

Inhaltsverzeichnis

1	Korpus.....	4
1.1	Typ: <i>Corpus</i> – Ridges Herbiology.....	4
2	Dokumente.....	6
2.1	Typ: Documents.....	6
3	Annotationsebenen – Transkription/Normalisierung.....	16
3.1	dipl.....	16
3.2	clean.....	18
3.3	norm.....	19
4	Annotationsebenen – Annotationen zu linguistischen Eigenschaften.....	20
4.1	pos.....	20
4.2	pos_klein.....	21
4.3	lemma.....	23
4.4	hyperlemma.....	24
4.5	foreign.....	25
4.6	foreign_trans.....	26
4.7	lang.....	27
4.8	Verbposition.....	28
4.9	Nebensatztyp.....	29
4.10	KOUS_Semantik.....	30
4.11	komp.....	32
4.12	komp_orth.....	33
4.13	prot.....	34
4.14	attr_gen.....	36
4.15	strD.....	37
4.16	personename.....	38
4.17	bemerkung.....	39
4.18	werkname.....	40
4.19	krankheitsname.....	41

4.20	form_krankheit.....	42
4.21	kraeutername.....	43
4.22	kraeutername_normiert	44
4.23	sprache_kraeutername.....	46
4.24	form_kraeutername.....	47
4.25	kraeuterzubereitung.....	48
4.26	form_zubereitung.....	49
4.27	nomen	50
4.28	form_nomen.....	51
5	Annotationsebenen – Strukturelle Annotation.....	52
5.1	lb.....	52
5.2	brace.....	53
5.3	brace_dir	54
5.4	p.....	56
5.5	p_n.....	57
5.6	p_rend	58
5.7	pb.....	59
5.8	pb_n.....	60
5.9	pb_rend	61
5.10	pb_ana.....	63
5.11	div1 – div5	64
5.12	div1_type – div5_type	65
5.13	div1_n – div5_n	67
5.14	unclear.....	68
5.15	atLeast	69
5.16	atMost	71
5.17	interpretation	72
5.18	figure	73
5.19	figure_rend.....	74
5.20	hi.....	76
5.21	hi_font	77
5.22	hi_rend	78
5.23	head.....	79
5.24	head_n.....	80
5.25	head_rend	81

5.26	note	83
5.27	ref	84
5.28	ref_target.....	85
5.29	ref_type	86
5.30	quote	87
5.31	list	89
5.32	list_type.....	90
5.33	item	91
5.34	xml_id.....	92
6	Annotationsebenen – Inhaltliche Annotation.....	93
6.1	definition	93
6.2	disease.....	94
6.3	term.....	95
6.4	author_ref	96
6.5	reader_ref.....	97
6.6	plant.....	98
6.7	property.....	100
6.8	name.....	101
6.9	name_type.....	102
7	Referenzen	104
8	Anhang.....	105
8.1	Transkription und Normalisierung	105
8.2	Fußnoten und Marginalien.....	113
8.3	Zeichentabelle	114
9	Dokumentation der Probleme während Normalisierung und Konsistenzkorrektur eines historischen Korpus – Ridges v.3.....	115
9.1	Normalisierung.....	115
9.2	Erstellung von „clean“	115
9.3	Konsistenzkorrektur	115
10	Benutzte Tools:.....	116
10.1	Makros.....	116
10.2	Python-Skripte.....	116

1 Korpus

1.1 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, Doreen Christen, Mascha Dayal, Antonia Dittberner, Cora Döhn, Imke Driemel, Olja Efremova, Gill-Maria Eichhorn, Judith Esser, Annegret Gerlach, Linda Giesel, Hagen Hirschmann, Sebastian Kiraga, Ewa Anna Kolbik, Kornél Kovács, Daisy Krüger, Anna-Maria Lehmann, Maria Lober, Anke Lüdeling, Laura Lueders, Samuele Maniscalco, Manuel Metzsig, Alexander Meyer, Sandra Müller, Vinzent Müller, Andrew Murphy, Johannes Mursell, Carolin Odebrecht, Akiko Okuda, Laura Perlitz, Katharina Reinig, Ina Riesler, Lena Rosin, Franz-Josef Sachse, Anna Sapronova, Simon Sauer, Claudia Schmidt, Iryna Sorokovska, Uwe Springmann, Kristina Stephan, Juliane Tiemann, Anna Tóth, Réka Tóth, Alexander Turtureanu, Vivian Voigt, Juliana Wekel, Karolina Zuchewicz (nachfolgend Studenten, Humboldt-Universität zu Berlin genannt)
Veröffentlichung:	30.09.2014
Anzahl der Texte:	29
Zeitraum:	1487-1870
diplToken:	154267
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. 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: 4.1

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

Art der Änderung: Für:
GartDerGesundheit_1487
Neue Annotationsebenen wurden hinzugefügt und zwar folgende: personename, werkname, krankheitsname, form_krankheit, krautername, krautername_normiert, sprache_krautername, form_krautername, krauterzubereitung, form_zubereitung, nomen, form_nomen, bemerkung

Für alle Dokumente:
Neue Annotationen wurden hinzugefügt und zwar folgende: attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

Alte Annotationen wurden korrigiert und zwar folgende: figure_rend, hi_rend, p_rend, pb_rend, pb_ana

Enthaltende Dokumente (Kürzel):
GartDerGesundheit_1487
ArzneiderKreutter_1532
ContrafaytKreuterbuch_1532
NewKraeuterbuch_1543
WieSichMeniglich_1552
Paradeiszgärtlein_1588
AlchemistischePraktik_1603
HortulusSanitatis_1609
KraeutterbuchCarrichter_1609
PflantzGart-c4_1639
PflantzGart_1639
PflantzGart-Vorrede_1639
Wund-Artzney_1652
ThesaurusSanitatis_1673
SonderbaresKraeuterbuch-I-II_1675
SonderbaresKraeuterbuch-II-21_1675
DerSchweizerischeBotanicus_1687
FloraSaturnizans_1722
MysterivmSigillorvm_1735
EinleitungZuDerKraeuterkenntnis_1764
Unterricht_1774

GrundrissKraeuterkunde_1792
 Flora-6_1821
 Flora-7_1821
 DieEigenschaftenAllerHeilpflanzen_1828
 NochEinigeWorte_1840
 VorlesungenUeberKraeuterkunde_1843
 DeutschePflanzennamen_1870
 FloraDerPreussischenRheinlande_1870

2 Dokumente

Alle Titel werden, den Standards der 'clean'-Normalisierung gemäß, nach Originalschreibweise realisiert, wobei Sonderzeichen, die der modernen Orthografie nicht entsprechen an den modernen Zeichensatz angepasst werden.

2.1 Typ: Documents

2.1.1 1870-FloraDerPreussischenRheinlande

Titel: Flora Der Preussischen Rheinlande
 Kürzel: FloraDerPreussischenRheinlande_1870
 Autor: Wirtgen, Philipp Wilhelm
 Jahr: 1870
 Ort: Bonn
 Verlag: NA
 Seitenangabe: 1-22
 Publikationstyp: Monographie
 dipl-Token: 7539
 Beschreibung: Kräutertext
 Annotationsebenen: clean, dipl, figure, head, hi, hi_font, hi_rend, lb, norm, note, p, pb, pb_n, quote, ref, ref_target, xml_id, author_ref, foreign, hyperlemma, lang, lemma, plant, pos, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.2 1870-DeutschePflanzennamen

Titel: Deutsche Pflanzennamen
 Kürzel: DeutschePflanzennamen_1870
 Autor: Grassmann, Hermann
 Jahr: 1870
 Ort: Stettin
 Verlag: NA
 Seitenangabe: 1-23
 Publikationstyp: Monographie
 dipl-Token: 10283
 Beschreibung: Kräutertext
 Annotationsebenen: atLeast, atMost, author_ref, clean, definition, dipl, div1, div1_type, div2, div2_type, foreign, foreign_trans, head, head_n, hi, hi_font, hyperlemma, lang, lb, lemma, name, name_type, norm, p, pb, pb_n, pos, property, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.3 1843-VorlesungenUeberKraeuterkunde

Titel: Vorlesungen Über Kräuterkunde
Kürzel: VorlesungenUeberKraeuterkunde_1843
Autor: Link , Heinrich Friedrich
Jahr: 1843
Ort: Berlin
Verlag: Verlag von C. G. Lüderitz
Seitenangabe: 28-48
dipl-Token: 7995
Beschreibung: Kräutertext
Annotationsebenen: author_ref, clean, dipl, foreign, foreign_trans, head, hi, hi_font, hi_rend, hyperlemma, lang, lb, lemma, norm, note, p, pb, pb_n, plant, pos, reader_ref, ref, ref_target, xml_id, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.4 1840-NochEinigeWorte

Titel: Noch einige Worte über den Befruchtungsakt und die Poljembryonie bei den höheren Pflanzen
Kürzel: NochEinigeWorte_1840
Autor: Meyen, Franz Julius Ferdinand
Jahr: 1840
Ort: Berlin
Verlag: Haude und Spenersche Buchhandlung (S. J. Joseephy.)
Seitenangabe: 24-50
Publikationstyp: Monographie
dipl-Token: 8042
Beschreibung: Kräutertext
Annotierte Ebenen: author_ref, clean, dipl, div1, div1_type, div2, div2_type, foreign, foreign_trans, head, head_rend, hi, hi_font, hi_rend, lb, lemma, name, name_type, norm, note, p, p_rend, pb, pb_n, pos, property, quote, reader_ref, ref, ref_target, ref_type, term, xml_id, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.5 1828-DieEigenschaftenAllerHeilpflanzen

Titel: Die Eigenschaften Aller Heilpflanzen
Kürzel: DieEigenschaftenAllerHeilpflanzen_1828
Autor: Sauer
Jahr: 1828
Ort: München
Verlag: NA
Seitenangabe: 3-21
dipl-Token: 6849
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, clean, dipl, disease, foreign, head, hi, hi_font, hyperlemma, interpretation, lang, lb, lemma, norm, p, pb, pb_n, plant, pos, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.6 1821-Flora-7

Titel: Flora oder Botanische Zeitung 7
Kürzel: Flora-7_1821
Autor: Sieber
Jahr: 1821
Ort: Regensburg
Verlag: Königl. botanische Gesellschaft
Seitenangabe: 97-112
Publikationstyp: Zeitschrift
dipl-Token: 3636
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, div1, div1_type, div2, div2_type, div3, div3_type, div4, div4_type, div5, div5_type, foreign, foreign_trans, head, head_n, hi_font, interpretation, lang, lb, lemma, norm, p, pb, pb_n, pos, property, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.7 1821-Flora-6

Titel: Flora oder Botanische Zeitung 6
Kürzel: Flora-6_1821
Autor: B. Wilbrand
Jahr: 1821
Ort: Regensburg
Verlag: Königl. botanische Gesellschaft
Seitenangabe: 81-96
Publikationstyp: Zeitschrift
dipl-Token: 3311
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, brace, brace_dir, clean, definition, dipl, div1, div1_type, div2, div2_type, foreign, foreign_trans, head, head_n, hi_font, interpretation, lang, lb, lemma, norm, note, p, pb, pb_n, pos, property, reader_ref, ref, ref_target, ref_type, term, unclear, xml_id, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.8 1792-GrundrissKraeuterkunde

Titel: Grundriss der Kräuterkunde zu Vorlesungen entworfen
Kürzel: GrundrissKraeuterkunde_1792
Autor: Willdenow, Carl Ludwig
Jahr: 1792
Ort: Berlin
Verlag: Haude und Spener
Seitenangabe: 282-305
Publikationstyp: Monographie
dipl-Token: 5094
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, div1, div1_type, div2, div2_n, div2_type, div3, div3_n, foreign, foreign_trans, head, hi, hi_font, hi_rend, interpretation, lang, lb, lemma, norm, p, pb, pb_n, pos, property, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.9 1774-Unterricht

Titel: Unterricht von der allgemeinen Kräuter- und Wurzelrocknung
Kürzel: Unterricht_1774
Autor: Eisen, Johann Georg
Jahr: 1774
Ort: Riga
Verlag: NA
Seitenangabe: 44-72
Publikationstyp: Monographie
dipl-Token: 3982
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, definition, dipl, div1, div1_type, div2, div2_type, div3, foreign, foreign_trans, head, head_n, hi, hi_font, hyperlemma, interpretation, item, lang, lb, lemma, list, list_type, norm, p, pb, pb_n, pos, property, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.10 1764-EinleitungZuDerKraeuterkenntnis

Titel: Einleitung Zu Der Kräuterkenntnis
Kürzel: EinleitungZuDerKraeuterkenntnis_1764
Autor: Oeder , George Christian
Jahr: 1764
Ort: Kopenhagen
Seitenangabe: 304-339
dipl-Token: 7115
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, foreign, head, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, p, pb, pb_n, plant, pos, reader_ref, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.11 1735-MysterivmSigillorvm

Titel: Mysterivm Sigillorvm
Kürzel: MysterivmSigillorvm_1735
Autor: Hiebner von Schneebergk, Israel
Jahr: 1735
Ort: Erfurt
Seitenangabe: 65-89
dipl-Token: 7864
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, disease, foreign, foreign_trans, head, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, p, pb, pb_n, plant, pos, quote, reader_ref, ref, ref_target, unclear, xml_id, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.12 1722-FloraSaturnizans

Titel: Flora saturnizans, Die Verwandtschaft des Pflanzen mit dem Mineral Reich
Kürzel: FloraSaturnizans_1722
Autor: Henckel, Johann Friedrich
Jahr: 1722
Ort: Leipzig
Verlag: Johann Christian Martini
Seitenangabe: 647-671
Publikationstyp: Monographie
dipl-Token: 6219
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, definition, dipl, div1, div1_type, div2, div2_type, div3, div3_type, foreign, head, head_n, hi, hi_font, interpretation, lang, lb, lemma, norm, note, p, pb, pos, property, reader_ref, ref, ref_target, ref_type, term, unclear, xml_id, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.13 DerSchweizerischeBotanicus_1687

Titel: Der Schweizerische Botanicus
Kürzel: V4-1687- DerSchweizerischeBotanicus
Autor: P. Timothei
Jahr: 1687
Ort: NA
Verlag: Franz Carl Roß
Seitenangabe: 54-79
Publikationstyp: Monographie
dipl-Token: 3510
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, head, hi_font, hi_rend, interpretation, lb, lemma, norm, pb, pb_n, pos, reader_ref, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition
Anmerkung: Aus der Originaltranskription der Studenten wurden 2 Seiten gelöscht, diese waren in der Transkription zugrunde liegenden PDF doppelt vorhanden und wurden von den Studenten auch doppelt annotiert.

2.1.14 1675-SonderbaresKraeuterbuch-II-21

Titel: Curioser Botanicus oder sonderbares Kräuter Buch
Kürzel: SonderbaresKraeuterbuch-II-21_1675
Autor: NA
Jahr: 1675
Ort: NA
Verlag: NA
Seitenangabe: II-21
Publikationstyp: Monographie
dipl-Token: 1907
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, clean, dipl, div1, div1_type, div2, div3, div3_type,

foreign, head, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, p, p_n, pb, pb_n, pos, property, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.15 1675-SonderbaresKraeuterbuch-I-II

Titel: Curioser Botanicus oder sonderbares Kräuter Buch
Kürzel: SonderbaresKraeuterbuch-I-II_1675
Autor: NA
Jahr: 1675
Ort: NA
Verlag: NA
Seitenangabe: I-II
Publikationstyp: Monographie
dipl-Token: 2319
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, clean, dipl, div1, div1_type, div2, div2_type, foreign, foreign_trans, head, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, p, p_n, pb, pb_ana, pb_n, pos, property, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.16 1673-ThesaurusSanitatis

Titel: Thesaurus Sanitatis
Kürzel: ThesaurusSanitatis_1673
Autor: Nasser, Adrian
Jahr: 1673
Ort: Nürnberg
Verlag: Johann Hoffmann
Seitenangabe: 497-524
Publikationstyp: Monographie
dipl-Token: 7041
Beschreibung: Kräutertext
Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, disease, foreign, foreign_trans, head, hi, hi_font, hyperlemma, interpretation, lang, lb, lemma, norm, p, pb, pb_n, plant, pos, reader_ref, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.17 1652-Wund-Artzney

Titel: Wund-Artzney
Kürzel: Wund-Artzney_1652
Autor: Hildani, Guilhelmi Fabricii
Jahr: 1652
Ort: Frankfurt am Main
Verlag: Friderich Greiffen
Seitenangabe: 218-223
Publikationstyp: Monographie

dipl-Token: 5252
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, disease, foreign, head, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, p, pb, pb_n, plant, pos, reader_ref, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.18 1639-PflantzGart-Vorrede

Titel: Pflantz-Gart (Vorrede)
 Kürzel: PflantzGart-Vorrede_1639
 Autor: Rhagor, Daniel
 Jahr: 1639
 Ort: Bern
 Verlag: Ben Stephan Schmid
 Seitenangabe: 1-10
 Publikationstyp: Monographie
 dipl-Token: 2319
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, disease, foreign, head, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, p, pb, pb_n, plant, pos, reader_ref, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.19 1639-PflantzGart-c4

Titel: Pflantz-Gart (Capitel 4)
 Kürzel: PflantzGart-c4_1639
 Autor: Rhagor, Daniel
 Jahr: 1639
 Ort: Bern
 Verlag: Ben Stephan Schmid
 Seitenangabe: 33-45
 Publikationstyp: Monographie
 dipl-Token: 2845
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, div1, div1_type, div2, div2_type, div3, div3_type, div4, div5, foreign, foreign_trans, head, head_n, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, p, p_rend, pb, pb_n, pos, property, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.20 1639-PflantzGart

Titel: Pflantz-Gart
 Kürzel: PflantzGart_1639
 Autor: Rhagor, Daniel
 Jahr: 1639
 Ort: Bern
 Verlag: Ben Stephan Schmid

Seitenangabe: 92-110
 Publikationstyp: Monographie
 dipl-Token: 4060
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, clean, dipl, figure, head, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, pb, pb_n, pos, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.21 1609-KraeuterbuchCarrichter

Titel: Kräuterbuch des Edelen und hochgelehrten herren
 Kürzel: PflantzGart_1639
 Autor: Carrichter, Bartholomei
 Jahr: 1609
 Ort: Straßburg
 Verlag: Antonium Bertram
 Seitenangabe: 47-75
 Publikationstyp: Monographie
 dipl-Token: 4992
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, head, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, pb, pb_n, pos, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.22 1609-HortulusSanitatis

Titel: Hortulus Sanitatis
 Kürzel: HortulusSanitatis_1609
 Autor: Durante, Castore; Peter Uffenbach
 Jahr: 1609
 Ort: Frankfurt am Main
 Verlag: Iona Rhodij
 Seitenangabe: 1-21
 Publikationstyp: Monographie
 dipl-Token: 6516
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, disease, foreign, foreign_trans, head, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, p, pb, pb_n, plant, pos, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.23 1603-AlchemistischePraktik

Titel: Alchemistische Praktik
 Kürzel: AlchemistischePraktik_1603
 Autor: Libavius, Andreas
 Jahr: 1603
 Ort: Frankfurt
 Verlag: Johann Saurm

Seitenangabe: 4-26
 Publikationstyp: Monographie
 dipl-Token: 5492
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, definition, dipl, div2, div2_n, div2_type, div3, div3_n, div3_type, div4, div4_type, figure, figure_rend, foreign, foreign_trans, head, head_n, hi, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, p, pb, pb_n, pb_rend, pos, property, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.24 1588-Paradeiszgärtlein

Titel: Paradeiszgärtlein
 Kürzel: Paradeiszgärtlein_1588
 Autor: Rosbachivm, Conradvm; u.a.
 Jahr: 1588
 Ort: Frankfurt am Main
 Verlag: Johann Spieß
 Seitenangabe: 1-43
 Publikationstyp: Monographie
 dipl-Token: 5056
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, figure, head, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, pb, pb_n, pos, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.25 1557-WieSichMeniglich

Titel: Wie sich meniglich
 Kürzel: WieSichMeniglich_1557
 Autor: von Bodenstein, Adam
 Jahr: 1557
 Ort: Basel
 Verlag: Bartholome Stähälin
 Seitenangabe: 28-47
 Publikationstyp: Monographie
 dipl-Token: 5520
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, disease, foreign, foreign_trans, head, hi_font, hyperlemma, interpretation, lang, lb, lemma, norm, p, pb, pb_n, plant, pos, reader_ref, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.26 1543-NewKraeuterbuch

Titel: New Kreüterbuch
 Kürzel: NewKraeuterbuch_1543
 Autor: Fuchs, Leonhard
 Jahr: 1543
 Ort: Basel
 Verlag: Michael Isingrin

Seitenangabe: 2-e4
 Publikationstyp: Monographie
 dipl-Token: 5272
 Beschreibung: Kräutertext
 Annotierte Ebenen: author_ref, clean, definition, dipl, div2, div2_type, div3, div3_type, figure, figure_rend, foreign, foreign_trans, head, hi_font, hyperlemma, lang, lb, lemma, norm, p, p_n, p_rend, pb, pb_n, pb_rend, pos, property, term, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.27 1532-ContrafaytKreuterbuch

Titel: Contrafayt kreüterbuch
 Kürzel: ContrafaytKreuterbuch_1532
 Autor: Brunnfelsz, Otho
 Jahr: 1532
 Ort: Straßburg
 Verlag: Hans Schotten
 Seitenangabe: 312-323
 Publikationstyp: Monographie
 dipl-Token: 4372
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, figure, head, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, pb, pb_n, pos, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.28 1532-ArtzneyDerKreutter

Titel: Artzney Buchlein der kreutter
 Kürzel: ArtzneyDerKreutter_1532
 Autor: Tallat, Johannes
 Jahr: 1532
 Ort: Leipzig
 Verlag: Michael Blum
 Seitenangabe: NA
 Publikationstyp: Monographie
 dipl-Token: 4630
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, head, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, note, pb, pos, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition

2.1.29 1487-GartDerGesundheit

Titel: Gart der Gesundheit
 Kürzel: GartDerGesundheit_1487
 Autor: von Cuba, Johannes

Jahr: 1487
 Ort: Ulm
 Verlag: NA
 Seitenangabe: NA
 Publikationstyp: Monographie
 dipl-Token: 4701
 Beschreibung: Kräutertext
 Annotierte Ebenen: atLeast, atMost, author_ref, clean, dipl, figure, head, hi_font, hi_rend, hyperlemma, interpretation, lang, lb, lemma, norm, pb, pos, reader_ref, term, unclear, attr_gen, komp, komp_orth, KOUS_Semantik, Nebensatztyp, pos_klein, prot, strD, Verbposition, personename, werkname, krankheitsname, form_krankheit, kraeutername, kraeutername_normiert, sprache_kraeutername, form_kraeutername, kraeuterzubereitung, form_zubereitung, nomen, form_nomen, bemerkung

3 Annotationsebenen – Transkription/Normalisierung

3.1 dipl

3.1.1 Typ: Layer - dipl

Name: dipl
 Dokumentation: Siehe Anhang Transkription und Normalisierung
 dipl-Token: 153732
 Beschreibung: 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. Grundsätzlich kann so auch entgegen modernen Orthographieregeln segmentiert oder transkribiert werden!

3.1.2 Typ: Preparationstep – dipl

Schritt: 1
 Aufbereitung: OCR/Transkript
 Aufbereitungsart: manuell
 Tool: NA
 Format: Text-Datei
 Segmentationsbasis der Annotation: eigenständige Segmentierung
 Qualitätsprüfer: NA
 Datum: 2013
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Erfassung des Transkriptes in TEI xml
 Aufbereitungsart: manuell
 Tool: Oxygen
 Format: TEI p5 subset
 Segmentationsbasis der Annotation: eigenständige Segmentierung
 Qualitätsprüfer: NA

Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	3
Aufbereitung:	Tokenisierung
Aufbereitungsart:	Automatisch
Tool:	Treetagger Version 3.2
Format:	Treetaggeroutput
Segmentationsbasis der Annotation:	eigenständige Segmentierung
Qualitätsprüfer:	LAUDATIO
Datum:	2013/14
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	4
Aufbereitung:	Mapping der TEI Elemente in Excel als Spalten
Aufbereitungsart:	semi-automatisch
Tool:	Perl Script
Format:	Excel
Segmentationsbasis der Annotation:	Tokenisiertes TEI xml
Qualitätsprüfer:	LAUDATIO
Datum:	2013/14
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	4
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	5
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

3.2 clean

3.2.1 Typ: Layer – clean

Name:	clean
Dokumentation:	Siehe Ridges-Website http://korpling.german.hu-berlin.de/ridges/index_de.html
clean-Token:	149310
Beschreibung:	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. Grafische Markierungen der einzelnen Texte wie Zeilenumbrüche werden aufgelöst und Sonderzeichen einiger Fonts wie Fraktur auf die heutigen Schriftsätze übertragen.

3.2.2 Typ: Preparationstep – clean

Schritt:	1
Aufbereitung:	Normalisierung, Annotation
Aufbereitungsart:	Automatisch
Tool	cleanV2.py
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	Vivian Voigt
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

3.3 norm

3.3.1 Typ: Layer – norm

Name: norm
Dokumentation: Siehe Anhang Transkription und Normalisierung
norm-Token: 148871
Beschreibung: Die **norm**-Ebene stellt einen weiteren Normalisierungsschritt dar, indem hier die Tokenisierung und die Orthografie einheitlich nach modernen Orthografierregeln (vgl. Duden) sowie automatisch angepasst werden. Unbekanntes lexikalisches Material wurde basierend auf graphematischen Regeln an die moderne Orthographie angepasst.

3.3.2 Typ: Preparationstep – norm

Schritt: 1
Aufbereitung: Normalisierung, Annotation
Aufbereitungsart: Manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: eigenständige Segmentierung
Qualitätsprüfer: LAUDATIO
Datum: 2013/14
Qualitätsprüfung: Manuelle Konsistenzprüfung
Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

4 Annotationsebenen – Annotationen zu linguistischen Eigenschaften

4.1 pos

4.1.1 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.

4.1.2 Typ: Preparationstep – pos

Schritt:	1
Aufbereitung:	Part-of-Speech-Tagging
Aufbereitungsart:	automatisch
Tool:	Treetagger Version 3.2
Format:	Treetagger-output
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	Semi-automatische Konsistenzprüfung mit DECCA (Dickinson & Meurers 2003)
Editor:	LAUDATIO

Schritt:	2
Aufbereitung:	unclear-POS-Tagging
Aufbereitungsart:	automatisch
Tool:	unclearPosTagging.bas
Format:	Excel 2010
Segmentationsbasis der Annotation:	pos
Qualitätsprüfer:	NA
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm

Annotation:
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.1.3 Typ: Annotationlayer – pos

Annotationsart:	Tokenannotation
Beschreibung:	Wortarten-Annotation nach dem Stuttgart-Tübingen-Tagset. Alle in norm als „unknown“ getaggten Token wurden in pos nachträglich mit „XY“ getaggt.

4.1.4 Typ: Annotationvalue – pos

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

4.2 pos_klein

4.2.1 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.

4.2.2 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: xls, xlsx
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: Hagen Hirschmann
 Datum: 2014
 Qualitätsprüfung: manuelle Korrektur
 Editor: Hagen Hirschmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

4.2.3 Typ: Annotationlayer – pos_klein

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

4.2.4 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

4.3.1 Typ: Layer – lemma

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

4.3.2 Typ: Preparationstep – lemma

Schritt: 1
Aufbereitung: Lemmatisierung
Aufbereitungsart: Automatisch
Tool: Treetagger Version 3.2
Format: treetaggeroutput
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.3.3 Typ: Annotationlayer– lemma

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

4.3.4 Typ: Annotationvalue – lemma

Wert:	Wertbeschreibung:
Normalisiertes Lemma	Lemmatisierung.

4.4 hyperlemma

4.4.1 Typ: Layer - hyperlemma

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

4.4.2 Typ: Preparationstep – hyperlemma

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2013/14
 Qualitätsprüfung: Manuelle Konsistenzprüfung
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch

Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.4.3 Typ: *Annotationlayer* – hyperlemma

Annotationsart:	Tokenannotation
Beschreibung:	In Fällen, in denen semantisch ein Begriff wie <i>Heumonat</i> nicht mehr transparent ist, wird ein modernes Äquivalent, z. B. <i>Heumonat – Juli</i> , angegeben.

4.4.4 Typ: *Annotationvalue* – hyperlemma

Wert:	Wertbeschreibung:
Text	Semantisches Äquivalent.

4.5 foreign

4.5.1 Typ: *Layer* – foreign

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

4.5.2 Typ: *Preparationstep* – foreign

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: NA
 Datum: 2013
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

4.5.3 Typ: *Annotationlayer* – foreign

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

4.5.4 Typ: *Annotationvalue* – foreign

Wert:	Wertbeschreibung:
foreign	Allgemeiner Tag für fremdsprachliches Material.

4.6 foreign_trans

4.6.1 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.

4.6.2 Typ: *Preparationstep* – foreign_trans

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: 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: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

4.6.3 Typ: *Annotationlayer* – foreign_trans

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

4.6.4 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.

4.7 lang

4.7.1 Typ: Layer – lang

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

4.7.2 Typ: Preparationstep – lang

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: dipl
Qualitätsprüfer: NA
Datum: 2013
Qualitätsprüfung: NA
Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

4.7.3 Typ: Annotationlayer – lang

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

4.7.4 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

4.8 Verbposition

4.8.1 Typ: Layer – Verbposition

Name: Verbposition
Dokumentation: NA

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

4.8.2 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: xls, xlsx
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: Hagen Hirschmann
Datum: 2014
Qualitätsprüfung: manuelle Korrektur
Editor: Hagen Hirschmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML

Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

4.8.3 Typ: Annotationlayer – Verbposition

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

4.8.4 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
Vletz	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
Vi	Verberststellung in einem gegebenen untergeordneten Satz mit subordinierender Konjunktion; analysiert als Token-Feature auf der Basis von pos=KOUS

4.9 Nebensatztyp

4.9.1 Typ: Layer – Nebensatztyp

Name:	Nebensatztyp
Dokumentation:	NA

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

4.9.2 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: xls, xlsx
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: Hagen Hirschmann
 Datum: 2014
 Qualitätsprüfung: manuelle Korrektur
 Editor: Hagen Hirschmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.9.3 Typ: Annotationlayer – Nebensatztyp

Annotationsart:	Tokenannotation
Beschreibung:	Annotation der Nebensatztypen.

4.9.4 Typ: Annotationvalue – Nebensatztyp

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

4.10 KOUS_Semantik

4.10.1 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.

4.10.2 **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:	xls, xlsx
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	Hagen Hirschmann
Datum:	2014
Qualitätsprüfung:	manuelle Korrektur
Editor:	Hagen Hirschmann

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

4.10.3 **Typ:** Annotationlayer – KOUS_Semantik

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

4.10.4 **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

4.11 komp

4.11.1 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.

4.11.2 Typ: Preparationstep – komp

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: Excel 2010
Format: xls, xlsx
Segmentationsbasis der Annotation: dipl
Qualitätsprüfer: Laura Perlitz
Datum: 30.09.2014
Qualitätsprüfung: manuelle Korrektur
Editor: Laura Perlitz

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch

Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.11.3 Typ: Annotationlayer – komp

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

4.11.4 Typ: Annotationvalue – komp

Wert:	Wertbeschreibung:
k	Kompositum

4.12 komp_orth

4.12.1 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.

4.12.2 Typ: Preparationstep – komp_orth

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: Excel 2010
 Format: xls, xlsx
 Segmentationsbasis der Annotation: komp
 Qualitätsprüfer: Laura Perlitz
 Datum: 30.09.2014
 Qualitätsprüfung: manuelle Korrektur
 Editor: Laura Perlitz

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.12.3 Typ: Annotationlayer – komp_orth

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

4.12.4 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)

4.13 prot

4.13.1 Typ: Layer – prot

Name: prot
 Dokumentation: Der Prototyp 1 (prot1) markiert alle Token, die eindeutig als Komposita 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 fäfi*).

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

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

4.13.2 Typ: Preparationstep – prot

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: Excel 2010
 Format: xls, xlsx
 Segmentationsbasis der Annotation: komp
 Qualitätsprüfer: Laura Perlitz
 Datum: 30.09.2014
 Qualitätsprüfung: manuelle Korrektur
 Editor: Laura Perlitz

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.13.3 Typ: Annotationlayer – prot

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

4.13.4 Typ: Annotationvalue – prot

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

4.14 attr_gen

4.14.1 Typ: Layer – attr_gen

Name:	attr_gen
Dokumentation:	Das Bezugsnommen 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.

4.14.2 Typ: Preparationstep – attr_gen

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	Excel 2010
Format:	xls, xlsx
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	Laura Perlitz
Datum	30.09.2014
Qualitätsprüfung:	manuelle Korrektur
Editor:	Laura Perlitz
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS

Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.14.3 Typ: Annotationlayer – attr_gen

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

4.14.4 Typ: Annotationvalue – attr_gen

Wert:	Wertbeschreibung:
gprä	pränominales Genitivattribut
gpost	postnominales Genitivattribut

4.15 strD

Typ: *Layer* – strD

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

4.15.1 Typ: Preparationstep – strD

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: Excel 2010
 Format: xls, xlsx
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Laura Perlitz
 Datum: 30.09.2014
 Qualitätsprüfung: manuelle Korrektur
 Editor: Laura Perlitz

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.15.2 Typ: Annotationlayer – strD

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

4.15.3 Typ: Annotationvalue – strD

Wert:	Wertbeschreibung:
strD	strukturelle Destruktion

4.16 personenname

Typ: *Layer* – strD

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

4.16.1 Typ: Preparationstep – personenname

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl

Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch

Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.16.2 Typ: Annotationlayer – personenname

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.16.3 Typ: Annotationvalue – personenname

Wert:	Wertbeschreibung:
String	Personennamen.

4.17 bemerkung

Typ: *Layer* – bemerkung

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

4.17.1 Typ: Preparationstep – bemerkung

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.17.2 Typ: Annotationlayer – bemerkung

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.17.3 Typ: Annotationvalue – bemerkung

Wert:	Wertbeschreibung:
String	Bemerkung.

4.18 werkname

Typ: *Layer* – werkname

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

4.18.1 Typ: Preparationstep – werkname

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014

Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.18.2 Typ: Annotationlayer – werkname

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.18.3 Typ: Annotationvalue – werkname

Wert:	Wertbeschreibung:
String	Werkname.

4.19 krankheitsname

Typ: *Layer* – krankheitsname

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

4.19.1 Typ: Preparationstep – krankheitsname

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA

Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.19.2 Typ: Annotationlayer – krankheitsname

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.19.3 Typ: Annotationvalue – krankheitsname

Wert:	Wertbeschreibung:
String	Krankheitsname.

4.20 form_krankheit

Typ: *Layer* – werkname

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

4.20.1 Typ: Preparationstep – form_krankheit

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: dipl
Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
Datum: 30.09.2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

4.20.2 Typ: Annotationlayer – form_krankheit

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.20.3 Typ: Annotationvalue – form_krankheit

Wert:	Wertbeschreibung:
String	NA

4.21 krautername

Typ: *Layer* – krautername

Name: krautername
Dokumentation:

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

4.21.1 **Typ:** Preparationstep – kraeutername

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: dipl
Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
Datum: 30.09.2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

4.21.2 **Typ:** Annotationlayer – kraeutername

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.21.3 **Typ:** Annotationvalue – kraeutername

Wert:	Wertbeschreibung:
String	Kräutername.

4.22 kraeutername_normiert

Typ: *Layer* – kraeutername_normiert

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

4.22.1 **Typ:** Preparationstep – kraeutername_normiert

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: dipl
Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
Datum: 30.09.2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

4.22.2 **Typ:** Annotationlayer – kraeutername_normiert

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.22.3 **Typ:** Annotationvalue – kraeutername_normiert

Wert:	Wertbeschreibung:
String	Normierter Kräuternamen.

4.23 sprache_kraeuternamen

Typ: *Layer* – sprache_kraeuternamen

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

4.23.1 Typ: Preparationstep – sprache_kraeuternamen

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: dipl
Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
Datum: 30.09.2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

4.23.2 Typ: Annotationlayer – sprache_kraeuternamen

Annotationsart:	Spannenannotation
-----------------	-------------------

Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.
---------------	--

4.23.3 Typ: Annotationvalue – sprache_kraeutername

Wert:	Wertbeschreibung:
String	Sprache des Kräuternamens.

4.24 form_kraeutername

Typ: *Layer* – form_kraeutername

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

4.24.1 Typ: Preparationstep – form_kraeutername

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.24.2 Typ: Annotationlayer – form_kraeutername

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.24.3 Typ: Annotationvalue – form_kraeutername

Wert:	Wertbeschreibung:
String	NA

4.25 kraeuterzubereitung

Typ: *Layer* – kraeuterzubereitung

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

4.25.1 Typ: Preparationstep – kraeuterzubereitung

Schritt: 1
Aufbereitung: Linguistische Annotation
Aufbereitungsart: manuell
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: dipl
Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
Datum: 30.09.2014
Qualitätsprüfung: NA
Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO

Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.25.2 Typ: Annotationlayer – kraeuterzubereitung

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.25.3 Typ: Annotationvalue – kraeuterzubereitung

Wert:	Wertbeschreibung:
String	NA

4.26 form_zubereitung

Typ: *Layer* – form_zubereitung

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

4.26.1 Typ: Preparationstep – form_zubereitung

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch

Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.26.2 Typ: Annotationlayer – form_zubereitung

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.26.3 Typ: Annotationvalue – form_zubereitung

Wert:	Wertbeschreibung:
String	NA

4.27 nomen

Typ: *Layer* – nomen

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

4.27.1 Typ: Preparationstep – nomen

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA

Editor: LAUDATIO
 Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

4.27.2 Typ: Annotationlayer – nomen

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.27.3 Typ: Annotationvalue – nomen

Wert:	Wertbeschreibung:
String	NA

4.28 form_nomen

Typ: *Layer* – form_nomen

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

4.28.1 Typ: Preparationstep – form_nomen

Schritt: 1
 Aufbereitung: Linguistische Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: Anke Lüdeling, Uwe Springmann
 Datum: 30.09.2014
 Qualitätsprüfung: NA
 Editor: Anke Lüdeling, Uwe Springmann

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML

Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

4.28.2 Typ: Annotationlayer – form_nomen

Annotationsart:	Spannenannotation
Beschreibung:	Nur in GartDerGesundheit_1487 annotiert.

4.28.3 Typ: Annotationvalue – form_nomen

Wert:	Wertbeschreibung:
String	NA

5 Annotationsebenen – Strukturelle Annotation

5.1 lb

5.1.1 Typ: Layer – lb

Name:	lb
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.

5.1.2 Typ: Preparationstep – lb

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	NA
Format:	Excel 2010

Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum:	2013
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.1.3 Typ: Annotationlayer – lb

Annotationsart:	Spannenannotation
Beschreibung:	Markierung einer ganzen Zeile.

5.1.4 Typ: Annotationvalue – lb

Wert:	Wertbeschreibung:
lb	Markierung pro ganzer Zeile.

5.2 brace

5.2.1 Typ: Layer – brace

Name:	brace
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.

5.2.2 Typ: Preparationstep – brace

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	31.05.2012
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.2.3 Typ: Annotationlayer – brace

Annotationsart:	Spannenannotation
Beschreibung:	Markierung einer Klammer, die Text über mehrere Zeilen zusammenfasst.

5.2.4 Typ: Annotationvalue – brace

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

5.3 brace_dir

5.3.1 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.

5.3.2 Typ: Preparationstep – brace_dir

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: NA
 Datum: 31.05.2012
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.3.3 Typ: Annotationlayer – brace_dir

Annotationsart:	Spannenannotation
Beschreibung:	Richtung der Klammer

5.3.4 Typ: Annotationvalue – p

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

5.4 p

5.4.1 Typ: Layer – p

Name:	p
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.

5.4.2 Typ: Preparationstep – p

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	31.05.2012
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.4.3 Typ: Annotationlayer – p

Annotationsart:	Spannenannotation
Beschreibung:	Markierung eines ganzen Paragraphen.

5.4.4 Typ: Annotationvalue – p

Wert:	Wertbeschreibung:
P	Markierung pro ganzem Paragraphen.

5.5 p_n

5.5.1 Typ: Layer – p_n

Name: p_n
 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.

5.5.2 Typ: Preparationstep – p_n

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: NA
 Datum: 31.05.2012
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS

Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.5.3 Typ: Annotationlayer – p_n

Annotationsart:	Spannenannotation
Beschreibung:	Markiert die Nummerierung eines Paragraphen.

5.5.4 Typ: Annotationvalue – p_n

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

5.6 p_rend

5.6.1 Typ: Layer – p_rend

Name: p_rend
 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.

5.6.2 Typ: Preparationstep – p_rend

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: Manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: NA
 Datum: 31.05.2012
 Qualitätsprüfung: NA

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO

Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.6.3 Typ: Annotationlayer – p_rend

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibung der Darstellung eines Paragraphen.

5.6.4 Typ: Annotationvalue – p_rend

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

5.7 pb

5.7.1 Typ: Layer – pb

Name: pb
 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.

5.7.2 Typ: Preparationstep – pb

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: LAUDATIO
 Datum: 2013
 Qualitätsprüfung: Manuelle Konsistenzprüfung
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.7.3 Typ: Annotationlayer – pb

Annotationsart:	Spannenannotation
Beschreibung:	Markierung einer ganzen Seite.

5.7.4 Typ: Annotationvalue – pb

Wert:	Wertbeschreibung:
pb	Markierung pro ganzer Seite.

5.8 pb_n

5.8.1 Typ: Layer – pb_n

Name: pb_n
 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.

5.8.2 Typ: Preparationstep – pb_n

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA

Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: LAUDATIO
 Datum: 2013
 Qualitätsprüfung: Manuelle Konsistenzprüfung
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.8.3 Typ: Annotationlayer – pb_n

Annotationsart:	Spannenannotation
Beschreibung:	Markierung der Nummerierung einer Seite.

5.8.4 Typ: Annotationvalue – pb_n

Wert:	Wertbeschreibung:
numerischer Wert	Seitenzahl.

5.9 pb_rend

5.9.1 Typ: Layer – pb_rend

Name: pb_rend
 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.

5.9.2 Typ: *Preparationstep* – pb_rend

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum:	2013
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.9.3 Typ: *Annotationlayer* – pb_rend

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

5.9.4 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.
vonDrachenwurtz.Cap.III.	Seite/n ist/sind einer Überschrift zugeordnet.
ohlZuMachen.	Seite/n ist/sind einer Überschrift zugeordnet.
zumBeltenZuDistilliren.	Seite/n ist/sind einer Überschrift zugeordnet.
wallerAußKräuternVndDergleichen	Seite/n ist/sind einer Überschrift zugeordnet.
auffsBelftZuDistilliren.	Seite/n ist/sind einer Überschrift zugeordnet.
außKräuternVndDergleichen	Seite/n ist/sind einer Überschrift zugeordnet.
wallerAußKräuternVndDergleichen	Seite/n ist/sind einer Überschrift zugeordnet.
amBeltenZuDistilliren.	Seite/n ist/sind einer Überschrift zugeordnet.

5.10 pb_ana

5.10.1 Typ: Layer – pb_ana

Name:	pb_ana
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.

5.10.2 Typ: Preparationstep – pb_ana

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA

Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.10.3 Typ: Annotationlayer – pb_ana

Annotationsart:	Spannenannotation
Beschreibung:	Interpretation von vermutlich nicht korrekten Seitenzahlangaben.

5.10.4 Typ: Annotationvalue – pb_ana

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

5.11 div1 – div5

5.11.1 Typ: Layer – div1 – div5

Name:	div1, div2, div3, div4, div5
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.

5.11.2 Typ: Preparationstep – div1 – div5

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell

Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.11.3 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/relANNIS Version angezeigt. Gilt pro Ebene (div1-div5).

5.11.4 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

5.12.1 Typ: *Layer* – div1_type – div5_type

Name:	div1_type, div2_type, div3_type, div4_type, div5_type
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.

5.12.2 Typ: *Preparationstep* – div1_type – div5_type

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.12.3 Typ: *Annotationlayer* – div1_type – div5_type

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibt den Typ/die Art des Kapitels/Unterkapitels. Die Einteilung könne 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).

5.12.4 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	
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

5.13.1 Typ: *Layer* – div2_n – div3_n

Name:	div2_n, div_3
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.

5.13.2 Typ: *Preparationstep* – div2_n – div3_n

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2

Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.13.3 Typ: *Annotationlayer* – div2_n – div3_n

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

5.13.4 Typ: *Annotationvalue* – div1_n – div5_n

Wert:	Wertbeschreibung:
Ziffer	Nummerierung der Sektion.

5.14 unclear

5.14.1 Typ: *Layer* - unclear

Name: unclear
 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.

5.14.2 Typ: *Preparationstep* – unclear

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der dipl

Annotation:
 Qualitätsprüfer: LAUDATIO
 Datum: 2013
 Qualitätsprüfung: Manuelle Konsistenzprüfung. Der Layer **reason**, der eine Begründung für unleserliche Elemente beschreiben sollte, wurde entfernt.
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.14.3 Typ: *Annotationlayer* – unclear

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

5.14.4 Typ: *Annotationvalue* – unclear

Wert:	Wertbeschreibung:
unclear	<p>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).</p> <p>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.</p>

5.15 atLeast

5.15.1 Typ: *Layer* – atLeast

Name: atLeast

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.

5.15.2 Typ: *Preparationstep* – atLeast

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	Manuell
Tool:	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum:	31.05.2012
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.15.3 Typ: *Annotationlayer* – atLeast

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

5.15.4 Typ: *Annotationvalue* – atLeast

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

5.16 atMost

5.16.1 Typ: Layer – atMost

Name:	atMost
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.

5.16.2 Typ: Preparationstep – atMost

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Datum:	2013
Qualitätsprüfer:	LADUATIO
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.16.3 Typ: Annotationlayer – atMost

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

5.16.4 Typ: Annotationvalue – atMost

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

5.17 interpretation

5.17.1 Typ: Layer– interpretation

Name:	interpretation
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.

5.17.2 Typ: Preparationstep – interpretation

Schritt:	1
Aufbereitung:	Linguistische Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum:	2013
Qualitätsprüfung:	Manuelle Konsistenzprüfung
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch

Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.17.3 Typ: *Annotationlayer*– interpretation

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

5.17.4 Typ: *Annotationvalue*– interpretation

Wert:	Wertbeschreibung:
Text	Angabe der Interpretation.

5.18 figure

5.18.1 Typ: *Layer* – figure

Name: figure
 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.

5.18.2 Typ: Preparationstep – figure

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: NA
 Datum: 2013
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm

Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.18.3 Typ: Annotationlayer – figure

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

5.18.4 Typ: Annotationvalue – figure

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

5.19 figure_rend

5.19.1 Typ: Layer – figure_rend

Name:	figure_rend
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.

5.19.2 Typ: Preparationstep – figure_rend

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool:	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.19.3 Typ: *Annotationlayer* – figure_rend

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

5.19.4 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

5.20.1 Typ: Layer – hi

Name:	hi
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.

5.20.2 Typ: Preparationstep – hi

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.20.3 Typ: Annotationlayer – hi

Annotationsart:	Spannen
Beschreibung:	Markierung eines irgendwie hervorgehobenen Bereichs.

5.20.4 Typ: Annotationvalue – hi

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

5.21 hi_font

5.21.1 Typ: Layer – hi_font

Name:	hi_font
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.

5.21.2 Typ: Preparationstep – hi_font

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum	2013
Qualitätsprüfung:	Manuelle Konsistenzprüfung: Korrektur der false positives, Auffüllung leerer Zellen mit der Standardschriftart des jeweiligen Textes
Editor:	Studenten, Humboldt-Universität zu Berlin, LAUDATIO
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der	norm

Annotation:
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.21.3 Typ: *Annotationlayer* – hi_font

Annotationsart:	Tokenannotation
Beschreibung:	Weist jedem Token eine Schriftart zu.

5.21.4 Typ: *Annotationvalue* – hi_font

Wert:	Wertbeschreibung:
antiqua	Text, der in Antiqua gedruckt ist.
gothic	Text, der in Gothic gedruckt ist.

5.22 hi_rend

5.22.1 Typ: *Layer* – hi_rend

Name: hi_rend
 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.

5.22.2 Typ: *Preparationstep* – hi_rend

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: LAUDATIO
 Datum: 31.05.2012
 Qualitätsprüfung: Manuelle Konsistenzprüfung: Korrektur der false positives
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm

Annotation:
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm

Annotation:
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.22.3 Typ: *Annotationlayer* – hi_rend

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

5.22.4 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:lem	Ein größerer Abstand zwischen Buchstaben, Sperrdruck.

5.23 head

5.23.1 Typ: *Layer* – head

Name: head
 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.

5.23.2 Typ: *Preparationstep* – head

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA

Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.23.3 Typ: Annotationlayer – head

Annotationsart:	Spannenannotation
Beschreibung:	Markierung einer Überschrift.

5.23.4 Typ: Annotationvalue – head

Wert:	Wertbeschreibung:
head	Allgemeine Markierung einer Überschrift.

5.24 head_n

5.24.1 Typ: Layer – head_n

Name:	head_n
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.

5.24.2 Typ: Preparationstep – head_n

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	Dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.24.3 Typ: Annotationlayer – head_n

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

5.24.4 Typ: Annotationvalue – head_n

Wert:	Wertbeschreibung:
Ziffer	Konkrete Nummerierung der Überschrift.

5.25 head_rend

5.25.1 Typ: Layer – head_rend

Name:	head_rend
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.

5.25.2 Typ: *Preparationstep* – head_rend

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.25.3 Typ: *Annotationlayer* – head_rend

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

5.25.4 Typ: *Annotationvalue* – head_rend

Wert:	Wertbeschreibung:
brown	Überschrift in brauner Farbe.

5.26 note

5.26.1 Typ: Layer – note

Name:	note
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.

5.26.2 Typ: Preparationstep – note

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	LAUDATIO
Datum:	2013
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.
Editor:	Studenten, Humboldt-Universität zu Berlin

Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.26.3 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.

5.26.4 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. Für Beispiele siehe 8.2
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

5.27.1 Typ: Layer – ref

Name:	ref
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.

5.27.2 Typ: Preparationstep – ref

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin

Schritt: 2

Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.27.3 Typ: Annotationlayer – ref

Annotationsart:	Spannenannotation
Beschreibung:	Referenz einer Fußnote.

5.27.4 Typ: Annotationvalue – ref

Wert:	Wertbeschreibung:
ref	Referenz einer Fußnote.

5.28 ref_target

5.28.1 Typ: Layer – ref_target

Name: ref_target
 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.

5.28.2 Typ: Preparationstep – ref_target

Schritt: 1
 Aufbereitung: Strukturelle Annotation
 Aufbereitungsart: manuell
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl

Annotation:
 Qualitätsprüfer: NA
 Datum: 2013
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.28.3 Typ: *Annotationlayer* – ref_target

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

5.28.4 Typ: *Annotationvalue* – ref_target

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

5.29 ref_type

5.29.1 Typ: *Layer* – ref_type

Name: ref_type
 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.

5.29.2 Typ: *Preparationstep* – ref_type

Schritte:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	2013
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.29.3 Typ: *Annotationlayer* – ref_type

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

5.29.4 Typ: *Annotationvalue* – ref_type

Wert:	Wertbeschreibung:
noteAnchor	Art der Referenz.

5.30 quote

5.30.1 Typ: *Layer* – quote

Name:	quote
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.

5.30.2 Typ: Preparationstep – quote

Schritt:	1
Aufbereitung:	Strukturelle Annotation
Aufbereitungsart:	manuell
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.30.3 Typ: Annotationlayer – quote

Annotationsart:	Spannenannotation
Beschreibung:	Markiert ein Zitat im Text.

5.30.4 Typ: Annotationvalue – quote

Wert:	Wertbeschreibung:
Quote	Ein Zitat im Text.

5.31 list

5.31.1 Typ: Layer – list

Name:	list
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.

5.31.2 Typ: Preparationstep – list

Aufbereitung:	Strukturelle Annotation
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.31.3 Typ: Annotationlayer – list

Annotationsart:	Spannenannotation
Beschreibung:	Eine Liste von Elementen.

5.31.4 Typ: Annotationvalue – list

Wert:	Wertbeschreibung:
List	Allgemeine Markierung einer Liste.

5.32 list_type

5.32.1 Typ: Layer – list_type

Name:	list_type
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.

5.32.2 Typ: Preparationstep – list_type

Aufbereitung:	Strukturelle Annotation
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

5.32.3 Typ: *Annotationlayer* – list_type

Annotationsart:	Spannenannotation
Beschreibung:	Beschreibung der Art der Liste.

5.32.4 Typ: *Annotationvalue* – list_type

Wert:	Wertbeschreibung:
Simple	Einfache Aufzählung.

5.33 item

5.33.1 Typ: Layer – item

Name:	item
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.

5.33.2 Typ: Preparationstep – item

Aufbereitung:	Strukturelle Annotation
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm

Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.33.3 Typ: Annotationlayer – item

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

5.33.4 Typ: Annotationvalue – item

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

5.34 xml_id

5.34.1 Typ: Layer – xml_id

Name: xml_id
 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.

5.34.2 Typ: Preparationstep – xml_id

Aufbereitung: Strukturelle Annotation
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: dipl
 Qualitätsprüfer: NA
 Datum: NA
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS

Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

5.34.3 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.

5.34.4 Typ: Annotationvalue –xml_id

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

6 Annotationsebenen – Inhaltliche Annotation

6.1 definition

6.1.1 Typ: Layer – definition

Name: definition
 Dokumentation: 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.

6.1.2 Typ: Preparationstep – definition

Aufbereitung: Inhaltliche Annotation
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: 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: Studenten, Humboldt-Universität zu Berlin
 Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML

Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

6.1.3 Typ: *Annotationlayer* – definition

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

6.1.4 Typ: *Annotationvalue* – definition

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

6.2 disease

6.2.1 Typ: *Layer* – disease

Name: disease
 Dokumentation: 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.

6.2.2 Typ: *Preparationstep* – disease

Aufbereitung: Inhaltliche Annotation
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der: norm
 Annotation:
 Qualitätsprüfer: NA
 Datum: NA
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch

Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

6.2.3 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. <i>der Bauch, der weh tut</i> oder <i>Bauchweh</i> .

6.2.4 Typ: Annotationvalue – disease

Wert:	Wertbeschreibung:
di	Benennung einer Krankheit

6.3 term

6.3.1 Typ: Layer – term

Name: term
 Dokumentation: 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.

6.3.2 Typ: Preparationstep – term

Aufbereitung: Inhaltliche Annotation
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: NA
 Datum: NA
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

6.3.3 Typ: Annotationlayer – term

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

6.3.4 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 author_ref

6.4.1 Typ: Layer – author_ref

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

6.4.2 Typ: Preparationstep – author_ref

Aufbereitung: Inhaltliche Annotation
 Tool: NA
 Format: Excel 2010
 Segmentationsbasis der Annotation: norm

Annotation:
 Qualitätsprüfer: NA
 Datum: NA
 Qualitätsprüfung: NA
 Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
 Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
 Aufbereitungsart: automatisch
 Tool: PAULA
 Format: PAULA-XML
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

Schritt: 3
 Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
 Aufbereitungsart: automatisch
 Tool: ANNIS
 Format: relANNIS
 Segmentationsbasis der Annotation: norm
 Qualitätsprüfer: LAUDATIO
 Datum: 2014
 Qualitätsprüfung: NA
 Editor: LAUDATIO

6.4.3 Typ: *Annotationlayer* – author_ref

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

6.4.4 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.

6.5 reader_ref

6.5.1 Typ: *Layer* – reader_ref

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

6.5.2 Typ: *Preparationstep* – reader_ref

Aufbereitung:	Inhaltliche Annotation
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

6.5.3 Typ: *Annotationlayer* – reader_ref

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

6.5.4 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.

6.6 plant

6.6.1 Typ: *Layer* – plant

Name:	plant
Dokumentation:	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.

6.6.2 Typ: *Preparationstep* – plant

Aufbereitung:	Inhaltliche Annotation
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	31.05.2012
Qualitätsprüfung:	Manuelle Konsistenzprüfung.
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

6.6.3 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. <i>das Kraut, das hinter dem Bahnhof wächst</i> oder <i>Kamille</i> .

6.6.4 Typ: *Annotationvalue* – plant

Wert:	Wertbeschreibung:
pl	Benennung einer Pflanze.

6.7 property

6.7.1 Typ: *Layer* – property

Name: property
Dokumentation: 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.

6.7.2 Typ: *Preparationstep* – property

Aufbereitung: Inhaltliche Annotation
Tool: NA
Format: Excel 2010
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 31.05.2012
Qualitätsprüfung: Manuelle Konsistenzprüfung.
Editor: Studenten, Humboldt-Universität zu Berlin

Schritt: 2
Aufbereitung: Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart: automatisch
Tool: PAULA
Format: PAULA-XML
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

Schritt: 3
Aufbereitung: Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart: automatisch
Tool: ANNIS
Format: relANNIS
Segmentationsbasis der Annotation: norm
Qualitätsprüfer: LAUDATIO
Datum: 2014
Qualitätsprüfung: NA
Editor: LAUDATIO

6.7.3 Typ: *Annotationlayer* – property

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

6.7.4 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.

6.8 name

6.8.1 Typ: Layer – name

Name:	name
Dokumentation:	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.

6.8.2 Typ: Preparationstep – name

Aufbereitung:	Inhaltliche Annotation
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

6.8.3 Typ: Annotationlayer – name

Annotationsart:	Spannenannotation
Beschreibung:	Markiert Eigennamen.

6.8.4 Typ: Annotationvalue – name

Wert:	Wertbeschreibung:
Name	Allgemeine Markierung von Eigennamen.

6.9 name_type

6.9.1 Typ: Layer – name_type

Name:	name_type
Dokumentation:	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.

6.9.2 Typ: Preparationstep – name_type

Aufbereitung:	Inhaltliche Annotation
Tool	NA
Format:	Excel 2010
Segmentationsbasis der Annotation:	dipl
Qualitätsprüfer:	NA
Datum:	NA
Qualitätsprüfung:	NA
Editor:	Studenten, Humboldt-Universität zu Berlin
Schritt:	2
Aufbereitung:	Konvertierung von Excel 2010 nach PAULA-XML
Aufbereitungsart:	automatisch
Tool:	PAULA
Format:	PAULA-XML
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO
Schritt:	3
Aufbereitung:	Konvertierung von PAULA-XML nach relANNIS
Aufbereitungsart:	automatisch
Tool:	ANNIS
Format:	relANNIS
Segmentationsbasis der Annotation:	norm
Qualitätsprüfer:	LAUDATIO
Datum:	2014
Qualitätsprüfung:	NA
Editor:	LAUDATIO

6.9.3 Typ: Annotationlayer – name_type

Annotationsart:	Spannenannotation
Beschreibung:	Spezifiziert den Typ der Eigennamen.

6.9.4 Typ: *Annotationvalue* – name_type

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

7 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.

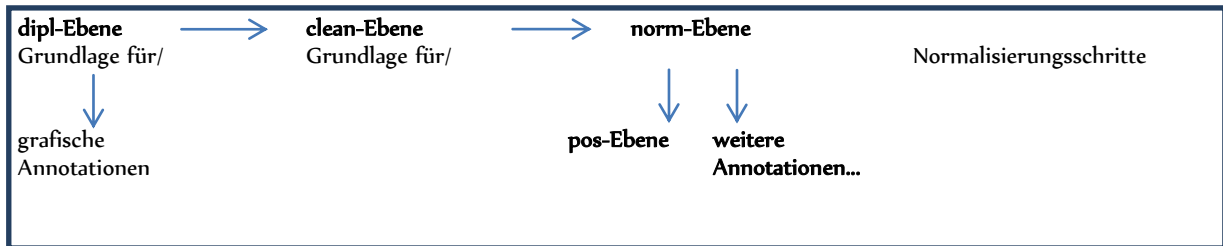
Krause, Thomas, Lüdeling, Anke, Odebrecht, Carolin & Amir Zeldes (2012). Multiple Tokenizations in a Diachronic Corpus. *Exploring Ancient Languages through Corpora Conference (EALC), 14.-16.Juni 2012. Oslo, Norway*.

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.

8 Anhang

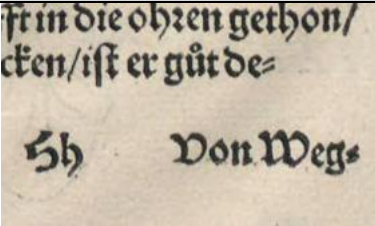



8.1 Transkription und Normalisierung



Die Transkription von Faksimiles stellt für die korpuslinguistische Aufbereitung zumeist die grundlegende, diplomatische Ebene (**dipl**). Diese Ebene soll sich grafisch hinsichtlich Orthografie, Getrennt- und Zusammenschreibung und Sonderzeichen möglichst nah am zugrunde liegenden Faksimile orientieren. Grundsätzlich kann so auch entgegen moderner Orthographieregeln segmentiert oder transkribiert werden! 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. Grafische 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 und die Orthografie einheitlich nach modernen Orthografiereregeln (vgl. Duden) 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.


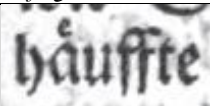




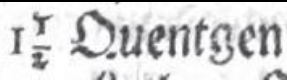
Nachfolgend werden die Richtlinien zur Transkription und Normalisierung im Einzelnen gegenübergestellt. Allgemein für jede Segmentierungsebene sowie Annotationsebene gilt, dass keine Leerzeilen oder Leerzeichen enthalten sein dürfen. Dazu wird für jeden Fall ein Beispiel aus dem Ridges Herbiology Version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

Nr.	dipl-Ebene	clean-Ebene	norm-Ebene
	Transkript des Faksimiles.	Erfolgt manuell im Format .xls. Tipp: Kopieren Sie sich die <u>dipl</u> -Ebene und verändern Sie dann die entsprechenden Stellen!	Erfolgt manuell im .xls Format. Tipp: Kopieren Sie sich die <u>clean</u> -Ebene und verändern Sie dann die entsprechenden Stellen!
1.0	 <p>Oft werden die ersten Zeichen bzw. ersten Wörter (im Bsp. <i>Von Wege</i>) einer neuen Seite bei einem Seitenumbruch in der vorhergehenden Seite in einem eigenen Absatz/in einer eigenen Zeile doppelt realisiert. Diese Zeichen bzw. das Wort wird nicht mit in das Transkript aufgenommen.</p>		
1.1	 <p>Das Zeichen für morphologische Worttrennung (oder Zeilenumbruch) wird mit U 2E17¹ übernommen.</p>	Die morphologische Trennung durch „f“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt. Komposita, die in der dipl-Ebene mit einem „f“ realisiert werden, behalten dieses Zeichen bei, um die allgemeine Suche nach Komposita in dieser Ebene zu ermöglichen. Für Erläuterungen zur rein graphischen Trennung, wie Zeilenumbrüchen, siehe Nummer 2.1.	Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschrieben. (Grundlage hierfür bildet der aktuelle Stand des Dudens .)
	<p>Dipl</p> <p>Dan̄ reiẒ</p>	<p>clean</p> <p><i>Dann-reisz</i></p>	<p>norm</p> <p><i>Tannenreis</i></p>
1.2	 <p>Das „f“ wird mit U 017F übernommen.</p>	Das lange „f“ wird durch ein reguläres „s“ ersetzt.	
	<p>Dipl</p> <p><i>Lefer</i></p>	<p>clean</p> <p><i>Leser</i></p>	<p>norm</p> <p><i>Leser</i></p>
1.3	<p>Gerade Längenzeichen oder Makrone über Konsonanten werden mit U 0304 übernommen.²</p>  <p>Längenzeichen werden mit U 0304 übernommen, wenn nicht klar ist, ob es sich um Längen- oder Makron handelt, wird zugunsten des Makrons entschieden.</p>	Längenzeichen werden entsprechend ihrer Funktion umgesetzt. Die kann u. a. aus Reichmann & Wegera (1993) entnommen werden. ³	
	<p>dipl</p> <p>Dan̄ reiẒ</p>	<p>clean</p> <p><i>Dann-reisz</i></p>	<p>norm</p> <p><i>Tannenreis</i></p>

¹ Sonderzeichen wie zum Beispiel das „f“ werden mit Hilfe von Unicodes in der dipl-Ebene realisiert. Eine Liste der zu verwendenden Unicodes ist angefügt.

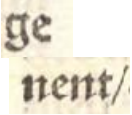



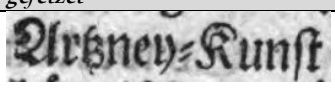
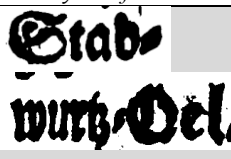
² Für alle Vokale gelten unäre Hex-Codes: Für „ā“ gilt U 0101 und für „Ā“ U 0100, für „ē“ U 0113 und für „Ē“ U 0112, für „ō“ U 014D und „Ō“ U 014C und für „ū“ U 016B und für „Ū“ U 016A.



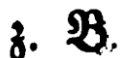
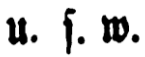
³ Nachweis für das Beispiel: Reichmann & Wegera 1993: 31.

1.4		Die Ligatur wird analog zu „sz“ aufgelöst.	Nach modernen Orthografieregeln wird für „sz“ das „ß“ verwendet.
	Wie auch u.a. in Beispiel 1.2 wird die Ligatur mit dem langen „f“ und „z“ übernommen.		
	dipl <i>fleißige</i>	clean <i>fleiszige</i>	norm <i>fleißige</i>
1.5		Umlaute mit superskribiertem „e“ werden wie moderne Umlaute des Deutschen realisiert.	
	Umlaute mit superskribiertem „e“ werden mit Vokal + U 0364 übernommen.		
	dipl <i>häuffte</i>	clean <i>häuffte</i>	norm <i>häufte</i>
1.6		Die Ligatur wird aufgelöst.	
	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.		
	dipl <i>hæc</i>	clean <i>haec</i>	norm <i>haec</i>
1.7		Dieses diakritische Zeichen wird in der clean-Ebene nicht mehr realisiert und durch den Segmentationsbasis der Annotationvokal ersetzt.	
	Dieses diakritische Zeichen wird mit U 016F in der dipl-Ebene übernommen. Für „ä“ gilt U 00E5.		
	Dipl <i>Güt</i>	clean <i>gut</i>	norm <i>gut</i>
1.8		Das kaufmännische Und wird übernommen.	Das kaufmännische Und wird übernommen.
	Das kaufmännische Und „&“ wird mit U 0026 übernommen.		
	Dipl &	clean &	norm &
1.9		Alle Akzente werden übernommen.	Alle Akzente werden übernommen.
	Alle Akzente werden übernommen. Hierbei ist zu beachten, dass der jeweilige unäre Hex-Code zu verwenden ist. ⁴ Für „ó“ gilt U 00F3.		
	Dipl <i>Quòd</i>	clean <i>quòd</i>	norm <i>quòd</i>
1.10		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		

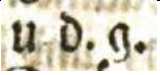

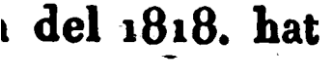
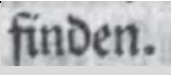
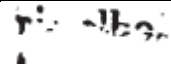
⁴ Dazu siehe Zeichentabelle im Anhang.



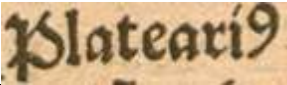


	segmentiert, um gemischte Brüche transparent zu halten.		
	Dipl	clean	norm
	1	1	1
	1/2	1/2	1/2
1.11	 <p>Dieses Symbol wird mit dem U 2299 übernommen.</p>	Dieses Symbol wird mit dem U 2299 übernommen.	Dieses Symbol wird mit dem U 2299 übernommen.
	Dipl	clean	norm
	⊙	⊙	⊙
1.12	 <p>Dies bedeutet et c.</p>		
	Dipl	clean	norm
	et. c.	et. c.	etc.
1.13	<p>diefe: haben</p> <p>Senkrechtstriche werden mit „ “ übernommen.</p>	Senkrechtstriche werden übernommen.	Senkrechtstriche werden übernommen.
	Dipl	clean	norm
	<i>diefe</i>	<i>diese</i>	<i>diese</i>
	/	/	/
	<i>Haben</i>	<i>haben</i>	<i>haben</i>
1.14	<p>La&tucis</p> <p>Die ct-Ligatur wird nicht übernommen.</p>	Die ct-Ligatur wird nicht übernommen.	Die ct-Ligatur wird nicht übernommen.
	dipl	clean	norm
	<i>Lactucis</i>	<i>Lactucis</i>	<i>Lactucis</i>
1.15	<p>γλεινός,</p> <p>Schriftsätze des Altgriechischen werden übernommen: Unicode-Tabellen unter http://www.unicode.org/charts/PDF/U0370.pdf sowie http://www.unicode.org/charts/PDF/U1F00.pdf</p>	Schriftsätze werden übernommen.	Schriftsätze werden übernommen.
2.1	<p>allge meinen</p> <p>Von Zeilenumbrüchen betroffene Elemente werden analog zur Primärquelle getrennt tokenisiert. Das Trennungszeichen „≠“ wird mit U 2E17 übernommen.</p>	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

	<i>allge meinen</i>	<i>allgemeinen</i>	<i>allgemeinen</i>
2.2	 <p>Im Falle einer Worttrennung aufgrund von Zeilenumbrüchen, die keine overte, grafische Markierung wie „=-/“ beinhalten, werden formal die getrennten Elemente als jeweils ein Token in der dipl-Ebene betrachtet.</p>	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
	<i>ge</i>	<i>ge</i>	<i>genannt</i>
	<i>nent</i>	<i>nent</i>	
2.3	 <p>Klitika werden als ein Token realisiert.</p>	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
	<i>ichs</i>	<i>ichs</i>	<i>ich</i>
			<i>es</i>
2.4	 <p>Auch entgegen moderner Orthografieregeln wird analog zur Textgrundlage tokenisiert.</p>		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
	<i>zufetzen</i>	<i>zusetzen</i>	<i>zu</i>
			<i>setzen</i>
2.5	 <p>Auch entgegen moderner Orthografieregeln wird analog zur Textgrundlage tokenisiert.</p>		Getrennt- und Zusammenschreibung wird nach modernen Orthografieregeln angepasst. (Grundlage hierfür bildet die aktuelle Ausgabe des Dudens.) Dazu wird die Tokenisierung verändert, indem aus zwei Token in dipl (oder clean) ein Token in der norm-Ebene gemacht wird.
	dipl	clean	norm
	<i>zusammen</i>	<i>zusammen</i>	<i>zusammengesetzt</i>
	<i>gesetzt</i>	<i>gesetzt</i>	
2.6	 <p>Komposita, egal welcher Komplexität oder Bildungsweise, mit Gleichheitszeichen (U 2E17) werden als ein Token realisiert.</p>	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
	<i>Arznei-Kunst</i>	<i>Arznei-Kunst</i>	<i>Arzneikunst</i>
2.7		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	Die Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschrieben. (Grundlage hierfür bildet die

	Von Zeilenumbrüchen betroffene Komposita werden analog zur Primärquelle getrennt tokenisiert, dabei bleibt die Kennzeichnung der morphologischen Worttrennung unberührt.	morphologische Trennung wird analog zu Nummer 1.1 mit einem Minus ersetzt. Das Kompositum wird in einer Spanne zusammengefasst wiedergegeben.	aktuelle Ausgabe des Dudens.)
	dipl	clean	norm
	<i>Stabwurtz</i>	<i>Stabwurtz-Oel</i>	<i>Stabwurtzöl</i>
	<i>wurtz Oel</i>		
2.8	 Auch entgegen moderner Orthografieregeln wird analog zur Textgrundlage tokenisiert.		Wortbildung und Großschreibung, die nicht der modernen Orthografieregeln entsprechen, werden angeglichen.
	dipl	clean	norm
	<i>Spannenlang</i>	<i>Spannen-lang</i>	<i>spannenlang</i>
2.9	 Auch entgegen moderner Orthografieregeln wird analog zur Textgrundlage tokenisiert.	Das lange „l“ wird durch ein reguläres „s“ ersetzt.	Nicht der modernen Orthografie entsprechende Bildungen werden angeglichen.
	dipl	clean	norm
	<i>zufetzen</i>	<i>zusetzen</i>	<i>zu</i>
			<i>setzen</i>
2/3.1	 Abkürzungen werden analog zur Textgrundlage tokenisiert. Das heißt, auch Setzfehler oder Spatien, die nicht modernen Orthografieregeln entsprechen, werden realisiert. ⁵		Abkürzungen werden nach Dudenrichtlinien umgesetzt. Dies hat zur Folge, dass Abkürzungen unterschiedlich tokenisiert werden können (bspw. Abkürzungen als zwei oder mehr Tokens wie <i>u. a. m.</i> vs. solche, die als ein Token realisiert werden, wie <i>usw.</i>). vgl. dazu Nummer 2/3.2 und 2/3.3. In Fällen, in denen keine Dudenrichtlinie zugeordnet werden kann, wird immer nach einem Punkt segmentiert.
	dipl	clean	norm
	<i>z.</i>	<i>z.</i>	<i>z.</i>
	<i>B.</i>	<i>B.</i>	<i>B.</i>
2/3.2	 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 <i>u. a. m.</i> vs. solche, die als ein Token realisiert werden, wie <i>usw.</i>). 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
	<i>u.</i>	<i>u.</i>	<i>usw.</i>
	<i>f.</i>	<i>s.</i>	
	<i>w.</i>	<i>w.</i>	

⁵ Allgemeine Bemerkung zu Abkürzung und ihrer Interpunktion: Die Interpunktion einer Abkürzung wird an das abgekürzte Element angehängt, auch wenn sie gleichzeitig einen Satz beendet.

2/3. 3			Abkürzungen werden nach Dudenrichtlinien umgesetzt. Dies hat zur Folge, dass Abkürzungen unterschiedlich tokenisiert werden können (bspw. Abkürzungen als zwei oder mehr Tokens wie <i>u. a. m.</i> vs. solche, die als ein Token realisiert werden, wie <i>usw.</i>). 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
	<i>u</i>	<i>u</i>	<i>u.</i> <i>d.</i>
	<i>d.g.</i>	<i>d.g.</i>	<i>g.</i>
3.1			Nach modernen Orthografieregeln werden Ordinalzahlen mit Punkt realisiert.
	dipl	clean	norm
	<i>I.</i> <i>Loth</i>	<i>I.</i> <i>Loth</i>	<i>I.</i> <i>Loth</i>
3.2			Um die Kardinalzahlen von Ordinalzahlen in der norm-Ebene getrennt betrachten zu können, wird die Interpunktion bei Kardinalzahlen weggelassen.
	dipl	clean	norm
	<i>1818.</i>	<i>1818.</i>	<i>1818</i>
3.3			
	dipl	clean	norm
	<i>finden</i>	<i>finden</i>	<i>finden</i>
3.4		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
3.5	Das Dicit-Zeichen wird in der dipl-Ebene mit dem "€"-Zeichen umschrieben (Euro-Zeichen).	Es wird in der clean automatisch mit "der" ersetzt	In der norm auch.

	Zwibel od Knob		
	dipl o€	clean oder	norm oder
3.6	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 (* **)
3.7	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.		Norm: immer als mehrere Token
3.8	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 :
3.9	Das Absatzzeichen wird in dipl-, clean- und norm-Ebene mit U+00B6 übernommen. 		
	dipl ¶	clean ¶	norm ¶
3.10	Das Zeichen, das der Ziffer 9 ähnelt und zumeist dazu diente, die Endung "us" zu bezeichnen wird in der dipl durch "\$" repräsentiert und in clean und norm durch "us" ersetzt. 	Ersetzung durch "us".	Ersetzung durch "us".
	dipl Plateari\$	clean Platearius	norm Platearius
3.11	 Dieses Zeichen wird durch vU+0364 realisiert.	'v' wird in der clean-Ebene durch 'ü' ersetzt.	'v' wird in der norm-Ebene durch 'ü' ersetzt.
	dipl 'v'ber	clean über	norm über
3.12	 Das ü mit schrägen Strichen wird durch U+0171 realisiert.	ü mit schrägen Strichen wird in der clean-Ebene durch 'ü' ersetzt.	ü mit schrägen Strichen wird in der norm-Ebene durch 'ü' ersetzt.
	dipl für	clean für	norm für
3.13	'i' mit Akzent wird in der dipl-Ebene mit U+00ED transkribiert und in clean und norm durch 'i' normalisiert.		

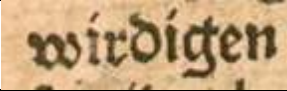
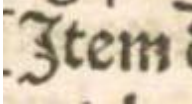
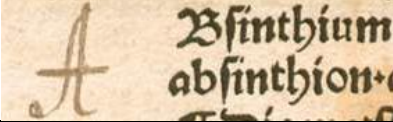
			
	dipl würdigem	clean würdigem	norm Würdigen
3.14	Die Zeichen für 'I' und 'J' sind in Fraktur-Texten graphisch in der Regel nicht voneinander unterscheidbar. Konvention in Ridges ist daher, das Zeichen in der dipl als 'J' darzustellen und erst in der norm-Ebene das Zeichen individuell nach moderner Orthografie zu interpretieren. 	Wie in der dipl-Ebene	Anpassung an moderne Orthografie.
	dipl Jtem	clean Jtem	norm Item
3.15	Handschriftlich Versalien die offensichtlich zum Text gehören (KEINE Anmerkungen oder Kommentare von Lesern), werden mit annotiert. 		
	dipl ABsinthium	clean ABsinthium	norm Absinth

Tabelle 1 Transkriptionsbeispiele

8.2 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.

8.2.1 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	dipl	not	xml_id
.			.		
PAftinaten			49)		
49)	ref	#f1	Paltinaten		
Wapp			find		
			unterschiedener		
			Arten		
			,		
			als		
			Paltinaca		
			Domeltica		
			,		
			fylveltris		
			,		
			marina		
			,		
			Syriac_		
			,		
			man		
			kan		
			Hn.		
			von		
			Franckenu		
			KrauterLexic.		
			voc.		
			Paltinaca		
			nachschlagen		
			,		
			wofelbten		
			der		
			Ulus		
			pharmacaticus		
			weiter		
			verfolget		
			wird		
			.	note	f1

Tabelle 2: Fußnotenreferenzierung

8.3 Zeichentabelle

Beschreibung	Zeichen	Additiver Hex-Code	Unärer Hex-Code
Schräger Doppelbindestrich	≋		2E17
Langes kleines „s“	ſ		017F
Makron	˜	0304 (n̄, m̄)	
a mit Makron	a˜		0101
A mit Makron	A˜		0100
e mit Makron	e˜		0113
E mit Makron	E˜		0112
o mit Makron	o˜		014D
O mit Makron	O˜		014C
u mit Makron	u˜		016B
U mit Makron	U˜		016A
i mit Makron	i˜		012B
I mit Makron	I˜		012A
Ligatur ae	æ		00E6
Ligatur AE	Æ		00C6
Ligatur oe	œ		0153
Ligatur OE	Œ		0152
a mit Akut	á		00E1
a mit Gravis	à		00E0

e mit Akut	è		00E8
e mit Gravis	é		00E9
i mit Akut	í		00ED
i mit Gravis	ì		00EC
o mit Akut	ó		00F3
o mit Gravis	ò		00F2
u mit Akut	ú		00FA
u mit Gravis	ù		00F9
Cedille klein	ç		00E7
Cedille groß	Ç		00C7
Superskribiertes „e“	e	0364	
a mit Kreis	â		00E5
u mit Kreis	û		016F
Absatzzeichen	¶		00B6

Tabelle 3: Zeichentabelle

9 Dokumentation der Probleme während Normalisierung und Konsistenzkorrektur eines historischen Korpus – Ridges v.3

9.1 Normalisierung

Die Annotationsebene <norm> wurde händisch von Studenten erledigt (im Rahmen des Kurses "Historische Korpuslinguistik") und im Anschluss manuell korrigiert.

9.2 Erstellung von „clean“

Die Annotationsebene <clean> wird automatisch erstellt durch das Skript *cleanV2.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.

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

(2) Das normalisierte *falsch* wurde durch Excel zum logischen Operator *FALSCH*

9.3 Konsistenzkorrektur

Die Konsistenzkorrektur wird semi-automatisch mithilfe des DECCA-Skriptes durchgeführt. Dubletten wurden mithilfe des *decca-pos-reduce*-Skriptes, einer Abwandlung des Originalen *decca-pos*-Skriptes, gelöscht. Alle n-Gramme wurden überprüft und korrigiert, wobei, aufgrund des hohen Arbeitsaufwandes die 2- und 1-Gramme nur nach Auffälligkeiten bzw. offensichtlichen Fehlern untersucht wurden, ohne Sichtung des Kontextes. Auf die 3-Gramme wurde zudem ein weiteres Skript (*nonfringe.py*) angewandt, das diejenigen Gramme ausgibt, bei denen das abweichende pos-Tagging nicht am Rand des Strings auftaucht, die Ausgabe dieses Skriptes wurde ausführlich und mit Betrachtung des Kontextes korrigiert, während die Korrektur der übrigen 3-Gramme analog zu den 1- und 2-Grammen ohne Sichtung des Kontextes erfolgte.

10 Benutzte Tools:

10.1 Makros

10.1.1 SucheUndVerbinde.bas

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

10.1.2 Lösche_Leerzeichen.bas

Das Makro Lösche_Leerzeichen.bas sucht und löscht in einem markierten Bereich Spatien am Ende eines Tokens, die laut Annotationsrichtlinien nicht gewünscht sind (ein Wert = ein Token ohne Spatium). Anschließend wird die Zelle, in der Spatien vorhanden waren, rot gefüllt. Somit kann das Makro nachverfolgt werden.

10.2 Python-Skripte

10.2.1 cleanV2.py

Die neuen Texte stellen 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.

10.2.2 DECCA

DECCA ist ein Tool zur semi-automatischen Konsistenzprüfung von Marcus Dickinson und Detmar Meurers, das n-Gramme ausgibt, in denen Zeichenketten der Länge n ausgegeben werden, die lexikalisch gleich sind, aber unterschiedlich getaggt wurden. Bei der Konsistenzprüfung von Ridges orientierten wir uns an den Heuristiken von Dickinson und Meurers 2003, der Workflow erfolgte, wie in 9.3 beschrieben.