

Korpusdokumentation des SiGS-Korpus der Hexenverhörprotokolle

Korpusdokumentation des DFG-Projekts

„Entwicklung der satzinternen Großschreibung im Deutschen.

Eine korpuslinguistische Studie zum Zusammenspiel

kognitiv-semantischer und syntaktischer Faktoren“

(Projektnummer 228096696)

Renata Szczepaniak, Fabian Barteld, Lisa Dücker, Stefan Hartmann, Klaus-Michael Köpcke, Marc
Schutzeichel

Leipzig, Juni 2023

SiGS – Korpusdokumentation

Inhalt

| | | |
|---------|--|----|
| 1 | Einleitung | 6 |
| 2 | Editorische Grundsätze | 7 |
| 2.1 | Streichungen | 7 |
| 2.2 | Nachträge | 7 |
| 2.3 | Abkürzungen | 7 |
| 2.4 | Unsichere Lesungen und Schreibversehen | 7 |
| 2.5 | Verschiedene Schriftarten | 8 |
| 2.6 | Randsatz | 8 |
| 3 | Tokenisierung | 8 |
| 3.1 | Das graphische Token | 9 |
| 3.2 | Das syntaktische Token | 9 |
| 3.2.1 | Regeln für die syntaktische Tokenisierung | 10 |
| 4 | Tokenbasierte Annotation | 14 |
| 4.1 | Großschreibung | 14 |
| 4.2 | PoS | 15 |
| 4.2.1 | Einführung | 15 |
| 4.2.2 | Zusätzliche Tags | 15 |
| 4.2.3 | Von STTS abweichende Regeln für die Verwendung von STTS-Tags | 21 |
| 4.2.4 | Umgang mit spezifischen sprachlichen Phänomenen | 23 |
| 4.2.5 | Liste der PoS-Tags | 24 |
| 4.3 | Belebtheit | 27 |
| 4.3.1 | Belebtheitskategorien | 27 |
| 4.3.1.2 | <i>Konkreta</i> | 28 |
| 4.3.1.3 | <i>Tierische Entitäten</i> | 29 |
| 4.3.2 | Annotation von Metaphern und Metonymien | 30 |
| 4.3.3 | Belebtheitsannotation von Komposita | 31 |
| 4.4 | Lemmatisierung | 31 |
| 4.4.1 | Einführung | 31 |
| 4.4.2 | Meta-Lemmata | 32 |
| 4.4.3 | Unterschiede zwischen syntaktischen und graphischen Tokens | 34 |

| | | |
|-------|--|----|
| 4.4.4 | Überblick: Grundformen der häufigsten Pronomina..... | 35 |
| 5 | Annotation der Satzgrenzen | 36 |
| 5.1 | Grundlagen | 36 |
| 5.2 | Flussdiagramm zum Vorgehen bei der Satzgrenzenannotation | 37 |
| 5.3 | Nebensätze und Hauptsätze | 37 |
| 5.3.1 | Relativsätze | 38 |
| 5.3.2 | Komplementsätze mit Verbzweitstellung | 39 |
| 5.3.3 | Eigenständige Nebensätze | 40 |
| 5.4 | Einschübe..... | 40 |
| 5.5 | Textuelle Gliederungszeichen..... | 41 |
| 5.5.1 | Feststellen der syntaktischen Integration | 41 |
| 5.5.2 | Freistehende Wortfolgen | 43 |
| 5.6 | Ellipsen..... | 43 |
| 5.7 | Ausfall von finiten Hilfs- und Modalverben | 44 |
| 5.8 | Abschnitte mit syntaktisch unklarer Struktur | 46 |
| 6 | Phrasenannotation..... | 48 |
| 6.1 | Präpositionalphrasen (PP) | 48 |
| 6.1.1 | Präpositionalphrasen ohne eingebettete Nominalphrasen | 48 |
| 6.1.2 | Präpositionalphrasen, die nicht annotiert werden..... | 50 |
| 6.1.3 | Namen mit Präpositionalphrasen..... | 50 |
| 6.1.4 | Koordination von Präpositionen..... | 50 |
| 6.2 | Nominalphrasen (NP) | 51 |
| 6.2.1 | Koordination und Apposition von Nominalphrasen innerhalb einer Präpositionalphrase 51 | |
| 6.2.2 | Relativa, „so“ als Relativpronomen | 51 |
| 6.2.3 | Fragepronomina (PWAV)..... | 51 |
| 6.2.4 | Formen mit „mal“ | 52 |
| 6.2.5 | Komplexe Relativpronomen („was für“)..... | 52 |
| 6.3 | Umgang mit lateinischem Wortmaterial | 52 |
| 6.3.1 | Eigennamen..... | 53 |
| 6.3.2 | Deutsche PP mit lateinischem Wortmaterial | 53 |
| 6.3.3 | Lateinische PP mit deutschem Wortmaterial | 53 |
| 6.3.4 | Deutsche Phrasen in lateinischen Sätzen (NP_lat und PP_lat)..... | 54 |
| 6.4 | Ziffern als Phrasenkopf | 54 |

| | | |
|--------|--|----|
| 7 | Bestimmung der Phrasengrenzen | 54 |
| 7.1 | Umgang mit Attributen..... | 54 |
| 7.1.1 | Präpositionalphrasen als Attribute..... | 54 |
| 7.1.2 | Ziffern als Attribute | 55 |
| 7.1.3 | Komplexe Zeitangaben (<i>gestern Abend</i>) | 55 |
| 7.2 | Quantifizierer | 55 |
| 7.3 | Koordination von Satzgliedern..... | 56 |
| 7.4 | Koordinationsellipse | 56 |
| 7.5 | Appositionen..... | 56 |
| 7.5.1 | Apposition bei Mengenangaben | 57 |
| 7.5.2 | Enge Appositionen mit geographischen Namen | 57 |
| 7.5.3 | Enge Apposition „mit Namen“ | 57 |
| 7.5.4 | Kombination aus Anrede und Titel..... | 57 |
| 7.6 | Komposita | 57 |
| 7.7 | Eigennamen | 58 |
| 8 | Phrasentyp (phrasentyp)..... | 59 |
| 8.1 | Phrasen im Satz..... | 59 |
| 8.2 | Nicht in einen Satz eingebettete Phrasen..... | 60 |
| 8.2.1 | Textuelle Gliederungszeichen/Absatzgliederung: | 60 |
| 8.2.2 | Listen | 60 |
| 8.2.3 | Randsatz | 60 |
| 8.2.4 | Anrede | 61 |
| 8.2.5 | Vergleichsphrasen | 61 |
| 9 | Einbettung von Phrasen (embedded) | 62 |
| 9.1 | embedded=false | 62 |
| 9.2 | embedded=true-attr..... | 62 |
| 9.3 | embedded=true-head..... | 63 |
| 9.4 | embedded=true-no_attr..... | 63 |
| 10 | Satzgliedfunktion (synfunc)..... | 63 |
| 10.1 | Adverbiale Bestimmung (adv)..... | 64 |
| 10.2 | potentielle Präpositionalobjekte (<i>pot_PPO</i>)..... | 64 |
| 10.3 | Synfunc=NA | 65 |
| 10.3.1 | Prädikatsnomen | 65 |

| | | |
|--------|--|----|
| 10.3.2 | Interrogativpronomina..... | 65 |
| 10.3.3 | Vergleichsphrasen..... | 66 |
| 10.3.4 | Annotation von „sich“..... | 66 |
| 10.4 | Elliptische NPs..... | 66 |
| 10.5 | Es im Vorfeld..... | 66 |
| 11 | Semantische Rolle (<i>sr</i>)..... | 67 |
| 11.1 | Proto-Agens..... | 67 |
| 11.1.1 | Agens in Aktivsätzen (<i>agens</i>)..... | 67 |
| 11.1.2 | Agens in Passivsätzen..... | 68 |
| 11.1.3 | Handlungskontrolle..... | 68 |
| 11.2 | Proto-Patiens (<i>patiens</i>)..... | 68 |
| 11.3 | Experiencer (<i>experiencer</i>) und Stimulus (<i>stimulus</i>)..... | 69 |
| 11.4 | Stativ (<i>stativ</i>)..... | 69 |
| 11.5 | mehrfache Rollenzuweisung bei einem Substantiv (<i>multi</i>)..... | 70 |
| 11.6 | unklare Rollenzuweisung (<i>ambig</i>)..... | 70 |
| 11.7 | keine Rollenzuweisung (<i>keine_rolle</i>)..... | 70 |
| 12 | Literaturverzeichnis..... | 71 |

1 Einleitung

Die in der hier vorliegenden Korpusdokumentation festgehaltenen Annotationsrichtlinien¹ wurden im DFG-geförderten Projekt „Entwicklung der satzinternen Großschreibung im Deutschen. Eine korpuslinguistische Studie zum Zusammenspiel kognitiv-semantischer und syntaktischer Faktoren“ (Projektnummer 228096696) für die Annotation von frühneuhochdeutschen Hexenverhörprotokollen entwickelt und verwendet. Annotiert wurden Texte aus der auf Handschriften basierenden Edition von Macha et al. (2005), die aus dem Projekt „Deutsche Kanzleisprache in Hexenverhörprotokollen der Frühen Neuzeit“ hervorgegangen ist.

Die dem Editionsprojekt entstammenden Textdateien der transkribierten und edierten Protokolle wurden zunächst durch ein Skript in ein TEI-kompatibles Format² konvertiert und in einem zweiten Schritt durch das Programm CAB³ automatisch vorannotiert. Diese vorannotierten Dateien wurden automatisiert vereinfacht und für die manuelle linguistische Mehr-Ebenen-Annotation vorbereitet.

Die linguistische Annotation wurde in einem nächsten Schritt nach den in diesem Handbuch dargestellten Richtlinien in dem Programm GATE⁴ manuell korrigiert und erweitert. Für die Umsetzung der Korrektur wurde das im Projekt entwickelte GATE-Plugin „Vertical Annotation Editor“⁵ verwendet, das eine Bearbeitung der Annotationen in Tabellenform erlaubt. Die genaue technische Umsetzung ist für die einzelnen Annotationsschritte jeweils in einem eigenen Abschnitt erläutert.

Für den Export der Daten wurde ein Skript entwickelt, das die Daten von GATE in das Format PAULA XML exportiert.⁶ Mit Hilfe des Tools PEPPER⁷ (Zipser & Romary 2010) wurden die Daten ins ANNIS-Format überführt und können so in das Visualisierungs- und Analysetool ANNIS⁸ eingelesen werden (Krause & Zeldes 2016). Das Korpus ist unter dem Namen "Hexenverhörprotokolle, Version 1.0" im paula- und relannis-Format im Laudatio-Repository⁹ veröffentlicht (<https://doi.org/10.34644/laudatio-dev-2TkBi4gBUnBk-z0fwp-w>).

Die Richtlinien selbst orientieren sich soweit wie möglich an zum Zeitpunkt der Erstellung existierenden Standards, die an die Spezifika der Hexenverhörprotokolle angepasst wurden. Zu diesen Spezifika gehören insbesondere strukturelle Mehrdeutigkeiten (vgl. Pauly et al. 2012). Diese konnten durch eine doppelte Tokenisierung in vielen Fällen aufgelöst werden (vgl. Barteld et al. 2014, Näheres s. Kap. Tokenisierung). Das PoS-Tagset basiert auf STTS (Schiller et al. 1999). HiTS (Dipper et al. 2013), eine explizit für ältere Sprachstufen des Deutschen entwickelte Variante von STTS, lag zu Beginn der Arbeit noch nicht vor. Bei späteren Anpassungen wurden aber Lösungen aus HiTS mit berücksichtigt.

¹ Wir möchten an dieser Stelle Kristina Albert, Annemarie Bischoff, Lisa Dücker, Julia Hübner, Aleksa Krieg, Johanna Legrum, Sophie Mühlberg, Anna Müller, Merle Pfau, Katja Politt, Nikolai Pudimat, Eleonore Schmitt, Aki Schünemann, Tanja Stevanović, Annika Vieregge, Fiona Walter und Nicholas Wieling für ihren Einsatz bei der Annotation der Texte danken.

² Text encoding initiative, <http://www.tei-c.org/Guidelines/P5/>

³ Cascaded Analysis Broker, <http://www.deutschestextarchiv.de/doku/software#cab>

⁴ General Architecture for Text engineering, <https://gate.ac.uk/>

⁵ https://github.com/fab-bar/GATE_Plugin-Vertical_Annotation_Editor

⁶ https://github.com/fab-bar/GATE-PAULA_Exporter

⁷ <http://corpus-tools.org/pepper/>

⁸ <https://corpus-tools.org/annis/>

⁹ <https://www.laudatio-repository.org>

Die Entwicklung der Richtlinien wurde nach dem MAMA-Schema (Pustejovsky & Stubbs 2012: 27f.) durchgeführt. Erste Texte wurden jeweils von zwei AnnotatorInnen unabhängig voneinander annotiert. Nach einem Vergleich der Abweichungen wurden die Richtlinien überarbeitet. Basierend auf den überarbeiteten Richtlinien wurden dann andere Texte von den AnnotatorInnen annotiert, bis die Übereinstimmung der Annotationen einen substantiellen Kappa-Wert (vgl. Artstein & Poesio 2008: 576) erreicht hat. Danach wurden weitere Texte je zweimal unabhängig annotiert. D.h. jede Annotation in jedem Text wurde von mindestens zwei AnnotatorInnen angefertigt.

2 Editorische Grundsätze

Die dem Korpus zugrundeliegenden Texte wurden ohne Veränderungen am Text aus der Edition von Macha et al. (2005) übernommen, d.h. dass die Editionsgrundsätze des SiGS-Projekts mit denen der zugrundeliegenden Edition übereinstimmen (vgl. Macha et al. 2005: XIII–XXVI).

Ein paar Unterschiede ergeben sich jedoch im Umgang mit für Handschriften typischen Besonderheiten:

2.1 Streichungen

Streichungen werden bei Macha et al. (2005) durch einen vorangestellten vertikalen Strich und ein nachgestelltes Pluszeichen gekennzeichnet (Bsp.1). Für den Umgang mit diesen Streichungen im SiGS-Korpus s. Kap.

Streichungen.

Bsp. 1, Oldesloe 1639¹⁰:

Aff[irrat] der Sögen das |beyn+
[INT]¹bein¹ entzwey geschlag[en]

2.2 Nachträge

Nachträglich hinzugefügte Wörter oder Textpassagen, die bei Macha et al. (2005) von Winkeln eingerahmt erscheinen (Bsp.1), können im SiGS-Korpus über die Annotationsebene „add“ gefunden werden. Die Annotation „add_place“ hat zwei Merkmalsausprägungen: „margin“ für Nachträge am Seitenrand und „above“ für interlineare Nachträge.

2.3 Abkürzungen

Abkürzungen werden bei Macha et al. (2005) aufgelöst und die ergänzten Wortteile in eckige Klammern gesetzt (Bsp.1). Im SiGS-Korpus werden die aufgelösten Wortformen gemeinsam mit dem übrigen Text als Grundlage für die Annotation verwendet. Das ergänzte Wortmaterial kann in ANNIS über die Annotationsebene *supplied* aufgerufen werden.

2.4 Unsichere Lesungen und Schreibversehen

Versehen in der Vorlage sind bei Macha et al. (2005) durch ein nachgestelltes [!] markiert. Diese Markierungen sind im SiGS-Korpus nicht enthalten. Auslassungen und unleserliche Stellen werden als

¹⁰ Auf die annotierten Texte wird mit einer Sigle aus Ortsname und Jahreszahl, wie sie in der Edition von Macha et al. (2005) angegeben sind, verwiesen.

... wiedergegeben. Unsichere Lesarten können über die Annotationsebene *unsicher* und die dort angesiedelte Annotation *Lesart=unsicher* abgerufen werden. Bei Macha et al. (2005) sind solche Lesarten durch nachgestelltes [?] gekennzeichnet.

2.5 Verschiedene Schriftarten

Die verschiedenen in den Handschriften verwendeten Schriftarten werden im SiGS-Korpus nicht voneinander unterschieden.

2.6 Randsatz

In der Edition von Macha et al. (2005) werden Einträge am linken Seitenrand in einer vom Haupttext getrennten Spalte abgebildet. Diese Marginalspalte ist im SiGS-Korpus nicht als solche ausgewiesen und kann nicht ausgeblendet werden. Einträge in dieser Spalte durchbrechen somit die lineare Darstellung von Sätzen in der Hauptspalte. Durch Aufklappen der Satzebene in ANNIS können Marginalia schnell als solche erkannt werden: Jeder Satz und jeder Eintrag in der Marginalspalte erhält eine einzigartige ID (s. Kap. Annotation der Satzgrenzen). Diese ist bei den einzelnen Teilen von diskontinuierlichen Sätzen identisch. Einschübe zwischen zwei Satzteilen mit identischer ID weisen auf einen Eintrag in der Marginalspalte hin.

3 Tokenisierung

Für die im Projekt „Entwicklung der satzinternen Großschreibung im Deutschen“ vorliegende Fragestellung ist es notwendig, zwei Arten von Tokens zu unterscheiden, die in den Protokollen nicht immer übereinstimmen: das graphische und das syntaktische Token (vgl. Barteld et al. 2014). Somit werden bei der Tokenisierung zwei Ebenen angelegt, die sich jeweils am graphischen oder syntaktischen Wort orientieren, und beide Tokenebenen erhalten in den darauffolgenden tokenbasierten Annotationsschritten getrennt ihre Annotation (jeweils PoS, Lemma und Belebtheit). In ANNIS sind die beiden Ebenen unter dem Namen *graphtoken* (graphische Token) und *syntoken* (syntaktische Token) zu finden.

Bsp. 2, Alme 1630:

| Text | <i>Auffm Teufelß dantz</i> | | |
|-----------------------|----------------------------|----------------|---------------------|
| graphtoken | <i>Auffm</i> | <i>Teufelß</i> | <i>dantz</i> |
| graphtoken_pos | <i>APPRART</i> | <i>NN</i> | <i>NN</i> |
| graphtoken_lemma | - | <i>teufel</i> | <i>tanz</i> |
| graphtoken_belebtheit | - | <i>belebt</i> | <i>unbelebt</i> |
| syntoken | <i>Auff</i> | <i>m</i> | <i>Teufelßdantz</i> |
| syntoken_pos | <i>APPR</i> | <i>ART</i> | <i>NN</i> |
| syntoken_lemma | <i>auf</i> | <i>D</i> | <i>teufelstanz</i> |
| syntoken_belebtheit | - | - | <i>unbelebt</i> |

3.1 Das graphische Token

Das graphische Token ist definiert als Folge von Graphen, die nicht durch Spatien oder Zeilenumbrüche¹¹ voneinander getrennt sind. Eine Besonderheit bilden die Interpunktionszeichen:

,;.:()[]!/?

Diese werden abgetrennt und als eigene graphische Tokens behandelt. Bei Worttrennungen am Zeilenende werden zwei graphische Tokens angesetzt. Mögliche Trennstriche, die die Worttrennung anzeigen, werden ins graphische Token integriert.

Bsp. 3, Georgenthal 1597:

| | | | | | | | | | | | |
|-------------------|--|---------------------|----------|------------|------------|------------|-------------|----------------|--------------|----------|--|
| Text | <i>Sie geantworttet, das man sie ver-↓brennen sollt,</i> | | | | | | | | | | |
| graphtoken | <i>sie</i> | <i>geantworttet</i> | <i>,</i> | <i>das</i> | <i>man</i> | <i>sie</i> | <i>ver-</i> | <i>brennen</i> | <i>sollt</i> | <i>,</i> | |

3.2 Das syntaktische Token

Das Ziel der syntaktischen Tokenisierung ist es, die kleinsten Einheiten für die syntaktische Analyse zur Verfügung zu stellen. Das syntaktische Token kann sich auf zwei Arten vom graphischen Token unterscheiden:

- Getrennschreibung eines syntaktischen Tokens
- Zusammenschreibung mehrerer syntaktischer Tokens

Die graphische Trennung eines syntaktischen Tokens kann z.B. durch das Vorliegen eines Zeilenendes bedingt sein (Bsp. 4).

Bsp. 4, Georgenthal 1597:

| | | | | | | | | | | | |
|-------------------|--|---------------------|----------|------------|------------|------------|-------------------|----------------|--------------|----------|--|
| Text | <i>Sie geantworttet, das man sie ver-↓brennen sollt,</i> | | | | | | | | | | |
| graphtoken | <i>sie</i> | <i>geantworttet</i> | <i>,</i> | <i>das</i> | <i>man</i> | <i>sie</i> | <i>ver-</i> | <i>brennen</i> | <i>sollt</i> | <i>,</i> | |
| syntoken | <i>sie</i> | <i>geantworttet</i> | <i>,</i> | <i>das</i> | <i>man</i> | <i>sie</i> | <i>verbrennen</i> | | <i>sollt</i> | <i>,</i> | |

Die Zusammenschreibung mehrerer syntaktischer Tokens liegt u.a. bei Klisen (*im, aufm, vom, zur* usw.) vor. Hier wird auf der Ebene des syntaktischen Tokens konsequent getrennt, unabhängig von der Klisenart.

Bsp. 5, Georgenthal 1597:

| | | | |
|-------------------|----------------|----------|-------------|
| Text | <i>Im hauß</i> | | |
| graphtoken | <i>Im</i> | | <i>hauß</i> |
| syntoken | <i>I</i> | <i>m</i> | <i>hauß</i> |

¹¹ Zeilenumbrüche werden in diesem Handbuch innerhalb von Tabellen mit ↓ markiert, Seitenwechsel mit ↓←.

3.2.1 Regeln für die syntaktische Tokenisierung

In den meisten Fällen sind die Grenzen von graphischem und syntaktischem Token identisch. Im Folgenden sind die Fälle geregelt, in denen die syntaktische Tokenisierung von der graphischen abweicht.

3.2.1.1 Komposita

Die Getrenntschreibung eines syntaktischen Tokens liegt häufig bei Komposita vor (vgl. Bsp. 6).

Bsp. 6, Georgenthal 1597:

| | | | | | |
|-------------------|----------------------------------|--------------|----------------------|---------------|----------|
| Text | <i>sey keine Millich diebin,</i> | | | | |
| graphtoken | <i>sey</i> | <i>keine</i> | <i>Millich</i> | <i>diebin</i> | <i>,</i> |
| syntoken | <i>sey</i> | <i>keine</i> | <i>Millichdiebin</i> | | <i>,</i> |

Eindeutige getrennt geschriebene Komposita werden auf die in Bsp. 6 demonstrierte Weise annotiert. Eindeutige Komposita liegen vor, wenn der vorausgehende Determinierer bzw. das vorangestellte Attribut ausschließlich mit dem Kopf des Kompositums kongruiert (wie in *der Gerichts Leute*). Strukturen, die keine Begleiter haben (*daß were weiber werck*, Ahaus 1608) bzw. bei denen die Begleiter mit beiden Gliedern des Kompositums kongruieren (*der Jesuiter kirchen*, Augsburg 1625) gelten als ambig. Sie werden ebenso wie die eindeutigen Komposita als ein syntaktisches Token annotiert und erhalten zusätzlich die Annotation *ambig=true* (vgl. Bsp.7).

Bsp. 7, Georgenthal 1597:

| | | | |
|-------------------|-----------------------------|------------------------|----------------|
| Text | <i>deß Gerichts knechts</i> | | |
| graphtoken | <i>deß</i> | <i>Gerichts</i> | <i>knechts</i> |
| syntoken | <i>deß</i> | <i>Gerichtsknechts</i> | |
| ambig | FALSE | TRUE | |

3.2.1.2 Partikelverben

Bei Partikelverben lassen sich bezüglich der Tokenisierung drei Realisierungen unterscheiden:

1. Bei Zusammenschreibung (Bsp. 8) wird sowohl graphisch als auch syntaktisch nur ein Token, ein Partikelverb, angenommen.

Bsp. 8, konstruiert, [+ graphisches Token] und [+ syntaktisches Token]:

| | | | |
|-------------------|-------------------------|-----------|-----------------|
| Text | <i>dass er mitsingt</i> | | |
| graphtoken | <i>dass</i> | <i>er</i> | <i>mitsingt</i> |
| syntoken | <i>dass</i> | <i>er</i> | <i>mitsingt</i> |

2. Als zwei graphische Tokens und ein syntaktisches Token wird die Kombination Verbpartikel¹² + Verb (Bsp. 9) analysiert, wenn sie in dieser Reihenfolge stehen und durch keine anderen graphischen Tokens unterbrochen werden.

¹² Vergleiche den Abschnitt Verbpartikel (PTKVZ) in den Richtlinien zur PoS-Annotation zur Abgrenzung von Verbpartikeln und anderen Wortarten.

Bsp. 9, konstruiert, [- graphisches Token] und [+ syntaktisches Token]:

| | | | | |
|-------------------|--------------------------|-----------|-----------------|--------------|
| Text | <i>dass er mit singt</i> | | | |
| graphtoken | <i>dass</i> | <i>er</i> | <i>mit</i> | <i>singt</i> |
| syntoken | <i>dass</i> | <i>er</i> | <i>mitsingt</i> | |

3. In allen anderen Fällen (z.B. Bsp. 10) liegen zwei graphische und syntaktische Tokens vor.

Bsp. 10, konstruiert, [- graphisches Token] und [- syntaktisches Token]:

| | | | |
|-------------------|---------------------|--------------|------------|
| Text | <i>Er singt mit</i> | | |
| graphtoken | <i>Er</i> | <i>singt</i> | <i>mit</i> |
| syntoken | <i>Er</i> | <i>singt</i> | <i>mit</i> |

3.2.1.3 zu als Infinitivmarker

Zu kann sowohl als Verbpartikel als auch als Infinitivmarker fungieren (zur Unterscheidung vgl. PoS-Regeln unter 4.2.4.3 Verbpartikeln). Als Infinitivmarker wird es als ein eigenes syntaktisches Token aufgefasst, auch bei Zusammenschreibung.

Bsp. 11, Alme 1630:

*Bleibt bei seiner Vrgicht bestendigh mit erpie
ten daruff zuleben Vnd zusterben Vnd eß
Allen nom[in]atis ins angesicht zu sagen*

| | | | | | | |
|-------------------|-------------------------------------|----------------|--------------|------------------|-----------|----------------|
| Text | <i>daruff zuleben Vnd zusterben</i> | | | | | |
| graphtoken | <i>daruff</i> | <i>zuleben</i> | <i>Vnd</i> | <i>zusterben</i> | | |
| syntoken | <i>daruff</i> | <i>zu</i> | <i>leben</i> | <i>Vnd</i> | <i>zu</i> | <i>sterben</i> |

3.2.1.4 Pronominaladverbien

Pronominaladverbien werden als ein syntaktisches Token analysiert, auch bei Getrenntschreibung. Als Pronominaladverbien werden nur *hier-*, *da(r)-*, *wo(r)-* + Präposition annotiert.

Bsp. 12, konstruiert:

| | | | | | | |
|-------------------|------------------------------|-------------|--------------|------------|---------------|--|
| Text | <i>Er habe da von gehört</i> | | | | | |
| graphtoken | <i>Er</i> | <i>habe</i> | <i>da</i> | <i>von</i> | <i>gehört</i> | |
| syntoken | <i>Er</i> | <i>habe</i> | <i>davon</i> | | <i>gehört</i> | |

3.2.1.5 Worttrennung am Zeilenende

Bei Worttrennungen am Zeilen- oder Seitenende werden zwei graphische Tokens, aber nur ein syntaktisches Token annotiert (vgl. oben Bsp. 3 und 4).

3.2.1.6 Klisen

Klisen werden auf der Ebene der syntaktischen Tokens aufgelöst. Es handelt sich meist um Präposition-Artikel- und Pronomen-Pronomen-Klisen. Die Tokenisierung wird so durchgeführt, dass das erste syntaktische Wort möglichst vollständig bleibt:

Bsp. 13, Flensburg 1608:

| | | | |
|-------------------|------------------|--------------|--------------|
| Text | <i>lhm felde</i> | | |
| graphtoken | <i>lhm</i> | <i>felde</i> | |
| syntoken | <i>lh</i> | <i>m</i> | <i>felde</i> |

Bsp. 14, Alme 1630:

| | | | | |
|-------------------|----------------------------|----------|-----------------|---------------|
| Text | <i>ins teuffelß Nahmen</i> | | | |
| graphtoken | <i>ins</i> | | <i>teuffelß</i> | <i>Nahmen</i> |
| syntoken | <i>in</i> | <i>s</i> | <i>teuffelß</i> | <i>Nahmen</i> |

3.2.1.7 Univerbierungen

Bei getrennt geschriebenen Univerbierungen lässt sich anders als bei Komposita nicht über formale Merkmale eines Einzelbelegs bestimmen, ob es sich um ein oder um zwei syntaktische Wörter handelt. Als univerbiert gelten die folgenden Kombinationen, wenn diese im Kontext eine von der freien syntaktischen Phrase abweichende Bedeutung zeigen, z.B. *in deme* als Subjunktion ‚indem‘. Im Korpus werden folgende Fälle als Univerbierungen annotiert. Alle anderen potentiellen Univerbierungen werden bei Getrenntschreibung als zwei syntaktische Tokens annotiert.

- *all weg*
- *all zeit*
- *zue mehrmaln*
- *offt mahlen*
- *zue dem*
- *in deme*
- *nach deme*
- *ein mahl*
- *mit sambt*
- *nichts desto weniger*
- *nach gehends*
- *bey sammen*
- *biß weiln*
- *zu forderst*
- *so baldt*
- *des wegen*
- *keines wegen*
- *Jeder zeit*
- *wider um*
- *oben an*

- *folgender maßen*
- *einß dan*
- *außer halb*
- *Jegen werttig*
- *Zu rucke*
- *alß dan*
- *zu Letst*

3.2.1.8 Graphische Doppelungen

Das erste Wort einer Seite wird häufig unten auf der vorherigen Seite wiederholt (Kustode oder Reklamante):

Bsp. 15, Georgenthal 1597:

Geist Ihr dieselben All weg Inn einem stotzen gepracht,
vnn d ↓ ←
Vnnd zu getragenn, Vnd Ime bepholen, deren nicht

Die graphische Tokendoppelung wird durch Auslassung auf der syntaktischen Tokenebene festgehalten. Als syntaktisches Token ausgelassen wird jeweils das zweite graphische Token.

| | | | | | | | | | | |
|-------------------|---|--------------|----------------|-----------------|----------|--------------|-------------|-----------|------------------|----------|
| Text | <i>Inn einem stotzen gepracht, vnn d ↓ ← Vnnd zu getragenn,</i> | | | | | | | | | |
| graphtoken | <i>Inn</i> | <i>einem</i> | <i>stotzen</i> | <i>gepracht</i> | <i>,</i> | <i>vnn d</i> | <i>Vnnd</i> | <i>zu</i> | <i>getragenn</i> | <i>,</i> |
| syntoken | <i>Inn</i> | <i>einem</i> | <i>stotzen</i> | <i>gepracht</i> | <i>,</i> | <i>vnn d</i> | \emptyset | <i>zu</i> | <i>getragenn</i> | <i>,</i> |

Ein besonderer Fall tritt bei Wiederholung mehrerer graphischer Tokens auf. In diesem Fall wird beim zweiten Vorkommen des graphischen Tokens ein Verweis auf das erste Vorkommen implementiert, sodass dieses eindeutig identifizierbar ist:

Bsp. 16, Georgenthal 1597:

Wiederholung zweier graph. Tokens, (hier mit Veränderung der Form)

| | | | | | | | | | | |
|-------------------------|---|---------------------|---------------|--------------|----------------|--------------------------|----------------------------|-------------------|---------------------|--------------|
| Text | <i>Ihr vnter dessen einen ↓ kleinen ↓ ← Ein kleinen Schlüssell Auffzuhebenn geben</i> | | | | | | | | | |
| graphtoken | <i>Ihr</i> | <i>vnter</i> | <i>dessen</i> | <i>einen</i> | <i>kleinen</i> | <i>Ein</i> | <i>kleinen</i> | <i>Schlüssell</i> | <i>Auffzuhebenn</i> | <i>geben</i> |
| Erstes Vorkommen | | | | | | Verweis auf <i>einen</i> | Verweis auf <i>kleinen</i> | | | |
| syntoken | <i>Ihr</i> | <i>vnter dessen</i> | | <i>einen</i> | <i>kleinen</i> | \emptyset | | <i>Schlüssell</i> | <i>Auffzuhebenn</i> | <i>geben</i> |

3.2.1.9 Streichungen

Gestrichene Wortformen (oder auch nur Buchstabenkombinationen) werden bei der syntaktischen Tokenisierung ausgelassen. Sie erhalten nur auf der graphematischen Tokenebene Annotationen.

4 Tokenbasierte Annotation

Auf den beiden Tokenebenen *graph tokens* und *synt tokens* sind folgende Annotationen angelegt:

1. Großschreibung
2. Wortart (PoS)
3. Belebtheit
4. Lemma
5. Editorische Informationen (vgl. Kap. Editorische Grundsätze)

Grundsätzlich werden diese Informationen sowohl für graphische als auch für syntaktische Tokens festgehalten. Unterschiede bei der Annotation der beiden Tokenarten werden jeweils in den einzelnen Abschnitten beschrieben.

4.1 Großschreibung

Tokens werden bezüglich der Großschreibung (Feature „orth“) wie folgt annotiert:

1. Typus Enthält mindestens eine Majuskel (orth=upperInitial, orth=mixedCaps oder orth=allCaps)
2. Typus Enthält keine Majuskel, aber enthält Minuskel (orth=lowerCase)¹³
3. Typus Besteht aus Ziffern und möglicherweise Interpunktionszeichen (orth=numeric)
4. Typus Enthält weder Majuskel noch Minuskel noch Ziffern (orth=punctuation)

Die Typen 1 und 2 erhalten eine Annotation für Lemma und PoS. Für den Typus 3 wird stets das Lemma *ZIFFER* vergeben (vgl. Kap. Zahlen und Ziffern). Für die PoS-Annotation werden bei Typus 3 zwei Fälle unterschieden:

1. Das Token ist Teil einer Phrase, aber nicht deren Kopf,
2. Das Token ist Kopf einer Phrase oder nicht in eine Phrase integriert.

Im ersten Fall erhält das Token einen Tag entsprechend seiner syntaktischen Verwendung eine PoS-Annotation, z.B. ADJA in „zum 17. Mal“. Im Beispiel „vor 3 Jahren Funcken Olmes“ (Alme 1630) trägt 3 den Tag CARD, analog dazu erhalten die Ziffern im Beispiel „5 oder 6 Schläge“ den Tag CARD. Im zweiten Fall wird der Tag „-“ vergeben (vgl. Bsp. 17, Bsp. 18 & Kap. Die Meta-Lemmata „*“ und „-“).

Bsp. 17, München 1600:

17. *Bey wem er alldort zhört*

Bsp. 18, Georgenthal 1597:

*Zum 4. Es habe der böse Geist Auch vff ein Zeit
den Hanßen Grymmen Inn ihr hauß gepracht*

¹³ Hierunter fallen zum gegenwärtigen Stand auch Tokens wie „5ten“, die als (Pseudo-)Initial eine Ziffer haben.

4.2 PoS

4.2.1 Einführung

Grundsätzlich wird für die PoS-Annotation das Stuttgart-Tübingen Tagset (STTS) verwendet. Generell gelten dabei die 1999 veröffentlichten Guidelines (Schiller et al. 1999). In diesem Handbuch werden ausschließlich die davon abweichenden Vorgänge thematisiert.

4.2.2 Zusätzliche Tags

Neben den in STTS verwendeten Tags werden die folgenden Tags verwendet (eine vollständige Übersicht aller verwendeten Tags findet sich unter Liste der PoS-Tags):

Tabelle 1: Im Korpus verwendete Tags, die nicht im STTS enthalten sind

| | |
|------------------------------------|--|
| NB | Nomen, Berufsbezeichnung |
| Lat | Lateinisches Wort |
| KOU | Nebensatzeinleitung (außer Relativpronomina); umfasst das, was in STTS als KOUS und KOUJ getaggt wird |
| KO | ambiger Satzeinleiter für Fälle, in denen KON und KOU nicht auseinander gehalten werden können |
| PAVL | linker Teil eines diskontinuierlichen Pronominaladverbs |
| PAVAP | rechter Teil eines diskontinuierlichen Pronominaladverbs |
| VI | Infinites Verb |
| VVIZU | zu-Infinitiv |
| VIMP | Verb im Imperativ |
| VF | Finites Verb |
| VPP | Partizip Perfekt eines Verbs |
| VA | Verb, bei dem ambig ist, ob es finit oder infinit ist |
| Nur beim graphischen Token: | |
| PRE | Wortteil, der mit einem Präfix beginnt |
| SUF | Wortteil, der mit einem Suffix beginnt |
| STEM | Wortteil, der mit einem Stamm beginnt |
| - | Ein Element, dem man als Ganzes keine Wortart zuordnen kann (z.B. bei Worttrennung am Zeilenende: <i>sol - ches</i>). |

Der Tag „NB“ ist eine Erweiterung, die der Entstehung von Beinamen aus Berufsbezeichnungen und dem darin begründeten unklaren Status vieler Berufsbezeichnungen Rechnung trägt.

„Lat“ steht für Lateinisch: STTS sieht hier eine Zuordnung zu „FM“ (*foreign material*) vor. Latein ist jedoch in den Hexenverhörprotokollen die einzige Fremdsprache, die dem Dialektkontinuum des Frühneuhochdeutschen gegenübersteht. Deshalb wurde hier ein präziserer Tag gewählt.

4.2.2.1 Regeln zur Verwendung von NE, NB und NN

Das Tag **NE** ist vorgesehen für:

- 1) Städtenamen, z.B. zu Regensburg,
- 2) Rufnamen und Beinamen, wenn sie nicht transparent sind, z.B. Hanns Scheiflinger,

- 3) transparente (möglicherweise auch motivierte) Namenszusätze, die in Kombination mit einem Rufnamen stehen, z.B. *Heintz Möller, Adam Schneiders*,
- 4) syntaktisch auffällige Zusätze in Phrasen, die Personen bezeichnen, z.B. ein unflektiertes Adjektiv wie in *die Schwart Trine*,
- 5) syntaktisch auffällige Nomina, die v.a. Personen, aber auch Orte bezeichnen, z.B. Vorkommen ohne Artikel wie in *Wolf kommt herein* (konstruiert). Auch die Verwendung des Verbs *nennen* führt zur Vergabe von NE, z.B. *heuberg genannt* (Riedlingen 1596). In diesem Beispiel ist der Eigenname zusätzlich durch die Artikellosigkeit gekennzeichnet,
- 6) Nomina, die von einem transparenten (daher ambigen) Personennamen (Beinamen bzw. Familiennamen) abgeleitet werden, z.B. die Diminution *das meierken* von *Meier* in *Aus wüllfte das meierken* (Alme 1630),
- 7) Nomina (NN), denen ein Anredeklassifikator vorausgeht, z.B. *Herr Wolf*.

Der Tag **NB** ist vorgesehen für:

- 1) Berufsbezeichnungen, die nicht mit einem Rufnamen verbunden sind, z.B. *Schulmeister, wirt*,
- 2) Berufsbezeichnungen, die durch einen Ortszusatz konkretisiert werden (Bedingung 1 muss erfüllt sein), z.B. *wirtten, od[er] Schencken zu Meliß* (Georgenthal 1597),
- 3) Berufsbezeichnungen, die neben dem Ruf- und Familien- bzw. Beinamen auftreten, z.B. *Melchior Anschutz schneider* (Georgenthal 1597), oder *der wirt lorenz konigk* (Georgenthal 1597).

Eigennamen treten oft in Reihungen oder komplexen Namensformeln auf. Weitere Tokens, die in Phrasen auftreten, mit denen Personen bezeichnet werden, erhalten den Tag, der ihnen auch in anderen Kontexten zukommt.

Beispiele der Annotation von Namen aus dem Protokoll Georgenthal 1597:

- 1) *Christina (NE) Thymen (NE) Von (APPR) Meliß (NE)*
- 2) *Bastian (NE) Fischbechers (NE) zu (APPR) Schönawen (NE)*
- 3) *Hanßenn (NE) Langenhains (NE) zu (APPR) hohen (ADJA) Kirchen (NN)*
- 4) *deß Gerichts- (NB) vnnd (KON) Amptschreibers (NB) Hanßen (NE) Heinrichs (NE) Körnners (NB)*
- 5) *vff (APPR) die (ART) Allte (ADJA) Heppen (NE) Schmidin (NE) zu (APPR) Benßhausenn (NE)*

Der Tag **NN** wird bei allen anderen nicht fremdsprachlichen Substantiven vergeben. Dazu zählen auch Bezeichnungen für Lokalitäten, die keine Siedlungsnamen sind wie etwa *Richtplatz, Domplatz, Tanzplatz* oder *an den Toren; an den Heertoren*.¹⁴

Auch *Teufel, Satan* und *Gott* sind als NN annotiert.

4.2.2.1 Die Verben

Abweichend vom STTS wird im SiGS-Korpus nicht zwischen Hilfs-, Modal- oder Vollverben unterschieden, sondern lediglich, ob es sich um einen Infinitiv (VINF, VIZU), ein Partizip (VPP), einen Imperativ (VIMP) oder ein finites Verb (VFIN) handelt.

Der Tag VIZU markiert auf der *graphtoken*-Ebene Fälle, in denen ein *zu*-Infinitiv ein graphematisches Wort darstellt. Auf der *syntoken*-Ebene werden diese Wörter dann in Partikel und Infinitiv getrennt (vgl. Kap. zu als Infinitivmarker).

Bsp. 19 (= Bsp. 11), Alme 1630:

| Text | <i>daruff zuleben Vnd zusterben</i> | | | | | |
|-----------------------|-------------------------------------|----------------|--------------|------------|------------------|----------------|
| graphtoken | <i>daruff</i> | <i>zuleben</i> | | <i>Vnd</i> | <i>zusterben</i> | |
| graphtoken_pos | PAV | VIZU | | KON | VIZU | |
| syntoken | <i>daruff</i> | <i>zu</i> | <i>leben</i> | <i>Vnd</i> | <i>zu</i> | <i>sterben</i> |
| syntoken_pos | PAV | PTKZU | VI | KON | PTKZU | VINF |

Wenn nicht entscheidbar ist, ob es sich um ein finites (d.h. VFIN oder VIMP) oder ein infinites (d.h. VINF, VPP oder VIZU) Verb handelt, wird der Tag VA vergeben. Dies ist u.a. in solchen Fällen wie in Bsp. 20 notwendig.

Bsp. 20, Schweinfurt 1616:

*Vff obbeschehene Ausßag Affræ
Riglerin, ist Susanna Pfisterin
befraget worden, welche solches
nachfolgender gestaldt
beantwortet.*

4.2.2.3 Die Konjunktionen

Die Tags KON, KOU und KO bezeichnen Konjunktionen und Subjunktionen. KON bezeichnet nebenordnende Konjunktionen wie *und*, KOU bezeichnet nebenordnende Subjunktionen wie *weil* und KO wird in Fällen verwendet, in denen (bspw. aufgrund elliptischer Strukturen) nicht zwischen KON und KOU entschieden werden kann.

Die Annotation wird anhand des topologischen Feldermodells und der Stellung des finiten Verbs getroffen: KOU stehen im Vorvorfeld des Satzes. Die unterordnenden Konjunktionen KOU besetzen die linke Klammer („Sie sind froh, **weil** sie zum Annotieren gehen.“), dies ist an der Spätstellung des finiten Verbs erkennbar.

¹⁴ Ortbezeichnungen werden in der Belebtheitsannotation über das Tag „konkret (Ort)“ erfasst (s. Kap. Konkreta).

Ist die Felderstruktur durch das Fehlen des finiten Verbs nicht bestimmbar, so wird – außer bei einigen unten aufgeführten Ausnahmen – der Tag KO vergeben.

Bsp. 21, Günzburg 1613:

da Ir ain gliedt nach dem andern heraus gerißten

KO wird auch beim graphischen Token vergeben, wenn eine Konjunktion als Teil einer Univerbierung vorliegt.

Bsp. 22, Alme 1630:

Vnd Alß

dan hab ihme der teuffel einen schwartzen gurtel bracht,

| | | |
|----------------|----------|------|
| graphtoken | alß | dann |
| graphtoken_pos | KO | ADV |
| syntoken | alß dann | |
| syntoken_pos | ADV | |

Dies gilt nicht, wenn die betreffende Univerbierung eine eindeutige Subjunktion enthält, z.B.

Bsp. 23, Jever 1592:

Auch Ir der Teuffel auferlecht, daß sie Ihren eigenen vatter, vor 3. Jairen darumb daß ir Moder ihr nicht Brautschatzes genuch geben wollen, eine kho vergeben,

Einige Junktionen sind immer unterordnend bzw. nebenordnend und werden daher unabhängig von der Realisierung des Prädikats (mit oder ohne finites Verb) mit KOU bzw. KON annotiert. Als eindeutig unterordnende Konjunktionen gelten *ob*, *weil* und *dass*. Sie erhalten immer den Tag KOU. Als eindeutig nebenordnend gelten *und* sowie *oder*. Sie erhalten immer den Tag KON. Auch bei der Verknüpfung zweier NPs mit einer Junktion liegt immer eine Nebenordnung (KON) vor.

4.2.2.4 Pronominaladverbien

Als Pronominaladverb (PAV) werden die Kombinationen aus *da(r)*, *hier* und *wo(r)* mit einer Präposition annotiert, *darauf*, *daneben*, *hierauf*, *hiermit*, *womit*. Zudem können auch Reduktionen auftreten, z.B. *drüber*.

Ist ein Pronominaladverb auf der *graphtoken*-Ebene so getrennt, dass der zweite Teil die vollständige Präposition enthält, werden die Elemente mit den Tags PAVL (Pronominaladverb, linker Teil) und PAVAP (Pronominaladverb, präpositionaler Teil) annotiert, z.B. *da-* (PAVL) ↓ *vber* (PAVAP).

Bsp. 24, Jever 1592:

da-

vber er so seer erschreckt

4.2.2.2 Die Verwendung der zusätzlichen Tags bei nicht übereinstimmenden Tokengrenzen

Grundsätzlich gilt, dass – wenn möglich – das graphische Token wie ein syntaktisches Token behandelt wird. Allerdings können hier Einheiten auftreten (z.B. bei der Worttrennung), die in STTS nicht berücksichtigt sind. Für sie werden zusätzliche Tags benötigt, die in der Tabelle in Kap. Liste der PoS-Tags entsprechend gekennzeichnet sind.

Bei der Nichtübereinstimmung von graphischen und syntaktischen Tokens (z.B. durch die Trennung am Zeilenende) lassen sich für die Analyse drei Fälle unterscheiden.

1. Das graphische Token ist kleiner als ein syntaktisches Wort:
Es wird einer der Tags PRE, SUF, STEM oder „-“ vergeben.
2. Das graphische Token kann auch als syntaktisches Wort angesehen werden:
Es wird der Tag vergeben, der dem syntaktischen Wort zukäme.
3. Das graphische Token ist größer als ein syntaktisches Wort:
Es wird entweder APPRART, VIZU oder „-“ vergeben.

In der Regel werden graphische Einheiten, die mit einem Bindestrich beginnen oder enden, nicht als potentielle syntaktische Wörter interpretiert. Im Fall von Präfixen und Suffixen wird aber PRE bzw. SUF vergeben, bei anderen Wortteilen „-“.

Bsp. 25, Hildesheim 1628:

| | | |
|-------------------|----------------------|-----------------|
| Text | <i>bey-↓gelassen</i> | |
| graphtoken | <i>bey-</i> | <i>gelassen</i> |
| graphtoken PoS | <i>PRE</i> | <i>STEM</i> |
| syntoken | <i>beygelassen</i> | |
| syntoken PoS | <i>VVPP</i> | |

Bsp. 26, Alme 1630:

| | | | |
|-------------------|------------------------|---------------|-------------|
| Text | <i>St. Jacobi tagh</i> | | |
| graphtoken | <i>St.</i> | <i>Jacobi</i> | <i>tagh</i> |
| graphtoken_pos | Lat | Lat | NN |
| syntoken | <i>St. Jacobi tagh</i> | | |
| syntoken_pos | NN | | |

4.2.2.3 Die Verwendung der zusätzlichen Tags bei übereinstimmenden Tokengrenzen

Bei elliptischen oder abgekürzten Wortformen wird auf der graphischen Ebene nur annotiert, was sichtbar ist. Auf der syntaktischen Ebene wird so annotiert, als wäre es die erschließbare Vollform (z.B. das Präfix *Ein-* im folgenden Beispiel).¹⁵

Bsp. 27, konstruiert:

| Text | <i>Ein- und Ausgang</i> | | |
|-------------------|-------------------------|------------|----------------|
| graphtoken | <i>Ein-</i> | <i>und</i> | <i>Ausgang</i> |
| graphtoken_pos | <i>PRE</i> | <i>KON</i> | <i>NN</i> |
| syntoken | <i>Ein-</i> | <i>und</i> | <i>Ausgang</i> |
| syntoken_pos | <i>NN</i> | <i>KON</i> | <i>NN</i> |

Bsp. 28, konstruiert, verbale Koordination:

| Text | <i>an und ausmachen</i> | | |
|-------------------|-------------------------|------------|------------------|
| graphtoken | <i>an</i> | <i>und</i> | <i>ausmachen</i> |
| graphtoken_pos | <i>PTKVZ</i> | <i>KON</i> | <i>VVFIN</i> |
| syntoken | <i>an</i> | <i>und</i> | <i>ausmachen</i> |
| syntoken_pos | <i>VVFIN</i> | <i>KON</i> | <i>VVFIN</i> |

4.2.3 Von STTS abweichende Regeln für die Verwendung von STTS-Tags

Um den Anforderungen eines historischen Sprachkorpus gerecht zu sein, ist die Vergabe von PoS-Tags gegenüber dem STTS-Tagset unter Berücksichtigung von syntaktischen Bedingungen eingeschränkt. Dies trifft insbesondere auf Relativpronomina, attributiv verwendete Indefinitpronomina sowie die Annotation von Substantivierungen zu.

4.2.3.1 PRELS und PRELAT

Bei den Pronomina *der*, *wer*, *welch* und *so* basiert die Entscheidung, ob es sich um ein Relativpronomen handelt, auf der Stellung des finiten Verbs. Diese Annotation erfolgt nur bei Verbletzstellung, s. (29a), bei Verbzweitstellung nicht (29b).

Bsp. 29a, Baden-Baden 1628:

*Ihren habe Er auch etwaß geben, alß wann
es drey bätzner gewesen, So (PRELS) aber
nur Ross koth war.*

¹⁵ Der in STTS verwendete Tag TRUNC wird nicht verwendet.

Beispiel 29b, Baden-Baden 1628:

*Sie solle sein sein, thun waß Er wöl-
le, So (ADV) wolle Ihnen alles genueg
geben.*

Fehlt das finite Verb und ist daher nicht bestimmbar, ob Verbletz- oder Verbzweitstellung vorliegt (29c), wird nicht „PRELS“ vergeben, sondern der Tag, der bei der nicht-relativischen Lesart vergeben werden würde, aber ergänzt um „_REL“.

Beispiel 29c, Baden-Baden 1628:

*Sie habe ain Stiffsohn gehabt, der (PDS_REL)
Bastian gehaisßen. denselben (PDS) habe Sie
vngefährlich vor 28. Jahren, in des
Sathanß namen geschlagen*

Bei *denselben* im obigen Beispiel liegt Verbzweitstellung des Finitums vor, daher ist keine relativische Lesart möglich. Im vorangegangenen Satz fehlt das finite Verb. Daher wird für *der* der Tag PDS_REL vergeben.

Die Doppeldeutigkeit, angezeigt durch „_REL“ kann bei folgenden Tags auftreten:

ADV *so*, PDS *der*, PDAT *dessen*, PWAV *wo*, PAV *dafür*

4.2.3.2 Indefinitpronomina in attributiver Stellung: PIDAT vs. PIAT

Die Annotation der attribuierenden Indefinitpronomina berücksichtigt ausschließlich die konkrete Realisierung: PIDAT wird im SiGS-Korpus nur vergeben, wenn unmittelbar hinter dem betreffenden Token (z.B. *all*) ein Artikel auftritt, z.B. *all die Leute*. Ansonsten wird PIAT vergeben (z.B. *alle Leute*). Im Gegensatz zu gegenwartssprachlichen Korpusannotationen können keine Tests durchgeführt werden, ob ein Token prinzipiell vor einem Artikel auftreten könnte.

4.2.3.3 Unterscheidung von Substantivierungen und elliptischen Konstruktionen

Ähnlich kann die Unterscheidung, ob ein Adjektiv als Nomen gebraucht wird oder vielmehr eine kopflose Nominalphrase vorliegt, für die frühneuhochdeutschen Texte nicht getroffen werden. Beispielsweise ist Beispiel 30 ambig zwischen ‚die Alte‘ (Substantivierung) und ‚die alte [Frau]‘ (Ellipse), sofern der Diskurskontext die elliptische Lesart erlaubt. Anders als im Gegenwartsdeutschen, wo die Groß- und Kleinschreibung zur Desambiguierung beiträgt, kann die Ambiguität in diesem Korpus nicht eindeutig aufgelöst werden. Deshalb werden Nominalphrasen im SiGS-Korpus immer so annotiert, dass in der Annotation ein nominaler Kopf enthalten ist. Es wird bei der Annotation also nicht zwischen elliptischen Nominalphrasen und Nominalphrasen mit Konversionsprodukten als Kopf unterschieden (so auch HiTS, S. 22ff.), sondern stets eine Substantivierung angenommen.

Bsp. 30, Augsburg 1625:

die altte (NN), die sie doch nicht Zue nennen wüsse, habe Ihrs geben

4.2.4 Umgang mit spezifischen sprachlichen Phänomenen

4.2.4.1 Abgrenzung zwischen unbestimmtem Artikel und Zahlwort

Ain wird als Zahlwort (CARD) annotiert, wenn eine Maßangabe folgt (wie z.B. bei *Ain* (CARD) *meil*, München 1600, *welche ihm 1* (CARD) *thaler geben*, Alme 1630). In den übrigen Fällen handelt es sich um den unbestimmten Artikel.

4.2.4.2 Analyse von Partizipien

Wenn nicht entschieden werden kann, ob das Partizip als Teil des Prädikats (VVPP) oder adjektivisch (ADJA/D) verwendet ist, erfolgt die Annotation unter Einbezug des Kontexts. Zusätzlich wird das DWB konsultiert, um zu überprüfen, ob ein Partizip bereits als Adjektiv lexikalisiert ist. In Bsp. 31 ist *Bekandt* als VPP annotiert. Grund dafür ist die vorausgehende Passage *Item Vorm Jahr ein grae katzen mit schwartzer materi vergeben* (VPP), die syntaktisch parallel konstruiert ist:

Bsp. 31, Alme 1630:

*Item Bekandt daß ihn Scheper Volpart
auff der Möne vor 3 Jahren die
wahrwulffschafft gelert hab*

4.2.4.3 Verbpunktikeln

Grundsätzlich werden nur Tokens als Verbpunktikel (PTKVZ) annotiert, die homograph mit Präpositionen sind (*mitkommen*, *mit Anna kommen*). Bei einem solchen Token handelt es sich um eine Verbpunktikel, wenn keine NP vorliegt, die von der Präposition regiert ist (Bsp. 32). Wenn eine solche vorliegt, handelt es sich um eine Präposition (Bsp. 33). In allen übrigen Fällen wurde (wenn getrennt geschrieben) die Wortart annotiert, die bei freier Verwendung angenommen werden kann, z.B. „*Rad* (NN) *fahren*.“

Bsp. 32, Leipzig 1640:

| Text | <i>Vrsul</i> | <i>Bilingen</i> | <i>will</i> | <i>mit</i> | <i>gehen</i> |
|-------------------|--------------|-----------------|-------------|------------------|--------------|
| graphtoken | <i>Vrsul</i> | <i>Bilingen</i> | | <i>mit</i> | <i>gehen</i> |
| graphtoken_pos | | | | PTKVZ | VINF |
| syntoken | <i>Vrsul</i> | <i>Bilingen</i> | | <i>mit gehen</i> | |
| syntoken_pos | | | | VINF | |

Bsp. 33, Reichenberg 1653:

| Text | <i>soll</i> | <i>mit</i> | <i>lr</i> | <i>gehen</i> |
|-------------------|-------------|------------|-----------|--------------|
| graphtoken | <i>soll</i> | <i>mit</i> | <i>lr</i> | <i>gehen</i> |
| graphtoken_pos | | APPR | | VINF |

PTKVZ wird auf der graphtoken-Ebene auch bei der Getrennschreibung von nominalisierten Partikelverben vergeben.

Bsp. 34, Höchst 1631:

| | | | | |
|-------------------|-----------------------------|------------------------|------------|----------------|
| Text | <i>Ihre mit ge- fertten</i> | | | |
| graphtoken | <i>Ihre</i> | <i>mit</i> | <i>ge-</i> | <i>fertten</i> |
| graphtoken_pos | | PTKVZ | PRE | STEM |
| syntoken | <i>Ihre</i> | <i>mit ge- fertten</i> | | |
| syntoken_pos | | NN | | |

4.2.4.4 Partikeln

Abseits der oben genannten Verbpartikeln sieht STTS keine Tags für Partikeln, etwa Diskurspartikeln, vor.¹⁶ In den meisten Fällen handelt es sich hierbei um Wörter, die gemäß STTS als ADV annotiert werden. Diese Praxis wird hier beibehalten, auch weil in vielen Fällen nicht zweifelsfrei festgestellt werden kann, um welche Art von Partikel es sich im konkreten Fall handelt.

Bsp. 35, München 1600:

er sei aber (ADV) leider zu weit khommen

4.2.5 Liste der PoS-Tags

| Tag | Bedeutung | Beispiel | Tagset |
|---------|---|---|--------|
| ADJA | Adjektiv, attributiv | <i>[das] weite [Land]</i> | STTS |
| ADJD | Adjektiv, prädikativ oder adverbial | <i>[es ist] weit, [sie laufen] weit</i> | STTS |
| ADV | Adverb | <i>vngefehr</i> | STTS |
| APPR | Apposition, Präposition | <i>zu, bei, an</i> | STTS |
| APPO | Apposition, Postposition | <i>[ihm] zufolge</i> | STTS |
| APZR | Apposition, rechter Teil einer Zirkumposition | <i>[um Ihrer] willen</i> | STTS |
| APPRART | Apposition-Artikel, Klise | <i>aufs, ins, im, zur</i> | STTS |
| ART | Artikel | <i>der, die, das, ein, eine, eines</i> | SIGS |
| CARD | Kardinalzahl | <i>zwei [Männer]</i> | STTS |
| ITJ | Interjektion | <i>ach</i> | STTS |
| NN | Nomen | <i>Haus</i> | STTS |

¹⁶ Vgl. dazu auch: Hirschmann, Hagen (i.V.): *Richtlinien zur Wortartenannotation von Adverb- und Partikelklassen – eine Granularisierung des STTS im Bereich von Modifikatoren. Technischer Report.*

| Tag | Bedeutung | Beispiel | Tagset |
|--------|--|--|--------|
| NE | Eigennamen | <i>Otto</i> | STTS |
| NB | Berufsbezeichnung | <i>Schäfer</i> | SIGS |
| KOU | Konjunktion, unterordnend | <i>denn, weil</i> | SIGS |
| KON | Konjunktion, nebenordnend | <i>und, oder, entweder ... oder</i> | STTS |
| KO | Konjunktion, neben- oder unterordnend (ambiger Fall, s. oben) | | SIGS |
| KOKOM | Vergleichspartikel | <i>[besser] als</i> | STTS |
| Lat | Lateinisch | <i>torqueatur</i> | SIGS |
| PPER | Pronomen, personal | <i>er, sie, es</i> | STTS |
| PPOSS | Pronomen, possessiv, ersetzend | <i>meins, deiner</i> | STTS |
| PPOSAT | Pronomen, possessiv, attributiv | <i>ihr [Haus], deine [Mutter]</i> | STTS |
| PDS | Pronomen, demonstrativ, substituierend | <i>dies/jenes [ist ein Buch]</i> | STTS |
| PDAT | Pronomen, demonstrativ, attributiv | <i>dieses/jenes [Buch]</i> | STTS |
| PRELS | Pronomen, relativ | <i>[derjenige], welcher</i> | STTS |
| PRELAT | Pronomen, relativ, attributiv | <i>[der Mann], dessen [Hut]</i> | STTS |
| PRF | Pronomen, reflexiv | <i>sich</i> | STTS |
| PIS | Pronomen, indefinit, substituierend | <i>keiner [kam], viele, man, niemand</i> | STTS |
| PIAT | Pronomen, indefinit, attributiv | <i>kein [Wort], irgendein [Glas]</i> | STTS |
| PIDAT | Pronomen, indefinit, attributiv und artikelfähig | <i>[ein] wenig [Wasser]</i> | STTS |

| Tag | Bedeutung | Beispiel | Tagset |
|--------|--|---------------------------------|--------|
| PWS | Pronomen, interrogativ | <i>wer, was</i> | STTS |
| PWAT | Pronomen, interrogativ | <i>wessen [Hut]</i> | STTS |
| PWAV | Pronomen, interrogativ/relativ, adverbial | <i>warum, wo, wann, worüber</i> | STTS |
| PAV | Pronominaladverb | <i>dadurch</i> | STTS |
| PAVL | Diskontinuierliches PAV, linker Teil | <i>da [nicht für]</i> | SIGS |
| PAVAP | Diskontinuierliches PAV, rechter (appositionaler) Teil | <i>[da nicht] für</i> | SIGS |
| PTKA | Partikel, beim Adjektiv oder Adverb | <i>am [schnellsten]</i> | STTS |
| PTKANT | Partikel, Antwort | <i>ja, nein</i> | STTS |
| PTKNEG | Partikel, Negation | <i>nicht</i> | STTS |
| PTKZU | Partikel, „zu“ | <i>zu</i> | STTS |
| PTKVZ | Partikel, Verb | <i>[er kommt] vorbei</i> | STTS |
| VFIN | Verb, Vollverb, finit | <i>[sie] gehen</i> | STTS |
| VIMP | Verb, Vollverb, Imperativ | <i>komm!</i> | STTS |
| VINF | Verb, Vollverb, Infinitiv | <i>gehen</i> | STTS |
| VIZU | Verb, Vollverb, Infinitiv mit „zu“ | <i>loszulassen</i> | STTS |
| VPP | Verb, Vollverb, Partizip II | <i>gesucht, vergessen</i> | STTS |
| \$. | Interpunktionszeichen | <i>, . ? ! ; : - [] ()</i> | STTS |
| PRE | Präfix, nur beim graphischen Token | Ge- <i>lehret</i> | SIGS |
| SUF | Suffix, nur beim graphischen Token | <i>Pein- lich</i> | SIGS |
| STEM | Wortstamm, nur beim graphischen Token | Guet- <i>lich</i> | SIGS |

| Tag | Bedeutung | Beispiel | Tagset |
|-----|---|----------------------------|--------|
| - | Graphisches Token kann nicht als Ganzes einer Wortart zugeordnet werden, nur beim graphischen Token | <i>Guetli</i> <i>ch</i> | SIGS |

4.3 Belebtheit

Die Referenten eines Nomens lassen sich auf einer Belebtheitsskala abbilden, die von unbelebt bis (maximal) belebt reicht (vgl. z.B. Comrie 1981). Belebtheit wird bei Nomina und Namen (PoS-Tags: NN, NE, NB) annotiert. Prinzipiell wird Belebtheit bei SIGS abhängig vom Lemma vergeben. Wenn die Belebtheit des Lemmas nicht mit der Belebtheit in einem konkreten Kontext zusammenfällt, wird zusätzlich auf der syntoken-Ebene die kontextuelle Belebtheit annotiert (vgl. Kap. Annotation von Metaphern und Metonymien).

4.3.1 Belebtheitskategorien

Die Kategorie *Belebtheit* ist wie folgt ausdifferenziert:

Abstrakta

abstrakt

abstrakt Maß

Konkreta

konkret

konkret Körperteil

konkret Ort

konkret Kollektivum

Tierische Entitäten

tierisch

tierisch Kollektivum

Menschliche Entitäten

menschlich

menschlich Kollektivum

Übermenschliche Entitäten

übermenschlich positiv

übermenschlich Teufel

4.3.1.1 Abstrakta

Abstrakta sind Wörter, die konzeptuelle Entitäten ohne physisch manifeste Extension bezeichnen. In den meisten Fällen liegen im Text Verbalabstrakta vor, die ein Handlungskonzept als Nomen ausdrücken. Diese werden als „abstrakt“ ausgezeichnet. Maße unterscheiden sich von den übrigen Abstrakta: Darunter fallen Abstrakta, die als Punkt, Intervall oder Einheit konzeptualisiert und daher (potenziell) wiederkehrend sind. Sie sind als „abstrakt Maß“ annotiert. So kehrt im Beispiel *drei Meter* die Einheit „Meter“ dreimal wieder, wohingegen „der dritte Meter“ nicht wiederkehrend ist.

Potenziell wiederkehrend heißt dabei nicht, dass Punkte/Intervalle/Einheiten in Mehrzahl auftreten müssen. So ist auch das Nomen *Nacht* im Beispiel *er arbeitet eine Nacht hindurch* als „abstrakt Maß“ annotiert, während es im Beispiel *er hat die Nacht durchgearbeitet* als „abstrakt“ annotiert ist.

Viele Maßeinheiten gehen historisch auf eine konkrete Grundbedeutung zurück oder sind mit einem konkreten Gegenstand verbunden (*Fuß, Elle, Uhr*). Hier wird je nach Kontext entweder als „abstrakt (Maß)“ (*habe 3 fueß ins Teuffelß Nahmen zuruck gehen*) und als „konkret“ (*Ihm tut der Fuß weh*) annotiert.

Auch nominalisierte Ordinal- und Kardinalzahlen zählen zu den Abstrakta:

Bsp. 36a, Günzburg 1613:

Vff das zwayvndzwainziqst

Diese können aber auch z.B. für Menschen stehen und tragen dann entsprechend auf der *syntoken*-Ebene die Belebtheitsannotation „menschlich“. Da immer ein Kopf angenommen wird, ist hier drei als Kopf der NP annotiert. Daher ist auch die Bestimmung der Belebtheit notwendig (vgl. Kap. Annotation von Metaphern und Metonymien):

Bsp. 36b, Riedlingen 1596:

| | |
|-----------------------|---|
| graphtoken | <i>khunden Annelin , Greta deß Meßmerß Tochter , vnd Eua Melchers Weib, so alle <u>drey</u> zu Vnlengen verschiner Jaren verbrent</i> |
| graphtoken_belebtheit | abstrakt |
| syntoken_belebtheit | menschlich Kollektivum |

Allgemeine Begriffe für gesellschaftliche Gruppen (Klerus, Zunft, Adel, Bürgertum, Arbeiterklasse) werden als „abstrakt“ annotiert.

Beispiel 37, Alme 1630:

*Daruff ein Junffer mit bögelen Alß ein
Von Adel in blauwen kleidern Vnd*

4.3.1.2 Konkrete

Konkrete Entitäten sind physisch manifest und können als gegenständlich oder dinglich in der Welt erfahren werden. Wörter, die solche Entitäten bezeichnen, sind als „konkret“ annotiert, es sei denn, die Entitäten sind lebendig. Die Kategorie konkret wird in drei Fällen (bei Orten, größeren Flächen und Körperteilen) weiter ausdifferenziert:

Beispiele:

„konkret“: *Pott, Pulver, Baum, Haus*

„konkret Ort“: *Georgenthal, Tambach*

„konkret Kollektivum“: *Gebirge, Siedlung*

„konkret Körperteil“: *Kopf, Auge, Füße*, auch Hyperonyme: *Glied*

4.3.1.2.1 Annotation von Orten

Die Entscheidung zwischen den einzelnen Tags für Konkreta wird zum einen hinsichtlich der Individuiertheit getroffen, andererseits stehen Landschaftsbezeichnungen Gebäuden und anderen konstruierten Bauten gegenüber.

So ist *Siedlung* als „konkret Kollektivum“ annotiert, während *Haus* wegen seiner Individuiertheit als „konkret“ annotiert wird. Gebäude, umgrenzte und von Menschenhand konstruierte Bauten sind als „konkret“ getaggt. Demgegenüber werden Flurnamen und Landschaften sowie Ortschaften als „konkret Kollektivum“ annotiert.

Bsp. 38, Alme 1630:

*Ihre dantz Leine sei an den Heertorn (konkret)
Vnd an die Galgen gebunden.*

Toponyme werden immer als „konkret Ort“ annotiert.

4.3.1.2.2 Annotation von Körperteilen

Menschliche und tierische Körperteile sind als „konkret Körperteil“ annotiert. Dazu zählen auch Hyperonyme wie *Glied*, nicht aber *Leib* oder *Körper*. Auch Körperflüssigkeiten sind nicht „konkret Körperteil“ annotiert, sondern tragen die Annotation „konkret“.

Bsp. 39, „konkret Körperteil“, Alme 1630:

*Herde Hanß weib sei die wehrwulffin
Vnd hab einen stuppelden stert.*

4.3.1.3 Tierische Entitäten

Wörter, die auf tierische Entitäten referieren, erhalten die Annotation „tierisch“. Wenn ein Kollektivum vorliegt, wird entsprechend die Annotation „tierisch Kollektivum“ vergeben. Dieser Tag wird nur dann vergeben, wenn morphologisch ein Singular vorliegt.

Bsp. 40a, Alme 1630:

Vnd Viehe zerreißen

-> Viehe ist „tierisch“, da morphologisch Pl.

Bsp. 40b, konstruiert:

Vnd Vieh zerreißen

-> Vieh ist tierisch (Kollektivum), da morphologisch Singular, aber semantisch Mehrzahl.

Weitere Beispiele:

„tierisch“: *Kröte, Laus, Schaf, Ziegen, Katze*

„tierisch (Kollektivum)“: *Ungeziefer, Herde, Getier*

4.3.1.4 Menschliche Entitäten

Die Gruppe der Substantive (NN, NB, NE), die als „menschlich“ annotiert ist, umfasst maximal handlungsfähige Entitäten, die nicht im unteren Sinne (s. Kap. Übermenschliche Entitäten) übermenschlich sind. Daher werden auch mythologische Entitäten (*Werwolf, Drache, Geist*) hierunter gefasst.

Falls ein Kollektivum vorliegt, wurde die Annotation „menschlich Kollektivum“ vergeben. Diese wurde in der Regel nur bei Substantiven im Singular vergeben. Eine Ausnahme bildet das Pluraletantum *Leute*, das wegen seiner morphologischen Besonderheit immer als „menschlich Kollektivum“ annotiert wird.

Bespiele:

„menschlich“: *Kind, Richter, Trine*

„menschlich (Kollektivum)“: *Gemeinschaft, Gruppe*

4.3.1.5 Übermenschliche Entitäten

Im SiGS-Korpus wird eine enge Definition übermenschlicher Entitäten zugrunde gelegt: *Gott, Jesus, Maria* und *Engel* sind als „übermenschlich positiv“ und *Teufel* und *Satan* als „übermenschlich Teufel“ annotiert.

Die Verkörperung von übermenschlichen Entitäten (Teufel in Menschengestalt etc.) wird wie eine Metapher/Metonymie behandelt. Aufgrund des Übergangscharakters der Bezeichnungen und Tabuwörter für den Teufel finden sich hier auch Personenbezeichnungen, die zuerst wie Namen erscheinen und daher auf der graphischen Ebene den Tag „menschlich“ erhalten, z.B. *Federhanß*. Auf der syntaktischen Ebene tragen solche Tokens dann die Annotation „übermenschlich Teufel“.

Entsprechend wurde die Bezeichnung *Vater*, wenn damit ‚Gott‘ gemeint ist, auf der graphischen Ebene als menschlich, auf der syntaktischen-Ebene aber als „übermenschlich (positiv)“ annotiert, ebenso wie *Jungfrau*, wenn damit Jungfrau Maria gemeint ist.

4.3.2 Annotation von Metaphern und Metonymien

Substantive wie *Amt* und *Gericht* können metonymisch für Personen, z.B. Inquisitoren oder Richter stehen. In solchen Fällen wird zwischen einer dekontextualisierten und einer *Belebtheit im Kontext* unterschieden. Als Grundlage für die dekontextualisierte Bedeutung dient die erste Bedeutungsangabe im Deutschen Wörterbuch der Brüder Grimm (DWB). Bei *Gericht* ist dies z.B. „recht, gerechtigkeit, gesetz, gottes gebot“.

Die dekontextualisierte Belebtheit wird auf der graphtoken-Ebene annotiert, die Belebtheit im Kontext ist auf der syntoken-Ebene angelegt.

Bsp. 41, Hamm 1592:

| <i>Demnach Erkennen Richter vnd <u>Gericht</u></i> | |
|--|------------|
| graphtoken_belebtheit (dekontextualisiert) | abstrakt |
| syntoken_belebtheit (im Kontext) | menschlich |

4.3.3 Belebtheitsannotation von Komposita

Bei getrennt geschriebenen Komposita erfolgt die Annotation der Belebtheit sowohl bei den graphischen als auch beim syntaktischen Token. Hierbei gilt, dass auf der graphtoken-Ebene die dekontextualisierte Belebtheit und auf der syntoken-Ebene die Belebtheit im Kontext annotiert wird:

Bsp. 42, Georgenthal 1597:

| | | |
|------------------------------------|---------------------------|------------------|
| graphtoken | <i>Gerichts</i> | <i>Schöpfung</i> |
| Belebtheit (dekontextualisiert) | abstrakt | menschlich |
| syntoken | <i>Gerichts Schöpfung</i> | |
| Belebtheit (im Kontext) | menschlich | |

4.4 Lemmatisierung

4.4.1 Einführung

Die Lemmatisierung dient dazu, die einzelnen Tokens jeweils einem Lexem zuzuordnen. Als Lexem-Inventar dient der Lemmabestand des Deutschen Wörterbuchs der Brüder Grimm (DWB)¹⁷. Zusätzlich werden einige vom SiGS-Projekt definierte Meta-Lemmata angesetzt, die jeweils einer Lexemklasse entsprechen. Dies geschieht vor allem bei Wortklassen, bei denen eine Lemmatisierung jedes Einzelfalls ein zu feinkörniges Ergebnis liefern würde (z.B. Namen und Zahlen). Wenn vorhanden, werden spezifische Lemmata aus dem DWB bevorzugt. Zur Unterscheidung sind die Lemmata aus dem DWB komplett in Minuskeln gehalten, während die Meta-Lemmata komplett in Majuskeln stehen bzw. „-“ oder „*“ sind.

Übersicht über die Meta-Lemmata:

| |
|---------------|
| NAME |
| ZAHL |
| LAT |
| ZIFFER |
| INTERPUNKTION |
| D |
| E |
| - |
| * |

¹⁷ <http://woerterbuchnetz.de/DWB>

Die Lemmaliste ist damit in sich geschlossen und wird nicht erweitert. Wenn ein Token auftritt, dem keines der vorhandenen Lemmata bzw. Meta-Lemmata zugeordnet werden kann, wird das Meta-Lemma „*“ vergeben.

Grundsätzlich wird bei der Lemmatisierung das zu annotierende Token auf die Grundform zurückgeführt. So wird bei Partizipien der zugrundeliegende Infinitiv als Lemma vergeben. Für Pronomina heißt dies, dass die flexivisch am wenigsten spezifizierte (Nenn-)Form gewählt wird, also z.B. „dies“ als Lemma für *diese* und *dieser*. Auch Personalpronomina im Plural werden auf eine Singularform zurückgeführt. So ist pluralisches *sie* immer als „er“ annotiert. Nur wenn sich im DWB keine allgemeine Form finden lässt, wird eine spezifische gewählt (Maskulinum). Pluralformen werden bei der maskulinen Nominativ-Singularform eingeordnet (s. Tabelle „Grundformen der Pronomina“). Movierungen (*Krämerin*, *Apothekerin*) werden mit dem movierten Lemma („krämerin“) annotiert oder, wenn (wie im Fall von *Apothekerin*) nur eine männliche Form des Lemmas im DWB vorhanden ist, mit „*“ markiert.

4.4.2 Meta-Lemmata

4.4.2.1 Zahlen und Ziffern

Ausgeschriebene Zahlen: Falls die Zahl im DWB aufgeführt ist, erhält sie das entsprechende Lemma (z.B. *eins*, *hundert*). Ansonsten wird das Meta-Lemma „ZAHL“ (z.B. bei *einhunderteinundzwanzig*) vergeben. Ziffern erhalten immer das Meta-Lemma „ZIFFER“.

4.4.2.2 Namen

Das DWB enthält auch Eigennamen (z.B. „johannes“, „marie“), sodass Tokens, die den Tag NE erhalten haben, mit dem entsprechenden Namen lemmatisiert werden können. Ist ein Name nicht im DWB enthalten, so wird dem Token das Meta-Lemma „NAME“ zugewiesen.

Genauso wird bei transparenten Personennamen verfahren. Wenn das entsprechende Lemma im DWB existiert, wird es dem Token zugewiesen:

Bsp. 43, Baden-Baden 1628:

herr Krebs (graphtoken_lemma/syntoken_lemma=krebs)

4.4.2.3 Lateinische Wörter

Lateinische Wörter erhalten jeweils das Meta-Lemma LAT. Grundsätzlich gilt die Faustregel: Ob ein Wort lateinisch ist oder nicht, wird anhand seiner Morphologie entschieden. Unbekannte Wörter sind teils in der Edition ausgezeichnet oder finden sich in einem (mittel)lateinischen Wörterbuch. Auch das DWB gibt Hinweise auf einige gebräuchliche lateinische Wendungen.

Beispiele:

- *anno perge*
- *Martii*
- *item*
- *illa*

Hier wird von einem engen Begriff des Latinismus ausgegangen: Lateinische oder latinisierte Wörter gelten als deutsch, wenn sie frnhd. Flexion zeigen. Ebenso gelten Wörter als deutsch, bei denen an einen frnhd. Wortstamm ein lateinisches (bzw. nach der Verbreitung des Vulgärlateinischen als französisch/romanisch erscheinendes) Wortbildungsaffix tritt (*buhlirt*). In beiden Fällen ist die Flexion entscheidend, so dass auch dieser Beleg nicht als LAT annotiert

4.4.2.4 Interpunktionszeichen

Sämtliche Interpunktionszeichen erhalten das Meta-Lemma „INTERPUNKTION“.

, ; . : () [] ! ? ~

4.4.2.5 Definite und indefinite Artikel (der/die/das, ein/eine/ein)

Der Definitartikel ist als „D“ (wie *der/die/das*) und der Indefinitartikel als „E“ (wie *ein/e*) lemmatisiert. Weitere Artikelwörter (z.B. *keine* oder Demonstrativartikel wie *dieser*) wurden nach den üblichen Regeln lemmatisiert.

4.4.2.6 Die Meta-Lemmata „*“ und „-“

Mit dem Meta-Lemma „*“ werden Tokens ausgezeichnet, die prinzipiell einer Wortform entsprechen, wenn das entsprechende Lemma nicht in der Lemmaliste des DWB enthalten ist. Dies veranschaulicht das folgende Beispiel 44, bei dem *Stutenbrei* einem Lexem „stutenbrei“ zugeordnet werden könnte, aber kein passendes Lemma im DWB existiert, weswegen „*“ vergeben wird.

Bsp. 44, Alme 1630, Lemma nicht im DWB vorhanden:

| | |
|-----------------------------------|-------------------|
| Text | <i>Stutenbrei</i> |
| graphtoken = syntoken | <i>Stutenbrei</i> |
| graphtoken_lemma = syntoken_lemma | * |

Mit dem Meta-Lemma „-“ werden Tokens ausgezeichnet, die keiner Wortform entsprechen. Dies kann nur beim graphischen Token der Fall sein, vor allem bei Worttrennungen (vielfach mit Trennstrich am Zeilen- oder am Kolumnenende (Bsp. 45), sowie bei Abkürzungen (Bsp. 46).

Bsp. 45, Alme 1630, Worttrennung:

| | | |
|------------------|----------------------|---------------|
| Text | <i>wehr-↓ wulfte</i> | |
| graphtoken | <i>wehr-</i> | <i>wulfte</i> |
| graphtoken_lemma | - | <i>wolf</i> |
| syntoken | <i>wehr- wulfte</i> | |
| syntoken_lemma | <i>werwolf</i> | |

Bsp. 46, Alme 1630, Abkürzung:

| | | | |
|------------|------------------------|---------------|-------------|
| Text | <i>St. Jacobi tagh</i> | | |
| graphtoken | <i>St</i> | <i>Jacobi</i> | <i>tagh</i> |

| | | | |
|------------------|-----------------|-----|-----|
| graphtoken_lemma | - | LAT | tag |
| syntoken | St. Jacobi tagh | | |
| syntoken_lemma | NAME | | |

4.4.3 Unterschiede zwischen syntaktischen und graphischen Tokens

In den meisten Fällen ist das Lemma auf der graphischen und der syntaktischen Ebene gleich.

Ein Unterschied zwischen den Lemmata auf den beiden Tokenebenen tritt in den folgenden Fällen auf:

1. Wenn die Grenzen von graphischem und syntaktischem Token nicht übereinstimmen.
2. Wenn diese Grenzen zwar übereinstimmen, aber die graphische Einheit nicht als vollständiges syntaktisches Wort aufgefasst werden kann (PoS: „-“, PRE, SUF).
3. Wenn das (graphische und syntaktische) Token als eigenständiges syntaktisches Token auftreten könnte, aber als Teil einer größeren Einheit gelten kann (z.B. PoS: PTKVZ)

Generell wurden bei der Vergabe des syntaktischen Tokens Abkürzungen und Auslassungen aufgelöst (dies gilt auch für Wortreste, die durch Klisen auf der syntaktischen Ebene entstehen), sodass das Lemma auf dieser Ebene maximale syntaktische Worteinheiten repräsentiert (s. zur Veranschaulichung das Token *Ein* im Bsp. 47a).

Beim graphischen Token hingegen wird nicht mehr und nicht weniger annotiert, als an dem Token selbst sichtbar ist. Als Konsequenz hieraus kann der Fall entstehen, dass das Token keinem Lexem zugeordnet werden kann. Für diesen Fall ist das Meta-Lemma „-“ vorgesehen.

Bei Wortteililellipsen wie *Ein- und Ausgang* erhält das graphische Token *Ein-* aus diesem Grund das Meta-Lemma „-“. Auf der syntaktischen Tokenebene hingegen wird die Auslassung aufgelöst:

Bsp. 47a, konstruiert, Koordination zweier Substantive mit Auslassung:

| Text | <i>Ein- und Ausgang</i> | | |
|---------------------|-------------------------|------------|----------------|
| graph. Token | <i>Ein-</i> | <i>und</i> | <i>Ausgang</i> |
| PoS | <i>PRE</i> | <i>KON</i> | <i>NN</i> |
| Lemma | - | <i>und</i> | <i>ausgang</i> |
| Synt. Token | <i>Ein-</i> | <i>und</i> | <i>Ausgang</i> |
| PoS | <i>NN</i> | <i>KON</i> | <i>NN</i> |
| Lemma | <i>eingang</i> | <i>und</i> | <i>ausgang</i> |

Die Analyse als Auslassung wird auch vorgenommen, wenn die Auslassung nicht graphisch gekennzeichnet ist, aber aus dem Kontext klar erkennbar ist, dass es sich um eine Wortteililellipse handelt. So wird im folgenden Beispiel *Rueben* als Erstglied von *Ruebengärtl* analysiert; da das Lemma *Rübengärtl* nicht im DWB verzeichnet ist, erhält es auf der syntaktischen Tokenebene das Meta-Lemma „*“:

Bsp. 47b, München 1600, Koordination zweier Substantive mit Auslassung:

| Text | <i>Rueben oder Pflanzgärtl</i> | | |
|---------------------|--------------------------------|-------------|--------------------|
| graph. Token | <i>Rueben</i> | <i>oder</i> | <i>Pflanzgärtl</i> |

| | | | |
|--------------------|---------------|-------------|--------------------|
| PoS | <i>NN</i> | <i>KON</i> | <i>NN</i> |
| Lemma | <i>rübe</i> | <i>oder</i> | <i>*</i> |
| Synt. Token | | | |
| | <i>Rueben</i> | <i>oder</i> | <i>Pflanzgärtl</i> |
| PoS | <i>NN</i> | <i>KON</i> | <i>NN</i> |
| Lemma | <i>*</i> | <i>oder</i> | <i>*</i> |

4.4.4 Überblick: Grundformen der häufigsten Pronomina

| Lemma | Flexionsformen |
|---------------------------------------|---|
| Indefinitpronomina | |
| <i>all</i> | <i>alle, alles, aller, ...</i> |
| Interrogativ-/Relativpronomina | |
| <i>welch</i> | <i>welche, welcher, welches, ...</i> |
| Demonstrativpronomina | |
| <i>dies</i> | <i>dies, diese, dieser, ...</i> |
| <i>der</i> | <i>der, den, dem, dessen, ... die [Pl.]</i> |
| <i>die</i> | <i>die, der, deren, derer, ...</i> |
| <i>das</i> | <i>das, den, dessen, ...</i> |
| <i>jener¹⁸</i> | <i>jener, jene, jenes, ...</i> |
| <i>solch</i> | <i>solcher, solche, solches, ...</i> |
| Possessivpronomina | |
| <i>mein</i> | <i>mein, meine, meines, meiner, uns, unser, unsere, ...</i> |
| <i>dein</i> | <i>dein, deine, deines, deiner, euch, euer, eure, ...</i> |
| <i>sein</i> | <i>sein, seine, seines, seine, ... ihr [Pl.]</i> |
| <i>ihr</i> | <i>ihr, ihre, ihres, ...</i> |
| Personalpronomina | |
| <i>ich</i> | <i>ich, mich, wir, uns,</i> |
| <i>du</i> | <i>du, dich ihr, euch</i> |
| <i>er</i> | <i>er, ihn, sich sie [Pl.]</i> |
| <i>sie</i> | <i>sie, sich, ihr</i> |
| <i>es</i> | <i>es, sich</i> |

¹⁸ Normalerweise wird zwischen den Genera unterschieden, das DWB führt aber nur „jener“ als Lemma.

5 Annotation der Satzgrenzen

Legende zu den Beispielen:

| | |
|---------|---|
| <satz> | steht vor dem Anfang eines annotierten Satzes |
| </satz> | steht nach dem Ende eines annotierten Satzes |

5.1 Grundlagen

Die Bestimmung der Satzgrenzen stützt sich auf die vorangegangene Annotation der Wortarten.

Für die Einteilung werden folgende Kriterien zugrunde gelegt:

1. Der Fokus liegt auf kleinsten vollständigen Einheiten. Vollständig ist eine Satzeinheit, wenn sie ein Vollverb mit allen Ergänzungen umfasst. (Der Umgang mit Ellipsen ist gesondert geregelt.)
2. Abhängige Nebensätze werden nicht als eigenständig aufgefasst. Als ein Satz wird also der Hauptsatz mit allen abhängigen Nebensätzen gezählt.
3. Hauptkriterium für Hauptsätze ist die V2-Stellung, entsprechend für Nebensätze V-letzt-Stellung.
4. Die Orientierung erfolgt an syntaktischen Einheiten vor Sinneinheiten.

Die typischerweise als Satz annotierten Einheiten sind also durch V2-Stellung gekennzeichnete Hauptsätze mit allen ihnen zugehörigen (teilweise auch satzförmigen) Ergänzungen und Angaben. Aufgrund der Regel der kleinsten Einheiten bilden aber auch alle Vollverben, bei denen eine Zweitstellung des Vollverbs oder des finiten Hilfsverbs als möglich gedacht werden kann (bei fehlendem finitem Verb), einen eigenen Satz. Zusätzlich können auch eigenständige Nebensätze auftreten, wenn sie nicht in einen Matrixsatz eingebettet sind.

Durch vier Parameter lassen sich die zu annotierenden Einheiten in unterschiedliche Klassen einteilen.

Legende:

| | | | |
|---|-----------------------------|-----|--------------------------------------|
| + | Parameter ist erfüllt | +/- | Parameter ist irrelevant |
| - | Parameter ist nicht erfüllt | -/- | Parameter kann nicht bestimmt werden |

| | Enthält Vollverb | Enthält flektiertes Verb | Ist vom Stellungstyp V1 oder V2 | Enthält Subjunktion |
|--------------------|------------------|--------------------------|---------------------------------|---------------------|
| Hauptsatz | + | + | + | - |
| Kopulasatz | - | + | +/- | +/- |
| Nebensatz 1 | + | + | - | +/- |
| Nebensatz 2 | + | - | -/- | + |
| Ambiger Satz | + | - | -/- | - |
| Syntaktisch unklar | - | -/- | -/- | +/- |

Allgemein wurden nur Sätze annotiert, die mindestens ein deutsches Wort (mit Ausnahme von Eigennamen) enthalten. Ganze lateinische Sätze oder Absätze werden nicht annotiert. Allerdings

können in deutschen Sätzen auch fremdsprachliche Ergänzungen, z.B. Nebensätze, vorkommen und dann als Teil eines Satzes annotiert werden.

Bsp. 48, Alme 1630:

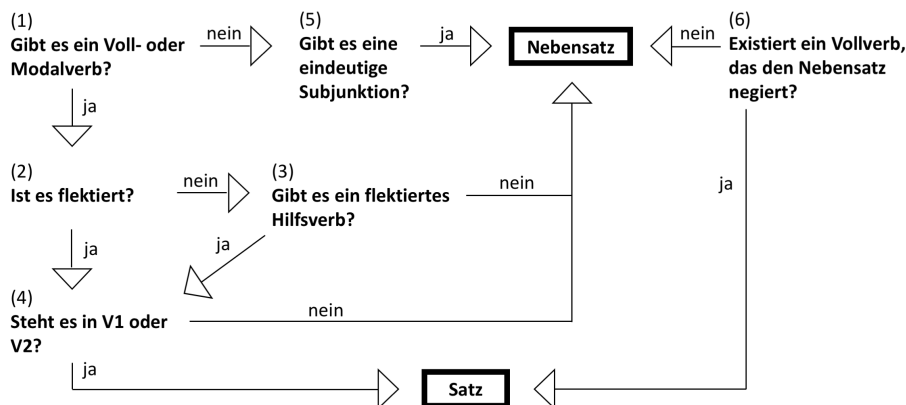
Ad 7. <satz>gestehet fugam et quidem vi tormentorum</satz>

Interpunktionszeichen sind als zum Satz, hinter dem sie stehen, zugehörig annotiert.

5.2 Flussdiagramm zum Vorgehen bei der Satzgrenzenannotation

Das folgende Flussdiagramm soll einen Überblick über das Vorgehen bei der Bestimmung der Satzgrenzen geben. Detaillierte Beschreibungen der einzelnen Annotationsregeln finden sich in den folgenden Abschnitten.

Fließschema zur Satzbestimmung:



5.3 Nebensätze und Hauptsätze

Grundsätzlich sind Hauptsätze durch V2-Stellung und Nebensätze durch Verb-Letzt-Stellung gekennzeichnet. Eine Ausnahme bilden Komplemente bei Verba dicendi und ähnlichen (s.u.).

Generell wurde die V2-Stellung immer dann angenommen, wenn sie syntaktisch möglich ist. So auch im folgenden Beispiel 49, wo dem Subjekt nur eine Reihung von Verben folgt. Da in den vorliegenden Daten auch eine vom nhd. Standard (finites Verb ganz rechts) abweichende Reihung in der rechten Klammer vorliegen kann, wurde hier die Analyse wie folgt vorgenommen:

Bsp. 49, Perleberg 1588:

Kontext:

<satz>war das der Rath, vnd die gerichtē alhie ihr darauff nachgetrachtet, </satz>

Satz:

<satz>vnd sie **wollen** greiffen laßen.</satz>

Analyse V2 nach dem Feldermodell:

| | | | |
|---------------------|---------------|------------|----------------|
| Vorvor- und Vorfeld | linke Klammer | Mittelfeld | rechte Klammer |
|---------------------|---------------|------------|----------------|

| | | | |
|---------|--------|------------------------------|----------------|
| vnd sie | wollen | [der Rath, vnd die gerichte] | greiffen laßen |
|---------|--------|------------------------------|----------------|

Alternativ mögliche, aber nicht verwendete Analyse als Nebensatz (V-Letzt):

| | | | |
|------------|---------------|--------------------------------|-----------------------|
| Vorvorfeld | linke Klammer | Mittelfeld | rechte Klammer |
| vnd | [das] | der Rath, vnd die gerichte sie | wollen greiffen laßen |

Bei elliptischen Sätzen kann es vorkommen, dass die V2-Stellung z.B. durch Fehlen des Subjekts oder auch des finiten Verbs nicht an der Oberfläche sichtbar ist. Auch hierbei wird im Zweifel für V2-Stellung entschieden. Die genaue Analyse solcher Sätze wird durch die im Abschnitt Ellipsen aufgeführten Richtlinien geregelt.

Eine weitere scheinbare Abweichung von der V2-Stellung kann bei Linksversetzung auftreten. Dies ist z.B. der Fall, wenn ein untergeordneter Nebensatz im Vorfeld durch ein Korrelat aufgegriffen wird und dadurch im Vorvorfeld steht. Der Nebensatz im Vorvorfeld wird nicht als eigener Satz analysiert.

Bsp. 50, Jever 1592:

```
<satz>vnd [Vorvorfeld: waß ir Sone konne ,][Vorfeld:Korrelat daß] habe sie Rixste Ihme auch
gelernt , </satz>
```

5.3.1 Relativsätze

Eindeutige Relativsätze werden wie andere Nebensätze in den übergeordneten Satz integriert. Bei als ambig getaggtten Relativpronomen (..._REL) wird bei einem nachgestellten Teilsatz immer eine Satzgrenze angenommen. Zudem wird hier immer das Feature „satz_ambig=true“ vergeben.

Bsp. 51, Baden-Baden 1628:

```
<satz>Sie habe ain Stiffsohn gehabt,</satz> <satz ambig=true>der
Bastian gehaisßen.</satz ambig=true>
```

Eine Ausnahme zu der Regelung, beim Fehlen des finiten Verbs einen eigenen Satz auszuzeichnen, bilden die mit „w“-Pronomen (z.B. PWAV, PWS, PWAT) eingeleiteten Sätze. Hierbei würde bei V2-Stellung eine Frage entstehen. Ist aus dem Kontext eindeutig erkennbar, dass es sich nicht um eine Frage handeln kann, wurde von einem Relativpronomen ausgegangen. Daher wurde in solchen Fällen kein eigener Satz annotiert, sondern er trägt entsprechend den genannten Regeln die Annotation „true“ für das Feature „satz_ambig“.

Bsp. 52, Alme 1630:

```
<satz ambig=true>so ihme einen goldgulden
geben welcher kuhdreck worden</satz ambig=true>
```

5.3.2 Komplementsätze mit Verbzweitstellung

In den Hexenverhörprotokollen finden sich viele Verben, die Komplementsätze annehmen, bei denen das Verb an zweiter Stelle steht, darunter Verba dicendi, die wörtliche oder indirekte Rede einleiten und auf die Wiedergabe der Verhörsituation oder den Einsatz von Frage- und Antwortkatalogen zurückzuführen sind. Es können darunter aber auch Verben gefasst werden, die eine Wahrnehmung bzw. eine Emotion versprachlichen und deren Gebrauch analog zu einem Verbum dicendi ist. Diese werden gleich behandelt.

Der zu diesen Verben (meistens Verba dicendi) gehörende Komplementsatz (häufig eine Redewiedergabe) wird auch bei V2-Stellung als abhängig gewertet und daher an den Satz angeschlossen. Folgen mehrere V2-Sätze, die als abhängig gewertet werden können, wird nur der erste als zugehörig angesehen. Die restlichen werden als eigenständigen Sätze annotiert:

Bsp. 53, München 1600:

<satz>Warumb er ... zu seinem weyb gesagt hab. Mein weyb was wöllen wier daraus machen,</satz><satz>Ich, sich wol, man sätzt vns nimmer aus.</satz>

Ebenso stellen Antwortpartikeln verbunden mit einem Verbum dicendi eine abgeschlossene Einheit dar:

Bsp. 54, Gaugrehweiler 1610:

<satz>Respondit Ja</satz> <satz>sie habe lme gesagt</satz>.

Wendungen wie „wie folgt“ sorgen hingegen dafür, dass ein folgender V2-Satz als eigene Satzeinheit gewertet wird und nicht als Komplementsatz.

Bsp. 55, Jever 1592:

<satz>vnd hatt darauf bekent wie folgt,</satz> <satz>Vor erst, Ihr eigenen Mutter habe eß Ir geleret</satz>

Statt eines Verbum dicendi kann auch ein Nominalisierungsverbgefüge auftreten, etwa „*machte ein Geständnis*“. Auch solche Konstruktionen können wie einfache Verba dicendi verwendet werden und nehmen dann einen V2-Satz als Komplement:

Bsp. 56, Perleberg 1588:

<satz>vnd darzu hette sie diese wortt ge-
braucht, dieb du soltt brengen das gerette,</satz>

Ähnlich wie Verba dicendi verhalten sich einige Präpositionalphrasen, die als Komplement einen V2-Attributsatz annehmen¹⁹:

Bsp. 57, Jever 1597:

¹⁹ Vgl. Abschnitt Attributsätze unter <https://grammis.ids-mannheim.de/systematische-grammatik/1114>, 22.05.2019

<satz> drauff were sie erschienen, in hoffnunge
sie were vnschuldich befunden worden, </satz>

5.3.3 Eigenständige Nebensätze

In seltenen Fällen kann ein durch V-letzt oder die Satzeinleitung eindeutig ausgezeichneter Nebensatz eigenständig eine vollständige syntaktische Einheit bilden. Dies ist daran zu erkennen, dass kein als Matrixsatz in Frage kommender Hauptsatz vorhanden ist.

Solche syntaktischen Einheiten finden sich vor allem in Listen von Anklagepunkten, beispielweise im Interrogationsteil der Protokolle.

Bsp. 58, München 1600:

1. <satz>Wie er haiß.</satz><satz> Wie allt er sey.</satz>
22. <satz>Ob er nitt dise gesölln auf den kirchtägen
vnd Jarmärkten neben seinen Sönen antreffen.</satz>

Bei eigenständigen Nebensätzen wird bei der Koordination von Vollverben keine Satzgrenze angenommen, wenn die Satzeinleitung nicht wiederholt wird.

Bsp. 59, München 1600:

25. <satz>wo sies bekommen und hintan haben</satz>

5.4 Einschübe

Einschübe (auch: Parenthesen) werden als eigene Einheit annotiert. Das hat zur Folge, dass sich Sätze (die auf der Ebene „satz_id“ durch eindeutige IDs gekennzeichnet sind) überlappen können.

Bsp. 60, Celle 1570:

<satz id="6269">do hab sie ihnen gebeten daß ehr den beutell in die hohe an einen bom
bringen wolte ,
<satz id="6354">das hette ehr gethan,</satz>
an welchem bom ehr hatt gehangen , biß an den heiligen Ostertag , </satz>

Bei Einschüben kann es sich – wie im obigen Beispiel – um einen eigenen Satz handeln oder auch um textuelle Gliederungszeichen, die nicht als Satz annotiert werden, aber dennoch den umschließenden Satz unterbrechen (s. Kap. Textuelle Gliederungszeichen).

Einschübe werden nicht als eigenständige Einheiten annotiert, wenn der eingeschobene Satz von einem anderen vollständig umgeben ist und (etwa aufgrund des Fehlens eines finiten Verbs) als „ambig=true“ annotiert ist. In solchen Fällen wurde der Einschub als Teil des übergeordneten Satzes annotiert.

5.5 Textuelle Gliederungszeichen

Zu den textuellen Gliederungszeichen gehören u.a. Aufzählungszeichen, Datumsangaben und Überschriften.

Hierbei lassen sich folgende Fälle unterscheiden:

- Textuelle Gliederungszeichen, die syntaktisch integriert sind, werden als dem Satz, in den sie integriert sind, zugehörig annotiert.
- Textuelle Gliederungszeichen, die nicht syntaktisch integriert sind, aber eine deutsche Nominalphrase enthalten, werden als eigenständiger Satz annotiert. Hierzu gehören auch die thematischen Angaben in den Randnotizen (etwa „Vieh“, „Totschlag“ in Baden 1628).
- In allen anderen Fällen (syntaktisch nicht integriert, keine deutsche Nominalphrase) werden die Einheiten ausgelassen und erhalten somit keine Annotation auf der Satzebene.

Als Gliederungssignale können auch Datumsangaben o.ä. fungieren, die nicht einzeln stehen, sondern syntaktisch in eine größere Einheit eingebunden sind. Es gibt daher Grenzfälle zwischen reinen textuellen Gliederungszeichen und Satzgliedern, die eine textgliedernde Funktion erfüllen.

Außerdem enthalten die Texte oft Gliederungssignale, die teils oder vollständig in lateinischer Sprache verfasst sind. Hierunter fallen insbesondere Verweise auf das juristische Verfahren, Abschnitte des bisherigen Protokolls (Frageteil, Anklagepunkte, Aussage) oder pronominale Verweise auf die beteiligten Personen. Das lateinische Wortgut ist eng mit den frühneuhochdeutschen Textelementen verbunden, wodurch auch hier Grenzfälle existieren.

Die folgenden Abschnitte geben Regeln zum Feststellen der syntaktischen Integration und der Sprache von Wörtern / Wortgruppen und nennen konkrete Beispiele.

5.5.1 Feststellen der syntaktischen Integration

Als syntaktisch integriert gelten Phrasen / Wörter, die

1. eine Leerstelle eröffnen, die von anderen deutschen Phrasen / Wörtern gefüllt ist, oder
2. eine syntaktische Funktion innerhalb eines Satzes ausüben.

Der erste Fall gilt z.B. bei Verba dicendi, die, auch wenn sie lateinisch sind und alleine vor einem V2-Satz stehen, als diesem zugehörig annotiert werden, da der folgende V2-Satz die vom Verb eröffnete Objektsleerstelle füllt (vgl. Abschnitt Komplementsätze mit Verbzweitstellung).

Bsp. 61, Güstrow 1615:

<satz>dicit das were rechtt vnnd wahr, </satz>

Der zweite Fall ermöglicht zuerst nur eine grobe Orientierung, da mögliche Gliederungssignale als fakultatives Adverbial nicht obligatorisch eine Funktion in einem Satz erfüllen müssen. Die in Perleberg häufig auftretende Phrase nach dem Muster „Vff den .XX. Articul“ könnte sowohl als zu einem Satz zugehörig gedacht werden („Vff den .XX. Articul hat sie gesagt ...“) als auch als eigenständig („Vff den .XX. Articul: sie hat gesagt ...“). Unterscheiden lassen sich beide Verwendungen durch die

Wortstellung: Im integrierten Fall steht die Phrase direkt vor dem finiten Verb, im nicht integrierten Fall steht das Subjekt zwischen dem finiten Verb und der Phrase.

Im Allgemeinen wurde die syntaktische Integration anhand des topologischen Feldermodells bestimmt: Eine Konstituente, die im Vorvorfeld steht und nicht im Vorfeld durch ein Korrelat wieder aufgenommen ist, gilt als syntaktisch nicht integriert. Außerdem ist ein Gliederungszeichen nicht integriert, wenn es im Nachfeld eines Hauptsatzes direkt vor einem Nebensatz erscheint (vgl. Bsp. 65 aus Georgenthal 1597 weiter unten).

Zu dieser allgemeinen Regel gibt es folgende Besonderheiten zu beachten:

- 1) Temporaladverbiale können auch gehäuft im Vorfeld auftreten. In diesem Fall wurden alle Temporaladverbiale zusammen als eine Konstituente analysiert.

Bsp. 62, Lemberg 1630:

<satz>Den 23. Aprilis 1630. Nach Mittag, Ist die Beclagtin Schmelichs
Eua wieder vorgepracht,</satz> <satz>vnd vf gegebene bedenckzeit erfragt
worden, waß sie sich nun bedacht, </satz>

Dies gilt jedoch nicht, wenn zwischen mehreren Adverbialen ein Absatz steht. In solchen Fällen wurden die Adverbiale, die vor dem Absatz stehen, als nicht integriert angesehen. Diese Adverbiale wurden nach den Regeln für Gliederungszeichen behandelt, d.h. vollständige NPs werden als eine eigene Satzeinheit annotiert. Zu beachten ist, dass auch hier lateinische NPs nicht gezählt werden.

Bsp. 63, Jever 1592:

Anno perge 92
<satz>Denn 10 Julii deß morgens vmb 6. schlegenn
ist Rixste Gralefs zuredede gestellet,</satz>

- 2) Bei Verba dicendi fehlt häufig das Subjekt im Hauptsatz. In diesem Fall wird angenommen, dass das Subjekt das Vorfeld besetzen würde, sodass das Gliederungssignal (hier „Vff den .10. Articul“) im Vorvorfeld steht und somit syntaktisch nicht in den Satz mit dem Verbum dicendi integriert ist:

Bsp. 64, Perleberg 1588

<satz>Vff den .10. Articul</satz>
<satz>(Sie) Sagt sie habe das nicht gethan,</satz>

Wenn das Subjekt vorhanden ist und nicht im Vorfeld steht, kann das mögliche Gliederungssignal in den Satz integriert sein:

<satz>Vff den .2.

Sagt sie also, das sie darbei gewesen,</satz>

- 3) In seltenen Fällen treten Gliederungszeichen nicht am linken oder rechten Rand einer Satzeinheit auf, sondern unterbrechen als Einschub eine syntaktische Einheit. Hierbei wird im Zweifel für eine syntaktische Integration entschieden. Nur in eindeutigen Fällen (17. im folgenden Beispiel) wird das Gliederungssignal eigenständig behandelt.

Bsp. 65, Georgenthal 1597

Nicht integriert, keine Nominalphrase, Einschubcharakter

<satz>Am 17. Martii hat die Christina Thymen Vor

Richter vnnd Schöpff[en], vff vorgehende befragunge

noch weiter güttlich gestand[en] </satz><satz ID="a17"> vnd beandt, </satz>

17.

<satz id="a17">Das sie Ihrer schwester sohn haß[en] Anschützens

verderbet, deme sie so viel gutter pferde, zu 40.

vnnd 50. Thalern vmbgepracht, das er darüber

zu einem armen Manne worden,</satz>

5.5.2 Freistehende Wortfolgen

Wenn eine Wortfolge nicht syntaktisch integriert ist, wurde sie nur dann als Satz annotiert, wenn sie eine (frühneuhoch-)deutsche Nominalphrase enthält.

Um als (frühneuhoch-)deutsch zu gelten, muss eine NP einen frnhd. Kopf aufweisen, bei dem Groß- und Kleinschreibung unterschieden werden kann. Ausgeschlossen ist dies bei Gliederungselementen, die nur aus einer Ziffer bestehen (z.B. „17.“) oder an der Position des Kopfs eine Ziffer haben („Vff den .2.“). Des Weiteren sind lateinische NPs ein Ausschlusskriterium. In beiden Fällen wird die entsprechende Wortfolge bei der Satzgrenzenannotation ausgelassen.

5.6 Ellipsen

Die Entscheidungen zur Annotation der Satzgrenzen beruhen auf dem Verb. Ein Vollverb konstituiert immer einen Satz. Diese Entscheidung wurde deshalb getroffen, weil in den Hexenverhörprotokollen

in vielen Fällen Auxiliarellipse vorliegt (vgl. Ausfall von finiten Hilfs- und Modalverben). Somit ist in einem Fall wie Bsp. 66 nicht klar entscheidbar, ob sich das Hilfsverb auf alle Vollverben bezieht oder nur auf das erste. Die kleinstmögliche Satzeinheit ist somit ein einzelnes Vollverb (häufig ein Partizip II).

Textstellen ohne Vollverb haben eine syntaktisch unklare Struktur und werden im folgenden Abschnitt behandelt (s. Kap. Ausfall von finiten Hilfs- und Modalverben).

Bsp. 66, Bamberg 1628:

<satz>aldta gessen</satz><satz> getrunken </satz><satz>vnd getantzet</satz>

Konjunktionen gehören zu dem Satz, dem sie vorangestellt sind. Auch Ergänzungen und Angaben des Vollverbs werden bei der Satzannotation zu dem Satz gerechnet. Sie können aber auch an der Oberfläche unvollständig sein. Neben der im nächsten Abschnitt behandelten Auxiliarellipse können folgende Formen der Ellipse vorliegen:

- Häufig fällt das Subjekt in einer Satzeinheit aus. Die fehlenden Ergänzungen wurden aus dem Kontext erschlossen.

Bsp. 67, Celle 1570:

<satz>vnd dasselbig hette Ihr Alheit vor 13 oder 14 Jaren gegebenn, widder die, die Ihr schalckheit tetenn, </satz> <satz>vnd habe eß verwarlich einhaltenn, In Ihrem vor schlossenen shap,</satz>

- Das ausgefallene Subjekt kann auch ein formales Subjekt sein. Zu ergänzen wäre im Beispiel 45, Georgenthal 1597 ein *es*.

Bsp. 68, Georgenthal 1597:

<satz>Viel weniger were der Drache Inn Ihrem hauß Auß vnd eingeflogenn,</satz><satz> (Es) sey Ihr auch vom Drachen Ihr lebetag nichts gepracht wordenn,</satz>

5.7 Ausfall von finiten Hilfs- und Modalverben

Eine Besonderheit der Hexenverhöprotokolle ist das Auslassen eines Auxiliars und damit der Ausfall des finiten Verbs. Hierdurch entstehen ambige elliptische Sätze, bei denen das ausgelassene Verb über Haupt- oder Nebensatz entscheiden würde:

Bsp. 69, Alme 1630:

<satz>dabei (**V2-Slot**) er bekenntlih verplieben (**V-Letzt-Slot**), alß folget ... </satz>

Das finite Verb könnte theoretisch an den markierten Stellen stehen, also sowohl V2 als auch V-Letzt einnehmen. Hierbei lassen sich zwei Fälle unterscheiden:

1. Die fragliche syntaktische Einheit befindet sich in einem anderen Satz (gemäß den oben definierten Regeln für die Satzannotation). Dies ist vor allem bei möglichen Relativsätzen der Fall.
2. Der fragliche Satz steht alleine.

Im ersten Fall wird so entschieden, dass es sich um einen Nebensatz handelt. Es wird also nur ein Satz annotiert. (Hervorgehoben sind in den folgenden Beispielen diejenigen Vollverben, die für die Entscheidung über die Satzgrenze ausschlaggebend sind.)

Bsp. 70, Riedlingen 1596:

<satz> Darauf sie zum Sibenden Georg Bauren Zu Möringen ain Schwarz Roß, mit ainem steckhen, daran sy thailß der Salben geschmiert geschlagen, das eß sey **gestorben**.</satz>

In zweiten Fall wird nach dem Prinzip der kleinsten möglichen Einheiten so entschieden, dass ein Hauptsatz annotiert wird:

Bsp. 71, Ellingen 1590:

<satz>dem [V2: hat] sie ein salben In ein Pier **gerurt**</satz>

<satz>das [V2: hat] er **gethrunken**</satz>

<satz>davon wegen [V2: hat] er außdorren vnndt sterben **muessen**</satz>

Ist eine eindeutige Nebensatzeinleitung vorhanden, wurde auch bei fehlendem finitem Verb Verbletzstellung angenommen.

Bsp. 72, Gaugrehweiler 1610:

<satz>Will nicht gestehen **das** sie vnd die

schaffhirtin die knie gegen einander

gekeret (VL: haben)</satz>

Weiterhin wird, wie unter Relativsätze erklärt, bei mit „w“-Pronomen (z.B. PWAV, PWS, PWAT) eingeleiteten Sätzen von Relativsätzen ausgegangen, wenn das finite Verb fehlt. Dementsprechend wird z.B. *Welliche* in Bsp. 73 als Relativpronomen interpretiert:

Bsp. 73 Jever 1592:

<satz>Darauf will (V2) sie Ihr

Sele opffern, auch Lebenn vnd sterbenn, Welliche

worte sie auch Rixstenn, in dem hinaußgehende,

weiln sie genow vor ir außreden konde, nach-

geschreiet</satz>

Im Zweifel wird aber für V2 entschieden, etwa bei dem folgenden Anschluss mit *inmaßen*:

Bsp. 74, Alme 1630:

<satz>Wahr das er den 7 Junii von Durriten Elßen vor ein Zauberer besagt, </satz>

<satz>Inmaßen die selb druff bestendigh hingerichtet.</satz>

Bei Kopulasätzen mit mehreren Prädikativen ohne Wiederholung des Kopulaverbs erfolgt im Allgemeinen keine Trennung in mehrere Teilsätze (vgl. *krumb, vnd noch bey dem Vattern*). Wenn ein Wechsel des Subjekts vorliegt, wird allerdings eine Satzgrenze angenommen (*die ander*):

Bsp. 75, Baden-Baden 1628:

<satz>die ain ist die Wöberin im Prüehel,</satz><satz>die ander aber noch ledig. krumb, vnd noch bey dem Vattern, Rudolph Schlosßers ledige dochter</satz>

5.8 Abschnitte mit syntaktisch unklarer Struktur

Ist kein Vollverb vorhanden – etwa bei Namenslisten –, so wurde für die Bestimmung der Satzgrenzen überprüft, ob durch das Auslassen der betreffenden Passage in der sie umgebenden Struktur eine unvollständige Einheit entstehen würde. Wenn ja, gilt die Passage als syntaktisch integriert und bildet bei der Annotation eine Einheit mit der sie umgebenden Struktur. Ansonsten wurde sie als Abschnitt mit syntaktisch unklarer Struktur angesehen und mit einer eigenen Satzannotation versehen, sobald sie die Mindestanforderungen erfüllt (vgl. Textuelle Gliederungszeichen). Beispielsweise fehlt der Einheit *Oben an einer tafel wehre gesessen* in Bsp. 76 das Subjekt, wenn man die folgende Reihe der Namen auslassen würde. Eine syntaktische Integration liegt auch vor, wenn die zu überprüfende Passage durch ein Korrelat („*volgendte*“) innerhalb des Satzes wieder aufgenommen wird.

Bsp. 76, Bamberg 1628:

<satz>Oben an einer tafel

wehre gesessen, **Cantzler Burgermeister**

Neydekher, doctor Geörg Adam haan</satz>

<satz>vnd volgendte wehren auch darbey

gewesen (...)</satz>

In Bsp. 77 ist hingegen *Bainschrauben* nach den obigen Regeln syntaktisch nicht integriert und bildet daher eine eigene Einheit.

Bsp. 77, Bamberg 1628:

<satz>empfindtet keinen schmerzen

im daumenstokh.</satz>

<satz>Bainschrauben,</satz> <satz>will gantz nichts
gestehen, könne vnd wisse nichts</satz>

6 Phrasenannotation

Im Rahmen des SiGS-Projekts wurden alle deutschsprachigen Nominalphrasen und Präpositionalphrasen auf der Annotationsebene *w_func* annotiert. Sie tragen Annotationen, die Auskunft geben über ihre Einbettung in einen Satz, ihren Phrasentyp, syntaktische Funktion und semantische Rolle. Bei diesen Annotationen wurde jeweils mit einem minimalen Set an Values gearbeitet: Die syntaktische Funktion (*synfunc*) ist entweder *Subjekt*, *Objekt*, *adverbiale Bestimmung*, *ambig* oder nicht bestimmbar (*NA*). Bei der semantischen Rolle stehen *agens*, *patiens*, *experiencer*, *stimulus*, *stativ*, *NA* und *ambig* zur Auswahl. Im Falle von *embedded* ist jeweils ein Boolesches „true“ oder „false“ auszuwählen und die Art der Einbettung wird durch die Ergänzung von „_attr“ für Attribute oder „_head“ für Phrasenköpfe spezifiziert.

Annotiert werden alle Nominalphrasen (NP) und Präpositionalphrasen (PP). Auch die in Präpositionalphrasen enthaltenen Nominalphrasen werden annotiert. Zu den NPs werden auch alle Pronomen gezählt.

6.1 Präpositionalphrasen (PP)

Auf der *w_func*-Ebene werden Präpositionalphrasen als Ganzes annotiert und beinhalten in der Regel eine in sie eingebettete Nominalphrase. Dies gilt nicht, wenn eine deutsche Präposition eine lateinische Nominalphrase regiert (vgl. Kap. 6.3.2 Deutsche PP mit lateinischem Wortmaterial).

(78) „*hette sie [in [8. Wochen]] nit gehabt*“ (Stralsund 1630)

Es werden nur Präpositionen annotiert, die eine untergeordnete Nominalphrase aufweisen. Wenn Präpositionen alleine auftreten, werden sie nicht annotiert.

(79) „*vnd hette ir Bole mit Ir darauf geseßen, vnd sie were **hinder** ... vnd hette Ihme vmbe daß leib gefaßet*“ (Jever 1592)

Wird eine Präpositionalphrase durch eine Präposition-Artikel-Klise (*aufm*, *am*, *im*, *durchs*) geöffnet, richtet sich die Annotation nach der syntaktischen Tokenisierung: Die Öffnung der NP geschieht immer an der Stelle, an der die Wortteile auf der *syntoken*-Ebene getrennt sind.

(80) „*Baltzar Meier hette daß mall noch kein gut [I[m **hoffe**]] gehabt*“ (Celle 1570)

6.1.1 Präpositionalphrasen ohne eingebettete Nominalphrasen

Präpositionalphrasen, die keine Nominalphrase enthalten, werden in der Regel nicht annotiert.

Dies betrifft zum Beispiel PPs mit eingebetteten Adverbien wie *von dannen*, *von hinnen*, *zue mehrmaln*, *biß dahero*, *vor deme*, *dauon wegen*, *dem nach* und *zu forderst*.

- (81) „*hab Ir Ir buol von dannen, biß
zu Irem Garten gehn Meringen daß glait
geben*“ (Riedlingen 1596)

In einigen besonderen Fällen werden bei SiGS allerdings Präpositionalphrasen annotiert, die *keine* Nominalphrase enthalten.

Dabei handelt es sich um Pronominaladverbien (Kap. 6.1.1.1), zusammengeschriebene Reziprokpronomina (Kap. 6.1.1.2), deutsche PPs mit lateinischem nominalem Kopf sowie PPs, in denen *ja* bzw. *nein* nicht attribuiert sind.

- (82) „*Wünden beede die händ, das mädlin wainet, sagt [von Ja]_{phrasentyp=PP_satz}, vnd seje Je war, die
muetter [von Nain]_{phrasentyp=PP_satz}, vnd betheürt Ihr vnschuld,*“ (Augsburg 1625)

6.1.1.1 Pronominaladverbien

Pronominaladverbien werden annotiert, wenn sie eine adverbiale Funktion ausüben. Es werden nur solche Pronominaladverbien annotiert, bei denen das Substantiv, das sie ersetzen, aus dem Kontext rekonstruiert werden kann. Pronominaladverbien, die für einen (Teil-)Satz stehen, werden nicht annotiert. Das gesamte Pronominaladverb wird - unabhängig davon, ob Getrennt- oder Zusammenschreibung vorliegt - als eine Präpositionalphrase ohne eingebettete Nominalphrase annotiert.

- (83) „*Ob ehr ihr nicht anfangs einen Reichstaler gegeben, Näscherei [dafuer] keuffen, vndt ihnen
zue Gaste bitten soltte.*“ (Ostrau 1628)
- (84) „*Hanß Freybergs Garttenhäußlein, [darin] bemelte ziesingische allein damals gewest*“
(Rosenburg 1618)

Pronominaladverbien, die Teil eines Verbs oder Funktionsverbgefüges sind, werden hingegen nicht annotiert:

- (85) „*hatt sie embsiglich
gebetten, sie herunder zu laßen*“ (Lemberg 1630)

Werden die beiden Bestandteile des Pronominaladverbs durch weiteres Wortmaterial voneinander getrennt, so wird eine `func_id` vergeben, um die beiden Teile miteinander zu verbinden.

6.1.1.2 Reziprokpronomina einander

Reziproke Pronomen wie *voneinander*, *nacheinander*, *gegeneinander* oder *miteinander* werden als Phrasen annotiert. Wenn sie getrennt geschrieben werden, wird eine PP mit einer eingebetteten NP annotiert. Wenn das Reziprokpronomen zusammengeschieden wird, wurde das gesamte Pronomen als eine PP ohne Einbettung annotiert.

(86) „wen sie [van [einander]_{phrasentyp=NP_praep, embedded=true-head}]phrasentyp=PP_satz *gescheiden wollen*“ (Jever 1592)

(87) „Caspar wittich, Clauß Gebhardt,
so beede Gärtner, wehren nach
solchem widerumb [voneinander]_{phrasentyp=PP_satz}
kommen.“ (Bamberg 1628)

6.1.2 Präpositionalphrasen, die nicht annotiert werden

Einen Sonderfall stellen die Präpositionen *bei* und *in* in der Bedeutung 'ungefähr, circa' dar. In solchen Fällen wurde keine Präpositionalphrase annotiert, sondern nur die enthaltene NP:

(88) „wie dann schier das ganz dorf aus-
vnd bei [70. Persohnen] gestorben“ (München 1600)

(89) „seye sie vnd Ire geSpillen in [die 20 mallen] In Keller gefahren“ (Ellingen 1590)

Die Phrasen zu *nichte* und zu *eigen* werden hingegen annotiert. Da diese Präpositionalphrasen Teil des jeweiligen Verbs sind (*etwas zu nichte machen*, *etwas zu eigen bringen*), bekommen die PPs folgende Annotationen: *synfunc=NA*, *semantische Rolle=keine_rolle*, *embedded=false*.

(90) „da Sandtman Ihr erstattung der [zu [nichte]] gemachten garben gethan“ (Güstrow 1615)

(91) „hette ir Mann, ir den
Milch Eimmer im ham [zu [eigenn]] gebracht“ (Jever 1592)

6.1.3 Namen mit Präpositionalphrasen

Ortsnamen, die mit einer Präposition an den Eigennamen angebunden sind, werden als Teil des Namens annotiert. Die Unterscheidung, ob es sich bei der PP um ein Attribut zum Eigennamen oder eine adverbiale Bestimmung handelt, wird in Abhängigkeit von dem dazugehörigen Verb getroffen. In Sätzen ohne Verb (z.B. bei Aufzählungen und Listen) kann die PP keine adverbiale Bestimmung sein und wird als Attribut zum Eigennamen annotiert. Wenn die PP ambig ist zwischen Attribut und adverbialer Bestimmung, wird die PP als adverbiale Bestimmung interpretiert.

(92) „[Lubbe [zu [Gaßmuß]]] sei Trummenschleger gewesen“ (Jever 1592)

(93) „diese Alheit sei Jetzo [bei [Ihrem Sohen] [dem kuhirtenn]]
[zu [Wesendorff]],“ (Celle 1570)

6.1.4 Koordination von Präpositionen

Wenn sich mehrere Präpositionen auf dasselbe Substantiv beziehen, wurde eine PP annotiert, die eine NP enthält.

(94) „vnd wolte [in die
oder vor [die augen]] sagen“ (Osnabrück 1636)

- (95) „*hab er dieselbig nacht, vnd [biß erst a[m morgen früe]], nit erwachen khinnen*“ (Riedlingen 1596)

6.2 Nominalphrasen (NP)

Es werden ausnahmslos alle Nominalphrasen mit deutschem Kopf annotiert, unabhängig von ihrem Satzgliedstatus. Das schließt sowohl NPs ein, die Teil eines Satzes sind, als auch Randsatz oder Listen von Namen. Es wird jeweils die gesamte NP inklusive aller Attribute annotiert.

6.2.1 Koordination und Apposition von Nominalphrasen innerhalb einer Präpositionalphrase

Treten in einer Präpositionalphrase Nominalphrasen in Koordination oder Apposition auf, so werden beide als zu der PP gehörig annotiert und erhalten jeweils eine eigene NP.

- (96) „*In [Ire ställ] vnd [häuser]] nit mehr einzegehen*“ (Riedlingen 1596)

- (97) „*Ihme gefolgt, ohne das sie [ahn [got] [den hern]] gedacht*“ (Hamm 1592)

- (98) „*Doch [[[vff [fernner Anhallten [deß Henckers]] vnd [befragung [deß Richters]]] gesagt,*“ (Georgenthal 1597)

Die koordinierten Nominalphrasen werden auch dann als einzelne Phrasen annotiert, wenn sie einen gemeinsamen Determinierer oder ein gemeinsames Attribut haben:

- (99) „*Sie wisse [von [keiner zauberi] oder [vorgift]], die vff oder vnden in der Erden vergrabenn*“ (Celle 1570)

Für die Koordination von Genitivattributen siehe Kap. 9.2 `embedded=true-attr`.

6.2.2 Relativa, „so“ als Relativpronomen

so wurde, wenn es als Relativpronomen auftritt, als Nominalphrase annotiert. Es wurde immer eine Phrasenannotation vorgenommen, wenn die POS-Annotation das Wort als *DPRELS* ausweist oder eine ambige POS-Annotation wie *ADV_REL* vorliegt.

- (100) „*die Alte Klaggesche mit ihrer Tochter trinen [so]_{phrasentyp=NP_satz, synfunc=subj, sr=stativ} zum scharpffenbergh wonnet*“ (Alme 1630)

6.2.3 Fragepronomina (PWAV)

Es wurden nur Fragepronomina annotiert, die für ein Substantiv bzw. eine NP stehen, wie bspw. *wer* und *was*. *Wie* und *warum* wurden hingegen nicht annotiert. Auch komplexe PWAV wie *wofür* oder *woran* werden nicht annotiert.

- (101) „*[wehr]_{phrasentyp=NP_satz, synfunc=subj, sr=agens} zue dier kömbt*“ (Ostrau 1628)

- (102) „*Gefragt, wofur das gutt wehre*“ (Stralsund 1630)

6.2.4 Formen mit „mal“

Konstruktionen mit *mal* wurden immer dann annotiert, wenn sie getrennt geschrieben auftreten. Wenn sie als ein Wort geschrieben wurden, wurden sie nicht als Nominalphrase annotiert. Wird *mal* von einem nicht-flektierbaren Wort attribuiert, so wird dies nicht als Teil der Phrase annotiert (vgl. Kap. 7.2).

- (103) „*der Geist hat ihr die zungen gehalten darumb sie [viel Mal] außgespitzt*“ (Georgenthal 1597)
- (104) „*vnd were des korns Jedesmal bey einem Achtel gewesen*“ (Georgenthal 1597)
- (105) „*daß widerholle Sie hiemit noch [mah]l darbey Sie biß ann Ihr End verbleiben wolle.*“ (Leonberg 1641)

In Fällen, bei denen *mal* durch einen Artikel eindeutig als nominal verwendet zu erkennen ist, spielt die Zusammenschreibung allerdings keine Rolle. Solche Konstruktionen wurden als Nominalphrase annotiert.

- (106) „*[Daß Drittemahl] seyen ihrer ohngeuehr dreyssig vff dem Hochheimer brachfeldt gewesewesen zu Frühlings zeitt*“ (Höchst 1631)

6.2.5 Komplexe Relativpronomen („was für“)

Neben den einfachen Relativpronomen („der“, „welche“ etc.) gibt es auch komplexe Umschreibung wie „was für“. Diese wurden annotiert und bei Getrenntschreibung gegebenenfalls per *func_id* zusammengeführt. Das dazugehörige Substantiv wurde als Teil des zweiten Phrasenteils annotiert. Die gesamte Phrase wurde als NP annotiert.

- (107) „*[Was]_{func_id=mittel} Sie [vor mittell]_{func_id=mittel} darzu gebrauchte*“ (Leipzig 1640)

6.3 Umgang mit lateinischem Wortmaterial

Lateinisches Wortmaterial wird nur dann annotiert, wenn es Teil einer Phrase mit deutschem Kopf ist. Dabei richtet sich die Entscheidung, ob ein Wort als lateinisch oder deutsch gilt, nach dem Grad der Entlehnung: Wenn ein lateinisches Wort eine deutsche Endung aufweist (z.B. *Inquisitin*) oder durch deutsche Affixe deriviert ist, dann gilt es als hinreichend deutsch. Trägt ein Wort eine lateinische Flexionsendung (z.B. *Junii*), dann wird es als lateinisches Wort interpretiert. Wenn ein Wort weder eine eindeutig lateinische, noch eine eindeutig deutsche Endung aufweist, dann gilt es als deutsch.

- (108) „*Wie [inquisitin] heiße,*“ (Leipzig 1640)
- (109) „*Sagt ihme ingleichem das Er im hautschmohr bey einem tantz geweßen, aber zuuor sey **Sancta Hostia** eingegraben worden.*“ (Bamberg 1628)

6.3.1 Eigennamen

Für Eigennamen gilt eine Sonderregel: Diese wurden annotiert, auch wenn es sich um latinierte Formen des Namens handelt oder sie lateinisch konjugiert werden. Latinisierungen des Namens *Jesus Christus* wurden allerdings nicht als deutsche Nominalphrasen annotiert.

(110) „[Vff [obbeschehene Ausßag [**Affræ Riglerin**]]“ (Schweinfurt 1616)

(111) „vnd Sagt das sie keinen bolen hab dan **Jesum Christum**“ (Erkelenz 1598)

6.3.2 Deutsche PP mit lateinischem Wortmaterial

Im Falle einer deutschen Präposition, die ein lateinisches Substantiv regiert, wird eine Präpositionalphrase, aber keine Nominalphrase annotiert:

(112) „alß sie gfraget Ob sie auch [**Jegen obbemelte indicia**] etwaß erheblichs so zu Ihrer defension dienlich sein möegte“ (Werl 1630)

Wenn eine Nominalphrase mit einem lateinischen Kopf in eine Phrase mit einem deutschen Kopf (z.B. eine PP) eingebettet ist, dann wird die PP annotiert, die NP aber nicht (Bsp. (113)).

Wenn eine Phrase mit deutschem Kopf als Attribut zu einer Phrase mit lateinischem Kopf steht, so wird die deutsche Attributphrase annotiert (Bsp. (114)).

(113) „[in [mein des Notarii gegenwart]]“ (Ostrau 1628)

(114) „alle drey verhoffte, [yber die [vonn Für [Cantzley] vsß] yberschickhte **Interrogatoria**, vnnd [andere wider Sie einkommene verdachts Puncten],] Ernstlich Examiniert vnnd verhört“ (Leonberg 1641)

6.3.3 Lateinische PP mit deutschem Wortmaterial

Wenn eine Phrase von einer lateinischen Präposition regiert wird und ein deutsches Substantiv enthält, wurde die deutsche Phrase annotiert, nicht aber die lateinische. Im Zweifelsfall wurde die Präposition *in* als lateinisch interpretiert (Bsp. (116)).

(115) *Inquisitional articul contra [Margrithen Petzschin, [von [Jeßnitz]]]*“ (Leipzig 1640)

(116) „**in specie** von Statz Wehrmanne berichtet worden,“ (Hildesheim 1628)

6.3.4 Deutsche Phrasen in lateinischen Sätzen (NP_lat und PP_lat)

Kommt in einem Satz mit lateinischem Vollverb eine deutsche NP oder PP vor, so wird diese annotiert, sofern sie einen deutschen Kopf aufweist. In solchen Fällen wird die NP bzw. PP als *phrasentyp=NP_latein* bzw. *phrasentyp=PP_latein* annotiert (vgl. Kap. 8). Dabei wird nicht unterschieden, ob das lateinische Wortmaterial tatsächlich einen Satz mit Vollverb bildet oder nicht.

(117) „26. [...] *Terreatur, et [nach[gestalt [seines erzaigens]]]*_{phrasentyp=PP_lat} *etiam super hoc torquendus est.*“ (München 1600)

(118) *“Abschrift Mößkhürchischen Ampts Prothocoll den 21. May Anno 1644 Præsentibus [herrn doctor Riescher]*_{phrasentyp=NP_lat} *[oberamtmanß]*_{phrasentyp=NP_lat} *[Stattmaman Georg Mierenß]*_{phrasentyp=NP_lat} *[Stattschreiber hertzogs]*_{phrasentyp=NP_lat} *vnnnd [kastenvogt Pfisters]*_{phrasentyp=NP_lat}“ (Messkirch 1644)

6.4 Ziffern als Phrasenkopf

Phrasen, deren Kopf eine Ziffer ist, wurden nicht als Nominalphrase annotiert (Bsp. (119) vgl. Kap. 7.1.2 zum Auftreten von Ziffern als Attribut). Tritt eine Ziffer als Kopf einer in eine Präpositionalphrase eingebetteten Nominalphrase auf, erhält die PP eine Annotation, die NP nicht (Bsp. (120)).

(119) *„bey deme sye [.18. khünder], von denen noch .9. im leben sein mochten“* (Reichenberg 1653)

(120) *„[Zum 14.] hab sein Ammanß zu Möringen hauß-frau, vngeuehr vor dreyen Jaren, ein Wesch vor irem hauß gehabt“* (Riedlingen 1596)

7 Bestimmung der Phrasengrenzen

In den folgenden Kapiteln sind die wichtigsten Regelungen zur Annotation der Phrasengrenzen auf der Annotationsebene *w_func* aufgelistet. Sie bilden die Grundlage für alle inhaltlichen Annotationen auf dieser Ebene.

7.1 Umgang mit Attributen

Attribute werden prinzipiell als Teil der von ihnen attribuierten Phrase annotiert. Das gilt allerdings nur, wenn das Attribut potentiell flektierbar ist.

7.1.1 Präpositionalphrasen als Attribute

Präpositionalphrasen, die als Attribute fungieren, wurden als Teil des entsprechenden Satzglieds annotiert (Bsp. (121) und (122)). PPs, deren Status als Attribut unklar ist, werden als eigenständige Satzglieder annotiert (Bsp. (123) und (124)). Sie erhalten dann die Annotation *synfunc=pot_PPO*. Das gilt auch und insbesondere für Ortsbezeichnungen. Auch hier wird in der Regel die kürzere Phrase bevorzugt (Bsp. (125)).

- (121) „Daruff [**ein Junffer [mit [bögelen]]]** Alß ein
Von Adel in blauwen kleidern Vnd
huuen zu ihme kommen“ (Alme 1630, bögelen: ‚Ringlein‘)
- (122) „ich sage [**Gott [i[m himel]]]** ab“ (Bamberg 1628)
- (123) „Sonsten hette sie ihr auch [**vo[m roten Rinde]] [den peserich]**, so pulverisiret, in wein eingegeben;“ (Stralsund 1630)
- (124) „die Salbe die Sie gehapt, hette Sie [**i[m hauß]] [auff [der kammer]] [auff[m balken]]** stehen gehapt“ (Wittgenstein 1629)
- (125) „vnd Spilltisch in Leezelten geht, Als [**zu [wistet]] [bey [wört]]** so Pfälzisch, [zu [Irlbach]], [zu [Altenpuech]] [ausserhalb [Straubing]]. [zu [Michelspuech]] ain Meil wegs [von [Plädling]], Item [zu [Plädling]], vnd [zu [osterhofen]]“ (München 1600)

7.1.2 Ziffern als Attribute

Attributiv verwendete Ziffern wurden als Teil des entsprechenden Satzglieds annotiert:

- (126) „dergleichen zu Alsentzen [**2 nacht]** beim scheffer vffgehaltenn“ (Gaugrehweiler 1610)

Konstruktionen vom Typ *xy Jahre alt sein* werden nicht annotiert:

- (127) „ist **55. Jahr** alt“ (Bamberg 1628)

7.1.3 Komplexe Zeitangaben (*gestern Abend*)

Bei komplexen Zeitangaben, die aus einem Adverb wie *gestern, heute* etc. und einem Substantiv wie *Morgen, Mittag* oder *Abend* bestehen, wurde die nominale Zeitangabe als Attribut zu *gestern* oder *heute* interpretiert.

- (128) „Sie hette nicht darauff liegen können
wie sie dan auch den Wulst [**gestern [abent]]**
auch darab schneiden wollen“ (Osnabrück 1636)

7.2 Quantifizierer

Quantifizierer wurden nicht als zur NP gehörig annotiert, wenn sie nicht flektierbar sind. Das heißt, *viel* und *alles* wurde als Teil der NP annotiert, aber Wörter wie *allerhand, mehr, bei, etwa, circa, so, gar* oder *genug* nicht. Auch *nichts* wurde als Attribut annotiert.

- (129) „Sie solle sein sein, thun waß Er wölle, so wolle Ihren [**alles] genueg geben.**“ (Baden-Baden 1628)
- (130) „aber auch nur [**Ross koth]** geweßen“ (Baden-Baden 1628)
- (131) „Bey den dántzen haben sich Jeder zeit [**Vil persohnen]** befunden“ (Messkirch 1644)
- (132) „vnd hette damit [**mannich menschen]** vnd [**viehe]** geholffen“ (Perleberg 1588)
- (133) „vnd wust [**nichts bößes]** von Ihnen“ (Helmstedt 1580)

7.3 Koordination von Satzgliedern

Bei Koordination von Satzgliedern mit „und“ oder „oder“ wurden zwei Satzglieder annotiert. Im Beispiel werden also *Gott* und *Alle seine heiligen* als einzelne Objekte annotiert. Das gleiche gilt auch für Konstruktionen mit „weder ... noch“ und „sowohl... als auch“.

(134) „*Ob sie mit demselben [Vnzucht] und [Hurerej] getrieben.*“ (Ostrau 1628)

Bei Koordination von Attributen, die zu demselben Phrasenkopf gehören, wurde eine Phrase annotiert:

(135) „*vnd also [seine vnd ihre vnschult] geklagtt*“ (Lemgo 1632)

Bei koordinierten Genitivattributen wurden hingegen zwei Phrasen annotiert:

(136) „*[[seines sohns] Ausßag] vndt [bekandtnus]*“ (Schweinfurt 1616)

7.4 Koordinationsellipse

Bei koordinierten Komposita wird teilweise ein Bestandteil in der Koordination weggelassen. In einem solche Fall werden die koordinierten Komposita zusammen als eine Phrase annotiert.

(137) „*Wer In disen vnnd obigen Iren mißenthatten Iren gespielen, Rathgebern, vnd mitgehilffen [von [Manns vnd weibs Persohnen]] gewesen,*“ (Günzburg 1613)

(138) „*Whar das nachgesetzte Als die Steuensch, Trine in den Boisen Gert Berendtz, Herman Osthoff vnd Peter Moller Sie Inquisitam [In vnd Außerhalb [der tortur]] fur eine zaubersche denunciirt*“ (Werl 1630)

7.5 Appositionen

Bei Apposition von zwei NPs wurden beide Phrasen als einzelne Satzglieder annotiert. Beide Phrasen erhalten dieselben Satzglied- und Rollenannotation.

(139) „*Aber sie habe diese nacht solches [vnserm herenn]_{synfunc=obj, sr=patiens} [Gott]_{synfunc=obj} abbetten das er Ihr solches verzeihenn wolle,*“ (Georgenthal 1597)

(140) „*Vff ihre Vnbestendigeitt ist M eister dietrich beholen [sie]_{synfunc=obj} [Peterschen]_{synfunc=obj} [inhafftirten]_{synfunc=obj} anzugreifen*“ (Lemgo 1632)

(141) „*[Enneke]_{synfunc=subj} [die Johanverschen]_{synfunc=subj} [mit [ihrer Tochter]_{synfunc=obj, embedded=true-head} [Trinen]_{synfunc=obj, embedded=true-head}]*“ (Alme 1630)

Wenn eine Konstruktion die Lesart als Apposition oder NP mit Demonstrativpronomen o.ä. zulässt, erfolgt die Annotation nach der POS-Annotation auf der w-Ebene (vgl. Kap. 4.2 PoS). Ist das fragliche Wort als attributiv annotiert (im Beispiel *PDAT* = attributives Demonstrativpronomen), wird die Konstruktion als eine zusammengehörige Phrase annotiert. Ist auf der w-Ebene die Wortart als *PDS*

(=substituierendes Demonstrativpronomen) o.ä. angegeben, wird die Konstruktion als Apposition gewertet und entsprechend zwei NPs annotiert.

(142) „Wahr, daß Meierinck auch [**dieselbig hille**] fur eine zeuberInne geschulden,“ (Ahaus 1608)

7.5.1 Apposition bei Mengenangaben

Bei Mengenangaben, die keine Interpretation als partitiven Genitiv zulässt, wird keine Apposition angenommen. Hierunter fallen vor allem Mengenangaben, bei denen der Kopf der Phrase keine Genitivendung zeigt. (*ein Stück Butter* vs. *Ein Haufen Goldes*).

(143) „da habe er Ir [**ain stuckh weiß gelt**] geben, so einem Taler gleich gesehen“ (Riedlingen 1596)

7.5.2 Enge Appositionen mit geographischen Namen

Bei Appositionen von geographischen Namen mit Wörtern wie *Stadt/Dorf* etc werden zwei getrennte NPs angenommen.

(144) „Zum .18. seye sy auf ainem geschmirbten steckhen, auf einen berg [**bey [dem dörrflin]**_{embedded=head} [**hohenberg**]_{embedded=head}], auf der Alb“ (Riedlingen 1596)

7.5.3 Enge Apposition „mit Namen“

Die Konstruktion [*mit Namen X*] wurde als enge Apposition interpretiert. Anders als bei der Apposition mit geographischen Namen wird bei Konstruktionen mit *mit Namen* der genannte Name als Attribut in einer eigenen eingebetteten NP annotiert.

„Daruff sein Bule zu ihm kommen in [*in [schwarzen seidenen kleideren]*], [**mit [Nahmen [luitger]**_{embedded=true-attr}_{embedded=true-head}_{embedded=false} *in weibes gestalt welche ihm 1 thaler geben*“ (Alme 1630)

„Berichtet sonst, das [**eine Frawe [mit [nahmen [Wathowersche]**_{embedded=true-attr}_{embedded=true-head}_{embedded=true-attr}_{embedded=false} *solche kunst mit der Cristallen gewußt habe,*“ (Stralsund 1630)

7.5.4 Kombination aus Anrede und Titel

Bestimmte Titel werden als Teil des Namens betrachtet, sodass hier keine Apposition angenommen wird. Dazu gehören Formulierungen wie *Herr Doktor XY, der Herr Bürgermeister* usw.

(145) „haben Ihre Gestrengen [**der herr Amptman**], [...], sye mit der tortur... zugreyffen [...] lassen“ (Höchst 1631)

(146) „[**her Amtman Keßell**] fragt in specie wie es eine bschaffenheit mit dem gurtell“ (Linz 1631)

7.6 Komposita

Komposita werden prinzipiell als eine zusammengehörige NP annotiert, auch wenn die Konstituenten nicht zusammengeschrieben werden (siehe Annotationsebene *syntoken*, Kap. 3.2.1.1 Komposita).

Straßennamen werden normalerweise als ein Kompositum annotiert, auch wenn sie strukturell ambig sind (Bsp. (148)).

(147) „*Hernach der behafftin [Ihr **bekhendtnuß puncten**] abermhals vurgelesen*“ (Hamm 1592)

(148) „*wie N. wilde [in [der **Francken straßen**]] etliche silber kleider vnd dergleichen sachen*“ (Stralsund 1630)

Konstruktionen, bei denen aufgrund von ambigen Kongruenzverhältnissen nicht eindeutig zu bestimmen war, ob es sich um ein Kompositum oder eine Konstruktion mit vorangestelltem Genitivattribut handelt, sind auf *der syntoken*-Ebene zusammengezogen und tragen die Annotation *ambig=true*. Diese potentiellen Komposita werden auch auf Phrasenebene als ein Satzglied interpretiert.

(149) „*Ich Kan [**Kein menschen broth**] nunmehr nicht darauff mehr erwerben*“ (Minden 1614)

7.7 Eigennamen

Bei der Bestimmung von Phrasengrenzen in den Hexenverhörprotokollen waren zwei Dinge besonders zu berücksichtigen. Zum einen liegen die Texte genau in der Zeit, in der sich feste Familiennamen entwickeln, gleichzeitig aber auch noch viele Beinamen kursieren.

Grundsätzlich gilt, dass Eigennamen, deren innere Struktur nicht eindeutig zu bestimmen ist, im Ganzen als eine Phrase ohne Binnenstrukturierung annotiert werden. Bei der Bestimmung der Struktur wurde auf die PoS-Annotation der Berufsbezeichnungen zurückgegriffen (vgl. Kap. 4.2.2.1 Regeln zur Verwendung von NE, NB und NN). Substantive, die die Annotation *pos=NB* trugen, wurden nicht als Teil des Eigennamens, sondern als Apposition zu diesem annotiert (vgl. Kap. 7.5).

(150) *Wiben Maes Telse* (Meldorf 1630)

(151) *Teden Hanß Jacobß Hans* (Meldorf 1630)

(152) *Melchior Anschutz schneider* (Georgenthal 1597)

Ist einem Personennamen eine Präpositionalphrase mit eingebettetem Ortsnamen nachgestellt, muss entschieden werden, ob es sich dabei um ein Attribut oder eine adverbiale Bestimmung des Orts handelt. Wenn möglich, wurde die PP nicht als Teil des Eigennamens interpretiert, sondern als adverbiale Bestimmung. Auch in ambigen Fällen wurde diese Annotation gewählt. In Sätzen ohne Verb (v.a. Listen) wurde die PP hingegen immer als Teil des Namens annotiert.

(153) [*die große Aleke Steineken [zu [Nider Allm]]*] (Alme 1630)

(154) „*Item bekandt, das sie die butterei vor .3. Jharen [von [peter Sophien]] [zu [verbitz]] **geleret***“ (Perleberg 1588)

8 Phrasentyp (phrasentyp)

Alle Spannen auf der w_func-Ebene wurden danach annotiert, ob es sich um eine Nominalphrase oder Präpositionalphrase handelt, und danach, ob sich diese Phrase in einem (finiten) Satz befindet oder nicht. Die Entscheidung, ob eine Phrase als Teil eines Satzes gilt, davon ab, ob in dem entsprechenden Satz ein Vollverb steht, das die Phrase regiert.

Für das Feature *phrasentyp* gibt es folgende mögliche Values:

| Values des Features <i>phrasentyp</i> | Erklärung |
|---------------------------------------|--|
| NP_satz | deutsche Nominalphrase in satzwertiger Struktur |
| PP_satz | deutsche Präpositionalphrase in satzwertiger Struktur |
| NP_nichtsatz | deutsche Nominalphrase in nicht-satzwertiger Struktur |
| PP_nichtsatz | deutsche Präpositionalphrase in nicht-satzwertiger Struktur |
| NP_praep | in eine deutsche Präpositionalphrase eingebettete deutsche Nominalphrase |
| NP_lat | deutsche Nominalphrase in lateinischem Satz (Kap. 6.3.4) |
| PP_lat | deutsche Präpositionalphrase in lateinischem Satz (Kap. 6.3.4) |

8.1 Phrasen im Satz

Eine Nominalphrase, die als Subjekt, Objekt oder auch in adverbialer Funktion Teil eines Satzes ist, wird als *phrasentyp=NP_satz* annotiert. Präpositionalphrasen erhalten die Annotation *phrasentyp=PP_satz*, wenn die Phrase Teil eines Satzes ist. Eine Konstruktion gilt auch dann als finiter Satz, wenn eine Auxiliarellipse vorliegt oder in einem Prädikatsatz das Kopulaverb fehlt.

Phrasen in einem finiten Satz:

- (155) *Wahr das [der Paull Hoikendorf]_{phrasentyp=NP_satz} einmahl
[ein laken]_{phrasentyp=NP_satz} verlohren, vnnnd da [deßen fraw]_{phrasentyp=NP_satz}
[sich]_{phrasentyp=NP_satz} darüber [i[m hauße]_{phrasentyp=NP_praep}]_{phrasentyp=PP_satz} beschweret“ (Güstrow
1615)*

Beispiel für Prädikativkonstruktion mit Verbellipse (NP_satz):

- (156) *„8. Ob[s]_{phrasentyp=NP_satz} nitt nur [sein anhang]_{phrasentyp=NP_satz} vnd nitt [sein Ehlichs weyb]
phrasentyp=NP_satz sey.“ (München 1600)*

Bei komplexeren Phrasengefügen wurden in eine PP eingebettete NPs als `phrasentyp=NP_praep` annotiert. Eingebettete PPs treten als `PP_satz` oder `PP_nichtsatz` auf, je nachdem, ob in dem Satz ein Vollverb vorhanden ist oder nicht.

(157) „*begert [die [bey [dem Rechtsgelernten]_{phrasentyp=NP_praep}]_{phrasentyp=PP_satz} außsprachte Aduis]_{phrasentyp=NP_satz} zueroffnen*“ (Hamm 1592)

8.2 Nicht in einen Satz eingebettete Phrasen

In den Hexenverhörprotokollen sind viele Phrasen enthalten, die nicht als Bestandteil eines Satzes gewertet werden. Darunter fallen viele Phrasen, die im Randsatz stehen oder Teil von Listen sind, ebenso wie Anreden. Auch Sätze ohne Vollverb fallen in diese Kategorie. Dies betrifft vor allem den Anfang der Gerichtsprotokolle. Für in solchen Konstruktionen enthaltene Phrasen wurde die Annotation `phrasentyp=NP_nichtsatz` bzw. `phrasentyp=PP_nichtsatz` vergeben.

Diese Annotation ist grundlegend für die weitere inhaltliche Annotation auf der Phrasenebene: Nur Phrasen, die als zu einem Satz gehörig annotiert wurden, erhalten differenzierte Annotationen für syntaktische Funktion und semantische Rolle.

8.2.1 Textuelle Gliederungszeichen/Absatzgliederung:

Textuelle Gliederungszeichen und textgliedernde Überschriften werden nur dann als zu einem Satz gehörig gewertet, wenn sie syntaktisch in den darauffolgenden Satz integriert sind (d.h. ob sie im Vorfeld des Satzes steht). Vgl. dazu die Satzannoation von textuellen Gliederungszeichen (s. 5.5 Textuelle Gliederungszeichen).

(158) „*[Zu[m Finfften]]_{Phrasentyp=PP_satz} hab Er Ir ain zaichen an der brust geben [...]*“ (Riedlingen 1596)

(159) „*[Zu[m vierten]]_{phrasentyp=PP_nichtsatz} Wann sy hab beichttet, vnd die beicht fürüber*“ (Riedlingen 1596)

8.2.2 Listen

Phrasen, die Teil von Listen sind, werden als nicht zu einem Satz gehörig annotiert. Solche Listen können im Text oder im Randsatz vorkommen.

(160) „20. *Ob er nitt die Meindel kenn, vnd wiewil diß namens ime bekannt seyen.*
21. *[Abrahamb kueffer]_{phrasentyp=NP_nichtsatz}, [Hans Kueffer]_{phrasentyp=NP_nichtsatz}, [Neunaugen]_{phrasentyp=NP_nichtsatz}, [Neuenfingerl]_{phrasentyp=NP_nichtsatz}, [Zuckerwastl]_{phrasentyp=NP_nichtsatz}, [Gilg]_{phrasentyp=NP_nichtsatz}.*“ (München 1600)

8.2.3 Randsatz

Als Randsatz bezeichnet man den Text, der links (oder rechts) vom eigentlichen Haupttext am Rand steht. Dabei handelt es sich meistens um Kommentare zum Text oder Indices, in denen einzelne Verbrechen der Angeklagten, die Namen von Zeugen oder von den Befragten beschuldigten Personen festgehalten werden, um zu später bei einer Aktendurchsicht schneller finden zu können. Es handelt

sich dementsprechend häufig um Stichwörter, die als NP_nichtsatz oder PP_nichtsatz annotiert werden:

- (161) „[*ehebruch*]_{phrasentyp=NP_nichtsatz}“ (Lemgo 1632)
(162) „[*Däntz*], *vnd* [*Wetther*]“ (Baden-Baden 1628)

8.2.4 Anrede

Nominale Anreden, die syntaktisch gesehen nicht zu dem darauffolgenden Satz gehören, werden als phrasentyp=NP_nichtsatz annotiert, da sie in diesem Satz keine syntaktische Funktion übernehmen können:

- (163) „zu *Ihr Appolonien* gesagt, O [*Appell*]_{phrasentyp=NP_nichtsatz} [*Appell*]_{phrasentyp=NP_nichtsatz} *wie wird dir* doch so vbell gehen, *dan sie wollen*“ (Gaugrehweiler 1610)
- (164) „*Petrus vnd Paulus gingen aus das kraut zusuchen, damit wolten sie das heilige dingk stillen vnd böten, do begegnete ihnen der herr Christus, Stehe* [*Petrus*]_{phrasentyp=NP_nichtsatz} [*Paulus*]_{phrasentyp=NP_nichtsatz}, *das kraut ist gefunden, die glocken sein geklungen, die Meßen sein gesungen*“ (Crivitz 1642)

Bei zweiteiligen Anreden mit Pronomen an erster Stelle wird eine Apposition angenommen und beide Teile getrennt annotiert:

- (165) „*bistu schon des teuffels* [*du*]_{phrasentyp=NP_nichtsatz} [*leidige Sack*]_{phrasentyp=NP_nichtsatz}“ (Osnabrück 1636)

Das gilt nicht, wenn der zweite Teil von einer Nominalisierung gebildet wird (etwa nominalisiertes Adjektiv). Dann wird die Anrede als eine NP annotiert:

- (166) „*höret her* [*gi grawen*], [*gy grunen*], [*schwarten*], [*witten*], [*bunten*],“ (Perleberg 1588)

8.2.5 Vergleichsphrasen

Bei Vergleichsphrasen erhalten nur die NPs annotiert, denen eindeutig ein Verb zugeordnet werden kann, die Annotation als NP_satz. Kann einer Phrase nicht eindeutig ein Verb zugeordnet werden, erhält es die Annotation *phrasentyp=NP_nichtsatz*. In Bsp. (167) ist „ein Plöwliches Zaichen“ Subjekt zum finiten Verb „befindet“. Das finite Verb zu „ein Kleeblath“ kann hingegen nicht eindeutig bestimmt werden und trägt die Annotation *phrasentyp=NP_nichtsatz*.

- (167) „*befindet sich in der rechten Seithen* [*ein Plöwliches Zaichen*]_{phrasentyp=NP_satz}, *wie* [*ein Kleeblath*]_{phrasentyp=NP_nichtsatz}“ (Bamberg 1628)

9 Einbettung von Phrasen (embedded)

Mit dem Feature *embedded* werden Informationen über die Einbettung einer Phrase kodiert. Dafür stehen vier Values zur Auswahl: *false*, *true-head*, *true-attr* und *true-no_attr*.

Phrasen, die nicht in größere Phrasen eingebettet sind, haben die Annotation *embedded=false* erhalten. Die anderen Values wurden dann verwendet, wenn eine NP oder PP in eine andere Phrase eingebettet ist. In diesem Fall wird durch den zweiten Teil des Values angezeigt, ob es sich bei der eingebetteten Phrase um ein Attribut handelt oder nicht. Bei Nominalphrasen, die in eine Präpositionalphrase eingebettet sind, wurde die NP als *embedded=true-head* annotiert. Phrasen, die als eingebettet annotiert wurden, erhalten keine weitere Annotation für semantische Rolle oder syntaktische Funktion (Ausnahme: Agens in Passivsätzen, vgl. Kap. 11.1.2).

(168) „[[**des Alten Peters**]_{embedded=true-attr, subj=NA} **weib**]_{embedded=false, subj=false} sei der leüchter, stehe auff kop Vnd hab das licht im hindersten.“ (Alme 1630)

9.1 embedded=false

Alle Nominalphrasen und Präpositionalphrasen, die nicht in eine andere Phrase eingebettet sind, bekommen die Annotation *embedded=false*. Das betrifft alle Phrasen, egal ob sie in einen finiten Satz eingebettet sind oder nicht.

(169) „gesteth
doch daß [vff ein Walpurg tage]_{embedded=false} [zu morgen]_{embedded=false}
alß [sie]_{embedded=false} [Melchior Cronsperger]_{embedded=false} nackent
gefunden [ein schüssell voll dhau]_{embedded=false} gefast vnd
[die zän]_{embedded=false} darmit gespüelet“ (Gaugrehweiler 1610)

9.2 embedded=true-attr

Alle Nominalphrasen und Präpositionalphrasen, die in eine andere Phrase eingebettet sind und als Attribut fungieren, werden als *embedded=true-attr* annotiert. Dabei kann es sich um NP_satz, NP_nichtsatz, PP_satz, PP_nichtsatz oder NP_praep handeln.

(170) „Habe auch [[**ermelten Stromengers**]_{embedded=true-attr} sohn] vorgeben, Er könde ein hosenbandt entzwey schneiden,“ (Schweinfurt 1616)

(171) „Hab 3 fueß [in[[**s teuffelß**]_{embedded=true-attr} Nahmen]] zurück gehen, Gott Vnd Alle seine heiligen Verleichenen Vnd sich dem teufel ergeben mußen.“ (Alme 1630)

(172) „Auß Wülfte köller Steffen, die NieMeiersche welcher man thodt, Clagges hanß welcher [ein lanck dingh [**mit questen**]]_{embedded=true-attr} Alß eine fane in der handt gehabt,“ (Alme 1630)

(173) „vnd hette [[**der Lazaruschen**] ihr] grawe kleider angehabt,“ (Crivitz 1642)

(174) „Derselbe habe Ihr [ein stücke [**goldes**]_{emb=true-attr}] geben“ (Jever 1592)

9.3 embedded=true-head

Nominalphrasen, die als phrasentyp=NP_praep annotiert sind, erhalten oftmals die Annotation embedded=true-head, da es sich bei ihnen nicht um Attribute handelt.

(175) „Ihr obrister [auff [**teufelß dantz**]_{embedded=true-head}] heiße
lucifer den sie [mit [**großer reuerentz**]_{embedded=true-head}]
Vnd [**hindersten küßen**]_{embedded=true-head} verehren müßen.“ (Alme 1630)

9.4 embedded=true-no_attr

Mit *embedded=true-no_attr* wurden Phrasen annotiert, die zwar in eine andere Phrase eingebettet sind, aber weder als Attribut noch als Phrasenkopf angesehen werden können. Das betrifft unter anderem in Präpositionalphrasen eingebettete Relativpronomina zu annotieren.

(176) „Darauff
sie abermals wieder vffgezogen, vnd ahn Ihr ersehen worden, daß sie
den Scharpffrichtter mitt der stangen so Ir durch Ihre zusammen gebundene
bain gelegtt, [vf [**welche**]_{sr=true-no_attr}] der Scharpffrichter sich gestellet vffgehoben,“ (Lemberg 1630)

10 Satzgliedfunktion (synfunc)

Im SiGS-Korpus sind Satzgliedfunktionen unter der Annotation *synfunc* zu finden. Dabei wurden die Funktionen Subjekt (*subj*), Objekt (*obj*), adverbiale Bestimmung (*adv*) und potentielles Präpositionalobjekt (*pot_PPO*) unterschieden. Die Objekte und adverbialen Bestimmungen wurden nicht weiter unterschieden. Diese Annotationen wurden prinzipiell nur an Phrasen vergeben, die in Sätze eingebettet sind. Bei Phrasen, die nicht Teil eines finiten Satzes sind, kann weder der Satzgliedstatus noch die semantische Rolle festgestellt werden. Phrasen, die für *phrasentyp* das Value *NP_nichtsatz* oder *PP_nichtsatz* haben, tragen also die Annotation *synfunc=NA*.

(177) „[**Pawel**]_{sr=keine_rolle, synfunc=NA} du spennest keine siede dabei“ (Güstrow 1615)

Wenn unklar ist, ob es sich bei einer NP um ein Objekt oder Subjekt handelt, wird *synfunc=ambig* vergeben. Zusätzlich steht das Value *NA* zur Auswahl. Dieses wird z.B. verwendet, um Prädikatsnomen zu annotieren (vgl. Kap. 10.3) oder dann, wenn eine Einheit zwar eine semantische Rolle innehat, ihr aber kein Satzgliedstatus zukommt. *NA* wird auch für alle Phrasen vergeben, die in eine andere Phrase eingebettet sind, also zum Beispiel alle Phrasen, die die Annotation *phrasentyp=NP_praep* haben oder *embedded=true*, sowie für Phrasen, die nicht in Sätze eingebettet sind.

10.1 Adverbiale Bestimmung (adv)

Adverbial verwendete Nominalphrasen tragen die Annotation *synfunc=adv*. Dabei wurde nicht zwischen den unterschiedlichen Arten von Adverbialen unterschieden.

adverbiale Bestimmung des Orts:

(178) „*auß Selbach*“ (Lemberg 1630)

adverbiale Bestimmungen der Zeit:

(179) „*Aber nichts be-
kennen wollen, biß sie [ein kleine weil]*_{synfunc=adv, sr=keine_rolle} *in der foltter gehalten*“ (Lemberg 1630)

adverbiale Bestimmungen der Art und Weise:

(180) „*Also wardt Sye [das vierte mahl] [obiger gestalt] auffgezogen, welche vier abzüge
sich zuemahlen nicht vber ein viertel
stundt erstreckhet*“ (Messkirch 1644)

adverbiale Bestimmungen des Grundes:

(181) „*Auch [wes vrsachen]*_{synfunc=adv, sr=keine_rolle} *solliches gethan habe,*“ (Günzburg 1615)

Das Wort *namens* wurde als adverbiale Bestimmung annotiert, die unabhängig von dem Namen und dem, den der Name bezeichnet, steht.

(182) „*[den ersten Mann] [Namens]*_{synfunc=adv, sr=keine_rolle} *[Jacob
Olbrecht]*_{synfunc=stativ, sr=NA} *hette Sie 2 Jahr gehabt,*“ (Leipzig 1640)

(183) „*daruon Ihre
zwue döchtern noch in leben, [nammens]*_{synfunc=adv, sr=keine_rolle}
*[Catharina]*_{synfunc=stativ, sr=NA} *vnnd [Maria]*_{synfunc=stativ, sr=NA}“ (Leonberg 1641)

10.2 Potentielle Präpositionalobjekte (pot_PPO)

Bei Präpositionalphrasen wurde nicht zwischen Präpositionalobjekt und adverbialer Bestimmung unterschieden; alle Präpositionalphrasen, für die *phrasentyp=PP_satz* und *embedded=false* gilt, tragen die Annotation *synfunc=pot_PPO*.

(184) „*Ihren Teuffelßdantz haben sie [an [der eggen]*_{synfunc=NA} *]*_{synfunc=pot_PPO}“ (Alme 1630)

- (185) „daneben Auch 5 Articulirte Persohnen womit sie in vngueten Ihrer bekendtnuß nach nichtz zuthuen gehapt; [**fur [eine zauberinne]**]_{synfunc=pot_PPO} besagt“ (Werl 1630)

Präpositionalphrasen, die in eine andere Phrase eingebettet sind, erhalten hingegen die Annotation *synfunc=NA* (vgl. Kap. 10.3).

10.3 Synfunc=NA

Mit der Annotation *synfunc=NA* wurden alle NPs und PPs annotiert, die nicht eindeutig als Subjekt, Objekt, adverbiale Bestimmung oder potentielle Präpositionalobjekte identifiziert werden konnten. Das betrifft somit alle eingebetteten Phrasen sowie alle Phrasen, deren Phrasentyp nicht *NP_satz* oder *PP_satz* ist. Im Folgenden sind die für das SiGS-Korpus einschlägigen Fälle gelistet. Für eine Auflistung der für das SiGS-Korpus typischen Fälle von Phrasen, die nicht in einen Satz eingebettet sind, vgl. Kap. 8.2.

10.3.1 Prädikatsnomen

Prädikatsnomen wurden als *phrasentyp=NA* annotiert. Dabei trägt immer die NP, die rechts vom Verb steht, die Annotation als *NA*. Diese Annotation wurde auch in Sätzen mit Subjektellipse oder Verbellipse vergeben.

- (186) „welche ihm 1 thaler geben Vnd sei [**perdedreck**]_{sr=stativ, synfunc=NA} worden, mit welcher er zuthuen gehabt, ist so [**kalter Natur**]_{sr=stativ, synfunc=NA} gewesen“ (Alme 1630)
- (187) „[**Funcken Olmes**]_{sr=stativ, synfunc=subj} sei [**der Spielman**]_{sr=stativ, synfunc=NA} gewesen.“ (Alme 1630)

- (188) „Vff erthailten Fürst Beuelch
~~neben yberschickhter Articul,~~
 Ist Inn bey sein Georg Abells
 vnnd Christoph Georgen vfm Sandt,
 [**béede**]_{synfunc=subj} [**deß Gerichts**]_{synfunc=NA} alhie, Maria, Jung Jacob Kautzen Witib
 vonn Görlingen, vnnd Ihre zwo
 döchtern Catharina vnnd Maria
 [**béede**]_{synfunc=subj} noch [**ledigstandt**]_{synfunc=NA, sr=keine_rolle,}
 alle drey verhoffte, yber die vonn Für
 Cantzley vsß yberschickhte Inter-
 rogatoria, vnnd andere wider
 Sie einkommene verdachts Puncten,
 Ernstlich Examinirt vnnd verhört“ (Leonberg 1641)

10.3.2 Interrogativpronomina

Bei Fragesätzen wird stets das Interrogativpronomen als *NA* annotiert und das Substantiv bzw. Pronomen als Subjekt.

- (189) „[**Wer**]_{synfunc=NA} [**sein vatter**]_{synfunc=subj}, vnd [**Muettr**]_{synfunc=subj} sey.“ (München 1600)

10.3.3 Vergleichsphrasen

Synfunc=NA wird außerdem an Vergleichsphrasen vergeben, wenn sie nicht eindeutig einem finiten Verb zugeordnet werden können (vgl. Kap. 8.2.5).

- (190) „*alda sie durch
einen schwartzen Mann in schwartzen kleidern wie [ein Paff]_{synfunc=NA} Irem Buhlen
gegeben worden,*“ (Lemberg 1630)

10.3.4 Annotation von „sich“

Wenn das Reflexivpronomen *sich* nicht als Objekt verwendet wird, trägt es ebenfalls die Annotation *synfunc=NA*. Das betrifft die Fälle, in denen das Reflexivpronomen vom Verb vorgegeben ist und in diesem Sinne kein eigentliches Objekt darstellt.

- (191) „*warzue sie [sich]_{synfunc=NA} nochmahß, wie zue vor bekanth,*“ (Friedberg 1620)
(192) „*was [sich]_{synfunc=NA} zue Berlitz Zugetragen*“ (Blankensee 1619)

10.4 Elliptische NPs

Bei elliptischen NPs ohne nominalen Kopf wurde die Ellipse trotzdem als Satzglied annotiert. Sie erhält ggf. die gleiche Annotation wie eine koordinierte voran- oder nachgestellte NP.

- (193) „*Sie hette [ein Pahr Elben]_{synfunc=obj} oder [zwölfte]_{synfunc=obj} mit ihrem Buhlen dem Teuffel
gezeuget,*“ (Ostrau 1628)

10.5 Es im Vorfeld

Im Vorfeld auftretendes *es* wird je nach Funktion annotiert. Handelt es sich um ein Vorfeld-*es*, dann werden sowohl das *es* als auch das nachgestellte Subjekt als Subjekte annotiert (Bsp. (194) und (195)). Expletives *es* trägt die Satzgliedannotation *synfunc=NA* (vgl. Bsp. (196)).

- (194) „*[Es]_{synfunc=subj, sr=stativ} wehren zue Ostraw vndt Wodern auch [zaubersche]_{synfunc=subj, sr=stativ}.*“
(Ostrau 1628)
(195) „*[Es]_{synfunc=subj, sr=agens} habe der böse Geist Auch vff ein Zeit
den Hanßen Grymmen Inn ihr hauß gepracht,*“ (Georgenthal 1597)
(196) „*Denn [es]_{synfunc=subj, sr=none} in der geburt
zeit, weill es ein groß feist kindt
gewesen, etwas hart zugangen,*“ (Ostrau 1628)

11 Semantische Rolle (sr)

Bei der Annotation wird eine weite Definition des Agens-Begriffs zugrunde gelegt, die weitestgehend mit Dowtys (1991) Begriff des Proto-Agens und Van Valins (2005) Konzept des *Actor* kompatibel ist. Beide subsumieren eine Reihe an thematischen Rollen, die über den klassischen Begriff des Agens im engeren Sinne hinausgehen. Die Agensrolle im engeren Sinn bildet jedoch den prototypischen Kern von Proto-Agens bzw. *Actor*.

Für die Zuweisung der einzelnen Rollen ist im Regelfall das Prädikatsverb ausschlaggebend, daher orientieren sich die folgenden Ausführungen ausschließlich an Verben. Auf eine explizite Annotation der semantischen Rollen bei Präpositionalphrasen wurde verzichtet. Auch Attribute haben keine *sr*-Annotation erhalten. Entsprechend tragen nur Phrasen mit der Annotation *phrasentyp=NP_satz* und *embedded=false* eine semantische Rolle, wie sie in Kap. 11.1 bis Kap. 11.6 beschrieben sind. Die übrigen Phrasen erhalten die Annotation *sr=keine_rolle*. Die einzige Ausnahme davon bildet das Agens in Passivsätzen (vgl. Kap. 11.1.2).

Die Annotation der semantischen Rollen erfolgte verbgeleitet – es wurden keine belebtheitsbezogenen Einschränkungen bei der Vergabe der Rollen gemacht. Es wurden also auch bspw. unbelebte Entitäten als Agens annotiert.

(197) alß [das wetter]_{sr=agens} In das schauff hauß alhier geschlagen (Ellingen 1590)

Treten Modalverben auf, so wurden an Subjekt und Objekte die semantischen Rollen vergeben, die sie beim Auftreten des Vollverbs ohne Modalverb erhalten hätten.

(198) „Bekennet mit Ja, Aber sie hette die Elben
von ihr ins waßer geschafft, vnd ins Teuffels
nahmen wieder vom Megdelein weg ge-
wunschet, das [sie]_{sr=experiencer} nun wieder sehen köntte“ (Ostrau 1628)

In Sätzen mit verneinter Proposition wurden die semantischen Rollen vergeben, die dem Vollverb entsprechen:

(199) „das hette sie nit gesagt“ (Stralsund 1630)

11.1 Proto-Agens

11.1.1 Agens in Aktivsätzen (*agens*)

Folgende Eigenschaften werden als typisch für ein Proto-Agens angesetzt (in Anlehnung an Dowty 1991: 572):

- willentliche Ausführung einer Handlung
- Bewusstsein
- Auslösung einer Handlung oder Bewirkung von Veränderungen eines anderen Partizipanten
- Bewegung (relativ zu anderen Partizipanten der Verbalhandlung)
- (Existenz ist unabhängig von der Verbalhandlung)

Entsprechend der Theorie der Proto-Rollen muss eine Nominalphrase nicht alle diese Eigenschaften aufweisen, um als *sr=agens* annotiert zu werden.

11.1.2 Agens in Passivsätzen

In Passivsätzen wird das Agens in der Regel durch eine Präpositionalphrase ausgedrückt. Die in dieser Präpositionalphrase enthaltene Phrase, die als *phrasentyp=NP_praep* annotieren ist, trägt die semantische Rolle *agens*.

(200) „wehre Er [**vo[m böesen Geist]**]_{sr=agens}]_{sr=keine_rolle} mit schlägen bel tractirt worden.“ (Bamberg 1628)

11.1.3 Handlungskontrolle

Wenn durch ein Verb beschrieben wird, dass etwas durch eine bewusste Handlung des Subjekts nicht geschieht oder eine bewusste Unterlassung einer Handlung vorliegt, wurde dies als Handlungskontrolle interpretiert und das Subjekt entsprechend als Agens annotiert (vgl. Primus 2012).

(201) „so Ihm gesagt der buddel kehme, [**er**]_{sr=agens} solt sich fest halten, sie wol in erretten.“ (Alme 1630)

(202) „weill [**er**]_{sr=agens} heinrich geschwiegen“ (Gaugrehweiler 1610)

11.2 Proto-Patiens (*patiens*)

Folgende Eigenschaften werden (in Anlehnung an Dowty 1991) als typisch für ein Proto-Patiens angesehen:

- Unterliegt einer Veränderung
- Affiziertheit durch einen anderen Partizipanten der Verbalhandlung
- unbewegt (relativ zu anderen Partizipanten der Verbalhandlung)
- (existiert in Abhängigkeit von der Verbalhandlung) (*Ich baue ein Haus*)

Eine Nominalphrase muss nicht alle diese Eigenschaften aufweisen, um im SiGS-Korpus als *patiens* annotiert zu werden. Es genügt, wenn eine dieser Eigenschaften erfüllt ist. Da SiGS mit Proto-Rollen arbeitet, fallen unter *patiens* auch Nominalphrasen, die man auch als Rezipient, Thema, Benefaktiv, Malefaktiv u.ä. bezeichnen könnte. Auch bei Sätzen mit unakkusativischen Verben wie *stürzen, fallen, sterben, ankommen* wurde das Subjekt als *sr=patiens* annotiert (vgl. Dowty 1991: 605).

11.2.1.1 Rezipient

Die Rezipientenrolle im SiGS-Korpus als Teil des Proto-Patiens aufgefasst. Entsprechend werden Subjekte von Verben, die den Erhalt eines Objekts anzeigen, als *sr=patiens* annotiert.

(203) „[**Es**]_{sr=patiens} hette aber noch die Tauffe **empfangen**.“ (Ostrau 1628)

11.2.1.2 Passiv

Alle Phrasen, die als Subjekte von Passivsätzen fungieren, wurden als Patiens annotiert, auch wenn im Falle eines Zustandspassivs eine Interpretation als *stativ* möglich wäre.

- (204) „[**Ihre dantz Leine**]_{sr=patiens} sei an den Heertorn Vnd an die Galgen gebunden.“ (Alme 1630)
(205) „wehre [er]_{sr=patiens} baldt darup in minschen gestaldt vorwandelet“ (Meldorf 1618)

11.3 Experiencer (experiencer) und Stimulus (stimulus)

Experiencer sind Partizipanten, die etwas mit ihren Sinnen wahrnehmen oder fühlen. Subjekte von Perzeptionsverben wie *sehen*, *hören*, *riechen*, sowie Verben des Denkens, Fühlens, Meinens wurden entsprechend als Experiencer annotiert. Der Auslöser von Gefühlen und Sinneswahrnehmungen wurde als *stimulus* annotiert. Das gilt auch für belebte Entitäten, die möglicherweise mit Absicht handeln.

- (206) „vmb das [Sie]_{sr=experiencer} mit Ime zuvnfriden worden“ (München 1600)
- (207) „daß [Er]_{sr=experiencer} [...] [den Trachen]_{sr=stimulus} durch seinen hoff fliegen sehen?“ (Rosenburg 1618)
- (208) „Illa, mochte
katzen muntze sein, dan [das selbe]_{sr=stimulus} möhten
[die katzen]_{sr=experiencer} gerne sehen vnd leiden;“ (Stralsund 1630)

11.4 Stativ (stativ)

Weist ein Partizipant einer Verbalhandlung keine der Eigenschaften von Agens, Patiens, Experiencer und Stimulus auf, so wurde die entsprechende Nominalphrase als *stativ* annotiert. Darunter fallen Besitz-, Existenz- und Positionsverben, sowie Verben, die Zustände im Allgemeinen ausdrücken. Auch Subjekte von Prädikatsätzen fallen in diese Kategorie (zur Annotation von Prädikatsnomen vgl. Kap. 10.3.1).

11.4.1.1 Existenzverben

Das Subjekt von Existenzverben wie *sein*, *sich befinden* und *wohnen* wird als *Stativ* annotiert.

- (209) „ist [keiner]_{sr=stativ} Alder Verhanden der so heist“ (Rosenburg 1618)
(210) „[diese Alheit]_{sr=stativ} sei Jetzo bei Ihrem Sohen dem kuhirtenn zu Wesendorff“ (Celle 1570)

11.4.1.2 Besitzverben

Besitzverben wie *besitzen* oder *haben* fordern ein Stativ in der Subjekt- und Objektposition.

- (211) „Ob [Sy]_{sr=stativ} [ein aigen gäbele]_{sr=stativ} gehabt“ (Augsburg 1625)
(212) „das vngeuarlich vur zehsechß Jharn do [sie]_{sr=stativ} [missell]_{sr=stativ} vnd [zweyspalt]_{sr=stativ} mit Ihrem hern gehabt“ (Hamm 1592)

11.4.1.3 Zuweisung von Eigenschaften (Kopulaverben)

Die Zuweisung von Eigenschaften kann durch prädikative Konstruktionen, Kopulaverben oder *haben* erfolgen. Auch *nennen*, *heißen* usw. wurden dazu gezählt. In solchen Fällen tragen Subjekt und Objekt die semantische Rolle *stativ*.

- (213) „Ob [*ihr Geist*]_{sr=stativ} hörnner vffm kopff gehabt“ (Georgenthal 1596)
(214) „[*Sie*]_{sr=stativ} heiße [*Grete*]_{sr=stativ}“ (Minden 1614)
(215) „[*Funcken Olmes*]_{sr=stativ, synfunc=subj} sei [*der Spielman*]_{sr=stativ, synfunc=NA} gewesen.“ (Alme 1630)

11.4.1.4 Positionsverben

Verben wie *sitzen*, *stehen*, *liegen*, *wohnen*, *vorhanden sein* gelten als Stativ.

- (216) „vor Bastian Hadels stallung [*drey stuckh [rindt Viech]*]_{sr=stativ} gestanden“ (Baden-Baden 1628)

11.5 Mehrfache Rollenzuweisung bei einem Substantiv (multi)

Bedingt durch Koordination von Verben kann es vorkommen, dass ein Substantiv gleichzeitig mehrere verschiedene semantische Rollen trägt. In solchen Fällen wurde das Value *multi* vergeben.

In Bsp. (217) trägt die NP *Cappelle* die Annotation *sr=multi*, da die beiden Verben unterschiedliche semantische Rollen als Subjekt fordern: *erschrocken* fordert einen Experiencer und *gezittert* ein Patiens.

- (217) „[*Cappelle*]_{sr=multi} Aber sehr erschrocken vnd gezittert“ (Helmstedt 1580)

11.6 Unklare Rollenzuweisung (ambig)

Wenn durch die Verben und den Kontext nicht eindeutig bestimmt werden konnte, welche semantische Rolle einer NP zugewiesen werden sollte, wurde diese als *sr=ambig* annotiert. Dies betrifft häufig das Personalpronomen *sie*, bei dem wie in Bsp. (218) rein formal nicht zwischen Nominativ und Akkusativ unterschieden werden kann.

- (218) „[*Die*]_{sr=ambig} [*sie*]_{sr=ambig} gleichfahls Vor der thüren auff der gasßen angetroffen vnd geschlagen“ (Baden-Baden 1628)

11.7 Keine Rollenzuweisung (keine_rolle)

Es gibt NPs, die keine semantische Rolle bekommen. Das betrifft die in Kap. 11 angesprochenen Fälle: Präpositionalphrasen, adverbiale Bestimmungen, eingebettete Nominalphrasen und alle Phrasen, die nicht als Teil eines finiten Satzes annotiert sind (vgl. Kap. 8.2), sowie expletives *es* und Prädikatsnomen. Auch wenn *sich* nicht als Objekt fungiert, sondern als fester Teil des Verbs, wird hier *keine_rolle* vergeben. Für diese Fälle siehe Kap. 10.3.4 zu Reflexivpronomina.

- (219) „Zum .16. Alß [[*Veltinß [von [Möringen]*]_{sr=keine_rolle}]_{sr=keine_rolle}]_{sr=keine_rolle} hauß-
frau]_{sr=agens}, [*nit Lang verschiner zeit*]_{sr=keine_rolle}, [*v[om Marckht]*]_{sr=keine_rolle}]_{sr=keine_rolle}
[*von [Riedlingen]*]_{sr=keine_rolle}]_{sr=keine_rolle} [*mit [Ir]*]_{sr=keine_rolle}]_{sr=keine_rolle} haimbwerts gangen,“
(Minden 1614)

- (220) „O ich leide [*eß*]_{sr=keine_rolle} vnschuldigh“ (Minden 1614)

12 Literaturverzeichnis

- Artstein, Ron & Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational linguistics* 34.4, 555–596.
- Barteld, Fabian, Renata Szczepaniak & Heike Zinsmeister. 2014. The definition of tokens in relation to words and annotation tasks. Tokens and words as basis for annotations. In Verena Henrich et al. *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13). December 12-13, 2014, Tübingen, Germany*. 250-257. (<http://tlt13.sfs.uni-tuebingen.de/tlt13-proceedings.pdf>).
- Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.
- Dipper, Stefanie et al. 2013. HiTS: ein Tagset für historische Sprachstufen des Deutschen. *Journal of Language Technology and Computational Linguistics* 28, 85-137.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67/3. 547-619.
- Ebert, Robert P. et al. 1993. *Frühneuhochdeutsche Grammatik*. Tübingen: Niemeyer.
- Jurish, Bryan. 2011. *Finite-state Canonicalization Techniques for Historical German*. Dissertation, Universität Potsdam, Januar, 2011. URN urn:nbn:de:kobv:517-opus-55789, URL <http://opus.kobv.de/ubp/volltexte/2012/5578/>.
- Krasselt, Julia et al. 2015. Guidelines für die Normalisierung historischer deutscher Texte/Guidelines for Normalizing Historical German Texts. *Bochumer Linguistische Arbeitsberichte* 15 (http://www.linguistics.ruhr-uni-bochum.de/bla/015-krasselt_etal2015.pdf).
- Krause, Thomas & Amir Zeldes. 2016. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31. 118-139. <https://doi.org/10.1093/lc/fqu057>
- Macha, Jürgen et al. (Hrsg.). 2005. *Deutsche Kanzleisprache in Hexenverhörprotokollen der Frühen Neuzeit*, 2 Bde. Berlin, New York: de Gruyter.
- Pauly, Dennis, Ulyana Senyuk & Ulrike Demske. 2012. Strukturelle Mehrdeutigkeit in frühneuhochdeutschen Texten. *Journal of Language Technology and Computational Linguistics* 27, 65-82.
- Primus, Beatrice. 2012. *Semantische Rollen*. Heidelberg: Winter.
- Pustejovsky, James & Amber Stubbs. 2012. *Natural Language Annotation for Machine Learning*. Beijing u.a.: O'Reilly.
- Schiller, Anne et al. 1999. *Guidelines für das Tagging deutscher Textcorpora mit STTS (kleines und großes Tagset)*. (<http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>).
- Oskar Wächter. 1882. *Das Hexenwesen. Vehmgerichte und Hexenprozesse in Deutschland*. Stuttgart.
- Van Valin, Robert D. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press
- Wilke, Anja. 2006. *Redewiedergabe in frühneuzeitlichen Hexenprozessakten*. Berlin, New York: de Gruyter.
- Zipser, Florian & Laurent Romary. 2010. A model oriented approach to the mapping of annotation formats using standards. *Workshop on Language Resource and Language Technology Standards, LREC 2010*. La Valette, Malta. (<https://hal.inria.fr/inria-00527799>).