



Universitätsverlag Potsdam

Manfred Stede (Hrsg.)

Handbuch Textannotation

Potsdamer Kommentarkorpus 2.0

Potsdam Cognitive Science Series | 8

Manfred Stede (Hrsg.)

Handbuch Textannotation

Potsdamer Kommentarkorpus 2.0

Universitätsverlag Potsdam

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de/> abrufbar.

Universitätsverlag Potsdam 2016

<http://verlag.ub.uni-potsdam.de/>

Am Neuen Palais 10, 14469 Potsdam

Tel.: +49 (0)331 977 2533 / Fax: 2292

E-Mail: verlag@uni-potsdam.de

Die Schriftenreihe **Potsdam Cognitive Science Series** wird herausgegeben von Johannes Haack, Dr. Christiane Wotschack und Dr. Michael Dambacher

ISSN (print) 2190-4545

ISSN (online) 2190-4553

Das Manuskript ist urheberrechtlich geschützt.

Hinweis: Nachfolgende Neuauflagen der Publikation werden nur online über den Publikationsserver der Universität Potsdam abrufbar sein.

Online veröffentlicht auf dem Publikationsserver der Universität Potsdam

URN [urn:nbn:de:kobv:517-opus4-82761](http://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-82761)

<http://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-82761>

Zugleich gedruckt erschienen im Universitätsverlag Potsdam

ISBN 978-3-86956-343-5

Abstract

The Potsdam Commentary Corpus is a collection of newspaper texts belonging to the ‚commentary‘ genre. The public part consists of 175 texts from *Märkische Allgemeine Zeitung* that have been manually annotated for syntax, coreference, connectives, and rhetorical structure. Further layers will be added to future releases of the corpus. This book assembles the annotation guidelines that have been used for that public part, as well as for other portions, where other layers of annotation have been experimented with. Most of the guidelines will be applicable to similar genres, and also to other languages.

Zusammenfassung

Das Potsdamer Kommentarkorpus ist eine Sammlung von Zeitungstexten, die dem Genre ‚Kommentar‘ zuzuordnen sind. Der öffentlich verfügbare Teil besteht aus 175 Texten aus der *Märkischen Allgemeinen Zeitung*, die hinsichtlich Syntax, Koreferenz, Konnektoren und Rhetorischer Struktur manuell annotiert wurden. Weitere Ebenen werden bei zukünftigen Korpusversionen hinzukommen. Dieses Buch enthält die Annotationsrichtlinien, die der Bearbeitung des öffentlichen Teils des Korpus zugrunde lagen, sowie auch anderer Teile, bei denen mit weiteren Annotationsebenen experimentiert wurde. Die meisten der Richtlinien werden auch für ähnliche Text-Genres und für andere Sprachen verwendbar sein.

Inhaltsverzeichnis

Abstract	v
Zusammenfassung	vii
Inhaltsverzeichnis	ix
Tabellenverzeichnis	xv
Abbildungsverzeichnis	xvii
Manfred Stede	
Einführung	1
Manfred Stede	
Hintergrund und Entwurfsentscheidungen	5
2.1 Entwurfsentscheidungen beim Korpusdesign	5
2.2 Diskurssegmentierung	7
2.3 Negation	9
2.4 Referentielle Ausdrücke und Koreferenz	10
2.5 Aboutness Topik	12
2.6 Konnektoren und ihre Argumente	15
2.7 Inhaltszonen	17
2.8 Rhetorische Struktur	17
2.9 Argumentationsstruktur	19
Manfred Stede, Sara Mamprin, Andreas	
Peldszus	
Diskurssegmentierung	23
3.1 Hintergrund	23
3.2 Annotationsvorgehen: Übersicht	24

3.3	Schritt 1: Teilung in Sinneinheiten	25
3.4	Schritt 2: Unterteilung der Sinneinheiten in Segmente	28
3.5	Schritt 3: Unterteilung zusammengesetzter Sätze . . .	32
3.5.1	Unterordnung: Nebensätze	33
3.5.2	Nebenordnung	38
3.5.3	Parenthetische Einschübe	40
3.6	Zusammenfassung der Typ-Bezeichner	42
3.7	Weitere Beispiele	42
3.7.1	Hauptsätze, Fragmente	42
3.7.2	Subjekt-, Objekt- und Prädikativsätze	43
3.7.3	Adverbialsätze	44
3.7.4	Attributsätze	44

André Herzog, David Kaupat, Manfred Stede

Negation	45
4.1	Hintergrund	45
4.2	Grundsätzliches Vorgehen	45
4.3	Negationsoperatoren und ihr Skopus	47
4.3.1	Lexikalische Negationsoperatoren	47
4.3.2	Morphologische Negationsoperatoren: Affixe	52
4.4	Hinweise	52
4.5	Abschwächungsoperatoren	53

Christian Chiarcos, Manfred Stede, Saskia

Warzecha		
Nominale referentielle Ausdrücke	55
5.1	Hintergrund: Begriffe	55
5.2	Markables: Typen und Ausdehnung	56
5.2.1	Begriffsbestimmung: PM	58
5.2.2	Begriffsbestimmung: SM	60
5.2.3	Begriffsbestimmung: NM	60
5.2.4	Keine Markables	61
5.2.5	Einige syntaktische Aspekte von Markables .	62
5.3	Attribute	63
5.4	Analyse eines Beispieltexts	66

Christian Chiarcos, Manfred Stede, Saskia

Warzecha

Nominale Koreferenz	71
6.1 Hintergrund: Begriffe	71
6.2 Annotationsanweisung	73
6.2.1 Ziel der Annotation	73
6.2.2 Annotationsverlauf	73
6.3 Einzelheiten	75
6.3.1 Antezedens-Entscheidung	75
6.3.2 Behandlung diskontinuierlicher Markables	75
6.3.3 Rekursive Einbettung	75
6.3.4 Gruppen-Referenz	76
6.4 Relationen	77
6.4.1 Anaphorische Relationen	78
6.4.2 Kataphern	79
6.4.3 Ambige Antezedenten	81
6.5 Attribute	82
6.5.1 Referentiality	82
6.5.2 Ambiguity	83
6.5.3 Anaphora-type	84
6.6 Prinzipien im Überblick	84
6.7 Analyse eines Beispieltexts	85

Sara Mamprin, Manfred Stede

Aboutness Topik	89
7.1 Annotationsschema	89
7.1.1 Überblick	89
7.1.2 Der Topik-Bereich	94
7.1.3 Multiple Topiks und Ambiguität	95
7.1.4 Segmente ohne Aboutness Topik	97
7.1.5 Identifikation von Aboutness Topiks (<i>ab</i>)	99
7.1.6 Identifikation von Frame-Setting Topiks (<i>fs</i>)	102
7.1.7 Identifikation von kontrastiven Topiks (<i>co</i>)	104
7.2 Annotationsvorgehen	105
7.3 Analyse eines Beispieltexts	107

Manfred Stede

Konnektoren und Argumente 111

8.1 Hintergrund und Ziel der Annotation 111

8.2 Konnektoren 112

 8.2.1 Subordinierende Konjunktionen 113

 8.2.2 Koordinierende Konjunktionen 113

 8.2.3 Adverbiale 114

 8.2.4 Präpositionen und Postpositionen 114

 8.2.5 Diskontinuierliche Konnektoren 115

8.3 Die Argumente von Konnektoren 116

 8.3.1 Argumenttypen 117

 8.3.2 Zuordnung des internen und externen Arguments zum Konnektor 119

8.4 Hilfestellung zur Annotation 121

 8.4.1 Konnektoren: Grundlegende Merkmale 121

 8.4.2 Wörter in nicht-Konnektor Lesart 121

 8.4.3 Bestimmung der Argumente 126

 8.4.4 Umgang mit Ellipsen 128

 8.4.5 Überlappungen, Einbettungen und multiple Konnektoren 129

 8.4.6 Details zur Notation 130

Andreas Peldszus, Manfred Stede

Inhaltszonen 133

9.1 Hintergrund 133

9.2 Das Inventar der Inhaltszonen 134

9.3 Die Annotationsschritte 136

 9.3.1 Vorbemerkungen zur Segmentierung 136

 9.3.2 Zentrale These 137

 9.3.3 Fazit 138

 9.3.4 Argumente für die These 138

 9.3.5 Argumente gegen die These 139

 9.3.6 Hintergrundinformation 141

9.4 Analyse eines Beispieltexts 141

Manfred Stede

Rhetorische Struktur 145

10.1 Hintergrund 145

10.2 Ziel der Annotation 146

10.3 Segmentierung 149

10.4 Relationen 152

 10.4.1 Format der Definitionen 153

 10.4.2 Primär pragmatische Relationen 156

 10.4.3 Primär semantische Relationen 163

 10.4.4 Textuelle Relationen 169

 10.4.5 Multinukleare Relationen 171

 10.4.6 Hinweise zur Relationszuweisung 173

10.5 Annotationsverlauf 178

10.6 Analyse eines Beispieltexts 181

Andreas Peldszus, Saskia Warzecha, Manfred Stede

Argumentationsstruktur 185

11.1 Allgemeines 185

11.2 Schritt 1: Segmentierung 187

 11.2.1 Bildung von EDUs 187

 11.2.2 Bildung von ADUs 189

11.3 Schritt 2: Identifikation der Gesamtthese 192

11.4 Schritt 3: Zuweisung der argumentativen Stimme . . 193

11.5 Schritt 4: Zuweisung der Funktion und des Bezugspunkts 195

 11.5.1 Stützen 195

 11.5.2 Anfechten 199

 11.5.3 Anfechtungen erwidern 201

11.6 Zusammenfassung: Übersicht der Annotationsschritte 203

11.7 Analyse eines Beispieltexts 204

Literaturverzeichnis 209

Tabellenverzeichnis

3.1	Typen der Diskurssegmente	41
7.2	Attribute für die Topik-Annotation	93

Abbildungsverzeichnis

2.1	Argumentationsschema nach Toulmin (1958)	20
10.2	Analysebeispiel RST (Exzerpt aus: George Packer: Suffering. <i>The New Yorker</i> 42, 2010)	146
10.3	Analysebeispiel RST 2 (Text fiktiv)	148
10.4	Beispieltext ‚Laubag‘ (maz-5297)	181
10.5	RST Analyse für Beispieltext ‚Laubag‘ (maz-5279) . .	184
11.6	Stützungsrelationen	196
11.7	Angriffe des Opponenten auf die Argumentation des Proponenten	199
11.8	Erwiderung des Proponenten auf die Anfechtungen des Opponenten.	201
11.9	Die argumentative Struktur des Beispieltextes	208

Einführung

Manfred Stede

Das *Potsdamer Kommentarkorpus* (engl. *Potsdam Commentary Corpus*; im Folgenden: PCC) ist eine Sammlung von deutschsprachigen Meinungstexten aus der Presse; im Zentrum stehen Kommentare aus der *Märkischen Allgemeinen Zeitung* (MAZ) und dem *Tagesspiegel am Sonntag* (TaS) aus den Jahren 2002–2007. Im Falle des TaS stammen die Kommentare aus der im Berliner Lokalteil erscheinenden Rubrik *Pro & Contra*, bei der jeweils eine aktuelle Streitfrage durch einen Hintergrundartikel erläutert und von zwei Autoren des Blattes (oder, selten, von Gästen) kommentiert wird. Die Fragestellung ist meistens eine ja/nein-Frage; einer der Beiträge argumentiert jeweils für eine positive Antwort, der andere dagegen.

Die einzelnen Texte wurden auf unterschiedlichen Beschreibungsebenen annotiert – teilweise automatisch (*part of speech*) oder teilautomatisch (Syntax), im Wesentlichen aber manuell. Je nach der Verfügbarkeit von Annotationen und der Publikationsfähigkeit der Primärtexte bilden wir Subkorpora, die wir teilweise öffentlich verfügbar machen. Dies gilt derzeit für 175 Kommentare aus der MAZ, die dank einer Vereinbarung mit der *Märkischen Allgemeinen* über die PCC Webseite zugänglich sind. Im Sommer 2014 wurden die Texte mit fünf Annotationsebenen als *PCC 2.0* bereitgestellt, dieses Korpus ist in (Stede u. Neumann, 2014) beschrieben, wo auch Verweise auf Publikationen zu bisherigen Anwendungen des PCC (automatisches rhetorisches Parsing, Untersuchung der Korrelation von rhetorischer Struktur und Koreferenz etc.) gegeben werden. Das PCC 2.0 löst die in (Stede, 2004) dokumentierte Version des Korpus ab, die bisher für Forschungszwecke individuell gegen Unterzeichnung einer Lizenzvereinbarung weitergegeben wurde.

Neben den 175 MAZ Kommentaren bildet *ProCon40* ein anderes wichtiges (aber bisher nicht öffentlich verfügbares) Teilkorpus, das sich aus 40 TaS Kommentaren zusammensetzt, die über PCC 2.0 hinaus auf weiteren Ebenen annotiert sind.

Das vorliegende Handbuch fasst die Beschreibungen und Annotationsanweisungen für die einzelnen Ebenen zusammen. Sie wurden in unterschiedlichen Annotationsstudien für verschiedene Teilkorpora verwendet; einige (Inhaltszonen, Argumentation) sind insbesondere für Pro & Contra Texte konzipiert. Inwieweit sich die hier zusammengestellten Richtlinien auch für andere Textsorten außerhalb des Bereichs „Kommentar“ einsetzen lassen, soll an dieser Stelle außer Betracht bleiben. Anhand des PCC wurden eine Reihe von Studien zur Bestimmung des *inter-annotator agreement* für einzelne Ebenen durchgeführt; Hinweise darauf finden sich auf der Webseite des Korpus.

Die meisten der Richtlinien sind für das PCC von Grund auf neu entstanden. Dies gilt nicht für diejenigen zu den Themen Koreferenz und Informationsstruktur. In den entsprechenden Kapiteln ist vermerkt, dass es sich um Weiterentwicklungen früherer Richtlinien – an denen auch andere Autoren beteiligt waren – handelt.

Dieses Handbuch enthält *kein* Kapitel zur syntaktischen Annotation. Der Grund ist, dass die PCC Texte auf der Syntax-Ebene direkt nach dem TIGER Schema (Brants u. a., 2002) und dessen Annotationsrichtlinien bearbeitet wurden.

Die einzelnen Annotationen liegen meist in verschiedenen Formaten vor, darunter die Originalformate der jeweiligen Annotationswerkzeuge, sowie das standoff XML-Format PAULA (Potsdamer Austauschformat für Linguistische Annotation) (Dipper, 2005). Die Ebenen sind also technisch voneinander unabhängig; auch inhaltlich sind sie nur teilweise aufeinander bezogen, wie unten erläutert wird. Dies entspricht dem Vorgehen bei einer *Mehrebenen-Annotation*, wie sie etwa in (Stede, 2007) beschrieben ist.

Nach einer Übersicht über grundlegende Entwurfsentscheidungen zum Korpus und zu einzelnen Ebenen (Kap. 3) geben die nachfolgenden Kapitel die einzelnen Annotationsrichtlinien wieder, wie

sie in der Korpuserstellung verwendet wurden und werden. Viele Richtlinien können unabhängig voneinander genutzt werden, aber in einigen Fällen nehmen Ebenen durch Definition der *Markables* aufeinander Bezug. Dies ist jeweils zu Beginn des Anweisungskapitels beschrieben und sei hier zusammengefasst:

- *Nominale Koreferenz* basiert auf *Nominale referierende Ausdrücke*.
- *Aboutness Topik* basiert auf *Diskurssegmentierung* und auf *Nominale referierende Ausdrücke*.
- *Rhetorische Struktur* basiert auf *Diskurssegmentierung*.
- *Argumentationsstruktur* basiert auf *Diskurssegmentierung*.

Die ‚Zielgruppe‘ für die Richtlinien, also die potenziellen Annotierenden, sind Studierende, die über eine linguistische Grundausbildung (1–2 Studienjahre) verfügen; einige der Richtlinien setzen aber auch nur sehr geringe Vorkenntnisse voraus.

Die einzelnen Anweisungskapitel beschränken sich auf die rein inhaltlichen Instruktionen, d. h., sie sind unabhängig von einem technischen Annotationswerkzeug. Hinweise zur Arbeit mit einem solchen Werkzeug werden in unseren Studien jeweils separat ausgegeben, bzw. die Annotierenden erhalten ein Training zu dem zu benutzenden Werkzeug.

Ein letzter Hinweis: Die verschiedenen Kapitel wurden nicht im Hinblick auf die Verwendung maskuliner, femininer und geschlechtsneutraler Formen homogenisiert, d. h. es finden sich in diesem Band verschiedene Vorgehensweisen. Für alle Kapitel sei deshalb an dieser Stelle betont, dass etwaige Personenverweise im generischen Maskulinum stets als geschlechtsneutral gemeint sind.

Hintergrund und Entwurfsentscheidungen

Manfred Stede

In den einzelnen Annotationskapiteln finden sich die konkreten *Anweisungen* – die AnnotatorInnen erhalten dort nur wenige Informationen über die theoretischen Hintergründe und Annahmen. Selbige sind deshalb im vorliegenden Kapitel gebündelt, das zunächst aber einige grundsätzliche Überlegungen zum Design des Gesamtkorpus vermittelt.

2.1 Entwurfsentscheidungen beim Korpusdesign

Das PCC ist offensichtlich in keinerlei Hinsicht ein *balanciertes* Korpus. Die Zielsetzung war, Texte insbesondere (wenn auch nicht ausschließlich) für die linguistische Untersuchung vor allem von pragmatischen Phänomenen der Subjektivität und Argumentation zu sammeln. Alle Exemplare gehören demselben Genre des Zeitungskommentars an, lassen sich dort aber verschiedenen Sub-Genres zuordnen (vgl. Schneider u. Raue, 1996). Ebenso wenig ist Balance durch die Abdeckung einer größeren Zahl von Zeitungen gegeben – ein medienlinguistisch orientierter Vergleich der Kommentierungsmuster verschiedener Zeitungen oder verschiedener Kommentar-Gattungen bildet derzeit keinen Untersuchungsgegenstand des PCC.

Die Textauswahl erfolgte aufgrund der folgenden Erwägungen:

- MAZ: Eine Regionalzeitung, die auf der Meinungsseite ein breites Spektrum von Themen (von lokalen Ereignissen bis

zur Weltpolitik) verhandelt und sich dabei einer relativ einfachen Sprache bedient – was für viele Analyse-Ebenen die Annotation erleichtert. Texte wurden über einen bestimmten Zeitraum pauschal gesammelt, ohne dass eine thematische oder sonstige Filterung erfolgt wäre.

- TaS: Für eine Argumentationsanalyse sind MAZ-Kommentare häufig wenig ergiebig, weil sie zwar subjektiv gefärbt sind, aber nicht unbedingt auf eine konkrete Einstellungsänderung beim Leser zielen. Pro- und Contra-Kommentare tun dies hingegen unausweichlich, weil sie explizit „gegeneinander“ gesetzt sind. Sie sind damit einer Argumentationsanalyse eher zugänglich, was ihre Aufnahme in das Korpus begründet. Die Texte wurden über einen bestimmten Zeitraum aus dem Lokalteil des TaS gesammelt, ohne dass eine thematische oder sonstige Filterung erfolgt wäre.

In beiden Textmengen stellt sich naturgemäß gelegentlich das Problem der Kontextabhängigkeit (temporal, lokal, thematisch): Kommentare bewerten meistens das, was andernorts in der Zeitung als Nachricht präsentiert wird. Im TaS ist dies immer ein Hintergrundbericht, der unmittelbar über den beiden Kommentaren erscheint. Diese Berichte wurden mitgesammelt, sind aber derzeit nicht Gegenstand der Annotation. In der MAZ werden entsprechende Nachrichtenmeldungen im jeweiligen Ressort erschienen sein; diese wurden aber nicht gezielt gesucht und mitgesammelt. – Die überwiegende Mehrzahl der PCC Texte erscheint aber aus sich heraus gut verständlich.

Das Untersuchungsziel ist wie gesagt die Analyse verschiedener Diskursphänomene und ggf. ihrer Korrelation mit syntaktischen Strukturen. Für eine solche Annotation von Texten auf verschiedenen Ebenen stellt sich stets die grundlegende Frage, ob Ebenen miteinander verzahnt oder voneinander unabhängig erstellt werden sollen. Der erste Weg wurde beispielsweise in der *Prague Dependency Treebank*¹ gewählt, wo eine zugrunde liegende linguistische

¹<https://ufal.mff.cuni.cz/pdt2.0/> (Zugriff: 30.9.2015)

Theorie die syntaktische Struktur als Bausteine der Koreferenzanalyse und auch der Diskursstruktur postuliert. Das PCC ist demgegenüber weitgehend modular angelegt: Die meisten Annotationen erfolgen unmittelbar auf dem Quelltext. Zwei Ausnahmen sind die Ebenen ‚referierende Ausdrücke‘ und ‚Diskurssegmente‘, die für einige andere Ebenen (siehe die Übersicht auf Seite 3) die Grundlage bilden. Insofern ist der Ansatz theorieneutral: Es geht nicht um eine geschlossene Konzeption linguistischer Elemente, sondern um eine Menge phänomenbezogener Einzelanalysen, deren Ergebnisse dann im Nachhinein miteinander korreliert werden können.

2.2 Diskurssegmentierung

Segmentierung ist ein grundlegender Annotationsschritt der Textanalyse, auf dem eine Reihe von Annotationsebenen aufbauen, im PCC namentlich Topik, Rhetorische Struktur und Argumentationsstruktur. Eine weitere für die Segmentierung relevante Domäne, die allerdings im PCC derzeit nicht untersucht wird, sind die ‚minimal units‘, innerhalb derer die Centering Theory (Grosz u. a., 1995) die verschiedenen Arten von *centers* definiert, um dann jeweils die Art des Übergangs zwischen benachbarten Segmenten zu bestimmen.

Die verschiedenen Analyseebenen nutzen allerdings nicht alle dieselben Segmentierungen: Beispielsweise ist die Beschreibungseinheit, innerhalb derer ein Topik zuzuweisen ist, nicht immer auch eine elementare Analyseeinheit der Argumentationsstruktur. Unsere Vorgehensweise ist deshalb, mit den Segmentierungsrichtlinien eine möglichst feinkörnige Gliederung zu erzielen, die jedem Segment auch einen (syntaktisch motivierten) *Typ* zuweist. Anhand der Typisierung können die anderen Annotationsebenen ihren eigenen Begriff von ‚minimalen Analyseeinheiten‘ definieren, indem sie ggf. Segmente, die qua ihres Typs für die Analyse-Ebene zu klein sind, an benachbarte größere Segmente anschließen. In dieser Weise legen beispielsweise die Richtlinien für die RST-Annotation fest, welche Arten von (im früheren Annotationsschritt identifizierten) Nebensätzen ‚minimale Einheiten‘ sind und welche nicht.

Allgemein gilt, dass der Arbeitsschritt einer Diskurssegmentierung aus zwei Gründen kompliziert werden kann: Einerseits wegen komplexer Sätze, in denen für bestimmte Zwecke die Teilsätze (‘clauses’), aber auch appositives Material als eigenständige Segmente zu definieren sind. Andererseits ergeben sich Schwierigkeiten durch syntaktisch unvollständige Einheiten, die nicht nur in der gesprochenen Sprache, sondern auch in geschriebenen Texten eine Rolle spielen; wir fassen diese unter dem Begriff ‚Fragment‘ zusammen.

Insbesondere aus computerlinguistischer Sicht ist der Gedanke attraktiv, die Segmentierung allein auf der Basis von formalen Kriterien, also frei von subjektiver Interpretation, zu definieren. Dieser Weg wurde beispielsweise in den Richtlinien von Jasinska ja u. a. (2007) beschrritten. Er führte dort zu sehr umfangreichen und komplexen Anweisungen, die die AnnotatorInnen vor nicht unbeträchtliche Schwierigkeiten stellen und die ein gründliches Training erfordern.

Im PCC, wo die letztendlichen Untersuchungsziele vorwiegend pragmatischer Natur sind, spielt subjektive Interpretation naturgemäß an verschiedenen Stellen eine Rolle (etwa bei der Zuweisung einer rhetorischen Struktur). Auch unsere Segmentierungsrichtlinien sind nicht völlig frei von subjektiven Kriterien; hierdurch werden sie kürzer und – hoffentlich – leichter handhabbar. Sie sind so konzipiert, dass die AnnotatorInnen zwar weitgehend syntaktische Kriterien anlegen, aber auch ein nurmehr intuitiv zu definierendes Konzept der ‚Sinneinheit‘ identifizieren (wie es dann eben z. B. für die Argumentation bedeutsam ist). Zentrale Aufgabe bei der Bestimmung von Sinneinheiten ist diese Klassifikation von Segmenten: Erfüllen sie ihre Funktion im Text eigenständig, oder nur im Verbund mit einem Nachbarsegment? Dafür sind die verschiedenen Typen ‚einleitendes Fragment‘, ‚abschließendes Fragment‘ und ‚Hauptsatzfragment‘ vorgesehen.

Für komplexe Sätze besteht die Problematik in der Definition von Kriterien, welche ihrer Bestandteile als separate Diskurssegmente aufzufassen sind. Zwei Beispiele mögen genügen, um die Perspek-

tivenabhängigkeit dieser Aufgabe zu illustrieren: Aus Sicht der Konnektoren-Analyse stellt sich die Frage, welche Art von Material von einem Konnektor verknüpft werden kann (vgl. Breindl u. a., 2011); für eine sprechaktororientierte Analyse hingegen ist die zentrale Frage, welcher Konstituente eine ‚illokutionäre Kraft‘ zugeschrieben werden kann. Diese beiden Sichten werden nicht immer zu denselben Antworten führen. So wird beispielsweise subordiniertes Material von verschiedenen Analyseebenen in unterschiedlicher Weise behandelt: Die Unterscheidung zwischen restriktivem und nicht-restriktivem Relativsatz wird für manche Zwecke wichtig sein, für andere nicht. Wir verwenden hier ein relativ feinkörniges Typinventar, das gleichwohl ‚theorieneutral‘ sein soll und übernehmen dabei weitgehend den Vorschlag von Bußmann (2002). Eine theoretische Grundlage für die auch von uns getroffene grundsätzliche Unterscheidung zweier Typen von Subordination wird in vielen Ansätzen vorgeschlagen, z. B. spricht van Valin jr. (2005) von *d-subordination* (subordiniertes Material ist ein Argument des Verbs) versus *ad-subordination* (subordiniertes Material bildet ein Adjunkt).

2.3 Negation

Negationen sind im Rahmen des PCC interessant, weil sie für die Analyse der (Inter-) Subjektivität eine wichtige Rolle spielen (vgl. Verhagen, 2005) und damit für genauere Betrachtungen der Argumentation bedeutsam sind.

Für die linguistische Untersuchung ist die Negation ein schwieriges Phänomen, bei dem Syntax, Semantik und Pragmatik gleichermaßen zu berücksichtigen sind. Der Skopus einer Negation kann der gesamte Satz sein oder nur eine Konstituente; in Korrekturkontexten kann es auch ein Wortteil sein (*Er hat das Studium nicht unter-, sondern abgebrochen*). Ebenso vielfältig ist das Spektrum der Negationsoperatoren; es ist in den Richtlinien benannt.

Die Annotation erfolgt direkt an der sprachlichen Oberfläche, ohne Vorgabe möglicher Skopus-Konstituenten, denn uns interessiert,

was die AnnotatorInnen semantisch/pragmatisch als „verneint“ wahrnehmen. Die so annotierten Skopi können dann posthoc zu den syntaktischen Strukturen in Beziehung gesetzt werden. Die Richtlinien fordern explizit dazu auf, den Negationsskopus *im Kontext* zu bestimmen, also bei der Entscheidung nicht nur den aktuellen Satz zu berücksichtigen.

Die große Mehrzahl von Negationsausdrücken in (PCC-) Texten scheint relativ unkompliziert, doch prinzipiell ist eine beachtliche Komplexität möglich, insbesondere Kombinationen von doppelten Negationen und *hedging*-Ausdrücken. In den Richtlinien wird dies aber nicht vertieft diskutiert, sondern nur kurz angesprochen und der Intuition der AnnotatorInnen überlassen.

2.4 Referentielle Ausdrücke und Koreferenz

Referenz-Annotation beschränkt sich im PCC derzeit auf nominale Koreferenz, es werden keine Sachverhaltsanaphern annotiert. Zudem beschränken wir die nominale Koreferenz auf die Fälle von Referenzidentität, kümmern uns also nicht um indirekte Wiederaufnahmen bzw. *bridging anaphora*. Die Richtlinien überlassen die Identitäts-Entscheidung weitgehend unkommentiert der Intuition der Annotierenden; auf Probleme der *near-identity* (vgl. Recasens u. a., 2012) gehen wir nicht ein. Gleiches gilt für die Problematik von generischen Ausdrücken und der ‚Spezifität‘ von Diskursreferenten; sie werden in der aktuellen Version der Richtlinien nicht vertieft diskutiert.

Der Ausgangspunkt für unsere Richtlinien war die Beschreibung des *PoCoS* Koreferenz-Schemas von Krasavina u. Chiarcos (2007), einige der dort getroffenen Entscheidungen wurden aber nicht unverändert übernommen, und einige Punkte wurden etwas weiter spezifiziert. Die wesentliche Änderung ist aber das Abtrennen der Richtlinien für referentielle Ausdrücke. Diese werden nun separat behandelt, weil sie im PCC die Basis für zwei andere Annotationsebenen darstellen: Koreferenz und Topik. Diese können

somit unabhängig voneinander, aber jeweils aufbauend auf den referentiellen Ausdrücken, annotiert werden (für das praktische Vorgehen wird es oft effektiv sein, die referentiellen Ausdrücke gemeinsam mit der übergeordneten Ebene, etwa der Koreferenz, zu annotieren). Wie im Koreferenz-Annotationsszenario üblich, verwenden wir für diese Ausdrücke hier den Begriff *Markable* (zu markierende Einheit).

Eine grundlegende, von Krasavina/Chiaros übernommene, Unterscheidung ist die zwischen *primären* und *sekundären* Markables: primäre Markables sind für die Lesenden *identifizierbar*, ihr Diskursreferent ist textuell, situationell oder per Welt-Wissen *gegeben* (,given'). Allerdings war die Unterscheidung bei Krasavina und Chiaros darüber hinaus durch die Zielsetzung der Annotation von referentiellen Ketten motiviert, für die nicht alle sekundären Markables annotiert werden müssen (sondern nur diejenigen, die wieder aufgenommen werden, also eine Kette beginnen). In den PCC Richtlinien hingegen werden sämtliche sekundären Markables annotiert, denn wir benötigen Vollständigkeit für die etwaige spätere Annotation des Aboutness Topik.

Darüber hinaus fordern die Richtlinien die explizite Markierung von *nichtreferentiellen* Markables, das sind beispielsweise Expletiva oder Nominalphrasen in idiomatischen Wendungen.

Insgesamt basiert die Definition der Markables und der Koreferenz auf einer Reihe von Prinzipien, die möglichst sprachübergreifend konzipiert sind. Diese Prinzipien sind in den Richtlinien erläutert; als Grundlage dafür wurden zunächst folgende Axiome festgelegt:

- (A) *Constituency*: Ein primäres oder sekundäres Markable muss eine unabhängige syntaktische Konstituente sein.
- (B) *Completeness*: Weder sub-tokens noch nicht-phrasale Nominale sind Gegenstand der Annotation, sondern allein syntaktische Wörter (Token) und Phrasen.
- (C) *Consistency*: Korrespondierende Features müssen auf eine korrespondierende Art analysiert werden.

- (D) *Maximal Analysis*: Längere anaphorische Ketten sind in der Annotation den kürzeren Ketten vorzuziehen.

Die Axiome A und B sind notwendige Vorbedingungen für die Korrespondenz einer syntaktischen Annotation mit der der Koreferenz. Axiom C impliziert, dass Annotationsprinzipien in einer Art und Weise formuliert werden müssen, welche intersubjektiv und sprachübergreifend stichhaltige Annotationsentscheidungen erlauben. Während A und B Bedingungen zur Markable-Identifikation definieren, beeinflusst C zudem die Präferenzauswahl unter potentiellen Antezedenten und motiviert die explizite Repräsentation anaphorischer Ambiguität.

2.5 Aboutness Topik

Dieses Kapitel behandelt eines der zentralen Phänomene der Informationsstruktur, nämlich das Topik. Für *Fokus* geben wir hier keine Richtlinien an, da seine Annotation in geschriebenen Texten sich in unseren Experimenten als ausgesprochen schwierig erwiesen hat (Ritz u. a., 2008). In Teilen des PCC wurde der *Informationsstatus* von referentiellen Ausdrücken annotiert, dazu wurden die Richtlinien des SFB 632 (Götze u. a., 2007) genutzt.

Je nach Perspektive werden dem *Topik* phonologische (z. B. Deakzentuierung), syntaktische (ausgezeichnete Positionen), semantische (Quantorenkopos) oder diskurspragmatische Merkmale zugeordnet. Die Phänomene werden in der Regel an gezielt konstruierten Beispielen untersucht, mit denen sich bestimmte Faktoren gut isolieren lassen. Eine Anwendung des Topikbegriffs auf authentische Texte erweist sich dann erwartungsgemäß als nicht trivial, da eine Vielzahl komplizierender Einflussfaktoren miteinander interagieren und die Kontexte beliebig vielschichtig sein können.

Für unsere Zwecke streben wir eine theorie-neutrale Annotation an, die mit anderen Beschreibungsebenen in Beziehung gesetzt werden kann. Dieses Anliegen teilen wir mit der Arbeit von Götze u. a. (2007), die den Ausgangspunkt für die hier vorliegenden Richtlinien bildete. Die zu annotierenden Merkmale und Werte wurden

praktisch unverändert übernommen, ebenso der Grundgedanke, Annotationsanweisungen entlang funktionaler Tests auszurichten und nicht durch Rekurrenz auf linguistische Merkmale an der Textoberfläche.

Während aber die SFB Richtlinien explizit nicht auf die Annotation eines Diskurstopik, sondern dezidiert auf das Satztopik zielen, ist für unsere Zielsetzungen eine kontextgebundene Analyse wichtig. Denn eine wesentliche Fragestellung für das PCC ist, wie das „topic management“ im Textverlauf funktioniert, auch wie Perspektivenwechsel in der Argumentation vorgenommen werden. Dies ist der Grund dafür, in diesem Kapitel auf den Begriff des *aboutness* Topik zu fokussieren.

Im allgemeinen wird in theoretisch-orientierten Arbeiten *givenness* als ein notwendiges Merkmal für Topiks betrachtet, etwa bei Krifka (2007) oder bei Erteschik-Shir (2007, S. 7 ff.). Givenness und aboutness bedingen einander jedoch nicht, und es erscheint nicht nützlich, bei einer Explikation der aboutness auch ein Merkmal des Informationsstatus hinzuzuziehen; dieses sollte stattdessen auf einer separaten Ebene behandelt werden.

Für eine Operationalisierung der aboutness greifen wir in den Topik-Richtlinien (Kapitel 7) in erster Linie auf den Vorschlag von Jacobs (2001) zurück, der seinerseits auf der Strawson/Reinhart Konzeption fußt. Die von Jacobs genannten drei Dimensionen informationelle Trennung, Prädikation und Adressierung sind in den Richtlinien erläutert. Das auch von Krifka (2007) als prominent eingestufte Merkmal der Adressierung im Sinne der ‚file card semantics‘ als Repräsentationssystem des ‚common ground‘ müsste für eine systematische Anwendung auf authentische Texte maßgeblich expliziert werden, was aber ein schwieriges Unterfangen darstellen dürfte. In den Richtlinien belassen wir es bei einem Appell an die Intuition der Annotierenden, das Eintragen einer Information auf der „richtigen“ Karteikarte – im Verbund mit den anderen Kriterien der aboutness – einzuschätzen.

Ein weiteres Problem bei der Behandlung von Topik in authentischen Texten ist die Definition der zugrunde liegenden linguis-

tischen Einheit, die ein Topik „haben darf“. Der idealsprachliche Hauptsatz scheidet hier als Explikandum aus, stattdessen ist festzulegen, wie einerseits mit Fragmenten unterschiedlicher Art, andererseits mit komplexen Sätzen umzugehen ist. Letzteres wird auf der theoretischen Seite als die Rekursivität von Topik-Strukturen untersucht. Für unsere Zwecke koppeln wir diese Entscheidungen an die Segmentierungsrichtlinien aus Kapitel 3 und bestimmen eine Reihe von Segment-Typen, in denen ein Topik zu lokalisieren ist (wir nennen sie TDs): Hauptsätze, Hauptsatzfragmente, Adverbialsätze und weiterführende Nebensätze. Dies stellt wohl nur eine Näherung dar und wäre ggf. zu verfeinern. Für die Annotation verzichten wir auf eine echte Rekursion, sondern geben vor, dass das Topik stets nur relativ zum kleinsten umgebenden „topik-fähigen“ Segment zu identifizieren ist.

Diese Vorgehensweise bei der Behandlung von komplexen Sätzen ist zum Teil durch Matic' u. a. (2014) motiviert, die sich mit dem Problem des Topiks in Satzkomplexen beschäftigen und grundsätzlich zwischen zwei Formen der Subordination unterscheiden (im Wesentlichen: Argumente versus Adjunkte) und für diese dann unterschiedliche Regeln der Topik-Fähigkeit formulieren. Dabei lassen sie auch *clauses* als Topiks zu, was gemäß unserer Richtlinien allerdings nicht möglich ist – wir beschränken uns in der aktuellen Version ausdrücklich auf nominale referierende Ausdrücke, so wie sie in Kapitel 5 definiert sind.

Im Fall von *Fragmenten*, der zweiten Problemquelle bei der Bestimmung der topikenthaltenden Einheit, ist zu entscheiden, ob sie Topiks enthalten; wenn ja, werden Sie dem zugehörigen HS/HSF hinzugefügt, ansonsten ignoriert.

Zur Frage etwaiger *multipler* Topiks betrachten Matic' u. a. (2014) den Fall eines Satzes mit zwei Pronomen (*I saw him at school*) und weisen darauf hin, dass manche Ansätze hier zwei Topiks konstatieren würden; im Sinne eines Strawson/Reinhart Konzepts von Topik allerdings sei es wichtig, sich für eines, das „main topic“ zu entscheiden. Dieser Auffassung schließen wir uns an und erweitern die Regel zu der Vorgabe, dass ein TD generell nur (und höchstens)

ein aboutness Topik haben soll. Die Annotierenden können aber ggf. weitere, nachgeordnete Kandidaten markieren. Als genrespezifische Regel fügen wir an, dass die in Kommentaren durchaus häufigen Pronomen der 1.ps.pl. im Falle mehrerer Kandidaten nicht zu präferieren sind.

Die gängige Charakterisierung von *Thetizität* als „all-new“ Satz rekuriert wiederum auf ein Informationsstatus (IS) Merkmal. Man könnte den thetischen Satz als Epiphänomen der IS-Annotation operationalisieren (ein TD, in dem auf keine *given* oder *accessible* Diskursreferenten referiert wird), doch ziehen wir es vor, dies im Zuge der Topik-Annotation als explizite Kategorie zu behandeln, die wir einerseits durch die „all-new“ Umschreibung charakterisieren, andererseits den gängigen Fragetest *Was ist passiert?/Was gibt's Neues?* aufführen.

Neben den aboutness Topiks sind zum Einen *frame-setting* Topiks, wie sie etwa von Krifka (2007) beschrieben werden, Gegenstand der Annotation. Zum Anderen behandeln wir *kontrastive* Topiks, die wir etwas anders als die Literatur nicht primär semantisch charakterisieren (partielle Antwort auf eine Frage; Alternativen offen lassend), sondern als textstrukturierendes Mittel: Zu einem gegebenen Oberbegriff werden im weiteren Verlauf mindestens zwei Unterbegriffe topikal eingeführt, die dann untereinander kontrastieren. Als paradigmatische Fälle von Ober-/Unterbegriff behandeln wir die Hyponymie, Meronymie und *set membership*.

2.6 Konnektoren und ihre Argumente

Als „Vorstufe“ zur Untersuchung von Textstruktur (im Sinne der RST, SDRT, o.ä.) erfasst die Annotation von Konnektoren die lexikalisch signalisierten Relationen zwischen benachbarten Textabschnitten. Die Definition des Konnektor-Begriffs übernehmen wir von Pasch u. a. (2003), mit der einen Ausnahme, dass wir auch einige Präpositionen als Konnektoren zulassen. Für die Richtlinien geben wir nur eine grobe Gliederung der syntaktischen Kategorien vor, ohne auf feinere Unterscheidungen und weitere Merkmale

(wie sie sehr ausführlich von Pasch u. a. herausgearbeitet wurden) einzugehen.

Wir nennen für jede Gruppe eine Reihe häufiger Konnektoren, weisen die Annotierenden aber darauf hin, dass diese nicht als vollständig zu betrachten sind, man beim Annotieren also nach weiteren Exemplaren Ausschau halten muss.

Die ‚Argumente‘ des Konnektors bestimmen wir hier nicht nach syntaktischen, sondern nach Inhaltlichen Gesichtspunkten: Wieviel Material wird benötigt, damit die vom Konnektor ausgedrückte Relation vollständig dargestellt ist? Natürlich ist dieses Kriterium nicht unproblematisch (z. B.: Wie definiert man ‚Vollständigkeit‘ bei fragmentarischen Strukturen bzw. Elision?), doch für unsere Zwecke ist eine diskursbezogene Sicht auf den Konnektor von größerem Interesse als eine rein syntaktische. Für die Abgrenzung der Argumente ergeben sich bei subordinierenden Konjunktionen und bei Präpositionen kaum Probleme. Bei nebenordnenden Konjunktionen und bei adverbialen Konnektoren können Argumente gelegentlich über die Satzgrenzen hinausreichen, hier muss also eine Verständnisprobe gemacht werden. Die zweite Komplikation resultiert aus modalen Einbettungen, wo wir die Annotierenden auffordern, wiederum aus dem Verständnis heraus nur „kleine“ Argumente zu markieren, also den einbettenden Modalausdruck nicht zu integrieren: *Ich glaube, (erst wird es regnen) und (später dann schneien.)*

Diese Annotationsebene weist große Ähnlichkeit zum Vorgehen in der *Penn Discourse Treebank* (Prasad u. a., 2008) auf. Da wir als weitere Ebene aber auch RST bearbeiten (Kapitel 10), verzichten wir anders als die PDTB bei der Konnektorenannotation auf die Zuweisung von Relationslabels; die Bedeutung des Konnektors muss hier also von den Annotierenden nicht expliziert werden. Dementsprechend werden auch keine unmarkierten Relationen (in der PDTB: „implicit connectives“) annotiert. Weniger zwangsläufig ist unsere Entscheidung, auch keine alternativen lexikalischen Signale (PDTB: „AltLex“) zu berücksichtigen. Dies könnte man tun, wenn der Dreh- und Angelpunkt der Annotation das allgemeine „Relationssignal“

wäre; wir interessieren uns hier aber ausschließlich für Konnektoren. Sie sind relativ klar definierbar, was den Annotationsaufwand erheblich reduziert, denn man muss bei der Bearbeitung eines Texts nicht nach einer offenen Klasse von Signalen für zugrunde liegende Relationen suchen.

2.7 Inhaltszonen

Die Konzeption der Inhaltszonen ist von der „move analysis“ nach Swales (1990) inspiriert. Im Mittelpunkt steht hier der Gedanke, dass durch ein Zonen-Inventar eine bestimmte Textsorte charakterisiert wird. Je nach Textsorte können die Zonen relativ groß und auch durch Layout-Merkmale gekennzeichnet sein; für viele Sorten sind die Zonen jedoch auf der Ebene des Einzelsatzes zuzuweisen. In der Computerlinguistik haben Teufel u. Moens (2002) das für wissenschaftliche Aufsätze demonstriert (und auch als automatische Klassifikation implementiert), wo die Zonen beispielsweise *Own Method* oder *Related Result* heißen.

Unser Ansatz folgt diesem Vorgehen. Nach Untersuchung vieler Exemplare von Pro-/Contra-Kommentaren hat sich das in den Richtlinien beschriebene Inventar von Inhaltszonen als geeignet erwiesen, die Texte jeweils vollständig mit einer Zonen-Zuweisung zu beschreiben. Diese Zonen können dann auch als Ausgangspunkt für eine genauere Analyse der Argumentationsstruktur (Kapitel 11) dienen (die Argumentationsrichtlinien sind jedoch von den Zonen unabhängig gehalten).

Voraussetzung für die Zonen-Analyse ist eine Diskurssegmentierung nach Kapitel 3, um die Rollen der Segmente auf verschiedenen Ebenen (etwa RST, Illokutionen, Zonen, Argumentation) vergleichen zu können.

2.8 Rhetorische Struktur

Die Richtlinien lehnen sich eng an die Darstellung der *Rhetorical Structure Theory* von Mann u. Thompson (1988) und Taboada u.

Mann (2006), sowie an die der RST Webseite² an. Sie weichen aber bei den Relationsdefinitionen in einigen Punkten von dieser „klassischen“ RST (im Folgenden: kRST) ab und geben (über die Beschreibung der Theorie hinaus) für die Annotierenden eine Vorgehensanweisung. Diese ist relativ konzise auf unsere Textsorte des Kommentars zugeschnitten und damit hoffentlich leichter umzusetzen als die der Richtlinien zur *RST Discourse Treebank* von Carlson u. Marcu (2001). Deren erheblich ausgeweitetes Relationsinventar und die einzelnen Definitionen weisen recht viele Unterschiede zu den hier vorgestellten Richtlinien auf.

Zu Beginn von Kapitel 10 findet sich eine knappe Zusammenfassung des Anliegens von RST, idealerweise sollten die Annotierenden zur Vorbereitung aber bereits die Arbeit von Mann u. Thompson (1988) oder die von Taboada u. Mann (2006) gelesen haben.

Die Abweichungen unserer Relationsdefinitionen von kRST betrifft zum einen die Einteilung in Gruppen. In kRST findet sich eine Dreiteilung in *subject-matter*, *presentational*, *multinuclear* Relationen. Die ersten beiden nennen wir hier (wie Sanders u. a. (1992)) *semantisch* und *pragmatisch*, und wir ordnen manche Relationen anders zu (etwa: Evaluation – der Autor bewertet einen Sachverhalt – sehen wir als pragmatische Relation, nicht als subject-matter/semantisch). Neben der Elaboration verwenden wir E-Elaboration, um zu unterscheiden, ob der komplette Sachverhalt des Nukleus, oder nur eine einzelne Entität näher beschrieben wird. Als eher textsortenspezifische Ergänzung benutzen wir eine gegenüber kRST neue Relation Reason, um Begründungen zu markieren, bei denen der Grund keine nachprüfbare Evidence, sondern wiederum eine subjektive Behauptung oder These darstellt. (Dies grenzen wir dann von Justify ab, indem diese Relation auf Rechtfertigungen im engeren Sinne beschränkt wird.) Umgekehrt gibt es kRST Relationen, die wir nicht übernommen haben, wie Disjunction; dies subsumieren wir unter die Conjunction.

²<http://www.sfu.ca/rst> (Zugriff: 30.9.2015)

Ein wesentlicher Unterschied ist die Einführung einer neuen Gruppe von Relationen: Als Textuelle Relationen bezeichnen wir diejenigen, bei denen weder beschriebene Sachverhalte noch Sprecherintentionen im Zentrum stehen, sondern Darstellungsentscheidungen: Preparation, Restatement und Summary. Diese Entscheidung ist von der Gruppe der *textual relations* bei Martin (1992) inspiriert.

Die grundsätzliche Rolle für eine RST-Analyse, wie auch in den Richtlinien kurz ausgeführt, sehen wir hier als die *Rekonstruktion der komplexen Autor-Intention* durch den Leser bzw. die Leserin. Dies setzt naturgemäß die Interpretation des Textes voraus, was mit einem gewissen Maß an subjektiver Einschätzung einhergeht. Wenn zwei Leser/innen desselben Texts zu unterschiedlichen RST-Analysen gelangen, bedeutet dies nicht automatisch eine Schwäche der Theorie – vielmehr kann es sich um unterschiedliche, möglicherweise gleichermaßen berechnigte, Rekonstruktionen handeln. Entscheidend ist dabei natürlich die Vermeidung von Belieblichkeit: Unterschiedliche Analysen, die auf den berechtigten Interpretationsspielraum zurückgehen, müssen unterschieden werden von Ambiguität, die aus unzureichenden Relationsdefinitionen oder unklaren Anweisungen resultiert. Eine sinnvolle Evaluation der Richtlinien im Sinne des *inter-annotator agreement* ist für eine Analyse-Ebene wie RST keineswegs trivial, und die Entwicklung aussagekräftiger Methoden stellt für die weitere Theoriebildung eine hochinteressante Aufgabe dar; einen aktuellen Versuch unternehmen Iruskieta u. a. (2014).

2.9 Argumentationsstruktur

Eine der bekanntesten jüngeren Arbeiten zur Argumentationstheorie ist das Schema von Toulmin (1958), das den grundsätzlichen Aufbau eines Arguments, bzw. dessen maximale Ausprägung, beschreibt; siehe Abbildung 2.1. Danach hat ein Argument sechs Komponenten: Aus einer Beobachtung (*Data*) wird ein Schluss (*Conclusion*) gezogen, denn es gibt eine allgemeingültige Regel

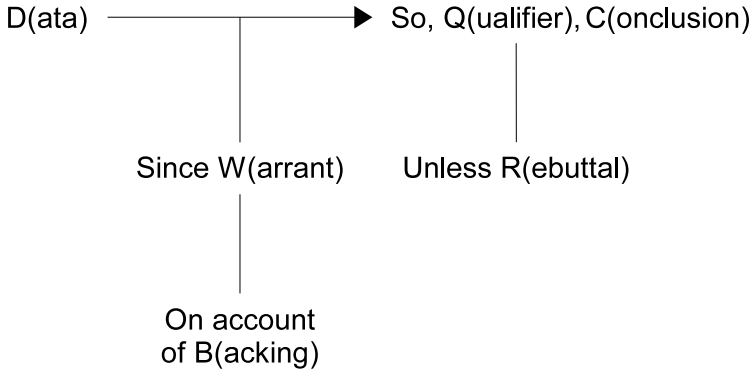


Abbildung 2.1: Argumentationsschema nach Toulmin (1958)

(*Warrant*), die man diesem Schluss zugrunde legen kann. Die Regel selbst kann durch eine Autorität (*Backing*) abgesichert werden. Möglicherweise ist der Schluss allerdings noch durch einen *Qualifier* moduliert: Er kann nur *vielleicht*, *wahrscheinlich* o.ä. gelten. Und schließlich kann auch ein mögliches Gegenargument (*Rebuttal*) Teil des Schemas sein – ein Gegenargument, das den Schluss aus der Beobachtung ungültig werden lässt oder zumindest angreift.

Ein Beispiel: Aus den aktuellen Arbeitsmarkt- und Konjunkturdaten (Data) ziehen wir den Schluss, dass sich die Wirtschaft in einem Abschwung befindet (Conclusion). Dem liege die (hier fiktive) Regel zugrunde, dass Beschäftigungsrückgang und fallende Umsatzerwartungen der Unternehmen auf einen Abschwung hindeuten (Warrant). Wir halten uns an diese Regel, weil sie von einer ganzen Reihe namhafter Wirtschaftsexperten immer wieder formuliert wird (Backing). Den Schluss ziehen wir, so die Behauptung, „mit großer Sicherheit“ (Qualifier). Er könnte nur dadurch gefährdet sein, dass die Konjunkturdaten allein bestimmten Artefakten der Weltwirtschaft geschuldet sind und derzeit nicht wirklich die Lage in unserem Land widerspiegeln (Rebuttal).

Das Toulmin-Schema beschreibt die Gestalt eines Arguments „an sich“, ist aber nicht unmittelbar für die Analyse argumentativer Texte verwendbar. Zu diesem Zweck benötigen wir ein *kompositionales* Schema, das es erlaubt, die einzelnen im Text angetroffenen Schluss-Züge zu einer Gesamtdarstellung des Arguments zusammenzufassen. Die „Züge“ bilden die Bausteine der Analyse: Ein Faktum stützt eine These; zwei Beobachtungen gemeinsam stützen eine These; ein mögliches Gegenargument attackiert einen Schluss, wird aber selbst auch widerlegt, usw. Ein Inventar solcher Züge wurde von Freeman (1991) vorgeschlagen, der Argumentation als einen inhärent dialogischen Prozess definiert, bei dem der Argumentierende die möglichen Einwände eines (virtuellen) Opponenten antizipiert und entkräftet. Das in Kapitel 11 erläuterte Schema lehnt sich eng an Freemans Arbeit an, wurde aber von uns zu einer dezidierten Annotationsprozedur für Pro-/Contra-Kommentare weiterentwickelt. Eine ausführlichere Darstellung der Motivation für die einzelnen Elemente findet sich in (Peldszus u. Stede, 2013).

Diskurssegmentierung

Manfred Stede, Sara Mamprin, Andreas Peldszus

Hinweise:

- Alle in diesem Kapitel genannten Textbeispiele ohne Quellenangabe sind fiktiv.
- Die vorliegenden Richtlinien verwenden eine Notation, wie sie „auf Papier“ oder im Text-Editor verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.

3.1 Hintergrund

Ein großer Teil diskursbezogener Analyse und Annotation bezieht sich auf zugrunde liegende „minimale Einheiten“, also auf eine Gliederung des Texts in *Diskurssegmente*. Im einfachen Fall hat man es mit einer Sequenz von Hauptsätzen zu tun, die trivialerweise dann auch die Segmente (gekennzeichnet durch eckige Klammern) bilden:

(3.1) *[Letzte Woche sind wir nach Chemnitz gefahren.] [Dort haben wir das Museum Gunzenhauser besucht.] [Es hat uns sehr gut gefallen!]*

Es bleibt allerdings nicht so einfach. In Texten begegnen wir im Hinblick auf Segmentierung zwei zentralen Komplikationen, die wir in den nachfolgenden Abschnitten besprechen werden.

- **Zusammengesetzte Sätze:** Wenn mehrere Hauptsätze nebeneinander verbunden sind, oder wenn zum Hauptsatz noch Nebensätze hinzutreten, muss entschieden werden, ob diese jeweils eigene Diskurssegmente bilden.

- Fragmentarische Einheiten: Fragmente unterschiedlicher Komplexität (ein einzelnes Wort, eine Phrase, etc.) können im Text die Position von Sätzen einnehmen, oder auch als Parenthesen bzw. Appositionen in Sätze eingefügt sein. Ihre Behandlung muss für die unterschiedlichen Fälle ihres Auftretens geklärt werden.

Wegen dieser Komplikationen teilen wir den Annotationsprozess in eine Sequenz aufeinander folgender Schritte der Identifikation von Sinneinheiten und der anschließenden feineren Einteilung in strukturelle Einheiten. Diesen Einheiten wird dann jeweils ein *Typ* zugewiesen. Anhand der vorgenommenen Typisierung können die in anderen Kapiteln behandelten Annotationsebenen ihren eigenen Begriff von „minimalen Analyseeinheiten“ definieren, indem sie ggf. Segmente, die qua ihres Typs für die Analyse-Ebene zu klein sind, an benachbarte größere Segmente anschließen.

3.2 Annotationsvorgehen: Übersicht

Die Annotation verläuft in drei Schritten, die jeweils den Text komplett behandeln:

1. Im ersten Schritt werden satzbeendende Interpunktionszeichen daraufhin geprüft, ob sie eine vollständige Sinneinheit abschließen, oder nur eine kaum interpretierbare, unvollständige Einheit. Es werden diejenigen Zeichen markiert, die dann insgesamt den Text in eine Folge von in sich geschlossenen Sinneinheiten gliedern.
2. Im zweiten Schritt wird jede einzelne Sinneinheit – soweit nötig – ihrerseits in eine Folge von strukturellen Einheiten gegliedert, die entweder Hauptsätze oder fragmentarische Einheiten darstellen.
3. Im dritten Schritt werden alle Hauptsätze daraufhin geprüft, ob sie durch Unterordnung (Nebensätze), Nebenordnung, oder Einbettung parenthetischer Einheiten noch einmal zu untergliedern sind.

Die in Schritt 1 nach semantisch/pragmatischen Kriterien markierten Sinneinheiten werden also in Schritt 2 und 3 ggf. weiter aufgegliedert.

Am Ende ist der Text vollständig in eine Sequenz von Sinneinheiten aufgeteilt. Jede Sinneinheit ist ihrerseits vollständig zerlegt in strukturelle Einheiten: Hauptsätze (HS) oder Fragmente (FR). Jedes Token des Textes muss also einem HS oder FR zugeordnet sein. In HS können andere Einheiten rekursiv eingebettet sein (Hauptsätze, Nebensätze, Parenthesen). Grundsätzlich unzulässig sind *Überschneidungen* zwischen Segmenten: Es darf nicht sein, dass der Anfang eines Segments A innerhalb von Segment B liegt, das Ende von A aber außerhalb von B. Mit anderen Worten, unsere Notation der eckigen Klammerung ist stets (durch Zählen der Klammern) eindeutig interpretierbar.

Im Folgenden besprechen wir die drei Schritte im Detail. Danach gibt Abschnitt 3.6 eine Übersicht über alle besprochenen Typ-Bezeichner und Abschnitt 3.7 diskutiert abschließend noch einige Annotationsbeispiele.

3.3 Schritt 1: Teilung in Sinneinheiten

„Satzartige“ Einheiten enden mit satzbeendenden Interpunktionszeichen. Aber nicht jedes Auftreten eines solchen Zeichens beendet auch eine vollständige „Sinneinheit“: Manche Einheiten sind oftmals in sich nicht interpretierbar, sondern nur im Verbund mit davor oder dahinter stehenden Einheiten. Darum ist ein Interpretationsschritt erforderlich, der diejenigen Interpunktionszeichen identifiziert, die tatsächlich eine vollständige Sinneinheit abschließen.

Identifikation einer Sinneinheit: Betrachte die Einheit zwischen zwei satzbeendenden Interpunktionszeichen (. / : ! / ?). Enthält sie einen Konnektor (Konjunktion, verbindendes Satzadverb), ignoriere diesen. Enthält sie anaphorische Pronomen oder Elipsen, ersetze diese durch ihr Antezedens. Ist die Einheit im Skopus einer zuvor eröffneten Modalität (wie in/direkte Rede, hypothetische Situation), füge diese Modalität der Einheit hinzu.

Prüfe nun, ob eine vollständige, in sich abgeschlossene Information vorliegt.

- Wenn ja, prüfe ob das nachfolgende Material eine neue Sinneinheit beginnt, oder aber der vorliegenden Sinneinheit angefügt werden muss (weil es sich um unselbstständige, ergänzende Information handelt). Füge solche nachfolgenden unselbstständigen Einheiten der aktuellen Sinneinheit hinzu.
- Wenn nein, prüfe in entsprechender Weise, ob durch Anfügung der nachfolgenden Einheit (das Material bis zum nächsten satzbeendenden Interpunktionszeichen) eine vollständige Sinneinheit entsteht; wenn nötig, setze diesen Prozess mit nachfolgenden Einheiten fort.

Wenn eine nach obiger Anweisung gefundene Sinneinheit zwei (oder mehr) durch Semikolon oder Komma getrennte Hauptsätze (oder nahezu vollständige Hauptsätze) enthält, dann prüfe ob an diesen Stellen auch eine Sinneinheitengrenze vorliegt. Wenn ja, markiere sie.

In diesen Richtlinien markieren wir solche Grenzen mit XX. Am Textanfang und am Textende wird kein XX gesetzt. Es folgen einige Beispiele:

(3.2) *Bube, Dame, König, As: Das sind die beliebten Spielkarten.*

Die Aufzählung vor dem Doppelpunkt wird erst durch den nachfolgenden Satz in eine Aussage eingebettet, daher liegt am Doppelpunkt keine Grenze vor.

(3.3) *Das ist das wichtigste: Man muss immer die Augen offenhalten.*

Hier steht vor dem Doppelpunkt zwar ein syntaktisch vollständiger Satz, aber „das“ ist kataphorisch, der Sachverhalt erschließt sich nur durch Vorgriff auf den nachfolgenden Satz. Daher liegt am Doppelpunkt keine Grenze vor.

(3.4) *Meier sagt immer, man soll die Augen offenhalten. XX Und gerade bei schlechtem Wetter immer vorsichtig fahren. XX Damit hat er recht!*

Der erste Satz ist eine vollständige, interpretierbare Einheit. Der zweite Satz ist eliptisch, wir ergänzen ihn gemäß obiger Anweisung mit *man soll*. Außerdem ergänzen wir die indirekte Rede, erhalten dann *Und Meier sagt, man soll gerade bei schlechtem Wetter immer vorsichtig fahren* – dies ist eine vollständige Einheit, sodass sie vom ersten Satz getrennt wird. Im dritten Satz verschiebt sich die Perspektive wieder zum Autor, die Anaphern *damit* und *er* füllen wir mit ihren Antezedenten, und wieder resultiert eine eigenständige Sinneinheit.

(3.5) *Gestern hat mein Freund Hans mich wieder besucht; XX er war ja immer schon eine treue Seele.*

Anstelle des Semikolons könnte auch ein Punkt stehen – das Thema wechselt, und wir bekommen eine neue, separate Information. Darum wird hier eine Abtrennung vorgenommen.

(3.6) *Der Kreisel ist Asbest verseucht. Nicht nur hier und da, sondern durch und durch. XX Zwar könnte man, wie beim Palast der Republik, den Bau bis aufs wacklige Stahlskelett entkleiden und neu aufbauen. XX Aber das würde mindestens 84 Millionen Euro, vielleicht auch das Doppelte kosten. (procon-kreiselpro)*

Der erste Punkt (nach *verseucht*) beendet einen Hauptsatz, der auch inhaltlich vollständig ist. Die nachfolgende Einheit jedoch ist nur in Verbindung mit diesem Satz verständlich, so dass der erste Punkt keine Sinneinheit beendet. Mit *zwar* beginnt ein neuer Gedanke, so dass davor die erste Sinneinheit-Grenze gesetzt wird. Die beiden Konnektoren „zwar..aber“ sind semantisch sehr eng

verbunden (ein „zwar“-Satz muss fast immer von einem „aber“ oder „doch“ Satz gefolgt werden). Unsere Definition besagt jedoch, dass wir Konnektoren ignorieren, daher ist zu prüfen, ob *Man könnte, wie beim Palast ... und neu aufbauen* einerseits und *Das würde mindestens ... kosten* andererseits jeweils eine vollständige Einheit bilden; mit Einsetzung des Antezedens (etwa: *der Umbau*) für die Anapher *das* ist das der Fall, so dass auch vor *Aber* eine Sinneinheit endet.

(3.7) *Die Topographie des Terrors, eben ausgegraben und freigelegt, stand für sich. XX (...) XX Dort hingen Fotos, XX dort gab es einen Katalog XX – mehr als genug, um wenigstens zu ahnen, welcher Terror von diesem Ort ausgegangen ist.* (procon-topographiecon)

Nach dem Bindestrich gibt der Autor eine Bewertung, die sich offenbar auf *Fotos* und *Katalog* bezieht, so dass sich die Frage stellt, ob alles zu einer umfangreichen Sinneinheit zusammengefasst werden sollte; wir behandeln solche Fälle aber analog zu den oben besprochenen von indirekter Rede oder anderen satzübergreifenden Modalitäten (die keine Zusammenfassung zu einer Gesamteinheit auslösen) und trennen die einzelnen Aussagen daher in Sinneinheiten auf.

3.4 Schritt 2: Unterteilung der Sinneinheiten in Segmente

In diesem Schritt werden jeder Sinneinheit ein oder mehrere strukturelle Segmente zugeordnet, die jeweils durch [] abzugrenzen und mit einem *Typ* zu etikettieren sind (in diesen Richtlinien durch Tieferstellung markiert). Im einfachsten Fall bestehen die aus Schritt 1 resultierenden Einheiten jeweils aus einem einzelnen Hauptsatz. Das ist im eingangs besprochenen Beispiel 3.1 gegeben: Jeder HS stellt auch eine eigene Sinneinheit dar. An den Satzgrenzen ist also jeweils noch ein XX zu ergänzen, zudem ist jede schließende Klammer mit einem _{HS} zu versehen. Das komplett annotierte Beispiel ist also:

(3.8) *[Letzte Woche sind wir nach Chemnitz gefahren.]_{HS} XX [Dort haben wir das Museum Gunzenhauser besucht.]_{HS} XX [Es hat uns sehr gut gefallen!]_{HS}*

Oftmals wird eine Sinneinheit hingegen komplex sein, weil sie mehrere Hauptsätze oder auch Fragmente enthält. Das entscheidende Merkmal für die Unterteilung sind wieder die satzbeendenden Interpunktionszeichen. Jede von einem solchen Zeichen gebildete Einheit muss auf ihren Hauptsatz-Status geprüft werden.

Identifikation eines Hauptsatzes: Betrachte das Segment zwischen zwei satzbeendenden Interpunktionszeichen (. / : / ! / ?), bzw. das Segment, das durch eine mit Semikolon oder Komma markierte Sinneinheitengrenze entstanden ist.

Weise ihm den Typ HS (Hauptsatz) zu, wenn es sich um einen strukturell vollständigen Satz (Aussagesatz, Fragesatz, Aufforderungssatz) handelt.

Besteht eine Sinneinheit nach dieser Erkennung allein aus (einem oder mehreren) HS, ist Schritt 2 für sie beendet.

Anderenfalls ist fortzufahren:

- Befindet sich in der Sinneinheit zumindest ein HS, sind die übrigen als fragmentarische Hauptsätze (HSF), einleitende Fragmente (FRE) oder beendende Fragmente (FRB) zu klassifizieren; siehe unten.
- Befindet sich in der Sinneinheit kein HS, ist darauf zu achten, dass zumindest ein fragmentarischer Hauptsatz zugewiesen wird, nämlich demjenigen Segment, das am ehesten als vollständige Informationseinheit bzw. Illokution interpretiert werden kann. Sind weitere Segmente vorhanden, sind sie als HSF, FRE oder FRB zu klassifizieren; siehe unten.

Identifikation eines Einleitenden Fragments: Betrachte das Segment zwischen zwei satzbeendenden Interpunktionszeichen (. / : / ! / ?).

Weise ihm den Typ FRE (einleitendes Fragment) zu, wenn es keine vollständige Information übermittelt und keinen eigenen Sprechakt darstellt, sondern die Funktion hat, das nachfolgende Segment vorzubereiten. Ein FRE ist nur in Verbindung mit dem nachfolgenden Segment (in Ausnahmefällen: mit den nachfolgenden Segmenten) interpretierbar.

Die „prototypischen“ FRE sind recht kurz, doch die letzten beiden Beispiele illustrieren, dass es auch relativ komplexe Einheiten (auch Nebensatzstrukturen, die nicht in einen HS eingebettet sind) sein können:

(3.9) [*Und dann:*]FRE [*Ist der Preis nicht sowieso unbezahlbar?*]HS

(3.10) [*Klar ist:*]FRE [*Die Situation wird sich nicht verbessern.*]HS

(3.11) [*Werner Meier.*]FRE [*Das ist der neue Star im deutschen Fußball.*]HS

(3.12) [*Vielen Ostdeutschen wurde wendebedingt jedoch eingeredet:*]FRE [*Ihr habt alles falsch gemacht!*]HS (maz-9852)

(3.13) [*Und selbst wenn schon 2012 eine europäische Metropole Ausrichter werden sollte:*]FRE [*Man muss die Muskeln früh spielen lassen, um...*]HS (procon-olympiapro)

Identifikation eines beendenden Fragments: Betrachte das Segment zwischen zwei satzbeendenden Interpunktionszeichen (. / : / ! / ?).

Weise ihm den Typ FRB (beendendes Fragment) zu, wenn es nicht satzwertig ist und keine vollständige Information übermittelt und keinen eigenen Sprechakt darstellt; sondern die Funktion hat, das vorangehende Segment zu ergänzen, darin eine Leerstelle zu füllen.

Beispiele für FRB:

(3.14) [*Wer wäre wohl bereit zu zahlen:*]HS [*die Lehrer, die Eltern, die Opernfreunde?*]FRB (procon-olympiacon)

(3.15) [*Dieensee zieht um.*]HS [*Ohne Wenn und Aber.*]FRB (maz-6993)

Relativ häufig treten knappe Fragmente auf, mit denen der Autor das vorangehende Segment kommentiert oder bewertet:

(3.16) [*Und dann kam der Rektor auf mich zu.*]HS [*Na prima.*]HSF

(3.17) [*Meier hielt es für einen schlechten Scherz.*]HS [*Kann man wohl sagen.*]HSF

Diese fassen wir *nicht* als Fortsetzung des vorangehenden Segments auf, denn es handelt sich um eigenständige Sprechakte; sie sind der nachfolgend beschriebenen Kategorie HSF zuzuordnen. Beachte den Unterschied zu 3.15: Auch dies könnte man als „Kommentierung“ verstehen, doch wird hier eine genauere Charakterisierung des im HS genannten Ereignisses vorgenommen – es findet keine Perspektivenwechsel zwischen den beiden Segmenten statt.

Identifikation eines Hauptsatzfragments: Betrachte das Segment zwischen zwei satzbeendenden Interpunktionszeichen (. / : / ! / ?) bzw. das Segment, das durch eine mit Semikolon oder Komma markierte Sinneinheitengrenze entstanden ist.

Weise ihm den Typ HSF (Hauptsatzfragment) zu, wenn es zwar kein strukturell vollständiger Satz ist, aber

- durch Rückgriff auf den vorangehenden Kontext zu einem solchen ergänzt werden kann (Auffüllen von Elipsen),
- oder einen eigenständigen Sprechakt darstellt.

Alle zuvor nicht als FRE oder FRB klassifizierten Segmente sollten nun der Kategorie HSF zugehörig sein. Bei der Interpretation

kann der Rückgriff auf den vorangegangenen Kontext auch über Sinneinheit-Grenzen hinweg erfolgen, wie die ersten beiden der folgenden Beispiele zeigen. Zudem illustrieren einige der Beispiele, dass HSF nicht zwingend ein Verb enthalten (satzwertig sein) müssen: Semantisch „leere“ Verbkomplexe wie *es ist* oder *dies gilt* dürfen bei der Interpretation ergänzt werden.

(3.18) [*Klaus fährt nach Berlin.*]_{HS} XX [*Kauft sich ein Motorrad.*]_{HSF}

(3.19) [*Der Kreisel ist nicht schön.*]_{HS} XX [*Ein Symbol der West-Berliner Filzwirtschaft in den späten Sechziger Jahren.*]_{HSF} (procon-kreiselpro)

(3.20) [*Der Bürgermeister verlas den Antrag.*]_{HS} XX [*Kichern und Prusten auf dem Fraktionsgestühl.*]_{HSF}

(3.21) [*Die Kahlschlagmentalität breitet sich aus.*]_{HS} [*In den Mietskasernen-Bezirken zum Beispiel.*]_{HSF}. (procon-bäumecon)

3.5 Schritt 3: Unterteilung zusammengesetzter Sätze

Im letzten Schritt sind alle bisher gefundenen (durch satzbeendende Interpunktionszeichen markierten) Segmente darauf hin zu prüfen, ob sie selbst weiter untergliedert werden müssen (wodurch eine rekursive Einbettung entsteht). Im Wesentlichen handelt es sich dabei um die Behandlung zusammengesetzter Sätze.

Prinzipiell gilt in diesem Schritt, dass wir für aneinandergereihte Segmente keine (semantisch oder anders motivierte) Hierarchie etablieren wollen, sondern die Strukturen so „flach“ wie möglich halten. Schematisch bedeutet das:

- NICHT: [[[Erst geschah X,] [dann geschah Y,]] [aber ich habe P,] [und dann habe ich Q]]]
- SONDERN: [[Erst geschah X,] [dann geschah Y,] [aber ich habe P,] [und dann habe ich Q]]]

Den ersten Fall der zu besprechenden Untergliederung stellt die strukturelle *Unterordnung* (Subordination) dar, wie sie prototypisch bei Hauptsatz/Nebensatz-Strukturen vorliegt. Ein Beispiel zur Illustration:

(3.22) [*Letzte Woche sind wir nach Chemnitz gefahren, [was im Süden von Sachsen liegt.]*_{ANR}]_{HS} XX [*Dort haben wir [(das war der Anlass der Reise)]*_{HS} *das Museum Gunzenhauser besucht.*]_{HS} XX [*Wie nicht anders zu erwarten:*]_{FRE} [*Es hat uns sehr gut gefallen!*]_{HS}

Der zweite Fall der Unterteilung zusammengesetzter Sätze betrifft die *Nebenordnung* (oft durch Konjunktionen) von Hauptsätzen oder Nebensätzen. Es wird sowohl der (in Schritt 2 behandelte) gesamte HS annotiert, als auch die eingebetteten:

(3.23) [[*Meier öffnete die Flasche,*]_{HS} [*und er trank sie in einem Zug leer.*]_{HS}]_{HS}

Der dritte Fall sind (ähnlich zu Schritt 2) Fragmente, die in Haupt- oder Nebensätzen eingebettet sein können; das betrifft vor allem Parenthesen und ähnliche Einschübe. Oben ist im Beispiel 3.22 bereits eine solche Parenthese annotiert.

Nachfolgend besprechen wir diese drei Fälle der Reihe nach. Zuvor noch der Hinweis, dass syntaktisch integrierte wörtliche Rede *nicht* als separates Segment zu behandeln ist, selbst wenn sie die Form eines Haupt- oder Nebensatzes aufweist:

(3.24) [*Meier verabschiedete sich mit einem fröhlichen ‚Es hat viel Spaß gemacht!‘ von den Kollegen.*]_{HS}

Der zitierte Satz ist weder ein Fall von Neben- noch von Unterordnung, und auch kein parenthetischer Einschub; vielmehr spielt er für den umgebenden HS die Rolle einer Nominalphrase und wird nicht als eigenes Segment aufgefasst.

3.5.1 Unterordnung: Nebensätze

In diesem Abschnitt behandeln wir Segment-Einbettungen, die durch verschiedene Arten von *Nebensätzen* entstehen, d. h. durch

Teilsätze, die „hinsichtlich Wortstellung, Tempus- und Moduswahl (...) abhängig vom übergeordneten Hauptsatz“ (Bußmann, 2002, p. 460) sind. Wir annotieren stets den *Gesamtsatz* als HS, und etwaige Nebensätze sind darin eingebettet. Also:

- NICHT: [Weil die Sonne scheint,] [legen wir uns an den Strand.]
- SONDERN: [[Weil die Sonne scheint,] legen wir uns an den Strand.]

Die nebensatzbeendenden Interpunktionszeichen sind dem Nebensatz-Segment zugeordnet, denn es gilt (wie implizit schon bei den Hauptsätzen angenommen) die Grundregel, dass Interpunktionszeichen nicht von einem unmittelbar adjazenten Wort abgetrennt werden sollen, d. h., es darf keine Grenze zwischen *scheint* und dem Komma liegen. Analog bilden wir keine Grenze zwischen Wort und dem satzbeendenden Punkt (oder anderem Interpunktionszeichen) – alle schließenden Klammern folgen also dem Satzende-Zeichen:

- NICHT: [Wir legen uns an den Strand, [weil die Sonne scheint].]
- SONDERN: [Wir legen uns an den Strand, [weil die Sonne scheint.]]

Für unser Typ-Inventar der Nebensätze übernehmen wir in diesem Abschnitt weitgehend die Klassifikation von Bußmann (2002). Neben einigen fiktiven Beispielen und denen aus PCC verwenden wir im Folgenden auch einige von canoo.net.³

3.5.1.1 Erweiterung von Satzgliedern

Wir unterscheiden Subjektsatz, Objektsatz, Prädikativsatz und Adverbialsatz, die jeweils in verschiedenen Formen auftreten können. Die ersten drei ersetzen ein Satzglied bzw. ergänzen es zu

³<http://www.canoo.net/services/OnlineGrammar/Satz/Komplex/Funktion/Uebersicht.html> (Zugriff: 30.9.2015)

einem satzwertigen Element; es handelt sich um Komplemente (also grammatisch notwendige Einheiten). Demgegenüber stellt der Adverbialsatz ein Adjunkt dar, das in einer semantischen oder pragmatischen Relation zum gesamten übergeordneten Satz steht.

Subjektsatz (Kürzel: SUB)

- dass-Satz: *[[Dass ihr hier seid,]_{SUB} freut uns sehr.]_{HS} (canoo)*
- ob-Satz: *[[Ob ich den Schlüssel wiederfinde,]_{SUB} ist fraglich.]_{HS} (canoo)*
- w-Satz: *[[Wann wir ankommen werden,]_{SUB} ist ungewiss.]_{HS} (canoo)*
- Infinitivkonstruktion, auch mit resumptivem Pronomen:
[[Ihr zu verzeihen,]_{SUB} fiel ihm schwer.]_{HS} (canoo)
[[Dass ihr hier seid,]_{SUB} das freut mich sehr.]_{HS}
- Uneingeleitet: *[Es ist besser, [du kommst noch einmal zurück.]_{SUB}]_{HS} (canoo)*

Bei den ersten vier Formen kann der Subjektsatz auch hinten stehen, wenn der Hauptsatz durch ein *es*-Korrelat eingeleitet wird. Beispiel: *[Es ist ungewiss, [wann wir ankommen werden]_{SUB}.]_{HS}*

Objektsatz (Kürzel: OBJ)

- dass-Satz: *[Wir verstehen, [dass ihr uns nicht begleiten könnt.]_{OBJ}]_{HS} (canoo)*
- ob-Satz: *[Sie wissen nicht, [ob sie auch eingeladen sind.]_{OBJ}]_{HS} (canoo)*
- w-Satz: *[Ich wundere mich, [warum du erst so spät kommst.]_{OBJ}]_{HS} (canoo)*
- Infinitivkonstruktion: *[Wir bedauern, [sie enttäuschen zu müssen.]_{OBJ}]_{HS} (canoo)*

- Uneingeleitet: [Er sagt, [er habe keine Zeit.]]_{OBJ}]_{HS} (canoo)
- wie-Satz: [Sie fühlte, [wie ihre Kräfte nachließen.]]_{OBJ}]_{HS} (canoo)
- Akkusativ mit Infinitiv: [Er hörte [ihn die Treppe hinaufkommen.]]_{OBJ}]_{HS} (canoo)

Auch beim Objektsatz sind oftmals verschiedene Abfolgen der Konstituenten möglich. Entscheidend für die Annotation ist die Identifikation der grammatischen Rolle (hier: Objekt).

Prädikativsatz (Kürzel: PRD)

Das typische Schema des Prädikativsatzes ist ‚Subjekt + Kopulaverb + Konjunktion‘.

[Diese Wohnung ist, [was ich schon immer suchte.]]_{PRD}]_{HS} (canoo)

[Die Frage ist, [ob die Sonne scheinen wird.]]_{PRD}]_{HS}

[Was mir dann klar wurde, war, [dass die Sonne scheinen würde.]]_{PRD}]_{HS}

Adverbialsatz (Kürzel: ADV)

Für alle nachfolgend genannten Unterkategorien gilt, dass stets ein bestimmter semantischer oder pragmatischer Zusammenhang zwischen zwei Sachverhalten/Propositionen ausgedrückt wird (einer im Nebensatz, einer im Hauptsatz). In der Annotation wird die Art des Zusammenhangs aber nicht vermerkt; wir verwenden nur den allgemeinen Typ ADV.

- Satz mit adverbialer Konjunktion wie *während, indem, weil, wenn, obwohl, sodass, damit* usw.
[Er ging schwimmen, [obwohl er eine leichte Grippe hatte.]]_{ADV}]_{HS} (canoo)
- Verberstsatz (temporale, konditionale Lesart)
[[Konnte der Kosovo-Krieg noch mit dem Kampf für die Menschenrechte in Europa begründet werden,]_{ADV} geht es nun um die militärische Solidarität mit den USA in einem weit entfernten Dritte-Welt-Land.]]_{HS} (maz)
[[Wäre ich Millionär,]_{ADV} würde ich sofort Dein Taschengeld erhöhen.]]_{HS}

- Erweiterter-Infinitiv-Konstruktion
[Sie beeilte sich, [um den Zug zu erreichen.]_{ADV}]_{HS}
[Sie beeilte sich, [ohne spürbar in Aufregung zu geraten.]_{ADV}]_{HS}
[Sie beeilte sich [anstatt den Tag geruhsam ausklingen zu lassen.]_{ADV}]_{HS}
 (Zu unterscheiden von modal eingebetteten Infinitiven, die nicht abgetrennt werden: *[Sie wünschte sich den Zug zu erreichen.]_{HS}*)
- Partizipialkonstruktion (den HS-Sachverhalt ergänzend)
[[In Paris angekommen,]_{ADV} suchten wir ein Hotel.]_{HS} (canoo)

3.5.1.2 Attributsatz

Der Attributsatz ersetzt nicht ein Satzglied, sondern er modifiziert ein overt gegebenes Satzglied, auf das er also explizit Bezug nimmt.

- Restriktiver Relativsatz (Kürzel ARR): Der Relativsatz beschreibt die Kopf-NP näher, damit sie identifizierbar wird
[Das Haus, [das dort drüben steht,]_{ARR} gehört meinem Onkel.]_{HS} (canoo)
- Nicht-restriktiver Relativsatz (Kürzel ANR): Der Relativsatz übermittelt weitere Information zur Kopf-NP, die aber nicht primär die Identifikation erlauben soll
[Wohin mit den vom Urlaub übrig gebliebenen Münzen, [die bald sowieso nichts mehr wert sind?]_{ANR}]_{HS} (maz-8838)
- Partizipialkonstruktion (Kürzel APK): ein einzelnes Satzglied modifizierend; im Gegensatz zur o. g. Adverbialsatz-Partizipialkonstruktion
[Spät werden jetzt vereinzelt Proteste laut, [angeregt und unterstützt allerdings auch nur von Einzelpersonen.]_{APK}]_{HS} (maz-9612)
- Sonstiger Attributsatz (Kürzel ATT): Es wird ein Satzglied modifiziert, aber nicht im Sinne der obigen drei Kategorien
[Die Idee, [einen Kaffee zu trinken,]_{ATT} ist prima.]_{HS}

*Es stellt sich die Frage, [ob die Preiserhöhung gerechtfertigt ist.]_{ATT}*_{HS}

3.5.1.3 Weiterführender Nebensatz (Kürzel: WEI)

Der weiterführende Nebensatz nimmt Bezug nicht auf ein einzelnes Element des vorangehenden (Teil-)Satzes, sondern auf dessen gesamte Aussage.

*[Er hat uns eingeladen, [was uns sehr freut.]_{WEI}*_{HS} (canoö)

3.5.1.4 Annotation unklarer Nebensätze (Kürzel: UNS)

Wenn ein Nebensatz vorliegt, die Typ-Bestimmung anhand der o. g. Kategorien aber nicht gelingt, ist das Kürzel UNS für „unklarer Nebensatz“ zu verwenden.

3.5.1.5 Im Zweifel: Verbalkomplexe abtrennen

Mitunter können Zweifel aufkommen, ob ein bestimmter Verbalkomplex im Sinne einer ‚clause‘ als Segment abzutrennen ist oder nicht. Etwa in diesem Fall:

(3.25) *Nein, man sollte mal schön alles so lassen wie es ist.*

In solcherlei Zweifelsfällen gelte die Präferenz für eine Abtrennung. Wenn nicht klar ist, welche Art von Nebensatz vorliegt, ist wie eben gesagt das Kürzel UNS zu verwenden.

3.5.2 Nebenordnung

Sind Hauptsätze oder andere satzwertige Einheiten durch eine nebenordnende Konjunktion (n. K.) verbunden oder durch Komma/Semikolon getrennt, so werden – zusätzlich zur übergeordneten HS/HSF-Annotation – alle als eigenständige Segmente annotiert. Als n. K. gelten prototypisch *und*, *oder* und *denn*. *Aber* kann ebenfalls als n. K. gebraucht werden. Zudem fassen wir *sondern* als n. K. auf. Die n. K. wird jeweils in das nachfolgende Segment

aufgenommen (d. h., sie leitet stets ein Segment ein, analog zur unterordnenden Konjunktion).

Beispiele für Hauptsatz(fragment)-Nebenordnung:

- (3.26) *[[Die erste Sammelbox war schneller als gedacht prall gefüllt und schwer,]_{HS} [und sie konnte bereits an den Initiator der Sammelaktion, den Internationalen Bund (IB), übergeben werden.]_{HS}]_{HS} (maz-8838)*
- (3.27) *[[Heute scheint die Sonne,]_{HS} [morgen soll es angeblich auch schön sein,]_{HS} [und übermorgen ist mir das Wetter egal.]]_{HS}]_{HS}*
- (3.28) *[[Der Mann bezahlte viele Handwerker nicht]_{HS} [und wurde voriges Jahr zu einer mehrjährigen Haftstrafe verurteilt.]]_{HSF}]_{HS} (maz-6193)*
- (3.29) *[Berlin braucht Hilfe.]_{HS} XX [[Braucht finanzielle Unterstützung]_{HSF} [- und bittet darum nicht nur,]_{HSF} [sondern klagt sie ein.]]_{HSF}]_{HS} (procon-olympiacon)*
- (3.30) *[[Bei schlechtem Wetter die Frei- und Sommerbäder zumachen]_{HSF} [und alle Hallen öffnen?]]_{HSF}]_{HSF}*

Beispiele für Nebensatz-Nebenordnung:

- (3.31) *[Leute, [die auf Kellermauern und alte Kacheln starren]_{ARR} [und ihren Gedanken nachhängen,]_{ARR} gelten als unfähig zu erfassen, [was geschehen ist.]]_{OBJ}]_{HS} (procon-topographiecon)*
- (3.32) *[Was dort fehlt, ist ein Dach unter dem freien Himmel.]_{HS} [Eine Gelegenheit, [sich auszuruhen,]_{ATT} [sich zu sammeln,]_{ATT} [in Büchern nachzulesen]_{ATT} [und das Gesehene mit Hilfe von Bild und Ton zu vertiefen.]]_{ATT}]_{HSF} (procon-topographiepro)*

Die n. K. sind oft fakultativ; auch Reihungen von HS/HSF ohne eine solche Konjunktion sind zu annotieren, wenn einzelne Teile des Gesamtsatzes die Kriterien von HS bzw. HSF erfüllen:

- (3.33) *[[Der Kreisel ist auch innen häßlich,]_{HS}, [zudem zugig]_{HSF} [und Energie verschleudernd.]]_{HSF}]_{HS} (procon-kreiselpro)*

Umgekehrt ist, wie eingangs erwähnt, die Präsenz einer n. K. kein *hinreichendes* Kriterium für zu annotierende Nebenordnung. Sie sind nur zu annotieren, wenn sie satzwertige Einheiten (Verbal-komplexe) verbinden. Durch verschiedene Grade der Elision kann es — mit oder ohne n. K. — Zweifelsfälle geben, als Faustregel gilt dann: eher keine Trennung vornehmen.

(3.34) [*Das Auto war schick und auch teuer.*]_{HS}

(3.35) [*Blauer Himmel, aber nur 17 Grad.*]_{HS}

3.5.3 Parenthetische Einschübe

Parenthetische Einschübe können durch Kommata, ein Paar von Bindestrichen oder Klammern begrenzt sein. Sie treten in vielen Formen auf, und wir annotieren hier ausschließlich *satzwertige* bzw. illokutionär-vollständige Einschübe: Nach Prüfung der Kriterien (vgl. Schritt 2) erhalten Hauptsätze und Hauptsatzfragmente wie gewohnt das Kürzel HS bzw. HSF.

Hat die Parenthese syntaktisch die Gestalt eines Nebensatzes, wählen wir ebenfalls das passende Kürzel gemäß der in Abschnitt 3.5.1 beschriebenen Klassifikation. Die Interpunktionszeichen selbst behandeln wir so, dass Klammern bzw. das Bindestrich-Paar stets dem Einschub zugeordnet werden.

Das erste der folgenden Beispiele illustriert das HS/HSF-Kriterium „Illokution“: Mit dem Einschub nimmt der Autor eine persönliche Bewertung des berichteten Sachverhalts vor, was einem eigenständigen Sprechakt gleichkommt.

(3.36) [*Gestern hat der Lehrer [– ganz schön lächerlich! –]*_{HSF} *mit blauen Briefen für verspätete Schüler gedroht.*]_{HS}

(3.37) [*Heute war, [so soll es ja auch sein,]*_{HS} *das Kind pünktlich in der Schule.*]_{HS}

(3.38) [*Heute war das Kind [(das öfters mal Probleme mit dem Aufstehen hat)]*_{ANR} *pünktlich in der Schule.*]_{HS}

Tabelle 3.1: Typen der Diskurssegmente

Hauptsatz	vollständig unvollständig		HS HSF
Nebensatz	Satzgliederweiterung	Subjektsatz	SUB
		Objektsatz	OBJ
		Adverbialsatz	ADV
Attributsatz	weiterführend	Prädikativsatz	PRD
		restr. Relativsatz	ARR
		nichtrestr. Relativs.	ANR
		Partizipialkonstr.	AKP
		sonstige	ATT
	unklar		WEI
			UNS
Fragment	einleitend		FRE
	beendend		FRB

Da wir lediglich „satzwertige“ Einschübe behandeln, werden beispielsweise nominale Appositionen, die unmittelbar eine NP modifizieren, nicht annotiert. Sie verbleiben einfach als Teil des übergeordneten Satzes:

(3.39) [*Herder, der Dichter, wirkte unter anderem auch in Weimar.*]_{HS}

Eine weitere Konsequenz der Beschränkung auf satzwertige Einschübe in diesem Schritt ist, dass es keine einleitenden oder beendenden Fragmente (FRE, FRB) gibt, die in übergeordnete (Haupt-/Neben-) Sätze eingebettet sind. Tritt also an der Peripherie eines Satzes fragmentarisches Material auf, wird es nicht gesondert annotiert.

(3.40) [*Klar, Gedichte Auswendiglernen ist langwierig.*]_{HS}

3.6 Zusammenfassung der Typ-Bezeichner

Tabelle 3.1 vermittelt einen Überblick über alle oben eingeführten Typen von Segmenten und ihre Kürzel.

3.7 Weitere Beispiele

Abschließend besprechen wir anhand einiger nicht-trivialer Beispiele (deren Form an authentische Textbeispiele angelehnt ist) den Umgang mit möglichen Zweifelsfällen.

3.7.1 Hauptsätze, Fragmente

(3.41) [*Meier stellte sich mit einem ‚Das wird schon gut gehen‘ hinter Schulze.*]_{HS}

Die zitierte Rede ist im obigen Beispiel zwar ein vollständiger Satz, aber für die übergeordnete clause wirkt sie wie ein Nominal und wird daher nicht segmentiert.

(3.42) [*Meier fiel hin.*]_{HS} [*Mit ärgerlichen Folgen:*]_{HSF} [*Das Knie war aufgeschlagen.*]_{HS}

Das mittlere Segment ist nicht nur FRE, denn es bewertet, bildet also eine eigenständige Illokution, mithin als HSF zu annotieren.

(3.43) [*In Spremberg verfahren die Leute offenbar nach der einfachen Rechnung:*]_{HS} [*Eins und Eins ist zwei.*]_{HS}

Das erste Segment kommentiert den Sachverhalt, stellt also eine eigene Illokution dar. *Die Rechnung* ist zwar kataphorisch, aber strukturell ist der nachfolgende Satz komplett, also ein HS und kein Fragment.

Hier noch zwei Beispiele für etwas ungewöhnliche HS:

(3.44) [*Es sei denn, der Kollege bekommt Muffensausen.*]_{HS}

(3.45) [*Handelte es sich doch um Bürger, [die Berlin nie mochten]*]_{ARR}]_{HS}

3.7.2 Subjekt-, Objekt- und Prädikativsätze

Subjektsätze können (z. B.) im Passiv stehen (3.46), eingebettet in Konditionalkonstruktionen stehen (3.47) oder nachgeordnete *dass*-Sätze sein (3.48):

(3.46) [*Es wird gemunkelt, [dass die Lage ernst ist]*_{SUB}]._{HS}

(3.47) [[*Gelingt es der Regierung, [die Stagnation zu überwinden]*_{SUB}]_{ADV}, *kann sie froh sein.*]_{HS}

(3.48) [*Bleibt zu hoffen, [dass sich viele Menschen anschließen]*_{SUB}]._{HS}

Objektsätze können (z. B.) als nachgestellte Konditionale erscheinen, wo aber das *wenn* im Sinne von *dass* verwendet wird (3.49), als vorangestellte indirekte Rede (3.50), oder nachgeordnet in *w*-Sätzen stehen (3.51):

(3.49) [*Man mag es Meier nicht verübeln, [wenn er seine Meinung deutlich sagt]*_{OBJ}]._{HS}

(3.50) [[*Gefährlich sei er nicht,*_{OBJ} *beeilte sich das Ministerium gestern mitzuteilen.*]_{HS}

(3.51) [*Was hat den SPD-Chef geritten, [eine solche Botschaft in die Welt zu senden]*_{HS}?]_{HS}

Prädikativsätze können (z. B.) in *so*, *X*-Konstruktionen stehen, in denen das *so* einen Vorwärtsverweis auf das Prädikat darstellt. In 3.53 ist *als* nicht temporal zu lesen, sondern als semantisch weitgehend leerer Verknüpfers, analog zu *dass*.

(3.52) [*Und natürlich ist es so, [dass man Kompromisse machen muss]*_{PRD}]._{HS}

(3.53) [*Es scheint fast so, [als wolle Meier nichts kaufen]*_{PRD}]._{HS}

3.7.3 Adverbialsätze

Hier zwei Beispiele für verkürzte Adverbialsätze; beide können durch Einsetzen von Auxiliaren vervollständigt werden (*Statt dass sie ihn dereguliert hat / Ob sie feenhaft schön oder pausbäckig barock sind*) und stehen dann in einer semantischen Relation zum Hauptsatz.

(3.54) *[[Statt dereguliert]_{ADV} hat sie den Arbeitsmarkt mit Gesetzen überfrachtet.]_{HS}*

(3.55) *[[Ob feenhaft schön oder pausbäckig barock –]_{ADV} wir mögen sie alle.]_{HS}*

3.7.4 Attributsätze

Auch wenn das *da* in 3.56 einen Teilsatz mit Hauptsatzwortstellung einleitet, ist es semantisch lediglich ein *in denen*, entspricht also einem Relativpronomen. Die Frage nicht-/restriktiv ist nicht immer leicht zu entscheiden, hier dürfte es sich aber um eine Definition *derjenigen Zeiten* handeln.

(3.56) *[Es gab Zeiten, [da verkaufte sich das sehr gut]_{ARR}.]_{HS}*

In den folgenden Beispielen besteht keine Relation zwischen zwei unabhängig interpretierbaren clauses, sondern die *Frage*, der *Mut* und die *Regel* werden genauer bestimmt:

(3.57) *[Meier beantwortete die Frage, [weshalb er so schlecht gelaunt sei,]_{ATT} nicht.]_{HS}*

(3.58) *[Es fehlte der Mut, [den Arbeitsmarkt zu regulieren]_{ATT}.]_{HS}*

(3.59) *[In Bonn hat man eine einfache Regel befolgt, [nämlich niemals Äpfel mit Birnen zu vergleichen]_{ATT}.]_{HS}*

Negation

André Herzog, David Kaupat, Manfred Stede

Hinweise:

- Beispiele ohne Quellenangabe in diesem Kapitel sind entweder fiktiv oder entstammen der Domäne ‚medizinischer Befundtext‘. (Solche Texte wurden neben dem PCC hinsichtlich ihrer Negationen annotiert.)
- Die vorliegenden Richtlinien verwenden eine Notation, wie sie „auf Papier“ oder im Text-Editor verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.

4.1 Hintergrund

Mit einer *Negation* bringt jemand zum Ausdruck, dass etwas (im sehr weiten Sinn ein ‚Sachverhalt‘) nicht zutrifft. Das zentrale Instrument dafür ist der *Negationsoperator*: eine lexikalische oder morphologische Einheit, deren Bedeutung darin besteht, Verneinung anzuzeigen. Das, was verneint wird, nennen wir den *Skopus* des Negationsoperators. Dies können ganze Sätze, Teilsätze oder kleinere Konstituenten sein. Die nachfolgenden Richtlinien erläutern anhand von Beispielen, wie in verschiedenen Fällen die Annotation von Negationsoperator und Skopus erfolgen soll. Ganz am Schluss gehen wir auf ein gelegentlich zu annotierendes drittes Element ein, sog. „Abschwächungsoperatoren“).

4.2 Grundsätzliches Vorgehen

Wir markieren Negationsoperatoren durch Unterstreichung und ihren Skopus durch eckige Klammern. Dabei wird der Negations-

operator in die Klammerung mit einbezogen, sofern er direkt neben bzw. innerhalb des Skopus steht, was in den meisten Fällen zutrifft. Interpunktionszeichen bleiben stets außerhalb des Skopus (es sei denn, der Skopus erstreckt sich über ein Interpunktionszeichen hinweg).

(4.60) *Die Untersuchung ergab [keine höhergradigen Herzrhythmusstörungen].*

Der Skopus ist in manchen Fällen nicht leicht zu identifizieren, es können Mehrdeutigkeiten vorliegen. Würde man das obige Beispiel mit geeigneter Intonation lesen und es vielleicht fortsetzen mit *Geringfügige liegen allerdings vor*, so wäre es angezeigt, lediglich *höhergradige* als Skopus zu annotieren. Bei der Textannotation haben wir naturgemäß keine Prosodie zur Verfügung, wohl aber den textuellen Kontext, der durchaus einflussreich sein kann. Die Annotation soll im Grundsatz diesen Regeln folgen:

1. Über den Skopus ist unter Berücksichtigung des gesamten Kontexts zu entscheiden.
2. Der Skopus ist primär nach semantischen/pragmatischen Gesichtspunkten zu bestimmen: Was wird ausgedrückt? Was möchte der/die Autor/in verneinen?
3. In Zweifelsfällen ist der Skopus tendenziell *eng* zu markieren.

Regel 3 folgend markiert man im obigen Beispiel nicht etwa den gesamten Satz als Skopus, auch wenn die Paraphrase *Es ist nicht so, dass die Untersuchung höhergradige Herzrhythmusstörungen ergab* hier plausibel ist. Aus demselben Grund entscheiden wir uns im folgenden (fiktiven) Beispiel für die Variante (a):

(4.61)

- (a) Klaus hat [keine Kartoffeln] gegessen.
- (b) Klaus hat [keine Kartoffeln gegessen].
- (c) [Klaus hat keine Kartoffeln gegessen].

Stets gilt aber das Prinzip der Berücksichtigung des Kontexts! Im Falle einer Fortsetzung durch *Stattdessen hat er ein Bier getrunken* wäre Variante (b) angemessener.

4.3 Negationsoperatoren und ihr Skopus

Wir nennen im Folgenden die typischen Kategorien von Negationsoperatoren, geben Beispiele an und diskutieren für jede Kategorie kurz das Auffinden des Skopus. Auch unter Berücksichtigung der Regel, Semantik und Pragmatik als primäre Entscheidungsgrundlage zu nutzen, wird er in aller Regel einer syntaktischen Konstituente entsprechen. Des Weiteren soll er, wie oben ausgeführt, im Zweifel semantisch „eng“ gefasst werden. Oft kann das durch eine Paraphrase mit der Formel *es ist nicht so, dass ...* und Umkehrung der Polarität geprüft werden, was wir für einige Beispiele illustrieren werden.

Einige der Beispiele werden auch deutlich machen, dass sowohl Negationsoperatoren als auch ein Skopus zweiteilig sein können, und dass der Operator nicht in jedem Fall adjazent zum Skopus sein muss. In diesen Fällen markieren wir die zusammengehörigen Teile durch numerische Koindizierung.

4.3.1 Lexikalische Negationsoperatoren

Artikel: *kein(e/er), keinerlei*

- (i) Skopus: das nachfolgende Substantiv, ggf. mit Ergänzungen

(4.62) *Der Patient zeigt [keine Auffälligkeiten].*

(4.63) *Auf dem Tisch liegen [keinerlei Briefe mit Adressaufkleber].*

(ii) Skopus: der gesamte (Teil-)satz

Mitunter ist syntaktisch betrachtet nur eine Nominalphrase im Skopus, semantisch gesehen erstreckt er sich jedoch über das Verb. Dies kann über den Paraphrasentest festgestellt werden. Die Annotation orientiert sich dann an der Semantik, das Verb wird also in den Skopus aufgenommen.

(4.64) [*Ich kann keine koronare Herzkrankheit nachweisen*].

Paraphrase: *Es ist nicht so, dass ich eine koronare Herzkrankheit nachweisen kann*. Es wird hier also nicht behauptet, dass nachgewiesen wurde, dass keine koronare Herzkrankheit vorliegt (das wäre der enge Skopus: *Ich kann nachweisen, dass es nicht so ist, dass eine koronare Herzkrankheit vorliegt*).

(4.65) [*Keine koronare Herzkrankheit nachweisbar*].

Paraphrase: *Es ist nicht so, dass eine koronare Herzkrankheit nachweisbar ist*. (Begründung analog zu Bsp. 4.64)

Präpositionen: *ohne, außer, (an)statt, anstelle (von)*

Skopus: die regierte NP

(4.66) *Kammerendteilveränderungen bei Elektrosystolie [ohne ischämiediagnostische Relevanz]*.

Substantivisch gebrauchte Pronomina: *niemand, nichts, keine/r/s*

(i) Skopus: identisch mit dem Negationsoperator

(4.67) *Ich esse [nichts]*.

(ii) Skopus: Pronominaladverb, elidiertes Material aus dem Kontext

(4.68) *Ich esse [nichts davon]*.

Bei klar erkennbarem Antezedens einer Ellipse annotieren wir dieses als Skopus. Das ist ein Fall von Nicht-Adjazenz, der entsprechend durch Indizes zu markieren ist (damit im Text ersichtlich ist, welcher Operator und welcher Skopus zusammengehören).

(4.69) *Gab es [Gänse]₁? Sie hat keine₁ gesehen*.

Koordinierende Konjunktionen: *weder ... noch ...*

Skopus: die koordinierten Konstituenten

(4.70) *Es gab [weder Butter noch Brot].*

Subordinierende Konjunktionen: *ohne dass ..., anstatt dass ..., ohne zu ...*

Skopus: der nachfolgende Nebensatz

(4.71) *Er warf den Ball weg, [ohne dass er ihn beachtet hätte].*

(4.72) [Anstatt dass er nach Golm fährt], bleibt er in Berlin.

(4.73) *Er ging über die Straße, [ohne nach rechts zu gucken].*

In Beispiel 4.73 ist eine Koindizierung der beiden Operateureile nicht nötig, da sie im selben Skopus liegen und die Zuordnung damit eindeutig ist.

Verben bzw. Verbkomplexe: *verneinen, bestreiten, ausschließen, vermeiden, etc.*

Die Abgrenzung zwischen „negierenden“ und „anderen“ Verben ist mitunter nicht einfach. Als Faustregel gilt: Wenn eine vollständige Proposition negiert wird, z. B. durch *bestreiten*, ist das Verb zu annotieren; ebenso, wenn die Existenz eines Objekts in Abrede gestellt wird. Das umfasst auch kausative Verben wie *entfernen* oder *von X befreien*.

(i) Skopus: Komplementsatz. Hier kann es vorkommen, dass Negationsoperator und Skopus nicht adjazent sind.

(4.74) *Ich bestreite energisch, [dass ich diesen Text geschrieben habe].*

(ii) Skopus: Satzglied (Subjekt, Objekt). Wir illustrieren hier auch den Fall eines phrasalen Verbs, bei dem entsprechend die beiden Teile zu unterstreichen sind.

(4.75)

(a) *Der Patient [verneint einen Leistungsknick].*

(b) *Der Patient [stellt einen Leistungsknick in Abrede].*

Das Subjekt ist hier nicht Teil des Skopus, denn der Paraphrasentest ergibt: *Der Patient sagt, es ist nicht so, dass es einen Leistungsknick gibt.* Im Vergleich dazu wäre die weitskopige Alternative *Es ist nicht so, dass der Patient einen Leistungsknick bestätigt* schwächer, weil aus dem Beispielsatz folgt, dass der Patient eine Äußerung getätigt hat. Die zweite Paraphrase wäre aber auch gültig, wenn keinerlei Aussage des Patienten vorläge.

(4.76) *[Das linke Rad fehlt] schon wieder.*

Hier bestätigt der Paraphrasentest, dass nur das Subjekt im Skopus sein sollte: *Schon wieder ist es nicht so, dass das linke Rad vorhanden ist* hat die gleichen Wahrheitsbedingungen; die Variante *Es ist nicht so, dass schon wieder das linke Rad vorhanden ist* hingegen nicht.

(4.77) *Der Patient sollte [größere Anstrengungen vermeiden].*

Paraphrase: *Der Patient sollte sich keinen größeren Anstrengungen unterziehen.* Das Modalverb ist nicht Teil des Skopus.

Substantive (vor allem Nominalisierungen): *Ausschluss, Abwesenheit, etc.*

Skopus: die regierte Nominalphrase, sofern explizit gegeben.

(4.78) *[Ausschluss einer koronaren Herzkrankheit].*

Paraphrase: *Es ist nicht so, dass eine koronare Herzkrankheit vorliegt.*

(4.79) *Wir müssen [die Abwesenheit des Trainers] feststellen.*

Paraphrase: *Wir müssen feststellen, dass es nicht so ist, dass der Trainer anwesend ist.*

(4.80) *Die Verurteilung erfolgte in [Abwesenheit].*

Adverbien: *keinesfalls, keineswegs, nicht, niemals, nie, nirgends, nirgendwo, etc.*

Skopus: das Verb (mit evtl. Auxiliaren) und die semantisch abhängigen Konstituenten (Argumente). Dies ist mitunter die gesamte clause, ggf. auch abhängige clauses wie erweiterte Infinitive. Adverbien und Adjunkte werden nicht einbezogen, sofern sie am Rand des Skopus stehen, ihre Aussparung also nicht zu einer Zweiteilung des Skopus führt. Im Umkehrschluss: Ein Adverb, das innerhalb des Skopus steht, verbleibt der Einfachheit halber darin.

(4.81) [*Es ist keineswegs zu spät*].

Paraphrase: *Es ist nicht so, dass es zu spät ist.*

(4.82) *Heute [muss ich nicht in die Schule].*

Paraphrase: *Heute ist es nicht so, dass ich in die Schule muss.*

(4.83) *Zum Glück [muss ich heute nicht in die Schule].*

Paraphrase: *Zum Glück ist es nicht so, dass ich heute in die Schule muss.*

(4.84) *Aber [beides rechtfertigt es nicht, das marode Gebäude zu sanieren]. (procon-kreiselp)*

Der Skopus wird aber dann zweigeteilt, wenn Verb/Auxiliar und Argumente deutlich voneinander getrennt sind, wie es häufig bei Elision passiert, u. a. bei Korrektur-Konstruktionen mit *sondern*. Die Zusammengehörigkeit eines zweigeteilten Skopus wird wiederum durch Koindizierung markiert.

(4.85) *Schließlich [ging es]₁ um eine profitable Nutzung auch der kleinsten Parzelle, und [nicht um Lebensqualität für deren Bewohner]₁. (procon-bäume)*

Paraphrase: *Es ist nicht so, dass es um Lebensqualität für die Bewohner ging.*

(4.86) *..., weil [es nicht um ein anderes]₁, sondern um ein sozialeres [Berlin gehen muss.]₁ (procon-senat)*

Paraphrase: *Es ist nicht so, dass es um ein anderes Berlin gehen muss.*

Satzäquivalent: *nein, niemals, nie*

Skopus: identisch mit dem Negationsoperator

(4.87) *Kommst Du mit ins Kino? [Niemals].*

4.3.2 Morphologische Negationsoperatoren: Affixe

- Präfix: *a(n), ab-, des-, dis- il-, in-, ir-, nicht-, miss-, non-, un-*
- Suffix: *-los, -frei*

Skopus: Das Wort, das das Affix enthält. Die Markierung erfolgt in gleicher Weise bei Derivationen.

(4.88) *Der Patient ist [beschwerdefrei].*

(4.89) *Die [Beschwerdefreiheit] des Patienten dauert an.*

4.4 Hinweise

Pronomen im Skopus: Wenn der Skopus des Negationsoperators ein Pronomen ist bzw. enthält, dann wird nur dieses Pronomen annotiert, das Antezedens wird nicht berücksichtigt. (Mit anderen Worten: Anaphernresolution ist nicht Teil dieser Annotation.)

(4.90) *Im Raum steht ein ungeheuerlicher Vorwurf. Ich [bestreite ihn].*

Ellipsen: Im Falle von elliptischen Konstruktionen, bei denen der Skopus des Negationsoperators komplett elidiert ist, ist zu prüfen, ob das Antezedens im Kontext erkennbar ist; wenn ja, wird es als Skopus markiert. Wenn nein, wird der Negationsoperator zugleich auch als Skopus markiert.

(4.91) *Bei einigen Patienten [traten Nebenwirkungen auf]₁, bei anderen nicht₁.*

Geschachtelte Verneinung: Negationen können ineinander verschachtelt sein. Die Zuordnung wird wie gewohnt durch numerische Koindizierung vorgenommen. Mitunter kann der Skopus der beiden Operatoren auch identisch sein.

(4.92) [*Der Lungenbefund ist keineswegs₁ [un₂auffällig]₂]₁.*

(4.93) [*Eine koronare Herzerkrankung ist durchaus nicht₁ ausgeschlossen₂]_{1,2}.*

4.5 Abschwächungsoperatoren

Durch bestimmte Verben (u. a. Modalverben), Partikeln oder Phrasen kann in einem fraglichen Satz mit Negation eine Abschwächung (engl. *hedging*) vorgenommen werden, mit der der Satz einen schwächeren epistemischen Status erhält, eine Vermutung oder Spekulation ausdrückt. In solchen Fällen wird die Negation wie gewohnt annotiert, und zusätzlich das Element, das für die Abschwächung verantwortlich ist (Abschwächungsoperator). Als Notation dafür wählen wir eine doppelte Unterstreichung. Bei Adjazenz wird der Operator in den Skopus aufgenommen, bei Nicht-Adjazenz und bei mehrteiligen Operatoren wird wie gewohnt mit Koindizierung gearbeitet.

(4.94) *Es sind [fast keine Symptome] zu beobachten.*

(4.95) *Die Krankheit [ist beinahe nicht zu erkennen].*

(4.96) *Eine Lungenentzündung [kann₁ nicht₁ ausgeschlossen₂ werden].*

Nominale referentielle Ausdrücke

Christian Chiarcos, Manfred Stede, Saskia Warzecha

Hinweise:

- Diese Richtlinien basieren auf dem PoCoS-Kernschema von Krasavina u. Chiarcos (2007); einige der dort getroffenen Entscheidungen wurden jedoch modifiziert.
- Textbeispiele, die in diesem Kapitel ohne Quellenangabe erscheinen, sind stets fiktiv.
- Die vorliegenden Richtlinien verwenden eine Notation, wie sie „auf Papier“ oder im Text-Editor verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.
- Für Diskussion und Verbesserungshinweise zu diesen Richtlinien danken wir André Herzog, David Kaupat und Sara Mamprin.

5.1 Hintergrund: Begriffe

Als *referentielle Ausdrücke* bezeichnen wir solche sprachlichen Formen, die benutzt werden, um im weiten Sinne auf einen Gegenstand, eine Person, oder einen Sachverhalt (oder jeweils auch mehrere) der „realen Welt“ oder einer „konzeptualisierten Welt“ (wie sie nur in unserer Vorstellung existiert) zu verweisen. Dabei schließen wir auch „Gattungsbegriffe“ (sog. generische Ausdrücke), wie zum Beispiel in *Der Wal ist bekanntlich ein Säugetier* mit ein. Als Stellvertreter des Gegenstandes/der Person/des Sachverhalts gilt der *Diskursreferent*: ein Objekt, das dann in formalen Repräsentationen des Diskurses verwendet werden kann. Das heißt: Sprachliche Ausdrücke verweisen auf abstrakte Diskursreferenten, denen ihrerseits ein Gegenstand (o. ä.) der Welt entspricht.

Das hier vorliegende Annotationsschema beschränkt sich auf *nominale* referentielle Ausdrücke, also auf (im weiten Sinne) Nominalphrasen. Im Folgenden wird (der Perspektive der Text-Annotation folgend) oft der Begriff des *Markable* für solche referentiellen Ausdrücke verwendet, die Gegenstand der Annotation sind. Zentrales Thema dieses Kapitels ist also die genaue Charakterisierung des Konzepts von *Markables*, wie es dann in späteren Annotationschritten der Bearbeitung von Koreferenz oder Informationsstatus zugrunde liegt. (Mit anderen Worten: Die Identifikation von referentiellen Ausdrücken ist in der Textanalyse normalerweise kein „Selbstzweck.“)

5.2 Markables: Typen und Ausdehnung

Die Annotationsaufgabe besteht darin, jeden Text in der Leseihenfolge zu bearbeiten und alle *Markables* zu identifizieren. Nach der Markierung eines *Markables* können ihm zudem verschiedene Attribute zugewiesen werden, die die Form des *Markables* charakterisieren. Diese werden später im Abschnitt 5.3 erklärt.

Wir unterscheiden zwischen drei Typen von *Markables*, die auch entsprechend als solche zu markieren sind:

- *Primäre Markables* (im Folgenden: PM) sind solche referentiellen Nominalphrasen, die qua ihrer Form ihren Diskursreferenten als *identifizierbar* markieren. Dies entspricht für uns der linguistischen Kategorie der *Definitheit*.
- *Sekundäre Markables* (im Folgenden: SM) sind alle anderen referentiellen Nominalphrasen – typischerweise indefinite Ausdrücke (NP mit indefinitem Artikel (*ein Hund*) bzw. ohne Artikel (*gutes Wetter*)).
- *Nichtreferentielle Markables* (im Folgenden: NM) sind Nominalphrasen, deren Funktion nicht im Verweis auf einen Diskursreferenten besteht. Dazu zählen z. B. NP in idiomatischen Wendungen wie *jemandem auf die Nerven gehen*.

Um die Abgrenzung zu verdeutlichen, werden die verschiedenen Arten von PM im folgenden Abschnitt genauer beschrieben, dann folgt eine kurze Charakterisierung der SM, und anschließend eine Liste der Fälle von NM. In den zur Illustration gegebenen Beispielen sind Markables jeweils in [eckigen Klammern] markiert. Der Typ (PM, SM, NM) ist der Kürze halber meist nicht mit angegeben. Mitunter sind zum Zweck der Übersichtlichkeit in einem Beispiel nicht alle Markables markiert, sondern nur diejenigen, deren Status gerade besprochen wird.

Für alle Arten von Markables stellt sich in gleicher Weise die Frage nach ihrer *Ausdehnung*. Dies regelt folgendes Prinzip:

PRINZIP Maximum

Gegenstand der Annotation sind stets Markables maximaler Größe. Das bedeutet:

1. Ein Markable beinhaltet alle Attribute seines Kopf-Nomens: Determinierer, Präpositionen, Relativsätze, Appositionen, Rechts- und Linksversetzungen, Modifikatoren, Komplemente.
[Hans – der immer schon eine Schwäche für Susanne hatte –] war auch da.
2. Eine Parenthese ist wie eine Apposition zu behandeln, wenn der Inhalt der Parenthese einen nominalen Kopf modifiziert, sonst nicht.
[Hans] – [er] hatte immer schon eine Schwäche für Susanne – war auch da.
3. Koordinierte NP/PP werden doppelt annotiert: alle Mitglieder separat, und die gesamte koordinierte NP/PP als weiteres Markable.
[[Hans] und [Susanne]]
Dies gilt entsprechend auch für durch Komma getrennte Aufzählungen.

Wie das Koordination-Beispiel illustriert, können Markables ineinander verschachtelt sein, was aufgrund des Maximum-Prinzips auch häufig vorkommt. Eine vollständig annotierte Version des ersten Beispiels ist:

[Hans – der immer schon [eine Schwäche [für Susanne]_{PM}]_{NM} hatte –]_{PM}
war auch [da]_{PM}.

5.2.1 Begriffsbestimmung: PM

Allgemein: PM sind solche referentiellen Nominalphrasen, die qua ihrer Form ihren Diskursreferenten als für die LeserInnen *identifizierbar* markieren. Wir unterscheiden die folgenden Kategorien:

(a) Pronomen

Personal-, Demonstrativ-, Possessivpronomen und Pronominaladverbien (z. B. *da*, *dort*, *daneben*, *dahin*) sowie *beide* und *alle* im nominalen Gebrauch (d. h. nicht als Determinierer)

[Ich] habe [sie] gestern gesehen.

Reflexivpronomen sind PM, wenn sie nicht verbgebunden sind, sondern als Satzglieder fungieren (ansonsten sind sie NM; siehe unten). Der Satzgliedstatus kann durch Umstellungsprobe ermittelt werden:

[Ich]_{PM} habe [mich]_{NM} gestern gewundert. (denn: *Mich habe ich gestern gewundert.)

[Ich]_{PM} habe [mich]_{PM} gestern gesehen. (denn: Mich habe ich gestern gesehen.)

Das demonstrative Pronomen ‚*solch*‘ sowie generisch gebrauchte Pronomen sind *keine* PM; siehe unten Abschnitt 5.2.3.

(b) Definite Beschreibungen

- Mit definitivem Artikel: [*das Haus*]
- Mit Determinierer ‚*beide*‘: [*beide Häuser*]
- Mit solchen Quantoren, die die ‚Gegenstandsmenge‘ eindeutig abgrenzen. Bei der Entscheidung hilft ein Ersetzungstest: Lässt sich ein definitiver Artikel oder ein

Demonstrativum einfügen, ohne dass sich die Bedeutung verändert? Beispiel: *alle Leute* → *alle diese Leute* → *definite Beschreibung* → PM

- Mit Possessivpronomen: *[[sein] Haus]*
- Definite Genitiv-Attributionen zählen als definite possessive Beschreibung, wenn sie den Determinierer ersetzen: *[[Haralds] Haus]*

(c) Eigennamen und Titel

Typische Arten von Eigennamen sind geographische Orte, Personen, Gesellschaften, Namen von Zeitungen und Zeitschriften, sowie Namen von verschiedenen sozialen, politischen und finanziellen Institutionen.

Beispiele für Personennamen:

- *[Bertolt Brecht]* (voller Name)
- *[Bert Brecht]* (gekürzter, voller Name)
- *[Bertolt]* (Vorname)
- *[BB]* (Abkürzung)
- *[der berühmte Brecht]* (über definite Beschreibung modifizierter Name)

Komplexe Eigennamen werden nur als einziges Markable behandelt und nicht weiter aufgeteilt (*[Heidelberger Druckmaschinen Vertrieb Deutschland GmbH]*). Titel in Verbindung mit Eigennamen werden als Bestandteil dieser Eigennamen annotiert, ebenso wie Namenszusätze (*[Dr. Müller]*, *[Dr. Martin Luther King, jr.]*). Alleinstehende Titel werden ebenfalls der Kategorie ‚Eigennamen‘ zugeschlagen und erhalten das entsprechende Attribut (siehe Abschnitt 5.3): *Es hat [dem Diplomingenieur] nicht gefallen..* Ist ein Eigennamen nicht der Kopf einer NP, dann wird die Phrase als definite/indefinite NP annotiert: *[ein Schweinchen namens Babe]*

(d) Zeit- und Datumsangaben

Diese beziehen sich auf einen Zeitpunkt oder Zeitintervall

und können im Text anaphorisch wieder aufgenommen werden; deshalb behandeln wir sie als PM. Dazu zählen Datumsformate ([12.3.2012]), Wochentags- und Monatsnamen, Jahreszahlen, etc.

5.2.2 Begriffsbestimmung: SM

Als SM sind alle referentiellen Nominalphrasen zu behandeln, die den Diskursgegenstand nicht als *identifizierbar* markieren. Das sind insbesondere:

- NP mit indefinitem Artikel, unabhängig davon ob es sich um eine *spezifische* Referenz handelt (*Da läuft [ein Fuchs]_{SM} über die Straße*) oder um eine *unspezifische* (*[Einen Fuchs]_{SM} habe ich zum letzten mal vor drei Jahren gesehen!*)
- NP mit einem Quantor, der die ‚Gegenstandsmenge‘ nicht eindeutig abgrenzt: [*manche Leute*], [*einige Pflanzen*], etc.
- Artikellose NP, insbesondere „bare plurals“ (*Ich habe [Kekse]_{SM} gegessen*), aber auch singuläre Ausdrücke wie in *Heute wird [gutes Wetter]_{SM} sein* und Mengenangaben wie [*dreißig Gramm*].

5.2.3 Begriffsbestimmung: NM

Es gibt einige Arten von NP, die auf den ersten Blick wie PM oder SM „aussehen“, aber nicht als solche zu annotieren sind, da es sich nicht um *referentielle* Ausdrücke handelt. In diesen Fällen ist der Typ NM (nicht-referentielles Markable) zuzuweisen. Dazu zählen insbesondere folgende Kategorien (die dem NM als Attribut mitzugeben sind, siehe Abschnitt 5.3).

- Expletive Ausdrücke (*es*-Pronomen) und pronominale Adverbien, die Regens von Relativsätzen sind
[Es]_{NM} galt zudem als sicher, dass ...

- NP in festen, konventionalisierten Redewendungen und entsprechenden Kollokationen
Er brachte mich auf [die Palme]_{NM} – ‚die Palme‘ ist nicht als referentiell zu lesen. Ähnlich: Und dann warf sie [die Flinte]_{NM} [ins Korn]_{NM}.
Aber: Markables in produktiven Metaphern werden nicht als NM, sondern als PM annotiert:
So lässt sich [das schlingende City-Schiff]_{PM} vielleicht doch noch auf einen erfolgversprechenden Kurs bringen. (maz-18914)
- Prädikativ gebrauchte NP, die einem Diskursreferenten einen weiteren „Typ“ zuschreiben
Der Chefarzt war [ein echter Fachmann]_{NM}.
Max Müller ist [der größte Mittelstürmer aller Zeiten]_{NM}!
- NP im Skopus einer Negation, die nicht wiederaufnahmefähig sind
Ich habe mir dann doch [kein neues Auto]_{NM} gekauft.
Unter bestimmten Umständen sind Wiederaufnahmen von negierten NP jedoch möglich; wenn eine solche erfolgt, ist diese NP als SM zu markieren
*[Ich]_{PM} habe [keinen Elch]_{SM} mehr. [Er]_{PM} ist gestern davongelau-
fen.*
- „Generische“ Pronomen wie zum Beispiel *wir, du, sie* (in Fällen, in denen keine spezifische Referenz vorliegt), *man, einer*
Meier sagte zu Müller: „[Man]_{NM} sollte jetzt gehen.“
Meier sagte zu Müller: „Letzes Jahr haben [sie]_{NM} hier ein Haus abgerissen.“
- Reflexivpronomen als verbgebundene Einheiten, die keinen Satzgliedstatus haben (siehe die Beispiele oben unter PM)

5.2.4 Keine Markables

Bestimmte Wörter können womöglich mit Markables im Sinne dieser Richtlinien verwechselt werden – sie sind aber keine Markables

und dementsprechend nicht zu annotieren. Das betrifft vor allem Pronominaladverbien wie *damit*, wenn sie als Konnektor fungieren: *[Auf dem Tisch]_{PM} liegt [eine Kneifzange]_{SM}. [Damit]_{PM} kann man viel anfangen.*

aber

[Ich]_{PM} habe [dir]_{PM} [den Brief]_{PM} gezeigt, damit [du]_{PM} [bescheid]_{NM} weißt.

5.2.5 Einige syntaktische Aspekte von Markables

(a) Präpositionalphrasen (P + Komplement)

[Das Flugzeug] landete [auf der Rollbahn].

Wenn eine NP als Komplement einer PP auftritt, muss (im Sinne des Prinzips ‚Maximum‘) der gesamten PP der Status eines Markables zugesprochen werden und nicht der eingebetteten NP allein.

(Achtung: Ambige Wörter wie ‚bis‘ und ‚als‘ sind immer nur dann in das Markable zu integrieren, wenn sie als Präpositionen und nicht als Konjunktionen fungieren.)

Postpositionen und Zirkumpositionen (wie ‚von Berufs wegen‘) werden wie Präpositionen behandelt.

(b) Relativpronomen

Relativpronomen werden mit dem gesamten Relativsatz als ein einziges Markable annotiert, doch Relativpronomen in Possessiv-Konstruktionen werden als Possessiv-Pronomen behandelt. Test: Ein (nicht-possessives) Relativpronomen ist genau dann gegeben, wenn es durch ‚welch‘ ersetzt werden kann:

*Und so schielten [die Israelis] [nach Washington], [an [dessen]/*welchem Tropf] [sie] hängen. (maz-19074)*

Und so schielten [die Israelis] [nach Washington, das/welches [sie] wirtschaftlich stützt].

(c) Koordination von Markables

Wie in Prinzip ‚Maximum‘ genannt, ist bei Koordinationen zusätzlich zu den einzelnen NP die gesamte koordinierte NP

zu annotieren. Wenn alle der einzelnen NP jeweils PM sind, bildet die gesamte NP ebenfalls ein PM, sonst ein SM.

Da läuft [[mein Hund]_{PM} oder [ein Fuchs]_{SM}]_{SM}.

[[Hans]_{PM} und [Maria]_{PM}]_{PM} haben heute schulfrei.

(d) **Diskontinuierliche Markables**

Markables bestehen nicht immer aus aneinander grenzenden Elementen; dies ist vornehmlich bei Relativsätzen der Fall, die nicht unmittelbar neben dem Kopfnomen stehen. Die beiden Teile sind dann per Koindizierung durch Kleinbuchstaben als zusammengehörig zu markieren: *Ich habe [das Eichhörnchen]_{PM,a} gesehen, [das quer durch den Garten lief]_{PM,a}.*

Ein weiteres Beispiel sind *split-NP*-Konstruktionen im Deutschen: *[Bücher]_{SM,a} hat Anna [drei]_{SM,a}.*

Diese beiden Phänomene können auch in Kombination auftreten, dann hat das Markable drei Bestandteile: *[Tee]_{PM,a} haben wir [den besten]_{PM,a} gekauft, [den wir kriegen konnten]_{PM,a}.*

Beachte, dass sich die PM-Eigenschaft (Identifizierbarkeit) hier auf den zunächst unspezifischen (also SM) *Tee* überträgt.

(e) **Elliptische Markables**

Eine NP kann durch Elision unvollständig sein und auf den ersten Blick nicht die Kriterien eines Markable erfüllen:

Ich hatte [zwei Stunden]_{PM} eingeplant, aber es wurden [drei]_{PM}.

Einzelne Zahlwörter sind normalerweise keine PM, doch wir vervollständigen in diesen Fällen die elidierte NP und erhalten damit ein PM.

5.3 Attribute

Es folgt eine Auflistung sämtlicher Attribute und ihrer möglichen Werte, die für Markables zu annotieren sind.

Achtung: Wenn Markables in einer Koordination verschiedene Features haben, dann sind für das übergeordnete Markable (das also „merkmalskonfligierende“ Markables umfasst) sämtliche Werte dieser Attribute als *other* zu annotieren.

NM-type (nur für nichtreferentielle Markables)

1. *none* (default)
2. *expl* – expletiver Ausdruck, Füllwort
3. *idiom* – NP in idiomatischer Wendung
4. *pred* – prädikativ gebrauchte NP
5. *other* – sonstige

Direct-speech

1. *text level* (default) – Referenz auf der Textebene
2. *dir* – Referenz auf oder innerhalb direkter Rede
3. *indir* – Referenz auf oder innerhalb indirekter Rede

Phrase-type

1. *np* (default) – nominale Phrase NP
2. *pp* – präpositionale Phrase PP
3. *other* – Phrasentyp, der nicht eindeutig klassifiziert werden kann

NP-form (gilt für NP & PP)

1. *none* (default)
2. *ne* – Eigennamen
3. *def-np* – definite NP
4. *indef-np* – indefinite NP
5. *pper* – Personalpronomen (*du, wir*)
6. *ppos* – Possessivpronomen (*seine, meine*)

7. *pds* – Demonstrativpronomen (*dieser, jener*)
8. *padv* – pronominale Adverbien (*davor, deswegen*)
9. *other*

Ambiguity

1. *not-ambig* (default) – Es besteht keine Unsicherheit.
2. *ambig-idiom* – Es besteht Unsicherheit, ob ein Markable als referentieller Ausdruck oder als Teil eines Idioms verstanden werden soll.
3. *ambig-expl* – Es besteht Unsicherheit, ob ein Pronomen ein Füllwort (Expletivum) und damit nicht referentiell oder aber doch anaphorisch ist.
4. *Ambig-other* – Sonstige Fälle von Ambiguität.

Complex-NP: Eine Beschreibung ist komplex, wenn sie mehr als eine Nominalphrase enthält.

1. *not-specified* (default)
2. *yes*
3. *no*

Entsprechender Test: Besteht eine Beschreibung aus mehr als einer nominalen (nicht pronominalen!) Phrase, so betrachte sie als komplexe Beschreibung. (Nominale Phrasen sind Phrasen, die über Kopf-Nomen verfügen, sowie als Komplement in Satzstrukturen fungieren können (Subjekt, Objekt)).

sein guter Freund und Finanzberater → typische komplexe NP

Grammatical-role: Oberflächenorientierte Definition grammatischer Rollen. Wird eine NP an der Prädikatposition verwendet, dann ist sie als ein Subjekt zu annotieren.

1. *not-specified* (default)
2. *SBJ subject* – Der Teil des Satzes oder clause, über den etwas gesagt wird; normalerweise Agens der Handlung; üblicherweise ein Nomen oder Pronomen (Nominativ)
3. *DIR-OBJ direct object* – Ein Nomen oder Pronomen (Akkusativ), das die Handlung des Verbes erfährt oder das Resultat der Handlung anzeigt. Test: direktes Objekt beantwortet die Fragen *Wen?/Was?* nach dem Verb.
4. *INDIR-OBJ indirect object* – Für die Annotation betrachten wir das indefinite Objekt als eines, das dem direktem Objekt vorausgeht und sagt, wem gegenüber oder für wen die Handlung des Verbes erfolgt und wer das direkte Objekt erhält/erfährt. Ein indirektes Objekt wird also nur dann annotiert, wenn auch ein direktes Objekt vorliegt. Indirekte Objekte antworten auf die Frage *Wem?* und werden häufig bei Verben des Kommunizierens oder Gebens gefunden. Sie sind immer ein Nomen oder ein Pronomen (Dativ), das nicht Teil einer PP ist.
5. *other* – PPs und eingebettete Elemente

5.4 Analyse eines Beispieltexts

Anhand eines Beispieltexts illustrieren wir noch einmal die Zuordnung der verschiedenen Typen von Markables, lassen die Zuweisung von Attributen hier jedoch beiseite.

1 Dass [er]_{PM} [2011]_{PM} [in Stockholm]_{PM} [den Literaturnobelpreis]_{PM}
2 erhielt, verwunderte nicht nur [wegen des Umfangs [des Gesamt-
3 werks [von Tomas Gösta Tranströmer]_{PM}]_{PM} – [es]_{PM} umfasst
4 gerade einmal [500 Seiten]_{SM} – sondern [es]_{PM} überraschte auch,
5 weil [man]_{NM} [dort]_{PM} [mit diesen Seiten]_{PM} [nach Jahrzehnten]_{SM}
6 erstmals wieder [Lyrik]_{SM} ausgezeichnet hat.
7 Keine Frage, auch [[von Günter Grass, der [den Preis]_{PM}
8 [1999]_{PM} entgegennahm]_{PM}, [Elfriede Jelinek, Preisträgerin
9 von [2004]_{PM}]_{PM} oder [Harold Pinter, Preisträger [2005]_{PM},]_{PM}]_{PM}
10 gibt es [Gedichte]_{SM}, aber geehrt wurden [diese drei]_{PM} explizit
11 [[für prosaische Arbeiten]_{SM} oder [[ihre]_{PM} Dramen]_{PM}]_{SM}.
12 Allerdings können [wir]_{NM} [uns]_{NM} an [2009]_{PM} erinnert fühlen,
13 als [Herta Müller, [zu [deren]_{PM} Werk]_{PM} neben [Romanen]_{SM} auch
14 [literarische Collagen]_{SM} zählen]_{PM}, [dem Komitee zufolge]_{PM}
15 [dafür]_{PM} geehrt wurde, dass mittels [Verdichtung]_{SM} [Landschaf-
16 ten [der Heimatlosigkeit]_{PM}]_{SM} gezeichnet werden.
17 Vielleicht zeigen [diese beiden Preisträger]_{PM} also, dass
18 [Verdichtung]_{SM} und [sprachliche Experimentierfreude]_{SM} gerade
19 hoch [im Literaturkurs]_{PM} stehen. [Das]_{PM} würde zum einen
20 erklären, weshalb [[Philip Roths]_{PM} Romane]_{PM} wieder nicht
21 gewürdigt wurden, zum anderen erklärt [es]_{PM}, weshalb [den
22 sich [an Bestsellerlisten]_{SM} orientierenden Lesern]_{PM} [bis zur
23 jeweiligen Auszeichnung]_{PM} [[Tomas Tranströmer]_{PM} und [Herta
24 Müller]_{PM}]_{PM} kein Begriff gewesen sein dürften.
25 Gratulieren wir also [beiden]_{PM} [zu [ihrem]_{PM} Verdienst]_{PM}
26 und bleiben wir gespannt [auf die Verleihungen [der nächsten
27 Jahre]_{PM}]_{PM}.

Wir diskutieren im Folgenden einige der Annotationsfälle, jedoch nicht alle – insbesondere dann nicht, wenn sich die besprochenen Phänomene innerhalb des Beispieltexts wiederholen.

(Z. 1) Jahreszahlen sind Zeitangaben und mithin PM. Das Substantiv *Stockholm* ist ein Eigenname, also ebenfalls ein PM, und wird inklusive der Präposition markiert, da nach Prinzip ‚Maximum‘ Markables maximaler Größe annotiert werden.

(Z. 2-3) *Wegen* leitet drei ineinander eingebettete PM ein: zwei definite Beschreibungen und ein Eigenname.

(Z. 3-4) *Es* wird als PM annotiert, da es einer Parenthese angehört, die keinen Kopf modifiziert. *500 Seiten* wird nicht als PM markiert, da dem Zahlwort kein Determinierer vorausgeht, wie es in Z. 10 bei *diese drei* der Fall ist. Stattdessen handelt es sich um ein SM.

(Z. 4) *Es* bezieht sich auf die Preisverleihung und bildet somit ein PM.

(Z. 5) Das *man* ist als generisches Pronomen nicht referentiell, also NM. *Dort* fungiert als Pronominaladverb und greift die Ortsbeschreibung *in Stockholm* wieder auf. *Jahrzehnte* ist ein ‚bare plural‘ und bildet zusammen mit der Präposition ein SM.

(Z. 6) *Lyrik* ist als artikelloser Gattungsbegriff ein SM (weitere finden sich in Z. 10-11).

(Z. 7-9) Die drei AutorInnen *Grass, Jelinek, Pinter* bilden jeweils für sich und zusätzlich als koordinierte Phrase auch ein gemeinsames PM. Die Einzel-PM sind jeweils nach dem Prinzip ‚Maximum‘ behandelt (mit Relativsatz bzw. Apposition).

(Z. 11) In der Koordination *für prosaische Arbeiten oder ihre Dramen* bildet der Gesamtausdruck ein SM, da nur *ihre Dramen* ein PM ist, nicht aber *prosaische Arbeiten*.

(Z. 12-13) Hier ist *wir* generisch zu lesen und deshalb ein NM. Die generische Lesart verhindert allerdings nur bei Pronomina den PM-Status, die Diskursreferenten *den Literaturnobelpreis* und *den sich an Bestsellerlisten orientierenden Lesern* zeigen jeweils ein Beispiel für generische NP im Singular/Plural, die trotzdem (hier aufgrund des definiten Determinierers) primäre Markables sind. Das Reflexivpronomen *uns* hat keinen Satzgliedstatus und ist daher ein NM. *Herta Müller* muss als maximales PM, also unter Einschluss der Apposition, markiert werden, die wiederum eine Reihe eingebetteter Markables enthält. *Als* wird nicht markiert, da es hier als Konjunktion, nicht als Präposition gebraucht wird.

(Z. 14) Das PM *dem Komitee zufolge* enthält sowohl den Artikel als auch die Postposition. *dafür* fungiert als referentielles Pronominaladverb und ist somit auch PM.

(Z. 19-21) *im Literaturkurs* sollte als produktive Metapher gelesen und als PM markiert werden, während *zum einen* und *zum anderen* im folgenden Satz trotz Definitheit nicht markiert werden, da die Begriffe lexikalisiert sind.

(Z. 21) *Es* ist hier kein Expletivum, sondern bezieht sich auf *das* in Z. 19, stellt also ein PM dar.

(Z. 23-24) In *Tomas Tranströmer und Herta Müller* werden beide PM zu einem neuen PM koordiniert.

(Z. 24) *Ein Begriff sein* ist eine idiomatische Wendung, es ergibt sich also kein Markable.

Nominale Koreferenz

Christian Chiarcos, Manfred Stede, Saskia Warzecha

Hinweise:

- Diese Richtlinien basieren auf dem PoCoS-Kernschema von Krasavina u. Chiarcos (2007); einige der dort getroffenen Entscheidungen wurden jedoch modifiziert.
- Alle in diesen Richtlinien genannten Beispiele ohne Quellengabe sind fiktiv.
- Die vorliegenden Richtlinien verwenden eine Notation, wie sie „auf Papier“ oder im Text-Editor verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgeben.
- Für Diskussion und Verbesserungshinweise zu diesen Richtlinien danken wir André Herzog, David Kaupat und Sara Mamprin.

6.1 Hintergrund: Begriffe

Koreferenz bezeichnet die Relation zwischen zwei Textelementen, die auf dieselbe Entität (denselben *Diskursreferenten*) verweisen. Die Textelemente nennen wir *referentielle Ausdrücke*; man spricht oft von *Wiederaufnahme*, wenn derselbe Diskursreferent mehrfach erwähnt wird.

Eine häufige Spielart von Koreferenz ist *Anaphorik*: Wenn ein referentieller Ausdruck allein nicht interpretierbar ist – sein Diskursreferent lässt sich nur unter Rückgriff auf einen vorher genannten koreferenten Ausdruck ermitteln – handelt es sich um eine *Anapher*. Der vorangehende koreferente Ausdruck ist dann das zugehörige *Antezedens*. Beispiel:

(6.97) *Susanne mag [das Turnen]₁ nicht, denn [es]_{2>1} ist sehr schwer.*

Wir notieren referentielle Ausdrücke in eckigen Klammern; wenn sie in einer Koreferenz-Relation stehen, erhalten sie einen numerischen Index, und das Zeichen $>$ zeigt die Bezugsrichtung (Anapher zu Antezedens) an.

Nicht-anaphorische Koreferenz liegt also vor, wenn zwei referentiellen Ausdrücken auch in Isolation der gleiche Diskursreferent zugeordnet wird. Das ist vorwiegend bei Eigennamen oder häufig auch bei definiten Nominalphrasen der Fall. Im nachfolgenden Beispiel gibt es zwei Koreferenz-Relationen, von denen eine anaphorisch ist.

(6.98) *Mittags eröffnete [der Bundespräsident]₁ [die Sitzung]₂, und am Abend hat [Joachim Gauck]_{3>1} [sie]_{4>2} wieder geschlossen.*

Wir nennen die Serie der Erwähnungen desselben Referenten eine *referentielle Kette*. Ob ein referentieller Ausdruck e zu einer Kette k gehört, kann mit einem Ersetzungstest geprüft werden (siehe auch Abschnitt 6.4.1): Wenn für jedes Substantiv s (Nomen, Eigenname) in k gilt, dass die Ersetzung von e durch s die Interpretation des Textes nicht verändert, so gehört e zur Kette k und es ist eine Koreferenz-Relation zum letzten Element der Kette zu annotieren.

Dieses Annotationsschema beschränkt sich auf *nominale* Diskursreferenten. Das bedeutet, es werden nur referentielle Ketten behandelt, deren Komponenten jeweils durch eine (im weiten Sinne) Nominalphrase ausgedrückt sind. Relationen, die durch sog. Sachverhaltsanaphern gestiftet werden, bei denen das Antezedens eine Verbalphrase, ein Satz, oder auch ein ganzer Teiltext sein können, werden hier nicht behandelt.

Im Folgenden wird (der Perspektive der Text-Annotation folgend) der Begriff des *Markable* für referentielle Ausdrücke verwendet. Die Bestimmung der Markables und ihres Status als *primär*, *sekundär* oder *nicht-referentiell* ist Gegenstand der Annotationsrichtlinien in Kapitel 5; dieser Bearbeitungsschritt wird für die folgenden Richtlinien vorausgesetzt.

Zwischen den Markables sollen nun also referentielle Ketten etabliert werden, anhand derer dann die referentielle Struktur eines

Textes bestimmt werden kann; diese macht explizit, welche Diskursgegenstände wie oft und in welcher Form behandelt werden und gibt Hinweise auf zentrale bzw. durchgehend zur Sprache kommende Diskursreferenten. Damit erlaubt sie oft Rückschlüsse auf die Gliederung des Textes in thematisch zusammenhängende Abschnitte.

6.2 Annotationsanweisung

6.2.1 Ziel der Annotation

Das Ziel der Analyse der referentiellen Struktur eines Textes ist eine Erfassung aller Paare von koreferenten Markables und damit der Etablierung von referentiellen Ketten. Wir unterscheiden dabei drei Annotationsebenen: *Markables*, *Relationen*, sowie *Attribute*. Da die Markables bereits vorliegen, ist das konkrete Ziel für die Koreferenzanalyse, jedes *primäre Markable* (siehe Abschnitt 5.2.1) hinsichtlich der Fragestellung zu annotieren, ob und welchem Antezedens-Markable es zugeordnet ist (siehe unten Abschnitt 6.3) und damit eine *Relation* (Abschnitt 6.4) zu etablieren. Außerdem ist zu entscheiden, welche spezifischen *Attribute* das behandelte Markable hinsichtlich der Koreferenz trägt (Abschnitt 6.5).

6.2.2 Annotationsverlauf

Schritt 1: Identifiziere die bereits markierten primären Markables (PM).

Für jedes PM α :

Schritt 2: Verbinde Markables durch Koreferenz-Verknüpfungen.

- (a) Prüfe ob α koreferent mit einem PM oder SM im vorangehenden Kontext ist; verwende dazu den in Abschnitt 6.4.1 beschriebenen Ersetzungstest.
- (b) Wenn Koreferenz festgestellt wird, schreibe α sein unmittelbares Antezedens zu (letzte Nennung). Notation: Index für α

,>' Index des Antezedens. (Hat das Antezedens noch keinen Index, etabliere einen neuen.)

- (c) Besteht keine Koreferenz im vorangehenden Kontext, prüfe ob α kataphorisch auf ein Element im nachfolgenden Kontext verweist (siehe dazu Abschnitt 6.4.2); wenn ja, annotiere die Relation. Notation: Index für α ,>' Index des nachfolgenden „Ante“zedens.
- (d) Liegt keine Koreferenz vor, so bekommt α keinen numerischen Index.

Schritt 3: Lege für α die Werte der Attribute fest.

Am Ende enthalten die Texte dann referentielle Ketten (deren erstes Glied ein primäres oder ein sekundäres Markable sein kann) sowie einzeln annotierte Markables (also solche, die im Text nicht wieder aufgenommen werden, engl. *singletons*). Mitglieder von Ketten tragen einen numerischen Index, singletons nicht. In welcher Reihenfolge die Indizes vergeben werden (also welche Nummern vergeben werden), ist unerheblich; entscheidend sind die Koreferenzbeziehungen. Der Übersichtlichkeit halber empfiehlt es sich aber, Indizes bei der Bearbeitung des Textes in einer aufsteigenden Reihenfolge zu vergeben.

Im folgenden Abschnitt werden zunächst einige wichtige Aspekte dieses Annotationsverlaufs genauer erläutert. Abschnitt 6.4 diskutiert den Unterschied zwischen anaphorischen und kataphorischen Relationen, und gibt dann Hinweise zum Umgang mit Ambiguität. Anschließend erklärt Abschnitt 6.5 die zu setzenden Attribute (Schritt 3 im obigen Verlauf). Am Ende dieser Richtlinien werden die Annotationsprinzipien noch einmal zusammengefasst, anschließend illustrieren wir die Annotationsentscheidungen anhand eines Beispieltexts.

6.3 Einzelheiten

6.3.1 Antezedens-Entscheidung

Die Zuweisung eines Antezedens wird in den meisten Fällen recht eindeutig sein. Sind aber mehrere Interpretationen gleichermaßen plausibel, so kann Prinzip 1 die Entscheidung regeln, wenn die Wahl zwischen einem PM und einem SM besteht.

PRINZIP 1

$[PM > SM]$

Gibt es je nach Interpretation mehrere Kandidaten für das Antezedens eines referentiellen Ausdrucks, so ist ein PM einem SM als Antezedens vorzuziehen. Diese Präferenz gilt unabhängig von der Distanz zum Antezedenskandidaten.

6.3.2 Behandlung diskontinuierlicher Markables

Wird ein diskontinuierliches Markable anaphorisch aufgenommen, so erstelle eine Relation von der Anapher zum *letztgenannten* Teil des Antezedens-Markables. Ist hingegen die Anapher diskontinuierlich, verbinde ihren *erstgenannten* Teil mit dem Antezedens.

Wir markieren hier die Zusammengehörigkeit der Markable-Teile durch Kleinbuchstaben und die Koreferenz durch Zahlen.

(6.99) *[Ich]_{PM,1} hatte [von dem Tee]_{PM,a} getrunken, [den [Anna]_{PM} [mir]_{PM,2>1} geschenkt hatte]_{PM,a,3}. Ehrlich gesagt schmeckte [er]_{PM,4>3} scheußlich.*

6.3.3 Rekursive Einbettung

Aufgrund des Prinzips der maximalen Ausdehnung von Markables (siehe Kapitel 5) kann es zu rekursiv eingebetteten Bezügen kommen. Zwar sind Relativpronomen keine Markables, doch im Falle einer im Relativsatz oder in der Apposition auftretenden Possessivkonstruktion verweist das Possessivpronomen auf die gesamte umgebende NP.

(6.100) *Und so schielten [die Israelis]_{PM,1} [nach Washington, [an
[dessen]_{PM,4>3} Tropic]_{PM,5} [sie]_{PM,2>1} hängen]_{PM,3}. (maz-19074)*

6.3.4 Gruppen-Referenz

Mit NPs im Plural (auch quantifizierenden NPs) und mit Plural-Pronomen (*sie, ihnen* etc.) kann man auf Gruppen von Antezedenten referieren:

(6.101) *[Deutschland]_{PM} [gegen Argentinien]_{PM} – [beide]_{PM} siegen mit
Hilfe [des Schiedsrichters]_{PM}.*

Hier bezieht sich *beide* auf *Deutschland* und *Argentinien*, jedoch gab es im Text bisher keine Phrase, die genau diese Gruppe bezeichnet, also auch kein Markable. Deshalb muss bei diesem Annotationsschritt eine weitere Art von Markables geschaffen werden: *Gruppen-Markables* (GM).

Bei der Notation orientieren wir uns an der für diskontinuierliche Markables. Die Mitglieder der Gruppe erhalten einen Index, der sich aus ‚G‘ und einer Ziffer zusammensetzt. Für das letztgenannte Mitglied der Gruppe wird die Koreferenz mit der Anapher wie gewohnt durch numerischen Index notiert:

(6.102) *[Deutschland]_{PM,G1} [gegen Argentinien]_{PM,G1} – [beide]_{PM,1>G1}
siegen mit Hilfe [des Schiedsrichters]_{PM}.*

Steht ein in einer Gruppe enthaltenes Markable auch noch als *einzelnes* Markable in einer Koreferenz-Relation (im Beispiel 6.102 könnte das etwa durch eine erneute Erwähnung von *Argentinien* im Nachfolgesatz geschehen), so wird ihm neben dem Gruppen-Index ein weiterer Index zugeschrieben, auf den die entsprechende Anapher dann verweisen kann.

GM können als Antezedens nominaler Markables dienen (nicht aber als Anaphern). Solche Gruppen können Markables umfassen, die sich bzgl. ihrer grammatischen Rolle, der Art des referentiellen Ausdrucks und anderer Merkmale unterscheiden. Daher wird eine Gruppe als Ganzes nicht weiter mit Attributen annotiert.

Werden Gruppen und ihre Mitglieder wiederholt erwähnt, kann es zu einer für die Annotation unklaren Situation kommen. Hier ein sehr kompaktes Beispiel:

(6.103) *Peter und Malte gingen spazieren. Beide trugen einen Hut. Peter war im Mantel unterwegs, Malte in der Regenjacke. Sie erreichten ...*

Bei der Annotation von *Sie* besteht die Versuchung, gemäß dem Kettenprinzip (Prinzip 3, siehe unten) die unmittelbar davor stehenden Erwähnungen von *Malte* und *Peter* erneut als Gruppe zu bündeln und als Antezedens zu markieren. Genau diese Gruppe wurde jedoch bereits zuvor durch die Anapher *beide* etabliert. Wir legen fest, dass in dieser Situation Koreferenz zwischen *sie* und *beide* zu annotieren ist, um die Identität der Gruppenverweise transparent zu machen.

Diese Frage der Vorrangigkeit der Identität von Gruppen vor dem weiter unten erläuterten Kettenprinzip (Prinzip 3) regelt das Prinzip 2 (eine Erweiterung von Prinzip 1):

PRINZIP 2

[*PM* > *SM* > *Gruppe*]

Wenn ein PM oder SM Antezedens eines Markables sein kann, so etabliere diese Verbindung, anstatt ein neues Gruppen-Markable zu deklarieren – selbst wenn das PM oder SM im Vergleich zum möglichen Gruppen-Markable sehr viel weiter links von der Anapher lokalisiert ist.

Weitere Bemerkungen zur Rangfolge der einzelnen Prinzipien folgen am Ende dieser Richtlinien.

6.4 Relationen

Die folgenden Anweisungen betreffen die zwischen Markables zu annotierenden Relationen. Wir differenzieren hier nicht zwischen anaphorischen und nicht-anaphorischen Koreferenz-Relationen.

Für die Annotation der Relation verwenden wir in diesem Abschnitt daher ‚anaphorisch‘ in einem weiten Sinne, der jede Wiederaufnahme umfasst.

6.4.1 Anaphorische Relationen

Wir annotieren gemäß diesem Schema ausschließlich Koreferenz-Relationen, die eine echte Identität zwischen Diskursgegenständen ausdrücken. Eine „lockerere“ Verbindung genügt also nicht. Dazu der folgende Test: Es gelten zwei gegebene nominale Beschreibungen als koreferent, wenn es möglich ist, sie untereinander durch die jeweils andere zu *ersetzen*. (Dabei können gewisse Umformungen nötig sein, etwa die Entfernung von Präpositionen aus Markables.)

Beispiel (viele hier nicht interessierende Markables sind unmarkiert):

(6.104) *Als 1999 die im Rahmen der Dorferneuerung neu gestaltete [Radeweger]_{PM,1} Ablage inklusive Seebücke mit viel Pomp eingeweiht wurde... Doch mit der Nachrüstung tut sich [Radewege]_{PM,2>1} schwer... Zu teuer, zu hässlich sei die Anlage, sagen die Meinungsführer [im Gemeinderat]_{PM,?} (maz-6488)*

Hier könnte [Gemeinderat] zwar möglicherweise als koreferent mit [Radewege]_{PM,2} gelesen werden, doch obwohl beide im Sinne einer Metonymie austauschbar sind, scheitert der Ersetzungstest für [Radewege]₁ – denn „neu gestaltete Gemeinderat Ablage“ ist in diesem Kontext nicht angemessen. Daher erhält Gemeinderat einen anderen Index.

Für die Annotation des Antezedens ist nach Prinzip 3 folgende Vorrangsregelung zu beachten:

PRINZIP 3 (Kettenprinzip)

[*Right Previous* > *Left Previous*]

Jedes anaphorische Markable hat genau ein Antezedens. Markiere immer die jüngste, ‚right-most‘ vorausgegangene Referenten-Erwähnung als Antezedens. Sämtliche Erwähnungen desselben Referenten ergeben eine linear geordnete Kette.

(Ausnahmen entstehen bei der Annotation von Ambiguität, wenn mehrere Kandidaten markiert werden.)

Wie eingangs erläutert, werden sog. Sachverhaltsanaphern zwar markiert (es handelt sich um PM), ihnen wird aber kein Antezedens zugewiesen, wenn dies nicht-nominal ist. Wenn die Sachverhaltsanapher selbst aber wiederholt auftritt, so soll zwischen ihnen eine anaphorische Relation markiert werden. Beispiel:

(6.105) *Gestern hat Bayern München schon wieder gewonnen.*

[*Das*]_{PM,1} *hat Jan ziemlich gestört. Marianne hingegen war*
[*davon*]_{PM,2>1} *begeistert.*

D. h., *Das* bekommt kein Antezedens, während zwischen *davon* und *Das* eine Relation markiert wird.

6.4.2 Kataphern

Es wird zwischen zwei Arten von vorwärts gerichteten Ausdrücken unterschieden: Diskurskataphern und syntaktischen Kataphern.

6.4.2.1 Diskurskataphern

Als Diskurskataphern werden jene vorwärtsgerichteten Ausdrücke bezeichnet, die nicht pronominal realisiert sind, sondern beispielsweise durch eine unterspezifizierte NP, die zum Zeitpunkt ihres Auftretens im Text noch nicht interpretiert werden kann. Der Autor mag auf diese Weise den Leser dazu ermutigen, weiterzulesen, um die fehlende Information zu erlangen. Beispiel:

(6.106) [*Die einstige Fußball-Weltmacht*]_{PM} *zittert [vor einem*
*Winzling]*_{SM}. *Mit seinem Tor zum 1 : 0 für die Ukraine stürzte der*

1,62 Meter große Gennadi Subow die deutsche Nationalelf vorübergehend in ein Trauma. Je kleiner die Kicker daherkommen, desto größer wird der Gegner geredet... (maz-10374)

Bei der Verarbeitung des ersten Satzes wird nur ein PM identifiziert. Erst wenn der Text weitergelesen wird, wird deutlich, dass sich *die einstige Fußballweltmacht* auf *die deutsche Nationalelf* bezieht, und darüber hinaus *vor einem Winzling* (als SM) je nach Interpretation entweder auf *die Ukraine* oder auf *der 1,62 Meter große Gennadi Subow*. (Ähnlich sind dann auch im nächsten Satz für *die Kicker* mehrere Interpretationen denkbar.)

Diskurskataphern werden wie normale Anaphern annotiert, also in Übereinstimmung mit dem Kettenprinzip (Prinzip 3): Es ist immer die jüngste vorausgegangene Referenten-Erwähnung (sofern es eine gibt) als Antezedens zu wählen. Für das Attribut *referentiality* wird der Wert *discourse-cataphora* angegeben.

6.4.2.2 Syntaktische Kataphern

(6.107) *Durch [seinen]_{PM,1>2} Anwalt hat [Herr Antar]_{PM,2} die Vorwürfe zurückweisen lassen.*

Syntaktische Kataphern (in der Regel pronominal realisiert) sind wie anaphorische, gerichtete Verbindungen zu annotieren, aber *in umgekehrter Richtung* (von links nach rechts zeigend). Die beiden Elemente treten in aller Regel im selben Satz auf. Für das Attribut *referentiality* ist, ebenso wie bei anaphorischen Relationen, der Wert *referring* anzugeben.

Ein weiteres Beispiel findet sich auch oben in 6.106, wo *seinem* ein kataphorischer Verweis auf *Gennadi Subow* ist.

Falls nicht eindeutig zwischen einer syntaktischen Katapher und einer Anapher entschieden werden kann, ist Prinzip 4 anzuwenden:

PRINZIP 4 (Kataphern auf Satzebene)

Kataphorische Relation innerhalb eines Satzes > Kettenprinzip
(right previous > left previous)

Falls die in Frage stehende kataphorische Relation die Satzgrenze nicht überschreitet, ist diese Interpretation vorzuziehen, anstatt, wie Prinzip 3 es erfordern würde, eine anaphorische Relation zum nächsten vorstehenden Referenten herzustellen. (Prinzip 4 ist somit höherrangig als Prinzip 3.)

(6.108) *Cannes ehrt erneut einen Großmeister des Films. Für [seinen]_{PM,1>2} neuen Film erhält [Michael Haneke]_{PM,2} bereits zum zweiten Mal die Goldene Palme.*

Hier befinden sich *seinen* und *Michael Haneke* in einem Satz, weshalb nach Prinzip 4 diese (vorwärtsgerichtete) Verbindung von *seinen* auf *Michael Haneke* der rückwärtsgerichteten, anaphorischen von *seinen* auf *einen Großmeister des Films* vorzuziehen ist. Die Koreferenzkette wird vollständig durch die Annotation der Referenz von *Michael Haneke* auf *einen Großmeister des Films*:

sein → *Michael Haneke* (syntaktische Katapher innerhalb des Satzes)
Michael Haneke → *einen Großmeister des Films* (Kettenprinzip)

6.4.3 Ambige Antezedenten

Herrschen Zweifel, ob etwas entweder mit einem Gegenstand aus dem vorausgegangenen Diskurs über eine anaphorische Relation verbunden ist, oder ob ein Markable eventuell gar kein Antezedens hat (expletive, idiomatische, oder generische Lesart), dann ist wie folgt vorzugehen:

PRINZIP 5

[anaphorische Relation > sonstige (generisch, prädikativ, etc) / idiomatisch / expletiv]

Die anaphorische Referenz ist vorzuziehen und somit eine Verbindung zu einem korrespondierenden Antezedens zu schaffen.

Wenn klar ist, dass es sich um eine anaphorische Relation handelt, aber das Antezedens nicht eindeutig oder mit klarer Präferenz identifiziert werden kann, wird im Sinne einer Disjunktion für jeden Kandidaten eine anaphorische Relation etabliert. An der

Anapher wird die Ambiguität durch ein Attribut markiert (siehe Abschnitt 6.5.2).

Achtung: Markables mit ambigen Tendenzen sollten möglichst nicht als Antezedens genutzt werden. (Dies ist eine Ausnahme vom Kettenprinzip.)

6.5 Attribute

Es folgt eine Auflistung sämtlicher Attribute für die Annotation. Die mit einem * versehenen Kategorien sind allein für SM gedacht, alle anderen für PM und SM gleichermaßen. NM werden hier nicht mit Attributen annotiert.

Achtung: Wenn Markables in einer Koordination verschiedene Werte haben, dann sind für das übergeordnete Markable alle Werte dieser Merkmale als *other* zu annotieren.

6.5.1 Referentiality

1. *not specified* (default) – Im Rahmen der Annotation wurde keine Entscheidung gefällt.
2. *referring* – Eine Diskurseinheit, die auf Basis des vorangegangenen Kontextes interpretiert werden kann.
3. *discourse-new* – Diskursentität, die zum ersten Mal erwähnt wird.
4. *discourse-cataphora* – Referenz auf eine neue Entität, welche in den Diskurs eingeführt wird, und zwar über die Bedeutung eines Ausdrucks mit einer unterspezifizierten Denotation.
5. *other* – Eine Entscheidung zwischen (1)-(4) ist nicht möglich; dies betrifft vor allem
 - generische Beschreibungen (*Wale sind Säugetiere./Der Präsident wurde immer schon durch die Stimmenmehrheit bestimmt*);

- prädikative Beschreibungen (*Nicht, dass beide eine Mehrheit für ihre Koalition suchten, war [das Ärgerliche in den vergangenen Tagen] ...*);
- Gruppen, deren Mitglieder verschiedene referentiality-features tragen.

Achtung: Wenn ein Markable sich auf eine Einheit im Text bezieht, die nicht zweifelsfrei als PM oder SM deklariert werden kann, dann ist für das Attribut ‚referentiality‘ der Wert ‚referring‘ zu wählen, aber kein Antezedens zuzuschreiben. Diese Ausnahme betrifft insbesondere diskurs-deiktische Verweise auf VPs und andere nicht-nominale Antezedenten.

6.5.2 Ambiguity

1. *not-ambig* (default) – Es besteht keine Unsicherheit.
2. *ambig-ante* – Es besteht Unsicherheit, welches das „richtige“ Antezedens für eine Anapher ist, deshalb sind Relationen zu mehreren Kandidaten etabliert.
3. *ambig-rel* – Es besteht Unsicherheit, ob eine anaphorische Relation besteht oder nicht. Dies ist mitunter bei definiten NPs der Fall, beispielsweise: (...) *Dieser Konflikt ist (...) Darum ist die Auseinandersetzung (...)* Wenn unklar ist, ob die Auseinandersetzung identisch ist mit dem Konflikt, soll die Relation etabliert und die Anapher mit diesem Attribut markiert werden.
4. *ambig-ante-rel* – Es besteht Unsicherheit *sowohl* hinsichtlich des Bestehens einer Relation *als auch* hinsichtlich des möglichen Antezedens. Dann sind die möglichen Relationen zu etablieren und dieses Attribut zu wählen.
5. *ambig-other* – Sonstige Fälle von Ambiguität.

6.5.3 Anaphora-type

1. *none* (default) – Ersterwähnung eines Referenten
2. *anaphorisch* – anaphorische Relationen

6.6 Prinzipien im Überblick

Wir geben hier noch einmal eine Zusammenfassung der genannten Prinzipien und erfassen sie anschließend in einem Ranking.

PRINZIP 1 (folgt aus Axiom (D))

[PM > SM]

Wenn PM bestehen, die Antezedens eines Markables sein können, dann muss zwischen diesen stets eine Verbindung geknüpft werden, statt etwa ein mögliches SM zu deklarieren – selbst in Fällen, in denen etwa ein PM im Vergleich zum möglichen SM viel weiter links von der Anapher lokalisiert ist.

PRINZIP 2

[PM > SM > groups]

Wenn ein PM oder SM Antezedens eines Markables sein kann, so etabliere diese Verbindung, anstatt etwa ein group-Markable zu deklarieren – selbst wenn das PM oder SM im Vergleich zum potentiellen group-Markable viel weiter links von der Anapher lokalisiert ist.

PRINZIP 3

[Right Previous > Left Previous] (Kettenprinzip)

Jedes anaphorische Markable hat genau ein Antezedens. Markiere immer die jüngste, ‚right-most‘ vorausgegangene Referenten-Erwähnung als Antezedens. All die Erwähnungen desselben Referenten bilden somit eine geordnete Kette.

PRINZIP 4

[Referenz zur Rechten im selben Satz > Kettenprinzip]

Für ein Markable, das sowohl als kataphorisches als auch anaphorisches Pronomen gelesen werden kann, gilt: Kann im selben Satz an

späterer Stelle ein Markable gefunden werden, das das Pronomen syntaktisch bindet, so ist eine kataphorische Beziehung zu annotieren, auch wenn eine vorangehende Äußerung ein anaphorisches Antezedens enthält.

PRINZIP 5

[anaphorische Relation > other (generic, predictive, etc)/idiomatic/expletive]

Die anaphorische Referenz ist vorzuziehen und eine Verbindung zu einem korrespondierenden Antezedens zu schaffen.

RANKING

Prinzip 4 (die Ausnahme von Prinzip 3) hat höheres Gewicht als Prinzip 2, welches wiederum höheres Gewicht als Prinzip 3 hat:

$$4 > 2 > 3$$

(P1 ist in P2 enthalten. P5 steht in keinem Konflikt zu irgendeinem der anderen Prinzipien und ist obligatorisch.)

6.7 Analyse eines Beispieltexts

Wir zeigen im Folgenden die Analyse eines Beispieltexts; die Zuordnung der Markables wurde (für dieses Beispiel) in Kapitel 5 vorgenommen. Es folgen dann Erläuterungen zu den Koreferenz-Entscheidungen.

-
- 1 Dass [er]_{PM,1>2} [2011]_{PM} [in Stockholm]_{PM,6} [den Literatur-
 2 nobelpreis]_{PM,10} erhielt, verwunderte nicht nur [wegen
 3 des Umfangs [des Gesamtwerks [von Tomas Gösta Tran-
 4 strömer]_{PM,2,G1}]_{PM,3}PM – [es]_{PM,4>3} umfasst gerade einmal [500
 5 Seiten]_{SM,8} – sondern [es]_{PM} überraschte auch, weil [man]_{NM}
 6 [dort]_{PM,5>6} [mit diesen Seiten]_{PM,7>8} [nach Jahrzehnten]_{SM} erst-
 7 mals wieder [Lyrik]_{SM} ausgezeichnet hat.
 8 Keine Frage, auch [[von Günter Grass, der [den Preis]_{PM,9>10}
 9 [1999]_{PM} entgegennahm]_{PM}, [Elfriede Jelinek, Preisträgerin von
 10 [2004]_{PM}]_{PM} oder [Harold Pinter, Preisträger [2005]_{PM}]_{PM,12} gibt
 11 [es]_{NM} [Gedichte]_{SM}, aber geehrt wurden [diese drei]_{PM,11>12} expli-
 12 zit für [prosaische Arbeiten]_{SM} oder [[ihre]_{PM,13>11} Dramen]_{PM}.

13 Allerdings können [wir]_{NM} [uns]_{NM} an [2009]_{PM} erinnert fühlen,
14 als [Herta Müller, [zu [deren]_{PM,15>14} Werk]_{PM} neben [Romanen]_{SM}
15 auch [literarische Collagen]_{SM} zählen]_{PM,14,G1}, [dem Komitee
16 zufolge]_{PM} [dafür]_{PM} geehrt wurde, dass mittels [Verdichtung]_{SM}
17 [Landschaften [der Heimatlosigkeit]_{PM}]_{SM} gezeichnet werden.
18 Vielleicht zeigen [diese beiden Preisträger]_{PM,15>G1} also, dass
19 [Verdichtung]_{SM} und [sprachliche Experimentierfreude]_{SM} gerade
20 hoch [im Literaturkurs]_{PM} stehen. [Das]_{PM,17} würde zum einen
21 erklären, weshalb [[Philip Roths]_{PM} Romane]_{PM} wieder nicht
22 gewürdigt wurden, zum anderen erklärt [es]_{PM,16>17}, weshalb
23 [den sich [an Bestsellerlisten]_{SM} orientierenden Lesern]_{PM} [bis
24 zur jeweiligen Auszeichnung]_{PM} [[Tomas Tranströmer]_{PM,18>2} und
25 [Herta Müller]_{PM,19>14}]_{PM,21} kein Begriff gewesen sein dürften.
26 Gratulieren wir also [beiden]_{PM,20>21} [zu [ihrem]_{PM,22>20}
27 Verdienst]_{PM} und bleiben wir gespannt [auf die Verleihungen [der
28 nächsten Jahre]_{PM}]_{PM}.

(Z. 1) Mit *er* liegt eine syntaktische Katapher vor, die auf *Tomas Gösta Tranströmer* referiert, weswegen hier die Relationsmarkierung vorwärtsgerichtet erfolgen muss. Wir vergeben dem Pronomen den Index 1, dem Eigennamen die 2 (was aber, wie oben erwähnt, eine willkürliche Entscheidung ist).

(Z. 4) *Es* verweist hier auf das *Gesamtwerk* (nicht auf den *Umfang des Gesamtwerks!*); der Index 3 ist also sorgfältig an die richtige schließende Klammer zu vergeben.

(Z. 5) Auch hier liegt ein *es* vor, welches sich aber nicht auf das Werk bezieht, sondern auf die Preisvergabe, die aber in Z. 1-2 als Sachverhalt durch eine VP beschrieben wurde; es liegt also keine nominale Koreferenz vor.

(Z. 6) *Dort* bezieht sich auf *Stockholm*, das also nun ebenfalls einen Index (6) erhält. Und *diese Seiten* ist hier als koreferent zu den *500 Seiten* in Z. 4-5 zu lesen.

(Z. 8) *Den Preis* lesen wir als referierend auf den Literaturnobelpreis als ‚Konzept‘ und damit als koreferent mit dem in Z. 1-2. (Eine andere Lesart isoliert die konkreten Instanzen (Tranströmers

Preis, Grass' Preis), die dann nicht koreferent wären; dies erscheint aber hier weniger plausibel.)

(Z. 11-12) *Diese drei* bezieht sich offenkundig auf die bereits als Koordination etablierte Menge der Preisträger, die also nun einen Index erhält; *ihre* ist dazu dann wiederum anaphorisch.

(Z. 14) *Deren* ist ein Beispiel für eine Einbettung eines Pronomens in eine NP *Herta . . . zählen*, auf die es sich als Ganzes bezieht.

(Z. 18) *Diese beiden Preisträger* muss im Kontext interpretiert werden; die naheliegendste Lesart ist ein Bezug auf *Tranströmer* (Z. 3-4) und *Müller*. Diese beiden erhalten also einen Gruppenindex G1.

(Z. 22) *Es* ist zwar eine Sachverhaltsanapher, jedoch koreferent mit *Das* in Z. 20 (seinerseits eine Sachverhaltsanapher); die Relation muss also markiert werden.

(Z. 24-25) Die Wiederaufnahme der Autorennamen ist jeweils zu markieren.

(Z. 26-27) *Beiden* bezieht sich auf *Tranströmer* und *Müller*. Für diese wurde in Z. 18 schon ein Verweis auf die entsprechende Gruppe annotiert; gemäß Prinzip 3 wählen wir jedoch die jüngste Erwähnung als Antezedens, und das ist die koordinierte NP in Z. 24-25, die nunmehr also den Index 21 erhält. Sodann kann *ihrem* als koreferent mit *beiden* annotiert werden. Das *Verdienst* ist ein typischer Fall von Mehrdeutigkeit bezüglich der Frage, ob Koreferenz im Sinne von Identität gegeben ist oder nicht: Ist das *Verdienst* identisch mit der Auszeichnung in Z. 24? Wir haben uns hier dagegen entschieden, da ein *Verdienst* eher eine Bewertung eines Vorgangs ist und daher nicht identisch mit dem Vorgang selbst. *Mithin* wird keine Ambiguität annotiert.

Aboutness Topik

Sara Mamprin, Manfred Stede

Hinweise:

- Diese Richtlinien sind eine sprach- und genrespezifische Weiterentwicklung der zweiten Version der Topik-Richtlinien des SFB 632 *Informationsstruktur* (Götze u. a., 2007). Deren grundsätzliche Vorgehen wurde übernommen, allerdings sind einige Detail-Entscheidungen unterschiedlich.
- Die vorliegenden Richtlinien basieren auf einer vorangegangenen Annotation von Diskurssegmenten (gemäß Kapitel 3) und referierenden Ausdrücken (gemäß Kapitel 5).
- Wir verwenden hier eine Notation, wie sie „auf Papier“ oder im Text-Editor verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.
- Alle Textbeispiele, für die keine Quelle angegeben ist, sind fiktiv.
- Für Diskussion und Verbesserungshinweise zu diesen Richtlinien danken wir Philippa Cook und Antje Sauer mann.

Abschnitt 7.1 erläutert den Hintergrund des Schemas und diskutiert dann im Detail die für die Topik-Annotationen zu treffenden Entscheidungen. Abschnitt 7.2 fasst anschließend das Prozedere des Annotationsvorgangs kurz zusammen. Abschließend zeigt Abschnitt 7.3 die Annotation eines vollständigen Beispieltexts.

7.1 Annotationsschema

7.1.1 Überblick

In der Linguistik greift die Diskussion um den Topik-Begriff je nach vertretenem theoretischen Ansatz auf prosodische, syntaktische, se-

mantische, sowie diskurspragmatische Faktoren zurück; wie diese verschiedenen Faktoren zusammenwirken, ist jedoch noch weitgehend ungeklärt. In den vorliegenden Richtlinien konzentrieren wir uns bewusst auf das Merkmal der *aboutness* und versuchen es für die Behandlung von Texten zu operationalisieren. Beim Annotieren sollen daher keine theoriespezifischen syntaktischen oder semantischen Kriterien angelegt werden, sondern die Richtschnur soll allein die Beschreibung von *aboutness* sein, die im Folgenden und später in Abschnitt 7.1.5 gegeben wird.

Neben dem *aboutness* Topik behandeln wir die nicht so häufig auftretenden „kontrastiven“ sowie die „rahmengebenden“, engl. *frame-setting* Topiks. Alle Arten werden unten genauer erläutert, aber eine erste Beschreibung ist:

- **Aboutness Topik:** die Konstituente, die auf diejenige Entität referiert, zu der die zentrale Prädikation des Satzes (bzw. des Segments)⁴ Information vermittelt. Wenn diese Entität X ist, fungiert der Satz als eine Antwort auf die Frage „Was ist mit X?“ (*What about X?*)
- **Kontrastives Topik:** Es hat eine diskurssteuernde Funktion, indem es nicht allein für den aktuellen Satz wirkt, sondern in Relation zu zwei mehr anderen Elementen in anderen Sätzen steht. Gemeinsam markieren diese einen Textbereich mit gemeinsamem Thema.
- **Frame-setting Topik:** Es setzt den kontextuellen Rahmen, innerhalb dessen die zentrale Aussage des Satzes zu interpretieren ist; oft ist dieser Rahmen eine temporale oder lokationale Information.

Die beiden Kategorien sind nicht exklusiv, d. h., ein Satz kann sowohl ein *aboutness* Topik als auch *frame-setting* Topiks haben. Ein kontrastives Topik kann gleichzeitig *aboutness*- oder *frame-setting* Topik sein; allerdings kann dieselbe Konstituente nicht gleichzeitig als *aboutness* Topik und als *frame-setting* Topik fungieren.

⁴Wir verwenden im Folgenden der Kürze halber nur ‚Satz‘, meinen damit aber ggf. auch kleinere Segmente, wie in Abschnitt 7.1.2 besprochen wird.

(7.109) [*Historisch gesehen*]_{fs}, *haben* [*die Stuarts*]_{ab} *einen legitimeren Anspruch auf die englische Krone.*⁵

Der Satz in Beispiel 7.109 enthält sowohl ein frame-setting Topik (markiert mit *fs*) als auch ein aboutness Topik (*ab*). Das *fs*-Topik *Historisch gesehen* ist ein typisches Beispiel eines satzinitialen Adverbials, das Informationen über den konzeptuellen Rahmen vermittelt, in dem die Aussage über das aboutness Topik *die Stuarts* zu interpretieren ist. Das frame-setting Topik beschränkt die Gültigkeit der Satzaussage: Sie gilt bezogen auf einen bestimmten Ort, auf einen Zeitpunkt oder ein Zeitintervall, oder – wie hier – auf einen konzeptuellen Bereich. Die in Beispiel 7.109 ausgedrückte Behauptung ist, dass die Stuarts einen größeren Anspruch (als eine oder mehrere hier nicht erwähnte andere Familien) auf die englische Krone haben, wenn die Frage aus der Perspektive des historischen Kontexts betrachtet wird; der Autor legt sich nicht fest, wie die Antwort ausfiele, wenn stattdessen beispielsweise eine juristische Bewertung vorgenommen wird, oder eine aus der Perspektive der aktuellen Politik.

Die meisten frame-setting Topiks sind Adverbiale, doch für die Annotation gibt es keine Beschränkung auf bestimmte syntaktische Konstituenten. Dasselbe gilt für kontrastive Topiks, die sich auf einen gemeinsamen Oberbegriff im Text beziehen und untereinander in einer Kontrast-Relation stehen. Wir markieren die Oberbegriff-Konstituente mit einem numerischen Index und das kontrastive Topik mit einem Zeiger auf diesen Index. Ein Beispiel:

(7.110) *Im Zoo gab es allerlei* [*interessante Tiere*]₁. *Zunächst besuchten wir* [*die Tiger*]_{co>1}, *mit denen wir viel Spaß hatten. Sie guckten so drollig. Nachmittags gingen wir* [*zu den Affen*]_{co>1}. *Dort hielten wir uns ...*

Für aboutness Topiks gilt die Einschränkung, dass es nur *nominale referierende Ausdrücke*, wie sie in Kapitel 5 bestimmt wurden, sein können. Für eine Definition von ‚aboutness‘ greifen wir hier

⁵<http://www.spiegel.de/spiegel/print/d-28957473.html> (Zugriff: 30.9.2015)

auf (Jacobs, 2001, S. 645) zurück, der drei Kriterien nennt. Diese sind im Prinzip voneinander unabhängig, doch „prototypische“ aboutness Topiks erfüllen alle drei:

- *Informationelle Trennung*: Das prototypische ab-Topik steht vor der Prädikation, die über es getroffen wird. Die semantische Verarbeitung des Satzes vollzieht sich in zwei Schritten: einer für das Topik, einer für den Rest des Satzes.
- *Prädikation*: Das prototypische ab-Topik füllt eine Position im Valenzrahmen des Prädikats (d. h. es ist nicht „nur“ ein Adjunkt).
- *Adressierung*: Das prototypische ab-Topik referiert auf eine Entität, die als „Adresse“ für die Speicherung von Informationen in der gemeinsamen Wissensbasis (*common ground*) von Autor und Leser fungiert, und die während der inkrementellen Diskursverarbeitung ständig aktualisiert wird.

Das als Adressierung beschriebene Kriterium wird häufig auch mit der „Karteikarten-Metapher“ erläutert: Leser bewegen sich linear durch den Text und tragen die aufgenommenen Informationseinheiten jeweils auf einer geeigneten (mentalen) Karteikarte ein, nämlich derjenigen, die sie für die Entität angelegt haben, um die es in der Informationseinheit geht. Ein illustratives Beispiel dafür wurde von Reinhart (1982) gegeben. Beide Sätze in 7.111 übermitteln quasi dieselbe Information, doch sie wird beim Lesen jeweils auf unterschiedlichen Karteikarten abgelegt:

- (7.111) (i) *Aristotle Onassis married Jacqueline Kennedy.*
(ii) *Jacqueline Kennedy married Aristotle Onassis.*

Zu beachten ist, dass mit den drei Kriterien „prototypische“ ab-Topiks beschrieben werden. Sie können anhand einfacher konstruierter Sätze gut demonstriert werden, doch im authentischen Text fällt die Beurteilung oft schwer. Bei der Topik-Annotation geht es deshalb darum, die im Satz angetroffenen Kandidaten für aboutness Topiks anhand der Kriterien zu bewerten und denjenigen zu wählen, der sie *am besten* erfüllt.

Tabelle 7.2: Attribute für die Topik-Annotation

Schritt 1: Topik-Bereich

Einheit: Diskurssegmente vom Typ HS, HSF, ADV, WEI

Attribut: **TD** Potenzieller „topic domain“

Schritt 2: Thetizität

Einheit: Diskurssegmente vom Typ HS, HSF, ADV, WEI

Attribute:	thetic	Thetische Einheit
	non-thetic	Nicht-thetische Einheit

Typ der Thetizität

Attribute:	thetic-event	Event-basierte thet. Einheit
	thetic-entity	Entität-basierte thet. Einheit

Schritt 3: Topik

Einheit für *ab*: Ref.Ausdruck

Einheit für *fs*, *co*: beliebig

Attribute:	ab	Aboutness Topik: > wovon handelt der Satz
	fs	Frame-setting Topik > Rahmen für die Gültigkeit der Satzaussage
	<i>n</i>	(numerischer Index: 1..) Oberbegriff für kontrastive Topiks > nachfolgend zu behandelndes Thema
	co>n	Kontrastives Topik > Diskursthema, abgeleitet vom Ober- begriff <i>n</i>

Tabelle 7.2 vermittelt einen Überblick über das Topik-Annotationsschema. Nachfolgend (Abschnitt 7.1.2) erläutern wir den ‚Bereich‘, in dem Topiks identifiziert werden sollen (welche linguistische Einheit „hat“ ein Topik?). Dann geht Abschnitt 7.1.3 der Frage nach, ob ein solcher Bereich auch mehrere Topiks haben kann, und Abschnitt 7.1.4 fragt analog, ob auch gar kein Topik vorliegen kann, dies ist das Phänomen ‚Thetizität‘. Die nächsten drei Abschnitte beschreiben im Detail die Kriterien für die Identifikation von aboutness Topiks (7.1.5), frame-setting Topiks (7.1.6) und kontrastiven Topiks (7.1.7).

7.1.2 Der Topik-Bereich

Theoretische Studien zu Topiks arbeiten in aller Regel mit (oft konstruierten) isolierten Sätzen oder sehr kurzen Diskursen, in denen die Topiks jeweils innerhalb eines Hauptsatzes identifiziert werden. Bei der Annotation von authentischem Text haben wir es aber mit ganz unterschiedlichen Arten von Sätzen oder auch Fragmenten zu tun, weshalb es nicht trivial ist, den Bereich zu definieren, innerhalb dessen ein Topik zugewiesen werden soll.

Unsere Definition basiert auf den Segmentierungsrichtlinien von Kapitel 3. Die Faustregel ist, dass wir Hauptsätze und fragmentarische Hauptsätze als Topik-Bereiche ansehen, dazu diejenigen Nebensätze, die semantisch und pragmatisch eher ‚autonom‘ sind, nämlich die Adverbialsätze und die weiterführenden Nebensätze.

Regeln für Topik-Bereiche:

1. Jedes Segment, das mit HS, HSF, ADV oder WEI gekennzeichnet ist, gilt als Topik-Bereich: Es kann Topiks aller drei Typen (ab, fs, co) enthalten.
2. Alle Topiks sind relativ zum kleinsten sie einschließenden Topik-Bereich zu annotieren. So ist beispielsweise in dem Satz [[Während [Peter]_{ab} auf das Ergebnis wartete]_{ADV}, spielte [Tom]_{ab} auf der Gitarre.]_{HS} die Konstituente *Peter* das ab-Topik des eingebetteten ADV Satzes, und *Tom* ist das ab-Topik des kompletten Hauptsatzes.

3. Wenn ein HS/HSF von einem FRE eingeleitet oder von einem FRB gefolgt wird, ist beim Annotieren zu prüfen, ob das Fragment dem HS/HSF Topik-Bereich zugeordnet werden muss: Das ist dann der Fall, wenn sich ein Topik-Kandidat im Fragment befindet, wie in diesen Beispielen:

- [Werner Herzog.]_{FRE} [Wieder einmal ein Großmeister des deutschen Films.]_{HSF}
- [Am Schlagzeug heute wieder:]_{HSF} [Thomas Lang.]_{FRB}

Es folgen Beispiele für verschiedene Fälle der Topik-Bereich (TD) Annotation (die Segment-Typen sind nicht immer mit angegeben). In 7.112 verbindet die Konjunktion „aber“ zwei Hauptsätze (HS). Jeder ist ein eigener Topik-Bereich. Analoges gilt in 7.113 für den mit *weil* eingeleiteten Nebensatz. In 7.114 ist das FRB in den TD des HSF integriert.

(7.112) [Er ist nicht ansehnlich,]_{TD} [aber von Asbest befreit, könnte er auch mit neuer, ansprechender Fassade ein Blickfang werden.]_{TD}
(procon-kreiselcon)

(7.113) [Sollen in Kreuzberg, Friedrichshain oder Prenzlauer Berg die Bäume fallen, [weil sie die damaligen Bauherren nicht für wichtig erachteten?]_{TD}]_{TD}
(procon-bäumecon)

(7.114) [[[Der Meier]_{ab} nun wieder.]_{FRB} [Immer einen Scherz auf den Lippen!]_{HSF}]_{TD}

7.1.3 Multiple Topiks und Ambiguität

In einfachen Fällen liegt ein einzelner Topik-Bereich vor, in dem (höchstens) ein ab-Topik oder ein fs-Topik zu markieren sind:

(7.115) [[Die Zahl der Straßenbäume]_{ab} sinkt stetig.]_{TD}
(procon-bäumecon)

Generell gilt jedoch, dass ein TD alle drei Arten von Topiks beherbergen kann. Wir wiederholen hier noch einmal Beispiel 7.110, nunmehr vollständig mit Segmenttypen, TDs und Topiks annotiert:

- (7.116) [[Im Zoo]_{fs} gab es allerlei [interessante Tiere]₁]_{HS-TD,thetic-entity}.
[Zunächst besuchten [wir]_{ab} [die Tiger]_{co>1}]_{HS-TD}, [mit denen wir
viel Spaß hatten]_{ANR}.
[[Sie]_{ab} guckten so drollig]_{HS-TD}.
[[Nachmittags]_{fs} gingen [wir]_{ab} [zu den Affen]_{co>1}]_{HS-TD}.
[Dort hielten [wir]_{ab} uns . . .]

Hier sehen wir TDs mit zwei und sogar drei verschiedenen Topiks.

Manchmal ist die Topik-Entscheidung nicht einfach. Gibt es innerhalb eines TD mehr als einen plausiblen Kandidaten, so liegt Topik-Ambiguität vor. Dies geschieht vorwiegend bei ab-Topiks, da Sätze durchaus mehrere Konstituenten haben können, die das *aboutness* Eigenschaften zu erfüllen scheinen. Dann können mehrere Kandidaten markiert werden, zwischen denen jedoch eine Rangfolge anzugeben ist („erfüllt die Kriterien besser“, wie oben besprochen). Im obigen Beispiel 7.116 ist das in dem weiterführenden Nebensatz bereits zu sehen. Hier ein Beispiel aus dem Korpus:

- (7.117) [[Die Zahl der Straßenbäume]_{ab} sinkt stetig,]_{TD} [und [die
abgeschwächte
Baumschutzverordnung]_{ab} wird [diesen Prozess]_{ab-2} noch
beschleunigen.]_{TD}
(procon-bäumecon)

In 7.117 ist der erste Satz einfach zu bearbeiten, da *die Zahl der Straßenbäume* der einzige und klare Kandidat für das ab-Topik ist. Hingegen gibt es im zweiten Satz zwei Kandidaten: Sowohl *die abgeschwächte Baumschutzverordnung* als auch *diesen Prozess* sind Entitäten, „über die etwas gesagt wird“. In diesem Fall ist *die Baumschutzverordnung* das Subjekt des Satzes und das satzinitiale Element, weshalb die Annotation es als besseren Kandidaten ausweist als den *Prozess*. Die Notationskonvention ist, den ersten Kandidaten wie gewohnt als *ab* zu markieren und den weniger plausiblen Kandidaten als *ab-2*. Im (recht unwahrscheinlichen) Fall, dass es noch einen dritten plausiblen Kandidaten gibt, erhält er die Annotation *ab-3*.

Eine in unserer Textsorte relativ häufige Quelle der Ambiguität ist das Auftreten des Pronomens *wir* als vager referierender Ausdruck, der im Sinne von „wir alle“ gebraucht wird. Gibt es innerhalb eines TD mehrere ab-Kandidaten, so sollen diese Pronomen in der Rangfolge nachgeordnet werden.

(7.118) *Die Kunstschätze sind in einem schlechten Zustand. Wir müssen sie restaurieren.*
(nach procon-bäumeopro)

Gemäß der genannten Regel ist der zweite Satz nicht eine Aussage über „uns“ sondern über die Kunstschätze (obwohl *wir* satzinitial auftritt). Die Annotation lautet also: [*Wir*]_{ab-2} *müssen* [*sie*]_{ab} *restaurieren*.

Sollte Ambiguität bei den anderen Topiks auftreten, wird dasselbe Schema benutzt: *fs* und *fs-2*, bzw. *co* und *co-2*.

7.1.4 Segmente ohne Aboutness Topik

Wir wenden uns nun der Frage zu, unter welchen Umständen Sätze (und andere Segmente) *kein* Topik haben. Allgemein gilt, dass *fs*-Topiks und *co*-Topiks relativ selten sind, während die meisten Sätze ein *ab*-Topik haben. Für manche Topik-Bereiche allerdings bietet sich kein *ab*-Kandidat an; diese können aufgefasst werden als Antworten auf die allgemeine Frage „Was ist passiert?“ oder „Was gibt’s neues?“

Sätze ohne *ab*-Topik und ohne *co*-Topik nennen wir „thetisch“ (in der Literatur manchmal auch „all-new“ oder „presentational“) und solche mit einem dieser beiden Topiks entsprechend „nicht-thetisch“. Daraus folgt, dass thetische Sätze durchaus ein *fs*-Topik haben können: Dann gibt es einen Rahmen, aber keine klare *aboutness* Entität.

Thetische Sätze treten in zwei Varianten auf: Die wesentliche Funktion eines *entity-central* thetischen Satzes (Beispiel 7.119) besteht in der Einführung einer neuen Entität (Person, Gegenstand, Institution, ...) in den Diskurs. Häufig, aber nicht immer, stellt diese

dann im nachfolgenden Satz das ab-Topik dar. Demgegenüber legt ein *event-central* thetischer Satz (Beispiel 7.120) das Gewicht weniger auf die Präsentation einer neuen Entität als auf die Darstellung einer kompletten Situation bzw. eines Sachverhalts.

(7.119) (Erster Satz eines Zeitungsartikels)

*[Detectives investigating phone hacking at News International have arrested a 31-year-old woman.]*_{TD,thetic-entity} *[The woman is believed to be ...]*_{TD,non-thetic}

(7.120) (Erster Satz eines Zeitungsartikels, procon-impfenpro)

*[Kein Kind weiß heute noch, was Pocken sind.]*_{TD,thetic-event}

Während dies relativ klare Fälle sind, ist die Entscheidung zwischen *entity-central* und *event-central* häufig nicht leicht zu treffen. Ggf. kann bei der Annotation darauf verzichtet werden, dann ist lediglich das allgemeine Attribut **thetic** zuzuweisen.

Die „all-new“ Eigenschaft von thetischen Sätzen ist im Sinne von „Hörer-neu“ zu verstehen, nicht als „Diskurs-neu“. Das bedeutet, dass eine Entität, die noch nicht im Text erwähnt wurde, durchaus ein aboutness Topik sein kann. Beispiel 7.121 ist der erste Satz eines Artikels in einer Zeitung, die in Berlin erscheint; die Stadt ist also den Lesern wohlbekannt, und der Satz trifft eine Aussage über Berlin – er ist also nicht thetisch.

(7.121) *Berlin ist dabei, seinen Ruf als eine der grünsten Metropolen der Welt zu verspielen.*

(procon-bäumecon)

Generell sind allerdings die texteinleitenden Sätze die „natürlichen“ Kandidaten für Thetizität. Im laufenden Text ist sie selten, kann aber auftreten, wenn beispielsweise das Thema wechselt, ein neues Argument präsentiert wird, eine neue Entität in den Diskurs eingeführt wird. Solche thematischen Verschiebungen sind im Text mitunter durch lexikalische Marker (kontrastive oder additive Konnektoren; Formeln wie *wie dem auch sei*) signalisiert. Zur Diagnose, ob der Satz thetisch ist, sind solche Signale zu entfernen und dann ist zu beurteilen, ob er eine plausible Antwort auf die Frage *Was*

gibt's Neues? / Was ist passiert? darstellt und gleichzeitig die Tests für ein ab-Topik (s. u.) keinen klaren Kandidaten dafür liefern.

(7.122) *Der Steglitzer Kreisel muss auf jeden Fall von Asbest befreit werden. Das kostet, der Abriss kostet, und der Neubau – wo immer er auch stehen mag – kostet erst recht. Dann entsteht vielleicht eine verödete Fläche an der Schloßstraße.*
(procon-kreiselcon)

In Beispiel 7.122 ist der durch den Konnektor *Dann* eingeleitete Satz thetisch, obwohl er keineswegs am Textanfang steht. Er präsentiert die Entität *eine verödete Fläche*, von der bisher nicht die Rede war, und stellt diese als Konsequenz des vorangegangenen Geschehens in den Raum. Zur Diagnose der Thetizität lassen wir wie gesagt den Konnektor beiseite und wenden den Fragetest an: *Was gibt's Neues? – Vielleicht entsteht eine verödete Fläche an der Schloßstraße.*

7.1.5 Identifikation von Aboutness Topiks (ab)

Aboutness Topiks werden im Text inkrementell annotiert: Die Entscheidung über das ab-Topik eines Satzes (oder anderen Bereichs) ist vor dem Hintergrund des vorangegangenen Kontexts zu treffen, aber *ohne* Betrachtung der nachfolgenden Sätze. Wie bereits erläutert ist das ab-Topik des Satzes diejenige Entität, über die eine Aussage getroffen wird, bzw. auf deren „Karteikarte“ im Verarbeitungsmodell des Diskursgedächtnisses (*common ground*) eine Eintragung vorzunehmen wäre. Ein ab-Topik muss daher für den Leser identifizierbar sein (entweder im Text vorerwähnt oder per Welt-Wissen bekannt). Die Einheiten, die ein ab-Topik bekommen können, wurden in Abschnitt 7.1.2 genannt; Kandidaten für den Topik-Ausdruck selbst sind ausschließlich nominale referierende Ausdrücke (NRA).

Zur Erleichterung der Entscheidung geben wir im Folgenden einige Paraphrasentests für die Identifikation des ab-Topiks an.

Test für Aboutness Topiks

Ein NRA X ist das ab-Topic eines Topik-Bereichs S , der X enthält, wenn

- S (mit Ersetzung von X durch ein geeignetes Pronomen) eine adäquate Antwort auf die Frage ‚Was ist mit X ?‘ wäre.

Beispiel:

Gestern hat meine Tochter ihr Abitur bestanden.

Was ist mit Deiner Tochter? – Gestern hat sie ihr Abitur bestanden.

- S kann in diesen intuitiv gleichbedeutenden Satz umgeformt werden:

Was X betrifft, S'

wo S' sich von S nur dadurch unterscheidet, dass X durch ein geeignetes Pronomen ersetzt wurde.

Beispiel:

Was meine Tochter betrifft, gestern hat sie ihr Abitur bestanden.

- S kann in die intuitiv gleichbedeutende Linksversetzungsstruktur umgeformt werden:

X, S'

wo S' sich von S nur dadurch unterscheidet, dass X durch ein geeignetes Personal- oder Demonstrativpronomen ersetzt wurde.

Beispiel:

Meine Tochter, die hat gestern ihre Abiturprüfung bestanden.

Die Paraphrase-Tests sind schwierig anzuwenden, wenn die Satzstruktur komplex ist und beispielsweise eingebettete Nebensätze zu beurteilen sind. In diesen Fällen bitte die in Abschnitt 7.1.1 genannten Kriterien direkt anwenden. Des weiteren kann (stille) Prosodie manchmal bei der Entscheidung helfen. Wir betrachten

noch einmal das oben genannte Beispiel 7.120. Wurden *Pocken* im Text bisher nicht erwähnt, wird der Satz typischerweise so gelesen: *Kein Kind weiß heute noch, was /Pocken sind*. Dies spricht für einen thetischen Satz. In einem Kontext, wo die *Pocken* bereits vorerwähnt sind, ist diese Prosodie angemessener: *Kein /Kind weiß heute noch, was Pocken /sind*. Dann wird es als Information über *Pocken* verstanden, also als Satz mit einem ab-Topik.

Gelegentlich ist es schwierig, eine Entscheidung unter möglichen Kandidaten für ein ab-Topik zu treffen. Wenn die Wahl schwer fällt, sollten alle Kandidaten annotiert werden, wie in Abschnitt 7.1.3 beschrieben. Es muss jedoch auch dann eine Entscheidung über die Rangfolge getroffen werden, also: welcher Kandidat erfüllt die ab-Topik Kriterien *am besten*? Dies wird dann durch die Zuweisung der Indizes markiert.

Es folgen einige Beispiele im Kontext, anhand derer wir die Anwendung der Tests zeigen.

(7.123) *Gestern traf ich Peter und Anne in Golm. [Peter]_{ab} trug rote Socken.*

Test: *Gestern traf ich Peter und Anne in Golm. Was Peter betrifft, er trug rote Socken.*

Wie eingangs festgestellt, stehen Topiks „prototypisch“ am Satzanfang. Dies muss aber nicht immer der Fall sein, zum Beispiel wenn die Satzstruktur komplex ist.

(7.124) (Text über die Wichtigkeit von Impfungen für Kinder)
Nur wer impfen lässt, hilft mit, dass [Impfungen]_{ab} eines Tages überflüssig werden.

Test: *Was Impfungen betrifft, nur wer impfen lässt, hilft mit, dass sie eines Tages überflüssig werden.*

(7.125) *Gestern haben wir die Mathearbeit zurück bekommen. Der große Frust setzte ein.*

Test: ??*Gestern haben wir die Mathearbeit zurück bekommen. Was den großen Frust betrifft, er setzte ein.*

Test: ??*Gestern haben wir die Mathearbeit zurück bekommen. Der große Frust, der setzte ein.*

In 7.125 könnte *der große Frust* auf den ersten Blick ein Kandidat für ein ab-Topik sein, doch anhand der Tests muss diese Hypothese verworfen werden. Ein Grund ist, dass *der große Frust* nicht das Kriterium der Adressierung erfüllt: Wir bekommen in dem Satz nicht etwa Information über Frust, sondern über *uns*, die Schüler der Klasse, die aber nicht ausgedrückt sind. Der Satz ist somit thetisch.

Aboutness Topik Ausdrücke müssen nicht immer (wie in den obigen Beispielen) definite/spezifische NPn sein. Für andere NPn kann in den Tests die extrahierte NP zu einer Plural-NP umgeformt werden. Hier ist ein Beispiel mit einer indefiniten generischen NP:

(7.126) *Ein Hund ist oft intelligent.*

Test: *Hunde, die sind oft intelligent.*

Eine Komplikation kann sich bei elidierten Konstituenten ergeben, etwa bei fehlenden Subjekten.

(7.127) *[[Aufgrund der dramatischen Kassenlage hat sie eine seit mehr als einem Jahr erarbeitete Kabinettsvorlage überraschend auf Eis gelegt]_{HS} [und vorgeschlagen, erst 2003 darüber zu entscheiden.]_{HSF}]_{HS}*

Wäre im HSF das Subjekt nicht elidiert, sondern z. B. pronominalisiert (*und sie hat vorgeschlagen, ...*), so wäre *sie* offensichtlich als ab-Topik zu markieren. Andere Kandidaten liegen hier nicht vor. Wir behandeln diese Fälle, indem das HSF *nicht* als thetisch markiert wird, aber auch *kein* ab-Topik erhält. Diese eigentlich unzulässige Kombination ist somit reserviert für Fälle, in denen eine elidierte Konstituente das ab-Topik „wäre“.

7.1.6 Identifikation von Frame-Setting Topiks (fs)

Frame-setting Topiks definieren den Bezugsrahmen, innerhalb dessen die im TD formulierte zentrale Aussage zu interpretieren ist. Im Gegensatz zu ab-Topiks gibt es niemals eine Prädikations-Relation zwischen dem fs-Topik und der einbettenden clause (das fs-Topik füllt keine Position im Valenzrahmen des Verbs).

Der Bezugsrahmen ist oft eine Zeit- oder Ortsangabe für die zentrale Aussage. Daher sind die typischen Arten von fs-Topik-Konstituenten temporale oder lokationale PPn, Adverbialphrasen sowie subordinierte Sätze, die Zeit- oder Ortsinformation vermitteln.

(7.128) *[In Berlin]_{fs} haben die Verhandlungspartner eine Regel nicht beachtet.*

Aber: Nicht *jede* solche Konstituente ist ein fs-Topik. Auch für fs-Topiks gilt das Kriterium der informationellen Trennung (s. Abschnitt 7.1.1) und daher stehen sie typischerweise am Satzanfang. Darüber hinaus gilt, dass die ausgedrückten Orts- oder Zeitangaben oft bereits Teil des gemeinsamen Wissens („common ground“) von Sprecher/ Autor und Adressat sind, oder aber sie können leicht daraus erschlossen werden. So wird Beispiel 7.129 am ehesten in einem Kontext auftreten, wo soeben der Zeitrahmen einer anderen Handlung besprochen wurde; dann ist *Um 8* ein klares fs-Topik. In 7.130 hingegen ist das temporale Adverbial ins Mittelfeld des Satzes integriert und liefert lediglich eine Nebeninformation.

(7.129) *[Um 8]_{fs} bin ich dann zum Frühstück gegangen.*

(7.130) *Ich bin dann um 8 zum Frühstück gegangen.*

Neben Zeit- und Ortsangaben können aber auch andere vorangestellte Adverbiale als fs-Topiks fungieren, insoweit sie Domänen oder Aspekte ausdrücken, innerhalb derer die nachfolgende Aussage zu interpretieren ist. Eine Konstruktion, um die Adverbiale entsprechend zu testen, ist eine Umformung zu *Was X angeht*:

(7.131) *[Finanziell]_{fs} sind wir in ziemlich guter Verfassung.*

Test: *Was die Finanzen angeht, sind wir in ziemlich guter Verfassung.*

(7.132) *Schnell hat sie sich noch einen Regenschirm gekauft*

Test: *??Was Schnelligkeit angeht, hat sie sich noch einen Regenschirm gekauft.*

7.1.7 Identifikation von kontrastiven Topiks (co)

Eine gängige semantische Charakterisierung des kontrastiven Topiks besteht darin, dass es nurmehr eine *partielle* Antwort auf eine vorangestellte Frage gebe:

(7.133) *Was machen [Deine Schwestern]₁?*
[Meine ältere Schwester]_{co>1} studiert Biologie. (...)

Im monologischen Text sind solche (rhetorischen) Frage-/Antwort-Szenarien sehr selten, doch kann ein Satz eine Ankündigung der Behandlung von Unter-Themen beinhalten und damit implizit eine entsprechende Frage aufwerfen. Im Zoo-Beispiel (7.110) löst der erste Satz die Frage *Welche Tiere denn?* aus, die dann vom weiteren Textverlauf in mehreren Schritten beantwortet wird. Den maßgeblichen Ausdruck haben wir dort als ‚Oberbegriff‘ bezeichnet, auch: *Hyperonym*. Die Relation zwischen *Tier* und *Tiger* ist also die der *Hyponymie*. Im obigen Beispiel 7.133 hingegen ist die Relation die der Mitgliedschaft in einer Menge. Ein häufiger dritter Fall ist die Teil-Ganzes Relation (*Meronymie*), die sich sowohl für Entitäten (7.134) als auch für Sachverhalte (7.135) definieren lässt.

(7.134) *[Die Sinfonie]₁ hat mich sehr beeindruckt. [Der Sonatensatz]_{co>1} war leicht und luftig, (...), und [das Finale]_{co>1} verströmte große Feierlichkeit.*

(7.135) *[Der heutige Staatsbesuch des polnischen Ministerpräsidenten]₁ verlief kompliziert. [Bei der Ankunft auf dem Flughafen]_{co>1} regnete es in Strömen. (...) Auch [der Empfang des Bundespräsidenten]_{co>1} zeichnete sich durch verschiedene Kuriositäten aus. (...)*

Entscheidend für das Vorliegen von kontrastiven Topiks ist also, dass eine der genannten Relationen zwischen einem (allgemein) „Oberbegriff“ und mindestens zwei (allgemein) „Unterbegriffen“ besteht. Die Unterbegriffe stehen dann untereinander in einer Kontrast-Relation; jede Aussage über einen der Unterbegriffe stellt eine partielle Fortführung des durch den Satz mit dem Oberbegriff

beschriebenen Sachverhalts dar (oder: ist eine partielle Antwort auf die implizit aufgeworfene Frage). Bei der Annotation erhält der Oberbegriff einen numerischen Index, auf den alle Unterbegriffe verweisen.

Die partiellen Antworten können sich durchaus jeweils über mehrere Sätze erstrecken, und in diesem Sinne kommt den kontrastiven Topiks eine diskursstrukturierende Funktion zu: Ein durch den Satz mit dem Oberbegriff eingeleiteter Teil des Texts wird thematisch untergliedert.

Kontrastive Topiks sind häufig gleichzeitig die ab-Topiks ihrer Sätze/Segmente. Das ist in allen oben gezeigten Beispielen 7.133–7.135 der Fall. Für die vollständige Annotation fügen wir das ab-Subskript hinzu:

[Der Sonatensatz]_{ab,co>1} war leicht und luftig.

7.2 Annotationsvorgehen

Benutze für die Annotation die in Tabelle 7.2 zusammengefassten Attribute.

Bearbeite den Text inkrementell, von Diskurssegment zu Diskurssegment vorangehend. Markiere jedes Segment vom Typ HS, HSF, ADV und WEI als Topik-Bereich (s. Abschnitt 7.1.2) mit dem Index TD. Für die Topik-Annotation ist das Segment nun vor dem Hintergrund des verstandenen vorangehenden Kontexts zu betrachten. Der nachfolgende Kontext hingegen ist ausschließlich im Hinblick auf das Vorliegen von kontrastiven Topiks zu betrachten; für ab- und fs-Entscheidungen soll er keine Rolle spielen.

Wende folgendes Prozedere auf jeden TD an:

1. Steht vor dem TD-Segment ein FRE, steht nach ihm ein FRB? Wenn ja, prüfe ob sie in den TD integriert werden müssen, weil sie Topik-Kandidaten enthalten (s. Abschnitt 7.1.2). Weiter mit Schritt 2.
2. Ist der TD thetisch („all new“)? D. h., wenn der TD vor dem Hintergrund des vorangehenden Kontexts verstanden wird,

antwortet er dann eher auf die Frage *Was gibt's Neues / Was ist passiert* als auf die Frage *Was ist mit X* (wo X ein referierender Ausdruck mit identifizierbarem Referenten ist)? (Näheres im Abschnitt 7.1.4).

- Ja: Markiere den TD als **thetic**. Prüfe, ob sich entscheiden lässt, ob **entity** oder **event** Thetizität vorliegt; wenn ja, benutze die speziellere Markierung **thetic-entity** bzw. **thetic-event**. Gehe zu Schritt 5.
 - Nein: Markiere den TD als **non-thetic** und gehe zu Schritt 3.
3. Führe die aboutness Topik Probe für alle referierenden Ausdrücke des TD durch (s. Abschnitte 7.1.1 und 7.1.5). Wieviele Kandidaten gibt es?
- 0: Fall 1: Es liegt ein elidiertes ab-Topik vor (vgl. Ende von Abschnitt 7.1.5). Dann wird keines markiert, weiter zu Schritt 5.
Fall 2: Kein elidiertes ab-Topik. Da der TD als non-thetic markiert ist, stimmt etwas nicht. Jedes nicht-thetische Segment benötigt mindestens ein ab-Topik. Gehe noch einmal zu Schritt 2.
 - 1: Markiere den Kandidaten als **ab** (anfügend, falls er bereits eine **co**-Markierung trägt) und gehe zu Schritt 5.
 - 2: Gehe zu Schritt 4
4. Erstelle eine Rangfolge der ab-Topik Kandidaten. Prüfe, welcher Kandidat am ehesten die Kriterien erfüllt und markiere ihn als **ab**, den nächsten als **ab-2**, und falls es noch mehr gibt, mit entsprechenden Zahlen. Sollte ein Kandidat bereits eine **co**-Markierung tragen, wird einfach angefügt.
- Für ineinander geschachtelte TDs gilt: Die Auszeichnung eines *ab* bezieht sich per Konvention immer auf den kleinsten umgebenden TD. Darum ist keine Ko-Indizierung zwischen TDs und **abs** erforderlich.
- Weiter mit Schritt 5.

5. Prüfe ob der TD eines oder mehrere fs-Topiks enthält. (Näheres in Abschnitt 7.1.6.) Beachte, dass ein Ausdruck, der bereits als **ab** markiert ist, nicht gleichzeitig auch ein **fs** sein kann. Gibt es mehrere Kandidaten, nimm eine Nummerierung analog zu der von Schritt 4 vor. Im Falle geschachtelter TDs erfolgt die Zuordnung so wie in Schritt 4 für **abs** beschrieben. Weiter mit Schritt 6.
6. Prüfe ob der TD einen Ausdruck enthält, der eine „Oberbegriff“ Rolle für den nachfolgenden Kontext spielt: Es müssen mindestens zwei „Unterbegriffe“ in unterschiedlichen Segmenten folgen, die wir dann als kontrastive Topiks marieren. (Näheres in Abschnitt 7.1.7.) Liegt eine solche Situation vor, markiere den Oberbegriff mit einem numerischen Index n und die Unterbegriffe mit $co>n$. Die Markierung von co -Topiks ist unabhängig von den anderen Kategorien, d. h. ein co -Topik sowie der „Oberbegriff“ kann bereits als ab -Topik oder fs -Topik markiert sein (muss aber nicht).

7.3 Analyse eines Beispieltexsts

Der Übersichtlichkeit halber beginnt ein TD immer in einer neuen Zeile. *Thetic* ist hier als *th* abgekürzt und *non-thetic* wird nicht verwendet. Jeder TD ohne *th* ist also implizit als *non-thetic* zu verstehen (und hat dementsprechend ein oder mehr ab -Topiks).

Stadtkern aufwerten

[[Im Streit zwischen der Stadt Wittstock und dem Landesstraßenbauamt Kyritz]_{fs} hat Bürgermeister Lutz Scheidemann einen klaren Sieg errungen.]_{TD-th-event}

[[Wittstock]_{ab} wird in einigen Jahren frei von Lkw-Verkehr sein,]_{TD}
[eine Umgehung entlastet [die Innenstadt]_{ab}.]_{TD}

[Klirrende Gläser in den Wohnzimmerschränken sind dann Geschichte.]_{TD-th-event}

[Wer sein Haus bereits saniert hat, muss auch nicht mehr befürchten, dass der Putz von der Fassade fällt.]_{TD-th-event}

[Mit der Einigung kommt aber auf [die Wittstocker]_{ab} eine große Herausforderung zu.]_{TD}

[[Sie]_{ab} sind aufgefordert, über [neue, sinnvolle Verkehrsabläufe]₁ in der Innenstadt nachzudenken.]_{TD}

[[Die Entlastungsstraße für Lkw]_{co>1} führt zwar um die Altstadt herum,]_{TD}

[doch [eine autofreie Stadt]_{co>1} wird es nicht geben.]_{TD}

[So braucht sich auch kein Geschäftsmann vor einer leeren Altstadt zu fürchten.]_{TD-th-event}

[Gut beraten ist allerdings, wer sich jetzt ernsthafte Gedanken um kundenfreundliche Parkmöglichkeiten macht.]_{TD-th-entity}

[Auch [Flaniermeilen in der Wittstocker Altstadt]_{co>1} sind gefragt.]_{TD}

[Denn ohne Lkw-Verkehr wird es ruhiger werden,]_{TD-th-event}

[[der Stadtbummel]_{ab} kann wieder Spaß machen.]_{TD}

[Deshalb müssen neue Ideen her, um [die Wittstocker Altstadt]_{ab} lebenswerter zu machen.]_{TD}

(maz-1679)

Der erste TD hat ein relativ klares fs-Topik, weil der *Sieg* sich auf den besagten Streit bezieht. Weniger klar ist, ob *Bürgermeister L. S.* ein ab-Topik darstellen sollte. Einerseits sprechen die Frage-/Umformungstests dafür, andererseits ist aber auch das Kriterium der Identifizierbarkeit zu prüfen. Das ist in diesem Fall eine

subjektive Entscheidung: Einwohner von Wittstock werden den Bürgermeister vermutlich kennen, viele andere Leser jedoch nicht. Letzteres war hier ausschlaggebend für die Entscheidung thetic-event. Im zweiten TD hingegen ist Wittstock bereits vorerwähnt und auf jeden Fall identifizierbar, analog dann die Innenstadt in TD 3. TD 4 und 5 nehmen keinen bekannten Diskursreferenten wieder auf, sondern schildern Beispielergebnisse für die Zukunft, anschließend werden *die Wittstocker* explizit erwähnt und bilden damit das ab-Topik von TD 6 und auch TD 7. Dort findet sich auch der abstrakte Ausdruck *neue, sinnvolle Verkehrsabläufe*, der nachfolgend mit konkreten Beispielen illustriert werden wird, er bildet damit einen Oberbegriff für drei kontrastive Topiks: die LKW-Entlastungsstraße, die Vorstellung einer autofreien Stadt und Flaniermeilen in der Altstadt. Alle drei sind für den nicht-Wittstocker eher nicht identifizierbar (dies ist wiederum eine subjektive Entscheidung), auf jeden Fall im Text nicht vorerwähnt, und daher hier nicht zusätzlich als ab-Topik markiert. Die beiden eingestreuten Sätze beginnend mit *So braucht* enthalten keine ab-Topik Kandidaten; der erste präsentiert einen event, der zweite stellt spezifisch eine Gruppe von (hypothetischen) Akteuren in den Raum (daher th-entity), die durchaus im Folgesatz weiter thematisiert werden könnten, was jedoch nicht geschieht. Der drittletzte TD handelt nicht vom Lkw-Verkehr, sondern von der Ruhe in der Stadt, der aber kein refrierender Ausdruck entspricht, daher th-event. Anschließend betrachten wir den *Stadtbummel* als identifizierbar (per bridging zu *Flaniermeilen*) und machen ihn daher zum ab-Topik, denn auch die Paraphrasentests fallen hier positiv aus: Der Stadtbummel, der kann wieder Spaß machen. Der Schlusssatz handelt dann nochmals allgemein von der *Wittstocker Altstadt*, wobei wegen der Einbettung der Paraphrasentest ein wenig unnatürlich klingt, aber noch akzeptabel ist: Die Wittstocker Altstadt: Neue Ideen müssen her, um sie lebenswerter zu machen.

Konnektoren und Argumente

Manfred Stede

Hinweise:

- Diese Richtlinien sind eine Überarbeitung und Erweiterung des Entwurfs von Annika Neumann und Andreas Wehr aus dem Jahr 2005.
- Die vorliegenden Richtlinien verwenden eine Notation, wie sie „auf Papier“ oder im Text-Editor verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.
- Alle in diesen Richtlinien genannten Textbeispiele ohne Quellenangabe sind fiktiv.

8.1 Hintergrund und Ziel der Annotation

Die Annotation zielt auf deutschsprachige Konnektoren und die Textspannen, die sie jeweils verbinden; diese nennen wir im Folgenden die *Argumente* der Konnektoren. Die zugrunde liegende Annahme ist, dass Konnektoren als (syntaktisch heterogene) Wortklasse grundsätzlich *semantisch zweistellig* sind, also zwei mehr oder weniger kurze Textspannen verbinden. Damit folgt der Ansatz im Großen und Ganzen dem Vorgehen der *Penn Discourse Treebank* (Prasad u. a., 2008) bei der Behandlung von „expliziten Konnektoren“.

Für einen Text ist die Annotation abgeschlossen, wenn jeder Konnektor identifiziert und seine Argumente markiert sind. Die Entscheidungen über Argument-Zuweisungen zu einem Konnektor sind stets unabhängig von anderen Konnektoren des Textes zu treffen. Es wird also insbesondere mit dieser Annotation keinerlei

Struktur aufgebaut; die Annotation ist vielmehr „flach“, weil jeder Konnektor individuell behandelt wird.

Abschnitt 8.2 erläutert zunächst den Begriff des Konnektors und die verschiedenen syntaktischen Typen. Dann beschäftigt sich 8.3 mit den Argumenten von Konnektoren. Konkrete Hinweise zum Annotationsvorgehen liefert Abschnitt 8.4; dies umfasst einerseits die Desambiguierung (d. h. die Feststellung, ob ein Kandidatenwort im jeweiligen Kontext wirklich als Konnektor gebraucht wird) und andererseits Hilfestellung für die Entscheidung über die Ausdehnung der Argumente.

8.2 Konnektoren

In diesem Abschnitt werden die verschiedenen Typen von Konnektoren zunächst informell vorgestellt, um die Aufgabenstellung für die Annotation zu verdeutlichen. Konnektoren lassen sich aufgrund ihrer syntaktischen Eigenschaften zunächst in vier Gruppen einteilen: Subordinierende Konjunktionen, koordinierende Konjunktionen, Adverbiale und Präpositionen. Zusätzlich behandeln wir mehrteilige, diskontinuierliche Konnektoren als eine separate Gruppe. Sie stellen zwar keine eigene syntaktische Kategorie dar, haben aber einige Besonderheiten. Manche Konnektoren sind mehr als einer Kategorie zugeordnet.

Unsere Unterteilung der Konnektoren dient als grobe Klassifikation – auf weitergehende syntaktische Merkmale und Untergruppierungen gehen wir nicht ein. Für jede unserer Gruppen geben wir eine Liste von Konnektoren an; die wesentliche Informationsquelle dafür ist das *Handbuch der deutschen Konnektoren* (Pasch u. a., 2003). Die Listen dienen aber nur der Orientierung und erheben aber keinen Anspruch auf Vollständigkeit! Auch ist nicht berücksichtigt, dass einige Konnektoren in mehreren Schreibweisen auftreten können (z. B. *so dass*, *sodass*, *so daß*, *sodaß*).

8.2.1 Subordinierende Konjunktionen

Subordinierende Konjunktionen verbinden zwei Sätze, von denen einer ein Hauptsatz und der andere ein Nebensatz ist. Der Konnektor befindet sich stets am Anfang des Nebensatzes. Bei den meisten Konnektoren sind beide Abfolgen von Hauptsatz und Nebensatz möglich, vgl. Bsp. 8.136 und 8.137.

In allen nachfolgenden Beispielen werden Konnektoren gefettet dargestellt.

(8.136) **Da** er krank war, blieb er daheim.

Er blieb daheim, **da** er krank war.

(8.137) **Obwohl** es regnet, gehe ich in den Park.

Ich gehe in den Park, **obwohl** es regnet.

Liste subordinierender Konjunktionen:

als; anderenfalls; anstatt dass; aufgrunddessen, dass; außer dass; außer wenn; bevor; bevor nicht; da; dadurch, dass; dafür, dass; falls; indem; insofern; insofern als; nachdem; nur dass; obgleich; obschon; obwohl; obzwar; seit; seitdem; so; sobald; so dass; sofern; soweit; statt; vorausgesetzt, dass; während; weil; wenn; wenn auch; wenngleich; weshalb; weswegen; wie; wogegen; wohingegen; womit; wonach; worauf; woraufhin; zumal

8.2.2 Koordinierende Konjunktionen

Werden zwei Ausdrücke durch eine koordinierende Konjunktion verbunden, so stehen beide auf derselben hierarchischen Stufe. Nicht nur komplette Sätze können koordiniert werden (vgl. 8.138), sondern auch Phrasen (vgl. 8.139).

(8.138) Du kannst ihm schreiben, **oder** du kannst ihn auch anrufen.

(8.139) Er kaufte sich vom Lottogewinn kein Auto, **sondern** machte mehrere Auslandsreisen.

Liste koordinierender Konjunktionen:

aber; das heißt; oder; sondern; und

8.2.3 Adverbiale

Die größte Gruppe der Konnektoren bilden die Adverbiale. Sie können satzinitial (8.140), final (8.141) oder medial (8.142) auftreten. Das eine Argument eines adverbialen Konnektors ist meist der Satz, in dem er auftaucht, das andere Argument steht meist im Satz (oder den Sätzen) davor.

(8.140) Gestern habe ich zuviel Wein getrunken. **Dadurch** habe ich heute furchtbare Kopfschmerzen.

(8.141) Gestern habe ich zuviel Wein getrunken. Heute habe ich furchtbare Kopfschmerzen **dadurch**.

(8.142) Gestern habe ich zuviel Wein getrunken. Heute habe ich **dadurch** furchtbare Kopfschmerzen.

Liste adverbialer Konnektoren:

aber; allein; allenfalls; allerdings; also; andererseits; anderenfalls; ansonsten; anstelle dessen; auch; aufgrund dessen; außer; außerdem; beispielsweise; bzw.; da; dabei; dadurch; dafür; dagegen; daher; danach; dann; darauf; daraufhin; darüber hinaus; darum; demgegenüber; denn; dennoch; des Weiteren; desgleichen; deshalb; dessen ungeachtet; deswegen; doch; ebenfalls; ebensowenig; ergo; ferner; folglich; gleichwohl; hierauf; hierdurch; hiernach; hingegen; infolge dessen; insofern; insofern als; inwiefern; inwieweit; inzwischen; jedoch; mithin; nämlich; nebenher; nichtsdestoweniger; nichtsdestotrotz; ob; obendrein; ohnedies; ohnehin; seitdem; seither; so; sodann; somit; sonst; soweit; später; statt; trotzdem; überdies; unterdessen; vielmehr; vorher; währenddessen; warum; weiterhin; wie; zudem; zugleich; zum Beispiel; zumal; zunächst; zusätzlich; zuvor; zwar

8.2.4 Präpositionen und Postpositionen

Präpositionale Konnektoren sind Konnektoren, die eine Nominalphrase regieren. Natürlich ist aber nicht jede Präposition in einem Text auch ein Konnektor; eine wichtige Bedingung ist, dass die

Nominalphrase einen Sachverhalt ausdrückt, syntaktisch also in der Regel eine Nominalisierung ist. Mehr dazu unten in Abschnitt 8.4.

Der Konnektor *wegen* kann sowohl als Prä- als auch als Postposition auftreten:

(8.143) **Trotz** des großen Andrangs bleibt das Lokal heute geschlossen.

(8.144) **Wegen** einer privaten Feier bleibt das Lokal heute geschlossen.

(8.145) Einer privaten Feier **wegen** bleibt das Lokal heute geschlossen.

Liste präpositionaler Konnektoren:

aufgrund; bei; durch; infolge; trotz; wegen

8.2.5 Diskontinuierliche Konnektoren

Eine Besonderheit sind Konnektoren, die an zwei Stellen im Satz, also diskontinuierlich auftreten können (z. B. *zwar–aber*) oder müssen (z. B. *weder–noch*). Es ist zu entscheiden, ob es sich um die gemeinsame Markierung derselben inhaltlichen Relation handelt – wenn ja, werden die beiden Elemente als Teile *desselben* Konnektors aufgefasst und entsprechend annotiert.

(8.146) Ich habe zugestimmt, **ohne** mir die Sache richtig überlegt **zu** haben.

Ein etwas komplizierterer Fall ist die *so-X-dass* Konstruktion. *So dass* kann kontinuierlich auftreten oder aber von einer Konstituente (Adverb, Adjektiv, VP) unterbrochen sein. Dann besteht die Möglichkeit, nur *dass* als Konnektor aufzufassen, oder aber *so* und *dass* als Teile eines diskontinuierlichen Konnektors. Da die Konstruktion produktiv (ohne Einschränkung über die X verwendbar) ist, entscheiden wir uns für letzteres:

(8.147) Das Kind hat **so** viel Sahne gegessen, **dass** ihm ganz übel wurde.

Liste diskontinuierlicher Konnektoren:

einerseits - andererseits; entweder - oder; insofern - als; ohne - zu;
so X - dass; um - zu; weder - noch; wenn - auch; zwar - aber;

8.3 Die Argumente von Konnektoren

Die vollständige Annotation von Konnektoren setzt auch das Identifizieren der jeweiligen Argumente eines Konnektors voraus. Diese werden oft auch als *Konnekte* bezeichnet.

Nach Pasch u. a. (2003) gilt für Konnektoren, dass sie propositionale Strukturen als Argument haben. Da wir im Gegensatz zu Pasch u. a. auch präpositionale Konnektoren behandeln, können aber auch bestimmte Nominalphrasen Argumente eines Konnektors sein.

Die propositionalen Strukturen beschreiben Sachverhalte, das sind typischerweise Zustände (8.148), Handlungen (8.149) oder Ereignisse (8.150). Beispiel 8.151 illustriert die „Verpackung“ eines Sachverhalts in einer Nominalphrase, der dann Argument eines Konnektors ist.

(8.148) Maria ist krank.

(8.149) Maria klappt ihren Laptop auf.

(8.150) Maria ist gestern vom Pferd gestürzt.

(8.151) **Aufgrund** ihres gestrigen Sturzes kann Maria heute nicht in die Schule gehen.

Die Bedeutung eines jeden Konnektors ist eine zweistellige Relation. Jeder Konnektor erwartet also zwei Argumente. Bei diesen Argumenten unterscheidet man gemäß Pasch u. a. (2003) zwischen *internem* und *externem* Argument. Das interne Argument ist syntaktisch eng mit dem Konnektor verbunden (bzw. bei Adverbien:

darin enthalten), das externe ist unabhängiger. In Beispiel (8.152) ist also das auf die subordinierende Konjunktion folgende Argument das interne, das andere ist das externe Argument. Achtung: Die Unterscheidung zwischen internem und externem Argument ist in diesen Fällen unabhängig von der linearen Abfolge der Argumente – vgl. (8.153).

(8.152) **Als** [Hans nach Hause kam,]_{intArg} [war die Tür verschlossen.]_{extArg}

(8.153) [Die Tür war verschlossen,]_{extArg} **als** [Hans nach Hause kam.]_{intArg}

8.3.1 Argumenttypen

Die möglichen Argumente von Konnektoren lassen sich grob wie folgt aufteilen: Es können vollständige Sätze sein, Reste von Sätzen, die durch das Entfernen von Satzteilen entstanden sind, oder Nominalphrasen.

8.3.1.1 Sätze

In den einfachsten Fällen werden die Argumente von Konnektoren durch einzelne Sätze gebildet. Dies können Haupt- oder Nebensätze sein. Beispiele haben wir oben gesehen (z. B. 8.152).

Im Fall von adverbialen Konnektoren ist es möglich, dass mehrere aufeinander folgende Sätze ein Argument bilden. Das tritt vor allem beim externen Argument auf. Die Entscheidung über seine Ausdehnung ist oft nicht ganz einfach, wir gehen darauf in Abschnitt 8.4 ein.

(8.154) [Das Meer tobte heftig an diesem Nachmittag. Der Wind hatte fast Orkanstärke.]_{extArg} **Darum** [sind die Fischer nicht ausgefahren.]_{intArg}

8.3.1.2 Nominalphrasen

Für präpositionale Konnektoren ist eine Nominalphrase, die einen Sachverhalt ausdrückt, das Argument des Konnektors. Wenn die Nominalphrase hingegen ein (im weitesten Sinne) Objekt ausdrückt (Person, Gegenstand, etc.), kann die Präposition kein Konnektor sein:

(8.155) **Wegen** [einer privaten Feier]_{intArg} [bleibt das Lokal heute geschlossen.]_{extArg}

(8.156) Wegen Hans bleibt das Lokal heute geschlossen.

Gleiches gilt für koordinierende Konjunktionen, die ebenfalls sachverhaltsausdrückende Nominalphrasen verbinden können und in diesem (und nur in diesem) Fall als Konnektor fungieren:

(8.157) [Der Aufzug der Demonstranten]_{extArg} **und** [das Abrücken der Polizei]_{intArg} waren praktisch gleichzeitig zu beobachten.

(8.158) Die Sonne und der Mond waren praktisch gleichzeitig zu beobachten.

8.3.1.3 Satzteile - Reste von Sätzen

Neben vollständigen Sätzen können die Argumente von Konnektoren auch „Satzreste“ sein, die durch Elision bestimmter Satzteile entstehen:

(8.159) **Weil** [sie sehr groß ist,]_{intArg} [interessierte sie sich für Basketball.]_{extArg}

(8.160) **Weil** [sehr groß,]_{intArg} [interessierte sie sich für Basketball.]_{extArg}

Hier entsteht ein Abgrenzungsproblem: Wann ist die Bedeutung eines Satzrests „noch“ ein durch Elision verkürzt dargestellter Sachverhalt, wann „nur“ ein Objekt? Darauf gehen wir unten in Abschnitt 8.4.4 ein.

8.3.2 Zuordnung des internen und externen Arguments zum Konnektor

Die lineare Positionierung von internem und externem Argument zum Konnektor lässt sich anhand der syntaktischen Kategorien präzise angeben.

8.3.2.1 Subordinierende Konjunktionen

Der Konnektor steht unmittelbar vor dem internen Argument (der Nebensatz). Das externe Argument (der Hauptsatz) kann davor oder dahinter stehen (vgl. Beispiele 8.136 und 8.137 oben); es ist aber auch möglich, dass Konnektor und internes Argument im Stil einer Parenthese in das externe Argument eingebettet sind:

(8.161) [Morgen werde ich, **sobald** [die Sonne scheint,]_{intArg} [an den Baggersee gehen.]]_{extArg}

8.3.2.2 Koordinierende Konjunktionen

Der Konnektor steht unmittelbar vor seinem internen Argument, das externe Argument geht beiden voraus.

(8.162) [Hans ging zum Basketball]_{extArg} **und** [verstauchte sich beim Spielen seinen Knöchel.]]_{intArg}

8.3.2.3 Adverbien

Adverbkonnektoren können an vielen Stellen im Satz positioniert sein. Der umschließende Satz oder Teilsatz (manchmal auch noch nachfolgendes Material) ist das interne Argument, das externe Argument geht diesem voraus.

(8.163) [Der Autoschlüssel ist in den Gully gefallen,]_{extArg} **deshalb** [muss ich jetzt zu Fuß nach Hause gehen.]]_{intArg}

(8.164) [Der Autoschlüssel ist in den Gully gefallen,]_{extArg} [jetzt muss ich **deshalb** zu Fuß nach Hause gehen.]]_{intArg}

8.3.2.4 Präpositionen

Der Konnektor steht unmittelbar vor dem internen Argument, das externe Argument kann davor oder dahinter stehen, oder das interne Argument einbetten.

(8.165) **Trotz** [des schlechten Wetters]_{intArg} [liege ich am See.]_{extArg}

(8.166) [Ich liege am See,]_{extArg} **trotz** [des schlechten Wetters.]_{intArg}

(8.167) [Ich liege **trotz** [des schlechten Wetters]_{intArg} am See.]_{extArg}

8.3.2.5 Diskontinuierliche Konnektoren

Die mehrteiligen Konnektoren treten immer in fixer Reihenfolge auf: Teil 1 vor Teil 2. So kann beispielsweise der *oder*-Teil niemals dem *entweder*-Teil vorangehen.

Das Argument, das Teil 1 einschließt, ist das interne; das Argument, das Teil 2 einschließt, ist das externe.

(8.168) [Der Gesetzgeber demonstriert damit **einerseits** scheinbar Härte,]_{intArg} [tut **andererseits** aber niemandem wirklich weh.]_{extArg}

(8.169) [Hier ist **entweder** eingebrochen worden,]_{intArg} **oder** [ich habe vergessen, die Tür zuzuschließen.]_{extArg}

(8.170) **Entweder** [ist hier eingebrochen worden,]_{intArg} **oder** [ich habe vergessen die Tür zuzuschließen.]_{extArg}

Manche zweiteiligen Konnektoren können auch „einteilig“ auftreten: z. B. *oder* ohne *entweder*; *andererseits* ohne *einerseits*. Dann verfahren wir bei der Argumentzuordnung wie gewohnt bei einteiligen Konnektoren: *andererseits* steht mit seinem internen Argument; etc.

8.4 Hilfestellung zur Annotation

Nach dem Überblick über das Annotationssziel werden nun einige weiterführende Erklärungen gegeben, außerdem diskutieren wir typische Problemfälle bei der Annotation.

8.4.1 Konnektoren: Grundlegende Merkmale

Wir haben bisher eher informell charakterisiert, was ein Konnektor ist. Als kompakte Definition geben wir jetzt nach Pasch u. a. (2003) an:

Merkmal 1: Ein Konnektor ist nicht flektierbar.

Merkmal 3: Die Bedeutung eines Konnektors ist eine spezifische zweistellige Relation.

Merkmal 4: Die Relate der Bedeutung eines Konnektors bezeichnen Sachverhalte.

Merkmal 5: Die Relate der Bedeutung eines Konnektors müssen durch Sätze bezeichnet werden können.

„Relat“ entspricht dem, was wir hier als „Argument“ bezeichnen. Das Merkmal 5 impliziert einen Paraphrasentest: Argumente eines Konnektors sind entweder bereits Sätze, oder aber sie sind intuitiv zu Sätzen umformbar, ohne dass sich die Bedeutung verändert.

Pasch u. a. (2003) verwenden auch noch das Merkmal 2: „Ein Konnektor vergibt keine Kasusmerkmale an seine syntaktische Umgebung.“ Dieses lassen wir aber hier beiseite, weil wir einige Verwendungen von Präpositionen (die ja Kasus vergeben) auch als Konnektoren betrachten.

8.4.2 Wörter in nicht-Konnektor Lesart

Viele der in den obigen Listen (Abschnitt 8.2) als Konnektor genannten Wörter haben auch eine nicht-Konnektor Lesart, sind also *ambig*. In der nicht-Konnektor Lesart hat ein solches Wort semantisch betrachtet keine zwei (Sachverhalt-) Argumente, also kann

es nach **Merkmal 3** kein Konnektor sein. Wir besprechen einige Beispiele.

8.4.2.1 Konnektor oder Anapher?

Wir betrachten zunächst das Beispiel *da*.

(8.171) Ich fahre nach Rügen, **da** ich eine Woche Urlaub habe.

(8.172) Ich fahre nach Rügen, *da* war ich als Kind auch schon oft.

Bei *da* geht die Unterscheidung zwischen Konnektor und Nicht-Konnektor einher mit dem Unterschied in der syntaktischen Kategorie (Subordinierende Konjunktion versus temporales/lokales Adverb), ist also recht einfach zu erkennen. Viele Adverbien sind jedoch bei gleichbleibender syntaktischer Kategorie ambig:

(8.173) Anfang September sollen der Landtagsverwaltung die Schlüssel für das Parlamentsschloss übergeben werden.

Dann folgt ein dreimonatiger Probebetrieb, *dann* werden Technik, Akustik und Klima- und Lüftungsanlagen im Plenarsaal, den Sitzungsräumen und Büros getestet.⁶

In diesem Beispiel drückt das erste **dann** eine temporale Abfolge zweier Ereignisse aus, nimmt also semantisch zwei Argumente: Auf die Schlüsselübergabe Anfang September folgt der Probebetrieb. Das zweite *dann* hingegen drückt keine Abfolge aus: Es ist nicht so, dass *nach* dem Probebetrieb der Techniktest stattfindet. Vielmehr ist *dann* hier eine temporale Anapher, die die NP *ein dreimonatiger Probebetrieb* wieder aufnimmt.

Es gibt eine Reihe weiterer Wörter, die einerseits als Konnektor, andererseits als Anapher gebraucht werden können. Hier ein Beispiel für die beiden Verwendungen von *dagegen*:

(8.174) Bei großer Hitze wird man durstig. **Dagegen** ist für Kälte kein solcher Effekt bekannt.

⁶Potsdamer Neueste Nachrichten, 20.8.13

(8.175) Bei großer Hitze wird man durstig. *Dagegen* kann man nichts machen.

Leider ist die Entscheidung nicht immer leicht zu treffen:

(8.176) Maria und Hans trafen sich an der Eisdielen um 19 Uhr, **als** Marias Lieblingsserie vorbei war.

In der einen Lesart nimmt *als* den Zeitpunkt *19 Uhr* anaphorisch auf und fungiert als Relativpronomen – ist also kein Konnektor. In der anderen Lesart drückt **als** die Gleichzeitigkeit zweier Ereignisse aus: das Treffen zwischen Maria und Hans, und das „Vorbeisein“ von Marias Lieblingsserie. Der Grundsatz für die Annotation ist: Wenn es für ein Wort sowohl eine Konnektor- als auch eine Nicht-Konnektor-Lesart gibt und keine klare Entscheidung möglich ist, dann behandeln wir das Wort als Konnektor.

8.4.2.2 Konnektor oder (nicht-anaphorisches) Adverb?

Nicht bei allen Kandidaten ist die nicht-Konnektor Lesart eine anaphorische. Ein oftmals schwieriger Fall ist *auch*.

(8.177) Maria wünscht sich ein neues Fahrrad. **Auch** hätte sie gern ein Smartphone.

(8.178) Maria wünscht sich ein paar technische ‚gadgets‘. Auch ein Smartphone hätte sie gern.

Durch die Position vor dem Verb wird in 8.177 klar, dass *auch* die beiden Sachverhalte ‚Marias Wunsch nach Fahrrad‘ und ‚Marias Wunsch nach Smartphone‘ durch eine additive Relation verbindet. Als Diagnose kann vor den zweiten Satz ein *Außerdem*: gestellt werden. Hingegen liegt keine solche Relation in 8.178 vor. Der *außerdem*-Test misslingt, weil der Wunsch nach dem Smartphone eine Spezialisierung des Wunschs nach ‚gadgets‘ ist, und kein weiterer Wunsch. Deshalb erhält man auch ein fragwürdiges Ergebnis, wenn *auch* vor das Verb gestellt wird:

(8.179) ? Maria wünscht sich ein paar technische ‚gadgets‘. Auch hätte sie gern ein Smartphone.

In 8.178 hat *auch* engen Skopus allein über die Nominalphrase: Maria wünscht sich so einiges, und darunter ist ein Smartphone. Weil der Skopus nicht über die VP reicht, liegt kein sachverhaltsverknüpfendes, semantisch zweitstelliges *auch* vor.

8.4.2.3 Konnektor oder Komplementierer?

Eine „Falle“ für die Konnektor-Analyse bilden Wörter wie *ob* und *dass* in ihrer Verwendung als Komplementierer, vor allem für Kommunikations- und Psych-Verben (aber nicht nur dort):

(8.180) Sie behauptete, dass sie krank sei.

(8.181) Dass die Welt eine Scheibe ist, habe ich noch nie angenommen.

(8.182) Ob das stimmt, werden wir ja noch sehen.

Nach o.g. Kriterium ‚Skopus mindestens über VPs‘ scheint in diesen Beispielen jeweils ein Konnektor vorzuliegen. Dem ist allerdings nicht so, denn wir fordern neben dem syntaktischen Kriterium eine semantische ‚Vollständigkeit‘ der beschriebenen Sachverhalte. In 8.180 ist das krank-sein der *Inhalt* der Behauptung, und nicht etwa unabhängig davon. Analog in 8.181: Welt-eine-Scheibe-sein ist der Gehalt der Annahme und nicht etwa ein Sachverhalt, der neben der Annahme besteht. (Ähnlich für 8.182.) Eine andere (aber eng verwandte) Diagnose besteht darin, gemäß **Merkmal 5** die möglichen Argumente auf Hauptsatzstatus hin zu prüfen. In 8.180 würde *dass* eine Relation herstellen zwischen „sie behauptet“ auf der einen Seite und „sie ist krank“ auf der anderen Seite. Während letzteres ein vollständiger Satz ist, ist „sie behauptet“ syntaktisch und semantisch unvollständig, es fehlt ein Argument des Verbs. Diese Diagnose greift auch für 8.181, etwas weniger klar für 8.182. Letztlich sollten im Zweifelsfall beide eben beschriebenen Kriterien angewandt werden.

Hier noch ein Korpus-Beispiel für eine Nicht-Konnektor Lesart, diesmal mit *weshalb*. Das vorwärtsverweisende *Es* am Satzanfang ist ein weiterer Hinweis darauf, dass der eingebettete Nebensatz die „Füllung“ des Verbs im Hauptsatz angibt, so dass keine zwei separaten, vollständigen Propositionen vorliegen.

(8.183) Es dürfte derzeit schwer zu vermitteln sein, weshalb ein Ressort pauschal von künftigen Einsparungen ausgenommen werden soll auf Kosten der anderen.
(maz00001)

Nicht zu vergessen: Alle eben besprochenen Wörter (und weitere) können aber auch als Konnektoren gebraucht werden:

(8.184) Meier musste sich zusammenreißen, **dass** er nicht anfang zu weinen.

(8.185) **Ob** es auch schneit, ich bleibe auf jeden Fall daheim.

(8.186) Es begann zu schneien, **weshalb** ich daheim blieb.

In diesen Beispielen ist *dass* bedeutungsähnlich zu *damit*, *ob* zu *wenn* und *weshalb* zu *so dass*, womit jeweils Sachverhaltsverknüpfungen hergestellt werden.

8.4.2.4 Konnektor oder Verb-Partikel?

Des weiteren haben einige Wörter mit Konnektor-Lesart auch eine Lesart als Partikel, die in enger Verbindung zum Verb steht, wie etwa in *als Maler arbeiten*. In solchen Fällen liegt natürlich kein Konnektor vor. Beispiele:

(8.187) Mein Nachbar arbeitet *als* Maler.

(8.188) Die Kollegin hat immer ihren Laptop *dabei*.

Gleiches gilt, wenn die Wörter Teile einer idiomatischen Wendung sind, wie in dem Ausruf *Ohne Wenn und Aber* (der bemerkenswerterweise eine Serie von vier Konnektor-Kandidaten darstellt).

Zusammengefasst: Jedes Wort der Listen aus Abschnitt 8.2 ist ein Konnektoren*kandidat*, muss aber im jeweiligen Kontext noch einmal darauf hin überprüft werden, ob es wirklich in der Konnektor-Lesart vorliegt! Gleiches gilt natürlich für Wörter, die nicht in den Listen stehen (denn diese haben ja nur orientierende Funktion).

8.4.3 Bestimmung der Argumente

Für Präpositionen und Konjunktionen dürfte die Bestimmung der Argumente in den meisten Fällen einfach sein. Es gibt jedoch schwierige Fälle bei der Festlegung der Ausdehnung von Argumenten – vor allem bei den Adverbien, mitunter auch bei Konjunktionen.

Für Adverbien ist das externe Argument in klaren Fällen der (Teil)-Satz, der dem (Teil)-Satz des internen Arguments (mit Konnektor) vorausgeht, wie in den obigen Beispielen 8.154, 8.163, 8.164.

In authentischen Texten stellt sich aber oft die Frage, ob nicht *mehr* als nur der vorangehende Satz das externe Argument bilden sollte. Für diese Entscheidung muss die semantische/pragmatische Relation, die der Konnektor ausdrückt, untersucht werden: Welches sprachliche Material ist erforderlich, damit diese Relation *vollständig* ausgedrückt ist?

(8.189) [Heute morgen habe ich mir das Fahrrad platt gefahren.
Und nachmittags bekam ich eine 4 in der
BWL-Prüfung.]_{extArg} [Es ist **also** wahrlich kein schöner
Tag.]_{intArg}

In manchen Fällen ist auch eine Ausdehnung des internen Arguments nötig, wenn der den Konnektor enthaltene Satz allein nicht für das Verständnis der Relation ausreicht. Zur Illustration können wir das obige Beispiel umformen:

(8.190) [Heute ist wahrlich kein schöner Tag.]_{extArg} **Denn** [heute
morgen habe ich mir das Fahrrad platt gefahren. Und
nachmittags bekam ich eine 4 in der BWL-Prüfung.]_{intArg}

Eine Ausnahme von der Forderung nach Vollständigkeit stellen anaphorische Pronomen dar. Wenn der Satz des externen Arguments ein solches Pronomen enthält, muss nicht nur wegen der Auflösung der Referenz noch weiteres vorangehendes Material hinzugefügt werden.

(8.191) Heute morgen habe ich meine Nachbarin getroffen. [Sie hat mir zwei Brötchen vom Bäcker mitgebracht.]_{extArg} [Ich habe mich **deshalb** artig bei ihr bedankt.]_{intArg}

Umgekehrt ist der Skopus manchmal auch *kleiner* als es vielleicht auf den ersten Blick scheint. Dies ist wichtig, wenn beispielsweise zwei konjugierte Verbkomplexe insgesamt unter einem Hauptsatz eingebettet sind:

(8.192) So ist etwa unklar, [wer Abfindungen erhalten soll,]_{extArg}
oder [was passiert, wenn zu wenig Lehrer die Angebote des vorzeitigen Ausstiegs nutzen.]_{intArg}
(maz00001)

Der Konnektor verbindet zwei Propositionen, die beide „unklar sind“, also dem semantischen (und hier auch syntaktischen) Skopus des Hauptverbs untergeordnet sind, darum dürfen auch nur die beiden Verbkomplexe die Argumente bilden, und nicht etwa das *So ist etwa unklar*.

Zusammengefasst geben wir diese „Faustregeln“ für die Abgrenzung der Argumente an:

- Ein Argument bildet immer eine kontinuierliche Textspanne, die lediglich durch ein anderes Argument unterbrochen werden darf. Etwaig eingebettete Parenthesen, Appositionen etc. sind stets Teil eines Arguments.
- Adverbien: Meistens wird das externe Argument dem internen *unmittelbar* vorausgehen. Das ist aber keine strikte Regel! Insbesondere bei solchen Adverbien mit anaphorischem Morphem (*deswegen*, *stattdessen*, usw.) geprüft werden, ob intervenierendes Material zu überspringen ist.

- Die Argumente sind korrekt bestimmt, wenn die durch den Konnektor ausgedrückte Relation mit ihnen inhaltlich vollständig beschrieben und verständlich ist. Die Prüfung auf Vollständigkeit geschieht nach (mentalem) Einsetzen des jeweiligen Antezedens für etwaig auftretende Pronomen.
- In der Regel verläuft eine Argumentgrenze entlang von (Teil-)Satzgrenzen (Ausnahme natürlich: die NP bei Präpositionen). Nur in gut begründeten Fällen sollte eine Argumentgrenze an anderer Stelle auftreten.
- Bei Argumenten in eingebetteten Sätzen ist auf die semantischen Skopusverhältnisse zu achten.
- Im übrigen gilt das Prinzip der Kürze: Die Argumente sollen nicht unnötig weit ausgedehnt werden.

8.4.4 Umgang mit Ellipsen

Wir haben oben am Beispiel von *auch* die Frage nach dem semantischen Skopus eines Konnektor-Kandidaten besprochen: Sachverhaltsverknüpfung ist „normalerweise“ nur möglich bei weitem, mindestens über eine VP reichenden Skopus.

(8.193) [Hans gewann eine Million Euro]_{extArg} **und** [reiste sofort um die ganze Welt.]_{intArg}

Hier liegen zwei offenkundig verschiedene Sachverhalte vor, die Elision des Subjekts im zweiten Konjunkt ändert für uns nichts an der „Vollständigkeit“ der beiden Argumente. Schwieriger kann die Entscheidung werden, wenn auch das Verb elidiert ist:

(8.194) Hans holt die Butter *und* das Brot.

Hier muss intuitiv entschieden werden, ob einer oder zwei Sachverhalte vorliegen – ob der „Hol-Vorgang“ hier ein und derselbe ist oder nicht; die sprachliche Formulierung lässt beide Interpretationen zu. Bestehen Zweifel, wie zu entscheiden ist, dann soll eher *ein*

Sachverhalt und damit eine Nicht-Konnektor-Lesart angenommen werden. Das bedeutet für dieses Beispiel: Hans holt etwas, und zwar die Butter und das Brot; der Skopus von *und* erstreckt sich „nur“ über zwei NPs, die keine Sachverhalte bezeichnen.

Wie in Abschnitt 8.3.1 erläutert, gibt es aber sachverhaltsbezeichnende NPs, die bei Präpositionen und eben auch bei Konjunktionen als Argumente fungieren.

(8.195) [Die Untersuchung der Wohnung]_{extArg} **und** [das Abschleppen des Autos]_{intArg} verschlangen einen ganzen Tag.

Schwierige Fälle der Elision können auftreten, wenn das zweite Konjunkt lediglich ein neues Attribut angibt, wie unten in Beispiel 8.196. Hier ist nicht klar, ob man von zwei verschiedenen ‚Sachverhalten‘ sprechen kann, doch stellt der Autor zwei Forderungen, die man als separat auffassen kann: Einerseits soll die Regierung sich entscheiden, andererseits soll sie das schnell tun. Entsprechend interpretieren wir die Konjunktion hier als Konnektor. In 8.197 führt die Elision des Objekts im ersten Konnekt zu einem verdächtig „kurzen“ Segment, doch liegen klar zwei verschiedene Sachverhalte vor, so dass wir auch hier von einem Konnektor ausgehen.

(8.196) [Die Regierung muss sich entscheiden,]_{extArg} **und** [zwar schnell.]_{intArg}

(8.197) [Hans liebte]_{extArg} **aber** [Maria verabscheute die Rosinen im Käsekuchen.]_{intArg}

8.4.5 Überlappungen, Einbettungen und multiple Konnektoren

Die Annotation von Konnektor+Argumenten erfolgt immer *individuell* für jeden aufgefundenen Konnektor. Das bedeutet vor allem, dass die Entscheidung über die Ausdehnung der Argumente unabhängig davon getroffen werden soll, ob möglicherweise noch

andere Konnektoren+Argumente intervenieren. Denn selbstverständlich können die Argumente von Konnektoren überlappen oder identisch sein.

(8.198) Heute nachmittag gehe ich ins Museum, **weil** ich nicht zur Arbeit muss. **Danach** mache ich ein Nickerchen.

Gelegentlich können auch zwei Konnektoren identische Argumente haben (dann gehen die beiden Konnektoren meist auch eine semantische Verbindung ein – wir sprechen dann von „multiplen Konnektoren“. Sie sind aber stets getrennt zu annotieren, wenn sie unterscheidbare Bedeutungsbeiträge leisten (Weglassprobe).

(8.199) Hans gewann eine Million Euro **und** reiste **danach** sofort um die ganze Welt.

8.4.6 Details zur Notation

Wie die letzten Beispiele zeigen, kann im Text die Zuordnung zwischen Konnektoren und Argumenten nicht immer einfach erfassbar sein. Um Eindeutigkeit zu erzielen, markieren wir alle Konnektoren und ihre beiden Argumente jeweils mit einem numerischen Index. Bei mehrteiligen Konnektoren erhalten beide Teile denselben Index.

Des weiteren muss per Konvention festgelegt werden, ob Konnektoren Teile ihrer Argumente sind oder nicht. Wir sind bereits (implizit) in den Beispielen nach dieser Regel vorgegangen: Steht ein Konnektor *am Rand* seines internen Arguments (das ist bei Konjunktionen immer, bei Adverbien manchmal der Fall), bleibt er außerhalb des Arguments. Adverbien, die innerhalb des internen Konnekts sind, verbleiben dort und werden nicht etwa „ausgeklammert“. Generell gilt, dass ein Argument immer eine kontinuierliche Textspanne sein muss und nicht unterbrochen werden kann.

Abschließend der Umgang mit Interpunktionszeichen: Zeichen am Ende eines Argumenten sind jeweils Teil des Arguments. Auch dies haben die Beispiele bereits illustriert.

Zur Verdeutlichung hier eine Erweiterung von Beispiel 8.198, nunmehr vollständig annotiert. Zu beachten ist hier auch, dass das

externe Argument von **Danach** nicht unmittelbar vor dem internen steht: Die temporale Abfolge bezieht sich auf die beiden Ereignisse Museumsbesuch und Nickerchen/Schwesterbesuch, nicht auf die dazwischen stehende Begründung.

(8.200) [[Heute nachmittag gehe ich ins Museum,_{extArg-1}]_{extArg-2}
weil₁ [ich nicht zur Arbeit muss.]_{intArg-1} **Danach**₂ [[mache
ich **entweder**₃ ein Nickerchen]_{intArg-3} **oder**₃ [besuche meine
Schwester.]_{extArg-3}]_{intArg-2}

Inhaltszonen

Andreas Peldszus, Manfred Stede

Hinweise:

- Diese Richtlinien gelten speziell für Pro-/Contra-Kommentare.
- Voraussetzung für die Anwendung der Richtlinien ist, dass der Text bereits in Diskurssegmente unterteilt ist (Kapitel 3).
- Die vorliegenden Richtlinien verwenden eine Notation, wie sie „auf Papier“ oder im Text-Editor verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.
- Alle hier genannten Textbeispiele ohne Quellenangabe sind fiktiv.

9.1 Hintergrund

Viele Texte bestehen „naturgemäß“ aus einer Reihe von klar unterscheidbaren Bestandteilen: Kochrezepte geben eine Zutatenliste sowie die Folge von Zubereitungsschritten an. Bedienungsanweisungen, vor allem die für technische Geräte, führen eine Folge von Schritten an, in die gut sichtbare Warnungen vor Fehlbedienung eingestreut sind. Beipackzettel von Medikamenten haben eine immer wiederkehrende grundsätzliche Struktur. Gerichtsurteile setzen sich aus Rubrum, Tenor, Tatbestand und Entscheidungsgründen zusammen, die ihrerseits eine mehr oder weniger standardisierte „Binnenstruktur“ aufweisen. Und so weiter. Für manche Textsorten ist diese Strukturierung stärker ausgeprägt (Beipackzettel, Urteil) als für andere (Tagebuch, persönliche Email).

Für die Art von Struktur, die wir hier behandeln, gelten zwei grundsätzliche Punkte:

- Es ist eine *funktionale* Struktur in dem Sinn, dass jeder der Bausteine eine konventionalisierte Rolle für den Gesamt-Zweck des Textes spielt.
- Es ist eine *textsortenspezifische* Struktur: Das Inventar der Bausteine und die Konventionen für ihre Abfolge im Text sind ein wichtiger Bestandteil der Charakterisierung einer bestimmten Textsorte.

Die „Bausteine“ nennen wir im Folgenden *Inhaltszonen*. Für den Zeitungskommentar scheint es auf den ersten Blick vielleicht etwas schwierig, über solche Zonen zu sprechen: Im Gerichtsurteil machen Layout und Überschriften die grobe Zonen-Einteilung bereits klar, aber wie steht es mit dem Fließtext eines Kommentars?

Doch auch an den Kommentar, speziell an den hier behandelten *Pro-/Contra-Kommentar* haben wir als LeserInnen bestimmte Erwartungen, welche funktionalen Bestandteile wir antreffen werden. Die Reihenfolge dieser Zonen allerdings ist kaum codifiziert, sondern weitgehend eine Frage der gewählten rhetorischen Strategie.

Im Folgenden erläutert Abschnitt 9.2 das Inventar der Zonen für unsere Textsorte. In 9.3 wird die Vorgehensweise beim Annotieren beschrieben. Zum Schluß illustriert Abschnitt 9.4 den Prozess anhand eines vollständigen Beispieltexts.

9.2 Das Inventar der Inhaltszonen

Wir definieren die folgenden Zonen zur Beschreibung der funktionalen Struktur von Pro-/Contra-Kommentaren. Einige Zonen sind für jedes Textexemplar obligatorisch, andere optional; das geben wir jeweils an. Erfahrungsgemäß gibt es keinen konventionalisierten Aufbau von Kommentaren, bestenfalls einige typische lokale Muster der Abfolge einzelner Zonen. Darauf gehen wir im Folgenden jedoch nicht ein. Wichtig ist: Es gibt keine einfache „Grammatik“ für den Aufbau von Kommentaren durch ihre Inhaltszonen.

Zentrale These: Das Segment, in dem der Autor konzise seine Position zur Fragestellung ausdrückt, die der Kommentar erörtert. Salopp gesagt, sagt der Autor in dieser Zone recht klar, ob er „dafür oder dagegen“ ist. In der Mehrzahl der Kommentare wird diese Zone vorhanden sein, sie zu identifizieren ist manchmal eine Frage der Abwägung unter verschiedenen Kandidaten. Nur wenige Kommentare argumentieren so implizit, dass sich kein geeigneter Kandidat für die zentrale These findet.

Pro Argument: Gründe, die der Autor zur Stützung der zentralen These angibt. Dies ist zwangsläufig in jedem Kommentar vorhanden und wird meist auch den größten Raum einnehmen. Pro-Argument-Zonen können mehrfach auftreten, sich also mit anderen Zonen abwechseln.

Con Argument: Gründe, die eigentlich *gegen* die Position des Autors sprechen, die er aber trotzdem nennt. Diese Gegen Gründe hält er für nicht stichhaltig genug, zu schwach, o. ä. In den meisten Fällen wird der Autor einen Gegen Grund im weiteren Verlauf des Textes wieder entkräften. Die Con Argument Zone ist optional: Manche Kommentare kommen ohne sie aus, argumentieren nur „in eine Richtung“.

Hintergrund: Informationen, die nicht unmittelbar an der Argumentation teilhaben, sondern primär dazu dienen, den Leser thematisch zu orientieren. Auch diese Zone ist optional.

Fazit: Viele Kommentare enden mit einer knappen Zusammenfassung der Position des Autors, oft sehr „griffig“ formuliert. Sofern die zentrale These bereits andernorts markiert wurde, markieren wir diese abschließende Zone als Fazit.

Eine Illustration dieser Zonen (Bearbeitung eines Beispieltexts) findet sich unten im Abschnitt 9.4.

9.3 Die Annotations Schritte

Wir geben die Annotationsanweisung in Form einer Abfolge von Schritten an, die auch in dieser Reihenfolge ausgeführt werden soll. Am Ende muss jedem Diskurssegment genau eine Zone zugeordnet sein. Das heißt, kein Segment bleibt von der Annotation unberücksichtigt, und keines hat mehr als eine Zonenzuordnung. (Mit anderen Worten: Es erfolgt eine vollständige Partitionierung des Textes.)

Unsere Notation nutzt eckige Klammern und numerische Indizes zur Markierung von Diskurssegmenten, über dem Index steht ggf. das Zonen-Label (oder, wenn die Rolle von Diskurssegment-Typen besprochen wird, dieser Typ).

9.3.1 Vorbemerkungen zur Segmentierung

Eine Inhaltszone kann sich über eines oder mehrere der bereits markierten Diskurssegmente erstrecken. Grenzen zwischen Zonen sind meistens Satzgrenzen, mitunter aber auch clause-Grenzen. Ein prototypisches Beispiel dafür sind *zwar-aber* Konstruktionen.

(9.201) [*Das Gebäude muss abgerissen werden.*]₁^{THESE} [*Zwar ist es ganz hübsch,*]₂^{CON} [*aber das Asbest lässt uns keine andere Wahl.*]₃^{PRO}

Aber auch alle anderen Segmentgrenzen sind an einer clause-Grenze, also innerhalb von Sätzen, möglich.

Mitunter trifft man auf die gegenteilige Situation: Eine einzelne Proposition ist über mehrere benachbarte Segmente verteilt. In aller Regel sind daran dann FRE oder FRB Segmente beteiligt. In diesem Fall weist man jedem dieser Segmente die Zone zu, die der Gesamt-Proposition entspricht. Im folgenden Beispiel elaborieren die Segmente 2 und 3 das Segment 1: Sie ergänzen eine allgemeine Aussage durch spezifischere Information; dabei besteht keine argumentative Stützungs- oder Angriffsrelation zwischen den Segmenten, weshalb sie derselben Zone zugeordnet werden.

(9.202) [*Berlin braucht mehr Parkplätze.*]₁^{HS} [*Jetzt, so schnell wie möglich.*]₂^{FRB} [*Und ohne das Verkehrskonzept neu zu erfinden.*]₃^{FRB}

Ein anderes Beispiel aus dem Text, den wir in Abschnitt 9.4 besprechen werden:

(9.203) [*Vermisst jemand den vermüllten Güterbahnhof, auf dem heute die „Spandau-Arcaden“ stehen?*]₁^{HS} [*Na also.*]₂^{FRB}

Die Bedeutung von Segment 2 hängt vollständig von der des Segments 1 ab; 2 ist nur im Zusammenhang mit 1 zu verstehen, beide müssen bei der Zonen-Analyse als eine Einheit betrachtet werden, also dieselbe Zone zugewiesen bekommen.

Beispiel 9.203 enthält eine *rhetorische Frage*. Für diese Konstruktionen gilt, dass sie vor den Zonen-Beurteilung „aufgelöst“ werden sollen, d.h. zu einer Aussage umgeformt werden sollen, die dem entspricht, was der Autor mit der Frage aussagen möchte. So ist unser Beispiel sicherlich gleichbedeutend mit *Niemand vermisst ja wohl den vermüllten Güterbahnhof, auf dem heute die „Spandau-Arcaden“ stehen.*

9.3.2 Zentrale These

Gibt es im Text – ohne die Überschrift einzubeziehen – ein oder mehrere Segmente, das/die in prägnanter Weise die Antwort des Autors auf die gestellte Frage ausdrückt/ausdrücken?

- Genau 1 Segment: Markieren Sie es als Zone ZENTRALE THESE.
- Mehr als 1 Segment: Ist eines dieser Segmente am Textende? Dann ist es wahrscheinlich besser als Zone FAZIT zu analysieren (s. nächster Schritt). Verbleibt 1 Segment, so markieren Sie es als ZENTRALE THESE; verbleiben mehrere Segmente, so prüfen Sie genau, ob möglicherweise eines eine Begründung für ein anderes darstellt – in diesem Fall ist nur dasjenige Segment, das begründet wird, die ZENTRALE THESE. Nur wenn wirklich mehr als ein Segment in gleicher Weise die „zentrale These“ ausdrückt, sind sie gleichermaßen zu markieren. (Dies wird selten der Fall sein.)

- Kein Segment: Kommt kein Textsegment in Frage, so scheint der Text nur implizit für seine Antwort zu argumentieren. In diesem Fall ist die Überschrift (mit der Antwort „ja/nein“) als ZENTRALE THESE zu markieren.

9.3.3 Fazit

Sofern das Segment am Textende noch nicht markiert ist: Dient es (bzw. die dortige Segmentfolge) als Zusammenfassung der Position des Autors, möglicherweise „pointiert“ oder rhetorisch geschickt formuliert? Dann ist dieses Segment als FAZIT zu markieren.

9.3.4 Argumente für die These

Jeder Pro/Contra Kommentar muss eine Position vertreten und diese Position begründen. Segmente, die diesen Zweck erfüllen, also zur argumentativen Abstützung der zentralen These dienen, werden als PRO ARGUMENT markiert. Ein Paraphrasentest, der in der Regel gut funktioniert, ist die Verknüpfung zwischen These und Stützung mit *denn* oder *weil*: Gibt der Satz „<These>, denn/weil <pro arg Kandidat>“ plausibel die Intention des Autors wieder, ist PRO ARGUMENT die richtige Zone. Für die Paraphrase müssen die Segmente, die <These> bzw. <Kandidat> entsprechen, ggf. etwas umgeformt werden, damit ein grammatikalischer und sinnvoller Satz entsteht; zum Beispiel sind etwaig vorhandene Anaphern durch ihr Antezedens zu ersetzen, und Konnektoren zu entfernen. (Ein Beispiel folgt in Kürze.)

Als Argumente für die These gelten nicht nur Segmente, die die These direkt stützen, sondern auch Segmente, die ein anderes PRO ARGUMENT stützen. Im folgenden Beispiel gilt Segment 3 also auch als PRO ARGUMENT, weil es Segment 2 stützt, welches seinerseits die ZENTRALE THESE 1 stützt:

(9.204) [*Berlin braucht mehr Parkplätze.*]₁^{THESE} [*Die vorhandenen Stellflächen sind längst mehr als ausgelastet.*]₂^{PRO} [*Sie wurden in den vergangenen Jahren nur selten erweitert.*]₃^{PRO}

Wir zeigen die behaupteten Stützungsbeziehungen anhand des *weil*-Tests:

Berlin braucht mehr Parkplätze, weil die vorhandenen Stellflächen längst mehr als ausgelastet sind. → 2 *stützt* 1

? Berlin braucht mehr Parkplätze, weil die vorhandenen Stellflächen in den vergangenen Jahren nur selten erweitert wurden. → 3 *stützt* 1 *nicht gut*

Die vorhandenen Stellflächen sind längst mehr als ausgelastet, weil sie in den vergangenen Jahren nur selten erweitert wurden. → 3 *stützt* 2

9.3.5 Argumente gegen die These

Oftmals begibt sich der Autor in einen „virtuellen Dialog“ mit der argumentativen Gegenseite und nennt ein potenzielles Argument gegen seine eigene Position, das von der Gegenseite vorgebracht werden könnte. Ein solches Gegenargument wird er dann in der Regel wieder entkräften, um insgesamt die Darstellung der eigenen Position kohärent zu halten. Ein Paraphrasentest für Gegenargument und Entkräftung ist „Zwar <con arg Kandidat> aber <pro arg Kandidat>“. Gibt dieser Satz die im Text abgebildete Intention des Autors wieder, sind die entsprechenden Zonen-Zuordnungen passend: Das Gegenargument erhält die Zone CON ARGUMENT; die Entkräftung des Gegenarguments hat den Effekt, die eigene These zu stützen, wird daher als PRO ARGUMENT markiert:

(9.205) [*Berlin braucht mehr Parkplätze.*]₁^{THESE} [*Zwar darf es keine zusätzlichen Anreize fürs Autofahren im Stadtzentrum geben.*]₂^{CON}
[*Aber irgendwo müssen die Autos, die eh schon da sind, ja auch hin.*]₃^{PRO}

Als Argument gegen die These gelten nicht nur Segmente, die die These direkt angreifen, sondern auch Segmente, die ein anderes CON ARGUMENT stützen. Im folgenden Beispiel gilt Segment 3 auch als CON ARGUMENT, weil es Segment 2 stützt, welches seinerseits die ZENTRALE THESE 1 angreift:

(9.206) [*Berlin braucht mehr Parkplätze.*]₁^{THESE} [*Zwar darf es keine
zusätzlichen Anreize fürs Autofahren im Stadtzentrum geben.*]₂^{CON}
[*denn niemand möchte noch mehr Verkehr in der Innenstadt.*]₃^{CON}
[*Aber irgendwo müssen die Autos, die eh schon da sind, ja auch
hin.*]₄^{PRO}

Der Autor kann sich mit einem vorgebrachten CON ARGUMENT unterschiedlich stark identifizieren. Er kann z. B. feststellen, dass das Gegenargument auch für ihn als ein berechtigter Einwand gilt, den er allerdings entkräften möchte. So in folgenden Schemata:

- THESE. Zwar CON, aber PRO.
- Obwohl CON, gilt THESE, denn PRO.
- THESE, auch wenn CON. Schließlich PRO.
- THESE. Ja, es ist richtig, dass CON. Aber PRO.

Andererseits kann der Autor sich vom Gegenargument etwas distanzieren, indem er es jemandem anderen zuschreibt, es als allgemein mögliche Annahme darstellt, oder sogar übersteigert zitiert.

- THESE. Zwar sagt XY, dass CON, aber PRO.
- THESE. XY befürchtet, dass CON. Jedoch PRO.
- THESE. Man könnte denken, dass CON, aber PRO.
- Oh ja, es ist ja sooo der Fall, dass CON. . . Was für ein Quatsch!
Es gilt PRO. Also THESE.

In seltenen Fällen kann es passieren, dass in einem Segment ein Gegenargument zwar Erwähnung findet, aber gleichzeitig im selben Segment noch entkräftet wird. Wenn dem so ist, soll das Segment *nicht* also CON-, sondern als PRO ARGUMENT markiert werden. Siehe in diesem Beispiel das Segment 5:

(9.207) [*Berlin braucht mehr Parkplätze.*]₁^{THESE} ... [*Der Hinweis, dass stattdessen ja auch Fahrradständer aufgestellt werden könnten, ist natürlich nicht ernstzunehmen.*]₅^{PRO} [*Davon gibt es nun wirklich genug.*]₆^{PRO}

Der Grund für die Entscheidung ist, dass die Gesamtfunktion des Segments durch die gleichzeitige Präsentation und Entkräftung eines Gegengrundes letztlich eine Stützung der zentralen These darstellt.

9.3.6 Hintergrundinformation

Segmente, die jetzt noch unmarkiert sind, werden die Funktion einer thematischen Einführung haben, oder an bestimmter Stelle Hintergrundinformation geben, die nicht unmittelbarer Bestandteil der Argumentation ist. Diese Segmente sind als HINTERGRUND zu markieren.

9.4 Analyse eines Beispieltexts

Zur Illustration der geschilderten Vorgehensweise und der Entscheidungen diskutieren wir hier die Bearbeitung eines Beispieltexts.⁷

⁷PCC-ProCon-einkaufszentren-con

-
- 0 Hat Berlin zu viele Einkaufszentren? *Nein!*
- 1 Ohgottogott! CON
- 2 All diese Einkaufscenter, überall nur kalter Stahl und hartes Glas. CON
- 3 Und die vielen Läden im Kiez – die haben doch keine Chance gegen diese Kommerzmonster. CON
- 4 So, jetzt aber Schluss mit solch jammernder Kleinstadt-Romantik. PRO
- 5 Berlin hat nicht zu viele, sondern noch lange nicht genug Einkaufszentren. THESE
- 6 Nehmen wir die Wilmersdorfer Straße in Charlottenburg. HGRND
- 7 Auch dort entsteht jetzt ein Einkaufszentrum mit vielen Stockwerken, HGRND
- 8 und keinem wird das Ding schaden. PRO
- 9 Im Gegenteil. PRO
- 10 Endlich verschwinden die Ramschläden mit ihren Prepaid-Handykarten und Kochtopfdckeln und das alte Parkhaus gleich mit. PRO
- 11 Ähnlich verhält es sich beim „Alexa“ am Fernsehurm. PRO
- 12 Will denn wirklich jemand den alten Parkplatz zurück, der sich dort befand? PRO
- 13 Vermisst jemand den vermüllten Güterbahnhof, auf dem heute die „Spandau-Arcaden“ stehen? PRO
- 14 Na also. PRO
- 15 Doch nicht nur Lücken in der Stadt werden durch Einkaufszentren geschlossen, sondern auch ganze Viertel aufgewertet PRO
-

-
- | | | |
|----|---|-----|
| 16 | (wie die Gropiusstadt etwa durch die -passagen oder die Gegend am Savignyplatz durch das Stilwerk), | PRO |
| 17 | wovon auch die kleineren Läden im Kiez wieder profitieren, | PRO |
| 18 | weil mehr Kundschaft kommt. | PRO |
| 19 | Gekauft wird dort, wo es den besten Service gibt, die besten Preise. | PRO |
| 20 | Wo es sauber ist, beleuchtet und die Wege kurz sind. | PRO |
| 21 | Und: Wo nicht überall Hunde hinmachen. | PRO |
| 22 | Das tun sie in Einkaufszentren jedenfalls nicht. | PRO |
-

Der erste Schritt besteht in der Identifikation der zentralen These. Nur Segment 5 kommt hierfür in Frage.

Ein Fazit am Ende des Texts, das die These kurz und knapp zusammenfassen würde, ist nicht vorhanden.

Da wenigstens eine These vorhanden ist, kann die Überschrift unannotiert bleiben.

Als nächstes suchen wir Argumente für die These. Es findet sich die Feststellung, dass das neue Einkaufszentrum niemandem schaden, sondern eher nützen wird, in Segment 8 und 9. Dies wird zusätzlich begründet in Segment 10. Weitere Beispiele dafür, dass durch Einkaufszentren unschöne Lücken im Stadtbild geschlossen wurden, folgen in den Segmenten 11–14, verpackt in rhetorische Fragen (die wir im Zuge der Analyse, hier nicht gezeigt, entsprechend umformulieren). Als weiteren Grund für mehr Einkaufszentren wird in Segment 15 der aufwertende Effekt für das umgebende Viertel angeführt und mit Beispielen (16) belegt. Davon sollen auch kleinere Läden profitieren (17–18), womit indirekt ein Einwand in Segment 3 entkräftet wird. Die Segmente 19 bis 22 führen das auf das Konsumentenverhalten zurück und stützen so indirekt die These.

Argumente gegen die These finden sich in den Segmenten 1–3, wo die Gegenposition übersteigert zitiert wird. Diese wird dann in Segment 4 mehr oder weniger begründet verworfen, weswegen dieses Segment letztlich für die eigene zentrale These arbeitet.

Die verbleibenden Segmente 6 und 7 können als Hintergrund markiert werden. Sie dienen dazu, erst einmal ein bestimmtes Beispiel bereitzustellen, anhand dessen sich dann das Argument entfalten kann.

Rhetorische Struktur

Manfred Stede

Hinweise:

- Voraussetzung für die Anwendung der Richtlinien ist, dass der Text bereits in Diskurssegmente unterteilt ist (Kapitel 3).
- Die vorliegenden Richtlinien erläutern die Vorgehensweise der RST-Analyse nur abstrakt. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.
- Alle Textbeispiele ohne Quellenangabe sind fiktiv.
- Für Diskussion und Verbesserungshinweise zu diesen Richtlinien danken wir Gisela Redeker.

10.1 Hintergrund

Die *Rhetorical Structure Theory* (Mann u. Thompson, 1988, Taboada u. Mann, 2006) geht davon aus, dass ein kohärenter Text durch eine Baumstruktur charakterisiert werden kann, deren Blätter die „minimalen Diskurseinheiten“ sind und deren interne Knoten durch je eine *Kohärenzrelation* bezeichnet sind. Vorgeschlagen werden ca. zwei Dutzend solcher Relationen, die also zwischen minimalen Einheiten oder rekursiv zwischen größeren Textabschnitten bestehen können. Für die meisten dieser Relationen gilt, dass die beiden miteinander verbundenen Einheiten nicht „gleichgewichtig“ sind, sondern eines jeweils die zentrale Funktion erfüllt und das andere eine nurmehr unterstützende Rolle spielt. Erstere Einheit heißt dann *Nukleus*, letztere *Satellit*.

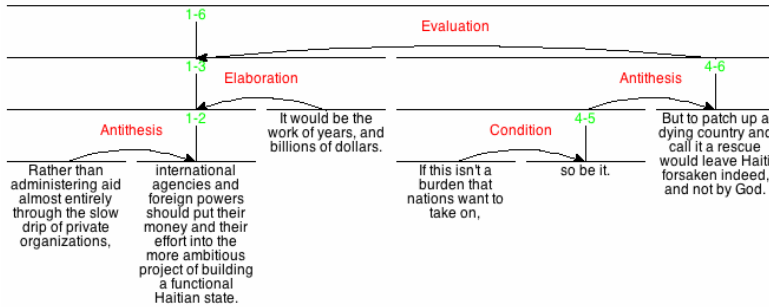


Abbildung 10.2: Analysebeispiel RST (Exzerpt aus: George Packer: *Suffering*. *The New Yorker* 42, 2010)

Zur Illustration zeigt Abb. 10.2 die RST-Analyse eines kurzen Textauszugs, die mit dem Werkzeug *RSTTool*⁸ vorgenommen wurde. Horizontale Linien bilden Textsegmente: entweder elementare (die „Blätter“ des Baumes) oder rekursiv Reihungen von kleineren Segmenten. Gebogene Pfeile verbinden den Satellit einer Relation mit ihrem Nukleus, der zudem durch eine vertikale Linie gekennzeichnet ist. Die Relation ist jeweils in rot angegeben.

Den Status einer solchen Analyse definieren wir als die *Rekonstruktion der Absicht des Autors aus Sicht des Lesers*. RST ist damit ein Instrument zur Explizierung des Textverständnisses: Die Baumstruktur macht klar, was nach Ansicht des Lesers der Autor mit dem Text intendiert hat. Somit ist der RST-Analyse eine gewisse Subjektivität inhärent – verschiedene Leser können durchaus zu etwas unterschiedlichen Analysen gelangen.

10.2 Ziel der Annotation

Die rhetorische Struktur eines Texts besagt, wie nebeneinander stehende Textabschnitte inhaltlich miteinander verbunden sind. Dies

⁸<http://www.wagsoft.com/rsttool> (Zugriff: 30.9.2015)

kann eine relativ lockere Verbindung wie die thematische Fortführung (Relation *Elaboration*) oder eine sehr spezifische, beispielsweise eine *wenn-dann*-Verbindung (Relation *Condition*) sein. Die Menge der Relationen ist im nachfolgenden Abschnitt beschrieben.

Wir gehen hier davon aus, dass die Segmentierung in Diskurs-einheiten bereits vorgenommen wurde (gemäß der Richtlinien in Kapitel 3). Die Abbildung von Diskurssegmenten auf die der RST Analyse zugrunde liegenden *minimal discourse units* besprechen wir in Abschnitt 10.3. Die eigentliche RST-Annotation besteht dann noch aus drei Teilaufgaben:

- Festlegen der *hierarchischen Struktur* des Texts: Welche benachbarten Abschnitte sind miteinander zu verbinden – wie entsteht daraus ein Baum, der den Text komplett überspannt?
- Festlegen der inhaltlichen Relation bei jeder Verbindung zweier Segmente.
- Festlegen des *Nukleus/Satellit*-Status der verbundenen Segmente. Dies ergibt sich bei den meisten Relationen automatisch durch die Wahl der Relation. Umgekehrt kann aber auch der Wunsch nach einer angemessenen Nukleus-Zuweisung die Menge der passenden Relationen einschränken.

Die Teilaufgaben werden in der Praxis nicht getrennt voneinander bearbeitet, sondern sind eng miteinander verwoben. Am Ende der Annotation muss der gesamte Text durch die Baumstruktur abgedeckt sein; es werden stets nur benachbarte Einheiten verbunden, so dass es nicht zu „kreuzenden Kanten“ im Baum kommt.

Der zweite Aspekt der Baum-Eigenschaft: Kein Knoten hat mehr als einen Elternknoten, das heißt: Jedes Segment darf nur eine Rolle für die rhetorische Struktur des Textes spielen. Es kann also nicht gleichzeitig als Satellit zu zwei verschiedenen Nuklei fungieren. Möglich ist allerdings umgekehrt, dass einem Nukleus mehrere verschiedene Satelliten zugeordnet sind, wie in Abb. 10.3 illustriert. Der letzte Satz stellt die zentrale Aussage dar und ist damit der zentrale Nukleus des Textes: Beginnt man oben im Baum, braucht



Abbildung 10.3: Analysebeispiel RST 2 (Text fiktiv)

man lediglich einer einzelnen Nukleus-Verbindung zu folgen, um bei diesem Segment zu landen. Die Motivation für die Aufforderung resultiert unmittelbar aus dem zweiten Satz, dieser ist das zweitwichtigste Segment des Textes. Dies kann durch einen Paraphrasentest überprüft werden, nach dem eine Aneinanderreihung der Nuklei eine Zusammenfassung des Textes ergeben sollte: „Weil Sie beim Zelten in der Natur vorsichtig sein sollen: Bewahren Sie kein Fleisch im Zelt auf.“

Die Motivationsrelation hält gemäß der Analyse zwischen den Sätzen 1–4 (Satellit) und 5 (Nukleus). Nun wird postuliert, dass sie damit aber automatisch auch zwischen dem Nukleus von 1–4 (also 2) und 5 hält. Diese Forderung an RST-Bäume wurde von Marcu (2000) als *strong nuclearity principle* formuliert und spielt gerade für argumentative Texte eine wichtige Rolle.

Innerhalb der Sätze 1–4 sieht die gezeigte Analyse keine weitere Hierarchisierung vor, sondern verbindet 1, 3 und 4 jeweils unmittelbar mit 2. Es ließe sich zwar argumentieren, dass die ersten drei Sätze die Braunbären behandeln und deshalb 1 der Background für 2–3 sein sollte, doch Satz 4 schließt wiederum gleichgewichtig an 2 an, so dass die Bildung eines Segments 1–3 hier strukturell unintuitiv wäre.

Durch die Zuweisung der Nukleus-Rollen erhalten also die einzelnen Textsegmente ein unterschiedliches Gewicht zugewiesen.

Die zentralen Segmente eines Texts nennen wir im Folgenden „stark-nuklear“: Der Weg von der Spitze des Baums zu ihnen führt nur oder fast nur über Nukleus-Verbindungen (im Bild also vertikale Linien).

Als nächstes wird der vorbereitende Schritt der Segmentierung erklärt und anschließend (Abschnitt 10.4) das Inventar der Relationen beschrieben und mit Beispielen illustriert. Abschnitt 10.5 erläutert dann das konkrete Annotationsvorgehen. Abschnitt 10.6 diskutiert schließlich die verschiedenen Annotationsentscheidungen anhand eines vollständigen Beispieltexts aus dem PCC.

10.3 Segmentierung

Die gemäß Kapitel 3 ausgeführte Diskurssegmentierung ist hierarchisch und muss für die RST Analyse nun zu einer flachen Struktur, einer Partitionierung in *elementary discourse units* (im Folgenden: EDUs, markiert durch eckige Klammern), umgestellt werden. Als Grundregel gilt dafür:

- Eine EDU entspricht einer erkennbaren, selbstständigen Sprechhandlung (Illokution). Diese muss aber nicht im engen Sinne strukturell „vollständig“ sein: Etwaige Elisionen sind bei der Beurteilung aufzufüllen, anaphorische (oder kataphorische) Verweise durch ihre Antezedenten zu ersetzen.

Diese Regel soll für die verschiedenen Segmenttypen wie folgt umgesetzt werden:

Einschübe: Aus rein formalen (und auch technischen) Gründen sollen bei der RST-Annotation stets benachbarte EDUs miteinander verbunden werden. Für parenthetische Einschübe, die in der Mitte eines Segments stehen, machen wir daher eine Ausnahme von der Grundregel und markieren sie auch dann *nicht* als EDU, wenn sie eigentlich eine Illokution darstellen. Alle nachfolgenden Anweisungen gelten *modulo* dieser Ausnahmeregelung.

Hauptsätze: Hauptsätze (HS) und -fragmente (HSF) bilden stets eine eigenständige EDU. (HSF stellen dabei eine gewisse „Grauzone“ für die Grundregel dar, da hier das Problem der Elision auftritt.)

Fragmente: FRE und FRB sind auf den Status einer selbstständigen Illokution hin zu überprüfen und dann entweder als eigene EDU zu markieren oder dem benachbarten HS/HSF anzufügen. Beispiele für unterschiedlich zu behandelnde FRE:

(10.208) [Übrigens: Der Bürgermeister wurde wiedergewählt.]

(10.209) [Kaum zu fassen:] [Der Bürgermeister wurde wiedergewählt.]

Im zweiten Fall drückt das FRE eine Beurteilung des nachfolgenden HS aus, der Autor vollzieht damit eine eigenständige Sprechhandlung.

Nebensätze: Für die verschiedenen Arten von Nebensätzen ist wie folgt vorzugehen:

- SUB, OBJ, PRD haben Satzgliedstatus und bilden in der Regel keine eigenständige EDU. Als Ausnahme sind aber Konstruktionen zu behandeln, in denen der Autor eine eingebettete Proposition persönlich bewertet bzw. beurteilt, wie hier:

(10.210) [Es ist kaum zu fassen,] [dass der Bürgermeister wiedergewählt wurde.]

- ARR und ATT dienen „nur“ zur näheren Charakterisierung eines Diskursreferenten, übermitteln also keine eigenständige Informationseinheit, bilden daher keine EDU.
- ANR und WEI kommunizieren selbstständige Informationseinheiten und bilden eine EDU.
- ADV stehen in einer durch den Konnektor markierten inhaltlichen Relation zum übergeordneten HS/HSF, enthalten

damit eine eigenständige Informationseinheit und bilden eine EDU.

- AKP und UNS müssen im Einzelfall anhand der Grundregel beurteilt werden.

Sind Nebensätze koordiniert, sind sie gemäß der obigen Regeln separat auf EDU-Status zu überprüfen. Eine Ausnahme entsteht, wenn koordinierte SUB, OBR oder PRD vorliegen: Diese werden eigentlich nicht als EDU abgetrennt, doch im Fall der Reihung zweier solcher Nebensätze (mit oder ohne Konjunktion) müssen auch zwei EDUs geschaffen werden. Wir trennen nach dem ersten Nebensatz, womit der zweite dann eine eliptische EDU darstellt. Beispiel:

(10.211) [Ich glaube, dass der Bürgermeister wiedergewählt wurde,] [der Stadtkämmerer hingegen nicht genug Stimmen bekommen hat.]

(Falls die Reihung mehr als zwei Nebensätze umfasst, werden auch die weiteren entsprechend abgetrennt.)

Präpositionalphrasen: Im Kapitel 3 bildet ein Verb die Voraussetzung für ein Segment. In der RST-Annotation werden darüber hinaus auch bestimmte Präpositionalphrasen als EDUs betrachtet. Dafür müssen zwei Bedingungen erfüllt sein:

- Die eingebettete Nominalphrase enthält ein Substantiv, mit dem auf einen Sachverhalt referiert wird (z. B., aber nicht unbedingt, ein nominalisiertes Verb).
- Die Präpositionalphrase steht in einer klar erkennbaren Kohärenzrelation (s. nächster Abschnitt) zum umgebenden Haupt- bzw. Nebensatz.

Typische Präpositionen für solche Fälle sind *trotz* und *wegen*, sowie bestimmte Verwendungen von *bei*. Beispiele:

- (10.212) [Der Bürgermeister wurde wiedergewählt,] [trotz einer überaus teuren Kampagne seiner Gegnerin.] (*Relation Konzession*)
- (10.213) [Der Bürgermeister wurde wiedergewählt,] [sicher wegen seiner vielen Auftritte im Radio.] (*Relation Reason*)
- (10.214) [Bei untergehender Sonne] [sollten Sie das Fernglas nicht benutzen.] (*Relation Condition*)

10.4 Relationen

Wir legen der Annotation Relationsdefinitionen zugrunde, die auf denen von der RST Webseite⁹ basieren, aber überarbeitet und durch Hinweise ergänzt werden, die speziell für die Textsorte *Kommentar* relevant sind. Die Relationen werden hier, ähnlich zur RST-Webseite, in Gruppen vorgestellt. (i) Primär *pragmatische* Relationen beschreiben, vor allem in der Textsorte *Kommentar*, die Argumentation des Autors: Was sind Behauptungen, wie werden diese durch Beobachtungen oder durch andere Behauptungen unterstützt? (ii) Primär *semantische* Relationen kommen zum Einsatz, wenn im Text ein (komplexer) Sachverhalt der Welt lediglich beschrieben wird, wobei dann beispielsweise eine kausale Relation zwischen zwei Ereignissen bestehen kann. (iii) *Textuelle* Relationen erfüllen einen organisierenden Zweck, sie erleichtern das Verständnis durch orientierende Information oder Wiederholung. In allen drei Gruppen bestehen die Relationen zwischen einem Nukleus und einem Satelliten; sie werden dementsprechend *mononukleare* Relationen genannt. (iv) *Multinukleare* Relationen hingegen tragen keine solche Gewichtung, sondern verbinden zwei oder auch mehr Nuklei.

Die Unterscheidung zwischen den ersten drei Gruppen ist also inhaltlich motiviert, während die vierte sowohl semantische als

⁹<http://www.sfu.ca/rst> (Zugriff 30.9.2015) – Diese Menge von Definitionen ist in der RST-Literatur als „Extended classical relations“ bekannt.

auch pragmatische Relationen enthält, deren Gemeinsamkeit das strukturelle Merkmal der Multinuklearität ist.

Vor allem infolge der Bereitstellung der verschiedenen Gruppen von Relationen in der RST ist es keineswegs ausgeschlossen, dass in einer Analysesituation für zwei Textsegmente mehr als eine Relation anwendbar erscheint, insbesondere solche aus verschiedenen Gruppen. Hinweise zur Behandlung solcher „Relationsambiguität“ folgen in Abschnitt 10.4.6.

10.4.1 Format der Definitionen

Relationsdefinitionen beschreiben die Rolle der verbundenen Segmente und den vom Autor mit der Verwendung der Relation beabsichtigten Effekt. Darüber hinaus geben wir jeweils Beispiele aus dem PCC, nennen typische Konnektoren zur Signalisierung der Relation und machen Bemerkungen, wiederum primär zur Verwendung für unsere Textsorte. Wir notieren das Format der Definitionen für die mononuklearen Relationen wie folgt:

- **N:** Charakterisierung des Typs und/oder der Funktion des Nukleus
(als Beschreibung der Haltung des Autors, nicht des Ausdrucks im Text)
- **S:** Charakterisierung des Typs und/oder der Funktion des Satellits
(als Beschreibung der Haltung des Autors, nicht des Ausdrucks im Text)
- **N/S:** Charakterisierung der Funktion der Nukleus/Satellit Kombination. Wenn es Beschränkungen oder Tendenzen für die textuelle Abfolge von N und S gibt, sind sie hier ebenfalls genannt.
- **Effekt:** Charakterisierung des mit der Verwendung der Relation vom Autor intendierten Effekts, formuliert als „vorher-nachher“ Veränderung.

- **Typische Konnektoren**

- **Beispiel:** N und S sind jeweils markiert. Wenn vorab ein Kontext charakterisiert wird, geschieht das in *Kursivschrift*. Am Ende ist die Quelle des Beispiels angegeben.
- **Bemerkung:** (optional)

Die Felder N, S und N/S bleiben frei, wenn es für eine Relation keine entsprechenden Beschränkungen gibt. Bei multinuklearen Relationen entfallen die Felder S und N/S.

Achtung: Bei den Beispielen ist mitunter die N/S-Verteilung nur im weiteren Diskurskontext nachvollziehbar, so dass die hier gezeigten Auszüge lediglich illustrativ wirken. (Z. B. bei Cause versus Result, Antithesis versus Contrast).

Bei den Angaben zu Konnektoren ist dreierlei zu beachten:

- Wir nennen hier lediglich Beispiele – in aller Regel gibt es weitere Konnektoren für die Relation, oft sogar viele.
- Konnektoren sind oftmals nicht eindeutig: Einerseits haben viele Wörter neben der Konnektor-Lesart noch weitere (meist als Partikel); andererseits können einige Konnektoren mehrere Relationen anzeigen, so z. B. *schließlich* (temporale Folge, abschließend/argumentierend).
- Manchmal können Konnektoren, die „eigentlich“ Relation X anzeigen, zu einer Relation Y „überinterpretiert“ werden. Zum Beispiel gehen temporale Abfolge und Kausalität miteinander einher, und manchmal ist die Kausalität wichtiger als die temporale Kernbedeutung des Konnektors:

(10.215) [Kaum haben sich ihre Verhandlungsführer in Berlin mit der SPD geeinigt,]_S reiben sich die Genossen in der Region verwundert die Augen.]_N (maz-16294)

Hier signalisiert *kaum* temporale Sequenz, doch die Einigung in Berlin ist wohl der Grund für das Augenreiben in der Region, daher ist die Relation Cause angemessen.

In der Analyse sollte man bei der Prüfung, ob für zwei Textsegmente eine bestimmte Relation anwendbar ist, so vorgehen:

- Gibt es einen Hinweis durch einen Konnektor, der die Relation anzeigt oder die Menge der möglichen Relationen zumindest einschränkt? – In manchen Fällen ist allerdings, wie gerade gesagt, die an der Oberfläche signalisierte Relation nicht die pragmatisch „wichtige“ – dies wird in Abschnitt 10.4.6 noch weiter besprochen.
- Interpunktionszeichen liefern zwar keine sehr klaren Hinweise auf bestimmte Relationen, doch es gibt Tendenzen, wie der Zusammenhang zwischen dem Semikolon und kontrastiven Relationen:

(10.216) [Schon bei der ersten Personalie vertraut er nicht
auf seine Macht;]_S [er sucht den Kompromiss.]_N
(Antithesis; maz-17706)

- Möchte der Autor durch die Juxtaposition der Segmente den jeweils genannten Effekt beim Leser erreichen? Dies ist eine notwendige Bedingung für die Anwendung einer Relation.
- Wenn die Relationsdefinition Beschränkungen für den Typ oder die Funktion von Nukleus, Satellit, oder ihrer Kombination nennt, sind diese erfüllt? Dies sind (soweit vorhanden) ebenfalls notwendige Bedingungen.
- Wird die mit der Relation verbundene Nukleus/Satellit-Verbindung der Rolle beider Segmente für die Textfunktion gerecht? Dies ist ein weniger striktes Kriterium als die beiden vorgenannten, kann aber oft die Entscheidung erleichtern, wenn mehrere Relationen anwendbar erscheinen.

Es folgen nun zunächst die Definitionen der mononuklearen Relationen (bei denen also ein Segment aus der Sicht des Autors jeweils wichtiger ist), die sich in primär pragmatische und semantische, sowie textuelle gliedern; danach werden die Definitionen der multinuklearen Relationen genannt, bei denen kein solches Ungleichgewicht zwischen den Segmenten vorliegt.

10.4.2 Primär pragmatische Relationen

In der Textsorte *Kommentar* werden relativ viele pragmatische Relationen verwendet, denn die Zielsetzung ist ja die Veränderung der Einstellung des Lesers, oder zumindest das Begründen der Einstellung des Autors (und damit verbunden das Werben um Unterstützung für diese Einstellung).

Die Relationen gliedern sich in solche, die dem Leser das Verständnis erleichtern (Background), eine positive Einstellung zu einem Sachverhalt hervorrufen (Antithesis, Concession), eine These des Autors begründen (Evidence, Reason), einen Sachverhalt aus Sicht des Autors bewerten (Evaluation), und Handlungen des Lesers auslösen bzw. erleichtern (Motivation, Enablement).

Background

- **N/S:** Das Verstehen von S erleichtert dem Leser das Verständnis für den Inhalt von N; S enthält orientierende Hintergrundinformation, ohne die N nicht oder nur schwer verständlich wäre. Im Text geht S meist dem N voraus, aber nicht immer. Ein Background-Satellit am Anfang eines Textes hat oftmals auch die Funktion, das Thema kurz einzuführen.
- **Effekt:** Die Fähigkeit des Lesers, den Inhalt von N zu verstehen, wird verbessert.
- **Typische Konnektoren:** (selten durch Konnektoren angezeigt)
- **Beispiel:** [Burkina Faso hieß bis 1984 noch Obervolta.]_S [Nach einer EMNID-Umfrage glauben viele Europäer bis heute, dass es sich um zwei verschiedene Länder handelt.]_N
- **Bemerkung:** Die Relation besteht eher selten zwischen EDUs, sondern in der Regel zwischen größeren Segmenten. Viele Kommentare sind so strukturiert, dass ein Background-Satellit den Textanfang bildet, also den inhaltlichen Ausgangspunkt für die nachfolgende Kommentierung darstellt.

Antithesis

- **N:** Autor schätzt den Inhalt von N als positiv oder wichtig ein.
- **S:** Autor schätzt den Inhalt von S (gegenüber N) als weniger positiv oder weniger wichtig ein.
- **N/S:** Die Inhalte von N und S sind nicht miteinander kompatibel – meist auf der Ebene ihrer *Bewertung*. Wegen der Inkompatibilität kann man nicht sowohl N als auch S gleichermaßen als positiv oder wichtig einschätzen. Eine Untergruppe bilden (oft durch *sondern* markierte) Korrektur-Beschreibungen, s. zweites Beispiel unten. Im Text geht S zumeist dem N voraus.
- **Effekt:** Leser schätzt den Inhalt von N positiv(er) oder wichtig(er) ein.
- **Typische Konnektoren:** *aber; jedoch; nicht–sondern; anstatt–zu; ...*
- **Beispiele:** *Versuch des Potsdamer Amtsgerichts, ein Grundstück zu verkaufen* [Einmal schien es schon geglückt zu sein.]_S [Doch der Käufer zahlte nicht.]_N (maz-6193)
[Ablehnung gab es nämlich nicht für den Mann,]_S [sondern für seinen Posten, der ihn zum Befürworter des Bombodroms machen muss.]_N (maz-17300)
- **Bemerkung:** Verwandte, aber abzugrenzende Relationen sind Concession und multinuklearer Contrast. Siehe dazu auch Abschnitt 10.4.6.

Concession

- **N:** Autor schätzt den Inhalt von N als positiv oder wichtig ein.

- **S:** Autor schätzt den Inhalt von S als weniger positiv oder wichtig ein, bestreitet ihn aber nicht.
- **N/S:** Autor konzidiert S und bestätigt implizit, dass S und N gewöhnlich/erwartungsgemäß nicht miteinander kompatibel sind; im aktuellen Fall sind sie jedoch kompatibel, und N wird betont.
- **Effekt:** Leser schätzt den Inhalt von N positiv(er) oder wichtig(er) ein.
- **Typische Konnektoren:** *zwar-aber; obwohl; trotzdem; dennoch; dabei; ..*
- **Beispiel:** [Sanitäre Anlagen gehören heute zum Standard großer Liegeplätze.]_S [Doch mit der Nachrüstung tut sich Radewege schwer.]_N (maz-6488)

Evidence

- **N:** Eine subjektive Aussage/Einschätzung/These, die der Leser möglicherweise nicht akzeptiert oder als nicht genügend wichtig oder positiv einschätzt.
- **S:** Eine Aussage, die der Leser wohl akzeptieren wird; in der Regel die „objektive“ Beschreibung eines Faktums.
- **N/S:** Durch das Verstehen von S akzeptiert der Leser die Aussage von N leichter, bzw. teilt die damit verbundene Einschätzung des Autors.
- **Effekt:** Leser glaubt eher, dass die in N getroffene Aussage zutrifft.
- **Typische Konnektoren:** (kausale Konnektoren)
- **Beispiel:** *Streit um das Nebeneinander der Schulfächer Religion und LER* [Und nun scheint sogar unsere Landesregierung

entschlossen, diese scheinbare Gleichbehandlung der beiden Fächer zu beseitigen.]_N [Stolpe, Reiche und Co. sagen zwar Ja zu einem möglichen Kompromissangebot aus Karlsruhe, dekretieren aber: Einen Wahlpflichtbereich LER/Religion kann es nicht geben.]_S (maz-6159)

- **Bemerkung:** Evidence verbindet oft ein längeres Satellit-Segment mit einem kürzeren Nukleus (der These).
-

Reason

- **N:** Eine subjektive Aussage/Einschätzung/These, die der Leser möglicherweise nicht akzeptiert oder als nicht genügend wichtig oder positiv einschätzt.
 - **S:** Eine subjektive Aussage/Einschätzung/These
 - **N/S:** Durch das Verstehen von S akzeptiert der Leser die Aussage von N leichter, bzw. teilt die damit verbundene Einschätzung des Autors.
 - **Effekt:** Leser glaubt eher, dass die in N getroffene Aussage zutrifft.
 - **Typische Konnektoren:** (kausale Konnektoren)
 - **Beispiel:** [Die Militärschläge der USA verlieren mit jedem Tag weiteren Bombardements schleichend an Glaubwürdigkeit.]_N [Mit flächenhafter Zerstörung sind weder die Taliban noch bin Laden zu treffen.]_S (maz-5701)
 - **Bemerkung:** Reason ist spezifischer als Evidence: der Unterschied besteht darin, ob S vom Autor als „objektiv“ dargestellt wird (Evidence) oder selbst eine subjektive Einschätzung ist (Reason).
-

Reason-N

- **N:** Eine subjektive Aussage/Einschätzung/These
- **S:** Eine subjektive Aussage/Einschätzung/These, die der Leser möglicherweise nicht akzeptiert oder als nicht genügend wichtig oder positiv einschätzt.
- **N/S:** Durch das Verstehen von N akzeptiert der Leser die Aussage von S leichter, bzw. teilt die damit verbundene Einschätzung des Autors.
- **Effekt:** Leser glaubt eher, dass die in S getroffene Aussage zutrifft.
- **Typische Konnektoren:** (kausale Konnektoren)
- **Bemerkung:** Dual zu Reason. Diese „Umdrehung“ von Reason wird in Kommentaren nur selten anzutreffen sein. Der Grund spielt für den Text eine wichtige Rolle, während die Folge eher nebensächlich ist.

Justify

- **N:** Eine subjektive Aussage/Einschätzung/These, die der Leser möglicherweise nicht akzeptiert
- **S:** Eine subjektive Darlegung einer (grundsätzlichen, z. B. moralischen) Haltung, oder ein Verweis auf eine bestimmte Position des Handelnden
- **N/S:** Das Verstehen von S erhöht die Bereitschaft des Lesers, die Präsentation von N durch den Autor zu akzeptieren
- **Effekt:** Leser hat eher Verständnis für die in N vorgebrachte Aussage/Einschätzung/These
- **Typische Konnektoren:** (kausale Konnektoren)

- **Beispiel:** [Dieses Gericht werde ich niemals essen.]_S [Ich bin Vegetarier.]_N
-

Evaluation-S

- **N:** Beschreibung eines Sachverhalts, oder eine subjektive Aussage (allerdings nicht aus Perspektive des Autors)
 - **S:** Eine subjektive Bewertung (positiv/negativ, erstrebenswert/nicht erstrebenswert) aus Perspektive des Autors
 - **N/S:** S bewertet N
 - **Effekt:** Leser erkennt die Bewertungsrelation zwischen S und N
 - **Typische Konnektoren:** (selten durch Konnektoren angezeigt)
 - **Beispiel:** [Seine Vergangenheit schien wie ein Fluch über dem Hotelkomplex zu liegen.]_S [Jahrelang hatte das Amtsgericht Potsdam umsonst versucht, es an den Mann zu bringen.]_N (maz-6193)
 - **Bemerkung:** Meist folgt das evaluierende Segment auf das evaluierte; manchmal ist es jedoch umgekehrt, wie im obigen Beispiel.
-

Evaluation-N

- **N:** Eine subjektive Bewertung (positiv/negativ, erstrebenswert/nicht erstrebenswert) aus Perspektive des Autors
- **S:** Beschreibung eines Sachverhalts, oder auch subjektive Aussage – allerdings nicht aus Perspektive des Autors
- **N/S:** N bewertet S

- **Effekt:** Leser erkennt die Bewertungsrelation zwischen S und N
 - **Typische Konnektoren:** (selten durch Konnektoren angezeigt)
 - **Bemerkung:** Dual zu Evaluation-S. Die Entscheidung zwischen den beiden erfolgt ausschließlich aus der Erwägung der relativen Wichtigkeit der Segmente für den Text.
-

Motivation

- **N:** Eine vom Leser auszuführende Tätigkeit.
 - **N/S:** S liefert ein Motiv für die Ausführung der in N beschriebenen Tätigkeit.
 - **Effekt:** Das Verstehen von S erhöht die Bereitschaft des Lesers, die in N genannte Tätigkeit auszuführen.
 - **Typische Konnektoren:** (kausale Konnektoren)
 - **Beispiel:** [Olympische Spiele stellen für jeden Austragungsort einen großen Gewinn dar.]_S [Berlin muss sich jetzt um die Spiele 2016 bewerben!]_N
 - **Bemerkung:** Diese Relation tritt in der Textsorte Kommentar selten auf; es kann manchmal vorkommen, dass „alle Leser“ der Zeitung metaphorisch zu einer Handlung ermuntert werden, wie im obigen Beispiel.
-

Enablement

- **N:** Eine vom Leser auszuführende Tätigkeit.
- **N/S:** Das Verstehen von S erleichtert dem Leser, die von N beschriebene Tätigkeit auszuführen.

- **Effekt:** Die Fähigkeit des Lesers, die Tätigkeit in N auszuführen, wird gesteigert.
 - **Typische Konnektoren:** *damit; ...*
 - **Beispiel:** [Wechseln Sie die Zündkerzen aus.]_N [Ein Vierkant-schlüssel befindet sich unter der Abdeckung.]_S
 - **Bemerkung:** Diese Relation tritt in der Textsorte Kommentar selten auf.
-

10.4.3 Primär semantische Relationen

Circumstance

- **S:** Ein (tatsächlich eingetretenes/eintretendes, also nicht hypothetisches) Ereignis bzw. ein Zustand.
 - **N/S:** S setzt einen Rahmen, in dem N zu interpretieren ist, z. B. die zeitliche oder räumliche Einordnung.
 - **Effekt:** Leser erkennt, dass S den Rahmen für die Interpretation von N liefert.
 - **Typische Konnektoren:** *als; während; wenn; ...* für zeitlichen Rahmen
 - **Beispiel:** [Als die Veag wegen der Liberalisierung des Strommarktes unter Druck geriet,]_S [hielt sie sich bei ihren Lieferanten schadlos.]_N (maz-5297)
-

Condition

- **S:** Eine hypothetische, künftige oder anderweitig irrealere Situation.

- **N/S:** Das Eintreten/die Realisierung von N hängt vom Eintreten/der Realisierung von S ab.
- **Effekt:** Leser erkennt, dass N von S abhängig ist.
- **Typische Konnektoren:** *wenn .. dann; sofern, ...*
- **Beispiele:** [Wenn immer mehr Kommunen finanziell ausbluten,]_S [wird die regelmäßige Schulbucheuerung zur Illusion.]_N (maz-00002)
[Kommt der Sanitärtrakt nicht spätestens zur nächsten Saison,]_S [droht Radwege den Wettbewerb um Freizeitkapitäne zu verlieren.]_N (maz-6488)
- **Bemerkung:** Der Konnektor *wenn* (ohne *dann*) kann sowohl Circumstance als auch Condition anzeigen. Entscheidend für Condition ist, ob N abhängig von einem hypothetischen Ereignis in S ist.

Otherwise

- **N:** Eine hypothetische, künftige oder anderweitig irrealer Situation.
- **S:** Eine hypothetische, künftige oder anderweitig irrealer Situation.
- **N/S:** Die Realisierung von N verhindert die Realisierung von S.
- **Effekt:** Leser erkennt den Verhinderungszusammenhang zwischen N und S.
- **Typische Konnektoren:** *ehe; bevor; sonst; ...*
- **Beispiele:** [Die Stadt wäre gut beraten, sich diesem moralischen Imperativ zu beugen.]_N [Anderenfalls müsste sie vor jeder Schule und jeder Kita Bremsbuckel aufpflastern.]_S (maz-5709)

[Der Gemeinderat sollte den unseligen Beschluss aufheben,]_N
[ehe die Bürger sich mit Taschenlampen versorgen.]_S (maz-15609)

Unless

- **N:** Eine hypothetische, künftige oder anderweitig irrealer Situation.
 - **S:** Eine hypothetische, künftige oder anderweitig irrealer Situation.
 - **N/S:** S beeinflusst die Realisierung von N: N wird nur realisiert, wenn S nicht realisiert wird.
 - **Effekt:** Leser erkennt die Abhängigkeit Realisierung von N von der Nicht-Realisierung von S.
 - **Typische Konnektoren:** *es sei denn; ...*
 - **Beispiel:** [Morgen wird der Satellit in den Pazifik stürzen.]_N
[Es sei denn, er verglüht doch noch vollständig in der Erdatmosphäre.]_S
-

Elaboration

- **N/S:** S liefert genauer Information bzw. Details zum Inhalt von N – jedoch nicht nur zu einem einzelnen, in N genannten Gegenstand (vgl. E-Elaboration). N geht S im Text voraus. Typische Zusammenhänge zwischen N und S sind Menge::Element, Ganzes::Teil, Abstraktion::Instanz, Vorgang::Einzelschritt.
- **Effekt:** Leser erkennt, dass S genauere Information zu N liefert.
- **Typische Konnektoren:** *besonders; beispielsweise; ...*

- **Beispiel:** [Diepensee siedelt um.]_N [Ohne Wenn und Aber.]_S
(maz-6993)
-

E-Elaboration

- **N/S:** S liefert genauere Information bzw. Details zu einem einzelnen, in N genannten Gegenstand. N geht S im Text voraus.
 - **Effekt:** Leser erkennt, dass S genauere Information zu einem in N genannten Gegenstand liefert.
 - **Typische Konnektoren:** (selten durch Konnektoren angezeigt)
 - **Beispiel:** [Heute wollen Tausende das Flair eines historischen Klassenzimmers kennen lernen,]_N [wie sie nur noch wenige in Deutschland existieren.]_S (maz-6728)
 - **Bemerkung:** Im Beispiel bezieht sich der *wie*-Nebensatz (S) nicht auf das gesamte Kennenlern-Ereignis, sondern nur auf das *Klassenzimmer*, also einen Gegenstand aus N.
-

Interpretation

- **N/S:** S leitet den Inhalt von N in einen anderen konzeptuellen Zusammenhang über. Dabei wird aber keine Bewertung vorgenommen, bzw. die Bewertung steht nicht im Vordergrund des Interesses. N geht S im Text voraus.
- **Effekt:** Leser erkennt den inhaltlichen, deutenden Zusammenhang zwischen N und S.
- **Typische Konnektoren:** *damit; mithin; ...*
- **Beispiel:** [In Potsdam findet heute der große Festzug zu Ehren von Friedrich II. statt.]_N [Damit sind die diesjährigen „Preußen-Tage“ beendet.]_S

- **Bemerkung:** Sobald S eine wesentliche Bewertung von N vornimmt, ist nicht Interpretation, sondern Evaluation zuzuweisen.
-

Means

- **N:** Eine Handlung/ Aktivität
 - **N/S:** S gibt Informationen, die die Realisierung/ Ausführung von N wahrscheinlicher/ einfacher machen (z. B. ein Instrument)
 - **Effekt:** Leser erkennt den Zusammenhang der höheren Wahrscheinlichkeit oder der Vereinfachung der Handlungsausführung
 - **Typische Konnektoren:** *dazu; damit; ...*
 - **Beispiele:** [Berliner fahren im August immer gern nach Lichtenrade.]_N [Dazu nehmen sie meistens die S25.]_S
[Der Parteivorsitzend hat gleich zwei CDU-Bürgermeister gegen sich aufgebracht]_S [und wird damit Ansehen und Rückhalt verlieren.]_N (maz-18160)
-

Cause

- **N:** ein realer Sachverhalt in der Welt.
- **S:** ein realer Sachverhalt in der Welt.
- **N/S:** der in N beschriebene Sachverhalt wird durch den in S beschriebenen Sachverhalt verursacht.
- **Effekt:** Leser erkennt den Kausalzusammenhang in der Welt.
- **Typische Konnektoren:** *weil; da; deshalb; ...*

- **Beispiel:** *Unerwartete Pressemeldung zur Verlegung des Ortes Diepensee.* [Überrascht reagierte auch Bürgermeister Jochen Wagner.]_N [Schließlich gaben die Stadtverordneten erst Montagabend grünes Licht für die weitere Erschließung des neuen Ortsteils Diepensee.]_S (maz-6993)
-

Result

- **N:** ein realer Sachverhalt in der Welt.
 - **S:** ein realer Sachverhalt in der Welt.
 - **N/S:** der in S beschriebene Sachverhalt wird durch den in N beschriebenen Sachverhalt verursacht.
 - **Effekt:** Leser erkennt den Kausalzusammenhang in der Welt.
 - **Typische Konnektoren:** *weil; da; deshalb; ...*
 - **Bemerkung:** Dual zu Cause. Die Entscheidung zwischen den beiden erfolgt ausschließlich aus der Erwägung der relativen Wichtigkeit der Segmente für den Text.
-

Purpose

- **N:** Eine Aktivität/Handlung.
- **S:** Eine hypothetische, künftige oder anderweitig irrealer Situation
- **N/S:** S wird durch die Ausführung von N realisiert
- **Effekt:** Leser erkennt den Zusammenhang der Realisierung von S durch die Ausführung von N.
- **Typische Konnektoren:** *durch; um .. zu; ...*

- **Beispiel:** [Um ihre Soldaten zu schützen,]_S [äußern sich die USA nicht zu Berichten über den Beginn der Boden-Operation in Afghanistan.]_N (maz-5701)
 - **Bemerkung:** Es besteht im weitesten Sinne ein Kausalzusammenhang; der Unterschied zu den Cause/Result Relationen besteht darin, dass S im Fall von Purpose klar als hypothetisch/unrealisiert signalisiert ist und die Intention des/der Handelnden darstellt.
-

Solutionhood

- **S:** Der Inhalt von S kann als „Problem“ aufgefasst werden.
 - **N/S:** Der Inhalt von N kann als Lösung des in S dargestellten Problems aufgefasst werden. N geht in der Regel S im Text voraus.
 - **Effekt:** Leser erkennt N als Lösung des Problems in S.
 - **Typische Konnektoren:** (selten durch Konnektoren angezeigt)
 - **Beispiel:** [Mit der Verabschiedung des Nichtraucherchutzgesetzes sitzen viele Kneipen in der Falle.]_S [Es empfiehlt sich, früh genug auf die Einrichtung abtrennbarer Räume zu achten.]_N
-

10.4.4 Textuelle Relationen

Die folgenden drei Relationen dieser Gruppe erfüllen primär eine organisierende Funktion für den Text und sein Verständnis, sind also nicht im engeren Sinne als semantisch (die Welt beschreibend) oder pragmatisch (die Einstellung des Lesers verändernd) einzustufen.

Preparation

- **N/S:** S geht N im Text voraus. S bereitet den Leser auf das Thema von N vor (ohne im engeren Sinne verständnisleichternde Information zu liefern).
 - **Effekt:** Leser wird durch S auf die Information in N vorbereitet.
 - **Typische Konnektoren:** (selten durch Konnektoren angezeigt; gelegentlich durch Doppelpunkt)
 - **Beispiel:** [Denn eines ist klar:]_S [Die Militärschläge der USA verlieren mit jedem Tag weiteren Bombardements schleichend an Glaubwürdigkeit.]_N (maz-5701)
 - **Bemerkung:** Die Relation ist zuzuweisen, wenn S keine „stärkere“ Funktion erfüllt als das Thema für N zu setzen, oder wenn S eine einleitende oder anderweitig gliedernde Formel enthält. S darf also keine (oder nur minimale) eigenständige Information übermitteln.
-

Restatement

- **N/S:** S folgt im Text auf N und wiederholt die Information von N „mit anderen Worten“. N und S haben ungefähr gleichen Umfang.
 - **Effekt:** Leser erkennt die wiederholende Funktion von S.
 - **Typische Konnektoren:** *anders ausgedrückt; ...*
 - **Beispiel:** [Der Bürgermeister hat den Stadtverordneten alle Informationen gegeben,]_N [also gewissermaßen die Karten auf den Tisch gelegt.]_S
-

Summary

- **N:** N umfasst mehr als nur eine EDU.
- **N/S:** S folgt im Text auf N und wiederholt die Information von N, ist jedoch kürzer.
- **Effekt:** Leser erkennt die zusammenfassende Funktion von S.
- **Typische Konnektoren:** *in Kürze; ...*

10.4.5 Multinukleare Relationen

Unter den multinuklearen Relationen ist Sequence als semantisch einzustufen, während alle anderen je nach Inhalt der Nuklei entweder semantisch oder pragmatisch sein können.

Contrast

- **N:** Genau zwei Nuklei. Beide sind gleichermaßen „wichtig“. Die Inhalte sind einander ähnlich, miteinander vergleichbar; sie sind aber nicht identisch, sondern unterscheiden sich in für den Autor wichtigen Aspekten.
- **Effekt:** Leser erkennt die Vergleichbarkeit der Nuklei und die Betonung des Unterschieds.
- **Typische Konnektoren:** *demgegenüber; hingegen; während; aber; ...*
- **Beispiel:** [Mein erstes Auto war ein Kleinwagen.]_N [Das zweite hingegen ein ausgewachsener Kombi.]_N

Sequence

- **N:** Die Nuklei beschreiben Sachverhalte der Welt, die in einer bestimmten zeitlichen Abfolge stattfinden.

- **Effekt:** Leser erkennt die temporale Relation zwischen den Nuklei.
 - **Typische Konnektoren:** *dann; anschließend; und; zuvor, ...*
 - **Beispiel:** [Um neun betrat die Lehrerin den Klassenraum.]_N [Fünf Minuten später verkündete sie, dass ein Test geschrieben wird.]_N
 - **Bemerkung:** Die Nennung der Ereignisse kann der zeitlichen Abfolge entsprechen (*dann*) oder gegenäufig sein (*zuvor*).
-

List

- **N:** Die in bestimmtem Sinne typgleichen Nuklei geben Informationen, die als zusammengehörig, aufzählend erkennbar sind, mithin eine gemeinsame Rolle für die Textfunktion spielen. Sie sind aber nicht alle durch Konjunktionen verbunden.
 - **Effekt:** Leser erkennt die gemeinsame Funktion der Nuklei.
 - **Typische Konnektoren:** Komma, Nummerierungen, ...
 - **Beispiel:** Was ich gestern getan habe? [Essen kochen,]_N [Kinder versorgen,]_N [Bad putzen.]_N
 - **Bemerkung:** Zur Behandlung elliptischer Formen wie im hier genannten Beispiel siehe Abschnitt 10.4.6.
-

Conjunction

- **N:** Die in bestimmten Sinne typgleichen zwei Nuklei geben Informationen, die als zusammengehörig, aufzählend erkennbar sind. Sie sind durch eine nebenordnende Konjunktion verbunden.
- **Effekt:** Leser erkennt die gemeinsame Funktion der Nuklei.

- **Typische Konnektoren:** *und; oder; je A, desto B; ...*
 - **Beispiel:** Sollte der Luftkrieg als bloße Drohkulisse gedacht gewesen sein, [verlöre er spätestens mit dem einsetzenden Winter und der unausbleiblichen humanitären Katastrophe seine Legitimation.]_N [Und die USA ihren Rückhalt in Westeuropa.]_N (maz-2318)
 - **Bemerkung:** Funktional sind List und Conjunction identisch; wenn die Oberflächen-Bedingung für Conjunction erfüllt ist, so ist diese Relation zuzuweisen.
-

Joint

- **N:** Die nicht unbedingt typgleichen Nuklei geben separate Informationen, stehen aber in keiner klar identifizierbaren semantischen oder pragmatischen Relation zueinander, haben gemeinsam auch nicht den Charakter einer Aufzählung. Nichtsdestotrotz besteht eine kohärente Verbindung, weil sie der übergeordneten Textfunktion gleichermaßen dienen.
- **Effekt:** Leser erkennt, dass jeder Nukleus „eine eigene Botschaft“ hat, die aber jeweils derselben Textfunktion dienlich sind.
- **Typische Konnektoren:** additive Konnektoren wie *zudem, auch*
- **Bemerkung:** Joint ist zu verwenden, wenn eine multinukleare Relation gesucht wird und keine der übrigen passt.

10.4.6 Hinweise zur Relationszuweisung

In vielen Fällen wird die Entscheidung für eine Relation nicht einfach sein, weil Mehrdeutigkeit vorliegt: Mehrere Relationen scheinen gleichermaßen verwendbar. Zudem sind auch die Entscheidungen über rekursive Einbettung, also die Hierarchisierung,

nicht immer klar. Für solche Fälle gibt dieser Abschnitt Hinweise zum Vorgehen.

10.4.6.1 Hierarchisierung

(1) Die Behandlung eliptischer Konstruktionen ist weitgehend mit der Segmentierung bereits entschieden, sei aber hier an einem Beispiel erläutert.

(10.217) Niemand weiß, warum sie nicht kommt, wer sie vertritt, wann oder ob sie wiederkommt. (maz-5876)

Die modale Einbettung *niemand weiß*, hat Skopus über den gesamten Satz. Dieser wird nur weiter segmentiert, wenn klare Hinweise auf eine Relation vorliegen, was hier durch die Aufzählung einzelner Propositionen gegeben ist. Segmentierungskriterium ist, dass eine Proposition im Wesentlichen vollständig ist. Dies ist bei *warum sie nicht kommt* oder *wer sie vertritt* gegeben, nicht aber beim stark eliptischen *wann* – das einzelne Fragepronomen wird nicht als separates Segment gesehen. Die Analyse ist demzufolge eine List-Relation mit drei Segmenten: (Niemand .. nicht kommt), (wer sie vertritt), (wann oder ob sie wiederkommt).

(2) Ein Satzendezeichen hat in der Regel, jedoch nicht immer Priorität bei der Hierarchisierung! Im folgenden Fall beginnt der Nukleus der Condition mit *verlöre* und endet mit *Westeuropa*; die beiden Conjunction-Nuklei sind also in den Condition-Nukleus eingebettet:

(10.218) [Sollte der Luftkrieg als bloße Drohkulisse gedacht gewesen sein,]_S [[verlöre er spätestens mit dem einsetzenden Winter und der unausbleiblichen humanitären Katastrophe seine Legitimation.]_N [Und die USA ihren Rückhalt in Westeuropa.]_N]_N

(3) Eine für die Hierarchisierung mehrdeutige Situation liegt vor, wenn beispielsweise eine These durch mehr als ein nachfolgendes Argument „abgestützt“ wird. Diese könnten entweder als einzelne Satelliten mit der These verbunden werden, oder zunächst durch

List (oder Conjunction, siehe unten) zusammengefasst werden, so dass es für die Abstützung nur einen Satelliten gibt. Wir legen fest: Wenn die Satelliten über dieselbe Relation anzubinden wären (z. B. Evidence), so hat ihre Zusammenfassung durch List/Conjunction Vorrang vor der Einzel-Anbindung.

10.4.6.2 Textsortenspezifität: Nuklearität sorgfältig zuweisen

Diese Richtlinien sind primär für die Annotation von Kommentaren konzipiert, so dass dem RST-Baum die Aufgabe zufällt, den vom Leser rekonstruierten „kommunikativen Plan“ des Autors abzubilden. Kommentare sind Meinungstexte und verfolgen in der Regel das Ziel, dem Leser eine bestimmte Meinung oder Sichtweise nahe-zulegen. Die RST-Struktur soll diese Absicht des Autors möglichst gut repräsentieren. Das bedeutet einerseits, dass Entscheidungen über Nuklearität konsequent die von Mann/Thompson so genannte *centrality for the author's purposes* darstellen sollen: Was für die kommunikative Absicht des Autors wichtig ist, sollte – möglichst auf allen Ebenen des Baums – Nukleus-Status besitzen. Dabei ist auch das von Marcu (2000) benannte *strong nuclearity principle* zu beachten, nach dem Nuklearitätsstatus im Baum propagiert wird. Dies ist sowohl *bottom up* als auch *top down* zu deuten: Ein als für den Text zentral eingestuftes Segment sollte nicht nur bei der unmittelbaren Verbindung zu Nachbarsegmenten nuklear sein, sondern das entstehende größere Segment sollte seinerseits dann ebenfalls die Rolle eines Nukleus besitzen. Umgekehrt betrachtet gilt, dass eine Relation zwischen zwei größeren Segmenten stets auch (oder gerade) zwischen den Nuklei der beiden Segmente bestehen soll.

Mitunter nimmt der Autor eines Kommentars auch eine eher lockere Aneinanderreihung von gleichermaßen wichtigen Aussagen vor, so dass eine klare Nukleus/Satellit Trennung willkürlich erschiene. Dann ist von den multinuklearen Relationen Gebrauch zu machen.

10.4.6.3 Textsortenspezifität: Vorrang für pragmatische Relationen

Im Sinne des eben genannten Primats für den „kommunikativen Plan“ ist neben der Nuklearitätsentscheidung auch eine generelle wichtigere Rolle für die pragmatischen Relationen anzunehmen. Wenn also beim Annotieren Mehrdeutigkeit zwischen einer pragmatischen und einer semantischen Relation festgestellt wird, ist stets die pragmatische zu bevorzugen, weil sie einen wertvolleren Beitrag bei der Rekonstruktion des Plans des Autors liefert.

10.4.6.4 Informationsreiche vor informationsarmer Relation

Ein letztes Prinzip dient vor allem der Entscheidung innerhalb der semantischen Relationen, deren Definitionen ein unterschiedliches Maß an Informativität im Sinne von implizit enthaltener Bedeutung aufweisen. So besagt Elaboration lediglich, dass das Thema des bisherigen Diskurses beibehalten und weiter spezifiziert wird. Wenn dies der Fall ist, darüber hinaus aber z. B. auch eine Kausalitätsbeziehung besteht, ist auf jeden Fall Cause/Result zu bevorzugen. In ähnlicher Weise lassen sich auch Unterschiede zwischen leicht „verwechselbaren“ semantischen/pragmatischen Relationen charakterisieren. So ist Circumstance relativ informationsarm, besagt lediglich das Vorliegen eines Begleitumstands; die Anwendbarkeit dieser Relation wird häufig mit der von Background überlappen, die jedoch die Verständniserleichterung ins Zentrum rückt: Die Information im Nukleus wäre erkennbar schlechter zu begreifen, wenn der Satellit nicht vorliegen würde. Background ist damit die informationsreichere Relation und im Zweifel gegenüber Circumstance zu bevorzugen.

10.4.6.5 Differenzierung verwandter Relationen

Adversativität/Kontrast: Die erste Entscheidung zwischen den bedeutungsähnlichen Relationen Contrast, Antithesis und Concession ist die der Multinuklearität. Kommt keinem der Segmente eine

größere „Wichtigkeit“ zu, so liegt Contrast vor. Unter den beiden mononuklearen Kandidaten ist Concession die etwas spezifischere Relation, die das überraschende Korrelieren zweier Handlungen oder Sachverhalte ausdrückt. Der Paraphrasentest mit *obwohl* ist in der Regel ein zuverlässiger Indikator: Wenn dieser Konnektor die beiden Segmente verbinden kann, liegt Concession vor; ansonsten die eher auf der Bewertungs- als auf der Sacherhaltsebene anzusetzende Relation Anthithesis.

Kausalität/Begründung: Eine ganze Reihe von Relationen lässt sich dem Bedeutungsfeld der Kausalität und Begründung zuordnen. Es gilt zunächst die Grundteilung in primär-semantische (Bericht über Kausalität „in der Welt“) und pragmatische (Begründung als Teil einer Argumentation mit dem Ziel der Einstellungsveränderung) Relationen. Innerhalb der semantischen Relationen sind Condition, Otherwise, Unless und Purpose für die Fälle eines nur hypothetischen oder künftigen, in diesem Sinne *nicht-faktischen* Ursache-Wirkung-Zusammenhangs vorgesehen („Wenn morgen die Sonne scheint, machen wir eine Radtour“). Purpose hat dabei eine „finale“ Bedeutung: der Nukleus bezeichnet eine Handlung, die zum Zweck des im Satelliten genannten Sachverhalts ausgeführt wird. Die Unterschiede zwischen Condition/Otherwise/Unless ergeben sich aus der Rolle von Negationen, wie in den Definitionen dargestellt. Werden Ursache und Wirkung aber als faktisch („real“) dargestellt, so sind Cause und Result zu verwenden, die sich allein durch die Wahl von Nukleus und Satellit unterscheiden.

Von den pragmatischen Relationen wird Motivation nur verwendet, wenn die Lesenden zu einer bestimmten Tätigkeit (Nukleus) aufgefordert werden und dies durch den Satellit begründet wird, dem mithin eine motivierende Rolle zukommt. Evidence/Reason/Justify zielen auf eine Veränderung der Einstellung der Lesenden und unterscheiden sich im Hinblick auf den Status der Begründung: Bei Evidence wird das Argument durch eine als mehr oder weniger „objektiv“ dargestellte Information über die Welt abgestützt („Partei X hat die Hälfte ihrer Mitglieder verloren

→ Partei X hat bei der kommenden Wahl keine Chance“). Bei Reason handelt es sich stattdessen um eine subjektive Einschätzung oder These des Autors („Partei X wird bei der kommenden Wahl keine Chance haben → Man sollte Partei X nicht wählen“). Justify ist dem ähnlich, doch ist der Grund keine spezielle These, sondern eine grundsätzliche Haltung des Autors (eine politische Einstellung, moralische Position, etc.), aus der heraus sich die Folgerung ergibt.

List/Conjunction/Joint: Diese drei Relationen haben recht ähnliche Anwendungsbereiche. Maßgeblich für List und Conjunction ist die Parallelität der Funktion der einzelnen Segmente: Sie müssen für die Intention des Gesamtsegments jeweils dieselbe Rolle spielen (prototypisch sind Aufzählungen von bestimmten Sachverhalten, die alle dieselbe Idee illustrieren sollen). Conjunction kommt zum Einsatz, wenn ein lexikalisches Signal (ein Konnektor wie *und*, *außerdem*, etc.) die Relation markiert; dies kann innerhalb eines Satzes, aber auch zwischen zwei Sätzen vorliegen. Gibt es keinen Konnektor, ist List zu verwenden. Für Joint gilt die Parallelitätsbedingung nicht: Diese Relation greift immer, wenn der Text zwei oder mehr „Punkte macht“, die in relativ loser Folge nebeneinanderstehen, aber nicht dieselbe Funktion für das übergeordnete Segment haben. Zudem muss natürlich das „Gewicht“ der Einzelsegmente gleich sein – sonst wäre ja eine geeignete mononukleare Relation zu wählen. Joint wird relativ selten auftreten; angebracht ist sie normalerweise ausschließlich bei der Verbindung größerer Textsegmente.

10.5 Annotationsverlauf

Beginnen Sie mit der Annotation erst, nachdem Sie den gesamten Text gründlich gelesen und den Verlauf seiner Argumentation verstanden haben. Dann gehen Sie in den nachfolgend beschriebenen Schritten vor. Zu beachten: Die Überschrift des Textes bildet eine

EDU, die *nicht* in den Baum integriert wird, also stets isoliert stehen bleibt (vgl. Abb. 10.5 auf Seite 184).

1. Zerfällt der Text in erkennbare thematische Einheiten? Markieren Sie etwaige Grenzen zwischen solchen Segmenten, die verschiedene Aspekte des Themas des Kommentars behandeln. Diese Grenzen werden später Anhaltspunkte für Grenzen auch im RST-Baum sein.
2. Entscheiden Sie, welche EDUs für den Text zentrale Rollen spielen, am Ende also im Baum stark-nuklear sein sollten. Wenn eine einzelne EDU als „Kernaussage“ klar heraussticht, markieren Sie diese sowie andere, ebenfalls besonders wichtige EDUs. Um das Ergebnis zu überprüfen, machen Sie dann den „Paraphrasentest“: Reihen Sie die markierten EDUs aneinander (unter Löschung etwaiger unangebundener Konnektoren und Ersetzung von Anaphern durch ihre Antezedenten) und beurteilen Sie, ob der entstehende Kurztext eine adäquate Zusammenfassung des Texts ist. Wenn nötig, überarbeiten Sie die Nukleus-Auswahlen.
3. Untersuchen Sie der Reihe nach jede EDU und ihre unmittelbaren Nachbarn. Gibt es eine klar erkennbare Relation zwischen zwei benachbarten EDUs? Dies ist insbesondere bei den meisten Haupt-/Nebensatz Verbindungen der Fall, und mitunter auch, wenn zwei syntaktisch unabhängige EDUs durch einen Konnektor verbunden sind.
Für zu verbindende Nachbar-EDUs:
 - Entscheiden Sie, ob eines der beiden Segmente für den Text ein größeres Gewicht besitzt als das andere, oder ob beide gleichberechtigt nebeneinander stehen.
 - Entscheiden Sie, welche Relation (s. Abschnitt 10.4) zwischen den beiden EDUs besteht. Dies wird durch die vorher getroffene Nuklearitätsentscheidung mit beeinflusst.

- Im Falle der multinuklearen Relationen können auch mehr als zwei nebeneinander stehende EDUs verbunden werden; prüfen Sie, ob dies zutrifft.
 - Notieren Sie die gefundene Relation.
4. Nachdem alle benachbarten EDU-Paare untersucht sind, fahren Sie mit der Betrachtung größerer Einheiten fort: Ein Konnektor kann auch mehr als eine EDU als Segment nehmen, und natürlich können Relationen sowohl zwischen EDUs als auch zwischen größeren Abschnitten *nicht* durch Konnektoren angezeigt sein. Identifizieren Sie zunächst die leicht erkennbaren Zusammenhänge zwischen solchen größeren Abschnitten. Wenn Sie in Schritt 1 thematische Grenzen markiert haben, ziehen Sie diese als Segmentgrenzen in Betracht. Bei der Zuweisung der Nuklearität kann nun die Rolle des Nukleus für den Gesamttext (Schritt 2) eine wichtige Rolle spielen – Ihre Entscheidungen zur Relationszuweisung an größere Segmente sollen berücksichtigen, dass am Ende die gewünschten EDUs streng-nuklear sind. Hilfreich für die Relationsentscheidung ist oft der Test, einen Konnektor einzufügen; z. B. bei der Probe, ob eine Begründung vorliegt, dem begründeten Segment ein „Also:“ voranzustellen. Generell empfiehlt es sich, dass Sie Ihre Hypothesen zunächst provisorisch notieren und dann prüfen, ob sie sich jeweils in eine adäquate Baumstruktur einfügen würden.
 5. Wenn Sie sich auf die Relationen zwischen größeren Abschnitten festgelegt haben, markieren Sie diese. Dabei sollten Sie tendenziell *bottom-up* vorgehen, also immer nur benachbarte Segmente (EDUs oder größere) verbinden und so sukzessive den Baum von unten nach oben aufbauen. Dementsprechend wichtig ist es für komplexe Texte, den Baum zuvor als Ganzes zu skizzieren.

(1) **Siamnesische Zwillinge**

(2) Die Lausitzer Braunkohle AG (Laubag) hat im abgelaufenen Geschäftsjahr mit der Förderung und dem Verkauf von Braunkohle erstmals Verluste gemacht. (3) Damit wird erneut deutlich, dass eine Neuordnung der Energiewirtschaft in Ostdeutschland überfällig ist. (4) Denn die Laubag und ihr wichtigster Kunde, die Veag, hängen wie siamesische Zwillinge voneinander ab. (5) Als die Veag wegen der Liberalisierung des Strommarktes unter Druck geriet und ihre Strompreise senken musste, (6) hielt sie sich bei ihrem Lieferanten schadlos. (7) Die Laubag musste Preiszugeständnisse machen, (8) die sie nun selbst in Bedrängnis bringen. (9) Einziger Weg aus dem Dilemma ist die Bildung eines neuen Energiekonzerns aus Rohstofflieferanten, Stromerzeugern und Endversorgern, (10) in dem Risiken besser verteilt werden können. (11) Die Idee, auch einen Gasversorger mit ins Boot zu holen, ist im Prinzip nicht schlecht. (12) Allerdings gilt auch hier, dass zu viele Köche den Brei verderben können. (13) Zumindest muss klar sein, wer im neuen Konzern das Sagen hat.

Abbildung 10.4: Beispieltext ‚Laubag‘ (maz-5297)

10.6 Analyse eines Beispieltexts

Zur Illustration des Annotationsprozesses gehen wir zum Schluss den in Abb. 10.4 gezeigten Text exemplarisch durch und erläutern die Entstehung seiner RST-Struktur. Die Nummern im Text kennzeichnen die EDUs.

Thematische Grenzen: EDUs 2–8 handeln von den Problemen der Laubag, die dann in 9–13 nicht mehr explizit zur Sprache kommt; stattdessen wird hier der Vorschlag zur Bildung eines neuen Konzerns gemacht. Nach 8 findet sich also eine thematische Grenze, die wahrscheinlich auch mit einer Grenze zwischen größeren Segmenten im RST-Baum einhergehen wird.

Nuklearität auf Textebene: Zu Textbeginn spielt 3 eine wichtige Rolle: Dies ist nicht nur die Darstellung des Textthemas, sondern auch eine subjektive These – eine zentrale Forderung des Autors. Die These wird in 9 fortgesetzt zu einem Lösungsvorschlag; dieses Segment bietet sich als zentraler Nukleus des Textes an. Das Schlussegment 13 fügt eine Forderung hinzu, ist also ebenfalls wichtig für den Text. Der entsprechende Paraphrasentest ergibt: „Es wird deutlich, dass eine Neuordnung der Energiewirtschaft in Ostdeutschland überfällig ist. Einziger Weg aus dem Dilemma ist die Bildung eines neuen Energiekonzerns aus Rohstofflieferanten, Stromerzeugern und Endversorgern. Es muss klar sein, wer im neuen Konzern das Sagen hat.“ Dies scheint in der Tat eine geeignete Zusammenfassung des Textes zu sein.

Lokale EDU-Verknüpfung: 5 beginnt mit einem Nebensatz, der wie in den meisten (aber nicht in allen) Fällen in der RST-Analyse unmittelbar mit dem Hauptsatz verbunden ist. Der Konnektor *als* signalisiert einen temporalen Zusammenhang; darüber hinaus könnte auch ein kausaler Zusammenhang vermutet werden, der hier aber implizit bleibt. Somit wählen wir die Relation Circumstance zwischen 5 (Sat) und 6 (Nuk). Auch die beiden Nebensätze 8 und 10 sind unmittelbar an die Hauptsätze anzubinden. 8 liefert weitergehende Information über die in 7 thematisierten Preisgegenstände, es liegt also E-Elaboration vor. Das gleiche gilt für 10 (Sat) und 11 (Nuk). — Ansonsten sind zwar Konnektoren gegeben, doch handelt es sich um Adverbiale, deren Skopus nicht sofort evident ist. Weitere Entscheidungen werden deshalb in diesem Schritt nicht getroffen.

Größere Einheiten und Hierarchisierung: In der ersten Texthälfte wird deutlich, dass 2 und 4 Gründe für 3 liefern, wobei 5–8 wiederum die Aussage in 4 begründet. 2 ist eine Beobachtung „der Welt“, steht also als Satellit in Evidence Relation zu 3. 4 hingegen ist keine objektive Beobachtung, sondern eine subjektive These, liefert also einen Reason für 3 (Nuk). Die Begründung für 4 durch

5–8 scheint eher objektiv (Beobachtung: Es wurden Preiszugeständnisse gemacht), es liegt also Evidence zwischen 5–8 (Sat) und 4 (Nuk) vor. 5/6 sowie 7/8 sind bereits verbunden. 7–8 erläutert die in 5 genannte Preissenkung genauer, somit setzen wir Elaboration zwischen 7–8 (Sat) und 5–6 (Nuk) an.

Am Textende dient die subjektive These 12 als Grund für 13, es besteht also die Relation Reason. Zwar ist *zumindest* ein einschränkender Konnektor, doch jenseits der Oberfläche ist die Begründungsrelation klar erkennbar. Wie durch *allerdings* angezeigt, besteht Adversativität zwischen 11 und 12–13. Wie der (funktionierende) Ersetzungstest durch *zwar–aber* bestätigt, wird 11 konzediert, 12–13 aber als stärkere Gegenthese aufgestellt. Damit liegt Concession zwischen 11 (Sat) und 12–13 (Nuk) vor. Um die eingangs festgestellte thematische Grenze zu respektieren, sollte 11–13 nun an 9–10 angebunden werden. Hier sind keinerlei Adversativität, temporale oder begründende Zusammenhänge erkennbar, stattdessen wird die These 9 durch den Spezialfall ‚Gasversorger‘ weiter ausgeführt, und insgesamt sind 11–13 sicherlich weniger nuklear als 9–10. Damit ist Elaboration zwischen 11–13 (Sat) und 9–10 (Nuk) eine adäquate Analyse.

Die letzte Entscheidung betrifft die Verbindung zwischen 2–8 und 9–13. Letztgenanntes Segment sollte der Nukleus sein, damit 9 in der Tat zum Text-Nukleus wird. Die Formulierung „Weg aus dem Dilemma“ legt nahe, dass ein Problem–Lösung Zusammenhang besteht: Die Energiewirtschaft muss neu geordnet werden. Aber wie? Durch einen neuen Konzern. Die Relation Solutionhood bildet dies gut ab.

Abbildung 10.5 zeigt den durch diese Entscheidungen erstellten RST-Baum.

Argumentationsstruktur

Andreas Peldszus, Saskia Warzecha, Manfred Stede

Hinweise:

- Diese Annotationsrichtlinien sind für Kommentare der Gattung „Pro & Contra“ konzipiert, die in knapper Form eine klare Position zu einer gegebenen Fragestellung beziehen und diese begründen.
- Die vorliegenden Richtlinien verwenden eine Notation, wie sie „auf Papier“ verwendet werden kann. Für das praktische Vorgehen mit einer speziellen Annotationssoftware werden separate Hinweise für deren Benutzung ausgegeben.
- Die Arbeit mit diesen Richtlinien setzt voraus, dass der zu bearbeitende Text gemäß Kapitel 3 segmentiert ist.
- Alle hier angegebenen Beispielsätze ohne Quellenangabe sind fiktiv.

11.1 Allgemeines

Die vorliegenden Annotationsanweisungen zielen auf Kommentare, in denen ein Autor relativ klar für oder gegen eine bestimmte *These* argumentiert. Das Ziel dieser Annotation ist ein schematisches Diagramm, das die argumentativen Zusammenhänge zwischen (Teil-)Sätzen identifiziert und damit insgesamt den Kern der Argumentationsstruktur des Texts abbildet. Dieser „Kern“ stellt insofern eine Abstraktion dar, als Argumentation durchaus subtil sein kann; z. B. kann die Verwendung emotional gefärbter Wörter „Nebenefekte“ erzielen. Solche werden in der Annotation vernachlässigt – unsere Aufmerksamkeit richtet sich auf die primären relationalen Zusammenhänge zwischen (Teil-)Sätzen.

Der Annotationsprozess gliedert sich in vier Schritte:

1. Der Text wird in minimale Einheiten (*argumentative discourse units*, kurz ADUs) segmentiert.
2. Die Gesamtthese, für die der Text argumentiert, wird bestimmt.
3. Für jedes einzelne Textsegment wird festgestellt, welche argumentative *Stimme* (Proponent für oder Opponent gegen die These) es repräsentiert.
4. Schließlich wird für jedes Segment die argumentative Funktion und der genaue Anknüpfungspunkt an die bestehende Struktur bestimmt.

Jeder der genannten Schritte wird in den folgenden Abschnitten einzeln beschrieben. Im letzten Teil dieser Anweisungen wird beispielhaft die Annotation eines längeren Textes erörtert.

Das Resultat des Annotationsprozesses ist eine Argumentationsstruktur: ein Graph, dessen Knoten den Segmenten des Textes entsprechen und deren Verküpfungen den argumentativen Zusammenhang zwischen den Segmenten repräsentieren. Der Graph soll Auskunft darüber geben, wo die Gesamtthese zu finden ist, für die in diesem Text argumentiert wird, und er soll zeigen, welche Gründe für die These angeführt werden, wobei es sich jeweils selbst wieder um zusammengesetzte Argumentationsstränge handeln kann. Der Graph bildet auch ab, welche Gegenargumente berücksichtigt werden und ob und wie diese entkräftet werden. Wie wir sehen werden, muss der Graph den Text nicht *vollständig* beschreiben – argumentativ irrelevante Teile sollen in dieser Strukturbeschreibung unberücksichtigt bleiben.

Zuletzt sei darauf hingewiesen, dass die verschiedenen Schritte im Annotationsprozess nicht immer unabhängig voneinander durchgeführt werden können: Entscheidungen in früheren Schritten bedingen solche in späteren Schritten. Es kann also mitunter auch passieren, dass eine Entscheidung aus früheren Schritten später noch einmal revidiert werden muss, bzw. dass diese gleich im Lichte späterer Möglichkeiten getroffen wird. Bitte lesen Sie

diese Annotationsanweisungen daher zunächst vollständig durch und vergewissern Sie sich, dass Ihnen die zu treffenden Unterscheidungen klar sind, bevor Sie mit dem eigentlichen Annotieren beginnen.

11.2 Schritt 1: Segmentierung

Der Segmentierungsschritt gliedert sich wiederum in zwei Teilaufgaben: Die Ableitung von elementaren Diskurseinheiten (EDUs)¹⁰ aus der gemäß Kapitel 3 bestimmten „allgemeinen“ Diskurssegmentierung, und die anschließende Bestimmung von spezifisch-argumentativen Einheiten (ADUs) aus den EDUs.

11.2.1 Bildung von EDUs

Die gemäß Kapitel 3 gebildete Segmentierung ist hierarchisch und muss für die Argumentationsanalyse zunächst zu einer flachen Struktur, einer Partitionierung in *elementary discourse units* (im Folgenden: EDUs, markiert durch eckige Klammern), umgestellt werden. Als Grundregel gilt dafür:

- Eine EDU entspricht einer erkennbaren, selbstständigen Sprechhandlung (Illokution). Diese muss aber nicht im engen Sinne strukturell „vollständig“ sein: Etwaige Elisionen sind bei der Beurteilung aufzufüllen, anaphorische (oder kataphorische) Verweise durch ihre Antezedenten zu ersetzen.

Diese Regel soll für die verschiedenen Segmenttypen wie folgt umgesetzt werden:

Einschübe: Aus rein formalen (und auch technischen) Gründen sollen bei der Annotation stets benachbarte EDUs miteinander verbunden werden. Für parenthetische Einschübe, die in der Mitte eines Segments stehen, machen wir daher eine Ausnahme von der

¹⁰Die hier beschriebene Bestimmung der EDUs verläuft genau so wie bei der Analyse gemäß RST (Kapitel 10).

Grundregel und markieren sie auch dann *nicht* als EDU, wenn sie eigentlich eine Illokution darstellen. Alle nachfolgenden Anweisungen gelten *modulo* dieser Ausnahmeregelung.

Hauptsätze: Hauptsätze (HS) und -fragmente (HSF) bilden stets eine eigenständige EDU.

Fragmente: FRE und FRB sind auf den Status einer selbstständigen Illokution hin zu überprüfen und dann entweder als eigene EDU zu markieren oder dem benachbarten HS/HSF anzufügen. Beispiele für unterschiedlich zu behandelnde FRE:

(11.219) [Übrigens: Der Bürgermeister wurde wiedergewählt.]

(11.220) [Kaum zu fassen:] [Der Bürgermeister wurde wiedergewählt.]

Im zweiten Fall drückt das FRE eine Beurteilung des nachfolgenden HS aus, der Autor vollzieht damit eine eigenständige Sprechhandlung.

Nebensätze: Für die verschiedenen Arten von Nebensätzen ist wie folgt vorzugehen:

- SUB, OBJ, PRD haben Satzgliedstatus und bilden in der Regel keine eigenständige EDU. Als Ausnahme sind aber Konstruktionen zu behandeln, in denen der Autor eine eingebettete Aussage persönlich bewertet bzw. beurteilt, wie hier:

(11.221) [Es ist kaum zu fassen,] [dass der Bürgermeister wiedergewählt wurde.]

- ARR und ATT dienen „nur“ zur näheren Charakterisierung eines Diskursreferenten, übermitteln also keine eigenständige Informationseinheit, bilden daher keine EDU.
- ANR und WEI kommunizieren selbstständige Informationseinheiten und bilden eine EDU.

- ADV stehen in einer durch den Konnektor markierten inhaltlichen Relation zum übergeordneten HS/HSF, enthalten damit eine eigenständige Informationseinheit und bilden eine EDU.
- AKP und UNS müssen im Einzelfall anhand der Grundregel beurteilt werden.

11.2.2 Bildung von ADUs

Im zweiten Teilschritt der Segmentierung geht es darum, aus den EDUs diejenigen zu identifizieren, die als „Bausteine“ der Argumentation, als *argumentative* Diskurseinheiten (ADUs) gelten können.

Im allgemeinen gilt: Jede EDU, die einen eigenständigen argumentativen Beitrag leistet, ist auch eine ADU. Dass es aber keine 1:1 Zuordnung ist, liegt an drei möglichen Konstellationen:

- Eine EDU muss in der Argumentationsstruktur übergangen werden, weil sie für das Argument nicht relevant ist. (→ Tilgen)
- Mehrere EDUs müssen zu einer ADU gebündelt werden, weil sie einzeln nicht vollständig wären. (→ Bündeln)
- Mehrere EDUs müssen zusammengefasst werden, weil eine spätere EDU nur eine Reformulierung einer früheren EDU liefert. (→ Wiederholen)

Diese Fälle werden im Folgenden genauer erläutert. Vorab jedoch zwei Hinweise zur „mentalen Umformung“ von Segmenten:

- Für die Argumentationsanalyse gilt generell: Die einzelnen Segmente müssen jeweils so betrachtet werden, als seien alle anaphorischen Bezüge (einschließlich Ellipsen) aufgelöst — die Segmente sollten also mental „vervollständigt“ werden.
(11.222) [*Dabei gehen ohnehin bloß die an solche Orte, die etwas wissen und sich erinnern wollen.*] [*Die brauchen keine Gegenwartsarchitektur über schreienden Mauern.*]

→ „Informierte Besucher brauchen keine
Gegenwartsarchitektur über schreienden Mauern.“
(topographie-con)

- Bei der Segmentierung sollte beachtet werden, dass rhetorische Fragen als die ihnen zugrunde liegende Aussage zu verstehen sind. So sollte die rhetorische Frage in Beispiel 11.223 aufgelöst werden zur Feststellung „Niemand zahlt gerne 12 Euro für die Warteschleife der Telekom.“

(11.223) Aber wer zahlt schon gerne 12 Euro für die
Warteschleife der Telekom?

Tilgen: Nicht jedes Segment im Text leistet einen argumentativen Beitrag. Ein Autor kann z. B. seinen argumentativen Text eröffnen, indem er in den ersten Sätzen erst einmal eine bestimmte Stimmung erzeugt, die er als gewinnbringend für die Überzeugungskraft seiner eigentlichen Argumentation ansieht. Auch ist es möglich, dass er für bestimmte Teile seiner Argumentation zunächst den Kontext setzt, indem er unkontroverse Hintergrundinformationen liefert, die nicht Teil der Argumentation im engeren Sinn sind (aber deren Verständnis erleichtern). Oder der Autor teilt im Verlauf seiner Darlegung „Seitenhiebe“ aus oder macht Nebenbemerkungen, die für den Kern der Argumentation aber nicht weiter relevant sind.

In allen diesen Fällen, wo der Text Segmente enthält, die man ausklammern kann, ohne dass dadurch mögliche Begründungen oder Gegenargumente unterschlagen werden, können die betreffenden Segmente getilgt werden. Für sie wird keine ADU und das heißt kein Knoten in der argumentativen Struktur angelegt. Ein Beispiel für ein nur Stimmungs-erzeugendes Segment ist die erste EDU von Beispiel 11.224.

(11.224) [Stifte gezückt und mitgeschrieben:]₁ [Diktate verbessern die Rechtschreibung kaum.]₂ [Die Pflicht, einmal im Jahr ein Diktat zu absolvieren, kann also getrost abgeschafft werden.]₃

Bündeln: Wenn ein Textsegment für sich genommen keinen vollständigen Satz umfasst, bzw. keine vollständige Aussage ausdrückt, es aber mit einem benachbarten Segment zu einem vollständigen Satz, bzw. einer vollständigen Aussage, kombiniert werden kann, sollten beide Segmente zu einer ADU gebündelt werden. Dies kann notwendig werden, wenn der Autor einen Satz aus rhetorischen Gründen, etwa zur Hervorhebung, durch Interpunktion in mehrere Teile aufteilt. In Beispiel 11.225 werden also die EDUs [1] und [2] zu einer ADU [1,2] gebündelt, welche besagt, dass das Gebäude durch und durch asbestverseucht ist.

(11.225) [Das Gebäude ist asbestverseucht.]₁ [Durch und durch!]₂

Eine Ausnahme: Wenn ein Segment unvollständig ist, weil es sich dabei um ein verkürztes (elliptisches) Satz-Konjunkt handelt, so muss entschieden werden, ob das verkürzte Segment auch einen eigenen argumentativen Beitrag erbringt, wenn die Ellipse aufgelöst würde. Falls ja, darf es als eigenständige ADU bestehen, ansonsten kann es mit dem anderen Satzkonjunkt zu einer ADU gebündelt werden. Ein Beispiel:

(11.226) [Dabei ist doch allgemein bekannt, dass derartige
Bauvorhaben am Ende immer mehr kosten als geplant]₁
[und dass Aufsichtsräte ihrer eigentlichen Aufgabe nur
selten gerecht werden.]₂

Zuerst löst man die Ellipse in Segment 2 auf zu „Dabei ist doch allgemein bekannt, dass Aufsichtsräte ihrer eigentlichen Aufgabe nur selten gerecht werden.“ und entscheidet dann, ob dies in der Argumentation eine Rolle spielt. Nur im positiven Fall darf es als eigenständige ADU bestehen.

(Segmente mit anaphorischen Pronomen gelten in diesem Sinne nicht als „unvollständig“, weil die Antezedenten mental einzusetzen sind.)

Wiederholung: Wenn zwei Segmente sich inhaltlich so gleichen, dass sie jeweils als Reformulierung des anderen gelten können,

sollten diese beiden Segmente zu einer ADU zusammengefasst werden. Dies tritt insbesondere bei der Gesamtthese des Textes (siehe Schritt 2) häufiger auf, wenn ein Autor sie zur Bekräftigung an einer anderen Stelle noch einmal wiederholt. Diese Form der Zusammenfassung wird also typischerweise eher bei nicht benachbarten Segmenten notwendig. In Beispiel 11.227 werden also die EDUs 1 und 5 als ein und dieselbe ADU (1=5) aufgefasst.

(11.227) [Es sollten wirklich mehr Menschen zur Wahl gehen.]₁
[Denn. . .]₂ [Außerdem. . .]₃ [und. . .]₄ [Die Wahlbeteiligung
muss also unbedingt gesteigert werden.]₅

Am Ende dieses Schrittes liegt eine Auflistung der ADUs vor, die notwendig sind, um die vom Autor vorgenommene Argumentation vollständig wiederzugeben. Diese werden im eigentlichen Annotationsprozess nun zu einem Graphen verbunden. Wenn im Folgenden von Segmenten gesprochen wird, sind damit ausschließlich ADUs gemeint.

11.3 Schritt 2: Identifikation der Gesamtthese

Zunächst muss von den Segmenten des Textes dasjenige ausgewählt werden, welches am ehesten die Gesamtthese des Textes wiedergibt. Es soll also die „Grundaussage“ des Textes gefunden werden, von der der Autor den Leser letztlich überzeugen möchte. Diese Gesamtthese steht für sich selbst, während alle anderen Textsegmente dazu dienen, durch Stützung und Anfechtung die Überzeugungskraft der Gesamtthese zu verstärken. Sie kann prinzipiell an jeder Stelle eines argumentativen Textes auftreten, sei es gleich zu Beginn, in der Mitte oder erst ganz zum Schluss.

Bei der Gesamtthese handelt es sich oft um Handlungsanweisungen oder -empfehlungen („Wir/Man/Peter sollte X tun.“). Genauso können aber auch andere Texthandlungen, wie Bewertungen („X ist schlecht.“) oder Vermutungen („Wahrscheinlich ist X der Fall.“) etc. als Gesamtthese fungieren.

Es kann mitunter der Fall sein, dass sich die Gesamtthese nur „zwischen den Zeilen“ des Textes findet: Zwar lässt sich die Gesamtthese aus dem Text eindeutig ablesen, aber kein Segment bringt sie klar und alleinig zum Ausdruck. In diesem Fall sollte ein neues Textsegment [GT] eingeführt werden, das sie explizit macht.

(11.228) [Im letzten Winter sind in Berlin wieder zwei Menschen auf offener Straße erfroren,]₁ [obwohl man es mit einem Anruf beim Kältebus eventuell hätte verhindern können.]₂ [Dabei kostet der Anruf nur ein paar Cent]₃ [und die Nummer - 0178-5235838 - ist doch schnell eingespeichert!]₄

So ist die Gesamtthese von Beispiel 11.228, dass im Winter in Notsituationen der Kältebus angerufen werden sollte. Dies ist allerdings nicht der explizite Inhalt einer der vier Sätze. Daher wird ein zusätzliches Segment eingeführt und explizit gemacht. [GT] = *Im Winter sollte in Notsituationen der Kältebus angerufen werden.*

11.4 Schritt 3: Zuweisung der argumentativen Stimme

Die Gesamtthese des Textes kann vom Autor kontrovers diskutiert werden. Er präsentiert eventuell nicht nur Gründe, die für die Annahme der These sprechen, sondern zieht auch mögliche Gegenargumente in Betracht. Dies kann man analog zu einem Streitgespräch verstehen, in dem ein Proponent eine These vertritt und diese gegen die Angriffe eines Opponenten verteidigt.

Der nächste Schritt besteht deswegen darin, für jedes Segment festzustellen, ob dessen Aussage im analogen Streitgespräch vom Verteidiger der Gesamtthese, also vom Proponenten vorgebracht werden würde, oder vom Opponenten, der die These und ihre Begründung kritisch hinterfragt.

In Beispiel 11.229 lässt der Autor eine mögliche Gegenstimme („Aber die Kleidung ist so günstig!“) zu Wort kommen, also würde das zweite Segment der Stimme des Opponenten zugeordnet wer-

den, während das erste Segment wie jede Gesamtthese die Stimme des Proponenten repräsentiert.

(11.229) [Den Textil-Discounter A&N sollte man nicht unterstützen.]₁ [auch wenn die Preise für Jeans und T-Shirts verlockend sind.]₂

Um diese Unterscheidung im Diagramm deutlich zu machen, repräsentieren wir jene Segmente, die zur Stimme des Proponenten gehören, durch einen kreisförmigen Knoten und eventuelle Segmente des Herausforderers durch einen quadratischen Knoten. Das Label eines Knoten ist stets die Nummer des repräsentierten Segments.

Beide Segmente im vorangegangenen Beispiel sind eher subjektiv, d. h. meinungstragend, was die Zuweisung der argumentativen Stimme meist vereinfacht. Andererseits gibt es auch oft Segmente, in denen objektiv ein nicht-kontroverser Fakt angeführt wird. Dem Segment selbst ist dabei mitunter nicht anzusehen, ob der Proponent oder der Opponent sich diesen Fakt argumentativ zu eigen macht. In diesem Fall müssen also die benachbarten Segmente mitberücksichtigt werden. Im folgenden Beispiel fungiert ein Segment einmal als Grund für die These und ist folglich dem Proponenten zuzuordnen (11.230) und das andere Mal als mögliches Gegenargument, wodurch es der Stimme des Opponenten angehört (11.231). Festzustellen ist dies nur im Kontext der anderen Segmente.

(11.230) [Der neue Stromtarif ‚Green Berlin‘ hat einen 95% igen Anteil von erneuerbaren Energien.]₁ [Er bietet also alles, was ein öko-bewußter Stromkunde sich wünscht.]₂

(11.231) [Der neue Stromtarif ‚Green Berlin‘ hat einen 95% igen Anteil von erneuerbaren Energien.]₁ [Trotzdem wird ein öko-bewußter Stromkunde damit nicht glücklich werden.]₂

Mitunter kann es passieren, dass eine Gegenstimme vom Autor genannt wird, diese aber nicht in einem eigenständigen Segment ausgedrückt wird, weil der Autor sie im selben Zug gleich wieder

entkräftet. In diesem Fall kommen also in einem Segment beide argumentativen Stimmen zu Wort. Solchen Segmenten weisen wir die übergeordnete Rolle zu, also die des Proponenten. Im Gegensatz zu Beispiel 11.233, wo das erste Segment die Opponentenstimme und das zweite die Proponentenstimme repräsentieren, ist in Beispiel 11.232 beides in einem Segment ausgedrückt, welches insgesamt dem Proponent zuzuordnen ist.

(11.232) [Die ach so teure Rettung der Banken ist unausweichlich gewesen.]

(11.233) [Ja, die Rettung der Banken war teuer.]₁ [Aber sie ist unausweichlich gewesen.]₂

Nach Abschluss dieses Schritts besteht die angestrebte Argumentationsstruktur aus Knoten (unterschiedlicher Art), die unverbunden nebeneinander stehen. Der letzte Annotationsschritt bestimmt deshalb die Art der Verbindung (also die Sorte des Pfeils) und die Stelle im Graphen, an die sich ein Knoten anschließt.

11.5 Schritt 4: Zuweisung der Funktion und des Bezugspunkts

Für ein Segment gibt es in der Argumentation (gemäß der argumentativen Stimme) zwei elementare Funktionen: „stützen“ und „anfechten“. Die Gesamtthese ist hiervon natürlich ausgeschlossen, denn alle anderen Segmente führen zu ihr hin. Jedes Segment soll nur *eine* Funktion haben, d. h. im Graphen geht von jedem Knoten maximal ein Pfeil aus. Im Zweifelsfall sollte man sich daher für die Funktion entscheiden, die für den Text von größerer Wichtigkeit ist.

11.5.1 Stützen

Es gibt mehrere Möglichkeiten, im Argument etwas zu stützen. Allen Möglichkeiten ist aber gemein, dass das, was gestützt wird,

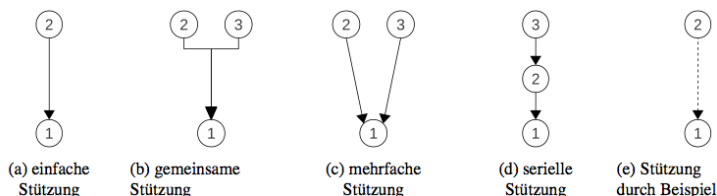


Abbildung 11.6: Stützungsrelationen

durch das, was stützt, glaubhafter gemacht werden soll. Die Stützung liefert eine Begründung für die Annahme einer Aussage, sie soll deren argumentative Kraft vergrößern. Indem der Autor ein Segment *A* durch ein anderes (*B*) stützt, beantwortet er indirekt die Frage des Lesers „Warum sollte ich *A* glauben/annehmen?“ Einen ersten Eindruck, ob es sich bei dem Segment *B* um eine Stützung von *A* handelt, gibt daher der warum-Test: Wenn das Gespräch „A. - Warum? - Weil B.“ kein geglückter Mini-Diskurs ist, dann ist es auch unwahrscheinlich, dass *B* eine Stützung von *A* ist.

Einfache Stützung: Ein einfaches Beispiel für eine Stützung wäre 11.234:

(11.234) [Wir sollten das Gebäude abreißen,]₁ [denn es ist völlig asbestverseucht.]₂

Die Stützungsrelation wird durch einen normalen Pfeil repräsentiert. Der Argumentationsstrukturgraph für dieses kurze Beispiel ist in Abbildung 11.6 (a) dargestellt. Es sei darauf hingewiesen, dass die Reihenfolge des stützenden und des gestützten Segments auch andersherum sein kann. So in Beispiel 11.235, wo erst das stützende Segment genannt wird und dann das dadurch gestützte. Im Graph wäre der Pfeil zwischen Segment 1 und 2 dann entsprechend andersherum.

(11.235) [Das Gebäude ist völlig asbestverseucht.]₁ [Wir sollten es daher abreißen.]₂

Gemeinsame Stützung: Neben dieser einfachen Stützung gibt es auch den komplexeren Fall einer gemeinsamen Stützung. Hier können zwei Segmente einzeln für sich genommen nicht zur Stützung einer Aussage beitragen, wohl aber gemeinsam, wenn sie beide vom Autor als zutreffend behauptet werden. Beispiel:

(11.236) [Wir sollten das Gebäude abreißen.]₁ [denn es ist völlig asbestverseucht,]₂ [und verseuchte Gebäude müssen abgerissen werden.]₃

Das Argument funktioniert nur, wenn beide Prämissen gleichzeitig wahr sind: Wenn verseuchte Gebäude nicht abgerissen werden müssten, spräche eine vorhandene Verseuchung nicht für den Abriss. Andersherum: Wenn verseuchte Gebäude abgerissen werden müssen, ein bestimmtes aber nicht verseucht ist, muss es auch nicht abgerissen werden. Im Graphen wird eine solche Struktur durch einen Stützungspfeil mit mehreren Startpunkten und einem gemeinsamen Zielpunkt angezeigt (siehe Abbildung 11.6 (b)).

Es ist zu bemerken, dass die Regel, die hier im dritten Segment explizit genannt ist, im vorherigen Beispiel 11.234 implizit bleibt und daher mit angenommen werden muss. In der Annotation werden aber nur explizit in Segmenten ausgedrückte Prämissen berücksichtigt.

Mehrfache Stützung: Von der gemeinsamen Stützung unterschieden wird der Fall, in dem mehrere Prämissen unabhängig voneinander stützen. Dies ist in Beispiel 11.237 der Fall:

(11.237) [Wir sollten das Gebäude abreißen,]₁ [denn es ist völlig asbestverseucht,]₂ [und die Anwohner in der Nachbarschaft haben es sowieso nie haben wollen.]₃

Selbst wenn das Gebäude nicht asbestverseucht ist, wäre die Unbeliebtheit bei den Anwohnern noch ein (mehr oder wenig starker)

Grund für den Abriss. In der Argumentationsstruktur wird diese Konstellation folglich durch zwei einzelne Stützungs Pfeile repräsentiert (siehe Abbildung 11.6 (c)).

Serielle Stützung: Natürlich sind auch andere Anknüpfungspunkte für ein stützendes Segment möglich. Das Abrissbeispiel könnte auch durch ein drittes Segment fortgesetzt werden, welches nicht die Gesamtthese stützt, sondern die Zwischenthese. In Beispiel 11.238 stützt der Proponent die Glaubhaftigkeit des zweiten Segments, indem er auf das Ergebnis der Expertenkommission verweist.

(11.238) [Wir sollten das Gebäude abreißen,]₁ [denn es ist völlig asbestverseucht.]₂ [Die Expertenkommission bescheinigte eine beträchtliche Kontamination.]₃

Auf diesem Weg ergibt sich rekursiv eine serielle Stützungsstruktur, wie sie in Abbildung 11.6 (d) zu sehen ist.

Stützung durch Beispiele: Ein besonderer Fall von Stützung liegt vor, wenn eine These dadurch gestärkt wird, dass auf ein Beispiel verwiesen wird:

(11.239) [Eine Bürgerinitiative kann die lokalen Autoritäten zwingen ein Gebäude abzureißen.]₁ [So hat es in München eine Gruppe geschafft, den Bürgermeister zum Abriss eines unansehnlichen, leerstehenden Bürogebäudes zu bewegen.]₂

Das konkrete Beispiel liefert „empirische Evidenz“ für das Zutreffen der These, ohne sie aus regelhaften Zusammenhängen abzuleiten. Insofern ist die Beispiel-Stützung meist weniger kräftig. Im Graphen wird diese spezielle Form der Argumentation durch einen gestrichelten Stützungs Pfeil dargestellt (siehe Abbildung 11.6 (e)).

Sowohl der Proponent als auch ein potenzieller Herausforderer kann sich einer Stützung seiner eigenen Behauptungen bedienen. Wir haben hier lediglich Stützungen auf Seiten des Proponenten

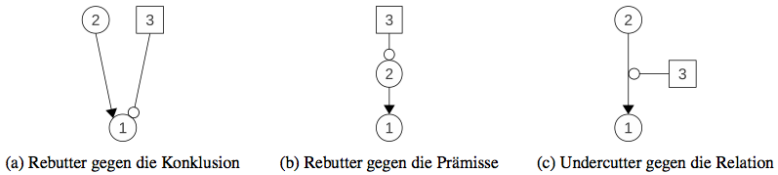


Abbildung 11.7: Angriffe des Opponenten auf die Argumentation des Proponenten

besprochen, dieselben Stützungs-Relationen sind aber ebenso zwischen den Segmenten des Opponenten möglich, z. B. wenn dieser für seinen Einwand noch zusätzliche Gründe ins Spiel bringt (siehe Beispieltext am Ende). Man sollte aber beachten, dass sowohl der Proponent wie auch der Opponent ausschließlich ihre jeweils eigenen Segmente stützen werden.

11.5.2 Anfechten

Entsprechend der Funktionsweise einer Stützung soll mit einer Anfechtung das Gegenteil erreicht werden: Das, was angefochten wird, soll durch die Anfechtung widerlegt oder in seiner argumentativen Kraft geschwächt werden. Alle Anfechtungen werden im Diagramm als Pfeil mit Kugelspitze dargestellt. Es wird zwischen zwei Sorten von Anfechtung unterschieden: Rebutter greifen Aussagen an, Undercutter greifen Beziehungen zwischen Aussagen an. Dieser Abschnitt erläutert diese Unterscheidung am Beispiel von Anfechtungen durch den Opponenten; der umgekehrte Fall der Erwidern von Anfechtungen durch den Proponenten wird im nachfolgenden Abschnitt behandelt.

Rebutter: Ein Rebutter ficht eine Aussage an. Er behauptet also, dass sie aus bestimmten Gründen nicht gilt. So im folgenden Beispiel:

(11.240) [Wir sollten das Gebäude abreißen,]₁ [denn es ist völlig asbestverseucht.]₂ [Andererseits hat man eine sehr gute Aussicht vom Dach.]₃

Die in Segment 3 angeführte gute Aussicht vom Dach spricht *gegen* einen Abriss des Gebäudes. Der entsprechende Argumentgraph ist in Abbildung 11.7 (a). Ein Rebutter kann unter Umständen auch als Argument *für die Negation* der angefochtenen Aussage verstanden werden: Die gute Aussicht stützt die Aufforderung, das Gebäude nicht abzureißen.

Wie auch bei den Stützungen, gibt es verschiedene mögliche Anknüpfungspunkte (wobei sich die Notation des Pfeils aber nicht ändert). Während im vorherigen Beispiel 11.240 die Konklusion angefochten wurde, wird im folgenden Beispiel 11.241 die Prämisse des Arguments attackiert, siehe Abbildung 11.7 (b).

(11.241) [Wir sollten das Gebäude abreißen,]₁ [denn es soll völlig asbestverseucht sein.]₂ [Aber eigentlich hat noch niemand eine genaue Einschätzung des Grads der Kontamination.]₃

Die Tatsache, dass es noch keine genaue Zahlen über den Grad der Kontamination gibt, spricht gegen die Vermutung, dass es völlig verseucht ist.

Undercutter: Ein Undercutter ficht im Gegensatz zum Rebutter nicht die Gültigkeit einer Aussage an, sondern eine Beziehung zwischen Aussagen, z. B. eine Stützungsrelation.

(11.242) [Wir sollten das Gebäude abreißen,]₁ [denn es ist völlig asbestverseucht.]₂ [Allerdings könnte man es auch sanieren.]₃

In diesem Beispiel wird eine Ausnahme als Gegenargument angeführt. Es wird weder angefochten, dass das Gebäude asbestverseucht ist, noch wird darüber eine Aussage getroffen, dass man es abreißen sollte, sondern es wird die Schlussfolgerung von Verseuchung zum Abriss angegriffen: Nur weil das Gebäude verseucht ist, muss man es ja nicht gleich abreißen. Man könnte es ja auch

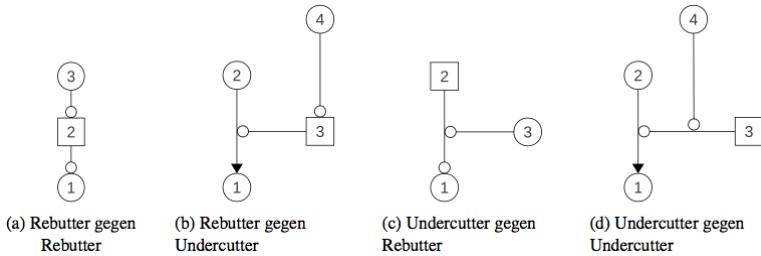


Abbildung 11.8: Erwidrung des Proponenten auf die Anfechtungen des Opponenten.

sanieren. Ein Undercutter wird im Argumentdiagramm folglich als Angriffspfeil gegen einen anderen Pfeil dargestellt, siehe Abbildung 11.7 (c).

Ob ein Einwand des Opponenten nun die Gültigkeit einer vom Proponent vorgebrachten Konklusion anficht (Rebutter), oder durch das Anführen einer Ausnahme die Gültigkeit der Schlussfolgerung von Prämisse zu Konklusion anficht (Undercutter), ist nicht immer offensichtlich, und dies zu entscheiden erfordert eine genaue Prüfung. Mitunter ist es dabei hilfreich zu testen, wie glücklich die Anfechtung ist, wenn die Prämisse nicht da wäre. Ein Undercutter ergibt nur Sinn, wenn es eine Schlussfolgerung gibt, die er untergräbt. Lässt man in Beispiel 11.240 die Prämisse weg (Segment 2), so ist die Anfechtung immer noch wirkungsvoll. Lässt man die Prämisse in Beispiel 11.242 weg, ergibt die Anfechtung keinen Sinn mehr, weil sie abhängig von der Prämisse ist. Wenn man die Prämisse weglassen kann, liegt also wahrscheinlich eher ein Rebutter gegen die Konklusion vor als ein Undercutter.

11.5.3 Anfechtungen erwidern

Bis zu diesem Punkt haben wir nur Anfechtungen des Opponenten gesehen, der sich gegen die Argumente des Proponenten wendet.

Natürlich setzt sich der Proponent auch zur Wehr und verteidigt seine Argumente, indem er im Gegenzug die Angriffe des Opponenten anführt. Dabei ergeben sich folgende Kombinationen:

Einen Rebutter rebutten: Der Opponent hatte einen Grund gegen eine Aussage vorgebracht. Der Proponent entkräftet diesen Einwand, indem er wiederum einen Grund gegen den Einwand vorbringt. So wird in Beispiel 11.243 der Mangel an gesichteten Touristengruppen vorgebracht, um zu zeigen, dass das Gebäude keine Touristenattraktion ist. Für den entsprechenden Graph siehe Abbildung 11.8 (a).

(11.243) [Wir sollten das Gebäude abreißen,]₁ [auch wenn es ja eine Touristenattraktion sein soll.]₂ [Ich hab da jedenfalls noch nie Touristengruppen gesehen.]₃

Einen Undercutter rebutten: Der Opponent hatte durch seinen Hinweis auf eine Ausnahmebedingung eine Schlussfolgerung angefochten. Im Beispiel 11.244: Wenn die Ausnahmebedingung gelten würde und Asbest harmlos wäre, dann müsste man ein verseuchtes Gebäude nicht abreißen. Der Proponent kann aber zeigen, dass die Ausnahmebedingung nicht erfüllt ist, indem er auf die fragwürdige Informationsquelle hinweist. Für den entsprechenden Graph siehe Abbildung 11.8 (b).

(11.244) [Wir sollten das Gebäude abreißen,]₁ [denn es ist völlig asbestverseucht.]₂ [Zwar soll es eine neue wissenschaftliche Studie geben, dass Asbest harmlos ist.]₃ [Aber das ist ja wohl eine Zeitungs-Ente.]₄

Einen Rebutter undercutten: Der Opponent hatte einen Grund gegen eine Aussage vorgebracht, bzw. einen Grund für die Negation der Aussage. Im Gegenzug zeigt der Proponent, dass der vermeintliche Grund diese Schlussfolgerung gar nicht erlaubt. Im Beispiel 11.245 kann der Proponent hinnehmen, dass das Gebäude tatsächlich eine Touristenattraktion ist. Aber er zeigt, dass dies kein

Grund gegen den Abriss ist, weil die Ausnahmebedingung erfüllt ist, dass der Neubau eine noch größere Attraktion sein wird. Für den Graphen siehe Abbildung 11.8 (c).

(11.245) [Wir sollten das Gebäude abreißen,]₁ [auch wenn es eine Touristenattraktion sein soll.]₂ [Die werden bestimmt eine neue und noch größere Attraktion dahin bauen.]₃

Einen Undercutter untercutten: Der Opponent hatte durch seinen Hinweis auf eine Ausnahmebedingung eine Schlussfolgerung angefochten. Der Proponent erwidert, indem er für die Ausnahme selbst wieder eine Ausnahme findet. Im Beispiel 11.246 zeigt der Proponent, dass die Möglichkeit einer Sanierung des Gebäudes wegen des zu hohen Preises irrelevant ist. Er prüft also gar nicht erst, ob die Ausnahmebedingung nun gilt oder nicht, ob eine Sanierung technisch möglich ist oder nicht, sondern er verwirft diese Möglichkeit von vornherein als unrealisierbar. Für den entsprechenden Graph siehe Abbildung 11.8 (d).

(11.246) [Wir sollten das Gebäude abreißen,]₁ [denn es ist völlig asbestverseucht.]₂ [Zwar könnte man es sanieren,]₃ [aber das wäre viel zu teuer.]₄

11.6 Zusammenfassung: Übersicht der Annotationsschritte

Vorbereitung: Den Text einmal komplett lesen.

Schritt 1: Segmentierung

(a) Den Text in EDUs teilen.

(b) EDU-Liste in ADU-Liste überführen:

- Tilgen
- Bündeln
- Wiederholung markieren

Schritt 2: Die Gesamtthese ermitteln: Wird sie im Text durch eine ADU expliziert? Sonst artifizielle ADU „[GT]“ bilden.

Schritt 3: Für jede einzelne ADU bestimmen, wessen Stimme sie repräsentiert, und entsprechende Knoten zeichnen:

- Proponent – rund
- Opponent – eckig

Schritt 4: Für jede ADU ihre Funktion und ihren Bezugspunkt bestimmen, entsprechende Pfeile zeichnen:

- Stützung
 - einfache Stützung
 - gemeinsame Stützung
 - Beispielsstützung
- Anfechtung
 - Rebuttal (gegen Aussagen)
 - Undercutter (gegen Relationen zwischen Aussagen)

Überprüfung:

- Ist der Knoten, zu dem alle Verbindungen hinführen, die in Schritt 1 ermittelte Gesamtthese?
- Geht von jedem Knoten (außer von der Gesamtthese) genau ein Pfeil ab?
- Stützen Proponent und Opponent jeweils nur ihre eigenen Knoten, und fechten sie nur die des jeweils anderen an?

11.7 Analyse eines Beispieltexts

Anhand eines Beispieltexts diskutieren wir im Folgenden das Vorgehen. Wir gehen von einer bereits erstellten EDU-Segmentierung aus.

- [1] Wer regelmäßig die Arbeit für eine kurze Ruhepause unterbricht, arbeitet konzentrierter und effektiver.
- [2] Das belegen schon lange wissenschaftliche Studien, die über längere Zeiträume die Leistungsfähigkeit in der Kreativ-Wirtschaft untersucht haben.
- [3] Unternehmen, die ihren Dichtern und Denkern Ruheräume zur Verfügung stellen, schnitten darin deutlich besser ab.
- [4] Auch Studierende sollten die Möglichkeit haben, zwischen den Vorlesungen einfach mal die Beine hochzulegen und den Kopf abzuschalten.
- [5] In einer ruhigen Umgebung, natürlich.
- [6] Zwar wird nicht jeder Studierende diese Möglichkeit wahrnehmen können,
- [7] denn so mancher muss in der Pause von einem Campus-Standort zum anderen pendeln.
- [8] Und Essen sollte man ja schließlich auch noch irgendwann.
- [9] Aber auch ein Nischen-Angebot hat seine Berechtigung.
- [10] Und mal unter uns:
- [11] Auch die Professoren freuen sich darüber, bei erfrischten und aufnahmefähigen Geistern Gehör zu finden.
- [12] Schaden wird es der Universität bestimmt nicht!

Schritt 1 – ADU-Segmentierung: Von den 12 EDUs können die meisten direkt in ADUs überführt werden, nur zwei verlangen besondere Aufmerksamkeit. Segment [5] ist unvollständig, kann aber mit dem vorhergehenden Segment zu einer Aussage zusammengefasst werden. Beide Segmente werden daher gebündelt und formen eine neue ADU [4,5]. Bei Segment [10] handelt es sich weder um die Gesamtthese oder dessen Stützung, noch um ein mögliches

Gegenargument. Es dient also keinem argumentativen, sondern vielmehr einem rhetorischen Zweck. Es kann folglich getilgt werden und wird im Graphen nicht auftauchen. Reformulierungen vorheriger Segmente sind in diesem Text nicht zu finden.

Schritt 2 – Gesamthese: Der Autor des Textes versucht seine Leser davon zu überzeugen, dass an der Universität für die Studierenden Ruheräume eingerichtet werden sollen. Dies wird am deutlichsten in der ADU [4,5] ausgedrückt.

Schritt 3 - Stimmenzuweisung: Von den übrigen ADUs sprechen sich die meisten direkt oder indirekt für die Gesamthese aus, bis auf die ADUs [6], [7] und [8]. In [6] wird ein mögliches Gegenargument präsentiert und noch durch [7] und [8] verstärkt. Dies sind die Einwände, die der gedachte Opponent vorbringen würde. Im resultierenden Graphen sind diese ADUs folglich als viereckige Knoten dargestellt, während alle anderen ADUs, nämlich diejenigen, die der Proponent vorbringt, als kreisförmige Knoten dargestellt werden.

Schritt 4 - Funktion und Bezugspunkt: Bevor die Gesamthese in [4,5] präsentiert wird, leitet der Autor den Text durch drei ADUs ein, die allesamt für die These sprechen. Wie geschieht dies genau? Handelt es sich um drei Gründe, die die These unabhängig voneinander stützen, oder stützen sie diese nur gemeinsam, oder stützen sie sich seriell? Dass Ruhepausen förderlich sind [1] stützt die Gesamthese auf jeden Fall unabhängig davon, ob es dazu Studien aus der Kreativ-Wirtschaft gibt [2] oder nicht. Eine gemeine Stützung ist also ausgeschlossen. Auch eine Mehrfachstützung durch voneinander unabhängigen Gründe kommt nicht in Frage, da die Gründe nicht gänzlich unabhängig voneinander sind. Der Autor möchte den Leser vom Bestehen des positiven Effekt der Ruheräume überzeugen und führt dazu die Studien an, *weil* sie eben dies zeigen sollen. Dass die Studien den angenommenen Effekt tatsächlich belegen, bestärkt er, indem er das gemessene Ergebnis der Studie resümiert. Es handelt sich hier also um eine serielle Stützungskette: [4,5] wird durch [1] gestützt, das durch [2] und das wiederum durch [3].

Nach der Gesamtthese wird ein mögliches Gegenargument präsentiert. Dabei muss entschieden werden, ob hier ein Rebutter oder ein Undercutter vorliegt und welches der Bezugspunkt ist. Zuerst kann man ausschließen, dass es sich hier um einen Rebutter gegen [1], [2] oder [3] handelt, denn weder der Effekt, noch die Belegkraft der Studie wird in Frage gestellt. Auch ein Undercutter gegen die Beziehung zwischen dem positiven Effekt [1] und dem Wunsch nach Ruheräumen [4,5] liegt nicht vor, da der Einwand der geringen Nutzung unabhängig von [1] ist: Es könnte andere Gründe (wie z. B. soziale Vorteile von gemeinsamen Mittagsschlaf) geben, doch die geringe Nutzung würde weiterhin genauso gegen die Ruheräume sprechen. Stattdessen handelt es sich bei [6] um einen Rebutter gegen die Gesamtthese. Der Autor präsentiert einen möglichen Grund, warum keine Ruheräume eingerichtet werden brauchen. [6] greift [4,5] direkt an.

Dieser Einwand wird durch [7] und [8] dann noch gestärkt. Beide stützen also [6]. In welcher Weise? Eine gemeinsame Stützung ist ausgeschlossen, denn Standortwechsel und Mensagang sind unabhängig voneinander gute Gründe, keine Zeit für ein Schläfchen zu haben. Für eine serielle Stützung fehlt der Zusammenhang (es sei denn, die Mensa ist an einem anderen Standort). Zudem signalisiert der Autor durch die satzeinleitende Konjunktion „und“, dass es sich um gleichrangige Gründe handelt. Folglich muss es sich hierbei um eine mehrfache Stützung handeln.

In [9] wird der mögliche Einwand schließlich vom Autor entkräftet. Zu klären ist auch hier, wie genau das geschieht. Die Entgegnung könnte sich als Rebuttal direkt auf [6], [7] oder [8] richten, aber es wird weder angefochten, dass nur ein Teil der Studierenden das Angebot nutzen können wird, noch warum das so ist. Auch ein Undercutter gegen die Stärkung des Einwands liegt nicht vor. Stattdessen wird die Relevanz des Einwands für die Gesamtthese angefochten. Selbst wenn das Angebot nur von einem Teil der Studierenden wahrgenommen werden kann, erfüllt es doch seinen Zweck. Die ADU [9] fungiert also als Undercutter.

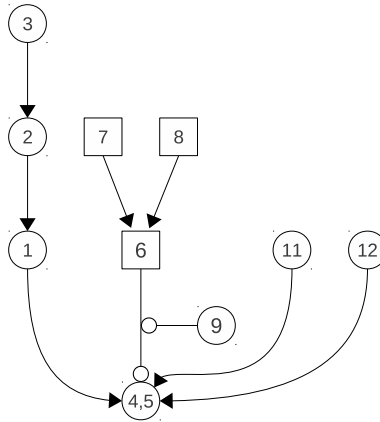


Abbildung 11.9: Die argumentative Struktur des Beispieltextes

Schließlich verbleiben noch die ADUs [11] und [12]. Beide sprechen für die Gesamthese. Eine gemeinsame Stützung kann wieder ausgeschlossen werden, da beide Gründe auch jeweils ohne den anderen funktionieren würden. Ob hier eine serielle oder eine mehrfache Stützung vorliegt, ist interpretationsabhängig: Wenn die Professoren sich einfach über aufmerksame Studierende freuen, unabhängig vom Schaden-Nutzen-Kalkül, dann handelt es sich um eine mehrfache Stützung. Alternativ könnte man unterstellen, dass die Professoren sich über geistige Frische freuen, *weil* sie der Universität nicht schadet. Dann würde die Gesamthese durch [11] und dies wiederum durch [12] gestützt. Solche Entscheidungen sollten aber immer daran festgemacht werden, wie stark das eine oder andere vom Autor signalisiert wurde. Da hier in [12] kein explizites Signal für eine zu unterstellende Beziehung ist, sollte man besser von einer mehrfachen Stützung ausgehen und damit den Weg des geringeren „interpretativen Aufwands“ gehen.

Die resultierende argumentative Struktur ist in Abbildung 11.9 dargestellt.

Literaturverzeichnis

- [Brants u. a. 2002] BRANTS, Sabine ; DIPPER, Stefanie ; HANSEN, Silvia ; LEZIUS, Wolfgang ; SMITH, George: The TIGER Treebank. In: *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol, 2002
- [Breindl u. a. 2011] BREINDL, Eva ; FERRARESI, Gisella ; VOLODI-NA, Anna: Einführung. In: *Satzverknüpfungen. Zur Interaktion von Form, Bedeutung und Diskursfunktion*. Berlin/New York : de Gruyter, 2011 (Linguistische Arbeiten)
- [Bußmann 2002] BUSSMANN, Hadumod: *Lexikon der Sprachwissen-schaft*. Stuttgart : Kröner, 2002
- [Carlson u. Marcu 2001] CARLSON, Lynn ; MARCU, Daniel: Discourse Tagging Reference Manual / Univ. of Southern California/ISI. 2001. – Forschungsbericht
- [Dipper 2005] DIPPER, Stefanie: XML-based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation. In: ECKSTEIN, Rainer (Hrsg.) ; TOLKSDORF, Robert (Hrsg.): *Proceedings of Berliner XML Tage*, 2005, S. 39–50
- [Erteschik-Shir 2007] ERTESCHIK-SHIR, Nomi: *Information structure: The syntax-discourse interface*. Oxford : Oxford University Press, 2007
- [Freeman 1991] FREEMAN, James B.: *Dialectics and the Macrostructure of Argument*. Berlin : Foris, 1991
- [Götze u. a. 2007] GÖTZE, Michael ; ENDRISS, Cornelia ; HINTER-WIMMER, Stefan ; FIEDLER, Ines ; PETROVA, Svetlana ; SCHWARZ, Anne ; SKOPETEAS, Stavros ; STOEL, Ruben ; WESKOTT, Thomas:

- Information structure. In: *Information structure in cross-linguistic corpora: annotation guidelines for morphology, syntax, semantics, and information structure* Bd. 7. 2007, S. 145–187
- [Grosz u. a. 1995] GROSZ, Barbara ; JOSHI, Aravind ; WEINSTEIN, Scott: Attention, Intentions, and the Structure of Discourse. In: *Computational Linguistics* 21 (1995), Nr. 2, S. 203–226
- [Iruskieta u. a. 2014] IRUSKIETA, M. ; CUNHA, I. da ; TABOADA, M.: Principles of a qualitative method for rhetorical analysis evaluation: A contrastive analysis English-Spanish-Basque. In: *Language Resources and Evaluation* 49 (2014), Nr. 2, S. 263–309
- [Jacobs 2001] JACOBS, Joachim: The dimensions of Topic-Comment. In: *Linguistics* 39 (2001), Nr. 4, S. 641–681
- [Jasinskaja u. a. 2007] JASINSKAJA, Katja ; MAYER, Jörg ; BOETHKE, Jutta ; NEUMANN, Annika ; PELDSZUS, Andreas ; RODRÍGUEZ, Kepa J.: *Discourse Tagging Guidelines for German Radio News and Newspaper Commentaries*. 2007. – Ms., Universität Potsdam
- [Krasavina u. Chiarcos 2007] KRASAVINA, Olga ; CHIARCOS, Christian: PoCoS: The Potsdam Coreference Scheme. In: *Proc. of the Linguistic Annotation Workshop (LAW) at ACL-07*. Prague, 2007
- [Krifka 2007] KRIFKA, Manfred: Basic notions of information structure. In: FÉRY, Caroline (Hrsg.) ; KRIFKA, Manfred (Hrsg.): *Interdisciplinary Studies on Information Structure* 6. Potsdam : Universitätsverlag, 2007, S. 13–56
- [Mann u. Thompson 1988] MANN, William ; THOMPSON, Sandra: Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. In: *TEXT* 8 (1988), S. 243–281
- [Marcu 2000] MARCU, Daniel: *The theory and practice of discourse parsing and summarization*. Cambridge/MA : MIT Press, 2000
- [Martin 1992] MARTIN, James R.: *English text: system and structure*. Philadelphia/Amsterdam : John Benjamins, 1992

- [Matić u. a. 2014] MATIĆ, Dejan ; GIJN, Rik van ; JR., Robert Van V.: Information structure and reference tracking in complex sentences: An overview. In: GIJN, Rik van (Hrsg.) ; HAMMOND, Jeremy (Hrsg.) ; MATIĆ, Dejan (Hrsg.) ; PUTTEN, Saskia van (Hrsg.) ; GALUCIO, Ana V. (Hrsg.): *Information structure and reference tracking in complex sentences*. Amsterdam : John Benjamins, 2014
- [Pasch u. a. 2003] PASCH, Renate ; BRAUSSE, Ursula ; BREINDL, Eva ; WASSNER, Ulrich H.: *Handbuch der deutschen Konnektoren*. Berlin/New York : Walter de Gruyter, 2003
- [Peldszus u. Stede 2013] PELDSZUS, Andreas ; STEDE, Manfred: From argument diagrams to argumentation mining in texts: a survey. In: *Int'l Journal of Cognitive Informatics and Natural Intelligence (IJCINI)* 7 (2013), Nr. 1, S. 1–31
- [Prasad u. a. 2008] PRASAD, Rashmi ; DINESH, Nikhil ; LEE, Alan ; MILTSAKAKI, Eleni ; ROBALDO, Livio ; JOSHI, Aravind ; WEBBER, Bonnie: The Penn Discourse TreeBank 2.0. In: *In Proceedings of LREC, 2008*
- [Recasens u. a. 2012] RECASENS, Marta ; MARTI, Antonia ; ORASAN, Constantin: Annotating Near-Identity from Coreference Disagreements. In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, 2012
- [Reinhart 1982] REINHART, Tanya: Pragmatics and Linguistics. An analysis of sentence topics. In: *Philosophica* 27 (1982), S. 53–94
- [Ritz u. a. 2008] RITZ, Julia ; DIPPER, Stefanie ; GÖTZE, Michael: Annotation of Information Structure: An Evaluation Across Different Types of Texts. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*. Marrakech, Morocco, 2008, S. 2137–2142
- [Sanders u. a. 1992] SANDERS, Ted ; SPOOREN, Wilbert ; NOORDMAN, Leo: Toward a taxonomy of coherence relations. In: *Discourse Processes* 15 (1992), S. 1–35

- [Schneider u. Raue 1996] SCHNEIDER, W. ; RAUE, P.: *Handbuch des Journalismus*. Hamburg : Rowohlt, 1996
- [Stede 2004] STEDE, Manfred: The Potsdam Commentary Corpus. In: *Proceedings of the ACL Workshop on Discourse Annotation*. Barcelona, 2004, S. 96–102
- [Stede 2007] STEDE, Manfred: *Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik*. Tübingen : Narr, 2007
- [Stede u. Neumann 2014] STEDE, Manfred ; NEUMANN, Arne: Potsdam Commentary Corpus 2.0: Annotation for Discourse Research. In: *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. Reikjavik, 2014, S. 925–929
- [Swales 1990] SWALES, John M.: *Genre analysis: English in academic and research settings*. Cambridge : Cambridge University Press, 1990
- [Taboada u. Mann 2006] TABOADA, Maite ; MANN, William: Rhetorical Structure Theory: Looking back and moving ahead. In: *Discourse Studies* 8 (2006), Nr. 4, S. 423–459
- [Teufel u. Moens 2002] TEUFEL, Simone ; MOENS, Marc: Summarizing Scientific Articles – Experiments with Relevance and Rhetorical Status. In: *Computational Linguistics* 28 (2002), Nr. 4, S. 409–445
- [Toulmin 1958] TOULMIN, Stephen: *The Uses of Argument*. Cambridge : Cambridge University Press, 1958
- [van Valin jr. 2005] VALIN JR., Robert D.: *Exploring the syntax-semantics interface*. Cambridge : Cambridge University Press, 2005
- [Verhagen 2005] VERHAGEN, Arie: *Constructions of intersubjectivity: Discourse, Syntax, and Cognition*. Oxford : Oxford University Press, 2005

Das Potsdamer Kommentarkorpus ist eine Sammlung von Zeitungstexten, die dem Genre ‚Kommentar‘ zuzuordnen sind. Der öffentlich verfügbare Teil besteht aus 175 Texten aus der Märkischen Allgemeinen Zeitung, die hinsichtlich Syntax, Koreferenz, Konnektoren und Rhetorische Struktur manuell annotiert wurden. Weitere Ebenen werden bei zukünftigen Korpusversionen hinzukommen. Dieses Buch enthält die Annotationsrichtlinien, die der Bearbeitung des öffentlichen Teils des Korpus zugrunde lagen, sowie auch anderer Teile, bei denen mit weiteren Annotationsebenen experimentiert wurde. Die meisten der Richtlinien werden auch für ähnliche Text-Genres und für andere Sprachen verwendbar sein.

ISSN 2190-4545
ISBN 978-3-86956-343-5



9 783869 563435