|  |  |
| --- | --- |
| Dokumentation des Korpus | Ridges Herbology Version 4.1  auf Grundlage des Datenmodells/ Metadatenframeworks nach LAUDATIO  Stand 19.03.2015  -deutsch- |
|  |  |

**Inhaltsverzeichnis**

[1 Korpus 4](#_Toc414620990)

[1.1 Typ: *Corpus* – Ridges Herbology 4](#_Toc414620991)

[2 Dokumente 6](#_Toc414620992)

[2.1 Typ: Documents 6](#_Toc414620993)

[3 Annotationsebenen – Transkription/Normalisierung 16](#_Toc414620994)

[3.1 dipl 16](#_Toc414620995)

[3.2 clean 18](#_Toc414620996)

[3.3 norm 19](#_Toc414620997)

[4 Annotationsebenen – Annotationen zu linguistischen Eigenschaften 20](#_Toc414620998)

[4.1 pos 20](#_Toc414620999)

[4.2 pos\_klein 21](#_Toc414621000)

[4.3 lemma 23](#_Toc414621001)

[4.4 hyperlemma 24](#_Toc414621002)

[4.5 foreign 25](#_Toc414621003)

[4.6 foreign\_trans 26](#_Toc414621004)

[4.7 lang 27](#_Toc414621005)

[4.8 Verbposition 28](#_Toc414621006)

[4.9 Nebensatztyp 30](#_Toc414621007)

[4.10 KOUS\_Semantik 31](#_Toc414621008)

[4.11 komp 32](#_Toc414621009)

[4.12 komp\_orth 33](#_Toc414621010)

[4.13 prot 35](#_Toc414621011)

[4.14 attr\_gen 36](#_Toc414621012)

[4.15 strD 37](#_Toc414621013)

[4.16 personenname 39](#_Toc414621014)

[4.17 bemerkung 40](#_Toc414621015)

[4.18 werkname 41](#_Toc414621016)

[4.19 krankheitsname 42](#_Toc414621017)

[4.20 form\_krankheit 43](#_Toc414621018)

[4.21 kraeutername 44](#_Toc414621019)

[4.22 kraeutername\_normiert 45](#_Toc414621020)

[4.23 sprache\_kraeutername 46](#_Toc414621021)

[4.24 form\_kraeutername 47](#_Toc414621022)

[4.25 kraeuterzubereitung 48](#_Toc414621023)

[4.26 form\_zubereitung 49](#_Toc414621024)

[4.27 nomen 50](#_Toc414621025)

[4.28 form\_nomen 52](#_Toc414621026)

[5 Annotationsebenen – Strukturelle Annotation 53](#_Toc414621027)

[5.1 lb 53](#_Toc414621028)

[5.2 brace 54](#_Toc414621029)

[5.3 brace\_dir 55](#_Toc414621030)

[5.4 p 56](#_Toc414621031)

[5.5 p\_n 57](#_Toc414621032)

[5.6 p\_rend 58](#_Toc414621033)

[5.7 pb 60](#_Toc414621034)

[5.8 pb\_n 61](#_Toc414621035)

[5.9 pb\_rend 62](#_Toc414621036)

[5.10 pb\_ana 64](#_Toc414621037)

[5.11 div1 – div5 65](#_Toc414621038)

[5.12 div1\_type – div5\_type 66](#_Toc414621039)

[5.13 div1\_n – div5\_n 68](#_Toc414621040)

[5.14 unclear 69](#_Toc414621041)

[5.15 atLeast 70](#_Toc414621042)

[5.16 atMost 71](#_Toc414621043)

[5.17 interpretation 72](#_Toc414621044)

[5.18 figure 73](#_Toc414621045)

[5.19 figure\_rend 75](#_Toc414621046)

[5.20 hi 76](#_Toc414621047)

[5.21 hi\_font 77](#_Toc414621048)

[5.22 hi\_rend 78](#_Toc414621049)

[5.23 head 80](#_Toc414621050)

[5.24 head\_n 81](#_Toc414621051)

[5.25 head\_rend 82](#_Toc414621052)

[5.26 note 83](#_Toc414621053)

[5.27 ref 85](#_Toc414621054)

[5.28 ref\_target 86](#_Toc414621055)

[5.29 ref\_type 87](#_Toc414621056)

[5.30 quote 88](#_Toc414621057)

[5.31 list 89](#_Toc414621058)

[5.32 list\_type 90](#_Toc414621059)

[5.33 item 91](#_Toc414621060)

[5.34 xml\_id 92](#_Toc414621061)

[6 Annotationsebenen – Inhaltliche Annotation 94](#_Toc414621062)

[6.1 definition 94](#_Toc414621063)

[6.2 disease 95](#_Toc414621064)

[6.3 term 96](#_Toc414621065)

[6.4 author\_ref 97](#_Toc414621066)

[6.5 reader\_ref 98](#_Toc414621067)

[6.6 plant 99](#_Toc414621068)

[6.7 property 100](#_Toc414621069)

[6.8 name 101](#_Toc414621070)

[6.9 name\_type 102](#_Toc414621071)

[7 Referenzen 104](#_Toc414621072)

[8 Anhang 105](#_Toc414621073)

[8.1 Transkription und Normalisierung 105](#_Toc414621074)

[8.2 Fußnoten und Marginalien 113](#_Toc414621075)

[8.3 Zeichentabelle 114](#_Toc414621076)

[9 Dokumentation der Probleme während Normalisierung und Konsistenzkorrektur eines historischen Korpus – Ridges v.3 115](#_Toc414621077)

[9.1 Normalisierung 115](#_Toc414621078)

[9.2 Erstellung von „clean“ 115](#_Toc414621079)

[9.3 Konsistenzkorrektur 115](#_Toc414621080)

[10 Benutzte Tools: 116](#_Toc414621081)

[10.1 Makros 116](#_Toc414621082)

[10.2 Python-Skripte 116](#_Toc414621083)

1. Korpus
   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 Kovàcs, Daisy Krüger, Anna-Maria Lehmann, Maria Lober, Anke Lüdeling, Laura Lueders, Samuele Maniscalco, Manuel Metzig, 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: | 31.03.2014 |
|  |  |
| 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 (**R**egister **i**n **D**iachronic **Ge**rman **S**cience) untersucht die Entstehung und Entwicklung der deutschen Wissenschaftssprache ab Mitte des 15. Jahrhunderts bis ins späte 19. Jahrhundert. Bis zum 16. Jahrhundert war die Sprache der Wissenschaft in Europa vorwiegend Latein und erst im Laufe des 15. Jahrhunderts beginnen deutsche Wissenschaftler, wissenschaftliche Texte (oder zumindest Texte mit wissenschaftlichen Inhalten für Laien) erstmalig auch auf Deutsch zu formulieren. Dazu mussten sie ein wissenschaftliches Register ‚erfinden‘ – sie mussten die Terminologie genauso wie angemessene Textstrukturen entwickeln und erproben. Wissenschaftliche Texte haben sich über die Jahrhunderte natürlich verändert. Dabei sind die meisten Veränderungen quantitativ und nicht qualitativ (kategorial). Man kann die Entwicklungen nur nachzeichnen, wenn man Texte oder Textsammlungen von vielen Zeitpunkten genau miteinander vergleicht. In unserem Projekt sollen wissenschaftliche Texte daher auf allen sprachlichen Ebenen (Syntax, Wortbildung, Lexik, Phraseologie, Textstruktur etc.) analysiert werden, um Entwicklungen und Tendenzen identifizieren und beschreiben zu können. In einem variationistischen Ansatz annotieren wir wissenschaftliche Texte mit korpuslinguistischen Methoden und nutzen quantitative Verfahren, um Veränderungen in den Merkmalen zu identifizieren.  Wichtig ist uns dabei die Mitwirkung unserer Studierenden, die im Rahmen ihres Studiums bei der Erstellung und der Annotation der Korpus-Ressourcen mit einbezogen werden. Die so entstandenen Ressourcen werden in mehreren Formaten unter einer Creative-Commons-Lizenz frei zur Verfügung gestellt.  Das RIDGES-Projekt wird von einem Google Digital Humanities Research Award unterstützt. Die originalen Faksimilia der Texte wurden in den meisten Fällen durch Google Books erhoben. |
|  |  |
| Forschungsfrage: | In unserem Projekt sollen wissenschaftliche Texte auf allen sprachlichen Ebenen (Syntax, Wortbildung, Lexik, Phraseologie, Textstruktur etc.) analysiert werden, um Entwicklungen und Tendenzen identifizieren und beschreiben zu können. In einem variationistischen Ansatz annotieren wir wissenschaftliche Texte mit korpuslinguistischen Methoden und nutzen quantitative Verfahren, um Veränderungen in den Merkmalen zu identifizieren. |
|  |  |
| Version: | 4.1 |
|  |  |
| Korpusarchitektur: | Das diachrone Korpus Ridges Herbology besitzt eine multiple Segmentierung, die es erlaubt, konfligierende Tokenisierung von multiplen Ebenen zu verarbeiten (Krause et al. 2012). Das Korpus ist mit Spannenannotationen versehen. |
|  |  |
| Art der Änderung: | Für: GartDerGesundheit\_1487 |
|  | Neue Annotationsebenen wurden hinzugefüht und zwar folgende: personenname, werkname, krankheitsname, form\_krankheit, kraeutername, kraeutername\_normiert, sprache\_kraeutername, form\_kraeutername, kraeuterzubereitung, 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-1-11\_1675 SonderbaresKraeuterbuch-11-21\_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 |

1. 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.

* 1. Typ: Documents
     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 |

* + 1. 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 |
|  |  |

* + 1. 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 |
|  |  |

* + 1. 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 |

* + 1. 1828-DieEigenschaftenAllerHeilpflanzen

|  |  |
| --- | --- |
| Titel: | Die Eigenschaften Aller Heilpflanzen |
| Kürzel | DieEigenschaftenAllerHeilpflanzen\_1828 |
| Autor: | Sauer |
| Jahr: | 1828 |
| Ort:  Verlag: | München  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 |
|  |  |

* + 1. 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 1774-Unterricht

|  |  |
| --- | --- |
| Titel: | Unterricht von der allgemeinen Kräuter- und Wurzeltrocknung |
| 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 |

* + 1. 1764-EinleitungZuDerKraeuterkenntnis

|  |  |
| --- | --- |
| Titel: | Einleitung Zu Der Kräuterkenntnisz |
| 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 |
|  |  |

* + 1. 1735-MysterivmSigillorvm

|  |  |
| --- | --- |
| Titel: | Mysterivm Sigillorvm |
| Kürzel: | MysterivmSigillorvm\_1735 |
| Autor: | Hiebnern 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 |
|  |  |

* + 1. 1722-FloraSaturnizans

|  |  |
| --- | --- |
| Titel: | Flora saturnizans, Die Verwandschafft 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 |

* + 1. DerSchweizerischeBotanicus\_1687

|  |  |
| --- | --- |
| Titel: | Der Schweiszerische Botanicus |
| Kürzel: | V4-1687- DerSchweizerischeBotanicus |
| Autor: | P. Timothei |
| Jahr: | 1687 |
| Ort: | NA |
| Verlag: | Franz Carl Rooß |
| 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 im der Transkription zugrunde liegenden PDF doppelt vorhanden und wurden von den Studenten auch doppelt annotiert. |

* + 1. 1675-SonderbaresKraeuterbuch-11-21

|  |  |
| --- | --- |
| Titel: | Curioser Botanicus oder sonderbares Kräuter Buch |
| Kürzel: | SonderbaresKraeuterbuch-11-21\_1675 |
| Autor: | NA |
| Jahr: | 1675 |
| Ort: | NA |
| Verlag: | NA |
| Seitenangabe: | 11-21 |
| 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 |

* + 1. 1675-SonderbaresKraeuterbuch-1-11

|  |  |
| --- | --- |
| Titel: | Curioser Botanicus oder sonderbares Kräuter Buch |
| Kürzel: | SonderbaresKraeuterbuch-1-11\_1675 |
| Autor: | NA |
| Jahr: | 1675 |
| Ort: | NA |
| Verlag: | NA |
| Seitenangabe: | 1-11 |
| 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 1609-KraeuterbuchCarrichter

|  |  |
| --- | --- |
| Titel: | Kräutterbuch 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 |

* + 1. 1609-HortulusSanitatis

|  |  |
| --- | --- |
| Titel: | Hortulus Sanitatis |
| Kürzel: | HortulusSanitatis\_1609 |
| Autor: | Durante, Castore; Peter Uffenbach |
| Jahr: | 1609 |
| Ort: | Frankfurt am Main |
| Verlag: | Ionae 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 |

* + 1. 1603-AlchemistischePraktik

|  |  |
| --- | --- |
| Titel: | Alchimistische Praktik |
| Kürzel: | AlchemistischePraktik\_1603 |
| Autor: | Libavius, Andreas |
| Jahr: | 1603 |
| Ort: | Frankfurt |
| Verlag: | Johann Saurn |
| 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 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 |

* + 1. 1532-ArtzneiDerKreutter

|  |  |
| --- | --- |
| Titel: | Artzney Buchlein der kreutter |
| Kürzel: | ArtzneiDerKreutter\_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 |

* + 1. 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, personenname, werkname, krankheitsname, form\_krankheit, kraeutername, kraeutername\_normiert, sprache\_kraeutername, form\_kraeutername, kraeuterzubereitung, form\_zubereitung, nomen, form\_nomen, bemerkung |

1. Annotationsebenen – Transkription/Normalisierung
   1. dipl
      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! |

* + 1. **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 |
| Qualtitätsprüfung: | 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 Spannen |
| 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 |

* 1. clean
     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. |
|  |  |

* + 1. **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 |

* 1. norm
     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 Orthografieregeln (vgl. Duden) sowie automatisch angepasst werden. Unbekanntes lexikalisches Material wurde basierend auf graphematischen Regeln an die moderne Orthographie angepasst. |

* + 1. **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 |

1. Annotationsebenen – Annotationen zu linguistischen Eigenschaften
   1. pos
      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. |

* + 1. **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 |
| Qualitätsprüfer: | LAUDATIO |
| Datum: | 2014 |
| Qualitätsprüfung: | NA |
| Editor: | LAUDATIO |

* + 1. **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. |

* + 1. **Typ:** Annotationvalue – pos

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

* 1. pos\_klein
     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. |

* + 1. **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: | xsl, xslx |
| 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 |

* + 1. **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. |

* + 1. **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 Interogativpronomen, die in einer NP vorkommen; übereinstimmend mit STTS |
| PWAV | Adverbiales Interrogativ- oder Relativpronomen; übereinstimment 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, übereinstimment 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) |

* 1. lemma
     1. **Typ:** *Layer* – lemma

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

* + 1. **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 |

* + 1. **Typ:** Annotationlayer– lemma

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

* + 1. **Typ:** Annotationvalue – lemma

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Normalisiertes Lemma | Lemmatisierung. |

* 1. hyperlemma
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – hyperlemma

|  |  |
| --- | --- |
| Annotationsart: | Tokenannotation |
| Beschreibung: | In Fällen, in denen semantisch ein Begriff wie *Heumonat* nicht mehr transparent ist, wird ein modernes Äquivalent, z. B. *Heumonat* – *Juli*, angegeben. |

* + 1. **Typ:** *Annotationvalue* – hyperlemma

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Text | Semantisches Äquivalent. |

* 1. foreign
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – foreign

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

* + 1. **Typ:** *Annotationvalue* – foreign

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| foreign | Allgemeiner Tag für fremdsprachliches Material. |

* 1. foreign\_trans
     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. |

* + 1. **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 |

* + 1. **Typ**: *Annotationlayer* – foreign\_trans

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

* + 1. **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. |

* 1. lang
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – lang

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

* + 1. **Typ:** Annotationvalue – lang

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Kürzel der entsprechenden Spache | 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 |

* 1. Verbposition
     1. **Typ:** *Layer* – Verbposition

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

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – Verbposition

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

* + 1. **Typ:** Annotationvalue – Verbposition

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

* 1. Nebensatztyp
     1. **Typ:** *Layer* – Nebensatztyp

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

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – Nebensatztyp

|  |  |
| --- | --- |
| Annotationsart: | Tokenannotation |
| Beschreibung: | Annotation der Nebensatztypen. |

* + 1. **Typ:** Annotationvalue – Nebensatztyp

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

* 1. KOUS\_Semantik
     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. |

* + 1. **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, xslx |
| 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 |

* + 1. **Typ:** Annotationlayer – KOUS\_Semantik

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

* + 1. **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 |

* 1. komp
     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. |

* + 1. **Typ:** Preparationstep – komp

|  |  |
| --- | --- |
| Schritt: | 1 |
| Aufbereitung: | Linguistische Annotation |
| Aufbereitungsart: | manuell |
| Tool | Excel 2010 |
| Format: | xls, xlsx |
| Segmentationsbasis der Annotation: | dipl |
| Qualitätsprüfer: | Laura Perlitz |
| Datum | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – komp

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

* + 1. **Typ:** Annotationvalue – komp

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| k | Kompositum |

* 1. komp\_orth
     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. |

* + 1. **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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – komp\_orth

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

* + 1. **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) |

* 1. prot
     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 ſaft*).  Der Prototyp 3 (prot3) markiert alle Zweifelsfälle, also Fälle, die weder eindeutig als Phrase noch als Kompositum zu identifizieren sind. Wieder liegen zwei freie Morpheme sowie Getrenntschreibung vor, jedoch kann man keine genaue Aussage darüber treffen, ob eine wortinterne Flexion vorgenommen wurde oder möglich wäre und ob das jeweilige Vorkommen syntaktisch erweiterbar ist (z.B. *widers Teuffels Biſſz*). 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. |

* + 1. **Typ:** Preparationstep – prot

|  |  |
| --- | --- |
| Schritt: | 1 |
| Aufbereitung: | Linguistische Annotation |
| Aufbereitungsart: | manuell |
| Tool | Excel 2010 |
| Format: | xls, xlsx |
| Segmentationsbasis der Annotation: | komp |
| Qualitätsprüfer: | Laura Perlitz |
| Datum | 31.09.2014 |
| 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 |

* + 1. **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. |

* + 1. **Typ:** Annotationvalue – prot

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

* 1. attr\_gen
     1. **Typ:** *Layer* – attr\_gen

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

* + 1. **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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – attr\_gen

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

* + 1. **Typ:** Annotationvalue – attr\_gen

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| gprä | pränominales Genitivattribut |
| gpost | postnominales Genitivattribut |

* 1. 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. |

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – strD

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markiert alle im Korpus enthaltenen koordinierten Komposita ("strukturelle Destruktionen"; z.B.: *gelb⸗ und Waſſerſucht*). |

* + 1. **Typ:** Annotationvalue – strD

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| strD | strukturelle Destruktion |

* 1. personenname

**Typ:** *Layer* – strD

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

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – personenname

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – personenname

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | Personennamen. |

* 1. bemerkung

**Typ:** *Layer* – bemerkung

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

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – bemerkung

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – bemerkung

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | Bemerkung. |

* 1. werkname

**Typ:** *Layer* – werkname

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

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – werkname

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – werkname

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | Werkname. |

* 1. krankheitsname

**Typ:** *Layer* – krankheitsname

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

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – krankheitsname

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – krankheitsname

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | Krankheitsname. |

* 1. 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. |

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – form\_krankheit

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – form\_krankheit

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | NA |

* 1. kraeutername

**Typ:** *Layer* – kraeutername

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

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – kraeutername

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – kraeutername

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | Kräutername. |

* 1. 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. |

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – kraeutername\_normiert

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – kraeutername\_normiert

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | Normierter Kräutername. |

* 1. sprache\_kraeutername

**Typ:** *Layer* – sprache\_kraeutername

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

* + 1. **Typ:** Preparationstep – sprache\_kraeutername

|  |  |
| --- | --- |
| Schritt: | 1 |
| Aufbereitung: | Linguistische Annotation |
| Aufbereitungsart: | manuell |
| Tool | NA |
| Format: | Excel 2010 |
| Segmentationsbasis der Annotation: | dipl |
| Qualitätsprüfer: | Anke Lüdeling, Uwe Springmann |
| Datum | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – sprache\_kraeutername

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – sprache\_kraeutername

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | Sprache des Kräuternamens. |

* 1. 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. |

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – form\_kraeutername

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – form\_kraeutername

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | NA |

* 1. kraeuterzubereitung

**Typ:** *Layer* – kraeuterzubereitung

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

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – kraeuterzubereitung

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – kraeuterzubereitung

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | NA |

* 1. 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. |

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – form\_zubereitung

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – form\_zubereitung

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | NA |

* 1. nomen

**Typ:** *Layer* – nomen

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

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – nomen

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – nomen

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | NA |

* 1. 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. |

* + 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 | 31.09.2014 |
| 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 |

* + 1. **Typ:** Annotationlayer – form\_nomen

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Nur in GartDerGesundheit\_1487 annotiert. |

* + 1. **Typ:** Annotationvalue – form\_nomen

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| String | NA |

1. Annotationsebenen – Strukturelle Annotation
   1. lb
      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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – lb

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markierung einer ganzen Zeile. |

* + 1. **Typ:** Annotationvalue – lb

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| lb | Markierung pro ganzer Zeile. |

* 1. brace
     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. |

* + 1. **Typ:** Preparationstep – brace

|  |  |
| --- | --- |
| Schirtt: | 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 |

* + 1. **Typ:** Annotationlayer – brace

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

* + 1. **Typ:** Annotationvalue – brace

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

* 1. brace\_dir
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – brace\_dir

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Richtung der Klammer |

* + 1. **Typ:** Annotationvalue – p

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

* 1. p
     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. |

* + 1. **Typ:** Preparationstep – p

|  |  |
| --- | --- |
| Schirtt: | 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 |

* + 1. **Typ:** Annotationlayer – p

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markierung eines ganzen Paragraphen. |

* + 1. **Typ:** Annotationvalue – p

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| P | Markierung pro ganzem Paragraphen. |

* 1. p\_n
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – p\_n

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markiert die Nummerierung eines Paragraphen. |

* + 1. **Typ:** Annotationvalue – p\_n

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

* 1. p\_rend
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – p\_rend

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Beschreibung der Darstellung eines Paragraphen. |

* + 1. **Typ:** Annotationvalue – p\_rend

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

* 1. pb
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – pb

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markierung einer ganzen Seite. |

* + 1. **Typ:** Annotationvalue – pb

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| pb | Markierung pro ganzer Seite. |

* 1. pb\_n
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – pb\_n

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markierung der Nummerierung einer Seite. |

* + 1. **Typ:** Annotationvalue – pb\_n

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| numerischer Wert | Seitenzahl. |

* 1. pb\_rend
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – pb\_rend

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

* + 1. **Typ:** *Annotationvalue* – pb\_rend

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| vonHaſelwurtz.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. |
| vonWermůt.Cap.I. | Seite/n ist/sind einer Überschrift zugeordnet. |
| vonDrachenwurtz.Cap.IIII. | Seite/n ist/sind einer Überschrift zugeordnet. |
| ohlZuMachen. | Seite/n ist/sind einer Überschrift zugeordnet. |
| zumBeſtenZuDiſtilliren. | Seite/n ist/sind einer Überschrift zugeordnet. |
| waſſerAußKräuternVndDergleichen | Seite/n ist/sind einer Überschrift zugeordnet. |
| auffsBeſtZuDiſtilliren. | Seite/n ist/sind einer Überschrift zugeordnet. |
| außKräuternVndDergleichen | Seite/n ist/sind einer Überschrift zugeordnet. |
| waſſerAußKräuternVndDergleichen | Seite/n ist/sind einer Überschrift zugeordnet. |
| amBeſtenZuDiſtilliren. | Seite/n ist/sind einer Überschrift zugeordnet. |

* 1. pb\_ana
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – pb\_ana

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

* + 1. **Typ:** Annotationvalue – pb\_ana

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

* 1. div1 – div5
     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. |

* + 1. **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 |

* + 1. **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**). |

* + 1. **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. |

* 1. div1\_type – div5\_type
     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. |

* + 1. **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 |

* + 1. **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**). |

* + 1. **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 | Anbaugebiete 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. |

* 1. div1\_n – div5\_n
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – div2\_n – div3\_n

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

* + 1. **Typ:** *Annotationvalue* – div1\_n – div5\_n

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Ziffer | Nummerierung der Sektion. |

* 1. unclear
     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. |

* + 1. **Typ:** *Preparationstep* – unclear

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

* + 1. **Typ:** *Annotationlayer* – unclear

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

* + 1. **Typ:** *Annotationvalue* – unclear

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

* 1. atLeast
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – atLeast

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

* + 1. **Typ:** *Annotationvalue* – atLeast

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

* 1. atMost
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – atMost

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

* + 1. **Typ:** Annotationvalue – atMost

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

* 1. interpretation
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer*– interpretation

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

* + 1. **Typ:** *Annotationvalue*– interpretation

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Text | Angabe der Interpretation. |

* 1. figure
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – figure

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

* + 1. **Typ:** Annotationvalue – figure

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

* 1. figure\_rend
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – figure\_rend

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

* + 1. **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 Staubwurz. |
| drawingOfAKamille | Abbildung einer Kamille. |
| drawingOfAHühnerdarm | Abbildung eines Hühnerdarms. |
| drawingOfAFilter | Abbildung eines Filters. |
| drawingOfAHelmet | Abbildung eines Helms. |
| drawingOfADill | Abbildung von Dill. |
| drawingOfAHauswurz | Abbildung von Hauswurz. |
| drawingOfAWaldenburgischerKolben | Abbildung eines Waldenburgischen Kolbens. |
| drawingOfADrachenwurtz | Abbildung von Drachenwurz. |
| drawingOfAGauchheyl | Abbildung von Gauchheyl. |
| drawingOfAHaselwurtz | Abbildung von Haselwurz. |

* 1. hi
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – hi

|  |  |
| --- | --- |
| Annotationsart: | Spannen |
| Beschreibung: | Markierung eines irgendwie hervorgehobenen Bereichs. |

* + 1. **Typ:** Annotationvalue – hi

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

* 1. hi\_font
     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. |

* + 1. **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 Annotation: | norm |
| Qualitätsprüfer: | LAUDATIO |
| Datum: | 2014 |
| Qualitätsprüfung: | NA |
| Editor: | LAUDATIO |

* + 1. **Typ:** *Annotationlayer* – hi\_font

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

* + 1. **Typ:** *Annotationvalue* – hi\_font

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

* 1. hi\_rend
     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. |

* + 1. **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 |
| 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 |

* + 1. **Typ:** *Annotationlayer* – hi\_rend

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

* + 1. **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 gedruck ist. |
| iniCap | Ein Text, der mit einem zusätzlichen Großbuchstaben gedruckt ist. |
| letter-spacing:1em | Ein größerer Abstand zwischen Buchstaben, Sperrdruck. |

* 1. head
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – head

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markierung einer Überschrift. |

* + 1. **Typ:** Annotationvalue – head

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| head | Allgemeine Markierung einer Überschrift. |

* 1. head\_n
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – head\_n

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

* + 1. **Typ:** Annotationvalue – head\_n

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Ziffer | Konkrete Nummerierung der Überschrift. |

* 1. head\_rend
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – head\_rend

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

* + 1. **Typ:** *Annotationvalue* – head\_rend

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| brown | Überschrift in brauner Farbe. |

* 1. note
     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. |

* + 1. **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 |

* + 1. **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. |

* + 1. **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 |
| 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. |

* 1. ref
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – ref

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Referenz einer Fußnote. |

* + 1. **Typ:** Annotationvalue – ref

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| ref | Referenz einer Fußnote. |

* 1. ref\_target
     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. |

* + 1. **Typ:** *Preparationstep* – ref\_target

|  |  |
| --- | --- |
| 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 |

* + 1. **Typ:** *Annotationlayer* – ref\_target

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

* + 1. **Typ:** *Annotationvalue* – ref\_target

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

* 1. ref\_type
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – ref\_type

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

* + 1. **Typ:** *Annotationvalue* – ref\_type

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| noteAnchor | Art der Referenz. |

* 1. quote
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – quote

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markiert ein Zitat im Text. |

* + 1. **Typ:** Annotationvalue – quote

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Quote | Ein Zitat im Text. |

* 1. list
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – list

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Eine Liste von Elementen. |

* + 1. **Typ:** Annotationvalue – list

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| List | Allgemeine Markierung einer Liste. |

* 1. list\_type
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – list\_type

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Beschreibung der Art der Liste. |

* + 1. **Typ:** *Annotationvalue* – list\_type

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Simple | Einfache Aufzählung. |

* 1. item
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – item

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

* + 1. **Typ:** Annotationvalue – item

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

* 1. xml\_id
     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. |

* + 1. **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 |

* + 1. **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 Kennzeichung f1, der zweiten f2 etc. zugewiesen, selbst wenn sie bspw. im Text die Nummer 86) darstellt oder mit \*\*) markiert ist.  Siehe auch und für Beispiele . |

* + 1. **Typ:** Annotationvalue –xml\_id

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

1. Annotationsebenen – Inhaltliche Annotation
   1. definition
      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. |

* + 1. **Typ:** *Preparationstep* – definition

|  |  |
| --- | --- |
| Aufbereitung: | Inhaltiche 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 |

* + 1. **Typ:** *Annotationlayer* – definition

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

* + 1. **Typ:** *Annotationvalue* – definition

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

* 1. disease
     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. |

* + 1. **Typ:** Preparationstep – disease

|  |  |
| --- | --- |
| Aufbereitung: | Inhaltiche 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 |

* + 1. **Typ:** Annotationlayer – disease

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

* + 1. **Typ:** Annotationvalue – disease

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| di | Benennung einer Krankheit |

* 1. term
     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. |

* + 1. **Typ:** Preparationstep – term

|  |  |
| --- | --- |
| Aufbereitung: | Inhaltiche 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 |

* + 1. **Typ:** Annotationlayer – term

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

* + 1. **Typ:** Annotationvalue – term

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| t | Kennzeichnet einen allgemeinen, technischen oder wissenschaftlichen Begriff (**t**erm) innerhalb der Kräuterkunde. |
| h | Kennzeichnet einen Begriff, der eine Pflanze bzw. ein Kraut (**h**erb) benennt/beschreibt. |
| d | Kennzeichnet einen Begriff, der eine Krankheit (**d**isease) benennt/beschreibt. |

* 1. author\_ref
     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. |

* + 1. **Typ:** *Preparationstep* – author\_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 |

* + 1. **Typ:** *Annotationlayer* – author\_ref

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

* + 1. **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. |

* 1. reader\_ref
     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. |

* + 1. **Typ:** *Preparationstep* – reader\_ref

|  |  |
| --- | --- |
| Aufbereitung: | Inhaltiche 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 |

* + 1. **Typ:** *Annotationlayer* – reader\_ref

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

* + 1. **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. |

* 1. plant
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – plant

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

* + 1. **Typ:** *Annotationvalue* – plant

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| pl | Benennung einer Pflanze. |

* 1. property
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – property

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

* + 1. **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. |

* 1. name
     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. |

* + 1. **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 |

* + 1. **Typ:** Annotationlayer – name

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Markiert Eigennamen. |

* + 1. **Typ:** Annotationvalue – name

|  |  |
| --- | --- |
| Wert: | Wertbeschreibung: |
| Name | Allgemeine Markierung von Eigennamen. |

* 1. name\_type
     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. |

* + 1. **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 |

* + 1. **Typ:** *Annotationlayer* – name\_type

|  |  |
| --- | --- |
| Annotationsart: | Spannenannotation |
| Beschreibung: | Spezifiziert den Typ der Eigennamen. |

* + 1. **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. |

1. 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.

1. Anhang
   1. Transkription und Normalisierung

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

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 Orthografieregeln (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 Herbology Version 4.0 Korpus gegeben, die konkrete Regel beschrieben und die weiteren Normalisierungsschritte gegenübergestellt (dipl – clean – norm). Die Richtlinien sind nach Schwerpunkten gruppiert: (1) Zeichensetzung/Sonderzeichen (2) Segmentierung/Tokenisierung (3) Interpunktion.

|  |  |  |  |
| --- | --- | --- | --- |
| **Nr.** | **dipl-Ebene** | **clean-Ebene** | **norm-Ebene** |
|  | Transkript des Faksimiles. | Erfolgt manuell im Format .xls.  Tipp: Kopieren Sie sich die dipl-Ebene und verändern Sie dann die entsprechenden Stellen! | Erfolgt manuell im .xls Format.  Tipp: Kopieren Sie sich die clean-Ebene und verändern Sie dann die entsprechenden Stellen! |
| 1.0 | Oft werden die ersten Zeichen bzw. ersten Wörter (im Bsp. *Von Weg⸗*) einer neuen Seite bei einem Seitenumbruch in der vorhergehenden Seite in einem eigenen Absatz/in einer eigene Zeile doppelt realisert. Diese Zeichen bzw. das Wort wird nicht mit in das Transkript aufgenommen. |  |  |
| 1.1 | Das Zeichen für morphologische Worttrennung (oder Zeilenumbruch) wird mit U 2E17 [[1]](#footnote-1) übernommen. | Die morphologische Trennung durch „⸗“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt.  Komposita, die in der dipl-Ebene mit einem „-“ realisiert werden, behalten dieses Zeichen bei, um die allgemeine Suche nach Komposita in dieser Ebene zu ermöglichen. Für Erläuterungen zur rein graphischen Trennung, wie Zeilenumbrüchen, siehe Nummer 2.1. | 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](http://www.duden.de/).) |
| **Dipl** | **clean** | **norm** |
| Dan̄ ⸗reiſz | *Dann-reisz* | *Tannenreis* |
| 1.2 | Das „ſ“ wird mit U 017F übernommen. | Das lange „ſ“ wird durch ein reguläres „s“ ersetzt. |  |
| **Dipl** | **clean** | **norm** |
| *Leſer* | *Leser* | *Leser* |
| 1.3 | Gerade Längenzeichen oder Makrone über Konsonanten werden mit U 0304 übernommen.[[2]](#footnote-2)    Längenzeichen werden mit U 0304 übernommen, wenn nicht klar ist, ob es sich um Längen- oder Makron handelt, wird zugunsten des Makrons entschieden. | Längenzeichen werden entsprechend ihrer Funktion umgesetzt. Die kann u. a. aus Reichmann & Wegera (1993) entnommen werden.[[3]](#footnote-3) |  |
| **dipl** | **clean** | **norm** |
| Dan̄ ⸗reiſz | *Dann-reisz* | *Tannenreis* |
| 1.4 | Wie auch u.a. in Beispiel 1.2 wird die Ligatur mit dem langen „ſ“ und „z“ übernommen. | Die Ligatur wird analog zu „sz“ aufgelöst. | Nach modernen Orthografieregeln wird für „sz“ das „ß“ verwendet. |
| **dipl** | **clean** | **norm** |
| *fleiſzige* | *fleiszige* | *fleißige* |
| 1.5 | Umlaute mit superskribiertem „e“ werden mit Vokal + U 0364 übernommen. | Umlaute mit superskribiertem „e“ werden wie moderne Umlaute des Deutschen realisiert. |  |
| **dipl** | **clean** | **norm** |
| *haͤuffte* | *häuffte* | *häufte* |
| 1.6 | Die Ligatur „æ“ wird mit U 00E6, für „Æ“ mit U 00C6 übernommen.  Das gleiche gilt ebenfalls für „Œ“ U 0152 und „œ“ U 0153. | Die Ligatur wird aufgelöst. |  |
| **dipl** | **clean** | **norm** |
| *hæc* | *haec* | *haec* |
| 1.7 | Dieses diakritische Zeichen wird mit U 016F in der dipl-Ebene übernommen.  Für “å“ gilt U 00E5. | Dieses diakritische Zeichen wird in der clean-Ebene nicht mehr realisiert und durch den Segmentationsbasis der Annotationvokal ersetzt. |  |
| **Dipl** | **clean** | **norm** |
| *Gůt* | *gut* | *gut* |
| 1.8 | Das kaufmännische Und „&“ wird mit U 0026 übernommen. | Das kaufmännische Und wird übernommen. | Das kaufmännische Und wird übernommen. |
| **Dipl** | **clean** | **norm** |
| & | & | & |
| 1.9 | Alle Akzente werden übernommen. Hierbei ist zu beachten, dass der jeweilige unäre Hex-Code zu verwenden ist.[[4]](#footnote-4) Für „ó“ gilt U 00F3. | Alle Akzente werden übernommen. | Alle Akzente werden übernommen. |
| **Dipl** | **clean** | **norm** |
| *Quòd* | *quòd* | *quòd* |
| 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 segmentiert, um gemischte Brüche transparent zu halten. |
| **Dipl** | **clean** | **norm** |
| 1 | 1 | 1 |
| 1/2 | 1/2 | 1/2 |
| 1.11 | Dieses Symbol wird mit dem U 2299 übernommen. | Dieses Symbol wird mit dem U 2299 übernommen. | Dieses Symbol wird mit dem U 2299 übernommen. |
| **Dipl** | **clean** | **norm** |
| ⊙ | ⊙ | ⊙ |
| 1.12 | Dies bedeutet et c. |  |  |
| **Dipl** | **clean** | **norm** |
| et.  c. | et  c. | etc. |
| 1.13 | Senkrechtstriche werden mit „|“ übernommen. | Senkrechtstriche werden übernommen. | Senkrechtstriche werden übernommen. |
| **Dipl** | **clean** | **norm** |
| *dieſe* | *diese* | *diese* |
| *|* | *|* | *|* |
| *Haben* | *haben* | *haben* |
| 1.14 | Die ct-Ligatur wird nicht übernommen. | Die ct-Ligatur wird nicht übernommen. | Die ct-Ligatur wird nicht übernommen. |
| **dipl** | **clean** | **norm** |
| *Lactucis* | *Lactucis* | *Lactucis* |
| 1.15 | 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> | Schriftsätze werden übernommen. | Schriftsätze werden übernommen. |
| 2.1 | Von Zeilenumbrüchen betroffene Elemente werden analog zur Primärquelle getrennt tokenisiert. Das Trennungszeichen „⸗“ wird mit U 2E17 übernommen. | Wörter, die von Zeilenumbrüchen betroffen sind, werden in der clean-Ebene ohne den (Doppel-) Bindestrich, der diesen anzeigt, sowie in einer Spanne zusammengefasst realisiert.  Die grafische Worttrennung wird somit aufgehoben. |  |
| **dipl** | **clean** | **norm** |
| *allge⸗* | *allgemeinen* | *allgemeinen* |
| *meinen* |
| 2.2 | 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. | Die Worttrennung ohne graphische Markierung ist nicht in jedem Fall transparent. Graphische Normalisierungen werden nicht vorgenommen. | Diese Art der Worttrennung, wenn sie transparent ist, wird nach modernen Orthographieregeln normalisiert und die graphische Trennung aufgehoben. |
| **dipl** | **clean** | **norm** |
| *ge* | *ge* | *genannt* |
| *nent* | *nent* |
| 2.3 | Klitika werden als ein Token realisiert. | Klitika werden als ein Token realisiert. | Klitika werden in dieser Ebene aufgelöst. Dazu wird die Tokenisierung verändert, in dem aus einem Token in dipl (oder clean) zwei Token gemacht werden! |
| **Dipl** | **clean** | **norm** |
| *Ichs* | *ichs* | *ich* |
| *es* |
| 2.4 | Auch entgegen moderner Orthografieregeln wird analog zur Textgrundlage tokenisiert. |  | Getrennt- und Zusammenschreibung wird nach modernen Orthografieregeln angepasst. Dazu wird die Tokenisierung verändert, indem aus einem Token in der dipl-Ebene (oder clean) zwei Token in der norm-Ebene gemacht werden! |
| **dipl** | **clean** | **norm** |
| *zuſetzen* | *zusetzen* | *zu* |
| *setzen* |
| 2.5 | Auch entgegen moderner Orthografieregeln wird analog zur Textgrundlage tokenisiert. |  | 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** |
| *zuſammen* | *zusammen* | *zusammengesetzt* |
| *geſetzet* | *gesetzet* |
| 2.6 | Komposita, egal welcher Komplexität oder Bildungsweise, mit Gleichheitszeichen (U 2E17) werden als ein Token realisiert. | Die morphologische Trennung durch „⸗“ bleibt erhalten, wird jedoch mit einem regulären „-“ ersetzt. | Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschrieben.  (Grundlage hierfür bildet die aktuelle Ausgabe des Dudens.) |
| **dipl** | **Clean** | **norm** |
| *Artzney⸗Kunſt* | *Artzney-Kunst* | *Arzneikunst* |
| 2.7 | Von Zeilenumbrüchen betroffene Komposita werden analog zur Primärquelle getrennt tokenisiert, dabei bleibt die Kennzeichnung der morphologischen Worttrennung unberührt. | Wenn ein Kompositum, das durch Gleichheitszeichen grundsätzlich getrennt wird, von einem Zeilenumbruch betroffen ist, so wird dieses Gleichheitszeichen in der clean-Ebene entfernt, das andere Gleichheitszeichen für die morphologische Trennung wird analog zu Nummer 1.1 mit einem Minus ersetzt. Das Kompositum wird in einer Spanne zusammengefasst wiedergegeben. | Die Gleichheitszeichen, die Kompositabildung anzeigen, werden durch Minuszeichen ersetzt oder das Kompositum wird analog zur modernen Orthografie zusammengeschrieben.  (Grundlage hierfür bildet die aktuelle Ausgabe des Dudens.) |
| **dipl** | **clean** | **norm** |
| *Stab⸗* | *Stabwurtz-Oel* | *Stabwurzöl* |
| *wurtz⸗Oel* |
| 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** |
| *Spannen⸗lang* | *Spannen-lang* | s*pannenlang* |
| 2.9 | Auch entgegen moderner Orthografieregeln wird analog zur Textgrundlage tokenisiert. | Das lange „ſ“ wird durch ein reguläres „s“ ersetzt. | Nicht der modernen Orthografie entsprechende Bildungen werden angeglichen. |
| **dipl** | **clean** | **norm** |
| *zuſetzen* | *zusetzen* | *zu* |
| *setzen* |
| 2/3.1 | Abkürzungen werden analog zur Textgrundlage tokenisiert. Das heißt, auch Setzfehler oder Spatien, die nicht modernen Orthografieregeln entsprechen, werden realisiert.[[5]](#footnote-5) |  | Abkürzungen werden nach Dudenrichtlinien umgesetzt. Dies hat zur Folge, dass Abkürzungen unterschiedlich tokenisiert werden können (bspw. Abkürzungen als zwei oder mehr Tokens wie *u. a. m* vs. solche, die als ein Token realisiert werden, wie *usw.*). vgl. dazu Nummer 2/3.2 und 2/3.3.  In Fällen, in denen keine Dudenrichtlinie zugeordnet werden kann, wird immer nach einem Punkt segmentiert. |
| **dipl** | **clean** | **norm** |
| *z.* | *z.* | *z.* |
| *B.* | *B.* | *B.* |
| 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 *u. a. m.* vs. solche, die als ein Token realisiert werden, wie *usw.*). vgl. dazu Nummer 2/3.1 und 2/3.3.  In Fällen, in denen keine Dudenrichtlinie zugeordnet werden kann, wird immer nach einem Punkt segmentiert. |
| **dipl** | **clean** | **norm** |
| *u.* | *u.* | *usw.* |
| *ſ.* | *s.* |
| *w.* | *w.* |
| 2/3.3 | Abkürzungen werden analog zur Textgrundlage tokenisiert. Das heißt, auch Setzfehler oder Spatien, die nicht modernen Orthografieregeln entsprechen, werden realisiert. |  | Abkürzungen werden nach Dudenrichtlinien umgesetzt. Dies hat zur Folge, dass Abkürzungen unterschiedlich tokenisiert werden können (bspw. Abkürzungen als zwei oder mehr Tokens wie *u. a. m.* vs. solche, die als ein Token realisiert werden, wie *usw.*). vgl. dazu Nummer 2/3.1 und 2/3.2.  In Fällen, in denen keine Dudenrichtlinie zugeordnet werden kann, wird immer nach einem Punkt segmentiert. |
| **dipl** | **clean** | **norm** |
| *u* | *u* | *u.* |
| *d.* |
| *d.g.* | *d.g.* | *g.* |
| 3.1 | Punktsetzungen bei Ordinalzahlen werden mit der Ziffer als ein Token realisiert. |  | Nach modernen Orthografieregeln werden Ordinalzahlen mit Punkt realisiert. |
| **dipl** | **clean** | **norm** |
| *1.* | *1.* | *1.* |
| *Loth* | *Loth* | *Loth* |
| 3.2 | Die Faksimilia weisen zum Teil Kardinalzahlen mit Interpunktion auf.  Diese Punktsetzungen werden mit der jeweiligen Ziffer als Token realisiert.  Ausnahmen bilden Zahlen, die mit einem satzbeenden Punkt auftreten. Hier wird wie gewohnt die Satzinterpunktion getrennt von der Kardinalzahl tokenisiert. |  | Um die Kardinalzahlen von Ordinalzahlen in der norm-Ebene getrennt betrachten zu können, wird die Interpunktion bei Kardinalzahlen weggelassen. |
| **dipl** | **clean** | **norm** |
| *1818.* | *1818.* | *1818* |
| 3.3 | Satzinterpunktion wie Kommata, Punkte, Virgeln oder Semikola werden als jeweils eigenständige Token realisiert. |  |  |
| **dipl** | **clean** | **norm** |
| *finden* | *finden* | *finden* |
| *.* | *.* | *.* |
| 3.4 | Nicht mehr lesbare Zeichen oder Zeichenketten werden mit einem Unterstrich markiert, unabhängig davon, wieviele Zeichen (in etwa) nicht mehr interpretiert werden können. | Durch die fehlende Transkription ist eine Normalisierung nicht mehr möglich, Dies wird mit dem Platzhalter-Tag „unknown“ angezeigt. | Durch die fehlende Transkription ist eine Normalisierung nicht mehr möglich, Dies wird mit dem Platzhalter-Tag „unknown“ angezeigt. |
| **dipl** | **clean** | **norm** |
| \_ | unknown | unknown |
| 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. |
| **dipl** | **clean** | **norm** |
| o€ | oder | 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 "$" repräsentiert und in clean und norm durch "us" ersetzt. | Ersetzung durch "us". | Ersetzung durch "us". |
| **dipl** | **clean** | **norm** |
| Plateari$ | Platearius | Platearius |
| 3.11 | Dieses Zeichen wird durch vU+0364 realisiert. | 'vͤ' wird in der clean-Ebene durch 'ü' ersetzt. | 'vͤ' wird in der norm-Ebene durch 'ü' ersetzt. |
| **dipl** | **clean** | **norm** |
| 'vͤ'ber | über | 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** | **clean** | **norm** |
| fűr | für | für |
| 3.13 | 'i' mit Akzent wird in der dipl-Ebene mit U+00ED transkribiert und in clean und norm durch 'i' normalisiert. |  |  |
| **dipl** | **clean** | **norm** |
| wírdígen | wirdigen | 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** | **clean** | **norm** |
| Jtem | Jtem | Item |
| 3.15 | Handschriftlich Versalien die offensichtlich zum Text gehören (KEINE Anmerkungen oder Kommentare von Lesern), werden mit annotiert. |  |  |
| **dipl** | **clean** | **norm** |
| ABſinthium | ABsinthium | Absinth |

Tabelle 1 Transkriptionsbeispiele

* 1. 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.

* + 1. Beispiele für die Referenzierung von Fußnoten

Siehe auch Abschnitt (xml\_id), (ref\_target), (ref) und (note).

|  |  |
| --- | --- |
|  |  |

Tabelle 2: Fußnotenreferenzierung

* 1. Zeichentabelle

|  |  |  |  |
| --- | --- | --- | --- |
| **Beschreibung** | **Zeichen** | **Additiver Hex-Code** | **Unärer Hex-Code** |
| **Schräger Dopppelbindestrich** | ⸗ |  | 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“** | ͤ | 0364 |  |
| **a mit Kreis** | å |  | 00E5 |
| **u mit Kreis** | ů |  | 016F |
| **Absatzzeichen** | ¶ |  | 00B6 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Tabelle 3: Zeichentabelle

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

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

* 1. 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, ÜBERPRÜFEN!!!

1. *1/2* wurde durch Excel zu *01.Februar*
2. Das normalisierte *falsch* wurde durch Excel zum logischen Operator *FALSCH*
   1. 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 Kontexts. 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.

1. Benutzte Tools:
   1. Makros
      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.

* + 1. 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.

* 1. Python-Skripte
     1. cleanV2.py

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

* + 1. 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.

1. Sonderzeichen wie zum Beispiel das „ſ“ werden mit Hilfe von Unicodes in der dipl-Ebene realisiert. Eine Liste der zu verwendenden Unicodes ist angefügt. [↑](#footnote-ref-1)
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. [↑](#footnote-ref-2)
3. Nachweis für das Beispiel: Reichmann & Wegera 1993: 31. [↑](#footnote-ref-3)
4. Dazu siehe Zeichentabelle im Anhang. [↑](#footnote-ref-4)
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. [↑](#footnote-ref-5)