



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Germanistische Institutspartnerschaft der Universitäten Hamburg und Vilnius
Korpusdidaktik für formelhafte (Fach)Sprache (KoDi-FS)
Workshop zu Korpora an der Universität Vilnius, 10.–11.10.2022

Das Deutsche Textarchiv und die Referenzkorpora zur deutschen Sprachgeschichte Erfahrungen und Anregungen zum Einsatz in der Lehre

Sarah Ihden
sarah.ihden@uni-hamburg.de

10.10.2022

Deutsches Textarchiv (DTA)

- Texte vom 16. bis frühen 20. Jh. aus verschiedenen Disziplinen und Gattungen (4 Genres: Belletristik, Gebrauchsliteratur, Wissenschaft, Zeitung)
- Kernkorpus enthält 1.468 Werke mit **150 Mio. Tokens** (Stand: 10.10.22)
(<http://www.deutschestextarchiv.de/doku/ueberblick>)
- Kernkorpus enthält überregional rezipierte Texte → bildet Entwicklung einer überregionalen hd. Umgangssprache seit dem Ende der frnhd. Sprachperiode ab
- Ziel: möglichst ausgewogene Verteilung der verschiedenen Disziplinen und Textsorten
- **Automatische** Annotation von Wortart (STTS) und Lemma (keine Korrektur!)
- über eigenes Suchtool abfragbar (<https://www.deutschestextarchiv.de/>) oder über DWDS (<https://www.dwds.de/>)



Text-Bild-Ansicht öffnen ...



Andolt, Ernst [d. i. Bernhard Abeken]: Eine Nacht. In: Deutscher Novellenschatz. Hrsg. von Paul Heyse und Hermann Kurz. Bd. 22. 2. Aufl. Berlin, [1910], S. 211–287. In: Weitin, Thomas (Hrsg.): Volldigitalisiertes Korpus. Der Deutsche Novellenschatz. Darmstadt/Konstanz, 2016.

BIBLIOGRAPHISCHE ANGABEN

URN: urn:nbn:de:kobv:b4-32197-6
 Titel: Eine Nacht
 Autor/in: Bernhard Abeken (GND, Wikipedia, ADB/NDB)
 Erscheinungsjahr: 1910
 Verlag/Drucker: Globus Verlag
 Ort: Berlin
 Band: Band 22
 erschienen in: Deutscher Novellenschatz
 Auflage: 2
 Bildnachweis: Bibliothek der Universität Konstanz, deu 838.29/h29

INFORMATIONEN ZUM WERK

Publikationstyp: unselbstständige Schrift innerhalb einer Reihe
 Verfügbarkeit: Text (TEI-XML-, HTML-, TCF-, E-Book-Fassung): Namensnennung 4.0 International (CC BY 4.0)
 Weitere Informationen: Nutzungsbedingungen.
 Schriftart: Fraktur
 Genre: Belletristik :: Novelle
 im DTA seit: 2017-07-13 11:40:29
 Korpus: DTAE (DTA-Erweiterungen) – Novellenschatz

GRUNDLAGE DIESES DIGITALISATS

Dieses Werk wurde im Rahmen des Moduls DTA-Erweiterungen (DTAE) digitalisiert. Weitere Informationen ...

Thomas Weitin: **Herausgeber**

Digital Humanities Cooperation Konstanz/Darmstadt: **Bereitstellung der Texttranskription.** (2017-03-14T12:28:07Z) *Bitte beachten Sie, dass die aktuelle Transkription (und Textauszeichnung) mittlerweile nicht mehr dem Stand zum Zeitpunkt der Übernahme des Werkes in das DTA entsprechen muss.*

Jan Merkt, Thomas Gilli, Jasmin Bieber, Katharina Herget, Anni Peter, Christian Thomas, Benjamin Fiechter: **Bearbeitung der digitalen Edition.** (2017-03-14T12:28:07Z)

Suche im Werk

Hilfe

Ansichten für dieses Werk

- Text-Bild-Ansicht
- alle Faksimiles
- DTAQ (Qualitätssicherung)

Download

XML (TEI P5) · HTML · Text
 TCF (text annotation layer)
 TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)
 XML (TEI P5 inkl. att.linguistic)

Metadaten

TEI-Header · CMDI · Dublin Core

Statistiken

Scans: 79
 Zeichen: ca. 133 577
 Tokens: ca. 20 004
 Oberflächentypen: ca. 5 434

Wortwolken

- Lemmata
- Lemmata (nur Nomen)
- Types
- Types (nur Nomen)

Voyant Tools

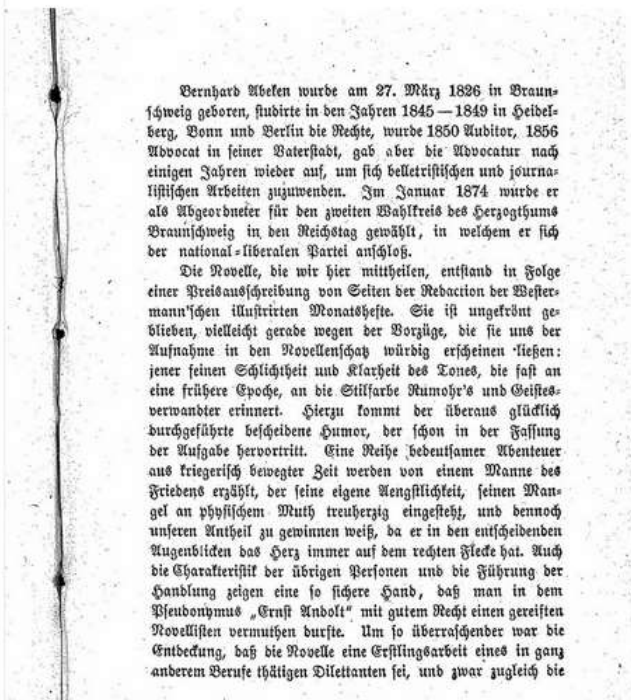
- transliterierter Text
- normalisierter Text
- lemmatisierter Text

Reiter *Texte* > *Verfügbar*

Andolt, Ernst [d. i. Bernhard Abeken]: Eine Nacht. In: Deutscher Novellenschatz. Hrsg. von Paul Heyse und Hermann I 211–287. In: Weitin, Thomas (Hrsg.): Volldigitalisiertes Korpus. Der Deutsche Novellenschatz. Darmstadt/Konstanz, 20

Bild: 0005 << vorherige Seite

+ - 1:1 fit 62%



nächste Seite >>

Bernhard Abeken wurde am 27. März 1826 in Braunschweig geboren, studierte in den Jahren 1845 — 1849 in Heidelberg, Bonn und Berlin die Rechte, wurde 1850 Auditor, 1856 Advocat in seiner Vaterstadt, gab aber die Advocatur nach einigen Jahren wieder auf, um sich belletristischen und journalistischen Arbeiten zuzuwenden. Im Januar 1874 wurde er als Abgeordneter für den zweiten Wahlkreis des Herzogthums Braunschweig in den Reichstag gewählt, in welchem er sich der national-liberalen Partei anschloß. Die Novelle, die wir hier mittheilen, entstand in Folge einer Preisausschreibung von Seiten der Redaction der Westermann'schen illustrierten Monatshefte. Sie ist ungekrönt geblieben, vielleicht gerade wegen der Vorzüge, die sie uns der Aufnahme in den Novellenschatz würdig erscheinen ließen: Jener feinen Schlichtheit und Klarheit des Tones, die fast an eine frühere Epoche, an die Stillfarbe Rumohr's und Geistesverwandter erinnert. Hierzu kommt der überaus glücklich durchgeführte bescheidene Humor, der schon in der Fassung der Aufgabe hervortritt. Eine Reihe bedeutsamer Abenteuer aus kriegerisch bewegter Zeit werden von einem Manne des Friedens erzählt, der seine eigene Ängstlichkeit, seinen Mangel an physischem Muth treuherzig eingestehet, und dennoch unseren Antheil zu gewinnen weiß, da er in den entscheidenden Augenblicken das Herz immer auf dem rechten Flecke hat. Auch die Charakteristik der übrigen Personen und die Führung der Handlung zeigen eine so sichere Hand, daß man in dem Pseudonymus „Ernst Andolt“ mit gutem Recht einen gereiften Novellisten vermuthen dürfte. Um so überraschender war die Entdeckung, daß die Novelle eine Erstlingsarbeit eines in ganz anderem Berufe thätigen Dilettanten sei, und zwar zugleich die

Suche im Werk

Hilfe

Informationen zum Werk

- Titeldaten
- Verfügbarkeit Text (TEI-XML-, HTML-, TCF-, E-Book-Fassung): Namensnennung 4.0 International (CC BY 4.0)
- Nutzungsbedingungen

Download dieses Werks

XML (TEI P5) • HTML • Text
 TCF (text annotation layer)
 TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)
 XML (TEI P5 inkl. att.linguistic)

Metadaten zum Werk

TEI-Header • CMDI • Dublin Core

Ansichten dieser Seite

- gerendertes HTML (normierte Zeichen)
- gerendertes HTML (Originalzeichen)
- zugrunde liegendes XML/TEI
- Reintext
- normalisierte Orthographie
- Faksimile in hoher Auflösung

Voyant Tools

- transliterierter Text
- normalisierter Text



Suchen im DTA (http://www.deutschestextarchiv.de/doku/DDC-suche_hilfe)

Anmelden (DTAQ) DWDS [dlexDB](#) [CLARIN-D](#)

DTA [Hilfe](#) Texte ▾ Projekt ▾ Dokumentation ▾ Impressum

in den Titeldaten im Korpus in der Dokumentation






Suche im Deutschen Textarchiv

Treffer 1 - 10 von 95444

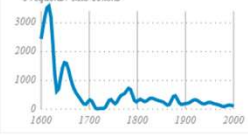
Neue Suche · KWIC [Hilfe](#)

10 Treffer pro Seite **Sortierung:** Datum aufsteigend/absteigend · zufällig

gehe zu: Anfang · -10 · -5 · vorherige · nächste · +5 · +10 · Ende

-  1: Abel, Heinrich Kaspar: Wohlerfahner Leib-Medicus der Studenten. Leipzig, 1699. #162
Und aus eben dem Fundament haben vormahls die Römer das Fieber wie einen **Gott** geehret / ihme auch Capellen erbauet und angebeten / nur damit es ihnen keinen Schaden thun möge. billiger aber ist es eine Furia als eine Göttin zu nennen. [mehr]
-  2: Achenwall, Gottfried: Abriß der neuesten Staatswissenschaft der vornehmsten Europäischen Reiche und Republicken. Göttingen, 1749. #296
Die Ursache fügt Art. 16. des Königlichen Gesetzes hinzu: damit die Könige von Dänemark und Norwegen es für den größten Ruhm schätzen, sich **Gott** zu unterwerfen, und vor die allerhöchste und grösste Macht halten, von dem allerhöchsten **Gott** durch die Diener seines Wortes zu glücklichem Antritt ihrer Regierung durch einen heiligen Segen geweyhet zu werden. [mehr]
-  3: Achenwall, Gottfried: Abriß der neuesten Staatswissenschaft der vornehmsten Europäischen Reiche und Republicken. Göttingen, 1749. #297
Nunmehr ist der König von Dänemark an kein menschliches Gesetz gebunden, er erkennt keinen Oberrn und Richter weder in geistlichen noch weltlichen Sachen als allein den einigen **Gott**, hat die höchste Gewalt Gesetze zu geben und abzuschaffen, ist die Qvelle aller Titul, Würden, Ehrenämter und Diensten; hat das höchste Recht des Krieges, Friedens, der Bündnisse und der Auflagen, die höchste Gewalt in geistlichen Sachen, und stehet bey ihm allein, alle Rechte und Regalien der Majestät zu seinem Nutzen und Besten anzuwenden. [mehr]
-  4: Arndt, Johann: Vom wahren Christenthumb. Bd. 3. Magdeburg, 1610. #8
Gott die gantze Seele zuergeben. [mehr]
-  5: Arndt, Johann: Vom wahren Christenthumb. Bd. 3. Magdeburg, 1610. #14
Das ist / spricht er / die beständige ruhe vnd sicherheit / wenn man von den vngestümnen Sturmwinden dieser Welt erlöset wird / vnd seine Augen vnd Hertz zu **Gott** erhebt von der Erden / vnd sich mit dem Gemüte zu **Gott** nahet / (mente DEO fit proximus) versteht auch das alles was vnter Menschlichen dingen für hoch vnnd köstlich gehalten wird / in seinem Herten vnd Gemüt verborgen liegt / also dz man nichts von der Welt wünschet vnd begert / weil ein [mehr]

Verlaufskurve DTA+DWDS
Frequenz / Mio Tokens



Suchergebnisse herunterladen: Text, Text/KWIC, JSON, YAML, ATOM 1.0, RSS 2.0, TCF 0.4.

Fahren Sie über die einzelnen Tokens mit der Maus, um folgende Informationen zu sehen:

- u: Originaltext, UTF-8-kodiert
- w: approximierter Latin-1-Text
- v: CAB-normalisierte Wortform
- l: Lemma (unflektierte Form)
- p: Part-of-Speech-Analyse

Dokumentation zu Suchabfragen unter Reiter *Dokumentation* > *Linguistische Suche*

Suche nach exakter Wortform (z.B. *@Gott*)



Suchen im DTA (http://www.deutschestextarchiv.de/doku/DDC-suche_hilfe)

Dokumentation zu Suchabfragen unter Reiter *Dokumentation* > *Linguistische Suche*

Suche nach Lemma (z.B. *Gott*)

Export der Suchergebnisse, z.B. als KWIC

The screenshot shows the DTA search interface. At the top, there is a search bar with the word 'Gott' entered. Below the search bar, there are options for where to search: 'in den Titeldaten', 'im Korpus' (selected), and 'in der Dokumentation'. To the right of the search bar are links for 'Texte', 'Projekt', 'Dokumentation', and 'Impressum'. Below the search bar, the page title is 'Suche im Deutschen Textarchiv' and it shows 'Treffer 1 - 10 von 246443'. There is a secondary search bar with 'Gott' and a 'Treffer pro Seite' dropdown set to '10'. Below this, there are navigation links: 'gehe zu: Anfang · -10 · -5 · vorherige · nächste · +5 · +10 · Ende'. The search results are listed in a table with 5 rows. Each row contains a small image of a book cover, a title, a description, and a '[mehr]' link. The first result is for 'Abel, Heinrich Kaspar: Wohlerfahmer Leib-Medicus der Studenten. Leipzig, 1699. #5'. The description for the first result mentions 'GOTT'. To the right of the search results, there is a graph titled 'Verlaufskurve DTA+DWDS' showing 'Erscheinung / Mio Tokens' from 1600 to 2000. Below the graph, there is a section for 'Suchergebnisse herunterladen' with options for Text, Text/KWIC, JSON, YAML, ATOM 1.0, RSS 2.0, and TCF 0.4. Below that, there is a section for 'Fahren Sie über die einzelnen Tokens mit der Maus, um folgende Informationen zu sehen:' with a list of options: 'u: Originaltext, UTF-8-kodiert', 'w: approximierter Latin-1-Text', 'v: CAS-normalisierte Wortform', 'l: Lemma (unklieferte Form)', and 'p: Part-of-Speech-Analyse'. A green arrow points from the text 'Export der Suchergebnisse, z.B. als KWIC' to the 'Text/KWIC' download option.

Das DTA im Digitalen Wörterbuch der deutschen Sprache (DWDS)

- Wortinformationssystem mit verschiedenen Ressourcen, z.B. Wörterbüchern, Textkorpora, statistischen Auswertungen
- enthält Referenzkorpora (DWDS-Kernkorpus, DTA-Kernkorpus), Zeitungskorpora (z.B. „Die ZEIT“), Webkorpora (z.B. Blogs) und Spezialkorpora (z.B. „Gesprochene Sprache“)
- über 47 Mrd. Tokens in historischen und gegenwartssprachlichen Textkorpora (<https://www.dwds.de/d/korpora>, Stand 10.10.22)
- viele der Korpora frei zugänglich, einige nur mit Anmeldung
- automatische Annotation der Wortarten nach dem STTS
- Link: <https://www.dwds.de/>

Korpusbelege DTA-Kernkorpus (1598–1913)

Himmel

Korpus: DTA-Kernkorpus (1598–1913)

Start: 1598 **Ende:** 1913





Textklassen: Belletristik Wissenschaft Gebrauchsliteratur

Anzeige: KWIC voll maximal

Sortierung: Datum aufsteigend

Treffer pro Seite: 10

1–10 von 40942 Treffern [21859 in DTA-Erweiterungen] [62801 in DTA-Kern+Erweit.]

- 1:  **Stuer, Barnim von:** Summarischer vnd kurtzer bericht Von den Türcken vrsprung. Magdeburg, 1601.
O Gott vnser Vater/ der du bist im **Himmel**/ wir bitten dich hertzlich/ erbarm dich vnser vnd streck jtz aus deinen Arm zu vnser beschirmung/ damit dein Erbschafft nicht komme vnter die Heyden vnd vnter dem Zepter des Blutdürstigen Tyrannen des Türckens/ welcher jtzunder dein Christliche Kirchen zerreist/ schrecket/ brennet/ betrübt vnd verfolgt: Sondern mache den Türcken sampt allen seinen Anhangt zuschanden vnd zu spotte/ damit sie nicht sprechen jhr mechtige Hand habe vns ermordet/ vnd wer ist nun jhr Gott?
- 2:  **Rollenhagen, Gabriel:** Vier Bücher Wunderbarlicher biß daher vnerhörter/ vnd vngleublicher Jndianischer reysen. Magdeburg, 1603.
Vier Bücher Wunderbarlicher biß daher vnerhörter/ vnd vngleublicher Jndianischer reysen/ durch die Lufft/ Wasser/ Land/ Helle/ Paradiß/ vnd den **Himmel**.
- 3:  **Rollenhagen, Gabriel:** Vier Bücher Wunderbarlicher biß daher vnerhörter/ vnd vngleublicher Jndianischer reysen. Magdeburg, 1603.
Der lieset alle Griechische vnd Lateinische Bücher durch/ vnd was er in denselben/ von **Himmel** vnd Erden/ Menschen/ Vieh/ Bewmen/ Kreutern/ Metallen vnd Edelgestein findet/ das fasset er auff kurtzest in ein besonder Buch/ vnd nennet dasselbige die Historia der Natur.
- 4:  **Rollenhagen, Gabriel:** Vier Bücher Wunderbarlicher biß daher vnerhörter/ vnd vngleublicher Jndianischer reysen. Magdeburg, 1603.
Jch hab allezeit in willens gehabt/ auch in meinen jtzigen gefehrlichen Kriegshendeln für gut angesehen/ dieweil ich wuste/ dz du zur Weißheit sonderlich lust/ vnd derohalben dich auff die Philosophia zu studiren begeben hettest/ an Dich etwas zu schreiben von

Himmel im DWDS-Wörterbuch

Belege in Korpora

Referenzkorpora

- DWDS-Kernkorpus (1900–1999) (11966)
- DWDS-Kernkorpus 21 (2000–2010) (1335)
- DTA-Kernkorpus (1598–1913)** (40942)

Metakorpora

- DTA-Kern+Erweit. (1465–1969) (66455)
- Historische Korpora (1465–1998) (278466)
- Referenz- und Zeitungskorpora (frei) (130210)

Zeitungskorpora

- Berliner Zeitung (1994–2005) (10717)
- Der Tagesspiegel (ab 1996) (8180)
- Die ZEIT (1946–2018) (31558)

Webkorpora

- Blogs (5843)

Spezialkorpora

- DTA-Erweiterungen (1465–1969) (25513)
- Archiv der Gegenwart (1931–2000) (421)
- Polytechnisches Journal (475)
- Filmuntertitel (7984)
- Gesprochene Sprache (122)
- DDR (1288)
- Politische Reden (1982–2020) (496)
- Bundestagskorpus (1949–2017) (2472)
- A. v. Humboldts Publizistik (dt. 1790–1859)

Das DTA
im DWDS



- Dokumentation der Suche unter: <https://www.dwds.de/d/korpussuche>
- Suche nach Lemma (z.B. **Haus** → findet **Haus, Häuser** etc.) oder exakter Wortform (z.B. **@Haus** → findet nur **Haus**)
- * als Platzhalter für null bis unendlich viele Zeichen, z.B. **Haus*** → findet **Haus, Hausmeister** etc.
- Wortgruppen in Anführungsstrichen, z.B. „**ein Haus**“ → findet **ein Haus, eines Hauses** etc.; „**@ein @Haus**“ → findet nur **ein Haus**
- Suche nach Wortart durch **\$p**; Verbindung von Wort und Wortart durch **WITH** oder **WITHOUT**, z.B. **Müller WITHOUT \$p=NE** → findet Müller nicht als Eigenname (sondern als Berufsbezeichnung)
- Filtern nach Metadaten, z.B. nach Textsorten über **textClass**, z.B.
Haus #HAS[textClass,/Roman/]

Die Referenzkorpora zur deutschen Sprachgeschichte



Die Referenzkorpora zur deutschen Sprachgeschichte (Dipper 2015: 522–523)

- aus der ursprünglich gemeinsam geplanten Initiative *Deutsch Diachron Digital* (DDD) hervorgegangene DFG-Projekte
- Referenzkorpus Altdeutsch (ReA), Mittelhochdeutsch (ReM), Frühneuhochdeutsch (ReF), Mittelniederdeutsch/Niederrheinisch (ReN), Deutsche Inschriften (ReDI)
- Korpusstruktur, Metadaten und Annotationen ähnlich → Ziel: korpusübergreifende Suchen
- publiziert im Tool ANNIS
- Annotationen: Lemma, Wortart, Flexionsmorphologie → halbautomatische Annotation, d.h. mit manueller Korrektur! (Hinweis: im ReF ist großer Teil nur automatisch annotiert)

Dipper 2015 = Dipper, Stefanie (2015): Annotierte Korpora für die Historische Syntaxforschung: Anwendungsbeispiele anhand des Referenzkorpus Mittelhochdeutsch. *Zeitschrift für Germanistische Linguistik* 43(3), 516–563.



Die Referenzkorpora zur deutschen Sprachgeschichte

- **ähnlich** in Korpusstruktur, Metadaten, Textaufbereitung und Annotation, **nicht identisch!**
- Beispiele:
 - unterschiedliche Tagsets: HiTS im ReA, ReM, ReF und ReDI, HiNTS im ReN (= Modifikation des HiTS) → Vergleich der Tags in **PoS-Vergleich_HiTS_HiNTS** unter https://www.fdr.uni-hamburg.de/record/9195/files/PoS-Vergleich_HiTS_HiNTS.pdf?download=1
 - teilweise unterschiedliche Text- und Annotationsebenen, z.B. **tok_dipl** und **tok_anno** im ReM, ReF, ReN, ReDI ggü. **edition** und **text** im ReA
 - teilweise unterschiedliche Metadaten, z.B. **topic** (z.B. **Religion**) im ReA, ReM, ReN, ReDI ggü. **text-type** (z.B. **RE**) im ReF

→ zu Gemeinsamkeiten und Unterschieden der Referenzkorpora s.

Ihden, Sarah / Schnelle, Gohar / Schröder, Ingrid / Zeige, Lars Erik: Der Verbund ‚Deutsch Diachron Digital – Referenzkorpora zur deutschen Sprachgeschichte‘. Strategien der Erschließung, Analyse und nachhaltigen Nutzung historischer Sprachdaten. In: Kupietz, Marc / Schmidt, Thomas: Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022. Tübingen (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 11).



Suchen in den Referenzkorpora in ANNIS

ANNIS interface showing search results for the query `pos="NE"`. The interface includes a search bar, a corpus list, and a list of search results with their respective paths and contexts.

Search query: `pos="NE"`

45 matches in 1 document

Visible: ren-anno

Name	Texts	Total
ReN_anno_2021-01-06_baltisch_13.2	1	1
ReN_anno_2021-01-06_baltisch_14.1	1	1
ReN_anno_2021-01-06_baltisch_14.2	1	1
ReN_anno_2021-01-06_baltisch_14.2-15.2	1	1
ReN_anno_2021-01-06_baltisch_15.1	2	2
ReN_anno_2021-01-06_baltisch_15.2	3	3
ReN_anno_2021-01-06_baltisch_16.1	2	2
ReN_anno_2021-01-06_elbstfaelisch_13.1-13.2	2	2
ReN_anno_2021-01-06_elbstfaelisch_13.2	3	3
ReN_anno_2021-01-06_elbstfaelisch_14.1	1	1

Search Results:

- Path: ReN_anno_2021-01-06_baltisch_13.2 > Nowg_Schra_Rig. (tok_dipl 6 - 16)
rechtes in deme hove to **nogarden** . D At si wetelic
- Path: ReN_anno_2021-01-06_baltisch_13.2 > Nowg_Schra_Rig. (tok_dipl 69 - 79)
deme hove der dhudesscen to **Nogarden** . bescreven is . alsdus
- Path: ReN_anno_2021-01-06_baltisch_13.2 > Nowg_Schra_Rig. (tok_dipl 128 - 138)
wanne se komet in de **Nv** . so solen se oldermanne
- Path: ReN_anno_2021-01-06_baltisch_13.2 > Nowg_Schra_Rig. (tok_dipl 141 - 151)
dhes hoves vnde sen te **peteres** vnder sic seluen de dar

Help us make ANNIS better! not logged in [Login](#)

pos="NE"

45 matches in 1 document

Visible: ren-anno

7 items with a total sum of 45 (query on ReN_anno_2021-01-06_baltisch_13.2)

rank	#1 morph	count
1	Masc.Dat.Sg	14
2	Masc.Nom.Sg	11
3	Masc.Gen.Sg	7
4	*.Dat.Sg	6
5	Fem.Akk.Sg	5
6	Neut.Dat.Sg	1
7	*.Dat.Akk.Sg	1



Export

Help us make ANNIS better! not logged in Login

Help/Examples Query Result x Frequency Analysis x Export x

pos="NE"

Query Builder

Search More History

45 matches in 1 document

Corpus List Search Options

Visible: ren-anno

Name	Texts	Top
ReN_anno_2021-01-06_baltisch_13.2	1	
ReN_anno_2021-01-06_baltisch_14.1	1	
ReN_anno_2021-01-06_baltisch_14.2	1	
ReN_anno_2021-01-06_baltisch_14.2-15.2	1	
ReN_anno_2021-01-06_baltisch_15.1	2	
ReN_anno_2021-01-06_baltisch_15.2	3	
ReN_anno_2021-01-06_baltisch_16.1	2	
ReN_anno_2021-01-06_elbstfaelisch_13.1-13.2	2	
ReN_anno_2021-01-06_elbstfaelisch_13.2	3	
ReN_anno_2021-01-06_elbstfaelisch_14.1	1	

Exporter: CSVMultiTokEx

Left Context: 5

Right Context: 5

Annotation Keys: ?

Parameters: ?

Parameters: *metakeys* - comma separated list of all meta data to include in the result (e.g. *metakeys*=title, documentname)

Perform Export Cancel Export Download

The CSV MultiTok Exporter exports only the values of the elements searched for by the user, ignoring the context around search results. The values for all annotations of each of the found nodes is given in a comma-separated table (CSV). This exporter will take more time than the normal CSV Exporter but it is able to export the underlying text for spans if the corpus contains multiple tokenizations.

Einsatz von DTA und Referenzkorpora in der Lehre



Referenzkorpora zur deutschen Sprachgeschichte

Möglichkeiten

- Breite der hoch- und niederdeutschen Varietäten von ca. 750–1650 → sprachstufenübergreifende variationssensitive Analysen v.a. grammatischer Phänomene → besonders geeignet für Seminare zu Sprachwandel (v.a. grammatischer Wandel)
- Metadaten: Suche möglich in ähnlichen Textgruppen und -sorten
- Annotation von PoS & Flexion: Suche möglich nach selbem Phänomen (ggf. abweichendes Tag)
- ANNIS: mächtiges Tool, umfangreiche Abfragemöglichkeiten

Herausforderungen

- nicht alle Korpora in einer gemeinsamen Instanz verfügbar
- Anpassung der Suchabfrage an jeweiliges Korpus (Metadaten, PoS-Tags, flexionsmorphologische Tags etc.)
- ANNIS: nicht intuitiv, Dokumentation (v.a. der Abfragemöglichkeiten, Operatoren etc.) könnte anschaulicher / verständlicher sein

Möglichkeiten

- zeitliche Breite: Kernkorpus v.a. 1600 bis 1900 → zeitlicher Anschluss an die Referenzkorpora
- Metadaten: Suche möglich in verschiedenen Textgruppen und -sorten → Anschluss an die Referenzkorpora
- DWDS: umfangreiche Abfragemöglichkeiten, sehr gute Dokumentation

Herausforderungen

- Annotation: Suchen eingeschränkter (nur PoS, nicht Flexion); gewisse Fehlerquote durch nur automatische Annotation
- Textgruppen: v.a. in früheren Zeiträumen (bis Mitte 18. Jh.) Korpus nicht nach Belletristik, Gebrauchsliteratur und Wissenschaft ausgewogen (Gebrauchsliteratur dominiert)

Vielen Dank für die Aufmerksamkeit

Anhang



Links zu den Referenzkorpora und den Handbüchern

ReA

Korpus in ANNIS: <https://korpling.german.hu-berlin.de/annis3/ddd>

Handbuch: <https://www.deutschdiachrondigital.de/manual/>

ReM

Korpus in ANNIS: <https://annis.linguistics.rub.de/REM/>

Handbuch: <https://linguistics.rub.de/forschung/arbeitsberichte/19.pdf>

ReF

Korpus in ANNIS: <https://annis.linguistics.rub.de/REF/>

Handbuch: <https://linguistics.rub.de/forschung/arbeitsberichte/24.pdf>

ReN

Korpus in ANNIS: <https://annis.fdm.uni-hamburg.de/annis-gui-3.6.0/ren>

Handbücher beim FDM: <http://doi.org/10.25592/uhhfdm.9195>



Links zu den Video-Tutorials zum ReN

Tutorials zur Nutzung des Referenzkorpus Mittelniederdeutsch/Niederrheinisch:

<https://www.slm.uni-hamburg.de/ren/korpus/tutorials.html>

Vergleich der Referenzkorpora zur deutschen Sprachgeschichte

ReA	ReM	ReF	ReN
alle überlieferten Texte (ohne Glossen) des Ahd. und As. (ca. 750–1050)	strukturierte Auswahl mhd. Texte (ca. 1050–1350)	strukturierte Auswahl frnhd. Texte (ca. 1350–1650)	strukturierte Auswahl mnd. und nr. Texte (ca. 1200–1650)
ca. 530.000 Token (Suche nach „edition“ in allen DDD-Subkorpora) (Stand: 26.08.22)	ca. 2,4 Mio. annotierte Token (Suche nach „tok_anno“ in allen Subkorpora) (Stand: 26.08.22)	3,4 Mio. annotierte Token, davon ca. 540.000 manuell korrigiert (Stand: 26.08.22)	ca. 1,5 Mio. annotierte Token + ca. 800.000 transkribierte Token (Suche nach „tok_anno“ in allen Subkorpora von „ren-anno“ bzw. „ren-trans“) (Stand: 26.08.22)

Zum Vergleich der Referenzkorpora (inkl. ReDI) s. auch:

Ihden, Sarah / Schnelle, Gohar / Schröder, Ingrid / Zeige, Lars Erik (i. E.): Der Verbund ‚Deutsch Diachron Digital – Referenzkorpora zur deutschen Sprachgeschichte‘. Strategien der Erschließung, Analyse und nachhaltigen Nutzung historischer Sprachdaten. In: Kupietz, Marc / Schmidt, Thomas: Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022. Tübingen (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 11).



Korpusstruktur

ReA	ReM	ReF	ReN
Keine nach Parametern strukturierte Textauswahl, sondern alle überlieferten Texte (ohne Glossen) des Ahd. und As.	Struktur nach den Parametern <i>Zeit</i> (50-Jahres-Abschnitte), <i>Sprachraum</i> (z.B. obd.), <i>Textart</i> (Vers, Prosa, Urkunden)	Struktur nach den Parametern <i>Zeit</i> (50-Jahres-Abschnitte), <i>Sprachraum</i> (z.B. obd.), <i>Textart</i> (Vers, Prosa, Urkunden)	Struktur nach den Parametern <i>Zeit</i> (50-Jahres-Abschnitte), <i>Sprachraum</i> (z.B. <i>lübisch</i>), <i>Feld der Schriftlichkeit</i> (z.B. Wissensvermittlung)
Metadaten u.a. zu <i>time</i> (z.B. 9,1 für 1. H. 9. Jh.), <i>language_area</i> (z.B. obd.), <i>form</i> (z.B. Prosa), <i>topic</i> (z.B. Religion)	Metadaten u.a. zu <i>time</i> (z.B. 12,1 für 1. H. 12. Jh.), <i>language-type</i> (z.B. oberdeutsch), <i>language-region</i> (z.B. westoberdeutsch), <i>language-area</i> (z.B. alemannisch), <i>genre</i> (z.B. P für Prosa), <i>topic</i> (z.B. Religion)	Metadaten u.a. zu <i>time</i> (z.B. 12,1 für 1. H. 12. Jh.), <i>language-type</i> (z.B. oberdeutsch), <i>language-region</i> (z.B. westoberdeutsch), <i>language-area</i> (z.B. hochalemannisch), <i>genre</i> (z.B. P für Prosa), <i>text-type</i> (z.B. RG für Rechts- und Geschäftstexte)	Metadaten u.a. zu <i>time</i> (z.B. 15/1 für 1. H. 15. Jh.), <i>language-area</i> (z.B. westfaelisch), <i>genre</i> (z.B. P für Prosa), <i>topic</i> (z.B. Religion), <i>topic_ReN</i> (z.B. W für Wissensvermittlung)



ReA	ReM	ReF	ReN
a) nach der handschriftengetreuesten gedruckten Edition (edition), b) in standardisierten Token, z.B. mit Normalisierung von Getrennt- und Zusammenschreibung (text)	a) nach der Handschrift (tok_dipl), b) nach der annotierten Worteinheit, d.h. mit normalisierter Getrennt- und Zusammenschreibung (tok_anno), c) in einer normalisierten Wortform, z.B. mit Normalisierung bestimmter Grapheme wie <u> und <v> (norm)	a) nach der Handschrift (tok_dipl), b) nach der annotierten Worteinheit, d.h. mit normalisierter Getrennt- und Zusammenschreibung (tok_anno)	a) nach der Handschrift (tok_dipl), b) nach der annotierten Worteinheit, d.h. mit normalisierter Getrennt- und Zusammenschreibung (tok_anno)



Annotationsebenen

	ReA	ReM	ReF	ReN
Textebenen	<code>edition</code> (z.B. <i>sint</i>) <code>text</code> (z.B. <i>sint</i>)	<code>tok_dipl</code> (z.B. <i>Uór</i>) <code>tok_anno</code> (z.B. <i>Uor</i>) <code>norm</code> (z.B. <i>vore</i>)	analog zum ReM geplant	<code>tok_dipl</code> (z.B. <i>böc</i>) <code>tok_anno</code> (z.B. <i>böc</i>) <code>token</code> (z.B. <i>*LN_böc*</i>) und weitere z.B. zu Auflösungen von Abkürzungen, Streichungen etc.
Lexikalische Ebenen	<code>lemma</code> (z.B. <i>sin</i>) <code>translation</code> (z.B. <i>sein</i>) <code>lang</code> (z.B. <i>goh</i>)	<code>lemma</code> (z.B. <i>vor(e)</i>) <code>lemmald</code> (z.B. <i>213246000</i>) <code>lemmaLemma</code> (z.B. <i>vor(e)</i>)	<code>lemma</code> (z.B. <i>gehorsam</i>) <code>lemmald</code> (z.B. <i>GG05091</i>)	<code>lemma_wsd</code> (z.B. <i>bôk¹</i>) <code>lemma</code> (z.B. <i>bôk</i>) <code>lemma_simple</code> (z.B. <i>bok</i>) <code>lemma_var</code> (z.B. <i>up,uppe</i> für <code>lemma_simple up(pe)</code>)
Wortartebenen	<code>pos</code> (z.B. <i>VAFIN</i>) <code>posLemma</code> (z.B. <i>VA</i>)	<code>pos</code> (z.B. <i>APPR</i>) <code>posLemma</code> (z.B. <i>AP</i>)	<code>pos</code> (z.B. <i>ADJD</i>) <code>posLemma</code> (z.B. <i>ADJ</i>)	<code>pos</code> (z.B. <i>NA</i>) <code>posLemma</code> (z.B. <i>NA</i>)



Annotationsebenen

	ReA	ReM	ReF	ReN
Flexions- morpho- logische Ebenen	<code>inflection</code> (z.B. <i>IND_PRES_PL_3</i>) <code>inflectionClass</code> (z.B. <i>IRR</i>) <code>inflectionClassLemma</code>	<code>inflection</code> (z.B. <i>Dat.Pl</i>) <code>inflectionClass</code> (z.B. <i>st.Neut</i>) <code>inflectionClassLemma</code>	<code>inflection</code> (z.B. <i>Pos.*.*.*</i>)	<code>morph</code> (z.B. <i>Neut.Nom.Sg</i>)
Syntaktische Ebenen	<code>clause</code> (z.B. <i>CF_U_M</i>) → als Spanne	<code>punc</code> (z.B. <i>DE</i>) → am Token	<code>tok_anno</code> (.) als zusätzliches Token	<code>bound_sent</code> (z.B. <i>Satz</i>) → als Spanne
Textstrukt- relle Ebenen	<code>document</code> (z.B. <i>(Althochdeutsche Benediktinerregel)</i>) <code>chapter</code> <code>verse</code>	<code>reference</code> (z.B. <i>46ra,1</i>)	<code>reference</code> (z.B. <i>001ra,01</i>)	<code>reference</code> (z.B. <i>2v,01</i>)
Sonstige	-	-	-	<code>comment</code> (z.B. <i>Irrelevanzkonditionalsatz</i>)