

Linke Satzperipherie, Mittelfeld und Nullsubjekte im Mittelniederdeutschen

Zur Untersuchung syntaktischer Phänomene mit dem CHLG

Hannah Booth, Anne Breitbarth und Melissa Farasyn

1 Einleitung

Die Syntax des Mittelniederdeutschen erfährt seit einigen Jahren erfreulicherweise die Aufmerksamkeit, die sie verdient, und die ihr so lange versagt geblieben ist (vgl. Peters 1973: 105). Pioniere auf dem Gebiet ihrer Erforschung sind u. a. Rösler (1997), Mähl (2004, 2014), Tophinke (2009), Petrova (2012, 2013), Wallmeier (2013, 2015) und Breitbarth (2014). In den letzten Jahren sind auch eine Reihe von Dissertationen zur Syntax des Mittelniederdeutschen entstanden (bspw. Farasyn 2018; Merten 2018; Ihden 2020), was für die Zukunft hoffen lässt. Im Rahmen der von der Deutschen Forschungsgemeinschaft geförderten Initiative ‚Deutsch Diachron Digital‘ (DDD) wurden in den letzten Jahren verschiedene große Referenzkorpora zu historischen Sprachstufen des Deutschen angelegt, die inzwischen auch alle öffentlich zugänglich sind. Auch für das Mittelniederdeutsche steht seit 2019 das Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650) (ReN) (ReN-Team 2021) zur Verfügung. Seit Ende 2020 ist auch ein syntaktisch annotiertes Subkorpus des ReN online verfügbar, das Corpus of Historical Low German (CHLG; www.chlg.ugent.be). Dieses soll im vorliegenden Beitrag kurz vorgestellt und seine Besonderheiten im Vergleich zu anderen Korpora anhand ausgewählter syntaktischer Fallstudien illustriert werden.

2 Das CHLG

2.1 Hintergrund

Die beiden Projekte für ein Korpus des Mittelniederdeutschen, das ReN und das CHLG, wurden zunächst unabhängig voneinander konzipiert und beantragt. Als für beide etwa zeitgleich Ende 2013 eine finanzielle Förderung bewilligt wurde, im Falle des ReN durch die DFG, im Falle des CHLG durch die belgische Hercules-Stiftung (die 2016 in das neue Forschungsinfrastrukturprogramm des Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO) eingegliedert wurde), wurde auf Initiative von Robert Peters beschlossen, die Ressourcen zu bündeln. Die MitarbeiterInnen des CHLG halfen bei der Annotation der Wortarten (POS-Tagging) und der Morphologie einiger Texte des ReN; im Gegenzug durfte das CHLG ausgewählte Texte des ReN syntaktisch annotieren.

Das ReN ist Teil des Korpus historischer Texte des Deutschen der Initiative Deutsch Diachron Digital (DDD), zu dem auch die Referenzkorpora Altdeutsch 750–1050 (ReA; <https://www.deutschdiachrondigital.de/rea/>), Mittelhochdeutsch 1050–1350 (ReM; <http://www.linguistics.rub.de/rem/>) und Frühneuhochdeutsch 1350–1650 (ReF; <http://www.linguistics.rub.de/ref/>) zählen. Es bietet Zugang zu ca. 1,45 Millionen Wortformen in 145 diplomatisch transkribierten, lemmatisierten und grammatisch annotierten Texten. Das CHLG bietet zusätzlich zur morphologischen Annotation des ReN für neunzehn der Texte (ca. 187.500 Wortformen) eine syntaktische Annotation nach dem Schema der Penn Parsed Corpora of Historical English (Santorini 2010), vgl. Booth et al. (2020). Es bietet somit auch eine mittelniederdeutsche Fortsetzung der Heliand Parsed Database (HeliPaD-Korpus, Walkden 2015, 2016), das auf Sievers' (1878) Edition der Handschrift Heliand C beruht. Beide, das CHLG und das HeliPaD, sind lemmatisiert und verwenden dasselbe Parsingschema, unterscheiden sich jedoch in der Annotation der Wortarten (POS-Tagging) und Morphologie, wie im Folgenden ausgeführt wird.

2.2 Aufbau und Annotation des CHLG

Da die syntaktische Annotation historischer Texte sehr arbeitsaufwändig ist, konnte nur ein Teil der Texte des ReN für das CHLG geparkt werden. Um dennoch eine ausgewogene Auswahl zu erreichen, wurde entschieden, das CHLG auf die Schreibsprachen Westfälisch, Ostfälisch, Nordniedersächsisch und Ostelbisch zu beschränken und baltische, südmärkische sowie niederrheinische Texte auszuschließen. Aus diesen vier Schreibsprachen wurden je vier bzw. fünf Texte für das CHLG geparkt, vgl. Tabelle 1. Wie auch im ReN besteht ein gewisses Ungleichgewicht in der Verteilung von Textsorten über den Zeitraum, was der Überlieferungslage geschuldet ist: typischerweise sind

Rechts- und Urkundentexte früher und über den gesamten Zeitraum belegt, während andere Textsorten erst im 15. Jh. hinzukommen.

Text	Zeitraum	Textsorte
Westfälisch		
Abdinghofer Arzneibuch	1451–1500	Wissenschaft
Herforder Rechtsbuch	1375	Recht
Soest, Schrae im Statutenbuch	1367	Recht
Rüthen, Statutarrecht	1300	Recht
Spiegel der Leyen	1444	Religion
Ostfälisch		
Braunschweiger Urkunden	1301–1500	Recht/Urkunden
Duderstadt, Braunschweiger Stadtrecht	1279	Recht
Engelhus, Chronica nova	1435	Literatur
Helmstedt, Zeno	1401–1450	Literatur
Nordniedersächsisch		
Agneta Willeken	1535	Privatbrief
Bremer Urkunden	1300–1350	Recht/Urkunden
Buxtehuder Evangeliar	1451–1500	Religion
Griseldis	1502	Literatur
Oldenburger Urkunden	1350–1500	Recht/Urkunden
Ostelbisch		
Flos unde Blankeflos	1401–1450	Literatur
Greifswalder Bürgersprache	1451	Recht
Rostocker Bürgersprache	1580	Recht
Schweriner Stadtbuch	1451–1500	Recht
Stralsunder Urkunden	1301–1500	Recht/Urkunden

Tabelle 1: Textauswahl des CHLG

Aufgrund der Zusammenarbeit mit dem ReN benutzt das CHLG nicht das von den anderen historischen Korpora der Penn-Familie gebrauchte POS-Tagset (Santorini 1990). Stattdessen baut die syntaktische Annotation, die bis auf kleine Anpassungen dem Penn-Standard folgt, auf dem HiNTS-Tagset des ReN auf,¹ das viel feinere Unterscheidungen zulässt und bereits eine ganze Reihe syntaktischer Merkmale mit einschließt. So unterscheidet das HiNTS zwischen 26 POS-Tags für determinative Elemente, wo das Penn-Tagset nur ein Tag D hat. Diese 26 Tags beginnen alle mit dem

¹ HiNTS steht für Historisches Niederdeutsch-Tagset (Barteld et al. 2018). Es handelt sich um eine Weiterentwicklung des Historischen Tagsets (HiTS) (Dipper et al. 2013), das wiederum auf dem Stuttgarter-Tübingen Tagset (STTS) (Schiller et al. 1999) aufbaut.

Grundlabel D, werden dann jedoch weiter nach verschiedenen Merkmalen in der Reihenfolge: Art des Determinativs–Artikelartigkeit–Position ausdifferenziert, die in Tabelle 2 aufgelistet sind, also bspw. DDARTA (D+D+ART+A) für ein definites, artikelartiges Determinativ, das seinem Bezugsnomen vorangeht, wie z. B. *dat* in ***dat land***. Ein Determinativ, das negativ, nicht artikelartig und prädikativ ist (also als Kopf eines nominalen Prädikats fungiert), wie z. B. *niemand* in *dat ist **niemand***, wird im HiNTS als DNEGD getagt (D+NEG+D).

Merkmal	möglicher Wert	Label
Art des Determinativs	definit	D
	indefinit	I
	negativ	NEG
	possessiv	POS
	relativisch	REL
	interrogativ	W
Artikelartigkeit	artikelartig	ART
	nicht artikelartig	—
Position	attributiv	A
	nachgestellt	N
	substituierend	S
	prädikativ	D

Tabelle 2: Kodierung verschiedener Merkmale von Determinativen im HiNTS

Dieser Annotation der Wortart (sowie der Flexionsmorphologie und der Lemmata, die auch im ReN zur Verfügung stehen) fügt das CHLG nun eine eigene Ebene der syntaktischen Annotation hinzu. Dabei wurde entschieden, dem Penn-Annotationsstandard zu folgen. Dabei handelt es sich um einen konstituenzbasierten Standard, im Unterschied zu dependenzbasierten Standards wie den Universal Dependencies (Nivre et al. 2016). Solche dependenzbasierten Ansätze wurden auch bereits für historische Korpora eingesetzt, bspw. für die Familie der Pragmatic Resources in Old Indo-European Languages (PROIEL) (Haug und Jøhndal 2008). Beide Systeme sind gleich mächtig und prinzipiell isomorph (Haug 2015; Taylor 2020). Es ist wichtig, im Auge zu behalten, dass Baumbanken Werkzeuge für eine möglichst präzise und effiziente Extraktion von Daten sind, keine exakte sprachwissenschaftliche Analyse.² Die Wahl des Penn-Systems im

² So werden im Penn-System beispielsweise weder VP-Knoten noch andere funktionale Projektionen im Mittelfeld annotiert; alle unmittelbaren Konstituenten eines Satzes sind direkte Töchter des Satzknotens. Dadurch können Konstituenten (Verben, Argumente und Ergänzungen) schnell und eindeutig über ihre Dominanz- und Präzedenzrelationen zueinander gefunden werden.

Falle des CHLG ist hauptsächlich durch die Kompatibilität mit den zahlreichen anderen historischen Korpora motiviert, wodurch dieselben Suchanfragen benutzt werden können, um mehrere Korpora zu durchsuchen. International ist das System auch bekannter und somit zugänglicher als andere Annotationsschemata. Zudem können bereits existierende Werkzeuge wie das Suchprogramm CorpusSearch (Randall 2005) oder das graphische Nutzerinterface (GUI) Annotald für die syntaktische Annotation (Beck, Ecaj und Ingason 2015) ohne weitere Modifikationen verwendet werden, da sie für das Penn-Annotationsschema entwickelt wurden. Für linguistische Argumente für die Verwendung des Penn-Systems verweisen wir auf Booth et al. (2020).

2.3 Wozu ein geparstes Korpus?

Historische Syntaxforschung war lange eine mühsame Disziplin, vor allem durch die aufwändige Sammlung von Belegen aus Handschriften, Drucken oder Ausgaben. Elektronische Textsammlungen brachten nur bedingt Erleichterung, da sich syntaktische Phänomene einfachen Volltextsuchen entziehen, sobald sie über die Abfolge einzelner Wörter hinausgehen und sich auf die Abfolge oder Interaktion mehr oder weniger umfangreicher Satzglieder beziehen. Die Entwicklung syntaktisch annotierter Korpora in den letzten Jahrzehnten hat einen entscheidenden Beitrag dazu geleistet, quantitative Untersuchungen syntaktischer Veränderungen in vielen Sprachen zu erleichtern und reproduzierbare und zuverlässige Untersuchungsergebnisse zu ermöglichen. Einen Überblick über diese Entwicklungen bieten beispielsweise Pintzuk, Taylor und Warner (2017).

Wie in Abschnitt 2.2 angesprochen, enthalten einige POS-Tags im HiNTS bereits gewisse syntaktische Informationen. Im oben diskutierten Beispiel der Determinative, aber auch bei anderen Wortarten, ist ein Teil des Tags für die Kodierung der Position des getagten Tokens hinsichtlich eines Bezugswortes (vorangestellt/attributiv, nachgestellt) reserviert. Somit ist im ReN allein schon auf der Ebene der Wortartannotation eine Bearbeitung gewisser syntaktischer Fragestellungen möglich. Die ANNIS-Suchabfragesprache AQL erlaubt auch die Arbeit mit Abstandsoperatoren, um Wortformen, Lemmata, Wortart- oder morphologische Annotationen von Tokens im Kontext anderer solcher Elemente zu suchen. Zusätzlich ermöglicht die Annotation von Satzgrenzen (`bound_sent`) im ReN – eine sehr willkommene Innovation gegenüber den anderen DDD-Korpora! – die Einschränkung einer Suche auf Einfachsätze. Syntax ist jedoch mehr als die direkte Adjazenz bzw. konkrete Abstände zwischen einzelnen Wörtern. Dem ReN wie auch den anderen DDD-Korpora fehlt die Möglichkeit, Konstituenten beliebigen Umfangs in Beziehung zueinander zu suchen. Der syntaktischen Funktion nominaler Konstituenten kann man sich eventuell über die morphologische Annotation der Kasusmarkierung annähern, doch auch hier besteht keine Möglichkeit, mit Konstituenten unterschiedlichen

Umfangs umzugehen. Im Folgenden sollen beispielhaft drei Phänomene besprochen werden, die einen Zugang zu linearen, hierarchischen und funktionalen Beziehungen zwischen Konstituenten unterschiedlichen Umfangs erfordern und deren Untersuchung daher in entscheidender Weise von der konstituentenstrukturellen Annotation des CHLG Gebrauch machen kann.

3 Mehrfache XP-Voranstellung in Sätzen mit Verbbewegung

Petrova (2012) untersucht an ausgewählten Texten die verschiedenen Möglichkeiten der Späterstellung des finiten Verbs im unabhängigen Satz im Mnd. Neben Abfolgen, die mit einer V2-Grammatik vereinbar sind und auch im Gegenwartsdeutschen möglich wären (z. B. Linksversetzungen, Freies Thema, scheinbare mehrfache Vorfelddbesetzung), bespricht sie auch Abfolgen, die nicht mit einer solchen Grammatik vereinbar sind. Dabei formuliert sie diagnostische Kriterien für die Feststellung von Verbbewegung in die linke Satzperipherie, um (unabhängige) Sätze ohne Verbbewegung, die in ihren Daten auch belegt sind, von solchen mit mehrfacher Vorfelddbesetzung zu unterscheiden. Im Wesentlichen identifiziert Petrova zwei größere Gruppen von V2-Verletzungen, einerseits Abfolgen, bei denen das Subjekt einer Adverbialbestimmung vorangeht (1a-1b) und andererseits solche, bei denen eine Adverbialbestimmung dem Subjekt vorangeht (1c-1e). Zudem scheinen direkte Objekte Subjekten im Vorfeld vorangehen zu können, (1f), allerdings ist nicht bei allen Belegen klar, ob Verbbewegung stattgefunden hat.

- (1) a. Subjekt – temporale PP
[Constantinus] [des morgenes vro] segede dit sinen vrunden (SW 118, 24)
 ‚Des Morgens früh sagte Constantinus dies seinen Freunden.‘
 (nach Petrova 2012: 173)
- b. Subjekt – kausale PP
[Kato de wise] [vor leide] dodede sic selven (SW 87, 27)
 ‚Cato der Weise tötete sich selbst aus Leid.‘
 (nach Petrova 2012: 173)
- c. temporale Adverbialbestimmung – Subjekt
[Nicht langhe darma] [der sone] volghete na vor ludeke (LChr 112, 3)
 ‚Nicht lange danach folgte der Sohn nach Richtung Lüttich.‘
 (nach Petrova 2012: 174)
- d. temporaler Adverbialsatz – pronominales Subjekt
[Do Alexander vornam ere missedat], [be] let se bemuren (SW 78, 4)
 ‚Als Alexander ihre Missetat vernahm, ließ er sie einmauern.‘
 (nach Petrova 2012: 174)

- e. temporaler Adverbialsatz – DP-Subjekt
[Do he starf], [de Romere] clageden in sere (SW 90, 17)
 ‚Als er starb, beklagten die Römer ihn sehr.‘
 (nach Petrova 2012: 174)
- f. direktes Objekt – Subjekt
[Eine warheit] [ich] wille dir sagen (SW 100, 29)
 ‚Eine Wahrheit möchte ich dir sagen.‘
 (nach Petrova 2012: 175)

Da Petrova nur die absoluten Zahlen der gefundenen Belege angibt³ und auch nur eine eingeschränkte Textauswahl untersucht – vier der sieben Texte bzw. Textsammlungen sind Chroniken und einer ein Reisebericht⁴ – stellt sich die Frage, wie der Befund in einem ausgewogeneren Korpus aussähe und wie häufig solche V2-Verletzungen überhaupt in unabhängigen Sätzen vorkommen. Eine solche Frage stellt eine Herausforderung für das ReN dar. Zwar ist es möglich, durch Bezug auf die Satzspanne (*bound_sent*) nach Phänomenen innerhalb eines Einfachsatzes oder Gliedsatzes zu suchen, allerdings können nur individuelle Tokens, ihre Eigenschaften (Lemma, Wortart, morphologische Eigenschaften) und ihr Abstand zu anderen Tokens gesucht werden. Dadurch würde man verschiedene Suchanfragen benötigen, um nach Abfolgen wie denen in (1) zu suchen. Je nachdem, ob ein Subjekt oder Objekt intern komplex ist (volle DP wie in (1c), evtl. noch mit Attributen erweitert, vs. pronominal wie in (1d)) oder eine adverbiale PP mehr als nur Präposition und Nomen enthält, müsste bspw. mit verschiedenen Abstandsangaben bzw. Abstandsbereichen zwischen Tokens gearbeitet werden. Die Gefahr, hierbei Belege zu übersehen, liegt auf der Hand. Für Abfolgen mit linksperipheren Adverbialsätzen wie (1e) ist eine ANNIS-Anfrage eigentlich gar nicht zu formulieren, da zwei Satzspannen nicht innerhalb einer Anfrage miteinander in Beziehung gesetzt werden können.

Mit einer konstituentenbasierten syntaktischen Annotation wie der des CHLG ist hingegen eine einzige Suchanfrage ausreichend. Man beschränkt die Suche auf unabhängige Sätze (*node = IP-MAT**), spezifiziert, dass die dritte Tochter dieses Satzknotens ein finites Verb sein soll ((*IP-MAT* iDomsNum 3 V*FIN*)), und stellt über weitere Spezifikationen sicher, dass die Anzahl der falschen Positive möglichst gering ausfällt.⁵

³ Relative Belegzahlen finden sich nur in Donhauser und Petrova (2009: 299), wo die Anzahl der uninvertierten Nachsätze nach vorangestellten Temporalsätzen in zwei mnd. Texten zu finden sind, die auch einen Teil von Petrovas (2012) Korpus ausmachen, nämlich in der ‚Sächsischen Weltchronik‘ (81,3% V3 nach Temporalsätzen) und Ludolfs von Sudheim ‚Reise ins Heilige Land‘ (4,62%).

⁴ Hinzu kommen eine Predigtsammlung sowie eine Sammlung von Rechtstexten aus Stadtbüchern des 14.-16. Jhs.

⁵ So muss man bspw. spezifizieren, dass die erste Konstituente keine koordinierende Konjunktion, META-Tags, die elliptische Konjunktionen über Indizes verbinden, oder textstrukturierende Zeichen

Da CorpusSearch auch immer die Gesamtzahl der Belege in der spezifizierten Umgebung (hier *node* = IP-MAT*) angibt, bekommt man zudem schnell einen Überblick über die Häufigkeit eines Phänomens.

Im CHLG finden sich alle von Petrova beschriebenen Abfolgen:⁶

- (2) a. Subjekt – temporale PP
 [*De engel godes*] [*na der tijd*] *stech neder an den dik*
 (Buxtehuder Evangeliar, 1451–1500)
 ‚Nach der Zeit stieg der Engel Gottes nieder an den Deich.‘
- b. Subjekt – temporaler Adverbialsatz
 [*Enoch*] [*alse he was lxx iar olt*] *ghewan Matuselem* (Engelhus, *Chronica nova*, 1435)
 ‚Enoch, als er 65 Jahre alt war, zeugte Matuselem.‘
- c. temporale Adverbialbestimmung – Subjekt
 [*Dan*] [*de richtere*] *sal den man kuden de in dem gherichte bescreghen ys to dreyweyrteyn nachten*. (Rüthener Statutarrecht, 1300)
 ‚Dann soll der Richter den Mann künden, der in dem Gericht zu dreimal 14 Nächten angeklagt ist.‘

(z. B. <I) enthält ((IP-MAT* iDomsNum 1 ! KON|META|...)) und es sich bei der zweiten Konstituente nicht etwa um die präverbale Negationspartikel *ne/en* handelt ((IP-MAT* iDomsNum 2 ! PTKNEG)), die als Klitik/Affix für die Berechnung der Verbposition keine Rolle spielt (vgl. Breitbarth 2014). Da Linksversetzungsstrukturen nicht als Verletzungen der V2-Regel aufgefasst werden, wurden Linksversetzungen (ia), von einem Resumptivum aufgenommene linksperiphere Adverbialsätze – oftmals korrelative Strukturen (ib-ic) –, manuell aussortiert.

- (i) a. [*Alle deghe*] *de to dussen iarmarkededen comet in de stat to Ruden*] **de** *solen hebbden syker gheleyde*. (Rüthener Statutarrecht, WF, 14.1)
 ‚Alle diejenigen, die zu diesen Jahrmärkten kommen in die Stadt zu Rüthen, die sollen sicheres Geleit haben.‘
- b. [*Do dat kint vorstarf*] **do** *wart der moder mit rechte syn erue toghedelet*. (Herford, WF, 14.2)
 ‚Als das Kind verstarb, da wurde der Mutter mit Recht sein Erbe zugeteilt.‘
- c. [*also my myn vader sant heft*] **also** *sende ik in* (Buxtehuder Evangeliar, NNS, 15.2)
 ‚Wie mich mein Vater gesandt hat, so sende ich euch.‘
- d. [*Swe den Dotslagh eder de wunnynghede doyt*] **De** *ne scal dar nerghe vrede hebbden* (Oldenburger Urkunden, NNS, 14.2-15)
 ‚Wer immer den Totschlag oder die Verwundung bewirkt, der soll dort nirgends Frieden haben.‘

Freie Relativsätze, die sehr häufig in Linksversetzungsstrukturen vorkommen, und die auch im Gegenwartsdeutschen möglich wären (id), wurden von vornherein in der Suchanfrage ausgeschlossen ((IP-MAT* Doms !CP-FRI*)).

⁶ Aufgrund stark abweichender Wortstellungsmuster (z. B. Häufung von Verbendstellung) wurden gereimte Texte (Flos unde Blankeflos, Spieghel der Leyen, Zeno) allerdings von dieser Untersuchung ausgeschlossen.

- d. temporaler Adverbialsatz – pronominales Subjekt
[Do dat geruchte vor Griseldis quam] [se] was trurich als yk meyne (Griseldis, 1502)
 ‚Als das Gerücht vor Griseldis kam, war sie traurig, wie ich meine.‘
- e. temporaler Adverbialsatz – DP-Subjekt
[DO Albert stur starf] [sin eldeste sone] nam sin berwede.
 (Herforder Rechtsbuch, 1375).
 ‚Als Albert Stur starb, nahm sein ältester Sohn sein Heergewäte [= Kriegsausrüstung].‘
- f. direktes Objekt – Subjekt
[Der moder rade] [de paepe] nympt gbelike dele der suster in der moder rade
 (Herforder Rechtsbuch, 1375).
 ‚(was) die Gerade der Mutter (betrifft), nimmt der Pfaffe gleiche Teile wie die [seine] Schwester.‘

Diese verteilen sich nicht ganz gleichmäßig über die Texte, wie aus Tabelle 3 ersichtlich wird, in der eine grobe Unterteilung in Rechtstexte i. w. S. und nicht-Rechtstexte vorgenommen wird.

Eine Besonderheit v. a. der mnd. Rechtssprache sind vorangestellte Konditionalsätze, bei denen eine Konstituente des Konditionalsatzes von einem resumptiven d-Pronomen wie in einer Linksversetzungsstruktur wieder aufgenommen wird (3). Man könnte solche Fälle als ‚Teilresumption‘ bezeichnen. Es kann sich dabei um ein Subjekt oder ein Objekt handeln.

- (3) a. Teilresumption – Subjekt
[Wert iemant dar ouer beargvonet] die schal sick mit seinem Eide Purgiren Bei Straffe des Rades (Rostocker Bürgersprache 1580)
 ‚Wird jemand deshalb verdächtigt, der soll sich mit seinem Eid für unschuldig erklären oder die vom Rat festgesetzte Strafe auf sich nehmen.‘
- b. Teilresumption – Objekt
[Were th dath Ienich timmerholt tho schepe queme vor de Stadt edder an de bauene] dat schall nen vorkoper kopen bynnen eneme daghe vnd ener nacht (Greifswalder Bürgersprache 1451)
 ‚Wäre es der Fall, dass irgendwelches Zimmerholz zu Schiff käme vor die Stadt oder in den Hafen, das soll kein Verkäufer kaufen innerhalb eines Tages und einer Nacht.‘

Solche Konstruktionen wurden vorläufig, wie auch bei Donhauser/Petrova (2009), mitgezählt, ebenso wie Konstruktionen, in denen eine Konstituente des vorangestellten Adverbialsatzes von einem Personalpronomen wiederaufgenommen wird, ähnlich wie bei einem Freien Thema (vgl. (2d)). In Tabelle 3 werden diese Belege

jedoch in Klammern angegeben, um sie ggfs. von anderen mehrfach besetzten Vorfeldern trennen zu können.⁷

Text	Mehrfache XP-Voranstellung			Total IP- MAT
	Su-XP-Vf	XP-Su-Vf	XP-XP -Vf	
Rechtstexte u. Urkunden				
Braunschweiger Stadtrecht 1279	0	0	0	132
Braunschweiger Urkunden 1301–1500	0	0	0	110
Rüthener Statutarrecht 1300	0	0(5)	0(5)	265
Soester Schrae 1367	0	1	1	262
Herforder Rechtsbuch 1375	1	2(2)	0(6)	830
Bremer Urkunden 1300–1350	0	0	0	48
Oldenburger Urkunden 1300–1530	0	0	0	231
Stralsunder Urkunden 1301–1500	1	0	0(2)	314
Greifswalder Bürgersprache 1451	0	0(6)	0(3)	148
Schweriner Stadtbuch 1451–1500	0	0	0(1)	327
Rostocker Bürgersprache 1580	0	0	0	44
Gesamt	2	3(+13)	1(+17)	2711
Literatur/Religion/Wissenschaft/Privatbrief				
Engelhus, Chronica 1435	15	0(2)	2	1532
Abdinghofer Arzneibuch 1451–1500	3	1(1)	0	1770
Buxtehuder Evangeliar 1451–1500	2	0(2)	0(2)	2038
Griseldis 1502	0	2	0	537
Agneta Willeken 1535	0	0(1)	0	188
Gesamt	20	3(+6)	2(+2)	6065

Tabelle 3: Mehrfache XP-Voranstellung in den Texten des CHLG (in Klammern: Teilresumption)

Allgemein lässt sich anhand der relativen Zahlen feststellen, dass mehrfache XP-Voranstellung nur sehr selten im CHLG vorkommt. Allerdings fallen zwei Dinge ins Auge. Erstens ist mehrfache XP-Voranstellung mit 0,5% aller unabhängiger Deklarativsätze in nicht-Rechtstexten weniger häufig belegt als in Rechtstexten und Urkunden (8,7%). Das Bild verändert sich jedoch, wenn die v. a. in Rechtstexten häufigen Teilresumtionen nach vorangestellten Adverbialsätzen nicht mitgezählt

⁷ In Breitbarth (2018) werden Argumente dafür angeführt, solche Strukturen als weiteres Zwischenstadium auf dem Weg zur Entwicklung von V2, zusätzlich zur Resumption (für das Mnd. vgl. Tophinke 2009), zu interpretieren.

werden: dann ist mehrfache XP-Voranstellung mit 0,4% doppelt so häufig in nicht-Rechtstexten belegt wie in Rechtstexten und Urkunden (0,2%). Zweitens ist die Art der mehrfachen XP-Voranstellung deutlich asymmetrisch verteilt. Abgesehen von den Teilresumptionsstrukturen ist die Abfolge XP-Su-Vf wie in (2d) und (2e), die für die Gegenwartssprache vor allem für das Kiezdeutsche (Wiese, Öncü und Bracker 2017; Wiese und Müller 2018; Wiese et al. 2020), zunehmend jedoch auch für standardnahes gesprochenes Deutsch beschrieben wurde (Schalowski 2015; Wiese und Müller 2018; Breitbarth 2022, 2023), im CHLG extrem selten (insgesamt 0,06%). Falls es hier also eine historische Kontinuität geben sollte, läge sie bestenfalls im Bereich der gesprochenen Sprache und somit außerhalb des Zugriffs der historischen Korpuslinguistik.⁸

Die im Gegenwartsdeutschen nicht mögliche Abfolge Su-XP-Vf ist viel stärker in nicht-Rechtstexten vertreten, und dort v. a. in einem Text, nämlich der Chronik von Engelhus (wo sie allerdings auch nur in 1% der unabhängigen Sätze vorkommt). Das ist vor allem deshalb interessant, weil Petrova (2012) v. a. Chroniktexte untersucht. Diese Konstruktion verbindet eindeutig eine Art Freies Thema mit einem Rahmensetzer: Ein Subjekt wird eingeführt (was X betrifft), dann folgt ein V2-Satz mit bspw. einem Rahmensetzer im Vorfeld (2a-2b).

Zusammenfassend kann festgestellt werden, dass die konstituentenbasierte syntaktische Annotation des CHLG eine schnelle und zuverlässige Überprüfung, Aktualisierung und Präzisierung älterer Forschungsergebnisse zur mehrfachen Vorfeldbesetzung erlaubt, sowie die Annäherung an neue Fragen wie bspw. zum Alter und zur historischen (Dis)Kontinuität bestimmter mehrfacher Vorfeldbesetzungen ermöglicht.

4 Spiegeleffekte im Mittel- und Nachfeld

Sogenannte ‚Spiegeleffekte‘, bei denen die im Mittelfeld (MF) bevorzugte Wortstellung im Nachfeld (NF) in umgekehrter Abfolge erscheint, sind in verschiedenen modernen westgermanischen Sprachen nachgewiesen worden (Koster 1974, 2001; Barbiers 1995; Haider 2000; Pittner 2004). Im modernen Niederländischen manifestiert sich dieser Effekt in der Wortstellung von PP-Argumenten und PP-Adjunkten: die bevorzugte Wortstellung im MF ist Adjunkt-Argument (4a) und im NF Argument-Adjunkt (4b).

⁸ Die Tatsache, dass im Gegenwartsdeutschen (innerhalb ebenso wie außerhalb des Kiezdeutschen) keine Teilresumtionen belegt sind, sowie weiterhin die Tatsache, dass Inversion nach vorangestellten Adverbialbestimmungen im Mnd. kaum belegt ist – Donhauser und Petrova (2009: 299) finden 0,93% in der ‚Sächsischen Weltchronik‘ und 1,54% in Ludolfs ‚Reise‘; in den hier untersuchten Texten im CHLG sind es ca. 7% aller Sätze mit vorangestellten Adverbialbestimmungen, allerdings stammen zwei Drittel dieser Belege (8/12) aus einem einzigen Text (Griseldis) – weist darauf hin, dass die mehrfachen Vorfeldbesetzungen im Mnd., inklusive der Abfolge XP-Su-Vf, eher darauf zurückzuführen sind, dass die V2-Stellung nach vorangestellten Adverbialbestimmungen noch nicht vollständig etabliert war.

- (4) a. *Hij beeft [tijdens de pauze] [aan zijn vader] gedacht*
 b. *Hij beeft gedacht [aan zijn vader] [tijdens de pauze]*
 ‚Er hat während der Pause an seinen Vater gedacht‘ (Koster 2001: 1)

Ein ähnlicher Effekt ist in Bezug auf die Abfolge mehrerer semantisch unterschiedlicher Adjunkte belegt, wie Haider (2000) für verschiedene westgermanische Sprachen zeigt, vgl. (5) (adaptiert aus Haider 2000: 96). Für das Neuhochdeutsche insbesondere gilt die Wortstellung in (5a) für das MF und die invertierte Wortstellung in (5b) für das NF (siehe auch Pittner 2004).

- (5) a. Zeit < Ort < Art&Weise – V
 b. V – Art&Weise < Ort < Zeit

Das Niederdeutsche befindet sich im kontinentalwestgermanischen Dialektkontinuum typologisch und geographisch gesehen zwischen dem Niederländischen und dem Hochdeutschen. Aufgrund dessen würde man erwarten, dass sich solche Spiegeleffekte auch im Niederdeutschen identifizieren lassen. Das ist bisher allerdings noch nicht untersucht worden, weder synchron noch diachron. Das Mnd. bietet sich nun als Ausgangspunkt für eine erste Untersuchung dieser Frage an, da in der Literatur bereits gezeigt wurde, dass die Verbal- und Satzklammer, die zu einem strukturell abgegrenzten MF (und NF) führt, im Laufe der Periode immer weiter ausgebaut wird, dahingehend, dass zunehmend mehr Konstituenten im Mittelfeld stehen und weniger ausgeklammert werden (Rösler 1997; Mähl 2014; Dreessen und Ihden 2015). Petrova (2013) zeigt, dass im Mnd. zumindest oberflächlich sowohl OV- als auch VO-Wortstellungen nachgewiesen sind. Daher findet man im Mnd. Sätze mit mehreren Konstituenten im MF (6a) sowie Sätze mit mehreren Konstituenten im NF (6b).

- (6) a. *Nemant schall [ocke] [des nachtes] [mit lichten ane lucerne] [In den stall] gan*
 (Greifswalder Bürgersprache, 1451)
 ‚Niemand soll auch des Nachts mit Lichtern ohne Laterne in den Stall gehen‘
 b. *vnde se schal sitten [openbare] [ane bode] [in deme bus]*
 (Stralsunder Urkunden, 1301–1500)
 ‚und sie soll sichtbar ohne Beistand in dem Haus sitzen‘

Allerdings ist die diachrone Entwicklung dieses Phänomens im Westgermanischen bisher noch nicht untersucht worden. Es ist zwar bekannt, dass die Wortstellung in älteren germanischen Sprachstufen im Vergleich zu den modernen Sprachen generell freier ist (bspw. Hinterhölzl und Petrova 2009; Bech und Eide 2014). Der Status der Wortstellung im MF ist jedoch bei mehreren älteren Sprachstufen umstritten,

wie bspw. Frey (2003) und Demske und Patak (2016) zu Adjunkten im MF im Frühneuhochdeutschen zeigen. Das Mnd. kann daher sowohl eine historische als auch eine typologische Lücke füllen.

Im Folgenden untersuchen wir, ob die oben genannten Spiegeleffekte im Mnd. zu erkennen sind. Das CHLG ist aufgrund seiner konstituentenbasierten Annotation ideal für die Untersuchung dieser Frage. Die hier präsentierte Vorstudie ist auf Hauptsätze mit einem strukturell abgegrenzten MF und NF, d. h. Hauptsätze mit einer Verbalklammer aus einem Hilfsverb und einem nicht-finiten Verb beschränkt.⁹ Aufgrund der konstituentenbasierten syntaktischen Annotation des CHLG und der Abfragesprache CorpusSearch (Randall 2005) lassen sich MF und NF gut identifizieren und isolieren. Dazu wird spezifiziert, welche Art(en) von Konstituenten im MF oder NF stehen sollen und auch jeweils wie viele davon. Um zum Beispiel die Abfolge von PP-Argumenten und PP-Adjunkten im MF zu untersuchen, kann eine Suchanfrage nach Hauptsätzen mit zwei oder mehr PPs im MF formuliert werden.¹⁰ Solche Suchanfragen sind nur durch die konstituentenbasierte Annotation möglich; als Voraussetzung muss man mehrere ganze phrasale Konstituenten (bspw. PPs und ADVPs) beschreiben können, die alleine durch wortgebundene POS-Tags wie bei der Annotation des ReN nicht identifizierbar sind. Durch solche Anfragen kann man also zuverlässige Datensätze mit sehr spezifischen und subtilen syntaktischen Merkmalen relativ schnell aus dem CHLG abrufen, die sich gut für weitere quantitative oder qualitative Untersuchungen eignen.

Als Erstes wurde untersucht, ob die Abfolge von PP-Argumenten und PP-Adjunkten im NF im Vergleich zur Abfolge im MF invertiert ist, wie im Niederländischen (4). Dazu wurden alle Hauptsätze mit mindestens zwei PPs im MF oder im NF herausgesucht. Aus den so gefundenen 81 Hauptsätzen mit mindestens zwei PPs im MF wurden manuell die 18/81 Sätze mit einem PP-Argument und einem PP-Adjunkt im MF herausgefiltert (in den übrigen Fällen handelte es sich um rein adverbiale PPs). Die Ausgangshypothese war, dass das Mnd. sich wie das Niederländische verhält, also im MF die Abfolge Adjunkt-Argument zeigt. Bei allen 18 Sätzen wurde diese Hypothese bestätigt. Einige Beispiele werden in (7) gezeigt.

- (7) a. *vnd*con*¹¹ wart dar [na sime cccxv iare] [in den paradis] ghevort*
 (Engelhus, Chronica nova, 1435)
 ‚und [er] wurde da nach seinem 365. Jahr in das Paradies geführt‘

⁹ Suchanfrage: (IP-MAT* iDoms VAFIN|VMFIN) AND (IP-MAT* iDoms VVINF|VVPP) AND (VAFIN|VMFIN precedes VVINF|VVPP)

¹⁰ Suchanfrage: (IP-MAT* iDoms VAFIN|VMFIN) AND (IP-MAT* iDoms VVINF|VVPP) AND (VAFIN|VMFIN precedes VVINF|VVPP) AND (IP-MAT* iDoms PP) AND (PP hasSister PP*) AND (VAFIN|VMFIN precedes PP) AND (PP precedes VVINF|VVPP) AND (VAFIN|VMFIN precedes PP*) AND (PP* precedes VVINF|VVPP)

¹¹ Siehe Abschnitt 5, insbesondere Fußnote 16.

- b. *vnde sik ne mach [mit rechte] nemant [to der vormundescap] teen*
(Herforder Rechtsbuch, 1375)
,und niemand kann rechtmäßig die Vormundschaft an sich ziehen‘
- c. *so en sculden se [binnen dessen Ses weken]... nigt [vpppe de straten] komen*
(Oldenburger Urkunden, 1350–1500)
,so sollten sie binnen diesen sechs Wochen nicht auf die Straße kommen‘
- d. *Se scolde ot ok [to den hilgen dagen To pinxten] [in de kerken] dragen*
(Helmstedt, Zeno, 1401–1450)
,Sie sollte es auch an den heiligen Tagen zu Pfingsten in die Kirche tragen‘

Was das Nachfeld betrifft, wurden 172 Hauptsätze mit mindestens zwei ausgeklammerten PPs gefunden. Auch diese Daten wurden manuell geprüft, um alle Belege mit einem PP-Adjunkt und einem PP-Argument im NF herauszufiltern ($n = 15$). 14/15 Sätze zeigen die gespiegelte Abfolge im NF, d. h. Argument-Adjunkt, vgl. die Beispiele in (8).

- (8) a. *dar schal ze vore ghaen [to rome] [dorch myner zele willen]*
(Stralsunder Urkunden, 1301–1500)
,da soll sie zuvor nach Rom gehen um meiner Seele willen‘
- b. *Ock schall dar nymant veghelsze mes edder ander vstedicheit bringen latben [vp dat grote market] [by iij punden]* (Greifswalder Bürgersprache, 1451)
,Auch soll da niemand Abfall, Mist oder anderen Unrat bringen lassen auf den großen Markt bei (Strafe von) drei Pfund‘

Bei einem Beispiel weicht die Abfolge der PPs im NF allerdings von diesem Muster ab, siehe (9). Es ist jedoch anzumerken, dass in diesem Beleg vier PPs im NF stehen, wobei das Argument von den Adjunkten flankiert wird, die Mehrzahl der Adjunkte jedoch, wie erwartet, dem Argument folgt (Adjunkt-Argument-Adjunkt-Adjunkt).

- (9) *Vnde *pro*¹² wart do geuoret [vp eynen witten telder] [yn dat pallas] [myt groter vroude] [myt volghende volke].* (Griseldis, 1502)
,Und [sie] wurde da geführt auf einem weißen Zelter in den Palast mit großer Freude mit nachfolgendem Volk‘

Da (9) die einzige Ausnahme ist, lässt sich schließen, dass die überwiegende Wortstellung im NF Argument-Adjunkt ist und dass daher tatsächlich ein Spiegeleffekt bezüglich der Abfolge von PP-Adjunkten und PP-Argumenten im Mnd. vorhanden ist.

Um zu untersuchen, ob im Mnd. auch ein Spiegeleffekt in Bezug auf die Abfolge von Temporal-, Modal- und Lokaladverbien (wie in (5)) belegt ist, wurden alle

¹² Siehe Abschnitt 5.

Hauptsätze herausgesucht, in denen mindestens zwei der folgenden Konstituenten entweder im MF oder im NF vorhanden sind: AdvP (ADVP), PP (PP), NP-Adjunkt (NP-TMP|NP-ADT). Auf diese Weise wurden 292 Sätze mit einem mehrfach belegten MF gefunden und 214 mit einem mehrfach belegten NF. Alle diese Sätze wurden manuell geprüft und solche mit mindestens zwei semantisch verschiedenen Adjunktklassen (Zeit, Ort, Art&Weise) im MF oder NF zurückgehalten.¹³ Für die Abfolge im MF gab es 25 relevante Belege. Die Mehrheit (22/25) entspricht dem Wortstellungsmuster in (5a), d. h. Zeit<Ort<Art&Weise; Temporaladverbien gehen Lokal- und Modaladverbien vorwiegend voran (10a), während Lokaladverbien vorwiegend Modaladverbien vorangehen (10b).

- (10) a. *vnde dat sal hey io [ses weken vore] [in allen kerken] dun kundigen*
(Soester Schrae, 1367)
,und das soll er wahrlich sechs Wochen vorher in allen Kirchen ankündigen lassen‘
- b. *So hebbe ik myn jngesegele [vor my vnde myne erue] [vor myne husfrowen vnde eren eruen] [witliken] gehenghet* (Stralsunder Urkunden, 1301–1500)
,So habe ich mein Siegel vor mir und meinem Erben, vor meiner Frau und ihren Erben wissentlich gehängt‘

Diese Wortstellungstendenzen finden sich im NF in umgekehrter Reihenfolge; die Mehrheit der relevanten Belege (19/22) entspricht dem Muster in (5b); Modaladverbien stehen vorwiegend vor Lokal- und Temporaladjunkten (11a) und Lokaladjunkte vorwiegend vor Zeitadjunkten (11b).

- (11) a. *man sal it don [in gemene hant] [iar vnde dach].*
(Braunschweiger Urkunden, 1301–1500)
,man soll es tun in gemeiner Hand Jahr und Tag‘
- b. *vnde *pro*¹⁴ is ghegebenen [in der stad to me sunde] [na gbodes bord dusent yar dre hundred yar] [yn deme achte vnde sostighbesten yare] [yn deme Daghe sunte Johannes baptisten]...* (Stralsunder Urkunden, 1301–1500)
,und [dies] ist gegeben in der Stadt bei dem Sund 1300 Jahre nach Gottes Geburt in dem 68. Jahre an dem Tag von Sankt Johannes Baptist‘

Mithilfe der syntaktisch annotierten Daten im CHLG lässt sich also zeigen, dass zwei verschiedene MF/NF-Spiegeleffekte im Mnd. belegt sind. Damit ist ein erster Schritt zur Untersuchung des diachronischen Status solcher Effekte im Westgermanischen gemacht. Wie mit der vorliegenden Vorstudie gezeigt wurde, haben solche Effekte – zumindest im Niederdeutschen – eine relativ lange Geschichte und scheinen in dieser

¹³ Pronominaladverbien wie *daran*, *darinne*, die im CHLG auf der syntaktischen Ebene auch als PP annotiert sind, wurden hier ausgeschlossen.

¹⁴ Siehe Abschnitt 5.

Hinsicht im Westgermanischen tief verwurzelt zu sein. Weitere Untersuchungen müssen ausweisen, wodurch solche Effekte zustande kommen und ob es tieferliegende strukturelle Gründe für ihre anscheinende diachrone und sprachübergreifende Stabilität gibt. Zumindest wurde gezeigt, dass Spiegeleffekte es auf jeden Fall wert sind, weiter erforscht zu werden, zumal über die theoretischen Grundlagen in Bezug auf Adjunkte, Wortstellung und syntaktische Struktur noch lange keine Einigkeit herrscht.

5 Nullsubjekte

Wie in den vorangegangenen Abschnitten bereits besprochen wurde, haben geparste Korpora den Vorteil, dass sie im Gegensatz zu Korpora, die nur auf der Wortebene annotiert sind, die vollständige Satzstruktur berücksichtigen. Das bedeutet, dass nicht nur die Wortebene, sondern auch die Beziehung oder Abhängigkeit von Konstituenten annotiert und durchsuchbar gemacht wird. Dadurch ist es möglich, einen Platz für ‚leere‘ Kategorien innerhalb von Sätzen zu schaffen, was es ermöglicht, auch elliptische Strukturen im Korpus zu finden. Das ist ein durchaus konventioneller Schritt, bspw. werden in der syntaktisch annotierten Potsdamer Baumbank des Frühneuhochdeutschen (Demske 2019) afinite Konstruktionen (Ellipse des finiten Hilfsverbs) mit einem separaten Knotenlabel (AS) markiert, da sie mit einer Volltextsuche nicht auffindbar wären.

Im CHLG werden aufgrund ähnlicher Erwägungen elliptische Kategorien mit einem Platzhalter an der Stelle des zu erwartenden lexikalischen Materials annotiert. Auf diese Art und Weise können beispielsweise referenzielle Nullsubjekte leichter gefunden werden, die mit *pro* annotiert werden, vgl. auch (9) und (11b) im vorigen Abschnitt.¹⁵

Obwohl referentielle Nullsubjekte in den germanischen Sprachen Gegenstand etlicher Studien sind (siehe bspw. Rosenkvist 2009 für eine Übersicht), wurde erst kürzlich entdeckt, dass sie auch im Mittelniederdeutschen vorkommen (Farasyn und Breitbarth 2016, 2021). Allerdings machen sie in diesen Studien, bei denen die Daten manuell annotiert wurden, nur einen sehr kleinen Teil der Gesamtzahl der

¹⁵ Darüber hinaus werden nicht overt realisierte Expletiva und Korrelate (ia) sowie reguläre Konjunktionsreduktionen (ib) (vgl. auch (7a) in Abschnitt 4) annotiert.

- (i) a. *ock *exp* ys better dat men dat crud in cleyne stuecke dar vnder lege*
(Abdinghofer Arzneibuch 1451–1500)
,[Es] ist auch besser, das Gewürz in kleinen Stücken darunter zu legen.⁶
- b. *Jtem so scalt u eme verbeden dat he nicht to vro bade vnde dat he nicht en ete to harde spyse vnde *con**
drincke neynen bosen dranck (Abdinghofer Arzneibuch 1451–1500)
,Genauso solltest du darauf achten, dass er nicht zu früh badet und dass er keine zu harten Speisen esse und [er] trinke kein schlechtes Getränk.⁶

Das Nullelement wird, der Tradition der Penn Parsed Corpora folgend, immer gleich am Satzanfang eingefügt, ungeachtet der vermutlichen tatsächlichen Position (in (ia) würde man vermutlich *auch ist es besser*... annehmen). Die Annotation *con* wird nur verwendet, wenn das ausgelassene Subjekt koreferentiell und im Numerus mit dem Subjekt des vorangehenden Konjunktus identisch ist (ib).

pronominalen Subjekte im Korpus aus. Das CHLG ermöglicht es nun, viel Zeit bei der Suche nach diesem seltenen, aber hochinteressanten linguistischen Phänomen einzusparen, indem man im Korpus nach dem Platzhalter *pro* sucht. Die Studien von Farasyn und Breitbarth (2016, 2021) haben ferner gezeigt, dass referentielle Nullsubjekte asymmetrisch über Haupt- (12a) und Nebensätze (12b) verteilt sind. Auch das kann ohne Probleme in einer Suchanfrage im CHLG formuliert werden, da als Suchdomäne IP-MAT* bzw. IP-SUB* spezifiziert werden kann.

- (12) a. *So we in my nicht ne bliuet de scal utgheworpen warden Vnde scal dorren also de winlode vnde se scolen ene to hope binden vnde werpen ene in dat vur Vnde *pro* scal vorbernen.* (Buxtehuder Evangeliar, 1451–1500)
 ‚Wer nicht in mir bleibt, der soll weggeworfen werden und verdorren wie die Rebe. Und sie sollen ihn zusammenbinden und ihn ins Feuer werfen. Und [er] soll verbrennen.‘
- b. *forder szo schole gy weten dath *pro* yuwen breff den ghy my by yuwen steffader senden entfangen hebbe* (Agneta Willeken 1535)
 ‚Außerdem solltet Ihr wissen, dass [ich] Euren Brief erhalten habe, den Ihr mir über euren Stiefvater gesendet habt‘

Eine Suche im CHLG liefert dabei ähnliche Ergebnisse wie die von Farasyn und Breitbarth (2016) beschriebenen: referentielle Nullsubjekte treten im Mnd. vorzugsweise in Hauptsätzen auf, vgl. Tabelle 4, was unterstreicht, dass die alte Asymmetrie, die bereits für das Althochdeutsche und Altsächsische charakteristisch war, im Mnd. bewahrt bleibt (vgl. Axel 2005, Walkden 2014).

	pro	Overt	Total
Hauptsatz	241 (2,2%)	10.526 (97,8%)	10.767
Nebensatz	67 (0,7%)	9.248 (99,3%)	9.315
Total	308 (1,5%)	19.774 (98,5%)	20.082

Tabelle 4: Referentielle Nullsubjekte in Haupt- und Nebensätzen im CHLG

Referentielle Nullsubjekte kommen besonders häufig in Koordinationsstrukturen vor, sowohl in Haupt- als auch Nebensätzen, ohne dass es sich dabei um Konjunktionsreduktion (*con*) handeln würde, da die Verben der koordinierten Sätze hinsichtlich der Kongruenzmerkmale nicht miteinander übereinstimmen. In (13) beispielsweise ist das Antezedens des referentiellen Nullsubjekts (3. Person Singular) im letzten Teilsatz das direkte Objekt des vorangehenden Satzes (*de iukfrowe*), nicht das Subjekt (*se*). Denn dieses, zudem in der 3. Person Plural, ist unter Koordinationsellipse ausgelassen (*con*), und ist mit dem Subjekt des ersten Teilsatzes korreferent (*se*).

- (13) *Se_i gingen benin myt groten wunder des volkes vnde *con*_i; vunden de iukefrowe; schaffen in deme huseken Vnde *pro*_j erschrak sere van deme seltzen gaste.* (Griseldis 1502)
 ‚Sie gingen unter großem Erstaunen des Volkes hinein und [sie] fanden die Jungfrau arbeiten in dem Häuschen und [sie] erschrak sehr vor dem sonderbaren Gast.‘

Obwohl Farasyn und Breitbarth (2016, 2021) anmerken, dass es im Mittelniederdeutschen grundsätzlich zwei Arten von Nullsubjekten gibt, nämlich Nulltopiks in Spez,CP (14a) und Null-Wackernagelklitika rechts von C / in Spez,TP (14b), werden diese im CHLG nicht gesondert annotiert. Zunächst einmal werden Hauptsätze in Übereinstimmung mit dem Penn-Format ohnehin nicht als CPs, sondern als IP-MATs annotiert. Nur in Sätzen mit w-Bewegung und in Nebensätzen gibt es eine CP-Ebene. Die syntaktische Annotation reflektiert also – wie auch bei der flachen Annotation des Mittelfeldes ohne VP-Knoten oder evtl. funktionale Projektionen – nicht die in der Literatur gängige Analyse der Satzstruktur, sondern ist ausschließlich dazu bestimmt, das Auffinden von Strukturen zu erleichtern, nicht dazu, eine vollständige theoretisch untermauerte Strukturanalyse zu liefern. Das gibt Nutzern des Korpus die Möglichkeit, eine eigene Analyse zu formulieren, statt sie von der Annotation aufgezwungen zu bekommen. Dennoch ist es möglich, mit einer differenzierten Suchanfrage die beiden Typen aufgrund ihrer Position in der Struktur zu unterscheiden.¹⁶ Auf diese Art und Weise lassen sich 131 (43%) Belege mit einem referentiellen Nullsubjekt in Spez,CP finden; die anderen 177 Belege befinden sich in der Position nach C (Spez,TP).

- (14) a. *Des vor wunderde sek dat her vnd *pro* bedden gerne wust wat he dar mede meynde* (Engelhus, Chronica nova, 1435)
 ‚Darüber wunderte sich das Heer und [sie] hätten gerne gewusst, was er damit meinte.‘

¹⁶ In Hauptsätzen kann man *pro* im Vorfeld / in Spez,CP mit Anfrage (i) suchen. Referenzielle Nullsubjekte in Spez,TP findet man im Hauptsatz Nebensatz mit Anfrage (ii) und im Nebensatz mit Anfrage (iii).

- | | |
|---|--|
| <p>(i) node: IP-MAT*
 query: (IP-MAT* iDoms NP-SBJ)
 AND (IP-MAT* iDoms VVFIN)
 AND (NP-SBJ dominates *pro*)
 AND (NP-SBJ precedes VVFIN)</p> | <p>(ii) node: IP-MAT*
 query: (IP-MAT* iDoms NP-SBJ)
 AND (IP-MAT* iDoms VVFIN)
 AND (NP-SBJ dominates *pro*)
 AND (VVFIN precedes NP-SBJ)</p> |
| <p>(iii) node IP-SUB*
 query: (IP-SUB* iDomsNum 1 NP-SBJ)
 AND (NP-SBJ dominates *pro*)</p> | |

- b. *Also lef beuet god de werlt ghehad dat *pro* sinen enbornen sone uor se geue vppe dat en iewelk de in eme louet nicht ne uorderue.*
 (Buxtehuder Evangeliar, 1451–1500)
 ‚So sehr hat Gott die Welt lieb gehabt, dass [er] seinen eingeborenen Sohn für sie aufgab, damit ein jeder, der an ihn glaubt, nicht zugrunde gehe.‘

Es mag eingewendet werden, dass die Annotation nicht-overten Materials vielleicht dem Anspruch des Penn-Schemas widerspricht, so wenig Analyse wie möglich in der Annotation selbst zu verankern. Da die Abfragesprache CorpusSearch den Gebrauch eines Negationsoperators erlaubt, könnte man also hypothetisch nach Sätzen ohne NP-SBJ-Zweig suchen.¹⁷ Diese Möglichkeit ergibt sich allerdings nicht. Selbst wenn sie bestünde, müsste in diesem Fall manuell zwischen referentiellen Nullsubjekten, Nullexpletiva bzw. Nullkorrelaten und Konjunktionsreduktionen unterschieden werden, was v. a. im Mnd. bei der verhältnismäßig geringen Anzahl von referentiellen Nullsubjekten und der sehr großen Anzahl von Koordinationsellipsen zu größerer Fehleranfälligkeit führen dürfte. Allerdings ist die Anwesenheit eines NP-SBJ-Knotens in jeder IP – außer in Imperativen und Infinitiven – im Penn-Schema ohnehin obligatorisch, sodass bei Abwesenheit eines overten Subjektes in jedem Fall ein kovertes Subjekt – *con*, *exp* oder eben *pro* – in der Annotation hinzugefügt werden muss.

6 Schlussbetrachtung

Im vorliegenden Beitrag sollte anhand dreier empirischer Phänomene gezeigt werden, dass die syntaktische Annotation des CHLG es erlaubt, schnell, zuverlässig und reproduzierbar Strukturen aufzufinden, die mit einem syntaktisch nicht annotierten Korpus wie dem ReN nicht ohne weiteres zu finden wären. Es ist unsere Hoffnung, dass möglichst viele Nutzer die Möglichkeiten zu schätzen wissen werden.

Danksagung: Das CHLG kam mithilfe einer großzügigen Infrastrukturförderung durch die Hercules-Stiftung (01.07.2014–31.12.2015, AUGE/13/02) und den Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (01.01.2016–31.12.2020, FWO G0F2614N) zustande. Besonderer Dank ist den Mitarbeitern des ReN-Pro-

¹⁷ Beispielsweise:

(i) node: IP*
 query: (IP* dominates ! NP-SBJ)

jekts für die Ermöglichung unserer Zusammenarbeit geschuldet, insbesondere Robert Peters und Ingrid Schröder, sowie Fabian Barteld, Katharina Dreessen, Sarah Ihden und Norbert Nagel für die praktische Umsetzung dieser Zusammenarbeit. Die Autorinnen danken weiterhin den anderen (ehemaligen) CHLG-Projektmitgliedern Aaron Ecay, Liliane Haegeman, Veronique Hoste und Mariya Koleva, sowie den externen Beratern des Projekts, Sheila Watts, George Walkden und Svetlana Petrova, für ihren jeweiligen Beitrag zu diesem Projekt.

Literaturverzeichnis

- Axel, Katrin: Null subjects and verb placement in Old High German. In: *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. Hrsg. von Stephan Kepser/Marga Reis. Berlin 2005 (Studies in Generative Grammar. 85), S. 27-48.
- Barbiers, Sjef: *The Syntax of Interpretation*. Leiden 1995. Diss. masch.
- Barteld, Fabian/Ihden, Sarah/Dreessen, Katharina/Schröder, Ingrid: HiNTS: A Tagset for Middle Low German. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC) 2018*, S. 3940-3945.
- Bech, Kristin/Eide, Kristine Gunn: *Information Structure and Syntactic Change in Germanic and Romance languages*. Amsterdam 2014 (Linguistik aktuell. 213).
- Beck, Jana/Ecay, Aaron/Ingason, Anton Karl: *Annotald*. Version 1.3.7. 2015.
- Booth, Hannah et al.: A Penn-style Treebank of Middle Low German. In: *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC) 2020*, S. 766-775. <http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.96.pdf>.
- Breitbarth, Anne: *The history of Low German negation*. Oxford 2014 (Oxford Studies in Diachronic and Historical Linguistics. 13).
- Breitbarth, Anne: Degrees of integration: Resumption after left-peripheral conditional clauses in Middle Low German. Vortrag DiGS 2018 (York) und SLE 2018 (Tallinn).
- Breitbarth, Anne: Prosodie, Syntax und Diskursfunktion von V>2 in gesprochenem Deutsch. In: *Deutsche Sprache* (2022,1), S. 1-30.
- Breitbarth, Anne: V3 after central adverbials in German: Continuity or change? In: *Journal of Historical Syntax* 7,6-19 (2023). <https://doi.org/10.18148/hs/2023.v7i6-19.138>
- Demske, Ulrike: *Referenzkorpus Frühneuhochdeutsch: Baumbank*. UP. Potsdam 2019. (<https://talar.sfb833.uni-tuebingen.de/erdora/cmd/DATEN-ZENTRUM/fnhd.UP>)

- Demske, Ulrike/Patak, Marianna: The syntax of adverbials in Early New High German. Handout, SLE 2016. Naples 2016.
- Dipper, Stefanie/Donhauser, Karin/Klein, Thomas/Linde, Sonja/Müller, Stefan/Wegera, Klaus-Peter: HiTS: ein Tagset für historische Sprachstufen des Deutschen. In: *Journal for Language Technology and Computational Linguistics* 28,1 (2013), S. 85-137.
- Donhauser, Karin/Petrova, Svetlana: Syntaktische Variation in Nachsatz-Konstruktionen des Mittelhochdeutschen. Sprachliche Evidenzen zur überlieferungsgeschichtlichen Diskontinuität im mhd. Prosa-Lancelot. In: *Germanistik zwischen Baum und Borke. Festschrift für Kari Keinistö zum 60. Geburtstag*. Hrsg. von Doris Wagner/Tuomo Fonsen/Henrik Nikula. Helsinki 2009 (*Mémoires de la Société Néophilologique de Helsinki*. 76), S. 289-303.
- Dreessen, Katharina/Ihden, Sarah: Korpuslinguistische Studien zur mittelniederdeutschen Syntax. In: *Jahrbuch für Germanistische Sprachgeschichte* 6,1 (2015), S. 249-275.
- Farasyn, Melissa/Breitbarth, Anne: Nullsubjekte im Mittelniederdeutschen. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur* 138,4 (2016), S. 524-559.
- Farasyn, Melissa/Breitbarth, Anne: Null subjects in Middle Low German: diachronic stability and change. In: *Syntactic features and the limits of syntactic change*. Hrsg. von Jóhannes Gísli Jónsson/Thórhallur Eythórsson. Oxford 2021 (*Oxford Studies in Diachronic and Historical Linguistics*. 43), S. 84-110.
- Farasyn, Melissa: Fitting in or standing out? Subject agreement phenomena in Middle Low German. Ghent 2018.
- Frey, Werner: Syntactic conditions on adjunct classes. In: *Modifying adjuncts*. Hrsg. von Ewald Lang/Claudia Maienborn/Catherine Fabricius-Hansen. Berlin 2003 (*Interface Explorations*. 4), S. 163-209.
- Haider, Hubert: Adverb Placement - Convergence of Structure and Licensing. In: *Theoretical Linguistics* 26,1-2 (2000), S. 95-134.
- Haug, Dag: Treebanks in historical linguistic research. In: *Perspectives on Historical Syntax*. Hrsg. von Carlotta Viti. Amsterdam 2015 (*Studies in Language. Companion Series*. 169), S. 185-202.
- Haug, Dag/Jøhndal, Marius: Creating a parallel treebank of the old Indo-European Bible translations. In: *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, S. 27-34.
- Hinterhölzl, Roland/Petrova, Svetlana: Information Structure and Language Change: New Approaches to Word Order Variation in Germanic. Berlin 2009 (*Trends in Linguistics*. 203).
- Ihden, Sarah: Relativsätze im Mittelniederdeutschen. Korpuslinguistische Untersuchungen zu Struktur und Gebrauch. Berlin/Boston 2020 (*Lingua Historica Germanica*. 23).

- Koster, Jan: Het werkwoord als spiegelcentrum. In: *Spektator* 3 (1974), S. 601-618.
- Koster, Jan: Mirror symmetry in Dutch. In: *Grammar in Progress. Articles Presented at the 20th Anniversary of the Comparison of Grammatical Models Group in Tilburg*. Hrsg. von Marc van Oostendorp/Elena Anagnostopoulou. Amsterdam 2001.
- Mähl, Stefan: Studien zum mittelniederdeutschen Adverb. Köln/Weimar/Wien 2004 (Niederdeutsche Studien. 49).
- Mähl, Stefan: Mehrgliedrige Verbalkomplexe im Mittelniederdeutschen: ein Beitrag zu einer historischen Syntax des Deutschen. Köln/Weimar/Wien 2014 (Niederdeutsche Studien. 57).
- Merten, Marie-Luis: Literater Sprachausbau kognitiv-funktional. Funktionswort-Konstruktionen in der historischen Rechtsschriftlichkeit. Berlin/Boston 2018 (Germanistische Linguistik. 311).
- Nivre, Joakim et al.: Universal Dependencies v1: A Multilingual Treebank Collection. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, European Language Resources Association (ELRA). Portoroz 2016, S. 1659-1666. <https://www.aclweb.org/anthology/L16-1262>.
- Peters, Robert: Mittelniederdeutsche Sprache. In: *Niederdeutsch – Sprache und Literatur*, Band 1: Sprache. Hrsg. von Jan Goossens. Neumünster 1973, S. 66-115.
- Petrova, Svetlana: Multiple XP-fronting in Middle Low German root clauses. In: *The Journal of Comparative Germanic Linguistics* 15,2 (2012), S. 157-188.
- Petrova, Svetlana: *The Syntax of Middle Low German*. Berlin 2013. Habil. masch.
- Pintzuk, Susan/Taylor, Ann/Warner, Anthony: Corpora and Quantitative Methods. In: *The Cambridge Handbook of Historical Syntax*. Hrsg. von Adam Ledgeway/Ian Roberts. Cambridge 2017, S. 218-240.
- Pittner, Karin: Where syntax and semantics meet: adverbial positions in the German middle field. In: *Adverbials: the interplay between meaning, context and syntactic structure*. Hrsg. von Jennifer R. Austin/Stefan Engelberg/Gisa Rauh. Amsterdam/Philadelphia 2004 (Linguistik Aktuell. 70), S. 253-287.
- Randall, Beth: *CorpusSearch2 User's Guide*. Philadelphia 2005. <http://corpussearch.sourceforge.net>.
- ReN-Team: Reference Corpus Middle Low German/Low Rhenish (1200–1650); Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650) (Version 1.1) [Data set]. 2021. <http://doi.org/10.25592/uhhfdm.9195>
- Rosenkvist, Henrik: Referential Null Subjects in Germanic Languages – an Overview. In: *Working Papers in Scandinavian Syntax* 84 (2009), S. 151-180.

- Rösler, Irmtraud: Satz, Text, Sprachhandeln: Syntaktische Normen der mittelniederdeutschen Sprache und ihre soziofunktionalen Determinanten. Heidelberg 1997 (Sprachgeschichte. 5).
- Santorini, Beatrice: Part-of-speech tagging guidelines for the Penn Treebank Project (3rd revision). Technical Report, University of Pennsylvania 1990.
- Santorini, Beatrice: Annotation manual for the Penn Historical Corpora and the PCEEC. Department of Linguistics, University of Pennsylvania. 2010. <https://www.ling.upenn.edu/hist-corpora/annotation/index.html>.
- Schalowski, Sören: Wortstellungsvariation aus informationsstruktureller Perspektive. Eine Untersuchung der linken Satzperipherie im gesprochenen Deutsch. In: Working Papers of the SFB 632. Interdisciplinary Studies on Information Structure (ISIS), vol. 18. Potsdam (2015). <https://publishup.uni-potsdam.de/frontdoor/index/index/docId/8008>.
- Schiller, Anne et al.: Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset). Universität Stuttgart 1999.
- Sievers, Eduard (Hrsg.): Heliand. Halle 1878 (Germanistische Handbibliothek. 4).
- Taylor, Ann: Treebanks in Historical Syntax. In: Annual Review of Linguistics 6,1 (2020). 195212. <https://doi.org/10.1146/annurev-linguistics-011619-030515>.
- Tophinke, Doris: Vom Vorlesetext zum Lesetext: Zur Syntax mittelniederdeutscher Rechtsverordnungen im Spätmittelalter. In: Oberfläche und Performanz. Untersuchungen zur Sprache als dynamischer Gestalt. Hrsg. von Angelika Linke/Helmuth Feilke. Tübingen 2009 (Germanistische Linguistik. 283), S. 161-183.
- Walkden, George: Syntactic Reconstruction and Proto-Germanic. Oxford 2014 (Oxford Studies in Diachronic and Historical Linguistics 12).
- Walkden, George: HeliPaD: the Heliand Parsed Database. Version 0.9. 2015. <https://doi.org/10.5281/zenodo.4395040>.
- Walkden, George. 2016. The HeliPaD. A parsed corpus of Old Saxon. In: International Journal of Corpus Linguistics 21,4 (2016), S. 559-571. <https://doi.org/10.1075/ijcl.21.4.05wal>.
- Wallmeier, Nadine: Sprachliche Muster in der mittelniederdeutschen Rechtssprache. Zum Sachsenspiegel und zu Stadtrechtsaufzeichnungen des 13. bis 16. Jahrhunderts. Köln/Weimar/Wien 2013 (Niederdeutsche Studien. 55).
- Wallmeier, Nadine: Konditionale Adverbialsätze und konkurrierende Konstruktionen in mittelniederdeutschen Rechtstexten. Niederdeutsches Jahrbuch 138 (2015), S. 7-26.
- Wiese, Heike/Müller, Hans G.: The hidden life of V3. An overlooked word order variant on verb-second. In: Non-canonical verb positioning in main clauses. Hrsg. von Mailin Antomo/Sonja Müller. Linguistische Berichte, Sonderheft 25 (2018), S. 201-223.

- Wiese, Heike/Öncü, Mehmet Tahir/Bracker, Philipp: Verb-dritt-Stellung im türkischdeutschen Sprachkontakt. Informationsstrukturelle Linearisierungen ein- und mehrsprachiger Sprecher/innen. In: *Deutsche Sprache* (2017,1), S. 31-52.
- Wiese, Heike et al.: Verb third in spoken German: A natural order of information? In: *Rethinking Verb Second*. Hrsg. von Rebecca Woods/Sam Wolfe. Oxford 2020, S. 682-699.