	178
6.10.	name	179
6.11.	name_type	181
6.12.	citation.....	183
7.	Metadaten.....	185
8.	Referenzen	188
1.	Anhang.....	189

1. Korpus

Typ: Corpus – Ridges Herbology

Name:	Ridges Herbology
Herausgeber:	Anke Lüdeling, Carolin Odebrecht, Amir Zeldes, Korpuslinguistik, Humboldt-Universität zu Berlin
Annotatoren:	Ilham Abed-Ali, Silke Andresen, Henriette Ast, Malte Belz, Cornelia Binnyus, Doreen Christen, Janin Czilwik, Mascha Dayal, Antonia Dittberner, Cora Döhn, Imke Driemel, Christian Ebert, Olja Efremova, Gill-Maria Eichhorn, Judith Esser, Annegret Gerlach, Linda Giesel, Hagen Hirschmann, Nikolaos Kartalis, Sebastian Kiraga, Ewa Anna Kolbik, Kornél Kovács, Daisy Krüger, Anna-Maria Lehmann, Eliese-Sophia Lincke, Maria Lober, Anke Lüdeling, Laura Lueders, Samuele Maniscalco, Maria Martynova, Kim Kristin Maser, Manuel Metzsig, Alexander Meyer, Sandra Müller, Vinzent Müller, Andrew Murphy, Johannes Mursell, Carolin Odebrecht, Akiko Okuda, Mareike Otto, Angelo Papenhoff, Laura Perlitz, Valerie Reichardt, Katharina Reinig, Ina Riesler, Lena Rosin, Franz-Josef Sachse, Anna Sapronova, Simon Sauer, Claudia Schmidt, Gohar Schnelle, Lucia Sohmen, Iryna Sorokovska, Uwe Springmann, Kristina Stephan, Juliane Tiemann, Anna Tóth, Réka Tóth, Alexander Turtureanu, Magdalena Vock, Vivian Voigt, Juliana Wekel, Karolina Zuchewicz (nachfolgend Studierende der Humboldt-Universität zu Berlin genannt)
Veröffentlichung:	07.06.2016
Anzahl der Texte:	36
Zeitraum:	1482-1914
diplToken:	182738
Sprache:	Deutsch
Register:	Kräutertexte
Zugang:	https://korpling.german.hu-berlin.de/annis3/
Projekt:	http://korpling.german.hu-Berlin.de/ridges/index_de.html
Projektbeschreibung:	Das RIDGES -Projekt (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.

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

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

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

Version: 5.0

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

Art der Änderung: Folgende Dokumente wurden hinzugefügt:
BuchDerNatur_1482_vonMegenberg
NewKreuetterBuch_1539_Bock
NewKreueterbuch_1563_Handsch
Phythologia_1662_Becher
TheatrumBotanicum_1696_Verzascha
ViridariumReformatum_1719_Valentini

Kraeuterbuch_1914_Losch

Neue Annotationen wurden hinzugefügt und zwar folgende:
bemerkung_lexik, citation, diachronie, dialekt, form_bezugskategorie,
form_des_relativpronomens, form_krankheit, form_nomen,
form_zubereitung, kraeutername_normiert, kraeuterzubereitung,
modifikation_bezugskategorie, nomen_nominativ, personennamen,
position_im_satz, position_zur_bezugskategorie, problem, referenz,
satztyp, verbstellung, werkname.

Komplette Überarbeitung der Metadaten inklusive Hinzufügen von
neuen Metadaten.

Automatisch erneuert wurden folgende Annotationen für alle Dokumente
im Korpus: clean, pos, lemma.

Die Ebene „hyperlemma“ wurde in allen Dokumenten in
„erläuterung“ umbenannt und ist ab sofort eine unsystematische Ebene.

Die Ebene „typeface“ wurde in allen Dokumenten in
„typeface“ umbenannt.

Ersetzung folgender Sonderzeichen in der dipl-Ebene durch das Skript
„normalizeDipl.sh“: $\bar{m} \rightarrow \tilde{m}$, $\bar{n} \rightarrow \tilde{n}$, $\bar{a} \rightarrow \tilde{a}$, $\bar{e} \rightarrow \tilde{e}$, $\epsilon \rightarrow \delta$, $\acute{u} \rightarrow \ddot{u}$, $\acute{i} \rightarrow \text{i}$

Korrektur der dipl- und der norm-Ebene in allen neuen Texten (s.o.) und
allen Texten der 4.1 Version von 1487 bis einschließlich 1652 durch
verschiedene Mitarbeiter der korpling-Arbeitsgruppe (LAUDATIO +
LangBank).

In den folgenden Dokumenten wurden die Deckblätter entfernt:

SonderbaresKraeuterbuch-1-11_1675_Anonymous
NochEinigeWorte_1840_Meyer

Punktuelle manuelle Korrektur der lb/pb-Ebene in:

MysteriumSigillorum_1735_Hiebner
Unterricht_1774_Eisen
FloraSaturnizans_1722_Henckel
SonderbaresKraeuterbuch-11-21_1675_Anonymus
Flora-6_1821_Wilbrand
EigenschaftenAllerHeilpflanzen_1828_Anonymous
DeutschePflanzennamen_1870_Grassmann
VorlesungenUeberKraeuterkunde_1843_Link

Vollständige manuelle Korrektur der lb/pb-Ebene sowie Korrektur der norm-Ebene (Segmentierungsfehler) in:

FloraDerPreussischenRheinlande_1870_Wirtgen

Manuelle Korrektur der Annotationsebenen „hi“, „typeface“ und „hi_rend“ in PflantzGart-c4_1639_Rhagor

Systematische, semi-automatische Überprüfung der Segmentierungsbasen aller Annotationen

Enthaltende (Kürzel):	Dokumente	BuchDerNatur_1482_vonMegenberg GartDerGesundheit_1487_vonCuba ArtzneyBuchleinDerKreutter_1532_Tallat ContrafaytKreuterbuch_1532_Brunfels NewKreuetterBuch_1539_Bock NewKreuterbuch_1543_Fuchs WieSichMeniglich_1557_vonBodenstein NewKreuetterbuch_1563_Handsch Paradeiszgaertlein_1588_Rosbach AlchymistischePractic_1603_Libavius HortulusSanitatis_1609_Uffenbach Kraeutterbuch_1609_Carrichter PflantzGart-c4_1639_Rhagor PflantzGart_1639_Rhagor PflantzGart-Vorrede_1639_Rhagor Wund-Artzney_1652_Greiff Phythologia_1662_Becher ThesaurusSanitatis_1673_Nasser SonderbaresKraeuterbuch-1-11_1675_Anonymous SonderbaresKraeuterbuch-11-21_1675_Anonymous SchweizerischerBotanicus_1687_vonRoll TheatrumBotanicum_1696_Verzascha ViridariumReformatum_1719_Valentini FloraSaturnizans_1722_Henckel MysterivmSigillorvm_1735_Hiebner EinleitungZuDerKraeuterkenntnisz_1764_Oeder Unterricht_1774_Eisen GrundrissKraeuterkunde_1792_Willdenow Flora-6_1821_Wilbrand Flora-7_1821_Sieber EigenschaftenAllerHeilpflanzen_1828_Anonymous
--------------------------	-----------	--

NochEinigeWorte_1840_Meyen
VorlesungenUeberKraeuterkunde_1843_Link
DeutschePflanzennamen_1870_Grassmann
FloraDerPreussischenRheinlande_1870_Wirtgen
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, clean, dipl, erlaeuterung, figure, hi_rend, lang, lb, lemma, norm, note, pb, pb_n, personename, pos, reader_ref, typeface, werkname

2.2. 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: 7539
 Beschreibung: Kräutertext
 Annotationsebenen: attr_gen, author_ref, clean, dipl, erlaeuterung, figure, foreign, head, hi, hi_rend, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, plant, pos, pos_klein, prot, quote, ref, ref_target, strD, typeface, Verbposition, xml_id

2.3. 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: 10283
 Beschreibung: Kräutertext
 Annotationsebenen: atLeast, atMost, attr_gen, author_ref, clean, definition, dipl, div1, div1_type, div2, div2_type, erlaeuterung, foreign, foreign_trans, head, head_n, hi, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, name, name_type, Nebensatztyp, norm, p, pb, pb_n, pos, pos_klein, property, prot, strD, term, typeface, unclear, Verbposition

2.4. 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, clean, dipl, erlaeuterung, foreign, foreign_trans, head, hi, hi_rend, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, plant, pos, pos_klein, prot, reader_ref, ref, ref_target, strD, typeface, Verbposition, xml_id

2.5. NochEinigeWorte_1840_Meyen

Titel: Noch einige Worte über den Befruchtungsakt und die Poljembryonie bei den höheren Pflanzen
 Kürzel: NochEinigeWorte_1840_Meyen
 Autor: Meyen, Franz Julius Ferdinand
 Jahr: 1840
 Ort: Berlin
 Verlag: Haude und Spenersche Buchhandlung (S. J. Joseephy.)
 Seitenangabe: 24-50
 Publikationstyp: Monographie
 dipl-Einheiten: 7984
 Beschreibung: Kräutertext
 Annotierte Ebenen: attr_gen, author_ref, clean, dipl, div2, div2_type, foreign, foreign_trans, head, hi, hi_rend, komp, komp_orth, KOUS_Semantik, lb, lemma, name, name_type, Nebensatztyp, norm, note, p, p_rend, pb, pb_n, pos, pos_klein, property, prot, quote, reader_ref, ref, ref_target, ref_type, strD, term, typeface, Verbposition, xml_id

2.6. **EigenschaftenAllerHeilpflanzen_1828_Anonymous**

Titel:	Die Eigenschaften Aller Heilpflanzen
Kürzel	EigenschaftenAllerHeilpflanzen_1828_Anonymous
Autor:	NA
Jahr:	1828
Ort:	München
Verlag:	NA
Seitenangabe:	3-21
dipl-Einheiten:	6849
Beschreibung:	Kräutertext
Annotierte Ebenen:	atLeast, atMost, attr_gen, clean, dipl, disease, erlaeuterung, foreign, head, hi, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, p, pb, pb_n, plant, pos, pos_klein, prot, strD, typeface, unclear, Verbposition

2.7. **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, clean, dipl, div1, div1_type, div2, div2_type, div3, div3_type, div4, div4_type, div5, foreign,

foreign_trans, head, head_n, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, p, pb, pb_n, pos, pos_klein, property, prot, strD, term, typeface, unclear, Verbposition

2.8. 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, brace, brace_dir, clean, definition, dipl, div1, div1_type, div2, div2_type, foreign, foreign_trans, head, head_n, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, pos, pos_klein, property, prot, ref, ref_target, ref_type, term, typeface, unclear, Verbposition, xml_id

2.9. 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, clean, dipl, div1, div1_type, div2, div2_n, div2_type, div3, div3_n, foreign, foreign_trans, head, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, p, pb, pb_n, pos, pos_klein, property, prot, strD, term, typeface, unclear, Verbposition

2.10. 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, clean, definition, dipl, div1, div1_type, div2, div2_type, div3, erlaeuterung, foreign, foreign_trans, head, head_n, hi, interpretation, item, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, list, list_type, Nebensatztyp, norm, p, pb, pb_n, pos, pos_klein, property, prot, reader_ref, strD, term, typeface, unclear, Verbposition

2.11. 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, clean, dipl, erlaeuterung, foreign, head, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, p, pb, pb_n, plant, pos, pos_klein, prot, reader_ref, strD, typeface, unclear, Verbposition

2.12. 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, clean, dipl, disease, erlaeuterung, foreign, foreign_trans, head, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, plant, pos, pos_klein, quote, reader_ref, ref, ref_target, strD, unclear, Verbposition, xml_id

2.13. 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, clean, definition, dipl, div1, div1_type, div2, div2_type, div3, div3_type, foreign, head, head_n, hi, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, pos, pos_klein, property, prot, ref, ref_target, ref_type, strD, term, typeface, unclear, Verbposition, xml_id

2.14. 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, clean, dipl, erlaeuterung, figure, head, hi_rend, lang, lb, lemma, norm, pb, pb_n, personenname, pos, typeface, werkname

2.15. 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 nützlich und ergetzlich

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

2.16. SchweizerischerBotanicus_1687_vonRoll

Titel: Der Schweizerische Botanicus
Kürzel: SchweizerischerBotanicus_1687_vonRoll

Autor: Timotheus von Roll
 Jahr: 1687
 Ort: NA
 Verlag: Franz Carl Roöß
 Seitenangabe: 54-79
 Publikationstyp: Monographie
 dipl-Einheiten: 3510
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, clean, dipl, head, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lb, lemma, Nebensatztyp, norm, pb, pb_n, pos, pos_klein, prot, reader_ref, strD, typeface, unclear, Verbposition
 Anmerkung: Aus der Originaltranskription der Studierende wurden 2 Seiten gelöscht, diese waren im der Transkription zugrunde liegenden PDF doppelt vorhanden und wurden von den Studierende auch doppelt annotiert.

2.17. SonderbaresKrauterbuch-11-21_1675_Anonymous

Titel: Curioser Botanicus oder sonderbares Kräuter Buch
 Kürzel: SonderbaresKrauterbuch-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, clean, dipl, div1, div1_type, div2, div3, div3_type, erlaeuterung, foreign, head, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, p, p_n, pb, pb_n, pos, pos_klein, property, prot, strD, term, typeface, unclear, Verbposition

2.18. 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, clean, dipl, div2, div2_type, erlaeuterung, foreign, foreign_trans, head, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, p, p_n, pb, pb_ana, pb_n, pos, pos_klein, property, prot, strD, term, typeface, unclear, Verbposition

2.19. 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, clean, dipl, disease, erlaeuterung, foreign, foreign_trans, head, hi, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, p, pb, pb_n, plant, pos, pos_klein, prot, reader_ref, strD, typeface, unclear, Verbposition

2.20. 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: clean, dipl, erlaeuterung, figure, head, hi_rend, lang, lb, lemma, norm, pb, pb_n, pos, reader_ref, typeface, werkname

2.21. 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, clean, dipl, disease, erlaeuterung, foreign, head, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, plant, pos, pos_klein, prot, reader_ref, referenz, strD, typeface, unclear, Verbposition

2.22. PflantzGart-Vorrede_1639_Rhagor

Titel: Pflantz-Gart (Vorrede)
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, clean, dialekt, dipl, div1, div2, div2_type, erlaeuterung, foreign, foreign_trans, head, hi, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, pb_rend, pos, pos_klein, prot, reader_ref, referenz, term, typeface, Verbposition

2.23. 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, citation, clean, dialekt, dipl, div1, div1_type, div2, div2_type, div3, div3_type, div4, div5, erlaeuterung, foreign, foreign_trans, head, head_n, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, p_rend, pb, pb_n, pos, pos_klein, property, prot, reader_ref, referenz, strD, term, typeface, unclear, Verbposition

2.24. 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: 4066
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, clean, dialekt, dipl, erlaeuterung, figure, head, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb,

lemma, Nebensatztyp, norm, note, pb, pb_n, pos, pos_klein, prot, referenz, strD, typeface, unclear, Verbposition

2.25. 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, clean, dipl, erlaeuterung, head, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, pb, pb_n, pos, pos_klein, prot, reader_ref, referenz, strD, term, typeface, unclear, Verbposition

2.26. HortulusSanitatis_1609_Uffenbach

Titel: Hortulus Sanitatis

Kürzel: HortulusSanitatis_1609_Uffenbach

Autor: Peter Uffenbach

Jahr: 1609

Ort: Frankfurt am Main

Verlag: Ionaë Rhodij

Seitenangabe: 1-21

Publikationstyp: Monographie

dipl-Einheiten: 6516
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, clean, diachronie, dialekt, dipl, disease, erlaeuterung, foreign, foreign_trans, head, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, plant, pos, pos_klein, prot, referenz, strD, typeface, unclear, Verbposition

2.27. AlchymistischePractic_1603_Libavius

Titel: Alchymistische Practic
Kürzel: AlchymistischePractic_1603_Libavius
Autor: Andreas Libavius
Jahr: 1603
Ort: Frankfurt
Verlag: Johann Saur
Seitenangabe: 4-26
Publikationstyp: Monographie
dipl-Einheiten: 5504
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, clean, definition, diachronie, dipl, div2, div2_n, div2_type, div3, div3_n, div3_type, div4, div4_type, erlaeuterung, figure, figure_rend, foreign, foreign_trans, head, head_n, hi, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebensatztyp, norm, note, p, pb, pb_n, pb_rend, pos, pos_klein, property, prot, reader_ref, referenz, strD, term, typeface, unclear, Verbposition

2.28. Paradeiszaertlein_1588_Rosbach

Titel: Paradeiszgärtlein
Kürzel: Paradeiszaertlein_1588_Rosbach

Autor: Conrad Rosbach
 Jahr: 1588
 Ort: Frankfurt am Main
 Verlag: Johann Spieß
 Seitenangabe: 1-43
 Publikationstyp: Monographie
 dipl-Einheiten: 5055
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, citation, clean, diachronie, dialekt, dipl, erlaeuterung, figure, head, typeface, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, lang, lb, lemma, Nebenstaztyp, norm, pb, pb_n, pos, pos_klein, prot, reader_ref, referenz, strD, term, typeface, unclear, Verbposition

2.29. NewKreueterbuch_1563_Handsch

Titel: New Kreüterbuch: Mit den allerschönsten vnd artlichsten Figuren aller Gewechß, dergleichen vormals in keiner sprach nie an tag kommen
 Kürzel: NewKreueterbuch_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, clean, dipl, erlaeuterung, figure, head, hi_rend, lang, lb, lemma, norm, pb, pb_n, personennamen, pos, reader_ref, referenz, typeface

2.30. 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:	5522
Beschreibung:	Kräutertext
Annotierte Ebenen:	atLeast, atMost, attr_gen, author_ref, clean, dipl, disease, erlaeuterung, foreign, foreign_trans, head, interpretation, komp, komp_orth, KOUS_Semantik, kraeuternormiert, lang, lb, lemma, Nebensatztyp, norm, p, pb, pb_n, personennamen, plant, pos, pos_klein, prot, reader_ref, strD, typeface, unclear, Verbposition

2.31. NewKreueterbuch_1543_Fuchs

Titel:	New Kreüterbuch
Kürzel:	NewKreueterbuch_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, clean, definition, dipl, div2, div2_type, div3, div3_type, erlaeuterung, figure, figure_rend, foreign, foreign_trans, head, komp, komp_orth, KOUS_Semantik, kraeuternormiert, lang, lb, lemma, Nebensatztyp, norm, p, p_n, p_rend, pb, pb_n, pb_rend, personennormiert, pos, pos_klein, property, prot, strD, term, typeface, Verbposition

2.32. 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, clean, dipl, erlaeuterung, head, hi_rend, lang, lb, lemma, norm, note, pb, pb_n, personennormiert, pos, reader_ref, referenz, typeface, unclear, werkname

2.33. 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, clean, dipl, erlaeuterung, figure, head, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, kraeuternamen_normiert, lang, lb, lemma, modifikation_bezugskategorie, Nebensatztyp, norm, pb, pb_n, personennamen, pos, pos_klein, prot, reader_ref, term, typeface, unclear, Verbposition

2.34. ArtzneiDerKreutter_1532_Tallat

Titel: Artzney Buchlein der kreutter
 Kürzel: ArtzneiDerKreutter_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, bemerkung_lexik, clean, dipl, erlaeuterung, form_bezugskategorie, form_des_relativpronomens, form_krankheit, form_nomen, form_zubereitung, head, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, kraeuternamen_normiert, kraeuterzubereitung, lang, lb, lemma, modifikation_bezugskategorie, Nebensatztyp, nomen_nominativ, norm, note, pb, personennamen, pos, pos_klein, position_im_satz,

position_zur_bezugskategorie, problem, prot, reader_ref, satztyp, term, typeface, unclear, Verbposition, verbstellung, werkname

2.35. 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: 4702
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, attr_gen, author_ref, bemerkung_lexik, clean, dipl, erlaeuterung, figure, form_bezugskategorie, form_des_relativpronomens, form_krankheit, form_nomen, form_zubereitung, head, hi_rend, interpretation, komp, komp_orth, KOUS_Semantik, kraeuternormiert, kraeuterzubereitung, lang, lb, lemma, modifikation_bezugskategorie, Nebensatztyp, nomen_nominativ, norm, pb, personennormiert, pos, pos_klein, position_im_satz, position_zur_bezugskategorie, problem, prot, reader_ref, satztyp, term, typeface, unclear, Verbposition, verbstellung, werkname

2.36. 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, clean, dipl, erlaeuterung, head, hi_rend, lang, lb, lemma, norm, pb, personennamen, pos, reader_ref, referenz, typeface

3. Annotationsebenen – Transkription/Normalisierung

3.1. dipl

Typ: *Layer* - dipl

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

Typ: Preparationstep – dipl

Schritt:	1
Aufbereitung:	OCR/Transkript
Aufbereitungsart:	semi-automatisch/manuell
Tool:	NA
Format:	Text-Datei
Segmentierung der Annotation:	eigenständig

Qualitätsprüfer: Korpling
Datum: 2015
Qualitätsprüfung: manuell
Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Tokenisierung
Aufbereitungsart: Automatisch
Tool: Treetagger Version 3.2
Format: Treetagger-Output

Segmentierung der
Annotation: eigenständig
Qualitätsprüfer: Studierende der Humboldt-Universität zu Berlin, korpling
Datum: 2015
Qualitätsprüfung: manuelle Konsistenzprüfung
Editor: Laura Perlitz

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

3.2. clean

Typ: *Layer* – clean

Name:	clean
Dokumentation:	Siehe Ridges-Website http://korpling.german.hu-berlin.de/ridges/index_de.html Beispiele: Siehe Anhang
clean-Einheiten:	177795
Beschreibung:	<p>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: <i>aufzwēdig</i> wird zu <i>auszwemdig auszwendig</i>).</p> <p>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 (<i>Gelb-Sucht</i> wird zu <i>Gelbsucht</i>; <i>MON- <lb> TANUM</i> wird zu <i>MONTANUM</i>).</p> <p>Trunkierte Elemente, die am Zeilenende stehen, werden bisher nicht als solche erkannt und daher automatisch mit dem ersten Element der folgenden Zeile zusammengezogen (<i>Speiſzund Nahrungsſaffts</i> wird zu <i>Speiszund (sic!)Nahrungssaffts</i>).</p> <p>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.</p>

Typ: Preparationstep – clean

Schritt: 1
Aufbereitung: Normalisierung, Annotation
Aufbereitungsart: Automatisch
Tool: cleanV22.py
Format: Clean-Skript-Output
Segmentierung der Annotation: Eigenständig
Qualitätsprüfer: Laura Perlitz
Datum: 2016
Qualitätsprüfung: manuelle Konsistenzprüfung
Editor: Laura Perlitz

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

3.3. norm

Typ: *Layer* – norm

Name: Norm
Dokumentation: Siehe Ridges-Website

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

Beispiele: Siehe Anhang

norm-Einheiten: 176855

Beschreibung: Die norm-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier Tokenisierung, Graphematik, Flexionsformen und Lexeme an den nhd. Sprachstand angepasst werden.

Graphematik:

orthographische Anpassung nach Duden

kreutter → Kräuter

Phonologie:

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

wänlîn → Wännlein

lehret → lehrt

Morphologie:

in die Nasen → in die Nase

das kal Haupt → das kahle Haupt

funden → gefunden

Lexikologie:

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

Vergefz → Vergess 'das Vergessen'

Wortbildung:

Ausgestorbene Wortbildungsmorpheme werden nach Möglichkeit durch nhd. Bildungen ersetzt.

halben → halber

stachelecht → stachelig

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

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

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

Kasus wurde normalisiert in folgenden Dokumenten:

Kraeuterbuch_1914_Losch

Phythologia_1662_Becher
TheatrumBotanicum_1696_Verzascha
ViridariumReformatum_1719_Valentini
BuchDerNatur_1482_vonMegenberg
NewKreueterbuch_1563_Handsch
NewKreuetterBuch_1539_Bock
ArzneiDerKreuter_1532_Tallat
ContrafaytKreuterbuch_1532_Brunfels
GartDerGesundheit_1487_vonCuba
NewKraeuterbuch_1543_Fuchs
WieSichMeniglich_1557_vonBodenstein

Kasus wurde nicht normalisiert in folgenden Dokumenten:

AlchymistischePractic_1603_Libavius
HortulusSanitatis_1609_Uffenbach
Kraeutterbuch_1609_Carrichter
Paradeiszaertlein_1588_Rosbach
PflantzGart_1639_Rhagor
PflantzGartVorrede_1639_Rhagor
PflantzGartc4_1639_Rhagor
WundArtzney_1652_Hildanus
DerSchweizerischeBotanikus_1687_vonRoll
DeutschePflanzennamen_1870_Grassmann
EigenschaftenAllerHeilpflanzen_1828_Anonymous
EinleitungZuDerKraeuterkenntnis_1764_Oeder
Flora6_1821_Willbrand
Flora7_1821_Sieber
FloraDerPreussischenRheinlande_1870_Wirtgen
FloraSaturnizans_1722_Henckel
GrundrissKraeuterkunde_1792_Willdenow
MysteriumSigillorvm_1735_Hiebner
NochEinigeWorte_1840_Meyen
SonderbaresKraeuterbuch-1-11_1675_Anonymous
SonderbaresKraeuterbuch-11-21_1675_Anonymous
ThesaurusSanitatis_1673_Nasser
Unterricht_1774_Eisen
VorlesungenUeberKraeuterkunde_1843_Link

Beispiele und Einzelentscheidungen siehe Anhang.

Schritt: 1
Aufbereitung: Normalisierung, Annotation
Aufbereitungsart: Manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation:
Eigenständig
Qualitätsprüfer: Korpling
Datum: 2015
Qualitätsprüfung: manuelle Korrektur
Editor: Studierende der Humboldt-Universität zu Berlin

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

4. Annotationsebenen – Annotationen zu linguistischen Eigenschaften

4.1. pos

Typ: *Layer* – pos

Name: Pos

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

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

Typ: Preparationstep – pos

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

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

Schritt: 3
Aufbereitung: Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart: Automatisch

Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	Korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: Annotationlayer – pos

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

Typ: Annotationvalue – pos

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

4.2. pos_klein

Typ: Layer – pos_klein

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

Typ: Preparationstep – pos_klein

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

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

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

Qualitätsprüfer: korpling
 Datum: 2016
 Qualitätsprüfung: NA
 Editor: Vivian Voigt

Typ: Annotationlayer – pos_klein

Annotationsart: Spannenannotation
 Beschreibung: Wortarten-Annotation teilweise nach dem Stuttgart-Tübingen-Tagset. Manche Tags aus dem STTS wurden zu neuen Tags verschmolzen.

Typ: Annotationvalue – pos_klein

Wert:	Wertbeschreibung:
ADJ	Adjektiv, verschmolzen aus den Tags ADJA und ADJD (aus STTS)
ADV	Adverb, verschmolzen aus den Tags ADV und PAV (aus STTS)
APPR	Präposition oder linke Zirkumposition, übereinstimmend mit STTS
APPRART	Präposition mit Artikel, übereinstimmend mit STTS
APPO	Postposition, übereinstimmend mit STTS
APZR	Rechte Zirkumposition, übereinstimmend mit STTS
ART	Artikel, verschmolzen aus den Tags ART, PIAT, PDAT, PIDAT und PPOSAT (aus STTS)
CARD	Kardinalzahl, übereinstimmend mit STTS
FM	Fremdsprachliches Material, übereinstimmend mit STTS
ITJ	Interjektion, übereinstimmend mit STTS
KOUI	Subordinierende Konjunktion mit 'zu' und Infinitiv, übereinstimmend mit STTS
KOKOM	Vergleichende Partikel ohne Satz, übereinstimmend mit STTS
KOORD	Koordinierende Konjunktion, äquivalent zu dem STTS-Tag KON
N	Nomen, verschmolzen aus den Tags NN, PDS, PPER, PIS, PPOSS, PWS und NE (aus STTS)

REL	Relativpronomen, verschmolzen aus den Tags PRELS und PRELAT (aus STTS)
PRF	Reflexives Personalpronomen, übereinstimmend mit STTS
PWAT	Attributives Interrogativpronomen; alternativ: für Interrogativpronomen, die in einer NP vorkommen; übereinstimmend mit STTS
PWAV	Adverbiales Interrogativ- oder Relativpronomen; übereinstimmend mit STTS
PAV	Pronominaladverb, übereinstimmend mit STTS
PTKVZ	Separierte Verbpartikel, übereinstimmend mit STTS
PTKZU	'zu' vor einem Infinitiv, übereinstimmend mit STTS
PTKNG	Negationspartikel, übereinstimmend mit STTS
PTKANT	Antwortpartikel, übereinstimmend mit STTS
PTKA	Partikel mit Adjektiv oder Adverb, übereinstimmend mit STTS
SUB	Subordinierende Konjunktion mit Satz, äquivalent mit dem TAG KOUS (aus STTS)
TRUNC	Trunkiertes Wort - erster Teil, übereinstimmend mit STTS
VFIN	Finites Verb, verschmolzen aus den Tags VVFIN, VMFIN und VAFIN (aus STTS)
VIMP	Imperativ, verschmolzen aus den Tags VVIMP, VMIMP, VAIMP (aus STTS)
XY	Nicht-Wort, das Sonderzeichen enthält; übereinstimmend mit STTS
ZEICHEN	Interpunktion, verschmolzen aus den Tags \$., \$, und \$ (aus STTS)

4.3. lemma

Typ: *Layer* – lemma

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

zusammengefasst.

Typ: Preparationstep – lemma

Schritt:	1
Aufbereitung:	Lemmatisierung
Aufbereitungsart:	Automatisch
Tool:	Treetagger Version 3.2
Format:	Treetagge-Output
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: Annotationlayer– lemma

Annotationsart:	Spannenannotation
Beschreibung:	Lemmatisierung: Normalisierung der flektierten Wortformen,

Zuordnung der entsprechenden Lexikoneinträge zu jeder Wortform.

Typ: Annotationvalue – lemma

Wert: Wertbeschreibung:
Normalisiertes Lemma Lemmatisierung.

4.4. foreign

Typ: Layer – foreign

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

Typ: Preparationstep – foreign

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

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

Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der
Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

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

Typ:*Annotationlayer* – foreign

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

Typ:*Annotationvalue* – foreign

Wert: Wertbeschreibung

foreign

Allgemeiner Tag für fremdsprachliches Material.

4.5. foreign_trans

Typ: *Layer* – foreign_trans

Name: foreign_trans

Dokumentation: Siehe Anhang

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

Typ: *Preparationstep* – foreign_trans

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: LAUDATIO

Datum: 2013

Qualitätsprüfung: Manuelle Konsistenzprüfung: Die Tags aus der Annotationsebene **definition**, die eine Markierung von Übersetzungen darstellen, wurden unter dieser Annotationsebene zusammengefasst.

Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der
Annotation:

basiert auf 'dipl'

Qualitätsprüfer:

korpling

Datum:

2016

Qualitätsprüfung:

NA

Editor:

Carolin Odebrecht

Schritt:

3

Aufbereitung:

Konvertierung von Excel 2013 nach PAULA-XML und ANNIS

Aufbereitungsart:

automatisch

Tool:

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

Format:

PAULA-XML, ANNIS

Segmentierung der

Annotation:

basiert auf 'dipl'

Qualitätsprüfer:

korpling

Datum:

2016

Qualitätsprüfung:

NA

Editor:

Vivian Voigt

Typ: Annotationlayer – foreign_trans

Annotationsart:

Spannenannotation

Beschreibung:

Beschreibung der Zielsprache und der Ausgangssprache einer Übersetzung.

Typ: *Annotationvalue* – foreign_trans

Wert:

Wertbeschreibung:

trans_to_german

Übersetzung eines Wortes in Deutsch.

trans_from_german

Übersetzung eines Wortes vom Deutschen.

trans_to_german_extended	Übersetzung einer Phrase in Deutsch.
trans_from_german_extended	Übersetzung einer Phrase vom Deutschen.

4.6. lang

Typ: *Layer* – lang

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

Typ: *Preparationstep* – lang

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	NA
Datum	2016
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der	

Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: Annotationlayer – lang

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

Typ: Annotationvalue – lang

Wert:	Wertbeschreibung:
Kürzel der entsprechenden Sprache	Getagged nach ISO Drei-Buchstaben-Code ISO 639-3 (http://www.sil.org/iso639-3/codes.asp?order=639_3&letter=%25)
lat	Latein
grc	Altgriechisch
spa	Spanisch
ita	Italienisch
ara	Arabisch
fra	Französisch
pol	Polnisch
nld	Niederländisch
eng	Englisch
rus	Russisch
goh	Althochdeutsch
dan	Dänisch

4.7. Verbposition

Typ: *Layer* – Verbposition

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

Typ: Preparationstep – Verbposition

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

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

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

Typ: Annotationlayer – Verbposition

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

Typ: Annotationvalue – Verbposition

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

	Satz mit subordinierender Konjunktion; analysiert als Token-Feature auf der Basis von pos=KOUS
V1	Verberststellung in einem gegebenen untergeordneten Satz mit subordinierender Konjunktion; analysiert als Token-Feature auf der Basis von pos=KOUS

4.8. Nebensatztyp

Typ: *Layer* – Nebensatztyp

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

Typ: Preparationstep – Nebensatztyp

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

Tool: Excel 2013
Format: Excel 2013
Segmentierung der
Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

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

Typ: Annotationlayer – Nebensatztyp

Annotationsart: Spannenannotation
Beschreibung: Annotation der Nebensatztypen.

Typ: Annotationvalue – Nebensatztyp

Wert: Wertbeschreibung:
Adverbial Adverbiale Funktion eines untergeordneten Satzes.

Attribut	Attributive Funktion eines untergeordneten Satzes.
Komplement	Komplementfunktion eines untergeordneten Satzes.

4.9. KOUS_Semantik

Typ: *Layer* – KOUS_Semantik

Name: KOUS_Semantik

Dokumentation: NA

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

Typ: Preparationstep – KOUS_Semantik

Schritt: 1

Aufbereitung: Linguistische Annotation

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

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: Hagen Hirschmann

Datum: 2014

Qualitätsprüfung: manuelle Korrektur

Editor: Hagen Hirschmann

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013
Segmentierung der
Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Carolin Odebrecht

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

Typ: Annotationlayer – KOUS_Semantik

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

Typ: Annotationvalue – KOUS_Semantik

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

4.10.komp

Typ: *Layer* – komp

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

Typ: Preparationstep – komp

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010

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

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

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

Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: Annotationlayer – komp

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

Typ: Annotationvalue – komp

Wert: Wertbeschreibung:
k Kompositum mit substantivischem Zweitglied

4.11.komp_orth

Typ: *Layer* – komp_orth

Name: komp_orth
Dokumentation: Ist ein Kompositum an einer Morphemgrenze durch einen Zeilenumbruch unterbrochen, wurde es entweder mit "lb1" oder "lb2" markiert, je nachdem, ob ein Bindestrich vorhanden ist oder nicht. Ist dies nicht der Fall und das Kompositum wurde innerhalb einer Morphemgrenze unterbrochen, wird es als zusammen geschrieben markiert, da davon ausgegangen wird, dass der Autor das Kompositum nicht getrennt geschrieben hätte, wenn an dieser Stelle kein Zeilenumbruch vorkäme. Mit "bs" wurden nur diejenigen Komposita markiert, die innerhalb einer Zeile mit einem Bindestrich versehen sind.
Beschreibung: Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – komp_orth

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

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

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

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

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – komp_orth

Annotationsart: Spannenannotation

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

Typ: Annotationvalue – komp_orth

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

4.12.prot

Typ: Layer – prot

Name: prot

Dokumentation: Der Prototyp 1 (prot1) markiert alle Token, die eindeutig als

Komposita mit substantivischem Zweitglied zu identifizieren sind. Folgende Kriterien müssen dabei erfüllt sein: Es liegen mindestens zwei freie Morpheme vor, die entweder zusammen oder mit Bindestrich geschrieben wurden. Es darf wortintern keine Flexion auftreten und das Kompositum kann syntaktisch nicht erweitert werden, beispielsweise durch Adjektiv- oder Präpositionalphrasen (z.B. *stabwurz*).

Der Prototyp 2 (prot2) fasst alle Tokens zusammen, die sehr wahrscheinlich Komposita sind. Diese bestehen ebenfalls aus zwei freien Morphemen, werden aber getrennt geschrieben oder wurden durch einen Zeilenumbruch unterbrochen und enthalten keine Bindestrichmarkierung. Sie dürfen wie Komposita des Prototyp 1 wortintern nicht flektierbar und syntaktisch nicht erweiterbar sein (z.B. *wermut saft*).

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 Getrenntschreibung vor, jedoch kann man keine genaue Aussage darüber treffen, ob eine wortinterne Flexion vorgenommen wurde oder möglich wäre und ob das jeweilige Vorkommen syntaktisch erweiterbar ist (z.B. *widers Teuffels Biffz*). Die Zweifelsfälle wurden nicht in der Ebene "komp" aufgenommen.

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

Typ: Preparationstep – prot

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

Qualitätsprüfung: manuelle Korrektur

Editor: Laura Perlitz

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – prot

Annotationsart: Spannenannotation

Beschreibung: Weist jedem Wert in der Ebene Kompositum einen Prototypen zu.

Der Prototyp gibt an, wie wahrscheinlich ein Vorkommen ein Substantivkompositum ist.

Typ: Annotationvalue – prot

Wert:	Wertbeschreibung:
prot1	sicher als Kompositum zu identifizieren
prot2	sehr wahrscheinlich Kompositum
prot3	Zweifelsfall; nicht in der Ebene "komp" markiert

4.13.attr_gen

Typ: Layer – attr_gen

Name:	attr_gen
Dokumentation:	Das Bezugsnomen und das Genitivattribut wurden in ein und derselben Spanne annotiert. Das Gleiche gilt für verschachtelte sowie koordinierte und durch weitere Phrasen erweiterte Genitivkonstruktionen (z.B. <i>Beobachtungen der nach der Bestäubung in den vegetabilischen Eyern erfolgenden Veränderungen</i>).
Beschreibung:	Unter linguistischen Annotationen werden u.a. pos-Tagging, Lemmatisierung und Markierung von fremdsprachlichem Material zusammengefasst.

Typ: Preparationstep – attr_gen

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

Qualitätsprüfung: manuelle Korrektur

Editor: Laura Perlitz

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – attr_gen

Annotationsart: Spannenannotation

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

Typ: Annotationvalue – attr_gen

Wert:	Wertbeschreibung:
gprä	pränominales Genitivattribut
gpost	postnominales Genitivattribut

4.14.strD

Typ: Layer – strD

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

Typ: Preparationstep – strD

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

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: Annotationlayer – strD

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

Typ: Annotationvalue – strD

Wert: strD
Wertbeschreibung: strukturelle Destruktion

4.15.bemerkung_lexik

Typ: Layer – bemerkung

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

Typ: Preparationstep – bemerkung_lexik

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

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

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

Typ: Annotationlayer – bemerkung_lexik

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

Typ: Annotationvalue – bemerkung_lexik

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

4.16.erlaeuterung

Typ: Layer – erlaeuterung

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

Typ: Preparationstep – erlaeuterung

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: NA

Datum: 2015

Qualitätsprüfung: NA

Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt: 2

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – erlaeuterung

Annotationsart: Spannenannotation

Beschreibung: Dies ist keine systematische Ebene. Sie dient lediglich der besseren Lesbarkeit. Wörter, deren Bedeutung sich aus dem Neuhochdeutschen nicht direkt erschließt (z.B. Krankheitsnamen

oder Wörter, die keine neuhochdeutsche Entsprechung haben), werden hier erklärt. Auch Wörter, deren Form es heute noch gibt, die aber ihre Bedeutung geändert haben (*dicke* für *oft*) werden erläutert.

Typ: Annotationvalue – erlaeuterung

Wert:	Wertbeschreibung:
Zeichenkette	Erläuterungen zur Lexik.

4.17.referenz

Typ: Layer – referenz

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

Typ: Preparationstep – referenz

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

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

Typ: Annotationlayer – referenz

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

Typ: Annotationvalue – referenz

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

4.18.form_krankheit

Typ: Layer – form_krankheit

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

Typ: Preparationstep – form_krankheit

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf ‘dipl’

Qualitätsprüfer: NA

Datum: 2014

Qualitätsprüfung: NA

Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf ‘dipl’

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

. **Typ:** Annotationlayer – form_krankheit

Annotationsart: Spannenannotation

Beschreibung: NA

. **Typ:** Annotationvalue – form_krankheit

Wert: Wertbeschreibung:

Zeichenkette NA

4.19.problem

Typ: Layer – werkname

Name: problem

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

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

Typ: Preparationstep – problem

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: NA

Datum: 2014

Qualitätsprüfung: NA

Editor: Anke Lüdeling, Uwe Springmann

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

Typ: Annotationlayer – problem

Annotationsart:	Spannenannotation
Beschreibung:	NA

Typ: Annotationvalue – problem

Wert:	Wertbeschreibung:
Zeichenkette	NA

4.20.krauternormiert

Typ: Layer – krauternormiert

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

Typ: Preparationstep – kraeutername_normiert

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

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

Typ: Annotationlayer – kraeutername_normiert

Annotationsart: Spannenannotation
Beschreibung: In dieser Ebene wird ein systematischer Kräuternamen gegeben. Manchmal ist der nicht eindeutig, dann wird in den

Annotationsebenen „erläuterung“ oder „bemerkung_lexik“ darauf hingewiesen. Es gibt keine Leerzeichen, die beiden Bestandteile von zweiteiligen Bezeichnungen werden durch einen Unterstrich verbunden.

Typ: Annotationvalue – kraeutername_normiert

Wert: Wertbeschreibung:
Zeichenkette Normierter, systematischer Kräuternamen.

4.21.kraeuterzubereitung

Typ: Layer – kraeuterzubereitung

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

Typ: Preparationstep – kraeuterzubereitung

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

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

. **Typ:** Annotationlayer – kraeuterzubereitung

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

. **Typ:** Annotationvalue – kraeuterzubereitung

Wert:	Wertbeschreibung:
Zeichenkette	Kräuterzubereitung, z.B. <i>safft des weremuts</i> → <i>safft_des_weremuts</i> <i>des roten beyfusz bletter</i> → <i>des_roten_beifuß_blätter</i>

safft von weremut → saft_von_wermut

weremutsafft → wermutsaft

weremut safft → wermutsaft

4.22.form_zubereitung

Typ: Layer – form_zubereitung

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

Typ: Preparationstep – form_zubereitung

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

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

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

. **Typ:** Annotationlayer – form_zubereitung

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

. **Typ:** Annotationvalue – form_zubereitung

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

4.23.nomen_nominativ

Typ: Layer – nomen_nominativ

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

zusammengefasst.

Typ: Preparationstep – nomen_nominativ

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

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

Typ: Annotationlayer – nomen_nominativ

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

Typ: Annotationvalue – nomen_nominativ

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

4.24.form_nomen

Typ: *Layer* – form_nomen

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

. **Typ:** Preparationstep – form_nomen

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

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

. **Typ:** Annotationlayer – form_nomen

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

. **Typ:** Annotationvalue – form_nomen

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
kompAN	AN-Komposita
kompVN	VN-Komposita

kompPN	PN-Komposita
derivat	Derivate
nom	Mit <i>nom</i> werden alle impliziten Nominalisierungen annotiert (Konversionen, Bildungen mit Ablaut, auch syntaktische Nominalisierungen etc.). Beispiele sind (<i>die</i>) <i>sucht</i> oder (<i>das</i>) <i>kalt</i> .
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 <i>lex</i> 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.
name	andere Namen außer Personennamen (für diese s. Annotationsebene „personename“.

4.25.satztyp

Typ: *Layer* – satztyp

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

Typ: *Preparationstep* – satztyp

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

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

Typ: Annotationlayer – satztyp

Annotationsart: Spannenannotation
Beschreibung: Annotation von Satztypen. Da wir nur flach annotieren und nicht beliebig Ebenen vervielfachen wollen, haben wir ein Problem bei

geschachtelten Strukturen (also Relativsätzen, die Relativsätze enthalten oder Relativsätzen, die Demonstrativsätze enthalten). Wir haben jeweils den äußersten (größten) Satz annotiert. In der Annotationsebene „bemerkungen_syntax“ steht ein Hinweis auf die Verschachtelung.

Typ: Annotationvalue – satztyp

Wert:	Wertbeschreibung:
rs	Klare Relativsätze, dabei sowohl w-Relativsätze als auch d-Relativsätze, z.B. <i>im sch^z adet auch nít ob aín thier das nít fínnig wár ín beiffet Wer dē beyfusz bei im tregt wen er wandert der wirt nít müde</i>
padv	Sätze, die durch ein Präpositionaladverb eingeleitet werden, z.B. <i>das stabwurz genúczt ver^z zeret vberflússig feuchtikaít díe ín den dármen sind dauo aín kranck haít komet genant colíca passíó</i> Hier ist oft unsicher, ob das überhaupt Relativsätze sind.
rsx	Relativsätze ohne Hauptsatz. Das kommt häufig in Überschriften vor, z.B. <i>Wer eynen kurtzen atthem hat</i>
rsdem	Ambige Fälle: Relativsatz oder Demonstrativsatz, z.B. <i>vnd nimpt traurigkeit die kompt von Melancoley</i>
padvpart	Sätze mit Präpositionaladverb und Partizip, z.B. <i>dauon genomen auf aín halb lot</i> Diese Sätze sind nicht ambig wie rsdem, sondern enthalten beides! Ob sie überhaupt Relativsätze sind, ist unklar.
dem	Demonstrativsätze. Hier sind alle Sätze gemeint, deren Subjekt ein Demonstrativpronomen ist, z.B. <i>der hilfft on zweifel</i>
part	Partizipien, die sich ähnlich verhalten wie Relativsätze, z.B. <i>Jtem stabwurz mít oleí vn salcz zefamē gestoffen vnd auff den pulz hende vnd fúsz gelegtist gút wider das fieber</i> Oft sind mehrere Partizipien verkettet (zB durch und). Die ganze Kette wird als eine Spanne annotiert. Die Form der Partizipstrukturen ist unterschiedlich, manchmal ist das Partizip der Kopf, manchmal eine Präposition.

4.26.position_im_satz

Typ: *Layer* – position_im_satz

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

Typ: Preparationstep – position_im_satz

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

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

Datum: 2016
Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: Annotationlayer – position_im_satz

Annotationsart: Spannenannotation
Beschreibung: Position des Relativsatzes im Hauptsatz

Typ: Annotationvalue – position_im_satz

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

4.27. position_zur_bezugskategorie

Typ: Layer – position_zur_bezugskategorie

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

Typ: Preparationstep – position_zur_bezugskategorie

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

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

Typ: Annotationlayer – position_zur_bezugskategorie

Annotationsart: Spannenannotation
Beschreibung: Position des Relativsatzes zur Bezugskategorie

Typ: Annotationvalue – position_zur_bezugskategorie

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

4.28.form_bezugskategorie

Typ: *Layer* – satztyp

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

Typ: Preparationstep – form_bezugskategorie

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

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

Typ: Annotationlayer – form_bezugskategorie

Annotationsart:	Spannenannotation
Beschreibung:	Form der Bezugskategorie des Relativsatzes

Typ: Annotationvalue – form_bezugskategorie

Wert:	Wertbeschreibung:
np	Nicht pronominale NP, z.B. <i>vn treibt aufz den stain der in der blafen wachft</i>
d-pron	<i>der, die, das, dieser</i> etc., z.B. <i>Wer dē beyfußz bei im tregt wen er wandert der wirt nit müde</i>
p-pron	Personalpronomina, z.B.

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

satz	Bei weiterführenden Relativsätzen, die sich auf den Sachverhalt im ganzen Bezugssatz beziehen.
null	Bei freien und asyndetischen Relativsätzen, die kein overttes Korrelat im Hauptsatz haben, z.B. <i>vn ~was nach dem getranck vberbeleibt vo kreütern fol man warm legen auf die milcz</i>

4.29. verbstellung

Typ: Layer – verbstellung

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

Typ: Preparationstep – verbstellung

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

Schritt: 2

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

Typ: Annotationlayer – verbstellung

Annotationsart:	Spannenannotation
Beschreibung:	Verbstellung im Relativsatz

Typ: Annotationvalue – verbstellung

Wert:	Wertbeschreibung:
v2	Verb-Zweit-Stellung, z.B. <i>Vn frawen die kinder find feügen solz len nit nützen eppich</i>
ve	Verb-End-Stellung, z.B. <i>vmb tuz gent willen die diße künigin an di fem kraut befand</i>
venf	Verb-End-Stellung mit besetztem Nachfeld, z.B. <i>Vn auch fast wol bekome den die den stain habē in den lenden</i> Ob das wirklich als Nachfeld analysiert werden soll, ist unerheblich, das hier meint einfach, dass es eine klare Verb-End-Stellung ist, aber noch etwas danach kommt.
amb	Ambig zwischen v2 und ve, z.B. bei intransitiven Verben.

4.30.form_des_relativpronomens

Typ: *Layer* – form_des_relativpronomens

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

Typ: Preparationstep – form_des_relativpronomens

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

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

Editor: Vivian Voigt

Typ: Annotationlayer – form_des_relativpronomens

Annotationsart: Spannenannotation

Beschreibung: Form der reltivsatzeinleitenden Kategorie

Typ: Annotationvalue – form_des_relativpronomens

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 fo er vber feld gat wird nít müd

w-phras *welch frau etc.*, z.B.
Weliche fraw aín tods kind ín írem leíb het die trinck vo beyfuß

4.31.modifikation_bezugskategorie

Typ: Layer – modifikation_bezugskategorie

Name: modifikation_bezugskategorie

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

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

Typ: Preparationstep – modifikation_bezugskategorie

Schritt: 1

Aufbereitung: Linguistische Annotation

Aufbereitungsart: manuell

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

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

Typ: Annotationlayer – modifikation_bezugskategorie

Annotationsart: Spannenannotation
Beschreibung: Modifikation der Bezugskategorie

Typ: Annotationvalue – modifikation_bezugskategorie

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.

4.32. diachronie

Typ: *Layer* – diachronie

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

Typ: Preparationstep – diachronie

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: Gohar Schnelle
Datum: 2016
Qualitätsprüfung: NA
Editor: Gohar Schnelle

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

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – diachronie

Annotationsart: Spannenannotation

Beschreibung: In dieser Ebene werden einzelne linguistische Phänomene und Formen unsystematisch einem linguistischen Bereich (Syntax, Morphosyntax, Phonologie, Graphematik, Morphologie) zugeordnet und dient der Findung von Beispielen für didaktische Zwecke.

Typ: diachronie

Wert:	Wertbeschreibung:
Zeichenkette	Unsystematische Wertzuordnung. Ordnet ein Phänomen einem linguistischen Themenbereich zu, z.B. Morphologie, Phonetik oder Syntax. Es werden zum Teil auch einzelne Phänomene benannt, z.B. Diphthongierung.

4.33.dialekt

Typ: Layer – dialect

Name: dialekt

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

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

Typ: Preparationstep – dialekt

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: Gohar Schnelle
Datum: 2016
Qualitätsprüfung: NA
Editor: Gohar Schnelle

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

Typ: Annotationlayer - dialekt

Annotationsart: Spannenannotation
Beschreibung: In dieser unsystematischen Ebene werden Wortformen, die

Aufschluss über die diatopische Qualität des Textes geben annotiert und einem Dialektbereich so genau wie möglich zugeordnet.

Typ: Annotationvalue – dialekt

Wert:	Wertbeschreibung:
nd	Niederdeutsch
md	Mitteldeutsch
obd	Oberdeutsch
wnd	Westniederdeutsch
ond	Ostniederdeutsch
omd	Ostmitteldeutsch
wmd	Westmitteldeutsch
bair	Bairisch
alem	Alemannisch
obs	Obersächsisch
thür	Thüringisch
wobd	Westoberdeutsch
rip	Ripuarisch
hess	Hessisch

5. Annotationsebenen – Strukturelle Annotation

5.1. lb

Typ: Layer – lb

Name:	lb
Dokumentation:	Siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden

auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – lb

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2015
Qualitätsprüfung:	manuelle Korrektur
Editor:	Studierende der Humboldt-Universität zu Berlin, korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	manuelle Korrektur
Editor:	Vivian Voigt

Typ: Annotationlayer – lb

Annotationsart: Spannenannotation
Beschreibung: Markierung einer ganzen Zeile.

Typ: Annotationvalue – lb

Wert: Wertbeschreibung:
lb Markierung pro ganzer Zeile.

5.2. brace

Typ: Layer – brace

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

Typ: Preparationstep – brace

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: LAUDATIO
Datum: 2012

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

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

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – brace

Annotationsart: Spannenannotation

Beschreibung: Markierung einer Klammer, die Text über mehrere Zeilen

zusammenfasst.

Typ: Annotationvalue – brace

Wert:	Wertbeschreibung:
brLeft	Text links der Klammer.
brRight	Text rechts der Klammer.

5.3. brace_dir

Typ: Layer – brace_dir

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

Typ: Preparationstep – brace_dir

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	LAUDATIO
Datum:	2012
Qualitätsprüfung:	Manuelle Konsistenzprüfung

Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der
Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – brace_dir

Annotationsart: Spannenannotation

Beschreibung: Richtung der Klammer

Typ: Annotationvalue – p

Wert:	Wertbeschreibung:
Left	Wenn die Richtung der Klammer nach links zeigt und ihre Arme sich gen rechts erstrecken

5.4. p

Typ: *Layer* – p

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

Typ: Preparationstep – p

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	LAUDATIO
Datum:	2012
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studierende der Humboldt-Universität zu Berlin

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

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

Typ: Annotationlayer – p

Annotationsart: Spannenannotation
Beschreibung: Markierung eines ganzen Paragraphen.

Typ: Annotationvalue – p

Wert: Wertbeschreibung:

Markierung pro ganzem Paragraphen.

5.5. p_n

Typ: Layer – p_n

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

Typ: Preparationstep – p_n

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
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 ‘dipl’	
Qualitätsprüfer:	korpling	
Datum:	2016	
Qualitätsprüfung:	NA	
Editor:	Carolin Odebrecht	
Schritt:	3	
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS	
Aufbereitungsart:	automatisch	
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpapper/pepper/download/stable/Pepper_2016.06.06.zip)	
Format:	PAULA-XML, ANNIS	
Segmentierung der Annotation:	basiert auf ‘dipl’	
Qualitätsprüfer:	korpling	
Datum:		2016
Qualitätsprüfung:		NA
Editor:		Vivian Voigt

Typ: Annotationlayer – p_n

Annotationsart:	Spannenannotation
Beschreibung:	Markiert die Nummerierung eines Paragraphen.

Typ: Annotationvalue – p_n

Wert:	Wertbeschreibung:
Ziffer oder Buchstabe	Entsprechende Nummerierung des Paragraphen durch Ziffern oder Buchstaben.

5.6. p_rend

Typ: Layer – p_rend

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

Typ: Preparationstep – p_rend

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

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

Typ: Annotationlayer – p_rend

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibung der Darstellung eines Paragraphen.

Typ: Annotationvalue – p_rend

Wert:	Wertbeschreibung:
initialCapital	Der Paragraph beginnt mit einem Großbuchstaben.
bigBoldType	Der Paragraph beginnt mit einem fettgedruckten Buchstaben.

5.7. pb

Typ: Layer – pb

Name:	pb
-------	----

Dokumentation: Siehe Anhang

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

Typ: Preparationstep – pb

Schritt: 1

Aufbereitung: Strukturelle Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf ‘dipl’

Qualitätsprüfer: korpling

Datum: 2015

Qualitätsprüfung: manuelle Korrektur

Editor: Studierende der Humboldt-Universität zu Berlin, korpling

Schritt: 2

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf ‘dipl’

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – pb

Annotationsart: Spannenannotation

Beschreibung: Markierung einer ganzen Seite.

Typ: Annotationvalue – pb

Wert: Wertbeschreibung:

pb Markierung pro ganzer Seite.

5.8. pb_n

Typ: Layer – pb_n

Name: pb_n

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

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

Typ: Preparationstep – pb_n

Schritt: 1

Aufbereitung: Strukturelle Annotation

Aufbereitungsart: manuell

Tool: Excel 2013

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

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

Typ: Annotationlayer – pb_n

Annotationsart: Spannenannotation
Beschreibung: Markierung der Nummerierung einer Seite.

Typ: Annotationvalue – pb_n

Wert: Wertbeschreibung:
numerischer Wert Seitenzahl

5.9. pb_rend

Typ: Layer – pb_rend

Name: pb_rend

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

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

Typ: *Preparationstep* – pb_rend

Schritt: 1

Aufbereitung: Strukturelle Annotation

Aufbereitungsart: manuell

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: LAUDATIO

Datum: 2013

Qualitätsprüfung: Manuelle Konsistenzprüfung

Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

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

Typ:*Annotationlayer* – pb_rend

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibung der Darstellung/Zuordnung einer Seite, mehrerer Seiten.

Typ:*Annotationvalue* – pb_rend

Wert:	Wertbeschreibung:
vonHafelwurtz.Cap.III.	Seite/n ist/sind einer Überschrift zugeordnet.
vonChamillen.Cap.VIII.	Seite/n ist/sind einer Überschrift zugeordnet.
vorrede.	Seite/n ist/sind einer Überschrift zugeordnet.
vorred	Seite/n ist/sind einer Überschrift zugeordnet.

vonStaubwurtz.Cap.II.	Seite/n ist/sind einer Überschrift zugeordnet.
vonEibisch.Cap.V.	Seite/n ist/sind einer Überschrift zugeordnet. Seite/n sind einer Überschrift zugeordnet.
vonWermüt.Cap.I.	Seite/n ist/sind einer Überschrift zugeordnet.
vonDrachewurtz.Cap.III.	Seite/n ist/sind einer Überschrift zugeordnet.
ohlZuMachen.	Seite/n ist/sind einer Überschrift zugeordnet.
zumBestenZuDiftilliren.	Seite/n ist/sind einer Überschrift zugeordnet.
wafferAußKräuternVndDergleichen	Seite/n ist/sind einer Überschrift zugeordnet.
auffsBefzZuDiftilliren.	Seite/n ist/sind einer Überschrift zugeordnet.
außKräuternVndDergleichen	Seite/n ist/sind einer Überschrift zugeordnet.
wafferAußKräuternVndDergleichen	Seite/n ist/sind einer Überschrift zugeordnet.
amBestenZuDiftilliren.	Seite/n ist/sind einer Überschrift zugeordnet.

5.10.pb_ana

Typ: *Layer* – pb_ana

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

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

Typ: Preparationstep – pb_ana

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

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

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

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

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ: Annotationlayer – pb_ana

Annotationsart: Spannenannotation

Beschreibung: Interpretation von vermutlich nicht korrekten Seitenzahlangaben.

Typ: Annotationvalue – pb_ana

Wert: Wertbeschreibung:

Int Entsprechende Angabe der vermutlich korrekten Seitenzahl.

5.11.div1 – div5

Typ: Layer – div1 – div5

Name: div1, div2, div3, div4, div5

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

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

Typ: Preparationstep – div1 – div5

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

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

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

Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: *Annotationlayer* – div1 – div5

Annotationsart: Spannenannotation
Beschreibung: Eine Sektion, ein Kapitel/Unterkapitel eines Dokuments. Die verschachtelte Tiefe wird explizit von der angegebenen Nummer von div in der PAULA/reIANNIS Version angezeigt. Gilt pro Ebene (**div1-div5**).

Typ: *Annotationvalue* – div1 – div5

Wert:	Wertbeschreibung:
div1	erste/s Sektion/Kapitel/Unterkapitel in Ebene div1.
div2	zweite/s Sektion/Kapitel/Unterkapitel in Ebene div2.
div3	dritte/s Sektion/Kapitel/Unterkapitel in Ebene div3.
div4	vierte/s Sektion/Kapitel/Unterkapitel in Ebene div4.
div5	fünfte/s Sektion/Kapitel/Unterkapitel in Ebene div5.

5.12.div1_type – div5_type

Typ: *Layer* – div1_type – div5_type

Name: div1_type, div2_type, div3_type, div4_type, div5_type
Dokumentation: http://korpling.german.hu-berlin.de/ridges/documentation_de.html
Beispiele:
Siehe Anhang
Beschreibung: Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren,

strukturellen Annotationen.

Typ: *Preparationstep* – div1_type – div5_type

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

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

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

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

Typ: *Annotationlayer* – div1_type – div5_type

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibt den Typ/die Art des Kapitels/Unterkapitels. Die Einteilung kann von einem ganzen Buch, über Kapitel bis hin zu Unterkapitel reichen. Dazu können auch registerspezifische Typen wie Ort des Anbaus oder Form einer Pflanze zählen. Gilt pro Ebene (div1-div5).

Typ: *Annotationvalue* – div1_type – div5_type

Wert:	Wertbeschreibung:
appendix	Anhang.
book	Ein ganzes Buch.
chapter	Kapitel.
description	Beschreibung einer Pflanze.
form	Form einer Pflanze.
herb	Kraut.
names	Liste von Namen.
name	Namen.
nature	Natur.
parts_preparation_and_uses	Zubereitung und Nutz.
places	Anbauggebiete einer Pflanze.
place	Orte.
power	Heilkraft
preface	Vorwort.

power	Wirkung einer Pflanze.
reproduction	Wiederherstellung.
season	Saison des Anbaus.
section	Unterkapitel.
species	Spezies.
title	Ein Titel.
time	Anbauzeit.
utensils	Werkzeuge.

5.13.div1_n – div5_n

Typ: *Layer* – div2_n – div3_n

Name:	div2_n, div_3
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (dipl -Ebene). So ist die Tokenisierung in der dipl -Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – div2_n – div3_n

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

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

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

Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013

Segmentierung der Annotation:
basiert auf 'dipl'

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

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

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling
Datum: 2015
Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: *Annotationlayer* – div2_n – div3_n

Annotationsart: Spannenannotation
Beschreibung: Markiert eine mit Ziffern oder Buchstaben nummerierte Sektion.

Typ: *Annotationvalue* – div1_n – div5_n

Wert: Wertbeschreibung:
Ziffer Nummerierung der Sektion.

5.14.unclear

Typ: *Layer* - unclear

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

Typ: *Preparationstep* – unclear

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

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

Typ: *Annotationlayer* – unclear

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

Typ: *Annotationvalue* – unclear

Wert:	Wertbeschreibung:
unclear	Nicht leserliche Zeichen oder Zeichenketten werden unabhängig von der Anzahl unleserlicher Zeichen, in der 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.

5.15.atLeast

Typ: *Layer* – atLeast

Name: atLeast

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

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

Typ: *Preparationstep* – atLeast

Schritt: 1

Aufbereitung: Strukturelle Annotation

Aufbereitungsart: Manuell

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: LAUDATIO

Datum: 2012

Qualitätsprüfung: Manuelle Konsistenzprüfung

Editor: Studierende der Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach Excel 2013

Aufbereitungsart: manuell

Tool: Excel 2013

Format: Excel 2013

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: korpling

Datum: 2016

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

Typ: *Annotationlayer* – atLeast

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

Typ: *Annotationvalue* – atLeast

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

5.16.atMost

Typ: *Layer* – atMost

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

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

Typ: Preparationstep – atMost

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Datum:	2013
Qualitätsprüfer:	LADUATIO
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

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

Typ: Annotationlayer – atMost

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

Typ: Annotationvalue – atMost

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

5.17.interpretation

Typ: Layer– interpretation

Name:	interpretation
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen-

und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: *Preparationstep* – interpretation

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

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

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

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

Typ: *Annotationlayer*– interpretation

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

Typ: *Annotationvalue*– interpretation

Wert:	Wertbeschreibung:
Text	Angabe der Interpretation.

5.18.figure

Typ: *Layer* – figure

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

strukturellen Annotationen.

Typ: Preparationstep – figure

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

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2015
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: Annotationlayer – figure

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

Typ: Annotationvalue – figure

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

5.19.figure_rend

Typ: *Layer* – figure_rend

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

Typ: *Preparationstep* – figure_rend

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

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

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

Typ: *Annotationlayer* – figure_rend

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

Typ: *Annotationvalue* – figure_rend

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

5.20.hi

Typ: *Layer* – hi

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

strukturellen Annotationen.

Typ: Preparationstep – hi

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

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

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

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

Typ: Annotationlayer – hi

Annotationsart: Spannen
Beschreibung: Markierung eines irgendwie hervorgehobenen Bereichs.

Typ: Annotationvalue – hi

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

5.21.typeface

Typ: Layer – typeface

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

Typ: *Preparationstep* – typeface

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2015
Qualitätsprüfung: Manuelle Konsistenzprüfung: Korrektur der false positives, Auffüllung leerer Zellen mit der Standardschriftart des jeweiligen Textes
Editor: Studierende der Humboldt-Universität zu Berlin, korpling

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

Typ: *Annotationlayer* – typeface

Annotationsart: Spannenannotation
Beschreibung: Weist jedem Token eine Schriftart zu. In manchen Texten kommt es vor, dass ein und dasselbe Token zum Teil in einem und zum Teil in einem anderen Font gedruckt wurde. Dies betrifft v.a. Wörter mit deutscher Flexionsendung und fremdsprachiger

Wurzel. Für diese Token wurde innerhalb von „typeface“ der Wert „mixed“ annotiert.

Typ:*Annotationvalue* – typeface

Wert:	Wertbeschreibung:
antiqua	Text, der in Antiqua gedruckt ist.
gothic	Text, der in Gothic gedruckt ist.
gothicF	Text, der in Gothic Fraktur gedruckt ist (fakultative Zusatzinformation, bisher nur in NewKreuerbuch_1563_Handsch).
gothicS	Text der in Gothic Schwabacher gedruckt ist (fakultative Zusatzinformation, bisher nur in NewKreuerbuch_1563_Handsch).
mixed	Text, der in verschiedenen Schriftsätzen gedruckt ist.

5.22.hi_rend

Typ: *Layer* – hi_rend

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

Typ:*Preparationstep* – hi_rend

Schritt:	1
Aufbereitung:	Strukturelle Annotation

Aufbereitungsart: manuell
 Tool: Excel 2013
 Format: Excel 2013
 Segmentierung der Annotation: basiert auf 'dipl'
 Qualitätsprüfer: korpling
 Datum: 2015
 Qualitätsprüfung: Manuelle Konsistenzprüfung: Korrektur der false positives
 Editor: Studierende der Humboldt-Universität zu Berlin

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

Typ:*Annotationlayer* – hi_rend

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

Typ:*Annotationvalue* – hi_rend

Wert: Wertbeschreibung:
 italics Text, der kursiv gedruckt ist.
 bold Text, der fett gedruckt ist.

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

5.23.head

Typ: *Layer* – head

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

Typ: Preparationstep – head

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

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

Typ: Annotationlayer – head

Annotationsart:	Spannenannotation
Beschreibung:	Markierung einer Überschrift.

Typ: Annotationvalue – head

Wert:	Wertbeschreibung:
head	Allgemeine Markierung einer Überschrift.

5.24.head_n

Typ: *Layer* – head_n

Name:	head_n
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen

und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – head_n

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

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

Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch

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

Typ: Annotationlayer – head_n

Annotationsart:	Spannenannotation
Beschreibung:	Zeigt die Nummerierung einer Überschrift an.

Typ: Annotationvalue – head_n

Wert:	Wertbeschreibung:
Ziffer	Konkrete Nummerierung der Überschrift.

5.25.head_rend

Typ: *Layer* – head_rend

Name:	head_rend
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die 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* – head_rend

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf ‘dipl’

Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: *Annotationlayer* – head_rend

Annotationsart: Spannenannotation
Beschreibung: Beschreibung der grafischen Darstellung der Überschrift, z. B. Farbe.

Typ: *Annotationvalue* – head_rend

Wert: brown
Wertbeschreibung: Überschrift in brauner Farbe.

5.26.note

Typ: *Layer* – note

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

Typ: *Preparationstep* – note

Schritt: 1

Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘dipl’
Qualitätsprüfer:	korpling
Datum:	2015
Qualitätsprüfung:	Manuelle Konsistenzprüfung. Die Annotationen der Ebene head_type mit dem tag „margin“ wurden in die Ebene note übernommen. Die Annotationen der Ebene head , die mit denen von head_type korrelieren, wurden entfernt (in Version 4.1) .
Editor:	Studierende der Humboldt-Universität zu Berlin

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

Typ: Annotationlayer – note

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

realisiert werden müssten, in die dipl-Ebene/Transkription aufgenommen.

Typ: Annotationvalue – note

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

5.27.ref

Typ: Layer – ref

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

Typ: Preparationstep – ref

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

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

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

Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: Annotationlayer – ref

Annotationsart: Spannenannotation
Beschreibung: Referenz einer Fußnote.

Typ: Annotationvalue – ref

Wert: Wertbeschreibung:
ref Referenz einer Fußnote.

5.28.ref_target

Typ: Layer – ref_target

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

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

Typ: *Preparationstep* – ref_target

Schritt: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell

Tool:	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: *Annotationlayer* – ref_target

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

Typ: *Annotationvalue* – ref_target

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

5.29.ref_type

Typ: *Layer* – ref_type

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

Typ: *Preparationstep* – ref_type

Schritte: 1
Aufbereitung: Strukturelle Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: Excel 2010
Segmentierung der Annotation: basiert auf 'dipl'

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

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

Segmentierung der Annotation:
basiert auf 'dipl'

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

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

Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: *Annotationlayer* – ref_type

Annotationsart: Spannenannotation

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

Typ: *Annotationvalue* – ref_type

Wert: Wertbeschreibung:
noteAnchor Art der Referenz.

5.30.quote

Typ: *Layer* – quote

Name: quote

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

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

Typ: Preparationstep – quote

Schritt: 1

Aufbereitung: Strukturelle Annotation

Aufbereitungsart: manuell

Tool: Excel 2010

Format: Excel 2010

Segmentierung der Annotation: basiert auf 'dipl'

Qualitätsprüfer: LAUDATIO

Datum: NA

Qualitätsprüfung: NA

Editor: Studierende der Humboldt-Universität zu Berlin

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

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

Typ: Annotationlayer – quote

Annotationsart: Spannenannotation
Beschreibung: Markiert ein Zitat im Text.

Typ: Annotationvalue – quote

Wert:	Wertbeschreibung:
Quote	Ein Zitat im Text.

5.31.list

Typ: *Layer* – list

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

Typ: Preparationstep – list

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

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

Typ: Annotationlayer – list

Annotationsart:	Spannenannotation
Beschreibung:	Eine Liste von Elementen.

Typ: Annotationvalue – list

Wert:	Wertbeschreibung:
List	Allgemeine Markierung einer Liste.

5.32.list_type

Typ: *Layer* – list_type

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

Typ: *Preparationstep* – list_type

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

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

Typ: *Annotationlayer* – list_type

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibung der Art der Liste.

Typ: *Annotationvalue* – list_type

Wert:	Wertbeschreibung:
Simple	Einfache Aufzählung.

5.33.item

Typ: Layer – item

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

Beispiele: siehe Anhang

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

Typ: Preparationstep – item

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht

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

Typ: Annotationlayer – item

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

Typ: Annotationvalue – item

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

5.34.xml_id

Typ: Layer – xml_id

Name:	xml_id
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Unter strukturellen Annotationen werden alle Annotationen zusammengefasst, die Aussagen über die graphische Strukturierung der Textgrundlage liefern. Dazu zählen Kennzeichnungen der Zeilen- und Seitenumbrüche, Seitenzahlen, unkenntliche/nicht lesbare Zeichen

und Marginalien sowie Fußnoten. Strukturelle Annotationen werden auf Grundlage der diplomatischen Ebene vergeben (**dipl**-Ebene). So ist die Tokenisierung in der **dipl**-Ebene Vorlage für die weiteren, strukturellen Annotationen.

Typ: Preparationstep – xml_id

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

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

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

	berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: Annotationlayer – xml_id

Annotationsart:	Spannenannotation
Beschreibung:	Bezeichnet die konkrete Nummer der Fußnote in einem Dokument, angefangen mit dem Zähler 1. Fußnoten sind in ihrer Indizierung unabhängig von ihrer tatsächlichen Nummerierung im Text. Der ersten Fußnote wird über xml_id die Kennzeichnung f1, der zweiten f2 etc. zugewiesen, selbst wenn sie bspw. im Text die Nummer 86) darstellt oder mit **) markiert ist. Siehe auch 5.28 und für Beispiele 8.2.

Typ: Annotationvalue –xml_id

Wert:	Wertbeschreibung:
fINT	Z (Zahl) bezeichnet die Nummer der Fußnote.

6. Annotationsebenen – Inhaltliche Annotation

6.1. definition

Typ: Layer – definition

Name:	definition
Dokumentation:	http://korpling.german.hu-berlin.de/ridges/documentation_de.html Beispiele: siehe Anhang
Beschreibung:	Inhaltliche Annotationen sind speziell auf das Register des Korpus

bezogen. Darunter fallen u. a. Annotationen von im Korpus enthaltenden Begriffen, Definitionen oder Referenzen.

Typ: *Preparationstep* – definition

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	LAUDATIO
Datum:	2013
Qualitätsprüfung:	Manuelle Konsistenzprüfung. Eingrenzung der in dieser Ebene beschriebenen Informationen, Markierung von Übersetzung in eine eigene Ebene (foreign_trans) ausgelagert.
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf ‘norm’
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-

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

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

Typ: *Annotationlayer* – definition

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

Typ: *Annotationvalue* – definition

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

6.2. disease

Typ: *Layer* – disease

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

Typ: *Preparationstep* – disease

Schritt: 1

Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	NA
Datum:	2012
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: Annotationlayer – disease

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

Typ: Annotationvalue – disease

Wert: Wertbeschreibung:
di Benennung einer Krankheit

6.3. term

Typ: Layer – term

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

Typ: Preparationstep – term

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

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

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

Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: korpling
Datum: 2016
Qualitätsprüfung: NA
Editor: Vivian Voigt

Typ: Annotationlayer – term

Annotationsart: Spannenannotation
Beschreibung: Linguistische Annotation, die einen technischen oder wissenschaftlichen Begriff markiert.

Typ: Annotationvalue – term

Wert:	Wertbeschreibung:
t	Kennzeichnet einen allgemeinen, technischen oder wissenschaftlichen Begriff (term) innerhalb der Kräuterkunde.
h	Kennzeichnet einen Begriff, der eine Pflanze bzw. ein Kraut (herb) benennt/beschreibt.
d	Kennzeichnet einen Begriff, der eine Krankheit (disease) benennt/beschreibt.

6.4. personenname

Typ: *Layer* – personenname

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

Typ: Preparationstep – personenname

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

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

Typ: Annotationlayer – personenname

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

Typ: Annotationvalue – personenname

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

6.5. werkname

Typ: Layer – werkname

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

Typ: Preparationstep – werkname

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Aufbereitungsart:	manuell
Tool	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'dipl'
Qualitätsprüfer:	NA
Datum	2015
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin; korpling
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.huberlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	NA basiert auf 'dipl'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: Annotationlayer – werkname

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

Typ: Annotationvalue – werkname

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

6.6. author_ref

Typ: *Layer* – author_ref

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

Typ: *Preparationstep* – author_ref

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2015
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

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

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

Typ: *Annotationlayer* – author_ref

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

Typ: *Annotationvalue* – author_ref

Wert:	Wertbeschreibung:
pron1pl	Referenz mit der ersten Person Plural.
pron1sg	Referenz mit der ersten Person Singular.
pron2sg	Referenz mit der zweiten Person Singular, z. B. im Imperativ.
pron3sg	Referenz mit der dritten Person Singular.
author	Explizite Referenz auf einen Autor.

6.7. reader_ref

Typ: *Layer* – reader_ref

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

Typ: *Preparationstep* – reader_ref

Schritt: 1
Aufbereitung: Inhaltliche Annotation
Tool: Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'norm'
Qualitätsprüfer: NA
Datum: 2015
Qualitätsprüfung: NA
Editor: Studierende der Humboldt-Universität zu Berlin

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

Typ: *Annotationlayer* – reader_ref

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

Typ: *Annotationvalue* – reader_ref

Wert: Wertbeschreibung:
pron1pl Referenz mit der ersten Person Plural.

pron2sg	Referenz mit der zweiten Person Singular, z. B. im Imperativ.
pron3sg	Referenz mit der dritten Person Singular.
pron2pl	Referenz mit der zweiten Person Plural.
address	Explizite Referenz auf den Leser.

6.8. plant

Typ: *Layer* – plant

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

Typ: *Preparationstep* – plant

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	LAUDATIO
Datum:	2012
Qualitätsprüfung:	Manuelle Konsistenzprüfung.
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013

Segmentierung der Annotation:

basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Carolin Odebrecht

Schritt: 3

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

Aufbereitungsart: automatisch

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

Format: PAULA-XML, ANNIS

Segmentierung der Annotation: basiert auf 'norm'

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Typ:*Annotationlayer* – plant

Annotationsart: Spannenannotation

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

Typ:*Annotationvalue* – plant

Wert: Wertbeschreibung:

pl Benennung einer Pflanze.

6.9. property

Typ: *Layer* – property

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

Typ: *Preparationstep* – property

Schritt:	1
Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	LAUDATIO
Datum:	2012
Qualitätsprüfung:	Manuelle Konsistenzprüfung.
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA

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

Typ:*Annotationlayer* – property

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

Typ:*Annotationvalue* – property

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

6.10. name

Typ: *Layer* – name

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

Typ: Preparationstep – name

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

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

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

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

Typ: Annotationlayer – name

Annotationsart: Spannenannotation
Beschreibung: Markiert Eigennamen.

Typ: Annotationvalue – name

Wert: Wertbeschreibung:
Name: Allgemeine Markierung von Eigennamen.

6.11.name_type

Typ: *Layer* – name_type

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

Typ: *Preparationstep* – name_type

Schritt: 1

Aufbereitung:	Inhaltliche Annotation
Tool	Excel 2010
Format:	Excel 2010
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	NA
Datum:	2012
Qualitätsprüfung:	NA
Editor:	Studierende der Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach Excel 2013
Aufbereitungsart:	manuell
Tool:	Excel 2013
Format:	Excel 2013
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Carolin Odebrecht
Schritt:	3
Aufbereitung:	Konvertierung von Excel 2013 nach PAULA-XML und ANNIS
Aufbereitungsart:	automatisch
Tool:	SaltNPepper, ExcelImporter (https://korpling.german.hu-berlin.de/saltnpepper/pepper/download/stable/Pepper_2016.06.06.zip)
Format:	PAULA-XML, ANNIS
Segmentierung der Annotation:	basiert auf 'norm'
Qualitätsprüfer:	korpling
Datum:	2016
Qualitätsprüfung:	NA
Editor:	Vivian Voigt

Typ: *Annotationlayer* – name_type

Annotationsart: Spannenannotation
Beschreibung: Spezifiziert den Typ der Eigennamen.

Typ: *Annotationvalue* – name_type

Wert:	Wertbeschreibung:
herb	Eigename eines Krauts.
scholar	Eigename einer Schule/Lehre.
plant	Eigename einer Pflanze.
person	Eigename einer Person.
flower	Eigename einer Blume.
tree	Eigename eines Baums.
gardener	Eigename eines Gärtners.
publisher	Eigename eines Verlegers.

6.12.citation

Typ: *Layer* – citation

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

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* – citation

Schritt: 1
Aufbereitung: Inhaltliche Annotation

Tool Excel 2013
Format: Excel 2013
Segmentierung der Annotation: basiert auf 'dipl'
Qualitätsprüfer: NA
Datum: 2016
Qualitätsprüfung: NA
Editor: Gohar Schnelle

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

Typ: *Annotationlayer* – citation

Annotationsart: Spannenannotation
Beschreibung: Unsystematische Ebene. Kennzeichnet und benennt Zitate (z.B. aus der Bibel), wodurch ein diachroner Vergleich Textstellen ermöglicht wird, denen ursprünglich identisches sprachliches Material zugrunde liegt.

Typ: *Annotationvalue* – citation

Wert: Wertbeschreibung:
Zeichenkette Sollte möglichst vollständige bibliographische Angaben zur Identifikation des

Zitats enthalten. Für genormte Bibliographien wie die Bibel oder Werke der klassischen Antike können Siglen verwendet werden. Tertium comparationis für Bibelzitate sind die Lutherbibel 1984 bzw. die katholische Einheitsübersetzung <http://www.die-bibel.de/online-bibeln>. Gekennzeichnet werden Bibelzitate nach folgendem Muster: Sigle, ggf. Buchnummer, Kapitel, Vers. Es werden die sogenannten Loccumer Siglen verwendet. (https://de.wikipedia.org/wiki/Wikipedia:Wie_zitiert_man_Bibelstellen). Für klassische Autoren: Thesaurus linguae latinae 2. Auflage 1990. Für mittelalterliche Autoren: *Index scriptorum novus mediae latinitatis ab anno DCCC usque ad annum MCC*.

7. Metadaten

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

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

Typ: *Preparationstep* - Metadaten

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

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

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

Format: PAULA-XML, ANNIS

Qualitätsprüfer: korpling

Datum: 2016

Qualitätsprüfung: NA

Editor: Vivian Voigt

Metadatum	Annotationswerte	Beschreibung
autor	Zeichenkette	Name des Autors des Textes.
bibl	Zeichenkette	Voller bibliographischer Eintrag für die Quelle inklusive der im Korpus annotierten Seitenzahlen.
datum	JJJJ (Int)	Datum der Veröffentlichung, normalerweise das Jahr.
verlag	Zeichenkette	Verlag/Verleger.
ort	Zeichenkette	Verlagsort.
titel	Zeichenkette	Titel der Arbeit, aus dem das Dokument extrahiert wurde.
uebersetzer	Zeichenkette	Übersetzer des Textes, wenn vorhanden.
uebersetztAus	it lat	Sprache, aus der der Text übersetzt wurde.
herausgeber	Zeichenkette	Herausgeber des Textes (wenn zusätzlich zum Verlag/Verleger angegeben).
version	5.0	Versionierung des Korpus.
aufgabe	Erstauflage Nichterstauflage	Wenn bekannt, wird hier angegeben, ob es sich bei dem Faksimile um eine Erstauflage handelt oder nicht.
band	Int	Nummer des Bandes, wenn vorhanden.
bereich	Wissenschaft Alltag	Angabe darüber, ob der Text eher Alltags- oder Wissenschaftsthemen behandelt.
thema	additiv aus: Al As B G K M R S	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, M: Medizin, R: Religion, S: Sprachwissenschaft. Beispielwerte sind "B", "BM" oder "BKM".
register	Kraeuterkunde	Register des Textes.
einMehrSpr	einsprachig mehrsprachig	Gibt an, ob größere Textstellen vorhanden sind, die nicht auf Deutsch verfasst wurden. Dies gilt nicht für einzelne Übersetzungen von Pflanzennamen.

originaldatum	JJJJ (Int)	Ist ein Text eine Nichterstaufgabe, dann kann hier das Jahr der Erstaufgabe stehen, wenn bekannt.
repositorium	URL	URL zum Repositorium des Faksimiles.
sprachtyp	mhd fnhd nhd	Angabe des Sprachtyps. mhd: mittelhochdeutsch; fnhd: frühneuhochdeutsch, nhd: neuhochdeutsch
sprachgebiet	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.
textgestaltung	Prosa Poesie gemischt	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.
gestaltungselemente	Endreim Endreim, Metrik	Wenn im Metadatum "textgestaltung" die Werte "Poesie" oder "gemischt" angegeben sind, dann wird hier eingetragen, welche Gestaltungselemente dafür verwendet wurden.
vorredeVorh	ja nein	Gibt an, ob eine Vorrede mit transkribiert wurde.
wermutVorh	ja nein	Gibt an, ob ein Textabschnitt zum Thema "Wermut" transkribiert wurde.
kraeutermonographiesammlung	ja nein	Gibt an, ob es sich bei dem Text um eine Kräutermonographiesammlung handelt, d.h. dass nacheinander verschiedene Kräuter abgehandelt werden.
korpusdokumentation	URL	URL zur Korpusdokumentation.

8. Referenzen

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

Empfohlene Literatur:

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

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

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

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

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

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

1. Anhang

Transkription und Normalisierung

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

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

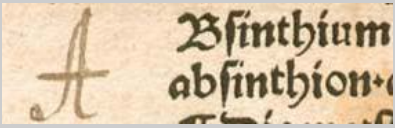
Die **clean**-Ebene enthält erste Normalisierungen hinsichtlich Sonderzeichen und grafischer Strukturierungen. So werden u. a. Ligaturen, die nach moderner Rechtschreibung nicht mehr verwendet werden, normalisiert. Graphische Markierungen der einzelnen Texte wie Zeilenumbrüche werden aufgelöst und Sonderzeichen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

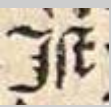
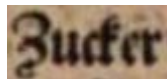

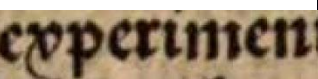
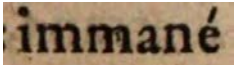
Die **norm**-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier die Tokenisierung, die Orthografie, Flexionsformen und Wortbildungen einheitlich dem nhd. Sprachstand angepasst werden. Durch die einzelnen Normalisierungsschritte und der daraus resultierenden multiplen Segmentierung ist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

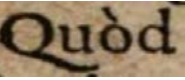
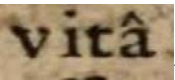

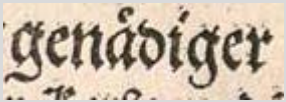

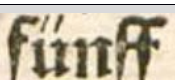
Nachfolgend werden die Richtlinien zur Transkription und Normalisierung im Einzelnen gegenübergestellt. Allgemein für jede Segmentierungsebene sowie Annotationsebene gilt, dass keine Leerzeilen oder Leerzeichen enthalten sein dürfen. Dazu wird für jeden Fall ein Beispiel aus dem Ridges Herbology Version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









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

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

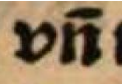

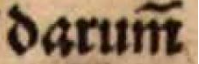
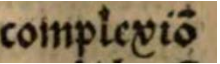



Nr.	dipl-Ebene	clean-Ebene	norm-Ebene
	Transkript des Faksimiles.	Wird automatisch durch ein Python-Skript (clean-skript.py) erstellt. Dieses ersetzt alle heute unüblichen Sonderzeichen durch heute verwendete Entsprechungen.	Erfolgt manuell im .xlsx Format nach der modernen neuen Rechtschreibung. Tipp: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


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

immané		
 Quòd	Quòd	Quòd
 vitâ	vitâ	vitâ
 Û 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
 genâdiger 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
 die 'i' mit schrägem Strich wird nicht vom 'i' mit Punkt unterschieden.		
dipl	clean	norm
die	die	die
 fünff Das ũ mit schrägen Strichen wird nicht vom ü mit Punkten unterschieden.	ũ 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
 ei	ein	ein
 samē	samen	Samen
 eynē	eynem	einem
 nēlich	nemlich	nämlich
- Verdopplung eines Buchstaben		
 hūmel	himmel	Himmel
 iñ	inn	in
- Abkürzung frequenter Wörter und Endungen		

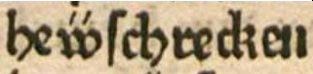
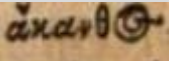

1 Nachweis für das Beispiel: Reichmann & Wegera 1993: 31.


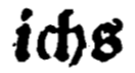
	unn (sic!)	und	
	umm (sic!)	umb	
	darumm (sic!)	darumb	
 complexiō	complexiom (sic!)	complexione	
Zwibel od Kn Das Dicit-Zeichen (Abkürzung für "der") wird in der dipl-Ebene mit ð (AltGr oder als Sonderzeichen)	Es wird in der clean automatisch mit "der" ersetzt	In der norm auch.	
dipl	clean	norm	
oð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

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

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

	übernommen.		
	Dipl	clean	norm
	⊙	⊙	⊙
	diefe: haben Senkrechtstriche werden mit „ “ übernommen.	Senkrechtstriche werden übernommen.	Senkrechtstriche werden übernommen.
	Dipl	clean	norm
	diefe	diese	diese
1.5	Die Ligatur „æ“ wird mit U 00E6, für „Æ“ mit U 00C6 übernommen. Das gleiche gilt ebenfalls für „Œ“ U 0152 und „œ“ U 0153.	Die Ligatur wird aufgelöst.	
	dipl	clean	norm
	hæc	haec	haec
	 Die ct-Ligatur wird nicht übernommen.	Die ct-Ligatur wird nicht übernommen.	Die ct-Ligatur wird nicht übernommen.
	dipl	clean	norm
	Lactucis	Lactucis	Lactucis
	 Dies ist eine Ligatur aus v̇ und v. Darstellung mit v+U+0308 Die Ligatur wird aufgelöst v̇v	v̇v	
	dipl	clean	norm



	heüvſchrecken	Heuschrecken
heüvſchrecken		
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	fleiszige	fleißige
Griechische Ligaturen werden nicht abgebildet, sondern bereits in der dipl aufgelöst:  Ligatur aus o und ç Als Hilfe: https://de.wikipedia.org/wiki/Griechisches_Alphabet#/media/File:Greek_alphabet_ligatures.jpg		
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

	&	&	&
1.6	Sprachliches Material mit anderen Schriftsystemen (z.B. Griechisch) wird in clean und norm nicht geändert http://www.unicode.org/chararts/PDF/U0370.pdf sowie http://www.unicode.org/chararts/PDF/U1F00.pdf		
	dipl	clean	norm
	ἄκανθος	ἄκανθος	ἄκανθος
1.7	Oft werden die ersten Zeichen bzw. ersten Wörter (im Bsp. Von Wegε) einer neuen Seite bei einem Seitenumbruch in der vorhergehenden Seite in einem eigenen Absatz/in einer eigenen Zeile doppelt realisiert (sog. Kustoden). Diese Zeichen bzw. das Wort wird nicht mit in das Transkript aufgenommen.		
2.0	 Auch entgegen moderner Orthographieregeln wird analog zur Textgrundlage tokenisiert.		Wortbildung und Großschreibung, die nicht der modernen Orthographieregeln entsprechen, werden angeglichen.
	dipl	clean	norm
	Spannenlang	Spannen-lang	spannenlang
2.1	 Klitika werden als ein Token realisiert.	Klitika werden als ein Token realisiert.	Klitika werden in dieser Ebene aufgelöst. Dazu wird die Tokenisierung verändert, in dem aus einem Token in dipl (oder clean) zwei Token gemacht werden!
	dipl	clean	norm
	Ichs	ichs	ich es

Auch entgegen moderner Orthographieregeln wird analog zur Textgrundlage tokenisiert.		Getrennt- und Zusammenschreibung wird nach modernen Orthografieregeln angepasst. Dazu wird die Tokenisierung verändert, indem aus einem Token in der dipl-Ebene (oder clean) zwei Token in der norm-Ebene gemacht werden!
dipl	clean	norm
zufetzen	zusetzen	zu
		setzen
Im Falle einer Worttrennung aufgrund von Zeilenumbrüchen, die keine overte, graphische Markierung wie „=-“ beinhalten, werden formal die getrennten Elemente als jeweils ein Token in der dipl-Ebene betrachtet.	Die Worttrennung ohne graphische Markierung ist nicht in jedem Fall transparent. Graphische Normalisierungen werden nicht vorgenommen.	Diese Art der Worttrennung, wenn sie transparent ist, wird nach modernen Orthographieregeln normalisiert und die graphische Trennung aufgehoben.
dipl	clean	norm
ge	ge	genannt
nent	nent	
Komposita, egal welcher Komplexität oder Bildungsweise, mit Gleichheitszeichen (U 2E17) werden als ein Token realisiert.	Die morphologische Trennung durch „≠“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt.	Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschrieben. (Grundlage hierfür bildet die aktuelle Ausgabe des Dudens.)
dipl	clean	norm
Artzney≠Kunſt	Artzney-Kunst	Arzneikunst

2.2	Von Zeilenumbrüchen betroffene Elemente werden analog zur Primärquelle getrennt tokenisiert. Das Trennungszeichen „≠“ wird mit U 2E17 übernommen.	Wörter, die von Zeilenumbrüchen betroffen sind, werden in der clean-Ebene ohne den (Doppel-) Bindestrich, der diesen anzeigt, sowie in einer Spanne zusammengefasst realisiert. Die grafische Worttrennung wird somit aufgehoben.	
	dipl	clean	norm
	allge≠	allgemeinen	allgemeinen
	meinen		
	Von Zeilenumbrüchen betroffene Komposita werden analog zur Primärquelle getrennt tokenisiert, dabei bleibt die Kennzeichnung der morphologischen Worttrennung unberührt.	Wenn ein Kompositum, das durch Gleichheitszeichen grundsätzlich getrennt wird, von einem Zeilenumbruch betroffen ist, so wird dieses Gleichheitszeichen in der clean-Ebene entfernt, das andere Gleichheitszeichen für die morphologische Trennung wird analog zu Nummer 1.1 mit einem Minus ersetzt. Das Kompositum wird in einer Spanne zusammengefasst wiedergegeben.	Die Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschieden. (Grundlage hierfür bildet die aktuelle Ausgabe des Dudens.)
	dipl	clean	norm
Stab≠	Stabwurtz-Oel	Stabwurzöl	
wurtz≠Oel			
2.3	Abkürzungen werden analog zur Textgrundlage tokenisiert. Das heißt, auch Setzfehler 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		Um die Kardinalzahlen von Ordinalzahlen in der norm-

	<p>Interpunktion auf. Diese Punktsetzungen werden mit der jeweiligen Ziffer als Token realisiert. Ausnahmen bilden Zahlen, die mit einem satzbeenden Punkt auftreten. Hier wird wie gewohnt die Satzinterpunktion getrennt von der Kardinalzahl tokenisiert.</p> 		<p>Ebene getrennt betrachten zu können, wird die Interpunktion bei Kardinalzahlen weggelassen.</p>
	dipl	clean	norm
	I.	I.	1
	Loth	Loth	Lot
	1818.	1818.	1818
	<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>		Norm: immer als mehrere Token
2.5	 <p>Satzinterpunktion wie Kommata, Punkte, Virgeln oder Semikola werden als jeweils eigenständige Token realisiert.</p>		
	dipl	clean	norm

	finden	finden	finden
	.	.	.
2.6	Fußnotenreferenzierungen werden in dipl- und clean-Ebene als mehrere Token und in der norm-Ebene als ein Token realisiert		Norm: ein Token
	dipl	clean	norm
	(*)	(*)	(* **)
	* *)	* *)	
3.1	Formen die in phonologischer Hinsicht fnhd. oder dialektalen Lautstand aufweisen werden in der norm den (standard)nhd Entsprechungen angepasst		
	dipl	clean	norm
	Hellenpein	Hellenpein	Höllenpein
	wänlin	wänlin	Wännlein
3.2	Fnhd. Flexionsformen werden durch die nhd. Entsprechungen ersetzt.		
	dipl	clean	norm
	in die Nasen das kal Haupt	in die Nasen das kal Haupt	in die Nase das kahle Haupt
	Es erfolgt keine Anpassung des Genus		
	dipl	clean	norm
	das Milz	das Milz	das Milz

3.3	Es erfolgt keine morphosyntaktische Anpassung (z.B. innerhalb der Nominalphrase)		
	dipl	clean	norm
	(es heilt) die versehrte Därmelein	die versehrte Därmelein	die versehrte Därmelein
3.4	Ausgestorbene Wortbildungsmorpheme werden nach Möglichkeit durch entsprechende nhd. Bildungen ersetzt. (siehe Beispieltabelle am Schluss)		
	dipl	clean	norm
	halben stachelecht	halben stachelecht	halber stachelig
3.5	Ausgestorbenes lexikalisches Material wird mit modernen Orthographieregeln übernommen.		
	dipl	clean	norm
	Vergefz	Vergefz	Vergess
3.6	Graphematisch wird grundsätzlich nach dem Duden normalisiert.		

Fußnoten und Marginalien

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

Beispiele für die Referenzierung von Fußnoten

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

dipl	ref	ref_target
.		
Paltinaten		
49)	ref	#f1
Wapp		

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

Tabelle SEQ Tabelle * ARABIC 1: Fußnotenreferenzierung

Zeichentabelle

Beschreibung	Zeichen	Eingabe über die Tastatur
Schräger Doppelbindestrich	≈	2E17
Langes kleines „s“	ſ	017F Alt s
Tilde	~	0303
Ligatur ae	æ	00E6
Ligatur AE	Æ	00C6
Ligatur oe	œ	0153
Ligatur OE	Œ	0152
Akut	´	dead key + <Vokal>
Gravis	`	shift+dead key+ <Vokal>
Cedille klein	ç	00E7
Cedille groß	Ç	00C7
Superskribiertes „e“	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	⊙	2299

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	für-	vor-
	<i>fürnehmlich</i>	<i>vornehmlich</i>
3	-für	-vor
	herfür	hervor
4	fnhd. etwan	nhd. <i>etwa</i> (Adverb) nhd. <i>etwas</i> (Pronomen)
5	-en	-Ø
	<i>endlichen</i>	<i>endlich</i>
	<i>sonsten</i>	<i>sonst</i>
	<i>selbsten</i>	<i>selbst</i>
6	-Ø	-en
	<i>ob</i>	<i>oben</i>
	<i>beizeit</i>	<i>beizeiten</i>
7	(Nicht)Normalisierung von Kasus	
	<i>mit Normalisierung: Man trinke des wassers</i>	mit Normalisierung von Kasus: <i>Man trinke das Wasser</i>
	<i>Man trinke des wassers</i>	ohne Normalisierung von Kasus: <i>Man trinke des Wassers</i>
8	Vom Nhd. abweichende starke oder schwache Verbalflexion bzw. Rückumlaut wird dem nhd. Stand angepasst	
	<i>gennent</i>	<i>genannt</i>
	<i>gebauen</i>	<i>gebaut</i>

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

Normalisierung

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

Erstellung von „clean“

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

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

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

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

Benutzte Tools:

Makros

SearchAndMerge.bas

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

ReplacePosOfUnclear.bas

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

Python-Skript

cleanV22.py

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

Shell-Skript

normalizeDipl.sh

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