

HIPKON

Historisches Predigtenkorpus zum Nachfeld

Sonderforschungsbereich 632, Teilprojekt B4:

Die Rolle der Informationsstruktur im Sprachwandel

Herausgeber:

Prof. Dr. Karin Donhauser, Humboldt-Universität zu Berlin
Dr. Eva Schlachter, Humboldt-Universität zu Berlin
Jun.-Prof. Dr. Marco Coniglio, Georg-August-Universität Göttingen

Annotatoren:

Jun.-Prof. Dr. Marco Coniglio
Anke Gehrlein
Carolin Odebrecht
Oxana Rasskazova
Dr. Eva Schlachter
Matthias Wirth

18.11.2015

Inhalt

1.	Motivation	2
2.	Korpusinhalt	3
3.	Erklärung des Tagsets	5
3.1	Primärquelle	5
3.2	Morphosyntaktische Annotation	6
3.2.1	<i>POS – Part of Speech</i>	6
3.2.2	<i>GF (Grammatische Funktion)</i>	10
3.2.3	<i>CAT – syntaktische Kategorie</i>	11
3.2.4	<i>SATZSTATUS</i>	12
3.2.5	<i>SATZKLAMMER</i>	12
3.2.6	<i>VERLINKUNG</i>	12
3.3	Annotation der Informations- und Diskursstruktur	13
3.3.1	<i>IS (Informationsstatus)</i>	13
3.3.2	<i>FOKUS</i>	15
3.3.3	<i>WIEDERAUFNAHME</i>	15
3.4	Metadaten	16
3.4.1	<i>Tokenspezifische Metadaten</i>	17
3.4.2	<i>Metadaten über Teilkorpora</i>	17
3.4.3	<i>Metadaten über das Gesamtkorpus</i>	17
4.	Abfrage und Langzeitspeicherung	17
4.1	<i>ANNIS</i>	17
4.2	<i>LAUDATIO</i>	18
	Quellenverzeichnis	19
	Literatur	21
	Tools	23

1. Motivation

Das Korpus HIPKON (**H**istorisches **P**redigten**k**orpus zum **N**achfeld) wurde zur Untersuchung der Nachfeldbesetzung in der Entwicklung des Deutschen erstellt. Zwar existieren schon verschiedene datenbasierte Untersuchungen zum Nachfeld, jedoch sind deren Ergebnisse kaum miteinander vergleichbar, da sich die zugrunde gelegten Daten hinsichtlich der Textsorte oder des Dialektgebiets unterscheiden bzw. die verwendete Terminologie verschiedene Phänomene erfasst. Das neue Korpus soll dementsprechend zum ersten Mal die Vergleichbarkeit der Daten über einen größeren Zeitraum hinweg ermöglichen.

Mit dem Begriff des Nachfelds wird hier die Position nach der rechten Satzklammer bezeichnet. Diese wird im Hauptsatz von den infiniten Verbteilen und Verbpartikeln, im Nebensatz vom infiniten Verb gebildet. Im Korpus wurden alle Hauptsätze erfasst, in denen das Nachfeld durch eine nicht-sententiale Konstituente besetzt ist. Die Positionierung von Nebensätzen im Nachfeld bleibt unberücksichtigt. Detailliertere begriffliche Unterscheidungen wie Ausklammerung, Rechtsversetzung, Nachtrag, Apposition u.Ä. (vgl. Altmann 1981) bleiben zunächst unberücksichtigt, jedoch werden vereinzelt vorkommende Appositionen als solche in der Annotation gekennzeichnet.

Laut Dudengrammatik (Wermke et al. 2009:886f) werden im Gegenwartsdeutschen besonders oft Präpositionalphrasen (1) und Vergleichsphrasen (2) ausgeklammert:

- (1) *Erstaunt wollte sie sich umsehen [in der Gegend].*
- (2) *Anna hat das Ergebnis schneller ausgerechnet [als alle anderen].*

Im Gegensatz dazu können aber zu Beginn der deutschen Überlieferung, im Ahd., nicht nur Präpositionalphrasen extraponiert werden, sondern auch Pronomen (3), Objekte (4) und Prädikativkonstruktionen (5).

- (3) „*dhazs ih fora sinem anthlutte hneige [imu dheodun], [...]*“
dass ich vor seinem Antlitz neige ihm die Völker
lat. ut subiciam ante faciem eius gentes (Is III.2, 153, zit. n. Schlachter 2009:234)
- (4) *thaz in mir habet [sibba]*
dass in mir habt (ihr) Frieden
lat. ut In me pacem habeatis (T 290,8, zit. n. Petrova 2009:266)

(5) *soso thie lihhazara sint [gitruobte]*

wie die Heuchler sind betrübt

lat. *sicut hypocritæ tristes* (T 68,23, zit. n. Petrova 2009:267)

Es stellt sich die Frage, wie und warum die Möglichkeiten der Nachfeldbesetzung im Laufe der Entwicklung des Deutschen immer restriktiver wurden. Erste Hypothesen bezüglich der einflussnehmenden Faktoren spiegeln sich in den verschiedenen Annotationsebenen wider. Erfasst werden neben den syntaktischen Funktionen vor allem Eigenschaften, die die Diskursebene betreffen: solche der Informationsstruktur, des Informationsstatus und der Relevanz für den Folgediskurs.

2. Korpusinhalt

Für das Korpus wurden Predigttexte aus dem Oberdeutschen (bairisch-alemannisch) ausgewählt, da nur diese Textsorte und dieses Dialektgebiet einigermaßen konstant über den gesamten Zeitraum vertreten sind. Jedes Teilkorpus hat einen Umfang von etwa 12.000 bis max. 17.000 Wortformen,¹ der zeitliche Abstand zwischen den Texten beträgt etwa 100 Jahre, so dass jedes Jahrhundert durch ein Teilkorpus repräsentiert wird. Lediglich für das 15. Jahrhundert konnte kein passender Text gefunden werden. Möglicherweise war die Überlieferung in dieser Zeit weder aus historischen noch aus theologischen Gründen von Interesse.

In Tabelle 1 wird eine Übersicht über die verwendeten Predigttexte, ihre Datierung und den Umfang in Wortformen gegeben.

¹ Für die Erstellung des Korpus wurde prinzipiell die Erfassung von 12.000 Wortformen anvisiert. Aus unabhängigen Gründen wurden für das 16. und 18. Jahrhundert mehr Wortformen erfasst. Dieses sollte man je nach der Art der Suchabfrage berücksichtigen.

Tabelle 1: Übersicht über die verwendeten Quellen

Text	Datierung	Wortformen
MHD		
Speculum Ecclesiae	1175-1200	12.300
Schwarzwälder Predigten	vor 1300	12.100
Nikolaus von Straßburg	1300-1350	12.350
FNHD		
Kaspar Adler (Evang. Prediger, Almosen, Von der Schul)	1523, 1533, 1523	17.614 (5.617, 4.961, 7.036)
Glückwünschungspredigt, Wahlpredigt	1620	12.708
NHD		
Heuschrecken, Ein Wort der Ermahnung	1751, 1747	14.914 (7.211, 7.211)

Die eigentliche Annotation wurde in EXMARaLDA (Extensible Markup Language for Discourse Annotation, Schmidt et al. 2009) vorgenommen, über ANNIS (Annotation of Information Structure, Zeldes et al. 2009; vgl. 4.1) ist eine Abfrage und Nutzung möglich.

Die mhd. Texte, die schon als digitale Textdateien vorlagen, wurden direkt nach EXMARaLDA importiert. Die fnhd. und nhd. Texte wurden dahingegen manuell aus digitalisierten Drucken transkribiert.²

Annotiert wurden nur die Belege, die für die Forschungsfrage relevant sind, also Hauptsätze mit einem komplexen Verbgefüge aus Auxiliar- oder Modalverb und einem infiniten Vollverb, dem noch eine oder mehrere

² Zur Beschreibung der in den Handschriften verwendeten Grapheme gibt die *Mittelhochdeutsche Grammatik* von Hermann Paul in *Kapitel 7 Schrift, Schreibung, Aussprache* (2007:23-28) eine Übersicht.

In den manuell transkribierten Texten wurde das tironische *et* nicht aufgelöst, sondern – wegen ihrer typographischen Ähnlichkeit – als rundes *r* wiedergegeben.

Konstituenten folgen. Auch die komplexen Verbgefüge mit verbalen Partikeln wurden aufgenommen.

Das Korpus ist ein *multi-layer* Korpus, d.h. jede Art der Information zur Morphosyntax oder Informationsstruktur ist einer eigenen Ebene zugeordnet. Dies ermöglicht eine spätere Verwendung für weitere Untersuchungen, da innerhalb des Annotationstools EXMARaLDA neue Ebenen hinzugefügt oder schon bestehende gelöscht oder modifiziert werden können.

3. Erklärung des Tagsets

Die Übersicht über das verwendete Tagset ist in vier Gruppen gegliedert: Die erste beinhaltet Annotationen zur Primärquelle (s. 3.1), die zweite Informationen zur Morphosyntax der aufgenommenen Belege. Hier wurde z.B. die Wortart, grammatische Funktion oder syntaktische Kategorie annotiert (s. 3.2). Die dritte Gruppe umfasst die Annotation der Informations- und Diskursstruktur (s. 3.3), die vierte enthält Metainformationen zu den Belegen wie Stellenangaben in der dazugehörigen Edition oder der Hand- bzw. Druckschrift (s. 3.4).

3.1 Primärquelle

TXT Diese Ebene beinhaltet den transkribierten Text, dabei wurden die Zeichen mit Diakritika und Überschreibungen in ihrer Anordnung möglichst genau wiedergegeben.

CLEAN Die Ebene CLEAN ist keine Normalisierungsebene im herkömmlichen Sinne. Hier werden Wörter mit Zeilenumbruch, die auf der Ebene TXT mit „=“ oder „-“ getrennt sind wieder zusammengefügt, um eine Suche nach diesen Wörtern zu erleichtern. Beispielsweise werden hier getrennt geschriebene Komposita zusammengefügt. Auf der anderen Seite werden pro- oder enklitisch angefügte Wörter wie z.B. Personalpronomen oder die Negationspartikel *en-/ne-* getrennt annotiert. Die Bereinigung betrifft nur die Belege und wird nicht für den ganzen Text angewendet.

SPRACHE Kennzeichnung des lateinischen Sprachmaterials mit „Latein“.

3.2 Morphosyntaktische Annotation

3.2.1 POS – *Part of Speech*

Die Annotation der Wortarten (*Part of Speech*, POS) folgt dem Tagset DDDTS-ahd (Deutsch diachron digital, Dipper et al. 2013),³ das auf der Basis des STTS (Stuttgart-Tübingen-Tagset, Schiller et al. 1999) mit einer Erweiterung für die Annotation historischer Texte entwickelt wurde.

³ Vgl. auch <https://referenzkorporaltddeutsch.wordpress.com/tagging-der-wortarten/>.

\$_	handschriftliche Interpunktionszeichen	APPR	Präposition, Zirkumposition links
ADJ	Adjektiv, attributiv, vorangestellt oder elliptisch	APPRART	Präposition mit Artikel
ADJD	Adjektiv, prädikativ oder adverbial	APZR	Zirkumposition rechts
ADJE	Adjektiv, attributiv, vorangestellt oder elliptisch, Teil eines Eigennamens	CARD	Kardinalzahl
ADJN	Adjektiv, attributiv, nachgestellt	CARDN	Kardinalzahl, attribulierend, nachgestellt
ADJNE	Adjektiv, attributiv, nachgestellt, Teil eines Eigennamens	CARDS	Kardinalzahl, substituierend
ADJO	Adjektiv, ordinal, attributiv, vorangestellt oder elliptisch	DD	Determinativ, definit/demonstrativ, vorangestellt oder elliptisch
ADJOE	Adjektiv, ordinal, attributiv, vorangestellt oder elliptisch, Teil eines Eigennamens	DDA	Determinativ, definit, artikelartig
ADJON	Adjektiv, ordinal, attributiv, nachgestellt	DDN	Determinativ, definit/demonstrativ, nachgestellt
ADJOS	Adjektiv, ordinal, substantiviert	DDREL	Determinativ, definit/demonstrativ, relativisch
ADJS	Adjektiv, substantiviert	DDS	Determinativ, definit/ demonstrativ, substituierend
ADV	Adverb	DDSREL	Determinativ, definit/ demonstrativ, substituierend, relativisch
ADVM	Adverb, multiplikativ (<i>zweimal</i>)	DI	Determinativ, indefinit, vorangestellt oder elliptisch
ADVNEG	Adverb, negativ	DIA	Determinativ, indefinit, artikelartig
ADVREL	Relativadverb		
APPO	Postposition		

DIN	Determinativ, indefinit, nachgestellt	DWREL	Determinativ, interrogativ, relativ
DINEG	Determinativ, indefinit, negativ, vorangestellt oder elliptisch	DWS	Determinativ, interrogativ, substituierend
DINEGN	Determinativ, indefinit, negativ, nachgestellt	DWSREL	Determinativ, interrogativ, substituierend, relativ
DINEGS	Determinativ, indefinit, negativ, substituierend	FM	fremdsprachliches Material
DIS	Determinativ, indefinit, substituierend	ITJ	Interjektion
DPOS	Determinativ, possessiv, vorangestellt oder elliptisch	KO	Konjunktion, unter- oder beordnend
DPOSD	Determinativ, possessiv, prädikativ oder adverbial	KOKOM	Vergleichspartikel (nur mit einer NP oder anderem Satzglied kombinier- bar)
DPOSN	Determinativ, possessiv, nachgestellt	KON	Konjunktion, neben- ordnend
DPOSS	Determinativ, possessiv, substituierend	KOUI	Konjunktion, unterordnend mit „zu“ und Infinitiv
DW	Determinativ, interrogativ	KOUS	Konjunktion, unter- ordnend mit Satz
DWG	Determinativ, (interrogativ,) generalisierend, attributiv	NA	Nomen, appellativ
DWGREL	Determinativ, (interrogativ,) generalisierend, relativ	NE	Nomen, Eigenname
		PAV	Pronominaladverb
		PAVREL	Pronominaladverb, relativisch
		PI	Pronomen, indefinit
		PINEG	Pronomen, indefinit, negativ

PPER	Pronomen, personal, irreflexiv	PWGREL	Pronomen, (interrogativ), generalisierend, relativ
PRF	Pronomen, personal, reflexiv	PWREL	Pronomen, interrogativ, relativ
PTKA	Partikel bei Adjektiv oder Adverb	VAFIN	Finites Verb, auxiliar
PTKANT	Antwortpartikel	VAIMP	Imperativ, auxiliar
PTKNEG	Negationspartikel	VAINF	Infinitiv, auxiliar
PTKREL	Relativpartikel	VAINFS	Infinitiv, auxiliar, substantiviert oder flektiert
PTKVZ	abgetrennter Verbzusatz	VAPP	Partizip Präteritum, auxiliar, im Verbalkomplex
PTKZU	<i>zu</i> vor Infinitiv	VAPS	Partizip Präsens, auxiliar, im Verbalkomplex
PW	Pronomen, interrogativ (<i>wer, was</i>)	VMFIN	finites Verb, modal
PWAV	Pronomen, interrogativ, adverbial (<i>wo, wann</i>)	VMIMP	Imperativ, modal
PWAVREL	Pronomen, interrogativ, adverbial, relativ	VMINF	Infinitiv, modal
PWG	Pronomen, (interrogativ), generalisierend (<i>swer</i>)	VMINFS	Infinitiv, modal, substantiviert oder flektiert
PWGAV	Adverb, (interrogativ), generalisierend (<i>swanne</i>)	VMPP	Partizip Präteritum, modal, im Verbalkomplex
PWGAVREL	Pronomen, (interrogativ), generalisierend, relativ	VMPS	Partizip Präsens, modal, im Verbalkomplex
		VVFIN	finites Verb, Vollverb
		VVIMP	Imperativ, Vollverb

VVINNF	Infinitiv, Vollverb	VVPS	Partizip Präsens, Vollverb, im Verbal-komplex
VVINFS	Infinitiv, Vollverb, substantiviert oder flektiert	VVPSA	Partizip Präsens, Vollverb, attribuierend, vorangestellt oder elliptisch
VVPP	Partizip Präteritum, Vollverb, im Verbal-komplex	VVPSD	Partizip Präsens, Vollverb, prädikativ oder adverbial
VVPPA	Partizip Präteritum, Vollverb, attribuierend, vorangestellt oder elliptisch	VVPSN	Partizip Präsens, Vollverb, attribuierend, nachgestellt
VVPPD	Partizip Präteritum, Vollverb, prädikativ oder adverbial	VVPS	Partizip Präsens, Vollverb, substantiviert
VVPPN	Partizip Präteritum, Vollverb, attribuierend, nachgestellt		

3.2.2 GF (Grammatische Funktion)

ADVB	alle Adverbiale jeder Spezifikation (modal, lokal, ...), keine weitere Spezifikation
APPOS	Apposition
ATTR	Attribut
DO	direktes Objekt, d.h. Akkusativobjekt
EXPL	Expletiva bzw. Proformen, die als „Platzhalter“ fungieren, wie in „ <i>Es ist ein Zeichen am Himmel erschienen.</i> “
GERUND	infinite, flektierte Verbform: „ <i>vn[d] h(at) fi machot gênde</i> “ (SchwPr 147,20)
INF	infinite, unflektierte Verbform
IO-GEN	indirektes Objekt im Genitiv
IO-DAT	indirektes Objekt im Dativ
PP-DIR	PP mit direktonaler Ergänzung

PP-SIT	PP mit situativer Ergänzung, e.g. <i>wohnt in Berlin</i>
PP-OBJ	PP-Objekt
PP-ADV	PP-Adverbial
PP-AGENS	Agens einer Passiv-Konstruktion
PRD	Prädikativ
SUBJ	Subjekt
VFIN	finites Verb
VGL	einen Vergleich einleitende Wörter wie <i>so, also, wie</i>
VOVF	Konstituenten im Vor-Vorfeld: Linkversetzungen, <i>Hanging Topics, nominativus pendens</i> , vokativische Ausdrücke <i>allíu mínú lid[er] i[n] mine[m] libe. díu fint mier reht zerzoge[n] vn[d] zerde[n]net úz aín and[er].</i> (SchwPr 157,17)
VZ_PTK	Verbzusatz von trennbaren Verben/Partikelverben
VZ_RFL	Verbzusatz: Reflexivpronomen, nur bei echten ⁴ Reflexivverben, z.B. <i>sich wundern</i>

3.2.3 CAT – syntaktische Kategorie

Auf dieser Ebene werden die syntaktischen Phrasen annotiert. Attribute werden nicht gesondert gekennzeichnet, da sie Teil der ausgezeichneten Phrase sind. Jedoch werden Relativ- und Adverbialsätze dann aufgenommen und als CP annotiert, wenn sie für die Wiederaufnahme relevant sind (siehe Abschnitt 3.3.3).

AP	Adjektivphrase
ADVP	Adverbialphrase
CP	Relativsätze und Adverbialsätze
NP	Nominalphrase

⁴ Echt reflexive Verben liegen dann vor, wenn Reflexivpronomen nicht weglassbar ist und der Austausch des Reflexivpronomens gegen ein anderes Pronomen oder Nominalphrase ebenfalls nicht möglich ist. Bei anderen Verben wie *waschen* (*sich waschen/ihn waschen*) wird die Ergänzung mit DO annotiert und auf der Ebene CAT wird dieses Element mit VP ausgezeichnet (vgl. Helbig & Buscha 2001:55f).

PP Präpositionalphrase

VP Verbalphrase

3.2.4 SATZSTATUS

Da im Korpus nur Hauptsätze annotiert wurden, gibt es lediglich vier spezifische Tags. Eingeschobene oder ausgeklammerte Nebensätze (sowohl Adverbial- als auch Relativsätze) wurden nicht gesondert ausgezeichnet, sondern mit in die Spanne aufgenommen und auf der Ebene CAT mit CP annotiert.

MAINDECL Hauptsatz, Deklarativsatz

MAINDECL-ELLI elliptischer Hauptsatz, Subjekt oder Verb fehlen

vn[d] wil fi zu[i]chtego[n] mit grôze[n] sclege[n] (SchwPr 17,28)

MAININTERR Hauptsatz, Interrogativsatz

MAINIMPER Hauptsatz, Imperativsatz

3.2.5 SATZKLAMMER

LI Kennzeichnung für die Position der linken Satzklammer

RE Kennzeichnung für die Position der rechten Satzklammer

3.2.6 VERLINKUNG

Auf dieser Ebene wird das syntaktische Verhältnis der postverbalen Konstituente zum Mittelfeld beschrieben. Hier wird erfasst, ob es sich bei der Konstituente um eine Koordination oder ein Attribut handelt. Die Konstituente wird als Spanne annotiert. Erscheinen mehrere Konstituenten im Nachfeld, werden diese durchnummeriert.

UNLINK keine Verlinkung zu einer Konstituente im Mittelfeld gegeben

Ir schult bittin umbe die heiligin [christ]inheit (Speccl 83,07)

LINK_ASSYM Verlinkung zu einer Konstituente im Mittelfeld gegeben, unkoordiniert

wir schulin got mere gehorfamin denni den mennifchin
(Speccl 78,30)

LINK_SYM Verlinkung zu einer Konstituente im Mittelfeld gegeben, koordiniert

*daru[m]b mieffen follich euangelisch prediger ketzer
gescholte[n] werdn[n] / vn[d] aufru[e]risch bu[e]ben /
(EvangP 05,13-15)*

LINK_APPOS wird vergeben, wenn eine Verlinkung zu einer Konstituente des Mittelfelds gegeben ist und die Phrase im Nachfeld eine Apposition zur Konstituente im Mittelfeld darstellt

jam mu[o]zzin wir ze tâl fîgen zuo dem tôde (Speccl 40,04)

3.3 Annotation der Informations- und Diskursstruktur

3.3.1 IS (Informationsstatus)

Auf dieser Ebene wurde der Informationsstatus der ausgeklammerten Konstituente(n) annotiert. Dabei wurden nur referierende Ausdrücke, keine idiomatischen Wendungen wie z.B. *bei Tag und Nacht* aufgenommen. Das Tagset wurde im Wesentlichen Götze et al. (2007) entnommen, wobei v.a. die Kategorie der zugänglichen Referenten (*accessible*, ACC) erweitert und umgebaut wurde. Ergebnisse und Überlegungen von Baumann und Riester (2012), Haug et al. (2014) und Taylor & Pintzuk (2014) wurden berücksichtigt. Ein besonderes Problemfeld stellten dabei abstrakte Nomina dar, deren Informationsstatus nicht immer zweifelsfrei bestimmt werden konnte. Daher wurde für diese Gruppe eine eigene Unterkategorie unter den zugänglichen Referenten geschaffen (ACC-ABSTR). Eine weitere Unterspezifizierung betrifft nur die mit ACC annotierten Elemente, bei denen eine Unterscheidung getroffen wird, ob sie bereits vorerwähnt sind (USED) oder ganz neu (UNUSED) in der Diskurseinheit auftreten.

In Relativsätzen wurden nicht die Relativpronomen auf ihren Informationsstatus hin annotiert, sondern nur das Bezugsnomen. Komplexe Nachfelder wurden wie folgt annotiert. In den Fällen, in denen im Nachfeld Genitiv- oder Präpositionalattribute (siehe (6)) sowie Possessivpronomina, (siehe (7)) enthalten sind, werden alle diese Bestandteile einzeln annotiert. Adjektivische Attribute werden in ihrer Gesamtheit mit dem Bezugsnomen annotiert, siehe (8).

(6) die Stimme von Gott

(7) seine Stimme

(8) schöne Stimmen

NEW	neu
GIV-ACTIVE	direkt im vorherigen Satz gegeben (auch Nebensatz zählt)
GIV-INACTIVE	wenn an einer vorherigen Stelle im Kapitel (nicht direkt im Vorsatz) gegeben
ACC-SIT	verfügbar durch den situativen Kontext, auch Personalpronomen der 1. und 2. Person
ACC-GEN	diese Kategorie ist den generischen Ausdrücken vorbehalten (<i>die Löwin ist gefährlich, wenn sie Junge hat</i>)
ACC-INF	Referent ist verfügbar über eine Teil-mengenbeziehung wie <i>der Garten</i> und <i>sein Eingang</i>
ACC-AGGR	Referent ist verfügbar durch die Zusammenführung von bereits explizit genannten Diskursreferenten (<i>Hund, Katze, Maus – die Tiere</i>)
ACC-UNI-UNUSED	verfügbare Unika (<i>Sonne, Mond, Gott</i>) und Orts- und Personennamen, die nicht vorerwähnt sind
ACC-UNI-USED	verfügbare Unika (<i>Sonne, Mond, Gott</i>) und Orts- und Personennamen, die schon vorerwähnt sind
ACC-ABSTR-UNUSED	Abstrakta, die noch nicht vorerwähnt sind
ACC-ABSTR-USED	Abstrakta, die vorerwähnt sind
[...]-USED-1	unmittelbar im vorherigen Satz erwähnt, nur für mit ACC ausgezeichnete Referenten
[...]-USED-X	vorherwähnt, aber nicht unmittelbar im vorherigen Satz, nur für mit ACC ausgezeichnete Referenten
NA	nicht anwendbar; für Elemente, bei denen keine Aussage hinsichtlich der Gegebenheit gemacht werden kann

3.3.2 FOKUS

Als Fokus wird der Bereich annotiert, der die relevante Information trägt. Zwar fällt dieser häufig mit neuer Information zusammen, dies muss aber nicht notwendigerweise der Fall sein (vgl. Molnár 1993:166f). Die relevante Information ergibt sich als das Ergebnis einer Alternativenauswahl (vgl. u.a. Rooth 1985, 1992, Krifka 2007:18ff), die letztlich durch den Kontext determiniert ist.

Generell werden nur die Fokuselemente im Nachfeld annotiert. Nur in Ausnahmefällen, wenn sich der Fokusbereich auch auf Konstituenten vor dem Infinitum erstreckt, wird auch dieser Teil als Fokus ausgezeichnet. Dieses Vorgehen ist mit der prinzipiellen Entscheidung begründet, in Zweifelsfällen, in denen sowohl enger als auch weiter Fokus möglich ist, immer für den weiten Fokus zu optieren. Ist im Nachfeld keine Konstituente Teil des Fokusbereichs, so werden die entsprechenden Konstituenten als Hintergrundinformation gekennzeichnet.

CF *contrastive focus* wird nur dann annotiert, wenn ein starker Kontrast bzw. eine kontrastive Auswahlmöglichkeit gegeben ist.

NF *new focus* wird annotiert, wenn die Konstituente im Fokus steht und (zum Teil) neue Informationen enthält

BGR *background information* ist eine Art „Restkategorie“, die „Nicht-Fokus“ bezeichnet, alle Elemente des Nachfelds, die nicht im Fokus stehen werden so bezeichnet

CF>, NF>, BGR> wird annotiert, wenn es mehrere Fokus- bzw. Hintergrundelemente im Nachfeld gibt

3.3.3 WIEDERAUFNAHME

Mit der Ebene WIEDERAUFNAHME soll die These von Vinckel-Roisin (2011) überprüft werden, nach der ein Motiv der Positionierung von Konstituenten im Nachfeld deren Wiederaufnahme im Folgediskurs ist. Auf dieser Ebene werden nur dann Annotationen vorgenommen, wenn eine oder mehrere ausgeklammerte Konstituente/n auch tatsächlich im weiteren Diskurs erneut erscheinen. Das Antezedens wird als KOREF1, die wiederaufnehmende Konstituente (Anapher) als KOREF2 bezeichnet. Dabei soll die aufgenommene Anapher innerhalb der nächsten drei Sätze erscheinen. Als Satz gilt eine

Einheit mit flektiertem Verb, Einheiten von unterbrechender direkter Rede werden ignoriert. Wird das Antezedens mehrere Male hintereinander wieder aufgenommen, wird nur die erste vorkommende Anapher annotiert. Eine Ausnahme bilden dazwischen tretende Relativpronomen, die als spezielle Form der Anapher (KOREF2-REL) aufgenommen werden. Dies ermöglicht die Unterscheidung zwischen Nachstellung einer Konstituente aufgrund eines folgenden Relativsatzes und Nachstellungen, die durch andere Arten der Wiederaufnahme bedingt sind. Die Annotation von Anaphern erfolgt nur auf den Ebenen: POS, CAT, GF, EDITION und HANDSCHRIFT und nur für das betreffende Satzglied.

- KOREF1** bezeichnet das Antezedens
- KOREF2** bezeichnet die Anapher
- KOREF2/1** bezeichnet Konstituenten, die sowohl Antezedens als auch Anapher sind
- KOREF2-SATZ** bezeichnet einen ganzen Satz, der als Anapher fungiert, z.B. mit *do*, *dc* eingeleitete Sätze
- daz felbe zeichin wart eroffint. wilen. do die iuden uzzer egypto vu[o]rin. (Speccl 13,19-20)*
- KOREF2-NULL** bezeichnet das erste Element in einem *topic drop*-Satz, die eigentliche Anapher im Satz fehlt
- nv ift chōmin unfer aller arzat. Ø hat die fiechin geheilet. (Speccl 10,04)*
- KOREF1-REL** bezeichnet das Bezugsnomen eines Relativsatzes
- KOREF2-REL** bezeichnet das Relativpronomen. Da Relativsätze immer koreferent sind, sind sie so leichter auffindbar und verzerren somit auch nicht die Ergebnisse

3.4 Metadaten

Metadaten liefern strukturierte Auskunft über das Gesamtkorpus, die einzelnen Teilkorpora sowie tokenspezifische Information, was den Umgang mit dem Korpus vereinfachen soll. Um dies zu ermöglichen, wurden die Metadaten strukturiert und wie folgt gespeichert.

3.4.1 Tokenspezifische Metadaten

Tokenspezifische Metadaten wurden als Spannenannotation in EXMARaLDA dargestellt. Die Information bezieht sich dabei nur auf den ausgewählten Token:

UEBERLIEFERUNGSTRAEGER

Auf dieser Ebene wird mit einer Spannenannotation der transkribierte Text TXT den entsprechenden Zeilen und Seiten des Überlieferungsträgers (Handschrift bzw. Druckschrift) zugeordnet.

EDITION

Diese Ebene ist nur für die Teilkorpora verfügbar, die ediert wurden: Speculum Ecclesiae, Schwarzwälder Predigten und Nikolaus von Straßburg. Mittels einer Spannenannotation wird der transkribierte Text TXT den entsprechenden Seiten und Zeilen der Edition zugewiesen.

3.4.2 Metadaten über Teilkorpora

Metadaten über Teilkorpora liefern Information über im diesem Teil verwendeten Predigten, deren Stellenangaben in Edition und Standort des Überlieferungsträgers. Diese Metadaten sind strukturiert in EXMARaLDA gespeichert und sind über ANNIS abrufbar.

3.4.3 Metadaten über das Gesamtkorpus

Metadaten über das Gesamtkorpus umfassen Information über das Projekt, die Motivation für die Korpuserstellung, Korpusdesign und Annotatoren. Als kurze Zusammenfassung ist diese Information über ANNIS abrufbar. Die längere Version ist als externer Text (Korpusdokumentation) über LAUDATIO (vgl. 4.2) verfügbar.

4. Abfrage und Langzeitspeicherung

4.1 ANNIS

Die Visualisierung des Korpus erfolgt über ANNIS (*ANNotation of Information Structure*), *open-source*, plattformabhängig und *browser*-basiert. Zur Dokumentation und zur Suchabfrage (*ANNIS query language*, AQL) siehe die aktuellen Anleitungen und Hinweise auf der ANNIS-Homepage.

<http://www.sfb632.uni-potsdam.de/annis/>

4.2 LAUDATIO

Um das Korpus dem interessierten (Fach-)Publikum zugänglich zu machen und um eine Langzeitarchivierung zu ermöglichen, wird das Korpus zusätzlich über LAUDATIO (Long-term Access and Usage of Deeply Annotated InformaTION, Odebrecht et al. 2015), ein *Open Access* Forschungsdaten-Repository für historische Korpora zugänglich gemacht.

<http://www.laudatio-repository.org/repository/>

Quellenverzeichnis

[Almosen] Kaspar Adler, Von Almosen geben

Von Almosen geben. Ein Sermon M. Caspers Adler... mit D. Mart. Luthers Vorrede.

Druckschrift: Wittenberg 1533, Bayrische Staatsbibliothek

<http://daten.digital-sammlungen.de/0003/bsb00030307/images/index.html?fip=193.174.98.30&id=00030307&seite=1>

[Ermahnung] Ein Wort der Ermahnung

Ein Wort der Ermahnung von der Unvergänglichkeit des Göttlichen Worts, Und das Gebet, wie dieses an dem ersten Tage Des Eintausend, Siebenhundert und Sieben und vierzigsten Jahres in allen sechs Evangelischen Kirchen Der Des Heil. Römischen Reichs freye Stadt Augspurg nach allen Vor- und Nachmittags-Predigten Gott zu Ehren gesprochen worden, Beydes aber noch weiter in denen Häusern zur Erbauung gebraucht werden solle

Druckschrift: Augsburg 1747, Universitäts- und Landesbibliothek Sachsen-Anhalt

<http://vd18.de/de-ulbsa-vd18/content/titleinfo/13756074>

[EvangP] Kaspar Adler, Evang. Prediger

Ain Sermon, darin zu allen Evangelischen Prediger ain frölichen tröstliche ermanung ist, das sy das allergröst hailtum das lebendig wort gotes fraidig und köck den dürstigen seelen fürlegen und sich von kainer creaturen schrecken lasen: vast nutzlich auß den propheten und apostlen gezogen

Druckschrift: Augsburg 1523, Bayrische Staatsbibliothek

<http://www.mdz-nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10161024-0>

[GlückwP] Glückwünschungspredigt

Vota Sacra Pro Ampliss: Reipub: Augusta: Senatu. Das ist/ Eine Christliche Glückwünschungspredigt/ auß dem 20. Psalmen Davids

Druckschrift: Augsburg 1620, Universitäts- und Landesbibliothek Sachsen-Anhalt

<http://nbn-resolving.de/urn%3Anbn%3Ade%3Agbv%3A3%3A1-18654>

[Heuschrecken]

Frag: Was Bedeuten die Heuschrecken? Eccho: Schröcken! Das ist: Moralische Kirchweyh-Predig Eingericht Auf dermahlig-gefährliche Welt- und Zeit-Umständ

Druckschrift: Salzburg / München 1751, Universitäts- und Landesbibliothek

<http://vd18.de/de-ulbsa-vd18/content/titleinfo/14169323>

[NikP] Nikolaus von Straßburg, Predigten

Deutsche Mystiker des 14. Jahrhunderts. Hg. v. Franz Pfeiffer, Bd. 1: Hermann von Fritzlar, Nikolaus von Straßburg, David von Augsburg, Leipzig 1845 (Nachdruck Aalen 1962), 261-305.

Handschrift: St. Florian, Stiftsbibliothek Cod. XI 284, kl. 4 , fol. 37ra1-85rb03

[Schul] Kaspar Adler, Von der Schul

Ain Sermon von der schul Christi: darinn ain yetlicher leeriunnger lernet, wa es im fälet an dem weg der sälligkait und wie wir noch so gar nit evangelisch seind

Druckschrift: Augsburg 1523, Bayrische Staatsbibliothek

<http://www.mdz-nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10161022-0>

[SchwPr] Schwarzwälder Predigten

Deutsche Predigten des XIII. Jahrhunderts. Zum erstenmal hrsg. von Franz Karl Grieshaber, 2. Bände, Stuttgart 1844 und 1846 (Nachdr. Hildesheim / New York 1978).

Handschrift: Freiburg, Universitätsbibliothek Hs. 460

[Speccl] Speculum Ecclesiae

Speculum ecclesiae. Eine frühmittelhochdeutsche Predigtsammlung (Cgm 39). Mit sprachlicher Einleitung neu hg. v. Gert Mellbourn (Lunder germanistische Forschungen 12). Lund / Kopenhagen 1944.

Handschrift: München, Staatsbibliothek Cgm 39

<http://daten.digitale-sammlungen.de/~db/0003/bsb00035336/images/>

[Wahlpredigt]

Wahlpredig/ Auß dem anderen Buch der Chronic. am 19. Cap. Von der schönen und Löblichen Regimentsbestellung/ deß frommen eyferigen/ und Gottseeligen Königs Josaphat/ sampt seiner getrewen/ und zumahl auch ernstlichen vermahnung an alle seine Richter und Amptleut

Druckschrift: Augsburg 1620, Universitäts- und Landesbibliothek Sachsen-Anhalt

<http://nbn-resolving.de/urn%3Anbn%3Ade%3Agbv%3A3%3A1-15803>

Literatur

Altmann, Hans

1981 *Formen der „Herausstellung“ im Deutschen. Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen.* Tübingen: Niemeyer.

Baumann, Stefan und Arndt Riester

2012 Referential and lexical givenness. Semantic, prosodic and cognitive aspects. In: *Prosody and Meaning*, Gorka Elodieta und Pilar Prieto (Hgg.), 119-162. Berlin: Mouton de Gruyter.

Götze, Michael, Thomas Weskott, Cornelia Endriss, Ines Fiedler, Stefan Hinterwimmer, Svetlana Petrova, Anne Schwarz, Stavros Skopeteas und Ruben Stoel

2007 Information Structure. In: *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics, and Information Structure*, Stefanie Dipper, Michael Götze und Stavros Skopeteas (Hgg.), 147-187. Potsdam: Universitätsverlag. [Interdisciplinary Studies on Information Structure 7.]

Haug, Dag T., Hanne Martine Eckhoff und Eirik Welo

- 2014 The theoretical foundations of givenness annotation. In: *Information Structure and Syntactic Change in Germanic and Romance Languages*, Kristin Bech und Kristine Gunn Eide (Hgg.), 17-52. Amsterdam/Philadelphia: Benjamins. [Linguistik Aktuell 213]

Helbig, Gerhard und Joachim Buscha

- 2001 *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. Berlin: Langenscheidt.

Krifka, Manfred

- 2007 Basic Notions of Information Structure. In: *Interdisciplinary Studies on Information Structure* 6, Caroline Féry, Gisbert Fanselow und Manfred Krifka (Hgg.), 13-55. Potsdam: Universitätsverlag Potsdam.

Molnár, Valéria

- 1993 Zu Pragmatik und Grammatik des TOPIK-Begriffes. In: *Wortstellung und Informationsstruktur*. Reis, Marga (Hg.), 155-202. Tübingen: Max Niemeyer Verlag.

Odebrecht, Caroline, Thomas Krause und Anke Lüdeling

- 2015 Austausch von historischen Texten verschiedener Sprachen über das LAUDATIO-Repository. In 37. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, DGfS-CL Poster Session. Leipzig. [\[online\]](#)

Paul, Hermann

- 2007 *Mittelhochdeutsche Grammatik*. 25. überarbeitete Auflage. Tübingen: Max Niemeyer Verlag.

Petrova, Svetlana

- 2009 Information structure and word order variation in the Old High German Tatian. In: *Information Structure and Language Change: New Approaches to Word Order Variation in Germanic*, Roland Hinterhölzl und Svetlana Petrova (Hgg.), 251-279. Berlin: Mouton de Gruyter.

Rooth, Mats

- 1985 Association with Focus. Diss., Amherst: University of Massachusetts.

Rooth, Mats

- 1992 A theory of focus interpretation. In: *Natural Language Semantics* 1, 75-116.

Schlachter, Eva

- 2009 Word order variation and information structure in Old High German: an analysis of subordinate clauses in Isidor. In: *Information Structure and Language Change: New Approaches to Word Order Variation in Germanic*, Roland Hinterhölzl und Svetlana Petrova (Hgg.), 223-250. Berlin: Mouton de Gruyter.

Taylor, Ann und Susan Pintzuk

- 2014 Testing the theory. Information structure in Old English. In: *Information Structure and Syntactic Change in Germanic and Romance Languages*, Kristin Bech und Kristine Gunn Eide (Hgg.), 53-78. Amsterdam/Philadelphia: Benjamins. [Linguistik Aktuell 213]

Vinckel-Roisin, Hélène

- 2011 Wortstellungsvariation und Salienz von Diskursreferenten: Die Besetzung des Nachfeldes in deutschen Presstexten als kohärenzstiftendes Mittel. In: *Zeitschrift für germanistische Linguistik* 39 (3), 378-405.

Wermke, Matthias et al. (Hgg.)

- 2009 *Duden – Die Grammatik. Der Duden in zwölf Bänden. Band 4. 8., überarbeitete Auflage.* Mannheim / Zürich: Dudenverlag.

Tools

Annotation and metadata: EXMARaLDA

Schmidt, Thomas und Kai Wörner

- 2009 EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. In: *Pragmatics* 19 (4), 565–582. [\[online\]](#)

Part-of-speech tagset: DDDTS-ahd, STTS

Dipper, Stefanie, Karin Donhauser, Thomas Klein, Sonja Linde, Stefan Müller, und Klaus-Peter Wegera

- 2013 HiTS: ein Tagset für historische Sprachstufen des Deutschen. *Journal for Language Technology and Computational Linguistics*, 28(1):85–137. Special Issue.

Schiller, Anne, Simone Teufel, Christine Stöckert und Christine Thielen

- 1999 Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset). [\[online\]](#)

Converter framework: SaltNPepper

Zipser, Florian und Laurent Romary

- 2010 A model oriented approach to the mapping of annotation formats using standards. Proceedings of the Workshop on Language Resource and Language Technology Standards, LREC 2010. [\[online\]](#)

Search and visualization interface: ANNIS

Zeldes, Amir, Julia Ritz, Anke Lüdeling und Christian Chiarcos

- 2009 ANNIS: A Search Tool for Multi-Layer Annotated Corpora. *Proceedings of Corpus Linguistics 2009*, July, 20–23, Liverpool. [\[online\]](#)

HIPKON

Coniglio, Marco, Karin Donhauser und Eva Schlachter

- 2015 HIPKON (Version 1.0), Humboldt-Universität zu Berlin. SFB 632 Teilprojekt B4
<http://www.sfb632.uni-potsdam.de/>
<http://hdl.handle.net/11022/0000-0000-2D18-4>