

Dokumentation der Annotationsrichtlinien für das Korpus „Frühneuzeitliche Fürstinnenkorrespondenzen im mitteldeutschen Raum“

Für die Annotation von 600 Briefen, die Fürstinnen aus dem ostmitteleutschen Raum während drei Jahrhunderten (1546 bis 1756) mit männlichen Briefpartnern wechselten, wurde jeder Brief mit einem eindeutigen Dateinamen und Metadaten versehen sowie auf einer Mehrebenenmatrix (mit dem Computerprogramm EXMARaLDA¹) mit linguistischen und formalen Informationen angereichert, die im Folgenden näher erläutert werden (verwendete Kürzel und Tags sind grau hinterlegt).

Abkürzungen der BriefschreiberInnen in den Dateinamen:

Die Dateinamen setzen sich aus dem Kürzel der Absenderin/des Absenders, dem Kürzel der Adressatin/des Adressaten sowie dem Datum in Jahr, Monat und Tag zusammen (jeweils getrennt durch Unterstrich).

- AD* = Anna Dorothea von Sachsen-Weimar (1657–1704), Äbtissin von Quedlinburg
Anna = Anna von Sachsen-Coburg (1567–1613), geb. Kursachsen, gesch. Sachsen-Coburg
Aug = August (1526–1586), Kursachsen
B = Bernhard von Anhalt (1571–1596)
BCS = Bernhardine Christiane Sophie von Schwarzburg-Rudolstadt (1724–1757), geb. Sachsen-Weimar-Eisenach
CA = Caroline Amalia von Sachsen-Hildburghausen (1700–1758), geb. Erbach-Fürstenau
DM = Dorothea Maria von Sachsen-Weimar (1574–1617), geb. Anhalt
Doehner = Gottlieb Benjamin Doehner, Assessor
DS = Dorothea Susanna von Sachsen (1544–1592), geb. Pfalz-Simmern
E = Elisabeth von Sachsen (1540–1594), geb. Pfalz-Simmern
EA = Ernst August I. von Sachsen-Weimar-Eisenach (1688–1748)
EAC = Ernst August II. Constantin von Sachsen-Weimar-Eisenach (1737–1758)
EAMS = Elisabeth Albertine von Mecklenburg-Strelitz (1713–1761), geb. Sachsen-Hildburghausen
EASL = Ernestine Albertine von Schaumburg-Lippe (1722–1769), geb. Sachsen-Weimar-Eisenach
ED = Eleonore Dorothea von Sachsen-Weimar (1602–1664), geb. Anhalt-Dessau
EF2 = Ernst Friedrich II. von Sachsen-Hildburghausen (1707–1745)
EF3C = Ernst Friedrich III. Carl von Sachsen-Hildburghausen (1727–1756)

¹ „Extensible Markup Language for Discourse Annotation“ (<http://www.exmaralda.org>) entwickelt am Sonderforschungsbereich „Mehrsprachigkeit“ (SFB 538) der Universität Hamburg, weitergeführt vom Hamburger Zentrum für Sprachkorpora (<http://www.corpora.uni-hamburg.de/>) und dem Archiv für gesprochenes Deutsch am Institut für Deutsche Sprache Mannheim (<http://agd.ids-mannheim.de/index.shtml>).

- F* = Friedrich von Sachsen-Weimar (1596–1622)
- F3* = Friedrich III., Kurfürst von der Pfalz (1515–1576)
- FW* = Friedrich Wilhelm von Sachsen-Weimar (1562–1603)
- J3* = Johann III. von Sachsen-Weimar (1570–1605)
- JC* = Johann Casimir von Sachsen-Coburg (1564–1633)
- JCh* = Johanna Charlotte von Sachsen-Weimar-Eisenach (1693–1751)
- JCPf* = Johann Casimir von der Pfalz (1543–1592)
- JE* = Johann Ernst von Sachsen-Eisenach (1566–1638)
- JE1* = Johann Ernst I. von Sachsen-Weimar (1594–1626)
- JE2* = Johann Ernst II. von Sachsen-Weimar (1627–1683)
- JF1* = Johann Friedrich der Ältere von Sachsen (1503–1554)
- JF2* = Johann Friedrich der Mittlere von Sachsen-Coburg (1529–1595)
- JG1* = Johann Georg I. von Anhalt-Dessau (1567–1618)
- L1* = Ludwig I. von Anhalt-Köthen (1579–1650)
- L6* = Ludwig VI., Kurfürst von der Pfalz (1539–1583)
- Maior* = Johann Maior, Superintendent von Sachsen-Weimar (1564-1654)
- MSW* = Maria von Sachsen-Weimar (1571–1610), Äbtissin von Quedlinburg
- SCA* = Sophie Charlotte von Sachsen-Weimar-Eisenach (1713–1743), geb. Markgräfin von Brandenburg-Bayreuth
- Sib* = Sibylla von Sachsen (1512–1554), geb. Jülich-Kleve-Berg
- VH* = Volkmar Happe (fl. 1678), Kanzler von Sachsen-Weimar
- W* = Wilhelm der Große von Sachsen-Weimar (1598–1662)
- WCh* = Wilhelmine Christiane von Schwarzburg-Sondershausen (1658–1712), geb. Sachsen-Weimar
- WE* = Wilhelm Ernst von Sachsen-Weimar (1662–1728)
- WvA* = Wolfgang von Anhalt (1492–1566)

Metadaten

Die Metadaten enthalten Informationen über die Korrespondenzpartner sowie formale Anmerkungen zu den jeweiligen Briefen. Sie können bei der Suche dazu genutzt werden, das Korpus (oder die Unterkorpora der einzelnen Korrespondenzen) nach bestimmten Parametern (wie Datum, Familienstand oder Annotationsart) einzugrenzen.

Die verwendeten Kategorien (in alphabetischer Reihenfolge) sind:

<i>arch</i> (archive)	= Nennung des Archivs, in dem das Dokument aufgefunden wurde
<i>class</i>	= Auskunft über die Standeszugehörigkeit der Absenderin/des Absenders
<i>comment</i>	= Informationen über die Schreiberhand sowie Nennung von AbsenderIn und AdressatIn mit Adelstitel und dem Verhältnis, das zwischen den Korrespondierenden besteht
<i>corpus</i>	= Information zur Annotationsart: NA = normale Annotation EA = erweiterte Annotation EH = eigenhändig FH = Schreibung von fremder Hand
<i>date</i>	= Datum des Briefs
<i>dob</i> (date of birth)	= Geburtsdatum der Absenderin/des Absenders
<i>length</i>	= Brieflänge in Zeichen (inklusive Leerzeichen)
<i>loc_1</i> (location of sender)	= Abfassungsort des Briefes
<i>loc_2</i> (location of addressee)	= Empfangsort des Briefes (soweit bekannt)
<i>marit</i> (marital status)	= Angaben zum Personenstand
<i>page</i>	= Seitenzahl(en) des Originaldokumentes
<i>pers1</i> (sender)	= Name der Absenderin/des Absenders (inklusive Lebensdaten und Adelstitel)
<i>pers2</i> (addressee)	= Name der Adressatin/des Adressaten (inklusive Lebensdaten und Adelstitel)
<i>projectName</i>	= lautet immer „Fürstinnenkorrespondenzen“
<i>prov</i> (provenance)	= regionale Herkunft der Absenderin/des Absenders
<i>relat</i> (relation)	= Verhältnis zwischen AbsenderIn und EmpfängerIn
<i>transcriptionName</i>	= Transkriptionsname – wie oben beschrieben – aus dem Namens Kürzel der Absenderin/des Absenders, dem Namens Kürzel der Adressatin/des Adressaten und dem Datum des Briefes (JJJJ_MM_TT)

Annotationsebenen

Für alle Ebenen der normalen Annotation werden die jeweils verwendeten Tagsets hier im Einzelnen aufgelistet. Im Anschluss daran sind die zusätzlich für die erweiterte Annotation verwendeten Spuren und Tags aufgeführt.

Normale Annotation:

- text* = **Tokenisierter Originaltext**, der sich so nah wie möglich an die Vorlage hält; vorgefundene Zusammenschreibungen werden hier getrennt aufgeführt, durchgestrichene Textpassagen durch „#“ ersetzt und Silbentrennung am Zeilenende nicht durchgeführt
- orig* = **Originalspur** (*original*): hier erscheinen originale Zusammenschreibungen, durchgestrichene Textpassagen und Silbentrennung am Zeilenende sowie Abkürzungen, die mit % dargestellt werden.
- norm* = **Normierte (neuhochdeutsche) Textspur**², die den Einsatz des Annotationswerkzeuges TreeTagger³ für die Lemmatisierung und die Part-of-Speech-Zuweisung erlaubt.
- lemma* = Die automatische **Lemmazuweisung** durch den Treetagger wurde durch manuelle Überarbeitung von Fehlzuweisungen bereinigt und optimiert. Fremdsprachliches Material und untergegangene Lemmata sind mit Rauten gekennzeichnet und auf der mean-Spur erläutert.²
- pos* = Die **Part-of-Speech-Zuweisung** erfolgte ebenfalls zunächst durch den Treetagger nach dem Stuttgart-Tübingen-Tagset (STTS)⁴ und wurde anschließend manuell überprüft. Der vom Treetagger zugewiesene Tag PROAV für Pronominaladverbien (statt POV wie von STTS vorgegeben) wurde nicht verändert. Potentiell können alle Tags mit dem Zusatz A versehen sein, der darauf hinweist, dass bei dem entsprechenden Token eine Abkürzung verwendet wurde, außerdem wurden movierte Personenbezeichnungen mit NNMOV bzw. NEMOV und Diminutiva mit NNDIM ausgezeichnet (insgesamt wurden in den Briefen 89 Tags eingesetzt).
- morph* = Auch die **morphologische Analyse** (*morphology*) erfolgte nach den Vorgaben von STTS⁵, wobei auf die Ausweisung der Deklinationsart bei Adjektiven verzichtet wurde. Unklare oder mehrdeutige Parameter sind mit einem Asterisk ausgezeichnet.

² Als Referenzwörterbücher wurden „Duden – Die deutsche Rechtschreibung“ (24. Aufl., 2006) und „Duden – Das große Wörterbuch der Deutschen Sprache“ (3. Aufl., 2000 – beide Mannheim: Bibliographisches Institut & F.A. Brockhaus) verwendet.

³ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

⁴ <http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html>

⁵ <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>

graph = Bei **graphematischen Besonderheiten** wurde die vorgefundene Schreibung durch Unterstrich der neuhochdeutschen gegenübergestellt.

phon = **Phonologische Besonderheiten** (*phonology*) erhielten die folgenden Tags:

Ana = Anaptyxe

Apo = Apokope

Enkl = Enklise

Prokl = Proklise

Met = Metathese

Syn = Synkope

≠Ana = nicht durchgeführte Anaptyxe

≠Apo = nicht durchgeführte Apokope

≠Met = nicht durchgeführte Metathese

≠Syn = nicht durchgeführte Synkope

quot = Bibelzitate und Sprichwörtliches wurden als **phraseologische Besonderheiten** (quotation) mit den Tags **BZ** und **SW** ausgezeichnet.

lex_gr = **Lexikalische und grammatische Besonderheiten** (*lexicology and grammar*) wurden auf dieser Spur ausgezeichnet, wobei bei den nachfolgenden sieben Grammatik-Tags jeweils eine Erläuterung auf der comment-Spur erfolgte:

Temp = Tempus (bei periphrastischen Zeitformen⁶)

Gen = Genus

Num = Numerus

Kas = Kasus

Mod = Modus

GV = Genus verbi (bei Passivkonstruktionen)⁶

Komp = Komparation

Pers = Person

D = dialektale Besonderheiten (wenn möglich mit Angabe des Dialektraumes)

G = andere grammatische Auffälligkeiten)

E1 = Endungsbesonderheiten

E2 = untergegangene Adverbendungen

NF = Nebenform mit Dudeneintrag; als Lemma erscheint die Hauptform

WB = Wortbildung (bei Komposita, Konversionen Ableitungen)

WS = Wortschatz

⁶ Angabe nur bei der erweiterten Annotation

mod_polite = Bei der Kennzeichnung von **Bescheidenheits- und Höflichkeitssignalen** (*modesty and politeness*) wurde jeweils zuerst festgehalten, ob es sich um eine Selbstnennung (n1), eine Nennung des Adressaten/der Adressatin (n2) oder die Erwähnung einer dritten Person bzw. Personengruppe (n3) handelt. Die Tags n1 und n2 sind außerdem noch spezifiziert, wobei n*-pn für pronominale und n*-form für formale Anreden steht. Anreden mit Verwandtschaftsbezeichnungen (n*-fam) und Namen (n*-nom) wurden getrennt getaggt, ebenso die Verwendung von anderen Bezeichnungen für die eigene oder angeredete Person (n*-voc [vocabulum = Bezeichnung, Benennung]). Emotionale und formale Komplimente (wie Bescheidenheitstopoi, Bekundungen der Dienstwilligkeit, Erhöhung der anderen und Geringersetzung der eigenen Person) mit x getaggt.

mean = Die **Bedeutung** (*meaning*) von untergegangenen oder fremdsprachigen Lemmata ist auf dieser Spur angegeben.

comment = Auf der Kommentar-Spur erfolgte eine genaue Bezeichnung der unter *lex_gr* getaggtten Auffälligkeiten

_ entspricht „statt“ (z. B. Acc_Dat = Akkusativ statt Dativ)

: entspricht „oder“ (z. B. Masc:Neut = Neutrum oder Masculinum)

- entspricht „ohne“ (z. B. -ART = ohne Artikel).

Die Tags für Periphrastische Tempusformen lauten:

Perf = Perfekt

Plq = Plusquamperfekt

Fut1 = Futur I

Fut2 = Futur II

formal = **Formale Besonderheiten** wurden auf dieser Spur ausgezeichnet (wenn nicht das gesamte Wort betroffen war, wurden die entsprechenden Bestandteile mit Doppelpunkt vor den Tag gesetzt):

BM = Binnenmajuskel

BS = Bindestrichschreibung

D = durchgestrichen

GK = Groß- bzw. Kleinschreibung

GS = Getrenntschreibung

HG = hochgestellte Textteile

KV = Veränderungen in der Konsonantenschreibung

L = lateinische Schrift

MJ = Schreibung in Majuskeln oder Kapitälchen

MOV = Movierung

MZ	=	Mehrfachschreibung bei Silbentrennung am Zeilenende
RE	=	Randergänzung
KA	=	kalligraphische Ausschmückung
SF	=	Schreibfehler
SU	=	Schreibung unklar
TG	=	tiefgestellte Textteile
TTF	=	Textteile fehlen
US	=	Unterstreichung
USF	=	Unterstreichung von fremder Hand
VV	=	Veränderungen in der Vokalschreibung
ZS	=	Zusammenschreibung

manu = Eigenhändige Textteile sind nicht näher ausgezeichnet. Text von fremder Hand wurde wie folgt beschrieben:

AFH	=	Abschrift von fremder Hand
DFH	=	Diktat
FH	=	fremde Hand
KEH	=	Kopie von eigener Hand
KFH	=	Kopie von fremder Hand

Erweiterte Annotation:

Für die erweiterte Annotation wurden fünf zusätzliche Auszeichnungsspuren verwendet, deren Tags im Folgenden aufgelistet sind.

grfunct = Das Tagset der **syntaktischen Analyse/grammatischen Funktion** (*grammatical function*) setzt sich folgendermaßen zusammen:

subj	=	Subjekt
voc	=	Vokativ/Anredenominativ
acc-o	=	Akkusativobjekt
dat-o	=	Dativobjekt
gen-o	=	Genitivobjekt
prep-o	=	Präpositionalobjekt
prsimpl	=	einfaches Prädikat
prcompl	=	komplexes Prädikat
pred	=	Prädikativ
attr	=	Attribut
tempadv	=	Temporaladverbiale

locadv	=	Lokaladverbiale
modadv	=	Modaladverbiale
causadv	=	Kausaladverbiale
textadv	=	Textadverbiale
satzadv	=	Satzadverbiale
neg	=	Negation
expans	=	Expansion

Bei den Nebensätzen sind Form und Funktion durch Doppelpunkt miteinander verbunden.

formal:

inf-ph	=	satzwertige Infinitivphrase
part-ph	=	satzwertige Partizipphrase
pronom	=	Pronominal-Nebensatz
subjunct	=	Subjunkional-Nebensatz
V1	=	uneingeleiteter Nebensatz mit Verberststellung
V2	=	uneingeleiteter Nebensatz mit Verbzweitstellung
V3	=	uneingeleiteter Nebensatz mit Verbdrittstellung
Vend	=	uneingeleiteter Nebensatz mit Verbendstellung
V*	=	uneingeleiteter Nebensatz ohne Finitum

functional:

adv_sc	=	Adverbial(neben)satz (Bestimmung auf clause-st)
attr_sc	=	Attribut(neben)satz (Bezug auf clause-st)
obj_sc	=	Objekt(neben)satz
pred_sc	=	Prädikativ(neben)satz
subj_sc	=	Subjekt(neben)satz
attr: subj_sc	=	Subjekt(neben)satz mit Platzhalter im übergeordneten Satz
attr: obj_sc	=	Objekt(neben)satz mit Platzhalter im übergeordneten Satz
attr: adv_sc	=	Adverbial(neben)satz mit Platzhalter im übergeordneten Satz

s_grfunct = Die syntaktische Analyse der Nebensätze (*grammatical function of sub-clauses*) erfolgte mit den Tags der grfunct-Spur.

clause-st = Die **Satzarten** (*clause-status*) erhielten die folgenden Tags:

Hauptsätze:

decl	=	Deklarativ-/Aussagesatz
desi	=	Desiderativ-/Wunschsatz
imper	=	Imperativ-/Aufforderungssatz

- inter = Interrogativ-/Fragesatz
- excl = Exklamativ-/Ausrufesatz
- par: decl = parenthetisch eingeschobener Aussagesatz
- par: desi = parenthetisch eingeschobener Desiderativsatz

Bei den Nebensätzen entsprechen die Tags für Form und Funktion denen der grfunct-Spur. Bei Attributnebensätzen wurde hier das attribuierte Satzglied des übergeordneten Satzes eingetragen. Bei Relativ- und Adverbialsätzen wurden außerdem folgende Tags verwendet:

Relativsätze:

- rel = freier Relativsatz
- rel: ... = attribuierender Relativsatz

Adverbialnebensätze:

- advers = Adversativsatz
- caus = Kausalsatz
- compar = Komparativsatz
- conc = Konzessivsatz
- cond = Konditionalsatz
- consec = Konsekutivsatz
- fin = Finalsatz
- mod = Modalsatz
- temp = Temporalsatz
- satz1 = themeneinführender Nebensatz und Redekommentar
- satz2 = weiterführender Nebensatz

complex = Um die **Satzkomplexität** (*sentence complexity*) darzustellen, wurden die Arten der Sätze mit Großbuchstaben gekennzeichnet und durchgezählt. Die Ziffern der Nebensätze zeigen die Abhängigkeiten an:

- Hauptsätze (A): A_1, A_2 etc.
- Nebensätze 1. Ordnung (B): B_1.1 (= Nebensatz 1 [1. Ordnung] zu Hauptsatz 1)
B_2.3 (= Nebensatz 3 [1. Ordnung] zu Hauptsatz 2)
- Nebensätze 2. Ordnung (C): C_1.2.3 (= Nebensatz 3 [2. Ordnung] zu Nebensatz 2 [1. Ordnung] zu Hauptsatz 1)
- etc.

phras = Die einzelnen Kategorien des **Briefstils** (*phrasing*) sind mit den nachfolgenden Tags versehen, Themen des Briefhauptteils wurden fortlaufend nummeriert und kombiniert (**res1...**).

intit1 = Intitulatio in Anfangsstellung = Selbstnennung des Absenders

salut = salutatio (förmlicher Gruß)

recep = receptio (Empfangsbestätigung)

recap = recapulatio (Wiederholung/Zusammenfassung eines Briefes/ Ereignisses)

honor = honoratio (Ehrerweisung, Komplimentsfloskeln, Entschuldigungen)

narr = narratio (Mitteilung)

consil = consilio (Ratschlag)

consol = consolatio (Trost)

grat = gratificatio (konkreter Dank)

petit = petitio (Aufforderung)

explor = exploratio (Erkundigung)

excus = excusatio (Entschuldigung)

affec = affectatio (Bekundung von Zuneigung)

prex = prex (Verwünschung, Fluch)

prom = promissum (konkrete Versprechen)

sanct = sanctio (Drohung)

apprec = apprecatio (Segenswunsch)

hort = hortatio (Ermahnung)

resp = responsio (Bitte/Forderung zu antworten)

salus1 = salus (Gruß an eine dritte Person)

salus2 = salus (Gruß von einer dritten Person)

datum = Orts- und Zeitangabe in der Datumszeile

intit2 = Intitulatio in Endstellung = Selbstnennung des Absenders