
Strukturelle Mehrdeutigkeit in frühneuhochdeutschen Texten

1 Einleitung

In der synchron orientierten Sprachwissenschaft ist in letzter Zeit ein verstärktes Interesse an syntaktisch annotierten Korpora zu erkennen. Für das Gegenwartsdeutsche wären hier das TIGER- bzw. NEGRA-Projekt zu nennen, in deren Rahmen bereits umfangreiche, syntaktisch annotierte Zeitungskorpora entstanden sind (BRANTS et al. 1999, 2002). Gerade in früheren Sprachstufen sind digitalisierte und linguistisch aufbereitete Korpora als Datenquelle allerdings von noch größerer Relevanz, denn eine introspektive Datenerhebung ist hier im Gegensatz zum Gegenwartsdeutschen nicht möglich, eine Untersuchung hat also notwendigerweise immer korpusbasiert zu erfolgen. Dementsprechend gibt es für bestimmte Sprachstufen wie das Mittelenglische mit der Penn-Datenbank schon größere, syntaktisch annotierte Korpora (KROCH/TAYLOR 2000). Auch für die einzelnen Sprachperioden des Deutschen werden derzeit Referenzkorpora aufgebaut, die mit linguistischen Informationen angereichert sind. Dazu gehören neben den DFG-Projekten zum Althochdeutschen (Berlin, Frankfurt/M. und Jena), Mittelhochdeutschen (Bochum und Bonn) und Frühneuhochdeutschen (Bochum, Halle und Potsdam) auch das GerManC-Projekt zum frühen Neuhochdeutschen (1650 – 1800) an der Universität Manchester und das ISWOC-Projekt (Information Structure and Word Order Change in Germanic and Romance Languages, BECH/EIDE 2011), das unter anderem syntaktische Informationen zum Althochdeutschen enthalten wird. Das hier vorgestellte Korpus ist ein syntaktisch annotiertes Korpus des Frühneuhochdeutschen, das im Rahmen eines Pilotprojekts von 2003 bis 2005 an der Universität des Saarlandes mit dem Ziel entstanden ist, an Texten, die sich sowohl durch große Varianz auf allen Ebenen des Sprachsystems als auch durch eine große Komplexität ihrer Phrasen und Sätze auszeichnen (ADMONI 1980), die Möglichkeiten einer halbautomatischen Annotation zu erproben. Basierend auf den Erfahrungen aus diesem Pilotprojekt sollen dann größere Textmengen aus dem Frühneuhochdeutschen im Baumbankformat aufbereitet und als annotiertes Referenzkorpus auf einer geeigneten Plattform frei zugänglich zur Verfügung gestellt werden. Eine solche Baumbank historischer Texte ermöglicht es dann, ausgesuchte Fragestellungen der historischen Syntax gezielter und auch in quantitativer Hinsicht zu untersuchen.¹ Darüber hinaus stellt die hohe Komplexität aus annotatorischer Sicht auch eine besondere Herausforderung dar, was die Qualität bzw. Konsistenz der Annotation angeht.

Wir werden im Folgenden das syntaktisch annotierte MERCURIUS-Korpus zum Frühneuhochdeutschen vorstellen und dabei sowohl auf die Textauswahl wie auch auf die gewählte Annotationsweise näher eingehen. Anhand von morphologischen Strukturen wie N-N-Komposita und Partikelverben sollen dann exemplarisch die Probleme disku-

tiert werden, die sich durch die Herausbildung dieser Lexeme aus syntaktischen Phrasen für die Annotation in den frühneuhochdeutschen Texten ergeben, und ein möglicher Umgang mit Problemen dieser Art aufgezeigt werden.

2 Eine Baubank frühneuhochdeutscher Zeitungstexte

2.1 Korpus

Das MERCURIUS-Korpus (DEMSKE 2007) – benannt nach einem der enthaltenen Texte – besteht aus bislang zwei Jahrgängen frühneuhochdeutscher Zeitungstexte, welche mit syntaktischen Informationen angereichert sind. Dabei handelt es sich um den ‘Nordischen Mercurius’ (M) von 1667 und die Monatsschrift ‘Annus Christi’ (AC), die 1597 erschienen ist. Damit stehen zwei Texte zur Verfügung, die beide aus einer eher späten Phase des Frühneuhochdeutschen stammen. Trotz der immer noch großen Varianz, die für diese Periode der deutschen Sprachgeschichte charakteristisch ist, erleichtert die Nähe zum Gegenwartsdeutschen den Einsatz eines Annotationsschemas, das für das Gegenwartsdeutsche entwickelt wurde. Das annotierte Korpus umfasst bislang etwa 170.000 Wortformen, wobei knapp über 130.000 Wortformen bzw. 7.500 Sätze aus dem Nordischen Mercurius und etwas mehr als 40.000 Wortformen und 1.000 Sätze aus dem Annus Christi stammen. Damit liegt für das Frühneuhochdeutsche ein erstes syntaktisch annotiertes Korpus vor, mit dem nicht nur qualitative, sondern auch quantitative Auswertungen mit vertretbarem Aufwand möglich sind.

2.2 Annotation

Bevor eine Annotation der jeweiligen Texte geleistet werden konnte, mussten die vorhandenen Zeitungsjahrgänge digitalisiert und segmentiert werden.² Die syntaktische Annotation erfolgte mittels des Werkzeugs ‘Annotate’, welches zur Annotation gegenwartsdeutscher Texte im Rahmen der TIGER- bzw. NEGRA-Baubank-Projekte entwickelt wurde (BRANTS et al. 1999, 2002). Das hybride Annotationsschema TIGER, das hierbei zur Anwendung kommt, unterscheidet zwischen drei syntaktischen Annotationsebenen, nämlich der Wort-, Phrasen- und Funktionsebene. Die Annotation der Wortarten erfolgt nach dem STTS (Stuttgart-Tübingen-Tagset, vgl. SMITH/EISENBERG 2000). Dem folgt – wie in Abb. 1 zu sehen ist – die Auszeichnung der Phrasenstruktur, welche rund umrahmt dargestellt wird, und schließlich die Einbindung einer funktionalen Ebene (eckig umrahmt). Die Annotation selbst wird halbautomatisch vorgenommen, was in diesem Fall heißt, dass dem Annotator die Lösungsvorschläge unterbreitet werden, die mit der höchsten Wahrscheinlichkeit zutreffen. Diese Vorschläge beruhen auf dem Markov-Modell (BRANTS 1999), welches mit statistischen Daten arbeitet, die aus bereits annotierten Teilen des Korpus stammen. Auf diese Weise werden zum einen auf der Wortartebene und zum anderen auf der phrasalen Ebene Vorschläge für potentiell richtige Strukturen unterbreitet. Bei den Vorschlägen auf der Wortartebene kann der Annotator entweder den unterbreiteten Vorschlag annehmen oder einen anderen Vorschlag aus einer nach Wahrscheinlichkeit geordneten Liste auswählen. Was die Vorschläge auf der

phrasalen Ebene angeht, kann der Annotator die vorgeschlagene Phrasenstruktur entweder akzeptieren oder sie entsprechend modifizieren. Während der Annotation wird dem Annotator jeweils ein vollständiger Satz präsentiert. Er kann jedoch jederzeit im Korpus blättern und vorausgehende oder nachfolgende Sätze ansehen. Die Annotation selbst wird stets von zwei Personen vorgenommen, die im sog. ‘double keying’-Verfahren jeweils denselben Teiltext selbständig annotieren und danach ihre Versionen computergestützt miteinander abgleichen. Auf diese Weise kann mit einem Mehr an Arbeitsaufwand höchstmögliche Konsistenz gewährleistet werden, wie BRANTS (2000) gezeigt hat.

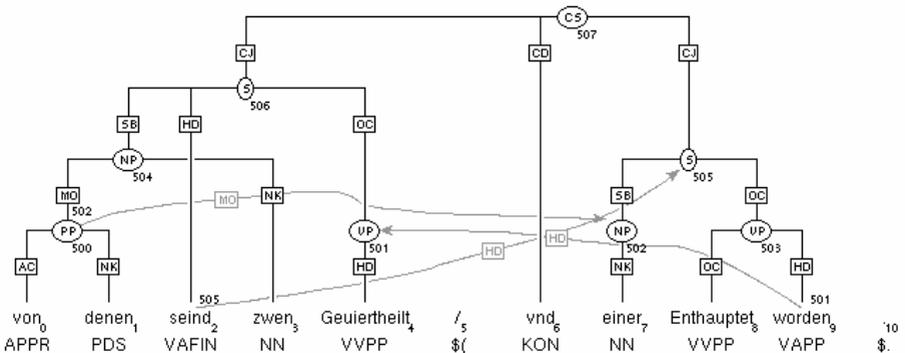


Abb. 1 Annotationsbeispiel

Ein wichtiges Merkmal ist weiterhin, dass versucht wurde, die Baumdiagramme möglichst übersichtlich zu gestalten und trotzdem alle wichtige Informationen zu integrieren. Diese Herangehensweise ist an den flachen Baumstrukturen zu erkennen. Ein Beispiel für eine solche Struktur stellt die Präpositionalphrase *von denen* in Abb. 1 dar. Präposition und nominaler Kopf sind Schwesterknoten einer Phrase, so dass Phrasen vom Typ PP problemlos gefunden werden können. Informationen über die interne Struktur der Präpositionalphrase enthält die flache Annotation dagegen nicht. Diese Vereinfachung erlaubt jedoch ein Annotieren größerer Textmengen, da die Bearbeitungszeit durch die Annotatoren wesentlich verkürzt werden kann.

Ferner besteht die Möglichkeit, sich überkreuzende Kanten wie auch sog. sekundäre Kanten zu erstellen, welche es ermöglichen, diskontinuierliche oder elliptische Strukturen darzustellen, die im Frühneuhochdeutschen häufig zu finden sind. So wird in der Abb.1 mit der sekundären Kante, die vom Verb *seind* zum Satzknos S führt, angezeigt, dass dieses Verb semantisch und strukturell nicht nur in den ersten, sondern – wenn auch nicht in dieser Flexionsform³ – ebenso in den zweiten Satz gehört, obwohl es an dieser Stelle ausgelassen wurde. Entsprechende Informationen werden auch durch die sekundären Kanten vom Verb *worden* und von der Präpositionalphrase *von denen* erfasst. Mit dem innerhalb des TIGER-Projekts geschaffenen Suchwerkzeug TIGERsearch kann

schließlich sichergestellt werden, dass speziell auf syntaktische Phänomene ausgerichtete Suchanfragen bearbeitet werden können (LEZIUS 2002, VOORMANN/LEZIUS 2002).⁴

3 Strukturelle Mehrdeutigkeit

In diesem Abschnitt soll es um ein Problem gehen, das sich verstärkt bei der Annotation historischer Texte stellt und den Zusammenhang von synchroner Variation und Sprachwandel betrifft. Im Besonderen geht es um die Frage, wie bei der Anreicherung eines Textes mit syntaktischen Informationen mit sprachlichen Mustern umzugehen ist, für die aufgrund des Kontextes nicht entschieden werden kann, ob es sich um die Ausgangs- oder die Zielstruktur eines sprachlichen Wandels handelt. Diese Frage soll im Folgenden exemplarisch für Univerbierungsprozesse im Deutschen diskutiert werden, die der Entstehung von N-N-Komposita und Partikelverben zugrunde liegen.

3.1 Uneigentliche Komposita

Wortbildungsmuster mit vermeintlichen Genitivformen als Erstglied sind im Gegenwartsdeutschen sehr verbreitet (1). Dass es sich bei der linken Komponente jedoch nicht um eine Genitivform handelt, zeigen Daten wie unter (1b): Feminine Nomina, die kein *-(e)s* in ihrem Paradigma haben, unterstützen die Annahme, dass es sich lediglich um eine morphologisch gesehen neutrale Fuge handelt und nicht um ein Flexionssuffix (RAMERS 1997).

- (1) a. Gottesbeweis
b. Liebesbeweis

Demgegenüber stehen ähnlich aussehende, jedoch phrasal zu interpretierende Formen mit pränominalem Genitiv (2). Genitivische Nominalphrasen können ihrem Kopfnomen vorangehen, wenn es sich um Eigennamen handelt (2a), bestimmte Pronomina (2b) oder – wie in (2c) – um belebte Individualnomina, wobei letztere Variante markiert ist. Sicher noch stärker markiert ist die pränominale Position für unbelebte Individualnomina (2d).

- (2) a. Julius Telefon
b. dessen Buch
c. ?des Königs Kleider
d. ??des Briefes Herkunft

Diese einfache Unterscheidung zwischen morphologischer und syntaktischer Struktur ist im Frühneuhochdeutschen jedoch häufig nicht möglich. So kann beispielsweise für Ausdrücke wie (3) nicht entschieden werden, ob es sich um eine syntaktische oder eine morphologische Struktur handelt. Hier könnte entweder eine durch einen pränominalen Genitiv modifizierte Nominalphrase (3a) oder ein Nominalkompositum (3b) vorliegen.

- (3) Kriegß Expedition⁵ (AC 177)⁶
a. [NP [NP Kriegß] Expedition]
b. [N [N Kriegß] [N Expedition]]

Die Getrennt- bzw. Zusammenschreibung bestimmter Strukturen ist dabei nur bedingt aussagekräftig. Oft liegen nämlich identische Strukturen in zweierlei Schreibung vor (4). Dementsprechend wird auch davon ausgegangen, dass sich die oben erwähnten Komposita per Reanalyse aus einem entsprechenden syntaktischen Muster entwickelt haben (DEMSKE 2001, 305).

- (4) a. KriegsWaffen (M 7153)
b. Kriegs Waffen (M 209)

Erschwerend kommt in diesem Fall hinzu, dass die für das Gegenwartsdeutsche erwähnten Beschränkungen hinsichtlich des Vorkommens pränominaler Genitive für das Frühneuhochdeutsche nicht gelten. Das heißt, dass es im Frühneuhochdeutschen grundsätzlich möglich ist, Nomina aller semantischer Subklassen als pränominale Modifikatoren zu benutzen. Theoretisch ist deshalb sowohl eine Analyse des Belegs unter (3) als morphologische wie auch als syntaktische Struktur möglich.

Bevor wir jedoch unsere Herangehensweise an dieses Annotationsproblem vorstellen, soll zunächst eine Übersicht über die im Korpus belegten Typen von Kompositionsbildungen gegeben werden. Aus (5) wird ersichtlich, dass nicht nur Nomina als 'Erstglieder' infrage kommen, sondern auch Vertreter anderer Wortarten. So ist das determinierende Element in (5a) ein Adverb, in (5b) eine Präposition und in (5c) ein Verb. Auch Negationspartikeln (5d), Konfixe (5e) bzw. Adjektive (5f) sind als Konstituenten von Komposita vertreten.

- (5) a. wider eroberung (AC 339)
b. Vor Jahre (M 435)
c. Aufziech Brucken (AC 414)
d. nicht annehmung (M 1596)
e. Vice Cancellario (M 6198)
f. Groß Vezier (M 2205)

Darüber hinaus sind die möglichen Kompositionsbildungen nicht auf Nominalkomposita beschränkt. Genauso finden sich nämlich auch Adverbien, welche durch Partikeln (6a) erweitert sein können, und auch in Verbindung mit Adjektiven kommen entsprechende Muster vor, in denen Zahlwörter (6b), Adjektive (6c), Verben (6d) oder Nomina (6e) als Modifikatoren links vom adjektivischen Kopf erscheinen.

- (6) a. als baldt (AC 224)
b. sechs tägliche Robat (AC 33)
c. nechst gelegne Spital (AC 44)
d. glaub würdig (M 1042)
e. Mast und Segel loß (AC 118)

Was die Annotation angeht, so sind die meisten der oben erwähnten Daten unproblematisch. In nahezu allen Fällen mit einem Adjektiv oder Adverb als Kopf muss von einer Kompositionsbildung ausgegangen werden, da sich eine syntaktische Modifikationsbeziehung schnell ausschließen lässt. So ist es bspw. ausgeschlossen, dass Adjektive durch Verben modifiziert werden (6d).

Allerdings gibt es auch schwieriger einzuordnende Fälle: Als Erstes möchten wir nochmals auf die anfangs genannte Struktur (3) – hier wiederholt als (7) – zurückkommen.

- (7) Doch begehren die Stände / daß zu dieser **Kriegß Expedition**, so wol auf die Gränitzhäuser zu Obristen / Rittmeistern / Hauptleuten vnnd andern Befelchshabern / geboren Behaimb / oder auß derer zu diesem Königreich gehörende / Länder / bürtige Personen dazu daugentlich / vor andern gefürdert vnnd gebraucht werden möchten (AC 177)
- a. [NP [NP Kriegß] Expedition]
 b. [N [N Kriegß] [N Expedition]]

Wir nehmen an, dass eine komplexe Phrase ausschließlich dann vorliegt, wenn mindestens eine der beiden folgenden Bedingungen zutrifft:

- (i) Das Erstglied wird in dem betreffenden Fall modifiziert respektive spezifiziert.
 (ii) Es liegt keine Adjazenz zwischen Erstglied und potentiellm Kopf vor.

Infolgedessen muss für Fälle wie (7) angenommen werden, dass es sich um ein komplexes Wort und keine komplexe Phrase handelt. Genau entgegengesetzt müssen hingegen Daten wie (8) gesehen werden. Hier können nur phrasale Strukturen vorliegen: In (8a) wird das Nomen *Türcken* durch einen Determinierer spezifiziert, während in (8b) und (8c) sogar beide Bedingungen für das Vorliegen einer komplexen Phrase erfüllt sind. Zum einen wird das jeweils linke Nomen durch einen Determinierer spezifiziert, zum anderen ist keine Adjazenz gegeben, denn in (8b) interveniert die Präpositionalphrase *auß spania* zwischen den beiden Nomina, in (8c) der adjektivische Modifikator *grosser*.

- (8) a. die darinn gelegehe Heiducken aber [...] **der Türcken Obristen** sampt noch
 13 zu Boden gelegt [...] (AC 567)
 b. in des Königs auß spania Gewalt (M 537)
 c. der Türcken grosser Verlust (M 5575)

In allen anderen Fällen, die nicht der Form ‘N + N’ entsprechen, wird entsprechend verfahren. So könnte in (9a) grundsätzlich auch ein attributives Adjektiv und damit eine komplexe Phrase vorliegen. Das fehlende Flexionssuffix kann hierbei nicht als hinreichendes Kriterium für eine Kompositionsanalyse betrachtet werden, denn im Frühneuhochdeutschen fehlen overt Kasusmarkierungen häufig. Ebenso könnte man für (9b) annehmen, dass das linke Adjektiv zusammen mit dem rechten eine komplexe Phrase bildet, zumal im Gegenwartsdeutschen bei ähnlich gelagerten Ausdrücken beide Interpretationen möglich sind, was sich in der Schreibung widerspiegelt (siehe z.B. *nahe*

gelegen vs. *nahegelegen*). Solche Fälle werden in der Baumbank jedoch analog zu den oben formulierten Bedingungen (i) und (ii) als Kompositionsbildungen behandelt.

- (9) a. Groß Vezier (M 2205)
b. nächst gelegne Spital (AC 44)

Ein ebenfalls mehrdeutiger Fall, der Ähnlichkeiten mit (7) aufweist, aber noch schwieriger zu beurteilen ist, liegt in (10) vor. Dadurch, dass auch das Kopfnomen *Tages* im Genitiv steht, kann aufgrund der Oberflächenstruktur nicht entschieden werden, ob sich der Determinierer auf das rechte oder auf das linke Nomen in der Nominalphrase bezieht. Es ist also anhand des Beispiels nicht zu klären, ob Bedingung (i) vorliegt oder nicht. In Abhängigkeit von der gewählten Interpretation kann die Analyse aussehen wie in (10a) bzw. (10b).

- (10) des Reichs Tages (M 105)
a. [_{NP} [_{NP} des Reichs] Tages]
b. [_{NP} des [_N Reichs Tages]]

Für diese Fälle gehen wir in der MERCURIUS-Baumbank davon aus, dass sich der Determinierer auf das Kopfnomen bezieht, der Sprachwandel also vollzogen ist. Dies bringt sowohl positive als auch negative Aspekte mit sich: Zum einen kann so zwar nicht garantiert werden, dass die vorgeschlagene Lösung für wirklich jedes Datum philologisch angemessen ist. Dies ist jedoch auch nicht unser Anspruch innerhalb der Korpusaufbereitung: Die Baumbank soll vielmehr als Datenquelle für die philologische Analyse bestimmter (ambiger) Strukturen dienen. Deshalb ist es wichtig, dass den Annotatoren klare Kriterien hinsichtlich der Annotation der Daten zur Verfügung stehen und dass alle einschlägigen Fälle mit einer TIGERsearch-Suche gefunden werden können. Die Auffindbarkeit kann in diesem Fall mittels der Suche nach dem dazugehörigen Label, das gleich vorgestellt werden soll, gewährleistet werden. Zum anderen wird so – und dies ist, was die Annotation angeht, der maßgebliche Punkt – eine konsistente und zügige Annotation gewährleistet.

Kommen wir nun zur technischen Seite der beiden unterschiedlichen Analysen: Eine Annotation als pränominaler Genitiv kann innerhalb des TIGER-Schemas problemlos geleistet werden. Dazu wird die jeweilige Phrase mit der Funktion GL (= Genitiv links) in die jeweilige Nominalphrase eingehängt (siehe hierzu Abb. 2). Liegt dagegen trotz Spatium eine morphologische Struktur vor, so liefert das Annotationsschema kein eigenes Label für die Auszeichnung der linken Wortteilkonstituente. Wir haben uns deshalb zur Einführung eines neuen Labels entschlossen. Mittels KOMPE (= Kompositions-Erstglied) kann auf Wortartebene markiert werden, dass es sich beim Kern der betroffenen Phrase um eine Kompositionsbildung handelt (vgl. Abb. 3). Damit sind die einschlägigen Belege im Korpus für entsprechende Recherchen identifizierbar.

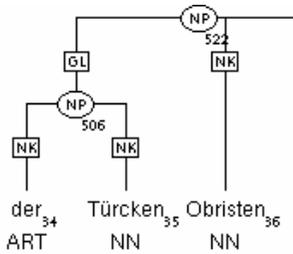


Abb. 2 Phrasale Struktur

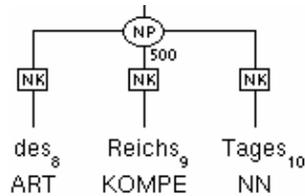


Abb. 3 Komposition

3.2 Partikelverben

Etwas anders gelagert sind die Probleme, die sich bei der Annotation von Partikelverben ergeben, weshalb diese hier gesondert behandelt werden. Partikelverben sind morphologisch komplexe Verben wie *anrufen*, deren Erstglieder – die sog. Verbpartikeln – formal gesehen identisch mit Präpositionen (11a), Adverbien (11b), Adjektiven (11c) oder Nomina (11d) sind. Meistens kann jedoch nicht von der Bedeutung der selbständigen Wörter auf die Bedeutung der Verbpartikeln geschlossen werden (DUDEN 2009, HELBIG/BUSCHA 2007).

- (11) a. aufstehen ('sich erheben' oder 'wach werden')
 b. zusammenschlagen ('jmdn. niederschlagen')
 c. schönreden ('beschönigen')
 d. teilnehmen ('bei etw. dabei sein')

Dadurch, dass Verbpartikeln formal wie frei vorkommende Wörter aussehen, kann es vor allem bei Partikelverben mit Adjektiven, Adverbien und Nomina als Erstglied zu einer Verwechslung mit korrespondierenden syntaktischen Strukturen kommen. Bei Präpositionen stellt sich dieses Problem nicht, da sie – wenn sie frei vorkommen – stets mit einem Komplement auftreten. Handelt es sich hingegen um deadjektivische Verbpartikeln wie in (12), müssen Bedeutungsunterschiede herangezogen werden, um zwischen morphologischer und syntaktischer Struktur zu differenzieren: In (12a) enthält das Partikelverb eine lexikalisierte bzw. idiomatisierte Bedeutung, die sich nicht kompositional aus den Partikelverbkonstituenten ergibt, vgl. *kaltstellen* und die Bedeutung 'verdrängen', 'beiseite schieben'. Bei der Adjektiv-Verb-Konstruktion in (12b) ist die Bedeutung transparenter, d.h. man kann aus den einzelnen Teilen *kalt* und *stellen* auf die Gesamtbedeutung dieser Konstruktion schließen.

- (12) a. Er **stellte** seinen Konkurrenten **kalt**. (Verb + Verbpartikel)
 b. Er **stellte** die Torte **kalt**. (Verb + prädikatives Adjektiv)

Ein weiteres Unterscheidungskriterium stellt im Gegenwartsdeutschen die Betonung dar. Dieses Kriterium wird meistens zur Unterscheidung von Partikelverb und selbständigem Adverb gebraucht: Bei Partikelverben liegt der Akzent normalerweise auf der Verbpartikel (13a), während bei syntaktischen Strukturen entweder nur das Verb oder beide Teile – Adverb und Verb – betont sein können (13b).⁷

- (13) a. Ich konnte alle meine Freunde **WIE**dersehen.
b. Er konnte nach der Augenoperation **wieder SE**hen/ **WIE**der **SE**hen.

Die mithilfe der oben beschriebenen linguistischen Kriterien herausgearbeitete Analyse spiegelt sich im Gegenwartsdeutschen in der Orthografie wider: Partikelverben werden zusammengeschrieben (14a), während Getrenntschreibung ein visuelles Zeichen dafür ist, dass es sich um syntaktische Strukturen handelt (14b).⁸

- (14) a. Er hat seinen Konkurrenten **kaltgestellt**. (Partikelverb)
b. Er wollte die Torte **kalt stellen**. (Verb + prädikatives Adjektiv)

Im Unterschied zum Gegenwartsdeutschen bereitet die Unterscheidung zwischen Partikelverb und syntaktischer Struktur in frühneuhochdeutschen Texten ein Problem, weil die Orthografie in diesem Fall wie auch schon bei den oben besprochenen N-N-Komposita häufig keinen Hinweis darauf gibt, ob eine morphologische oder syntaktische Struktur vorliegt. Wie (15) zeigt, kann beispielsweise die Verbindung von Präposition und Verb sowohl zusammen- als auch getrenntgeschrieben vorkommen, obwohl es sich in diesem Fall wegen des fehlenden präpositionalen Komplements um ein Kompositum handeln muss.

- (15) a. [...] der Bremischen Friedenshandlung **beyzuwohnen** (M 2174)
b. [...] den FriedensTractaten **bey zu wohnen** (M 1323)

Für die Fälle, in denen sich Verben mit Nomina, Adjektiven und Adverbien verbinden, stellt sich jedoch die Frage, wie bei fehlender sprachlicher Kompetenz entschieden werden kann, ob ein Partikelverb vorliegt oder nicht, da die Prosodie in den historischen Textkorpora als diagnostisches Kriterium ausscheidet und semantische Unterschiede oftmals schwierig zu erkennen sind. Dass es sich hier nicht um ein zu vernachlässigendes Problem weniger Einzelfälle handelt, ergibt sich aus der Beobachtung, dass viele im Gegenwartsdeutschen gebräuchliche Partikelverben erst im Verlauf der jüngeren deutschen Sprachgeschichte als Produkte von Inkorporation, also einem Prozess entstanden sind, bei dem ein Verbstamm einen anderen mit ihm in Beziehung stehenden Wortstamm morphologisch eingliedert, wodurch ein komplexes Verb gebildet wird (EISENBERG 2006, 234). So hat sich das komplexe Verb *standhalten* wohl ursprünglich aus der syntaktischen Objekt-Verb-Konstruktion *Stand halten* entwickelt. Viele komplexe Verben mit Adjektiven und Adverbien als Erstglieder wie *krankliegen*, *bereitstellen*, *hinaufgehen* sind ebenfalls aus einst syntaktischen Konstruktionen entstanden (EISENBERG 2006, 332ff). Für zahlreiche frühneuhochdeutsche Belege von Partikelverbkandidaten ist folglich zu entscheiden, ob der Inkorporationsprozess bereits abgeschlossen ist oder im Einzelfall

noch syntaktische Strukturen vorliegen. In Fällen wie (16) muss folglich geklärt werden, wie eine morphosyntaktische Annotation aussehen muss, die den sprachlichen Fakten gerecht wird, ohne deren Analyse vorwegzunehmen. Anders ausgedrückt muss es für jeden Nutzer der Baumbank möglich sein, in dem annotierten Korpus qua TIGERsearch alle für die Herausbildung von Partikelverben einschlägigen Varianten qualitativ und quantitativ zu erfassen. Gleichzeitig muss gewährleistet sein, dass den Annotatoren unzweifelhafte Kriterien für die Annotation von Mustern wie (16) an die Hand gegeben werden.

(16) [...] vor den Holländern wol **weg gekommen** sey (M 4248)

3.2.1 Annotation von Partikelverben: Generelles

Was die tatsächliche Vorgehensweise beim Annotieren angeht, so wird hier ähnlich wie bei den im vorausgehenden Abschnitt diskutierten Kompositionsfällen verfahren, mit der Ausnahme, dass eine zusätzliche Annahme getroffen werden muss. Von einer Verbpartikel und somit einer morphologischen Struktur sprechen wir also, wenn

- (i) die mutmaßliche Verbpartikel und das Verb in der Grundstruktur adjazent stehen,⁹
- (ii) keine Modifizierung bzw. Spezifizierung der mutmaßlichen Verbpartikel vorliegt (bei Nomen, Adjektiven oder Adverbien als Verbpartikel) bzw. bei der mutmaßlichen Verbpartikel kein Komplement auftritt (bei Präpositionen als Verbpartikel),
- (iii) das fragliche Lexem als Partikelverb im heutigen Deutsch belegt ist.

Die Fälle in (17) werden somit als Partikelverben analysiert und das Erstglied mit dem Wortart-Tag PTKVZ (= abgetrennter Verbzusatz) versehen (vgl. dazu Abb. 4).

(17) a. [...] bald soll **mit getheilet** werden. (M 2794)
 b. [...] darauff sie folgenden Tag **fort ziehen** wollen / (AC 787)

Bei den Daten in (18) wird entsprechend der genannten Kriterien jedoch von phrasalen Strukturen ausgegangen: In (18a) und (18b) liegt keine Adjazenz vor, außerdem treten die fraglichen Konstituenten entweder als Kern einer Präpositionalphrase auf, vgl. *mit* in (18a) sowie Abb. 5, oder sie werden durch weitere Phrasen modifiziert wie *fort* in (18b). In (18c) stehen *leid* und *thun* zwar adjazent, *leid* wird jedoch durch das Indefinitpronomen *kein* spezifiziert, weswegen man in diesem Fall von einer vollständigen Nominalphrase ausgehen muss.

(18) a. [...] einige Stücke **mit** sich führete. (M 372)
 b. [...] sich [...] **fort** nach Niemiecrowice begeben [...] (M 2387)
 c. [...] niemandt kein **leid** thun [...] (AC 294)

Wie bereits bei den Kompositafällen können auch hier etwaige Zweifelsfälle mit TIGERsearch gefunden werden, indem man nach Wörtern mit dem Wortartlabel PTKVZ sucht, die adjazent zu einem Verb stehen.¹⁰

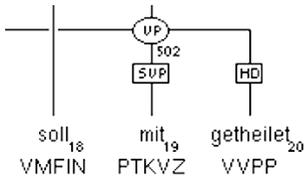


Abb. 4 Morphologische Struktur

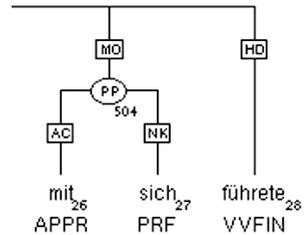


Abb. 5 Syntaktische Struktur

Am Anfang dieses Abschnitts wurde die gewählte Vorgehensweise, Partikelverben im Frühneuhochdeutschen unter anderem vor dem Hintergrund gegenwartsdeutscher Sprachkompetenz zu annotieren, angesprochen. Allerdings finden sich in älteren Texten des Deutschen auch Konstruktionen, die gemäß der Kriterien (i) und (ii) als Partikelverben zu analysieren sind, ohne eine Entsprechung im Gegenwartsdeutschen zu haben (19), wofür es unterschiedliche Gründe geben kann: So kennt man aus dem Gegenwartsdeutschen das Partikelverb *vortragen*, nicht aber *für tragen* (19a). Ebenso werden in älteren Texten Lexeme verwendet, die im Gegenwartsdeutschen fehlen, wie z.B. *anhero* in (19b). Und schließlich können auch komplexe Verben vorkommen, bei denen wir das Erstglied als solches aus dem Gegenwartsdeutschen kennen, die Kombination aus dieser Partikel mit dem darauffolgenden Verb im Gegenwartsdeutschen jedoch kein usualisiertes Wort mehr ist, wie z.B. im Fall von *nieder säbeln* in (19c).¹¹

- (19) a. [...] wie [...] denselben Ständen **für getragen** worden. (AC 15)
 b. [...] etliche der ihrigen Officirer **anhero zu schicken** [...] (M 5909)
 c. [...] auch andere [...] **nieder säbelten**. (M 370)

Für derartige Fälle wird von uns vorgeschlagen, diese als Partikelverben zu annotieren, falls sich ein Bezug zu einem Partikelverb im Gegenwartsdeutschen herstellen lässt. So wird *für tragen* in (19a) als Partikelverb analysiert, da es im Sinne vom Gegenwartsdeutschen *vortragen* gebraucht wird. Auch *anhero schicken* in (19b) wird als Partikelverb analysiert, weil es hier in der Bedeutung von 'hierherschicken' gebraucht wird, das im Gegenwartsdeutschen als Partikelverb angesehen wird. Bei *nieder säbeln* kann man ebenfalls von einem Partikelverb im Sinne von gegenwartsdeutschen Partikelverben *niederschlagen* bzw. *niederstechen* ausgehen. Weitere ähnliche Zweifelsfälle sind in (20) aufgeführt:

- (20) a. [...] derohalben sie außtruckentlichen beuelch haben / höchst gedachten Römischen Keyser / von dieser Fürsten wegen / zuermahnen / daß er / die erwünschte gelegenheit nit **fürüber gehn** lasse / [...] (AC 579)
 b. Von Cerigo wird berichtet/ daß die 40. Türkische Galeen / so den Succurs

in Candia gebracht / sich bereits wiederumb zu rücke begeben hätten mehr grosse Häupter und Völcker abzuholen und **über zuführen**.¹² (M 423)

Dagegen werden die Fälle unter (21) nicht als Partikelverben annotiert, da sie sich im Gegenwartsteutschen nicht mit Partikelverben, sondern mit syntaktischen Strukturen in Verbindung bringen lassen: So ist aus dem Kontext klar ersichtlich, dass *dahinden geblieben* in (21a) sich auf die syntaktische Struktur *dahinten/hinten geblieben* und *dahinden lassen* in (21b) sich ebenfalls auf eine solche Struktur *dahinten/hinten lassen* zurückführen lassen und nicht mit den formal sehr ähnlichen gegenwartsdeutschen Partikelpräfixverben *hinterbleiben* und *hinterlassen* zu verwechseln sind. Solche Fälle werden folglich als syntaktische Strukturen annotiert.

- (21) a. Die vnserer seind zu Gran / den 25. Aprilis / glücklich wider ankommen / deren mit mehr als 8. **dahinden geblieben** [...] (AC 288)
 b. [...] hat doch nach langem Scharmützlen / mit schaden der seinigen weichen / vnd seinen Rennfahnen **dahinden lassen** müssen. (AC 496)

Unter der dargestellten Vorgehensweise lassen sich nahezu alle Fälle erfassen. Nichtsdestotrotz gibt es im Kontext der Partikelverben noch einige Konstruktionen, deren Annotation problematisch bleibt. Auf diese Konstruktionen soll im Folgenden eingegangen werden.

3.2.2 Annotation von Partikelverben: Sonderfälle

Doppelpartikel oder Adverb? Aufgrund der bereits formulierten Kriterien könnte man in Fällen wie (22) bei *mit* davon ausgehen, dass es mit der anderen Verbpartikel eine Art Doppelpartikel bilden könnte, wie etwa das Verb *miteinbeziehen* im Gegenwartsteutschen.

- (22) a. Die Römisch-Catholische Städte sind zwar von den Protestantischen sehr ersuchet worden / sich dieser Sachen **mit anzunehmen** [...] (M 1690)
 b. [...] Etliche haben im abziehen ihrer Hanschke das Fleisch von allen Fingern **mit abgezogen**. (M 1715)

Unseres Erachtens ist jedoch die Analyse von *mit* als Adverb im Sinne von ‘auch’, ‘ebenfalls’ in diesen Fällen aus folgenden Gründen vorzuziehen: Zum einen scheint das Vorkommen von *mit* als Adverb im Deutschen – im älteren (23a) wie im heutigen (23b) – im Gegensatz zu anderen Präpositionen ziemlich gebräuchlich zu sein (mehr dazu vgl. auch die Wörterbucheinträge zum Lexem *mit* im Deutschen Wörterbuch von Jakob und Wilhelm GRIMM (1999) für das ältere und im DUDEN-Wörterbuch (2007) für das heutige Deutsch):

- (23) a. [...] **mit** nach Genua gezogen [...] (AC 364)
 b. Er wollte **mit** nach oben gehen/ **mit** Tee trinken/ sich **mit** auf die Prüfung vorbereiten.

Im Gegensatz dazu kommen im Gegenwartsdeutschen kaum Doppelpartikeln mit *mit* vor,¹³ schon gar nicht solche, die aus *mit* und einer weiteren Verbpartikel mit einfachem oder komplexem Adverb als Erstglied zusammengesetzt werden, wie die Beispiele aus dem Korpus in (24) zeigen. Das spricht unserer Meinung nach gegen die Lexikalisierung dieses Sprachmusters im Deutschen, weswegen wir uns in diesen Fällen für eine Annotation von *mit* als Adverb (s. Abb. 6) entschieden haben:

- (24) a. [...] 30. der reichsten Kauffleute [...] **mit** weggeführt [...] (M 6794)
 b. [...] eine schöne Koppel Türkischer Pferde **mit** über nehmen [...]¹⁴ (M 5992)
 c. [...] viel 1000. Seelen nebest dem Herrn Michowsky **mit** hinweg geführt [...] (M 401)

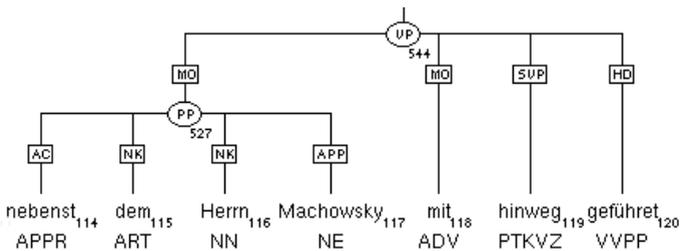


Abb. 6 *mit* als Adverb

Verbpartikel oder Adverb? Verben in Verbindung mit *da*-Adverbien wie *dahin* sollen hier insofern als besondere Gruppe hervorgehoben werden, als dass diese Adverbien ebenfalls im Gegenwartsdeutschen im Zusammenhang mit bestimmten Verben – entweder als Verbpartikel (25a) oder als Richtungsadverb (25b) – auftreten können:

- (25) a. dahindämmern
 b. dahin bringen

Der Unterschied zwischen morphologischer und syntaktischer Struktur im Gegenwartsdeutschen wird bei diesem Typ semantisch vorgenommen: Bei einer Konstruktion wie *dahin* + Verb wird dann von einer syntaktischen Struktur ausgegangen, wenn *dahin* in dieser Verbindung eine transparente Bedeutung besitzt, d.h. im Sinne von ‘an diesen Ort’, ‘so weit’ gebraucht wird (26a). Ansonsten ist bei *dahin* von einer Verbpartikel aus-

zugehen (26b). In diesem Fall weist die gesamte Konstruktion meistens eine lexikalisierte Bedeutung auf (RAT FÜR DEUTSCHE RECHTSCHREIBUNG 2006).

- (26) a. Wir können zu Fuß **dahin gehen** (im Sinne von ‘an diesen Ort gehen’)
 b. Die Tage sind **dahingegangen** (im Sinne von ‘verstreichen’)

Was die Analyse und Annotation im Frühneuhochdeutschen angeht, so soll hier das gleiche Kriterium angewendet werden. D.h. bei der Bestimmung der *da*-Wörter entscheidet die Bedeutung: Bildet *dahin* zusammen mit dem Verb eine gemeinsame lexikalisierte Bedeutung, dann wird es als Verbpartikel (PTKVZ) annotiert (27). Wenn *dahin* jedoch eine transparente Bedeutung im Sinne von ‘an diesen Ort’ besitzt, wird es als Adverb (ADV) annotiert (28).

- (27) **morphologische Struktur**
 Alle Consilia allhier gehen noch **dahin**_{PTKVZ} [...] (M 434)
- (28) **syntaktische Struktur**
 Man hat nun vor/ den Handel auch mit denen von Algiers/ Tripoli/ Tunis und Biserta zu stabiliren/ und einen Ambassadeur mit etlichen Schiffen **dahin**_{ADV} zu senden [...] (M 673)

Verbpartikel oder Präpositionalphrase? Dieser Abschnitt beschäftigt sich mit dreigliedrigen Strukturen im Frühneuhochdeutschen, wie sie in (29) zu sehen sind:

- (29) a. Bald darnach haben die Kriegsleut [...] / 20. Türcken nidergehawet / vnnd 13. gefangen / vnnd solche **zu ruck** gebracht. (AC 954)
 b. DJe Spanisch Armada, [...] ist auff den Spanischen Gränitzen angelangt / vnd durch vngewitter [...] **zu Grund** gangen. (AC 1059)

Aufgrund der Verbindung aus Präpositionalphrase und Verb könnte man diese Konstruktionen den syntaktischen Strukturen zuordnen. Aus dem Gegenwartsdeutschen weiß man jedoch, dass die Strukturen in (29) unterschiedliche Wege eingeschlagen haben: Die Verben mit adjazentem *zu ruck* (29a) haben sich durch Univerbierung zu den Partikelverben *zurückbringen*, *zurückgehen*, *zurückfahren* usw. entwickelt. Bei Konstruktionen wie in (29b)¹⁵ handelt es sich dagegen nicht um Partikelverben, da hier der Prozess der Univerbierung noch nicht abgeschlossen zu sein scheint: Das zeigt sich beispielweise darin, dass diese Konstruktionen im Gegenwartsdeutschen immer noch als syntaktische Strukturen analysiert werden, die aus einem Verb und einem adverbialen Modifikator bestehen. Lediglich bei den ersten beiden Gliedern sind zwei verschiedene Analysen – als Adverb oder als Präpositionalphrase – erlaubt.¹⁶

- (30) a. zugrunde_{ADV}/ [PP [zu_{APPR}] [Grund_{ENN}]] gehen *vs.* *zugrundegehen
 b. zuwege_{ADV}/ [PP [zu_{APPR}] [Weg_{ENN}]] bringen *vs.* *zuwegebringen

Vor diesem Hintergrund sollen hier unterschiedliche Analysen für die beiden Strukturen unter (29) angenommen werden: Die Fälle in (29a) werden als Partikelverben und somit

als morphologische Strukturen analysiert (s. dazu Abb. 7). Dafür spricht auch – freilich als eher schwächeres Argument –, dass man bei Fällen wie (29a) neben dem getrennten *zu ruck* auch das zusammengeschriebene *zuruck* (31) belegt findet:

- (31) [...] die gelegenheit der Bawren aber ware jhme vnbekannt / derowegen er wider **zuruck gezogen**. (AC 296)

Konstruktionen wie in (29b) werden hingegen als syntaktische Strukturen annotiert. Eine vorgenommene Suche nach ähnlichen Fällen im Korpus (diese ist bei solch wenigen und speziellen Fällen wie in (31) im Gegensatz zum gesamten Bereich der Partikelverben relativ schnell zu leisten) führt zu dem Ergebnis, dass solche Konstruktionen in unserem Korpus immer mit einem Spatium aufgetreten sind, was dafür spricht, die Erstglieder als Präpositionalphrasen (32a) (s. dazu Abb. 8) und nicht als Adverb-Komposita (32b) zu analysieren.

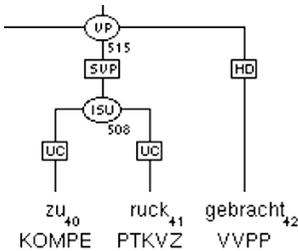


Abb. 7 Partikelverb

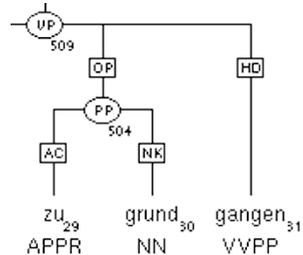


Abb. 8 Präpositionalphrase

- (32) a. [PP [zu_{APPR}] [Grunde_{NN}]] gehen
 b. *[AVP [zu_{KOMPE}] [Grunde_{ADV}]] gehen

4 Zusammenfassung

Wir haben im vorliegenden Artikel das MERCURIUS-Korpus vorgestellt, welches bislang aus zwei mit syntaktischen Informationen angereicherten frühneuhochdeutschen Zeitungstexten besteht. Um die spezifischen Probleme aufzuzeigen, die sich bei der syntaktischen Annotation historischer Korpora ergeben, haben wir uns im vorliegenden Beitrag mit sprachlichen Strukturen beschäftigt, denen sowohl Wortstatus als auch phrasaler Status zugeschrieben werden kann. Für Partikelverben und (andere) Kompositionsbildungen wurde gezeigt, wie solche Muster in einem zeitlich vertretbaren Rahmen dennoch deskriptiv adäquat annotiert werden können.

Anmerkungen

¹Wie das händisch bereits in den Arbeiten geschehen ist, die in der Reihe ›Bausteine zur Sprachgeschichte des Neuhochochdeutschen‹ ab 1964 erschienen sind.

²Näheres zur Digitalisierung im Allgemeinen und der Segmentierung der Satzgrenzen im Speziellen findet sich in DEMSKE et al. (2004) und DEMSKE (2007).

³Diese als nicht-parallele Koordinationsellipse bekannte Konstruktion ist im Frühneuhochdeutschen sehr gebräuchlich.

⁴Mit TIGERsearch als Suchwerkzeug kann eine Suche nicht nur nach einzelnen Lexemen, Wortarten oder Phrasen getätigt werden, sondern es können auch spezielle und explizite Suchanfragen ausgeführt werden, wie z.B. die Suche nach einer Nominalphrase, die ein vorangestelltes Genitivattribut enthält, oder nach einem Satz, der einen Komplementsatz enthält.

⁵Wir werden bei den Korpusbeispielen im Folgenden nur dann einen Kontext mitliefern, wenn ein solcher erforderlich ist, um die entsprechende Konstruktion verstehen zu können. Ansonsten werden wir – mit Ausnahme des zu Illustrationszwecken gezeigten Satzes in (7) – stets nur die fragliche Struktur selbst abbilden.

⁶Angegeben ist jeweils nach der Abkürzung für den spezifischen Text die Satznummer im elektronischen Korpus.

⁷Mehr zu diesen und ähnlichen Unterscheidungskriterien vgl. RAT FÜR DEUTSCHE RECHTSCHREIBUNG (2006) und ZUR NEUREGELUNG DER DEUTSCHEN RECHTSCHREIBUNG (2011).

⁸Die orthografische Regelung ist hierbei selbstverständlich nur als Reflex der linguistischen Verhältnisse zu sehen. Dies heißt, dass selbst ein fälschlicherweise mit Spatium versehenes *kaltstellen* trotzdem aufgrund linguistischer Kriterien wie der semantischen Verschiebung ein Partikelverb bleibt, und zwar unabhängig von der Schreibung.

⁹Gemeint sind hier die Fälle, in denen Verbpartikel und Verb topologisch gesehen zusammen in der rechten Satzklammer auftreten.

¹⁰Zu Näherem bzgl. einer ähnlichen Suchanfrage siehe MÜLLER (2006).

¹¹So kommt *niedersäbeln* bspw. im mehrbändigen DUDEN-Wörterbuch (1999) mit mehr als 200.000 Stichwörtern gar nicht mehr vor.

¹²Das Verb *über führen* wird hier im Sinne vom gegenwartsdeutschen Partikelverb *herüberführen* und nicht dem Partikelpräfixverb *überführen* gebraucht.

¹³Bei der Recherche in Online-Duden (www.duden.de) stößt man nur auf wenige Doppelpartikelverben mit *mit* wie bspw. *miteinbeziehen* oder *miteinrechnen*.

¹⁴*Über* wird hier im Sinne von 'herüber' verwendet.

¹⁵Ähnliche Fälle wären *zu Wege/zuwege*, *zu Stande/zustande*, *zu Lasten/zulasten bringen*.

¹⁶Die Label in den nachfolgenden Beispielen wurden aus Annotate übernommen: PP steht für 'Präpositionalphrase', APPR für 'Präposition', NN für 'Nomen', AVP für 'Adverbphrase', KOMPE für 'Kompositions-Erstglied' und ADV für 'Adverb'.

Quellen

[M] = Mercurius 1667. Nordischer Mercurius. Welcher kürzlich vorstellet/was in diesem 1667. Jahre an Novellen aus Europa einkommen ist. Hamburg 1667.

[AC] = Annus Christi 1597. Historische erzöhlung/der fürnembsten Geschichten vnd handlungen/so in diesem 1597. Jahr (...) abgelaufen (...). Rorschach 1597. Nachdruck: Walluf-Nedeln: Sändig 1977.

Literatur

Admoni, V. G. (1980). Zur Ausbildung der Norm der deutschen Literatursprache im Bereich des neuhochdeutschen Satzgefüges (1470-1730). Ein Beitrag zur Geschichte des Gestaltungssystems der deutschen Sprache. Akademie-Verlag, Berlin.

Bech, K. und Eide, K. G. (2011). The annotation of morphology, syntax and information structure in a multilayered diachronic corpus. JLCL, 26(2):13–24.

- Brants, S. (1999). Tagging and Parsing with Cascaded Markov Models – Automation of Corpus Annotation. DFKI, Saarbrücken Dissertations in Computational Linguistics and Language Technology Bd. 6.
- Brants, T. (2000). Inter-annotator agreement for a German newspaper corpus. In Proceedings of Second International Conference on Language Resources and Evaluation LREC-2000. Athen.
- Brants, S., Dipper, S., Hansen, S., Lezius, W., und Smith, G. (2002). The TIGER treebank. In Proceedings of the Workshop on Treebanks and Linguistic Theories, 24–41, Sozopol.
- Brants, T., Wojciech, S., und Uszkoreit, H. (1999). Syntactic annotation of a german newspaper corpus. In Proceedings of the ATALA Treebank Workshop, 69–76, Paris.
- Demske, U. (2001). Merkmale und Relationen: Diachrone Studien zur Nominalphrase des Deutschen. de Gruyter, New York, Berlin.
- Demske, U. (2007). Das Mercurius-Korpus: Eine Baumbank für das Frühneuhochdeutsche. In Kallmeyer, W. und Zifonun, G., Hgg., Sprachkorpora – Datenmengen und Erkenntnisfortschritt, 91–104. de Gruyter, Berlin, New York.
- Demske, U., Frank, N., Laufer, S., und Stiemer, H. (2004). Syntactic interpretation of an early new high german corpus. In Kübler, S., Nivre, J., Hinrichs, E., und Wunsch, H., Hgg., Proceedings of the Third Workshop on Treebanks and Linguistic Theories (TLT 2004), 175–182, Tübingen.
- Deutsches Wörterbuch von Jacob und Wilhelm Grimm (1999). Nachdruck der Erstausgabe von 1854-1961. Deutscher Taschenbuchverlag, München.
- Duden (1999). Duden – Das große Wörterbuch der deutschen Sprache in 10 Bänden. 3., völlig neu bearbeitete und erweiterte Auflage. Band 6. Dudenverlag, Mannheim, Leipzig, Wien, Zürich.
- Duden (2007). Duden – Deutsches Universalwörterbuch. 6., überarbeitete und erweiterte Auflage. Dudenverlag, Mannheim, Wien, Zürich.
- Duden (2009). Duden – Die Grammatik. 8., überarbeitete Auflage. Dudenverlag, Mannheim, Wien, Zürich.
- Eisenberg, P. (2006). Grundriss der deutschen Grammatik. Band 1: Das Wort. 3. Auflage. Metzler, Stuttgart, Weimar.
- Helbig, G. and Buscha, J. (2007). Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht. 6. Auflage. Langenscheidt, Berlin, München, Wien.
- Kroch, A. und Taylor, A. (2000). Penn-Helsinki Parsed Corpus of Middle English.
- Lezius, W. (2002). TIGERSearch – Ein Suchwerkzeug für Baumbanken. In Busemann, S., Hg., Proceedings der 6. Konferenz zur Verarbeitung natürlicher Sprache, Saarbrücken.
- Müller, S. (2006). Quantitative Korpusanalyse für die Grammatiktheorie: Introspektion vs. Korpus. In Kallmeyer, W. und Zifonun, G., Hgg., Sprachkorpora – Datenmengen und Erkenntnisfortschritt, 70–90. de Gruyter, Berlin, New York.
- Ramers, K. H. (1997). Die Kunst der Fuge: Zum morphologischen Status von Verbindungselementen in Nominalkomposita. In Dürscheid, C., Schwarz, M., und Ramers, K. H., Hgg., Sprache im Fokus. Festschrift für Heinz Vater zum 65. Geburtstag, 33–46. Niemeyer, Tübingen.
- Rat für deutsche Rechtschreibung (2006). Deutsche Rechtschreibung. Regeln und Wörterverzeichnis. Entsprechend den Empfehlungen des Rats für deutsche Rechtschreibung. Überarbeitete Fassung des amtlichen Regelwerks mit den Nachträgen aus dem Bericht 2010.

- Smith, G. und Eisenberg, P. (2000). Kommentare zur Verwendung des STTS im NEGRA-Korpus. Manuskript.
- Voormann, H. und Lezius, W. (2002). TIGERin – Grafische Eingabe von Benutzeranfragen für ein Baumbank-Anfragewerkzeug. In Busemann, S., Hg., Proceedings der 6. Konferenz zur Verarbeitung natürlicher Sprache.
- Zur Neuregelung der deutschen Rechtschreibung ab 1. August 2006 – Nachtrag (2011). Sprachreport. Extra-Ausgabe. Institut für Deutsche Sprache, Mannheim.