

Maja Bärenfänger

# **Ebenen des Themas**

**Zur Interaktion von Thema, Text und Wissen**

Inaugural-Dissertation  
zur Erlangung des Doktorgrades der Philologie  
des Fachbereichs 05 - Sprache, Literatur, Kultur  
der Justus-Liebig-Universität Gießen  
Dekanin: Prof. Dr. Cora Dietl  
1. Gutachter: Prof. Dr. Henning Lobin  
2. Gutachter: Prof. Dr. Helmuth Feilke  
Tag der Disputation: 1. März 2011

*Für meine Eltern*

# Inhalt

<b>1</b>	<b>EINLEITUNG</b>	<b>3</b>
<b>1.1</b>	<b>Einführung in das Untersuchungsgebiet</b>	<b>2</b>
1.1.1	Zur Problematik des Terminus und Konzepts Thema	2
1.1.2	Geschichte und Tradition des Konzepts Thema	4
1.1.3	Verwendungsweisen von „Thema“	9
1.1.4	Disziplinen und Anwendungsbereiche, für die das Konzept Thema relevant ist	11
<b>1.2</b>	<b>Forschungsfragen und Zielsetzung</b>	<b>14</b>
<b>1.3</b>	<b>Überblick über den Aufbau der Arbeit</b>	<b>18</b>
<b>2</b>	<b>THEMA-KONZEPTE: BESTANDSAUFNAHME UND KATEGORISIERUNG</b>	<b>19</b>
<b>2.1</b>	<b>Thema als grammatisches Konzept</b>	<b>20</b>
2.1.1	Thema und Satzposition	21
2.1.2	Thema und grammatische Rolle	23
2.1.3	Thema und thematische/semantische Rolle	24
2.1.4	Fazit	26
<b>2.2</b>	<b>Thema als semantisches Konzept</b>	<b>27</b>
2.2.1	Thema als Gegenstand, über den etwas gesagt wird	28
2.2.1.1	Thema als Mitteilungsgegenstand	28
2.2.1.2	Thema als fortlaufend konstituierter zentraler Gegenstand des Textes	29
2.2.2	Thema als abstrakter Kern des Textinhalts	37
2.2.3	Fazit	41
<b>2.3</b>	<b>Thema als kommunikativ-pragmatisches Konzept</b>	<b>42</b>
2.3.1	Thema als das kommunikativ am wenigsten dynamische Element des Satzes	43
2.3.2	Thema als Fragliches	45
2.3.3	Thema als Mittel zur Strukturierung von Diskursen: Thematische Textmuster	50
2.3.4	Fazit	57
<b>2.4</b>	<b>Kognitiv orientierte Thema-Konzepte</b>	<b>59</b>
<b>2.5</b>	<b>Das Konzept Thema in computerlinguistischen Arbeiten</b>	<b>63</b>

---

<b>3</b>	<b>THEORETISCHE GRUNDLAGEN</b>	<b>68</b>
<b>3.1</b>	<b>Texttheorie</b>	<b>69</b>
3.1.1	Texte als komplexe sprachliche Handlungen	69
3.1.2	Texte als komplexe indexikalische Zeichen	71
3.1.2.1	Zum Begriff der „kommunikativen Situation“	71
3.1.2.2	Indexikalische Zeichen	75
3.1.2.3	Zur Semiose einfacher und komplexer indexikalischer Zeichen	79
3.1.3	Texte als auf mehreren Ebenen strukturierte Äußerungseinheiten	82
3.1.3.1	Informationsstruktur	87
3.1.3.2	Lexikalische Struktur	90
3.1.3.3	Referenzielle Struktur	93
3.1.3.4	Relationale Struktur	95
3.1.3.5	Visuelle und logische Textstruktur	97
3.1.3.6	Generische Texttypstruktur	99
<b>3.2</b>	<b>Kognitive Grundlagen der Textproduktion und -rezeption</b>	<b>100</b>
3.2.1	Wissensbestände beim Kommunizieren	100
3.2.2	Kognitive Schemata und Frames	102
3.2.3	Kognition und Textproduktion: Frames als Listen von Themen	110
<b>3.3</b>	<b>Zum Kommunikationsprozess: Textproduktion und -rezeption</b>	<b>114</b>
3.3.1	Textproduktion	114
3.3.2	Textrezeption	121
<b>3.4</b>	<b>Zusammenfassung</b>	<b>127</b>
<b>4</b>	<b>EIN EBENEN-BASIERTES THEMA-MODELL</b>	<b>131</b>
<b>4.1</b>	<b>Einleitung</b>	<b>132</b>
4.1.1	Einführung in das Ebenen-basierte Thema-Modell	132
4.1.2	Einordnung der zentralen Thema-Konzepte aus Kapitel 2 in die drei Ebenen des integrativen Thema-Modells	138
<b>4.2</b>	<b>Zum Zusammenhang von Thema und Wissen</b>	<b>141</b>
4.2.1	Thema und Sachwissen: Thema als Sachverhaltskomplex	142
4.2.2	Thema und Textmusterwissen: Globale und lokale thematische Muster	153
4.2.3	Thema und Sprachwissen: sprachliche Möglichkeiten der Signalisierung, Etablierung und Strukturierung des Themas	156
4.2.4	Thema und Wissen über die Kommunikationssituation	158
4.2.5	Wissensressourcen für die maschinelle Themenidentifikation	159

---

<b>4.3</b>	<b>Reflexionen und Realisationen des Themas im Text</b>	<b>170</b>
4.3.1	Thema als zentraler Gegenstand	171
4.3.1.1	Grammatische und informationsstrukturelle Markierung des Themas	172
4.3.1.2	Koreferenz: Wiederaufnahme zentraler Diskursentitäten	179
4.3.1.3	Thematischer Wortschatz, lexikalische Ketten und metasprachliche lexikalische Thema-Marker	185
4.3.1.4	Visuelle und logische Thema-Marker	198
4.3.2	Zusammenhänge zwischen Aspekten des Themas	203
4.3.2.1	Kohärenzrelationen	203
4.3.2.2	Texttypstrukturkategorien	217
4.3.3	Zur „Verrechnung“ der Reflexionen des Themas	225
<b>4.4</b>	<b>Die Rolle des Themas für die Produktion und Rezeption von Texten</b>	<b>235</b>
4.4.1	Thematisieren als textorganisierende Handlung	236
4.4.2	Thema als „Verstehenshorizont“	244
4.4.3	Architektur und Komponenten eines Systems zur maschinellen Themenidentifikation	247
<b>5</b>	<b>ZUSAMMENFASSUNG UND AUSBLICK</b>	<b>254</b>
<b>6</b>	<b>LITERATUR</b>	<b>265</b>
<b>6.1</b>	<b>Forschungsliteratur</b>	<b>265</b>
<b>6.2</b>	<b>Quellen zur Textsammlung „Waldsterben“</b>	<b>288</b>
<b>7</b>	<b>ANHANG: BEISPIELTEXTE IM ORIGINAL</b>	<b>289</b>
<b>7.1</b>	<b>Beispiel 1: „Gewalt gegen Castor nimmt zu“</b>	<b>289</b>
<b>7.2</b>	<b>Beispiel 2: „Heringe“</b>	<b>290</b>

## Danksagung

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die mich auf unterschiedliche Weise unterstützt und zum Entstehen dieser Arbeit beigetragen haben.

Besonders herzlich bedanken möchte ich mich bei dem Betreuer meiner Arbeit, Prof. Dr. Henning Lobin. Ohne seine intensive langjährige fachliche und moralische Unterstützung sowie sein Verständnis dafür, dass sich nach der Geburt eines Kindes nicht nur die Arbeitszeiten ändern, hätte ich diese Arbeit vermutlich niemals beendet. Und vermutlich hätte ich sie auch nie begonnen, wenn er mein Literaturwissenschafts- und Linguistikstudium nicht mit dem Seminar „Einführung in die Texttechnologie“ gekreuzt und mir so den Weg in meine fachliche Wahlheimat Computerlinguistik und Texttechnologie gezeigt hätte.

Mein besonderer Dank gilt auch Prof. Dr. Helmuth Feilke, dem Zweitgutachter der Arbeit, dem ich viele wertvolle Anregungen verdanke, die diese Arbeit sehr bereichert haben. Auch für die guten Diskussionen mit Prof. Dr. Gerd Fritz möchte ich mich herzlich bei ihm bedanken, ebenso wie für seine hilfreichen Anmerkungen zu meiner Arbeit.

Großen Dank schulde ich Frank Binder, Dr. Malgorzata Dynkowska, Dr. Harald Längen, Dr. Georg Rehm und Andrea Schüttler, die einzelne Kapitel oder die gesamte Arbeit Korrektur gelesen haben und deren fundierte kritische Hinweise mich oft noch einen Schritt tiefer in die Materie getrieben haben.

Für anregende Diskussionen, Literaturhinweise, fachliche Ratschläge, technische Unterstützung, die Begleitung in einer oder mehreren Entstehungsphasen der Dissertation sowie absolut unersetzliche moralische Unterstützung möchte ich mich außerdem bedanken bei: Dr. Iris Bons, Dr. Malgorzata Dynkowska, Mirco Hilbert, Dr. Harald Längen, Prof. Dr. Joybrato Mukherjee, Dr. Georg Rehm, Prof. Dr. Dietmar Rösler, Prof. Dr. Britt-Marie Schuster, Prof. Dr. Torsten Steinhoff sowie den Mitgliedern der linguistischen Promotionskolloquien am Fachbereich 05 der Justus-Liebig-Universität Gießen.

Desweiteren möchte ich meinen bislang noch nicht genannten Arbeitskollegen an der Professur für Angewandte Sprachwissenschaft und Computerlinguistik, insbesondere Ralf Grünspahn und Silvia Baumgart, meinen Kollegen am Zentrum für Medien und Interaktivität (ZMI) sowie in der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“ für die gute Zusammenarbeit danken.

Last but not least: Von ganzem Herzen danke ich meiner Familie, insbesondere meinen Eltern und Großeltern, dafür, dass sie mich so bedingungslos unterstützt und an mich geglaubt haben. Mein allergrößter Dank gilt allerdings meinem Mann Martin und meinem Sohn Ben – danke, dass ihr mich und meine Arbeit mit so viel Liebe und Geduld mitgetragen und mich ermuntert (Martin) bzw. aufgemuntert (Ben) habt, wenn mal wieder alles schlecht war!

## Einleitung

„Worum geht’s?“ Eine wichtige, wenn nicht *die* zentrale Frage für jeden, der in eine kommunikative Situation gerät – sei es als Beobachter (z.B. bei einer Fernsehtalkshow) oder als (potentiell) aktiver Beteiligter (z.B. bei einer face-to-face-Diskussion, aber auch als potentieller Leser eines Textes). Sie ist zumindest eine der ersten Fragen, die wir uns stellen, und ihre Beantwortung ist mitentscheidend dafür, ob wir weiterzappen bzw. uns abwenden oder ob wir uns – zumindest kurzzeitig – in die Situation einklinken. Wir können diese Frage laut an jemand Dritten richten oder sie uns selbst stellen – in diesem Fall müssen wir hoffen, durch das aktive Verfolgen und Interpretieren der kommunikativen Situation selbst eine Antwort auf unsere Frage zu erhalten. Dabei ist es ein Gebot der Ökonomie, möglichst schnell herauszufinden, was das Thema der Talkshow oder des Textes ist – und ob es uns interessiert oder wir unsere Zeit besser verwenden können. Meistens ist dies für einen Muttersprachler kein Problem – wir verstehen in der Regel recht schnell, worum es in einer kommunikativen Situation geht. Um das Thema zu ermitteln, beziehen wir viele verschiedene Faktoren mit ein, sprachliche (in erster Linie die Lexik, aber auch Textmuster) und außersprachliche (den Kommunikationskontext, visuelle Mittel etc.). Gleichzeitig stellen wir einen Bezug zu unseren kommunikativen Erfahrungen her und nutzen unser sprachliches und außersprachliches Wissen und Können, um eine erste Hypothese über das Thema des Textes zu bilden. An genau dieser Aufgabe allerdings scheitern Computer bis heute. Auch ein Modell zur Operationalisierung der Analyse des Themas, das eine zentrale Grundlage für die maschinelle Identifikation des Themas darstellt, wurde bisher nicht in zufriedenstellender Weise entwickelt. Selbst die Frage, was das Thema<sup>1</sup> (als linguistisches Konzept<sup>2</sup>) überhaupt ist, lässt sich nicht ohne Weiteres beantworten, ohne Widerspruch von der einen oder anderen Seite zu ernten. In der vorliegenden Arbeit soll deshalb der Versuch unternommen werden, einerseits einen Beitrag zur linguistischen Thema-Forschung zu liefern und zur Klärung o.g. Frage beizutragen, und andererseits computerlinguistische und sprachtechnologische Anwendungen zur Identifikation und Extraktion des Themas in Texten auf eine solidere theoretische Grundlage zu stellen. In diesem Zusammenhang soll sowohl ein Vorschlag zur Operationalisierung der Analyse des Themas gemacht werden als auch ein formales, auf texttechnologischen Methoden basierendes Modell zur Repräsentation und Identifikation des Themas skizziert werden.

---

<sup>1</sup> Der Ausdruck „Thema“ wird in dieser Arbeit mehrheitlich metasprachlich verwendet, d.h. zur Bezugnahme auf ein abstraktes linguistisches Konzept. Aus Gründen der Lesbarkeit wurde dennoch davon abgesehen, diese metasprachliche Verwendung von „Thema“ durch Anführungszeichen zu signalisieren.

<sup>2</sup> Die Ausdrücke „Konzept“ und „Begriff“ werden im Folgenden synonym und im Sinne von Schwarz/Chur als „mentale Informationseinheit, in der wir Wissen über die Welt abspeichern“ verwendet (Schwarz und Chur 2004, 222).



## 1.1 Einführung in das Untersuchungsgebiet

### 1.1.1 Zur Problematik des Terminus und Konzepts Thema

Das Konzept Thema stellt eines der zentralen und gleichzeitig umstrittensten Konzepte in der Textlinguistik dar: Kaum ein Werk zur Textlinguistik kommt ohne ein Kapitel oder einen längeren Abschnitt zum Thema aus (siehe z.B. Dressler 1973; Coseriu 1980; Heinemann und Viehweger 1991; Vater 2001; Gansel und Jürgens 2002; Heinemann und Heinemann 2002; Adamzik 2004b; Stede 2007). Trotz der vielfältigen und intensiven Auseinandersetzungen mit dem Konzept existiert allerdings weder ein Konsens über die Definition des Themas noch über geeignete Analyseverfahren:

Obwohl Thema und thematischer Zusammenhang von Texten und Kommunikationen in den letzten Jahren zunehmend die Aufmerksamkeit von Textlinguisten gefunden haben, kann bis heute von einem einigermaßen befriedigendem Konzept für die Behandlung dieses Gegenstands in der Textlinguistik kaum die Rede sein. (Fritz 1982, 205)

In fact, 'topic' could be described as the most frequently used, unexplained term in the analysis of discourse. (Brown und Yule 1983, 70)

Trotz der Bedeutung, die dem Aspekt ‚Thema‘ in der Textlinguistik, wie erwähnt, zugemessen wird, existiert bislang dafür keine einheitliche oder allgemein akzeptierte Definition. (Lötscher 1987, 3)

The notion of 'topic' has remained somewhat murky in linguistics. (Givón 1992, 10)

'Topic' is a commonly used term for a variety of approaches to a range of functional concepts. Few terms in linguistics, however, are so widely used without proper justification. (Schlobinski und Schütze-Coburn 1992, 89)

In der Sprachwissenschaft finden sich unterschiedliche, z.T. vom Alltag weit entfernte Themabegriffe. Von einem Konsens kann keine Rede sein. (Hoffmann 2000, 344)

Studies in this area have been characterised by terminological profusion and confusion because very different positions have been taken on the appropriate criteria for the definition and identification of the notions of Theme/Topic (and related categories). (Gómez-González 2001, 4)

Obwohl wir im Normalfall kein Problem haben, zumindest grosso modo spontan anzugeben, was wir für das Thema eines bestimmten Textes halten, sind wir mit großen Schwierigkeiten konfrontiert, wenn es darum geht, mithilfe rein linguistischer Methoden zu bestimmen, was das Thema eines Textes ist. Man kann wohl sagen, dass es der Textlinguistik bis heute nicht gelungen ist, klar operationalisierte Verfahren zur eindeutigen Bestimmung bzw. ‚Destillation‘ von Textthemen zu entwickeln. (Linke et al. 2004, 237)

Wie aus den Zitaten deutlich wird, beziehen sich die Autoren sowohl auf das „Konzept“ bzw. den „Begriff“ Thema (siehe Fritz 1982; Givón 1992; Hoffmann 2000; Gómez-González 2001), als auch auf den Ausdruck bzw. Terminus („term“) Thema (siehe Brown und Yule 1983; Schlobinski und Schütze-Coburn 1992). Die Autoren konstatieren alle in unterschiedlicher Weise den Mangel einer befriedigenden und allgemein akzeptierten linguistischen Definition oder den Mangel eines befriedigenden Analyseverfahrens (siehe Linke et al. 2004). Mit ihrer unterschiedlichen Rede vom „Terminus“ bzw. „Konzept“ Thema verweisen sie gleichzeitig aber implizit noch auf einen weiteren zentralen Punkt in

der Problematik um das Thema: In der linguistischen Welt existieren nicht nur verschiedene Thema-Konzepte, sondern auch unterschiedliche Termini, mit denen auf diese Konzepte Bezug genommen wird – die konzeptuelle Diversität spiegelt sich in einer terminologischen Diversität, und umgekehrt (für einen terminologischen Überblick siehe Gülich et al. 1977, 89; Gómez-González 2001, 6f.; Kruijff-Korbayova und Steedman 2003, 254). Die vier ähnlichsten Ausdrücke, die in diesem Kontext zur Referenzierung der diversen noch zu spezifizierenden Thema-Konzepte auf Satz- oder Textebene verwendet werden, sind „Thema“, „Theme“, „Topik“ und „Topic“. Theme und Topic sind dabei weder englischsprachige Äquivalente zu Thema und Topik noch beziehen sie sich konsistent auf jeweils ein spezifisches Konzept Thema.<sup>3</sup> Mit Thema, Theme, Topik oder Topic wird, je nach Autor, auf den zentralen Gegenstand (im Satz oder Text) oder eine Folge von wiederaufgenommenen Diskursentitäten<sup>4</sup> referiert (siehe Kap. 2.2.1), ebenso wie auf die „Makroproposition“ des Textes (siehe Kap. 2.2.2), ebenso wie auf die impliziten Fragen, die der Autor durch den Text zu beantworten sucht (siehe Kap. 2.3.2). Umgekehrt werden unterschiedliche Ausdrücke gebraucht, um auf dasselbe sprachliche Phänomen Bezug zu nehmen: Als Ausgangspunkt der Satzaussage werden Thema oder Topik genannt (siehe Kap. 2.2.1); die Satzkonstituente in Satzerstposition ist Theme, Topik oder Topic (siehe Kap. 2.1.1); die bekannte Diskursentität im Satz ist Thema, Theme oder Topic (siehe Kap. 2.4).

Vertieft wird diese terminologische und konzeptuelle Konfusion noch dadurch, dass das Konzept Thema<sup>5</sup> für die Beschreibung sprachlicher Phänomene sowohl auf Satz- als auch Text- bzw. Diskursebene<sup>6</sup> herangezogen wird. Unterscheiden lassen sich hier Thema-Konzepte auf der Ebene der Grammatik, der Informationsstruktur, der Semantik, der Lexik, der Pragmatik sowie kognitiv orientierte Arbeiten zum Thema. Das Konzept wird also zur Beschreibung von Phänomenen auf sehr unterschiedlichen sprachlichen Ebenen verwendet. Mit Recht kann hier deshalb konstatiert werden, dass die Forschungslage zum Thema sehr heterogen und widersprüchlich ist.<sup>7</sup>

---

<sup>3</sup> Der deutsche Terminus „Topik“ wird beispielsweise sowohl zur Bezeichnung einer Einheit auf Textebene gebraucht – „Topik“ als kognitives Konzept mit der größten Verbindungsdichte zu anderen Konzepten (vgl. de Beaugrande und Dressler 1981, 89) – als auch zur Referenz auf sprachliche Entitäten im Satz – unterschiedliche allerdings: Topik als Element in Satzerstposition (vgl. z.B. Welke 2002; Dürr und Schlobinski 2006, siehe auch den Begriff der Topikalität (bzw. Topicality), mit dem auf die an das Satzthema geknüpften Wortstellungsregularitäten bzw. die vom Satzthema gesteuerte Gliederung und Perspektivierung des Satzes rekuriert wird), als unbetontes, nicht fokussiertes Satzelement (im Konzeptpaar „Topik-Fokus“) oder als Satzelement, über das im Satz Informationen gegeben werden (im Konzeptpaar „Topik-Kommentar“) (vgl. z.B. Brandt und Rosengren 1992; Altmann 1993; Molnár 1993).

<sup>4</sup> Eine Diskursentität wird hier definiert als ein Objekt in der Diskurswelt, das mittels einer sprachlichen Einheit referenziert wird. In gleicher Bedeutung wird auch der Terminus „Diskursreferent“ gebraucht (siehe auch Fußnote 44).

<sup>5</sup> Der Einfachheit halber rede ich weiterhin vom Konzept Thema (Singular), obwohl mittlerweile deutlich geworden sein sollte, dass es *das* Konzept Thema nicht gibt.

<sup>6</sup> „Diskurs“ wird in der vorliegenden Arbeit aufgefasst als sprachliche Einheit, die größer als ein Satz ist; damit sind sowohl geschriebene Texte als auch gesprochene Dialoge gemeint.

<sup>7</sup> „The literature on topic is large and often contradictory.“ (Glanzberg 2002, 353).

### 1.1.2 Geschichte und Tradition des Konzepts Thema

Das linguistische Konzept Thema wird häufig (siehe Schlobinski und Schütze-Coburn 1992; Molnár 1993; Hoffmann 2000; Gómez-González 2001) auf Arbeiten von Hermann Paul (1898) und Georg v. d. Gabelentz (1972 [1891]) zurückgeführt, deren Ideen später von dem Begründer der Prager Schule, Vilém Mathesius, weiterentwickelt wurden. Paul und v. d. Gabelentz führten das Konzept des psychologischen Subjekts und psychologischen Prädikats als Gegenpol zum grammatischen Subjekt und Prädikat ein. Sie definieren das psychologische Subjekt als „das, worüber der Sprechende den Hörenden denken lassen, worauf er seine Aufmerksamkeit hinleiten will, das psychologische Prädikat dasjenige, was er darüber denken soll“ (Paul 1898, 112). Das psychologische Subjekt wird außerdem als kontextuell weniger relevant (wichtig)<sup>8</sup> als das psychologische Prädikat betrachtet. Mit dieser unterschiedlichen Relevanz (Gewichtung) der beschriebenen Kategorien verbinden sowohl Paul als auch v. d. Gabelentz eine bestimmte Abfolge im Satz. Von der Gabelentz ging davon aus, dass das psychologische Subjekt im Satz am Anfang stehe (vgl. von der Gabelentz 1972 [1891], 369f.). Paul vertrat dagegen die Auffassung, dass dies erstens nicht für alle Sprachen gelte und dass es zweitens auch innerhalb einer Sprache Ausnahmen von der Regel „Subjekt vor Prädikat“ gebe, z.B. könne „lebhafter Affekt die Veranlassung werden, dass sich das Präd. an die Spitze drängt“ (Paul 1898, 114). Wie im Folgenden gezeigt werden soll, wurde diese Koppelung von funktionaler Kategorie und linearer Abfolge im Satz in den folgenden Jahrzehnten vor allem in grammatisch orientierten Thema-Konzeptionen immer wieder aufgegriffen, wobei beide Auffassungen – das Thema steht eher am Anfang vs. das Thema kann in bestimmten Fällen auch rechtsversetzt sein – vertreten wurden.

In den 20er Jahren des 20. Jahrhunderts wurden durch Hermann Ammann und Vilém Mathesius für das psychologische Subjekt/das psychologische Prädikat die Begriffe „Thema“ und „Rhema“ eingeführt (vgl. Hoffmann 2000, 344). Damit war der Grundstein für die Bindung des linguistischen Terminus Thema an satzbezogene Phänomene gelegt.<sup>9</sup> Mathesius verwendet das Begriffspaar „Thema-Rhema“ im Kontext seines Konzepts der „Funktionalen Satzperspektive“. Er definiert Thema wie folgt:

Der zweite Teil des Satzes enthält die Basis der Mitteilung oder Thema [sic], nach der älteren Terminologie das psychologische Subjekt, d.h. die verhältnismäßig bekannten oder auf der Hand liegenden Dinge, von denen der Sprecher ausgeht. In den ruhig vorgetragenen Sätzen wird zuerst das Thema und erst dann der Mitteilungskern [das Rhema; M.B.] gesetzt (objektive Abfolge), während bei einer erregten Behauptung eine umgekehrte Abfolge vorherrscht (subjektive Abfolge). (Mathesius 1971 [1929], 6f.)

<sup>8</sup> Das Merkmal „wichtig“ wird in diesem Kontext für die Information mit dem höchsten Informationsgehalt bzw. der höchsten kommunikativen Dynamik gebraucht.

<sup>9</sup> Diese stark ausgeprägte Forschungstradition der Assoziation des linguistischen Konzepts Thema mit der Satzebene, die dem Alltagsverständnis von Thema und dem Verständnis des Konzepts Thema von Text- und Gesprächslinguisten konträr gegenübersteht, hat meiner Ansicht nach mit dazu beigetragen, dass es bis heute keinen Konsens über eine Definition des Konzepts gibt.

Diese Definition umfasst drei zentrale Merkmale, die zuvor bereits für die Charakterisierung des psychologischen Subjekts gebraucht wurden und die grundlegend für viele der in den folgenden Jahrzehnten vorgeschlagenen Thema-Definitionen sind:<sup>10</sup>

1. Semantik: Charakterisierung des Themas als (semantische) Basis der Mitteilung
2. Informationsstruktur: Thema als Bekanntes<sup>11</sup>
3. Syntax: Abfolge „Thema vor Rhema“ als syntaktische Normalfolge

Während die hier skizzierte Tradition des Thema-Konzepts an das Ende des 19. (Begriffsbildung „psychologisches Subjekt“ vs. „psychologisches Prädikat“) bzw. den Anfang des 20. Jahrhunderts (Begriffsbildung „Thema vs. Rhema“) datiert wird, also als sehr jung gelten kann, lässt sich das linguistische Thema-Konzept auch auf sehr viel ältere Wurzeln zurückführen:

Der Begriff Textthema baut auf dem Themabegriff der antiken Rhetoriker auf, die unter Thema ‚zu behandelnde strittige Proposition‘ verstanden, während der Begriff Satzthema letztlich auf den Subjektsbegriff der antiken Logiker und Grammatiker zurückgeht, die damit logisch den Gegenstand, über den eine Aussage gemacht wird und der grammatisch gewöhnlich im Nominativ steht und das Verb in Person und Numerus regiert, bezeichneten. (Lötscher 1987, 233f.)

Der Begriff des „psychologischen Subjekts“ sowie Ammans und Mathesius Thema-Konzept lassen sich demnach in die Traditionslinie des logischen Subjektsbegriffs der antiken Logiker und Grammatiker einordnen.

Ende der 1960er Jahre wurde das von Mathesius spezifizierte linguistische Konzept Thema vom Satz auf den Text übertragen. František Daneš, ein Vertreter der Prager „Funktionalen Satzperspektive“ (siehe Daneš 1970; Beneš und Vachek 1971; Mathesius 1971 [1929]; Sgall et al. 1973; Firbas 1992), entwickelte ein Modell zur Beschreibung der „thematischen Progression“ im Text, das auf den Bezügen der Satzthemen bzw. -rhemen in aufeinander folgenden Sätzen basiert (Daneš 1970, 1976).<sup>12</sup> Aus seinem Ansatz resultiert kein eigenständiges Textthema-Konzept, allerdings legt die von ihm vorgeschlagene Ausweitung der Analyse des Themas vom Satz auf den Text nahe, dass sich Mathesius auf den Satz bezogene Thema-Begriff ohne große konzeptuelle Veränderungen auf den Text über-

---

<sup>10</sup> Die meisten Arbeiten in der Tradition von Mathesius beschäftigen sich dabei mit der Informationsstruktur, einer in der Schnittstelle von Grammatik und Pragmatik wurzelnden Strukturebene, die die kommunikative Gewichtung der einzelnen Satzteile beschreibt, z.B. in „Thema vs. Rhema“, „Topic vs. Comment“, „Topik/Theme/Topic/Ground“ vs. „Fokus/Focus/Emphasis“, „bekannt/given/known/ salient“ vs. „neu/new/unknown/non-salient“, „Vordergrund“ vs. „Hintergrund“ (siehe z.B. Molnár 1993; Reis 1993; Lambrecht 1994; Firbas 1995; Fries 1995; Garrod 1995; Martin 1995; Matthiessen 1995; Ramm und Villiger 1995; Jäger und Oshima 2002; Green 2004; Gundel und Fretheim 2004; Jasinskaja et al. 2004).

<sup>11</sup> Die Diskussion um das (umstrittene und schwierig operationalisierbare) Merkmal der Bekanntheit – was kann als „bekannt“ gelten, wie wird „Bekanntheit“ erklärt (kontextuell, kognitiv) – wird im Kapitel 2.4 detaillierter dargestellt.

<sup>12</sup> In ähnlicher Form wurde später auch in den Arbeiten in der Tradition der Systemisch-Funktionalen Linguistik die Entwicklung von aufeinander folgenden Satzthemen analysiert, um so Aussagen über die thematische Progression des Textes machen zu können (vgl. Martin 1995, 228ff.).

tragen lässt. Forschungsarbeiten in der Tradition von Daneš (1970; 1976), die einen Fokus auf die Analyse aufeinander folgender Mathesiusscher Satzthemen legen, um so Aussagen über das Textthema und die thematische Struktur bzw. die thematische Progression des Textes zu gewinnen, sind bis heute sehr populär und stellen einen Kernbereich der Forschungen zum Textthema bzw. der thematischen Struktur dar – siehe z.B. die Arbeiten zu „topic continuity“ (Givón 1983a; 1983b; 1992) oder „thematischen Organisation“ (Hoffmann 1997; 2000; Lenz und Storrer 2003) sowie einige Arbeiten zur thematischen Entwicklung im Rahmen der Systemisch-Funktionalen Linguistik (Martin 1995; Ramm und Viliger 1995; Ventola 1995; Lavid 2000; Taboada und Lavid 2003; Schwarz et al. 2008).

Die Entscheidung der Linguisten der Funktionalen Satzperspektive sowie der Systemisch-Funktionalen Linguistik, den logisch-grammatisch geprägten Satzthema-Begriff auf den Text zu übertragen und so implizit das grammatisch bzw. logisch motivierte Satzthema-Konzept mit dem rhetorischen Textthema-Konzept gleichzusetzen, war für die Diskussion um den Begriff Thema folgenscher: Meines Erachtens stellt sie eine der Ursachen für die vielfältigen und heterogenen Thema-Konzepte in der Linguistik dar.

Neben den in der Tradition des logisch-grammatischen Satzthema-Konzepts stehenden Arbeiten zum Textthema entstanden seit Anfang der 1970er Jahre zahlreiche Forschungsarbeiten zum Textthema in der Tradition des rhetorischen Thema-Begriffs (siehe z.B. Keenan und Schieffelin 1976; Schank 1977b; Rothkegel 1984b; Lötscher 1987; Fritz 1994; Brinker 1997; Schröder 2003). Die klassische Rhetorik war in erster Linie orientiert auf die „schmückende und verfeinernde Ausformung eines Themas“ in den „5 Verarbeitungsphasen des Themas“ („inventio“, „dispositio“, „elocutio“, „memoria“ und „actio et pronuntiatio“); sie war demnach auf Textganzheiten fokussiert und kann deshalb in gewisser Weise als „Vorläufer pragmatischen und insbesondere auch textlinguistischen Vorgehens“ gelten (vgl. Heinemann und Viehweger 1991, 20f.). Den auf dem rhetorischen Thema-Konzept basierenden Arbeiten ist gemein, dass sie sich mit der thematischen Strukturierung von Texten bzw. Diskursen beschäftigen. Die Perspektive auf diese thematische Strukturierung ist dabei allerdings sehr heterogen: Es lassen sich semantische, kommunikativ-pragmatische und kognitive Textthema-Konzeptionen unterscheiden.

In den semantisch orientierten Arbeiten wird das Textthema als einer der zentralen Aspekte, die zur Kohärenz des Textes beitragen, betrachtet. Das Textthema gilt als semantischer Kern des Textes, von dem aus der Text als Ganzes entfaltet wird. Dieser semantische Kern wird dabei als Teil einer impliziten Textbasis betrachtet. Verankert ist diese Sichtweise in der von den Vertretern der Generativen Grammatik vorgeschlagenen Unterscheidung von Oberflächen- und Tiefenstruktur, wobei das Konzept Thema der Texttieferstruktur zugeordnet wird. Metzelin/Jaksche (1983) sowie van Dijk (1980) fassen das Textthema in diesem Kontext als „Basisproposition“ bzw. „Makroproposition“ des Textes auf, als Proposition, die an der Spitze einer hierarchisch geordneten Zusammenstellung der Propositionen des Textes steht. Dressler (1973) und Agricola (1976; 1979) definieren das Textthema dagegen enger als Sachverhalt aus Prädikaten und Aktan-

ten<sup>13</sup>. Allen Ansätzen gemeinsam ist die Auffassung, dass ein Text auf seinen semantischen Kern, das Textthema, kondensiert werden könne und dass dieser Kern durch ein auf dem Text operierendes Analyseverfahren extrahierbar sei.

Ein gänzlich anderer Zugang zum Textthema bildete sich seit Mitte der 1970er Jahre im Zuge der kommunikativ-pragmatischen Wende und dabei insbesondere im Rahmen der Analyse gesprochener Sprache heraus. Die verschiedenen, methodisch und theoretisch divergenten Forschungsarbeiten teilen die Grundannahme, dass das Thema als sprachlicher Akt des Thematisierens und im Kontext seiner Produktions- und Rezeptionsbedingungen beschrieben werden müsse (siehe z.B. Keenan und Schieffelin 1976; Schank 1977a; Brown und Yule 1983; Givón 1983b; Yule und Mathis 1992; Fritz 1994). Zwar wird teilweise auch hier die sequenzielle Entwicklung des Themas von Äußerung zu Äußerung in den Blick genommen (z.B. bei der Analyse der kollaborativen Themenentwicklung)<sup>14</sup>, im Gegensatz zu den Arbeiten in der Tradition der Funktionalen Satzperspektive oder der Systemisch-Funktionalen Linguistik wird jedoch davon ausgegangen, dass das Konzept Thema auf Satz- bzw. Äußerungsebene bedeutungslos sei.

Eine vierte Hauptrichtung bei der Forschung zum Thema entwickelte sich seit Ende der 1970er Jahre und verstärkt seit Anfang der 1980er Jahre im Rahmen einer kognitiv orientierten Linguistik. Das Thema wird in diesem Forschungszusammenhang häufig definiert sowohl unter Rekurs auf kognitive Schemata bzw. Frames<sup>15</sup> als auch kommunikative Prinzipien, so dass man hier auch nicht von rein kognitiven Thema-Konzepten sprechen kann. Beispiele für solche vielschichtigeren Erklärungsansätze, die sowohl die Kommunikationsbedingungen von Texten als auch deren kognitive Korrelate in den Blick nehmen, sind die Arbeiten von Schank (1977b) und Rothkegel (1984a; 1997). Sie beschreiben Regeln für das Thematisieren als sprachliche Handlung und stellen dabei eine Verbindung von dem in Schemata organisiertem Wissen und möglichen Themen, die ein Autor behandeln kann, her. Insofern hier, wie in den kommunikativ-pragmatischen Ansätzen, das Thematisieren als sprachliche Handlung verstanden wird, können diese Arbeiten auch als Spezialfälle der kommunikativ-pragmatischen Themaforschung betrachtet werden. Für eine automatisierte Analyse des Themas interessant sind allerdings vor allem die kognitiv orientierten Forschungen, die sich mit der Frage beschäftigen, ob das, was thematisiert werden kann, auf einer abstrakten Ebene vorhersagbar ist – z.B. über allgemeine kognitive Schemata bzw. Frames (Schank 1977b; Rothkegel 1984a; 1997) oder

---

<sup>13</sup> Aktanten sind semantisch gesehen Entitäten, die am Prozess der Handlung beteiligt sind. Sie sind von der Valenz des Verbs abhängig. Syntaktisch gesehen sind Aktanten Nominalphrasen (vgl. Bußmann 2002, 58f.). Eine spezifischere Definition von Aktant findet sich bei Agricola (1976; 1979) (siehe Kapitel 2.2.2, insbesondere Fußnote 55).

<sup>14</sup> Grundlegend für die hier angesprochenen, den Dialog oder Text als Sequenz fokussierenden Ansätze ist u.a. die Auffassung vom Thema als „Fraglichem“, das den sequenziellen Aufbau des Diskurses leitet (siehe Keenan und Schieffelin 1976; Hellwig 1984a, b; van Kuppevelt 1995; Büring 1997).

<sup>15</sup> Kognitive Schemata und Frames werden in Kapitel 2.7.2 ausführlich behandelt und differenziert.

über texttypspezifische<sup>16</sup> Schemata bzw. „Superstrukturen“ (van Dijk 1980; 1985).

Mehr oder weniger losgelöst von der linguistischen Forschung ist die computerlinguistische oder sprachtechnologische Forschung zum Thema. Während es in der Linguistik vorwiegend um eine theoretische Ausarbeitung des Themas geht, liegt der Fokus computerlinguistischer und sprachtechnologischer Arbeiten vor allem auf der maschinellen Analyse des Themas in Texten. Es erstaunt deshalb wenig, dass das Thema in diesen Arbeiten selten explizit definiert und stattdessen häufiger intensional bestimmt wird. Das Thema wird dabei oft zweckorientiert als Einheit der Textoberfläche aufgefasst, die sich über Merkmale wie Rekurrenz identifizieren lässt (siehe Kapitel 2.5).

Über die sprachliche Form des Themas gibt es sehr unterschiedliche Auffassungen. Fritz (1994, 193) geht beispielsweise davon aus, dass das Thema selbst nicht in sprachlicher Form im Text vorkommt, sondern „ein Aspekt des Verständnisses“ ist. In den satzbezogenen, informationsstrukturellen Arbeiten wird das Thema vorwiegend als Nominalphrase (NP) bestimmt. Unter allen Nominalphrasen des Textes wird die thematische NP dabei über eine Reihe von Merkmalen herausgefiltert – wobei diese Merkmale je nach Ansatz divergieren. Als zentrale Merkmale erscheinen Satzerstposition, „Bekanntheit“, „Aboutness“ und die semantische Rolle AGENS (siehe Kapitel 2.1 und 2.2.1.1 sowie Halliday 1967; Li und Thompson 1976; Lerner 1977; Givón 1983b; Halliday 1985; Lambrecht 1994). In den textbezogenen Ansätzen, die vor allem die sequenzielle Entwicklung des Themas in aufeinanderfolgenden Sätzen betrachten, wird das Thema häufig ebenfalls als Nominal- oder Präpositionalphrase identifiziert (siehe Kapitel 2.2.1.2 sowie Daneš 1976; Givón 1983a; Hoffmann 1997). Unter allen NPs des Textes werden dabei die rekurrenten NPs bzw. die NPs, die über eine semantische Relation mit der NP des vorausgehenden Satzes verknüpft sind, als thematisch aufgefasst. Daraus resultiert in den Ansätzen zur sequenziellen Themenentwicklung meist eine Liste von Themen (bzw. Satzthemen), nicht aber ein singuläres Textthema. Im Gegensatz dazu definieren die semantischen Ansätze das Thema als Basis- oder Makroproposition bzw. Prädikat-Aktanten-Struktur, d.h. als singuläres Textthema (siehe Kapitel 2.2.2 sowie Dressler 1973; Agricola 1979; Wulff 1979; van Dijk 1980; Metzelin und Jaksche 1983). Diese Basis- oder Makroproposition bzw. Prädikat-Aktanten-Struktur lässt sich nach Ansicht der Autoren mittels ausgearbeiteter Analyseverfahren ermitteln. In der Praxis erweisen sich diese Verfahren jedoch oft als wenig objektiv (zur Kritik an diesen Modellen und Verfahren siehe Kapitel 2.2.2). Eine dritte Repräsentationsform des Themas stellt die Auffassung vom Thema als „Fraglichem“ dar (siehe Kapitel

---

<sup>16</sup> Der Begriff „Texttyp“ wird in dieser Arbeit im Sinne von Textklasse gebraucht, also als Sammelbegriff für eine Menge von Texten, die relevante Eigenschaften miteinander teilen und sich von anderen Mengen von Texten mit anderen relevanten Eigenschaften abgrenzen lassen. Texttypen sind demnach sowohl Textsorten (z.B. Kochrezept), als auch weniger spezifische Textklassen, die sich z.B. durch die Art ihres Vertextungsmusters unterscheiden (z.B. deskriptive, appellative, argumentative, explikative Texte etc). Für einen Überblick über die Begriffe „Texttyp“ und „Textsorte“ und ihre Verwendungsweisen in der linguistischen Forschung siehe Heinemann (2000b).

2.3.2 sowie Keenan und Schieffelin 1976; Hellwig 1984a, b; von Stutterheim 1994; van Kuppevelt 1995; Büring 1997). In diesem Rahmen wird das Thema entweder als propositionale Präsupposition, die aus einer impliziten Frage, die der Autor zu beantworten sucht, hervorgeht, oder als die implizite Frage selbst aufgefasst. Im Gegensatz zu den vorgenannten Repräsentationsformen handelt es sich hier also nicht um eine sprachliche Einheit, die mehr oder weniger direkt mittels spezifischer Analyseschritte aus dem Text extrahierbar ist. Wie vom sprachlichen Material ausgehend die Ermittlung der „Fragen“ methodisch erfolgen kann, ohne in Subjektivität zu münden, bleibt unklar. Eine vierte, alternative Repräsentationsform des Themas findet sich in einigen der kognitiv orientierten Forschungen zum Thema. Hier wird das Thema als kognitives Konzept aufgefasst, d.h. als mentale und nicht sprachliche Einheit (siehe Kapitel 2.4 sowie de Beaugrande und Dressler 1981; Rothkegel 1984a; van Dijk 1985). Im Gegensatz zur Repräsentation des Themas als „Fraglichem“ können bei diesem Ansatz jedoch die im Text vorkommenden, auf kognitive Konzepte verweisenden sprachlichen Ausdrücke (insbesondere Nominalphrasen) genutzt werden, um das kognitive Konzept zu ermitteln, das im Zentrum aller referenzierten Konzepte steht und das als solches als Thema des Textes definiert wird.

### 1.1.3 Verwendungsweisen von „Thema“

Zum Schluss dieses Überblicks über die verschiedenen Forschungstraditionen und Thema-Konzepte soll noch kurz auf die Verwendung des Ausdrucks „Thema“<sup>17</sup> in der Alltagsdomäne eingegangen werden. Das Wort „Thema“ gehört zu den am Häufigsten verwendeten Substantiven des Deutschen. In der korpusbasierten Wortgrundformenliste DEREWO des Instituts für deutsche Sprache in Mannheim, die 40.000 Lemmata umfasst, gehört „Thema“ zur Häufigkeitsklasse 8 und steht in Hinblick auf seine absolute Häufigkeit an 191. Stelle.<sup>18</sup> In dem hier exemplarisch für ein modernes Wörterbuch ausgewählten Duden-Universalwörterbuch (o.J.) werden folgende Verwendungsweisen von „Thema“ genannt:<sup>19</sup>

- ein interessantes, beliebtes, heikles, aktuelles, politisches, literarisches T.
- dieses T. ist tabu, ist erschöpft
- das ist für mich kein T. (das steht nicht zur Diskussion)
- ein T. berühren, [eingehend] behandeln, aufgreifen, anschneiden, fallen lassen
- das T. wechseln
- im Aufsatz das T. verfehlen

---

<sup>17</sup> Der Ausdruck „Thema“ wird in den folgenden Absätzen in Anführungszeichen gesetzt, um ihn vom Konzept Thema zu differenzieren.

<sup>18</sup> DEREWO, Version derewo-v-40000g-2009-12-31-0.1; [<http://www.ids-mannheim.de/kl/derewo/>]; (veröffentlicht am 31.12.2009; zuletzt abgerufen am 09.11.2010).

<sup>19</sup> Das Duden-Universalwörterbuch unterscheidet zwei Lesarten von „Thema“: „Thema“ als „Gegenstand“ und „Thema“ als „Melodie“. Die hier genannten Verwendungsweisen von „Thema“ beziehen sich auf die erste Lesart.



- vom T. abkommen, abschweifen
- das gehört nicht zum T.
- zum eigentlichen T. zurückkommen

Im Online-Wörterbuch zur deutschen Gegenwartssprache am Institut für deutsche Sprache, „elexiko“ (2003ff.)<sup>20</sup>, werden darüber hinaus u.a. folgende weitere Verwendungsweisen von „Thema“ aufgeführt:

- eine Debatte über das Thema [z.B. *Religionsunterricht*]
- weitere Informationen zu dem Thema [z.B. *Allergien*]
- Ausführungen zum Thema [z.B. *Innehalten – Umgang mit Stress*]
- Dissertation zum Thema [z.B. *Videoüberwachung*]
- Fragen zum Thema [z.B. *Krebs*]
- ein Vortrag zum Thema [z.B. *Das dritte Ohmsche Gesetz*]
- das Thema [z.B. *anonyme Anzeige*] behandeln
- das Thema ausführlich diskutieren
- sich dem Thema [z.B. *Gesundheit*] widmen
- sich mit bestimmten Themen auseinandersetzen
- sich intensiv mit dem Thema [z.B. *Liebe und Leid*] beschäftigen
- über ein Thema sprechen
- vom Thema abgelenkt werden
- sich zu diesem Thema äußern
- zu aktuellen Themen Stellung beziehen
- das Thema [z.B. *Kongresswahlen in den USA*] ist abgehakt
- das Thema ist brisant und aktuell
- das Thema [z.B. *Gewalt*] steht im Mittelpunkt

Der Ausdruck „Thema“ wird in diesen Verwendungsweisen häufig gebraucht als Gegenstand, auf den eine metasprachliche Handlung vom Typ „Kommunizieren“ bezogen ist. Als ein solcher Gegenstand kann er mit Adjektiven („interessantes, beliebtes, heikles, aktuelles, politisches, literarisches“ etc.) oder Adverbien näher bestimmt werden, aber auch als Teil einer Verbalphrase („das Thema behandeln“, „über ein Thema sprechen“) auftreten. Die metasprachliche Handlung, die sich auf den „Gegenstand“ (das Thema) bezieht, kann sowohl als allgemeines Kommunikationsverb („diskutieren“, „äußern“, „sprechen“, „Stellung beziehen“) als auch als Nominalisierung eines solchen Verbs („Debatte“, „Ausführung“, „Frage“, „Vortrag“) realisiert sein. Neben diesen allgemeinen Kommunikationsverben findet sich im Zusammenhang mit „Thema“ eine Gruppe von Verben, die gesprächs- oder themenstrukturierend wirken („anschnitten“, „fallen lassen“,

---

<sup>20</sup> In „eLexiko“ werden drei Lesarten von „Thema“ unterschieden: „Thema“ als „Grundgedanke“, „Grundmelodie“ und „semantische Einheit“. Die Lesart vom „Thema“ als „Grundgedanken“ wird dabei als Hauptlesart beurteilt, die beiden anderen Lesarten werden dagegen als Differenzierungen der ersten Lesart aufgefasst. Die hier genannten Verwendungsweisen von „Thema“ beziehen sich auf die erste Lesart.

„wechseln“, „abkommen vom“, „abschweifen“, „zurückkommen auf“, „ablenken“, „abhaken“). Mit diesen Verben wird „auf Situationen Bezug genommen (...), in denen ein Sprecher durch seine Äußerung die Gestaltung des Gesprächs oder die Behandlung des Gesprächsthemas beeinflusst“ (Harras et al. 2004, 386). In diesem Zusammenhang sind auch die satzwertigen Phrasen „das gehört nicht zum Thema“ oder „im Aufsatz das Thema verfehlen“ interessant – an ihnen wird deutlich, dass Themen ein Gespräch oder einen Text inhaltlich restringieren und dass die Einhaltung der Themaorientierung eingeklagt werden kann bzw. die Nicht-Einhaltung zu negativen Konsequenzen führen kann.

Die genannten Verwendungsweisen von „Thema“ zeigen deutlich zwei Dinge: Erstens wird das Wort „Thema“ metakommunikativ gebraucht zur Strukturierung von Gesprächen: ein Thema wird 1. „eingeführt“ oder „vorgegeben“ (z.B. bei einem Aufsatz), 2. „behandelt“, 3. „abgehakt“ oder „fallen gelassen“ – mit diesen Verbalphrasen lassen sich gleichzeitig auch die Phasen der kommunikativen Behandlung von Themen charakterisieren: Themaeinführung oder Themavorgabe, Themabehandlung, Themabeendigung. Diese Schritte können unterbrochen sein durch einen Themawechsel (das Thema „wechseln“) oder die Behandlung von Nebenthemen (ein Thema „berühren“, vom Thema „abschweifen“), von diesem Themawechsel kann anschließend wieder „zurückgekehrt“ werden zum Hauptthema. Über den Akt des Thematisierens können Diskurse demnach organisiert und strukturiert werden. Zum Zweiten kann durch die Verwendung von „Thema“ in Verbindung mit einem postnominalen Erweiterungsnomen (Bsp.: „das Thema *Babypause*“) ein konkreter Gegenstand metakommunikativ als Thema „gesetzt“ werden, so dass dieser zum konkreten Thema erhobene Gegenstand als inhaltliches Zentrum des Gesprächs definiert wird. Offensichtlich ist dabei das, was inhaltlich zu diesem Gegenstand gesagt werden kann, kognitiv vorstrukturiert, ansonsten wäre es überflüssig, vom „Verfehlen“ oder „Abschweifen vom“ Thema zu sprechen. Das gesetzte Thema wirkt demnach beschränkend<sup>21</sup> auf die inhaltliche Struktur des Diskurses ein.

Diese beiden Erkenntnisse, die sich aus der alltäglichen Verwendungsweise von Thema ableiten lassen, müssen auch in der linguistischen Diskussion des Thema-Konzepts Berücksichtigung finden und in die Definition und Analyse des Themas einfließen.

#### **1.1.4 Disziplinen und Anwendungsbereiche, für die das Konzept Thema relevant ist**

Das Konzept Thema spielt in mehreren Disziplinen und Anwendungsbereichen eine zentrale Rolle (siehe Abbildung 1). Ohne Frage ist die Disziplin, die sich am Ausführlichsten mit dem Konzept Thema beschäftigt, die Linguistik. Während in den linguistischen Teildisziplinen Textlinguistik, Semantik, Grammatik, Pragmatik und kognitive Linguistik vor allem die theoretische Auseinandersetzung mit dem Thema im Zentrum steht, liegt der Fokus der Computerlinguistik in

---

<sup>21</sup> Im weiteren Verlauf der Arbeit wird auf das in der Computerlinguistik gebräuchliche Konzept der „Constraints“ mit der deutschen Übersetzung „Beschränkung“ Bezug genommen.

erster Linie auf der Anwendung.<sup>22</sup> Sowohl für die automatische Informations- oder Themenextraktion, die Textklassifikation/Textkategorisierung als auch das automatische Textzusammenfassen ist die Identifizierung („Topic Identification“), Extraktion („Topic Extraction“) und Benennung (teilweise auch die Typisierung) von Themen von besonderem Interesse. Hierzu werden sowohl statistische und mathematische Methoden eingesetzt als auch regelbasierte Verfahren, die auf Oberflächenmerkmalen und lexikalischen Merkmalen des Textes operieren (z.B. Position und Salienz von Wörtern). Daneben werden auch wissensbasierte Ansätze zur Identifizierung und vor allem Typisierung und Benennung von Themen genutzt. Dabei werden häufig Wörter mit Hilfe von lexikalisch-semantischen Wissensquellen auf Konzepte abgebildet, um so zu generelleren Themenangaben zu gelangen. Für die Aufgabe der Textsegmentierung spielt die Benennung von Themen dagegen keine Rolle, hier geht es vor allem um die Bestimmung von Themengrenzen bzw. Themenwechseln („Topic Chaining“, „Topic Tracking“). Dabei wird vor allem auf statistische Methoden gesetzt, über die die semantische Ähnlichkeit von Abschnitten des Textes berechnet wird. Problematisch an nahezu allen computerlinguistischen Arbeiten ist die fehlende Verknüpfung dieser Anwendungen und Methoden mit einer theoretischen Auseinandersetzung mit dem Konzept Thema. Aus dieser fehlenden theoretischen Fundierung resultiert in den meisten Ansätzen ein vereinfachter lexikalischer Thema-Begriff, der dazu führt, dass beispielsweise die Anwendungen „Themen-“ und „Keyword-Extraktion“ häufig keine großen Unterschiede aufweisen und dem besonderen Status des Themas, der ihn von einfachen Schlüsselwörtern eines Textes unterscheidet, nicht Rechnung getragen wird.

---

<sup>22</sup> Auch wenn die Computerlinguistik grundsätzlich eine Teildisziplin der Linguistik ist, sind insbesondere die methodischen, aber auch theoretischen Unterschiede so groß, dass die Computerlinguistik im Folgenden eher als Nachbardisziplin und nicht als Teildisziplin der Linguistik betrachtet wird.

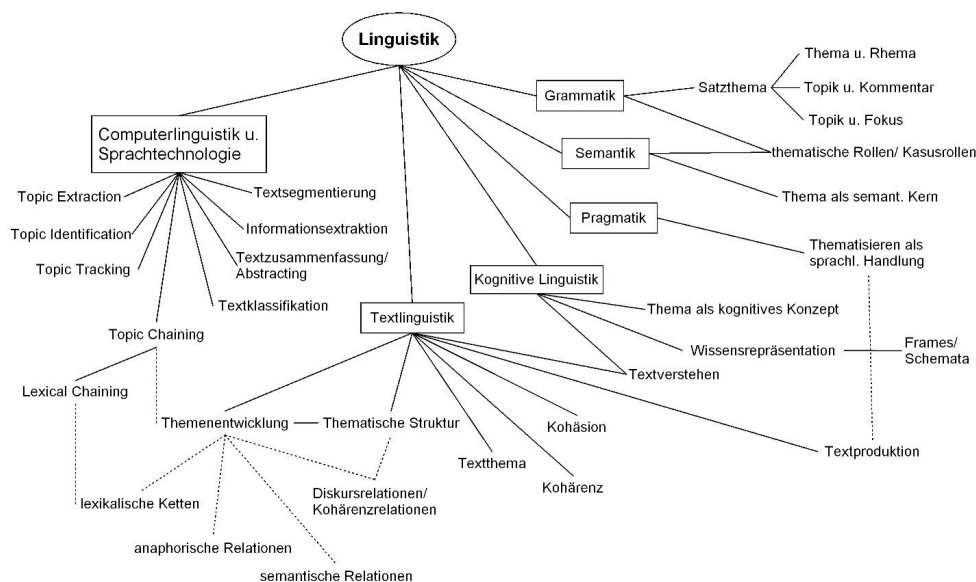


Abbildung 1: Überblick über die Disziplinen, Gegenstandsbereiche und Anwendungen, die sich mit dem Thema beschäftigen

Innerhalb der Linguistik lässt sich das Thema in einer Vielzahl von Teildisziplinen verorten: Sowohl in der Grammatik, der Semantik, Pragmatik, der kognitiven Linguistik und natürlich vor allem der Textlinguistik spielt das Thema eine Rolle. Während in der Grammatik vor allem Satzthemen, ihre Ausdrucksform und kommunikative Funktion betrachtet werden, aber auch (an der Schnittstelle zur Semantik) thematische Rollen<sup>23</sup>, findet in den anderen Teildisziplinen vor allem eine Auseinandersetzung mit dem Textthema statt. In jeder Teildisziplin wird das Konzept Thema dabei anders definiert (siehe Kapitel 2). Die im Hinblick auf die Definition des Textthemas ergiebigste Teildisziplin der Linguistik ist dabei die Textlinguistik. Neben der eigentlichen Definition des Textthemas wird hier auch die Funktion des Themas für den Text beleuchtet – in diesem Zusammenhang wird vor allem nach dem Status der thematischen Struktur sowie nach der Rolle des Textthemas für die Kohärenz des Textes<sup>24</sup> gefragt.

<sup>23</sup> Alternativ: Kasusrollen, Theta-Rollen, semantische Rollen – siehe hierzu Kap. 2.1.3.

<sup>24</sup> Die Verfasserin geht davon aus, dass Kohärenz von Kommunikanten auf Basis ihres individuellen Wissens und Könnens konstruiert wird und dass sich diese Kohärenz interindividuell unterscheiden kann. Dennoch kann vereinfachend davon gesprochen werden, dass auch Texte selbst kohärent (oder nicht kohärent) sind – gemeint ist damit, dass Texte die Konstruktion von Kohärenz und damit den Aufbau eines sinnvollen mentalen Modells des Textes in unterschiedlicher Weise unterstützen: Texte, die die Kohärenzbildung durch ihren globalen Aufbau, den Einsatz

## 1.2 Forschungsfragen und Zielsetzung

Aus dem bisher Gesagten lässt sich zusammenfassend festhalten:

1. Der Terminus „Thema“ ist in der Linguistik mit sehr heterogenen Vorstellungen besetzt, die sich beträchtlich voneinander unterscheiden und zu heterogenen Thema-Definitionen führen, in denen „Thema“ teilweise zur Bezugnahme auf Satz-, teilweise auch textbezogene sprachliche Phänomene gebraucht wird.
2. Der Ausdruck „Thema“ wird im Alltag im Kontext metakommunikativer Äußerungen gebraucht. In Verbindung mit gesprächs- und themenstrukturierenden Verben spielt er eine Rolle bei der auf die inhaltliche Ebene bezogenen metakommunikativen Strukturierung von Diskursen. Als Bezugspunkt für ein postnominales Erweiterungsnomen dagegen definiert er einen konkreten Gegenstand als Thema und stellt diesen Gegenstand so in den inhaltlichen Mittelpunkt des Diskurses.
3. Das Thema spielt in einer Vielzahl von Forschungsbereichen (Grammatik, Semantik, Pragmatik, kognitive Linguistik, Textlinguistik, Computerlinguistik) und Anwendungen (automatische Textsegmentierung, Textzusammenfassung, Textklassifikation, Informationsextraktion) eine wichtige Rolle.
4. Zwischen der Linguistik und Computerlinguistik findet bezüglich der Definition und des Status des Themas sowie der Analysemethodik kaum ein Austausch statt. Bis heute ist es weder in der Linguistik noch Computerlinguistik gelungen, ein zufriedenstellendes operationalisierbares Analyseverfahren zu entwickeln, das gleichzeitig auf einem linguistisch fundierten Thema-Begriff basiert.

Zunächst stellt sich die Frage, warum es eigentlich so vielfältige und heterogene Thema-Konzepte gibt. Eine mögliche Erklärung könnte die generelle Diversität der linguistischen Theorieansätze sein, die auch bei der Definition anderer linguistischer Konzepte (z.B. dem „Textsorten“-Begriff) für einen Mangel an Konsens verantwortlich ist. Eine alternative Erklärung könnte jedoch auch im Thema-Konzept selbst begründet sein. In dieser Arbeit wird deshalb danach gefragt, welche Rollen das Thema auf den verschiedenen Beschreibungsebenen des Textes spielt, und ob in den unterschiedlichen Thema-Konzepten jeweils eine bestimmte Ebene des Textes in den Blick genommen wird. Dabei soll insbesondere der Frage nachgegangen werden, ob sich die Konzepte zwar auf unterschiedliche sprachliche (und außersprachliche) Ebenen beziehen, letztlich aber durch *ein* umfassendes Thema-Konzept, dass die verschiedenen Perspektiven auf das Thema in einem Ebenen-basierten Thema-Modell integriert, beschrieben werden könnten.

In diesem Kontext wird auch nach dem Zusammenhang von Satz- und Textthema gefragt. Handelt es sich um einen rein terminologischen Zusammenhang

---

sprachlicher Mittel oder die Aktivierung von Wissen gut unterstützen, werden als kohärent bezeichnet; Texte, die dies nicht tun, als weniger kohärent.

### **Forschungsfragen**

1. Warum existieren so viele heterogene Thema-Konzepte? Lassen sich die existierenden Thema-Konzepte zu einem integrativen Ebenen-basierten Thema-Konzept kombinieren?
2. Wie interagieren Thema, Text und Wissen? Ist das Thema im Text irgendwie enthalten bzw. aus dem Textmaterial heraus ermittelbar? Ist das Thema als Kern einer Wissensstruktur aufzufassen?
3. Wie können die Ebenen des Themas für die maschinelle Identifikation des konkreten Themas eines Textes genutzt werden und wie könnte ein System zur maschinellen Repräsentation und Identifikation der Ebenen

### **Ziele**

1. Problematisierung des Status des Konzepts Thema
2. Entwicklung eines Ebenen-basierten integrativen Thema-Konzepts und Beschreibung der Interaktionen zwischen diesen Ebenen
3. Darstellung der computerlinguistischen Methoden und Ansätze, die zur Repräsentation und Identifikation der Ebenen des Themas im Text genutzt werden können, und Skizzierung eines Software-Systems zur maschinellen Analyse und Identifikation des Themas

(„zufällig“ wird der gleiche Ausdruck für zwei verschiedene Konzepte verwendet) oder gibt es auch einen konzeptuellen Zusammenhang zwischen Satz- und Textthema (wie z.B. van Kuppevelt (1995) und Jäger/Oshima (2002) annehmen)? Sind auch Satz- und Textthema nur Bestandteile eines globaleren Thema-Konzepts?

Geht man von der Hypothese aus, dass sich das Thema auf mehrere sprachliche und außersprachliche Ebenen bezieht und über die für die unterschiedlichen sprachlichen Ebenen spezifischen Mittel im Text „reflektiert“ wird, stellt sich die Frage, ob der Terminus Thema selbst dann überhaupt noch auf eine notwendige linguistische Kategorie verweist. Schlobinski/Schütze-Coburn (1992) stellen in diesem Zusammenhang die These auf, dass das Thema in gewisser Weise eine überflüssige Kategorie sei. Durch die Assoziierung von sprachlichen Phänomenen im Bereich Grammatik, Semantik und Pragmatik mit einem Thema-Konzept bestehe die Gefahr, dass die zugrunde liegenden syntaktischen, semantischen oder pragmatischen Kategorien verschleiert würden:

The use of ‘topic’ is in some sense unnecessary. Despite its intuitive appeal, ‘topic’ as a cover term leads to unhelpful representations and obfuscates the underlying syntactic, semantic, or pragmatic categories on which it is based. (Schlobinski und Schütze-Coburn 1992, 89).

In der vorliegenden Arbeit soll – unabhängig von der Frage nach der Bedeutung des Thema-Konzepts für die Linguistik – ein integratives, Ebenen-basiertes Modell des Themas entwickelt werden, das zeigen soll, dass die unterschiedlichen linguistischen Thema-Konzepte letztlich gar nicht so heterogen und unvereinbar sind, wie es zunächst scheint, sondern dass sie nur verschiedene Ebenen des Textes (zu denen auch die kommunikative Ebene der Textproduktion und -rezeption sowie die kognitive Ebene zu rechnen sind) in den Blick nehmen. Bei der Entwicklung des integrativen Thema-Modells ist insbesondere danach zu fragen, welche Rolle das Thema auf den verschiedenen Beschreibungsebenen des Textes spielt. In diesem Kontext muss auch geklärt werden, wie Thema, Text und Wissen im Prozess der Textproduktion und -rezeption interagieren und in welchem Verhältnis Thema und Text („Ist das Thema im Text irgendwie enthalten bzw. aus dem Text heraus ermittelbar?“), Thema und Wissen („Ist das Thema als Kern einer Wissensstruktur aufzufassen?“) sowie Text und Wissen stehen.

Eine weitere zentrale Frage, die in der vorliegenden Arbeit beantwortet werden soll, ist schließlich, inwiefern ein Ebenen-basiertes Thema-Modell als Grundlage für die maschinelle Identifikation des konkreten Themas eines Textes genutzt werden kann, d.h. ob sich ein komplexes, linguistisch fundiertes Thema-Modell als Ausgangspunkt für die Entwicklung eines computerlinguistischen Software-Systems zur Identifikation des konkreten Themas eines Textes eignet. Ein solches, von einem linguistischen Thema-Konzept ausgehendes, theoretisch fundiertes System zur maschinellen Thema-Identifikation wäre für die Computerlinguistik ein echtes Novum, da bisherige Verfahren und Systeme zur Themenidentifikation den Thema-Begriff weitgehend „naiv“ verwenden und sich in der Regel nicht auf linguistische Thema-Konzepte beziehen oder diese gar als Grundlage für die maschinelle Textverarbeitung bzw. Themenidentifikation nutzen. Aus den ermittelten Ebenen des Themas sollen in einem weiteren Schritt schließlich Cues<sup>25</sup> und Constraints für die maschinelle Identifikation des Themas abgeleitet werden. Während Cues die oberflächenbasierten Reflexionen des Themas im Text darstellen, stellen Constraints „Beschränkungen“ dar, die sich aus den Ebenen der Kognition und Kommunikation ableiten lassen, beispielsweise texttypspezifische Beschränkungen oder schemabasierte inhaltliche Beschränkungen (z.B. im Hinblick auf mögliche Aspekte von Themen). Constraints bzw. Wissensbestände spielen dabei eine ebenso wichtige Rolle für die Identifikation des konkreten Themas eines Textes wie die Cues. Sowohl Cues als auch Constraints können mit Hilfe texttechnologischer Verfahren bzw. Auszeichnungssprachen repräsentiert werden. Während die oberflächenbasierten Reflexionen des Themas im Text als XML<sup>26</sup>-basierte Multi-Ebenen-Annotationen repräsentiert werden können, sollen für die wissensbasierten Beschränkungen formale Modellierungen gewählt werden, die sich an der Theorie kognitiver Frames orientieren. Abschließend soll die Architektur eines Software-Systems zur maschinellen Analyse und Identifikation des konkreten Themas eines Textes

---

<sup>25</sup> Der Begriff „Cues“ wird in der vorliegenden Arbeit synonym zu „Reflexionen im Text“ und synonym zu „indexikalischen Zeichen“ (im Sinne von Antos 2009, 408, siehe auch Kapitel 3.1.2.2) verwendet.

<sup>26</sup> XML: Extensible Markup Language (W3C – World Wide Web Consortium 2006).

skizziert werden, das die in dieser Arbeit entwickelten theoretischen Annahmen und Festlegungen berücksichtigt und die genannten Cues und Constraints zur Identifikation des Themas nutzt. Ziel eines solchen Systems soll es sein, die formal repräsentierten Ebenen des Themas und deren Interaktionen so auszuwerten, dass hieraus Aussagen über die im Text behandelten Aspekte des Themas und deren Zusammenhänge gewonnen werden können.

Die vorliegende Arbeit beschränkt sich auf konzeptionell und medial schriftliche Texte (vgl. Koch und Oesterreicher 1994), und speziell auf expositorische Texte. Bei diesen Texten handelt es sich um primär darstellungs- bzw. sachverhaltsbezogene, informative Texte, deren Funktion in erster Linie referenzieller Natur ist, während die konative (bei Bühler: *appellative*), emotive (bei Bühler: *expressive*), poetische, phatische und metasprachliche Funktion nur eine untergeordnete Rolle spielen (siehe auch Jakobson 1971, 147-152; Bühler 1978 [1965]). Der Grund für diese Einschränkung ist der, dass das Thema nicht in allen Texten gleichermaßen zentral ist – in sachverhaltsbezogenen Texten spielt das Thema eine wesentlich wichtigere Rolle für die Textproduktion und -rezeption als in Texten, die eine primär poetische oder emotive Funktion haben. Sachorientierte informative Texte sind auf ein übergeordnetes Thema (oder einen Themenkomplex) ausgerichtet und im Hinblick auf dieses Thema strukturiert, während z.B. emotiv orientierte Texte inhaltlich und thematisch wesentlich freier sind und die Orientierung am Thema nicht das zentrale Strukturierungsprinzip darstellen muss. Die Kohärenz ist in diesen Texten weniger thematisch als vielmehr emotiv und situativ gebunden. Trotz der Einschränkung dieser Arbeit auf Texte, deren Funktion die Verständigung über einen bestimmten Inhalt ist (referenzielle Funktion) – sowohl durch Informieren, Beschreiben, Erklären als auch Erzählen, insofern das „Was“ und nicht das „Wie“ im Zentrum steht – lassen sich die Ergebnisse dieser Arbeit auf andere Texttypen übertragen, sofern die referenzielle Funktion in diesen Texten nicht völlig irrelevant ist. Solange Texte (auch) inhaltlich orientiert sind (im Sinne der Funktion der Darstellung bei Bühler), liegt gleichzeitig eine thematische Textstrukturierung vor – diese kann allerdings mehr oder weniger deutlich ausgeprägt sein. Um die Relevanz des Thema-Begriffs für die Strukturierung von Texten deutlich zeigen zu können, gehe ich in dieser Arbeit von primär thematisch strukturierten Texten aus – sowohl das entwickelte Thema-Konzept als auch die dargestellte Analyseverfahren können jedoch auch auf Texte angewandt werden, bei denen die thematische Strukturierung gegenüber anderen Strukturierungen im Hintergrund steht. In diesem Fall sind jedoch Interaktionen mit und Überlagerungen durch andere Aspekte wie das Verhältnis der Kommunikanten, der ästhetischen Intention etc. zu berücksichtigen.



### 1.3 Überblick über den Aufbau der Arbeit

In Anbetracht der deutlichen Unterschiede zwischen den existierenden Thema-Konzepten stellt sich die berechnigte Frage, ob es auch Gemeinsamkeiten zwischen diesen Konzepten gibt und welche dies sind. In Kapitel 2 werden deshalb zunächst die zentralen Thema-Konzepte erörtert und – als Teil der Antwort auf die Frage nach den Gemeinsamkeiten dieser Konzepte – anhand der linguistischen Beschreibungsebenen Grammatik, Semantik, Pragmatik und Kognition kategorisiert.

In Kapitel 0 werden die theoretischen Grundlagen zu den Bereichen Text, Kommunikation und Kognition herausgearbeitet. Dieses Kapitel hat die Funktion, eine theoretische Basis für das in Kapitel 3 entwickelte Ebenen-basierte Thema-Modell zu schaffen.

In Kapitel 3 soll auf Basis der in Kapitel 3 entwickelten theoretischen Vorstellungen, ein Ebenen-basiertes Thema-Modell entwickelt werden. Dieses Modell ist darauf ausgerichtet, dass sämtliche Ebenen des Themas formal repräsentiert und mittels computerlinguistischer Methoden maschinell ausgewertet werden können. In diesem Kapitel sollen Theorie und Praxis miteinander verzahnt werden, indem sowohl dargestellt wird, welche Cues und Constraints bzw. Wissensbestände aus theoretischer Sicht für die maschinelle Ermittlung des konkreten Themas eines Textes und seiner Aspekte nötig sind, als auch, wie diese Cues mittels aktueller computerlinguistischer Verfahren maschinell ermittelt bzw. Constraints und Wissensbestände mittels computerlinguistischer Formalismen repräsentiert und so einer maschinellen Verarbeitung zugänglich gemacht werden können.

Das in Kapitel 4 entwickelte Ebenen-basierte Thema-Modell (für einen Überblick siehe Abbildung 7, S. 128) basiert zum einen auf der Analyse und Kategorisierung der linguistischen und computerlinguistischen Thema-Konzepte (Kapitel 2), zum anderen auf dem in Kapitel 3 entwickelten Beschreibungsmodell der Textstruktur (für einen Überblick siehe Abbildung 5, S. 80). Das Ebenen-basierte Thema-Modell wird auf Basis des in Kapitel 3 entworfenen Modells des Zusammenspiels von Textproduktion/-rezeption, Wissen und Text (Kapitel 3.4, siehe insbesondere Abbildung 6, S. 120) spezifiziert durch ein in Kapitel 3.4.3 zusammenfassend dargestelltes Prozessmodell zum Zusammenhang der Ebenen des Themas, das als Vorbild für die Architektur eines Software-Systems zur maschinellen Themenidentifikation dienen kann. Um die theoretische Validität des entwickelten Modells zu demonstrieren, wird in Kapitel 3.1.2 darüber hinaus demonstriert, wie die in Kapitel 2 dargestellten linguistischen und computerlinguistischen Thema-Konzepte in die Ebenen des Thema-Modells integrierbar sind (für einen Überblick siehe Tabelle 5, S.129f.).

## 2 Thema-Konzepte: Bestandsaufnahme und Kategorisierung

Im Folgenden sollen die unterschiedlichen Forschungsrichtungen zum Konzept Thema dargestellt und im Hinblick auf die sprachlichen Beschreibungsebenen „Grammatik“, „Semantik“, „Pragmatik“ und „Kognition“ kategorisiert werden (siehe Tabelle 1).<sup>27</sup> Da es zum einen bereits mehrere umfassende Überblicke zu Forschungen zum Konzept Thema gibt (siehe Lötscher 1987; Schlobinski und Schütze-Coburn 1992; Hoffmann 2000; Gómez-González 2001) und zum anderen die Anzahl der hier skizzierten Thema-Ansätze bzw. Forschungsarbeiten recht groß ist, wird nicht im Detail auf die einzelnen Arbeiten eingegangen, sondern lediglich in Form von Kurzcharakterisierungen. Diese dienen als Basis für das in Kapitel 4 vorgestellte Ebenen-basierte Thema-Modell, das die existierenden, teils disparat erscheinenden Thema-Konzepte in ein integratives Thema-Modell einordnen wird.

Bei diesem Kategorisierungsansatz wird nicht grundsätzlich zwischen satz- und textbezogenen Themenkonzeptionen unterschieden, sondern Satz- und Textthema werden innerhalb der linguistischen Kategorisierungsebenen gemeinsam abgehandelt. Dadurch soll die Möglichkeit geschaffen werden, Gemeinsamkeiten der satz- und textbezogenen Konzeptionen deutlicher hervorzuheben, als dies bei einer separierenden Darstellung möglich wäre. Neben dem gewählten Kategorisierungsansatz sind natürlich andere Kategorisierungsmöglichkeiten denkbar. Der gewählte Ansatz hebt sich allerdings in grundsätzlicher Weise von den üblichen Kategorisierungen des Konzepts Thema ab. Typischerweise werden in Überblicksarbeiten zum Thema Kategorisierungen wie „Thema als Informationskern“, „Thema als Makroproposition“, „Thema als Fragestellung“ gewählt (siehe Lötscher 1987; Adamzik 2004b) oder „topic as given/known information“, „topic as point of departure“, „topic and discourse topic“ (siehe Schlobinski und Schütze-Coburn 1992). Mit derartigen, sehr spezifischen Einteilungen, die meist genau ein Thema-Konzept repräsentieren, lässt sich natürlich Kritik an der Zusammenfassung verschiedenartiger Thema-Konzepte unter einem Oberbegriff bzw. einer Einordnungskategorie vermeiden. Allerdings ist man dadurch einer Durchdringung der Gemeinsamkeiten und Unterschiede von Gruppen von Thema-Konzepten auch keinen Schritt näher gekommen. Genau dies soll die Darstellung und Einordnung der Thema-Konzepte in Kapitel 2 jedoch leisten: Sie soll die Grundlage für eine Klassifizierung der Ebenen des Themas und damit einer umfassenden, integrativen Sichtweise auf das Konzept Thema bilden.

---

<sup>27</sup> Eine teilweise Überschneidung der Kategorien bzw. die prinzipiell mögliche Mehrfach-Zuordnung einiger Ansätze lässt sich bei dem gewählten Kategorisierungsmodell allerdings nicht umgehen, weil viele der grammatischen oder semantischen Konzeptionen zugleich pragmatisch oder funktional orientiert sind. Eine Einordnung eines Ansatzes als grammatisch oder semantisch bedeutet deshalb nicht automatisch, dass dieser Ansatz überhaupt nicht kommunikativ-pragmatisch ausgerichtet ist.

<b>Ebene</b>	<b>Thema-Konzept</b>	<b>Kapitel</b>
Grammatik	Thema als Satzkonstituente: Thema und Satzposition	2.1.1
	Thema als Satzkonstituente: Thema und grammatische Rolle	2.1.2
	Thema als Satzkonstituente: Thema und semantische/thematische Rolle	2.1.3
Semantik	Thema als Mitteilungsgegenstand	2.2.1.1
	Thema als fortlaufend konstituierter zentraler Gegenstand im Text	2.2.1.2
	Thema als abstrakter Kern des Textinhalts	2.2.2
Pragmatik	Thema als das kommunikativ am wenigsten dynamische Element des Satzes	2.3.1
	Thema als Fragliches	2.3.2
	Thema als Mittel zur Strukturierung von Diskursen: Thematische Textmuster	2.3.3
Kognition	Thema als bekannte Information bzw. aktivierte Diskursentität Thema als Kern eines konzeptuellen Netzes	2.4

Tabelle 1: Übersicht über Thema-Konzepte

In Ergänzung zu den dargestellten linguistischen Thema-Konzepten wird in Kapitel 2.5 auch auf die in computerlinguistischen Arbeiten verwendeten Thema-Konzepte eingegangen. Im Gegensatz zu den linguistischen Arbeiten sind diese allerdings häufig nur implizit über die Darstellung der sprachlichen Merkmale, die zur automatischen Analyse des Themas verwendet werden, erschließbar und nicht über explizite Definitionen oder Charakterisierungen.

## 2.1 Thema als grammatisches Konzept

Das Thema wird in einer Vielzahl von Forschungsarbeiten als Konzept charakterisiert, das ein (oder mehreren) Konstituenten des Satzes zugeordnet werden kann. Wie bereits in der Einleitung dieser Arbeit hervorgehoben wurde, finden sich solche satzbezogenen Thema-Konzepte traditionell vor allem in funktionalen Grammatiktheorien<sup>28</sup> wie der Prager Schule zur „Funktionalen Satzperspektive“ (siehe Daneš 1970; Beneš und Vachek 1971; Mathesius 1971 [1929]; Firbas 1992, 1995; Eroms 2000) und der „Systemic Functional Linguistics“ (Systemisch-Funktionalen Linguistik) (siehe Halliday 1967, 1985; Berry 1995; Martin 1995).<sup>29</sup>

<sup>28</sup> Ausgangspunkt dieser Arbeiten ist die Überlegung, dass die Satzstruktur auch durch pragmatische Aspekte beeinflusst wird, da ein Textproduzent immer die Wahl zwischen verschiedenen, grammatisch gleich korrekten Realisierungsvarianten hat und sich bewusst für eine spezifische Variante entscheidet. Die Wahl einer solchen Variante hängt dabei ab von der kommunikativen Intention des Textproduzenten. Je nachdem, welche Teile des Satzes besonders hervorgehoben werden sollen, entscheidet sich ein Textproduzent für eine spezifische Wortstellung.

<sup>29</sup> Allerdings nehmen auch Grammatiker anderer Schulen auf das Konzept Thema Bezug. Chomsky vertritt beispielsweise die Auffassung, dass das Thema (als Teil des Begriffspaares Topic-Comment) eine Rolle spielt für die Beschreibung der grammatischen Oberflächenstruktur. Er definiert das Thema (bzw. „Topic“) als „leftmost NP immediately dominated by S in the surface structure and that is, furthermore, a major category“ (Chomsky 1965, 221). Topic-Comment

Das Thema wird in funktionalen Grammatikmodellen häufig als Teil des bipolaren Begriffspaares „Thema-Rhema“<sup>30</sup> betrachtet und zur Differenzierung unterschiedlich gewichteter Satzteile verwendet, dementsprechend wird es als funktional determinierte grammatische Einheit aufgefasst (vgl. Reis 1993; Lambrecht 1994, 2ff.; Welke 2002, 125). Im Gegensatz zur Bestimmung anderer grammatischer Einheiten (z.B. grammatisches Subjekt) oder grammatisch-semantischer Einheiten (z.B. AGENS) ist die Zuweisung der Kategorie Thema zu einem bestimmten Element des Satzes jedoch deutlich schwieriger. Diese Schwierigkeiten hängen zusammen mit den vagen Definitionen des Themas, z.B. Thema als „Ausgangspunkt der Aussage“ (Dürr und Schlobinski 2006, 134), als “point of departure of the message” oder “starting-point for the message” (Halliday 1985, 39) sowie als „das, worüber etwas mitgeteilt wird“ (Daneš 1970, 591) oder als “what the clause is going to be about” (Halliday 1985, 39). Diese Definitionen können als Aussagen über das Thema als syntaktisches Konzept aufgefasst werden (syntaktische Konkretisierung: Thema als Element an Satzerstposition, Thema als nominales Element, Thema als grammatisches Subjekt), aber auch als Aussagen über das Thema als semantisches Konzept: Thema als Satzgegenstand, als Referenzakt im Sinne von Searle (1996 [1969], 148) (im Gegensatz zum Prädikationsakt). Im Folgenden soll zunächst die grammatische Perspektive auf das Satzthema in den Fokus genommen werden; in Kapitel 2.2.1 soll es um die semantische Interpretation gehen. Vorweg muss festgehalten werden, dass das Thema im Folgenden nicht definiert wird als Element in Satzerstposition, grammatisches Subjekt oder AGENS, sondern dass es über die Merkmale „+satzerstposition“, „+subjekt“, „+agens“ näher bestimmt werden kann, da es häufig mit diesen korreliert.<sup>31</sup>

### 2.1.1 Thema und Satzposition

Die Relation „Thema-Rhema“ (oder “Topic-Comment”) manifestiert sich nach Ansicht funktionaler Grammatiker auf der sprachlichen Oberfläche in Form einer spezifischen Abfolge im Satz. In den meisten funktionalen Grammatiktheorien oder Arbeiten zur Informationsstruktur des Satzes wird die Abfolge „Thema-vor-Rhema“ als unmarkierte Abfolge<sup>32</sup> bestimmt (Halliday 1967; Li und

---

werden als Oberflächenrealisationen der auf der Tiefenstruktur zu verortenden Subjekt-Prädikat-Relation aufgefasst. Chomsky führt nicht weiter aus, welchen Status er dem Topic des Satzes zuweist, so dass im Folgenden nicht weiter auf diesen Ansatz eingegangen wird.

<sup>30</sup> Statt Thema und Rhema (Halliday 1967; Daneš 1970, 1976; Halliday 1985; Fries 1995), verwenden Chomsky (1965), Gundel (1977; 2004), Chafe (1976), Reinhart (1982), Lambrecht (1994), Sgall (1973), Reis (1977) und andere ähnliche Begriffspaare wie Topic-Comment oder Topik-Fokus. Auf die Unterschiede der Begriffspaare Topic-Comment, Thema-Rhema und Topik-Fokus wird in Kapitel 2.6.3.1 ausführlicher eingegangen.

<sup>31</sup> Neben den genannten Merkmalen bieten sich aber auch zahlreiche andere grammatische Formen zur Realisierung des Themas an, z.B. Rechtsversetzung, Definitheit, Passivkonstruktionen etc. (siehe z.B. Li und Thompson 1976; Gundel 1977; Givón 1992, 1993; Lambrecht 1994; Dik 1997). Einen Überblick über grammatische Kodierungsmöglichkeiten gibt Gómez-González (2001, 11f.). Zu Thema-Markern in anderen Sprachen siehe Fußnote 35.

<sup>32</sup> „Wenn zwei Satzglieder A und B sowohl in der Abfolge A – B wie in der Abfolge B – A auftreten können, und wenn B – A nur unter bestimmten testbaren Bedingungen auftreten kann, denen A

Thompson 1976; Lernerz 1977; Reis 1977; Halliday 1985; Wöllstein-Leisten et al. 1997). Insbesondere die Satzerstposition wird als wichtiges Merkmal zur Identifizierung des Themas aufgefasst.<sup>33</sup> Dürr/Schlobinski (2006, 134) gehen für das Deutsche davon aus, dass die Erstposition im Satz durch ein Thema („Topik“) besetzt wird. In der Regel ist dieses Topik als grammatisches Subjekt realisiert; wird dagegen ein anderes Satzglied in die Erstposition (bzw. Topik-Position) gebracht, spreche man von „Topikalisierung“.

Halliday vertritt die Auffassung, dass die Funktion des Themas darin liege, einen „starting-point for the message“ (Halliday 1985, 39) zu schaffen. Im Englischen werde diese Funktion durch die Satzerstposition realisiert („the theme is what comes first in the clause“ (Halliday 1967, 212)); in anderen Sprachen könne dies jedoch anders sein. Neben der Funktion, einen kataphorischen Bezugspunkt bzw. Ausgangspunkt für den aktuellen Satz zu bilden, stellt das Thema gleichzeitig auch ein Mittel zur Anknüpfung an den vorangegangenen Satz dar (anaphorische Funktion des Themas). Diese Anknüpfungsfunktion basiert auf der Eigenschaft des Themas, eine bereits bekannte Information zu repräsentieren (siehe hierzu auch Kapitel 2.4). Die kommunikative Maxime „Beginne bei dem, was du beim Hörer als bekannt voraussetzen kannst!“ (Welke 2002, 125) lässt sich auf Basis der Annahme, dass das Thema die bekannte Information ist, demnach als Stützung der „Thema-vor-Rhema“-Abfolge anführen.

Für das Englische geht Halliday (1985, 38ff.) davon aus, dass das Thema im Satz an erster Stelle steht. Er unterscheidet zwischen einfachen und multiplen Themen. Die einzelnen Bestandteile des multiplen Themas werden mit Hilfe der drei Metafunktionen „textual“, „interpersonal“ und „ideational“, die auch als die „three aspects of the meaning of the clause“ definiert werden, weiter kategorisiert in „textual“, „interpersonal“ und „ideational Themes“ (vgl. Halliday 1985, 53).<sup>34</sup> Unter „textual Themes“ werden die textinternen und deiktischen Bezüge subsumiert, z.B. Konjunktionen oder Subjunktionen. „Interpersonal Themes“ sind Elemente, in denen die Haltung des Textproduzenten zum Ausdruck kommt, wie z.B. bewertende Adverbien. Als „ideational Themes“ werden die auf den Inhalt ausgerichteten Elemente verstanden. Halliday bezeichnet das „idea-

---

– B nicht unterliegt, dann ist A – B die ‚unmarkierte Abfolge‘ und B – A die ‚markierte Abfolge‘.“ (Lernerz 1977, 27).

<sup>33</sup> In gesprochenen Texten wird zur Bestimmung des Themas darüber hinaus die Intonation genutzt (siehe z.B. Jasinskaja et al. 2004).

<sup>34</sup> „Ideational meaning“ repräsentiert Informationen, die Erfahrungen mit der realen und imaginären Welt darstellen, und wird als eigentlicher Inhalt des Satzes bestimmt. „The ideational function of the clause is that of representing (...) ‘processes’: actions, events, processes of consciousness, and relations“ (Halliday 1985, 53). „Interpersonal meaning“ bezieht sich auf die kommunikativen Aktionen von Sprecher und Hörer. „The interpersonal function of the clause is that of exchanging roles in rhetorical interaction: statements, questions, offers and commands, together with accompanying modalities“ (Halliday 1985, 53). Unter „textual meaning“ werden der sprachliche sowie der situative Kontext verstanden. „The textual function of the clause is that of constructing a message“ (Halliday 1985, 53). Diese Unterscheidung basiert auf der Theorie der Systemisch-Funktionalen Grammatik, die davon ausgeht, dass eine Äußerung als funktionale Einheit gesehen werden muss und drei unterschiedliche Metafunktionen hat: Durch die „textual metafunction“ wird das „Theme“-System begründet, durch die „interpersonal metafunction“ das „Mood“-System, durch die „ideational metafunction“ das „Transitivity“-System.

tional Theme” auch als “Topical Theme”, das im Satz realisiert ist als “some entity functioning as Subject, Complement or circumstantial Adjunct” (Halliday 1985, 54). Da Ideational Themes bzw. Topical Themes als einzige thematische Elemente vorhanden sein müssen (“there is always an ideational element in the Theme”, Halliday 1985, 53), stellen sie die relevantesten Themabestandteile dar.

Andere Linguisten widersprechen der Auffassung, das Thema lasse sich über seine Position im Satz bestimmen. Ausdrücke in Satzerstposition könnten in ihrer Funktion für den Text variieren und sowohl thematisch als auch rhematisch besetzt werden (siehe Givón 1983a, 19f.; Schlobinski und Schütze-Coburn 1992, 97; Lambrecht 1994, 117). Gestützt wird diese Auffassung durch Untersuchungen von Gundel (1977, 30-33) zum Englischen und Russischen. Gundel belegt anhand von Beispielen, dass Satzerstposition weder eine notwendige noch hinreichende Bedingung für Thematizität ist und dass Themen sowohl durch Links- als auch Rechtsversetzung realisiert werden können. Givón (1983a, 20) begründet dies mit dem psychologischen Prinzip „Wende dich erst der dringenden Aufgabe zu“ (“Attend first to the most urgent task”). Wenn das Thema offensichtlich ist, ist es dringender, Aussagen zum bzw. über das Thema zu machen (“Comment”). Wenn das Thema dagegen weniger offensichtlich ist, ist es dringender, das Thema zu etablieren. Da in der Regel das Thema „bekannt“ ist, geht Givón davon aus, dass die Abfolge “Comment-Topic” die unmarkierte ist (vgl. Givón 1983a, 19).

### 2.1.2 Thema und grammatische Rolle

Obwohl das Begriffspaar „Thema-Rhema“ ursprünglich als Opposition zu den grammatischen Kategorien „Subjekt“ und „Prädikat“ eingeführt wurde (als „psychologisches Subjekt“ bzw. „Prädikat“, siehe Kapitel 1.1.2), lassen sich Thema und grammatisches Subjekt nach Ansicht von Chomsky, Givón und Lambrecht dennoch aufeinander beziehen:<sup>35</sup> Chomsky (1965) beschreibt den Zusammenhang zwischen Thema und Subjekt als einen zwischen Oberflächen- und Tiefenstruktur bestehenden: Thema und Rhema (bzw. Topic und Comment) seien

---

<sup>35</sup> In den sprachtypologischen Arbeiten von Li/Thompson (1976) wird im Gegensatz hierzu die Position vertreten, dass sich Subjekt und Thema durch ihre Funktion für den Text unterscheiden. Während das Thema (im Satz) bezogen ist auf den Text als Ganzen und einen inhaltlichen Rahmen für die Prädikation (den Comment-Teil des Satzes) bildet, ist das Subjekt ausschließlich bezogen auf den Satz selbst, insofern es eine Orientierung bzw. einen “point of view” für die durch das Verb bezeichnete Aktion o.ä. schafft (vgl. Li und Thompson 1976, 463ff.). Allerdings stellen Li/Thompson die Hypothese auf, dass die Rolle des Themas grammatisch ebenso bedeutsam sei wie die des Subjekts und dass sowohl “topic-comment” als auch “subject-predicate” als grammatische Relationen zu betrachten seien (vgl. Li und Thompson 1976, 459). Sprachen würden sich allerdings bezüglich der Regeln und Strategien zur Satzbildung unterscheiden, d.h. bezüglich der Rolle, die Subjekt und Thema als grammatische Kategorien für die Satzbildung spielen: In manchen Sprachen (z.B. Mandarin) sei das Thema ausschlaggebend für die Wortstellung, in anderen Sprachen (z.B. Indo-germanische Sprachen) das Subjekt. Daneben gebe es Sprachen (z.B. Japanisch und Koreanisch) für die beide Konzepte gleich wichtig seien. Schlobinski/Schütze-Coburn (1992, 105) wenden ein, dass die beschriebenen Unterschiede zwischen Tp- und Sp-Sprachen lediglich auf unterschiedliche Formen der Markierung (syntaktisch-semantic vs. morphosyntaktisch) zurückzuführen seien.

grammatische Relationen der Oberflächenstruktur, die mit der Subjekt-Prädikat-Relation der Tiefenstruktur korrespondierten.<sup>36</sup> Dementsprechend geht Chomsky davon aus, dass Thema und Subjekt häufig (jedoch nicht immer) miteinander korrelieren. Auch Givón (1983a, 22) legt dar, dass Thema und Subjekt (“primary topic”) stärker miteinander verknüpft sind als Thema und direktes Objekt (“secondly topic” [sic]) oder Thema und andere grammatische Funktionen – allerdings beschreibt er den Zusammenhang als einen, der vom Subjekt bzw. dem direkten Objekt ausgeht: Das Subjekt sei mehr Thema als das direkte Objekt und dieses wiederum sei „thematischer“ als das indirekte Objekt, und nicht: Das Thema ist häufiger Subjekt als Objekt (vgl. Givón 1992, 11).<sup>37</sup> Lambrecht (1994, 132) geht ebenfalls davon aus, dass (zumindest im Englischen) das Subjekt das unmarkierte Thema ist. Sprachbenutzer hätten eine unbewusste Tendenz, die grammatische Subjekt-Prädikat-Struktur mit der pragmatischen Topic-Comment-Struktur in Einklang zu bringen. Empirisch belegt werden könne dies durch die Tatsache, dass in kohärenten Diskursen die Mehrzahl der grammatischen Subjekte als Pronomen realisiert sei, die als ein zentrales Mittel zur Signalisierung von thematischer Kontinuität im Diskurs betrachtet werden können (siehe auch Kapitel 2.2.1.2). Laut Welke (2002, 126) stimmen Thema und Subjekt im Deutschen allerdings nur in ca. 60% der Fälle überein.<sup>38</sup>

### 2.1.3 Thema und thematische/semantische Rolle

Das Thema wird auch in Zusammenhang gebracht mit der semantischen Rolle AGENS. So konstatiert Givón (1983a, 22), dass Satzglieder in der Rolle des AGENS häufiger Thema seien als Satzglieder in anderen Rollen.

Das Konzept der Kasusrollen (Fillmore), semantischen Rollen (Montague), thematischen Rollen (Gruber, Jackendoff, Chomsky) oder Theta-Rollen<sup>39</sup> (Chomsky) wurde seit den 60er Jahren in zahlreiche syntaktische und semantische Theorien, z.B. die generative Grammatik, generative Semantik oder die Lexical-

<sup>36</sup> Welke geht ebenfalls von einer Korrespondenz zweier Ebenen aus, allerdings rekurriert er dabei nicht auf die Konzepte der Oberflächen- und Tiefenstruktur. Er vertritt stattdessen die Auffassung, dass es zwei Ebenen der Perspektivierung im Satz gebe, von denen die Thema-Rhema-Gliederung (Ebene der äußeren Perspektivierung) eine sei, die aus der Bedeutung des Verbs folgende inhärente Perspektivierung die zweite. Beide Perspektivierungsebenen könnten übereinstimmen – in diesem Fall korrelieren Subjekt und Thema – sie könnten einander aber auch widersprechen; die Thema-Rhema-Gliederung habe in diesem Fall die Funktion der Korrektur und Umperspektivierung (vgl. Welke 2002, 131).

<sup>37</sup> Givón ist allerdings kein Vertreter einer auf den Satz bezogenen Thema-Konzeption. Er beschäftigt sich zwar mit der Thematizität im Satz, geht aber davon aus, dass das Thema nur im Diskurskontext von Bedeutung sei: “Put plainly and in operational terms, the topic is only ‘talked about’ or ‘important’ if it remains ‘talked about’ or ‘important’ through a number of successive clauses.” (Givón 1992, 12). Siehe auch Gundel (2004, 192) und Lambrecht (1994, 120).

<sup>38</sup> Givón geht dagegen von einer Übereinstimmung von 80-90% aus (vgl. Givón 1979, 210, zit. n. Welke 2002, 126).

<sup>39</sup> In der Government & Binding Theorie werden Thetarollen etwas abweichend von semantischen/ thematischen Rollen definiert. Es wird angenommen, dass jedes Verb ein Raster mit Thetarollen aufweist. Einstellige Verben haben nur eine Rolle im Thetaraster, zweistellige zwei usw. Wichtig ist, dass die thematischen Rollen im Raster geordnet werden: Die erste Rolle ist AGENS, die zweite THEMA/THEME, die dritte ZIEL (vgl. von Stechow 1991, 135f.).

Functional Grammar integriert (vgl. Carlson 1984, 259). Von ihrem Status her sind thematische Rollen in der Schnittmenge von Grammatik und Semantik anzusiedeln. Sie beschreiben sowohl ein grammatisches als auch zugleich semantisches Phänomen<sup>40</sup>, nämlich die Tatsache, dass die Argumente eines Prädikats unterschiedliche Funktionen bzw. Rollen einnehmen können, und zwar nicht nur grammatisch, sondern auch semantisch (Handelnder/Agens, Erfahrender/Experiens, Nutznießer/Benefaktiv etc.<sup>41</sup>). Grundgedanke ist, dass sich Rollen nicht unabhängig vom Satz bestimmen lassen, sondern erst durch das Prädikat (alternativ – in Abhängigkeit von der Theorie – auch durch das Verb) und dessen syntaktisch-semantischen Aussagerahmen (Argumentstellen und semantische Funktionen) determiniert werden (vgl. von Polenz 1985, 167ff.). Prädikate bzw. Verben bilden demnach einen Aussagerahmen für den Satz, der spezifische semantische Leerstellen (Typen von thematischen Rollen) eröffnet.

Wie stellt sich nun der Zusammenhang zwischen dem Konzept Thema und dem Phänomen der durch das Prädikat determinierten thematischen Rollen dar? Das Satzthema wird als Gegenstand definiert, über den in der Prädikation etwas mitgeteilt wird. Wie bereits dargelegt, fällt es häufig mit dem grammatischen Subjekt zusammen. In diversen Rollenhierarchien, durch die die Reihenfolge der Rollen im Satz und teilweise der Zusammenhang mit syntaktischen Funktionen dargestellt wird, wird in der Regel die Rolle AGENS an die erste Stelle gesetzt und mit dem grammatischen Subjekt verknüpft (vgl. Welke 2002, 98f.). Grund für diese Reihenfolge ist nach Welke, dass die Rolle AGENS das erste Argument des Verbs sei und als solches die Perspektivierung des Sachverhalts bedinge. Signalisiert werde dies durch die syntaktische Spitzenposition. Auch Thema und Rhema sind Kategorien, die eine Gewichtung und Perspektiviertheit der Satzmitteilung ausdrücken. Thematische Rollen und Thema-Rhema-Strukturierung haben demnach ähnliche Funktionen für den Aufbau des Satzes bzw. die Strukturierung der Informationen im Satz. Thematische Rollen sind jedoch sprachinhärent determiniert durch die syntaktisch-semantische Valenz des Prädikats bzw. Verbs, die Konzepte Thema und Rhema dagegen durch die Intention des Produzenten sowie den Kontext der Äußerung.

Givón (1983a, 22) stellt die Hypothese auf, dass das Thema häufig in der thematischen Rolle des AGENS auftrete. Allerdings bleibt vollkommen unklar, wie die Koppelung von Subjekt, Agens, Thema und Satzerstposition theoretisch begründet werden kann. Die Zuordnung des Themas zu einer spezifischen thematischen Rolle scheint meiner Ansicht nach eher eine sekundäre Folge der Beziehung zwischen Thema und Subjekt auf der einen Seite und Subjekt und thematischer Rolle auf der anderen Seite zu sein.

Zum Zusammenhang zwischen der spezifischen Rolle THEME und dem Satzthema gibt es offenbar noch keine Untersuchungen – zumindest haben entsprechende Recherchen nichts hierzu ergeben. Aufgrund der terminologischen Über-

---

<sup>40</sup> Zu dieser Diskussion siehe z.B. Carlson (1984) und Rauh (1988).

<sup>41</sup> Über die Anzahl und Benennung der Rollen gibt es sehr unterschiedliche Auffassungen, was daran liegt, dass sich „semantische Rollen (...) anscheinend nicht scharf voneinander abgrenzen: Sie sind vage und prototypisch bestimmt.“ (Welke 2002, 95).



schneidung (“THEME” – Thema/“Theme”) sind an dieser Stelle jedoch einige Gedanken zur konzeptionellen Gemeinsamkeit angebracht: Die Rolle THEME wurde zuerst von Gruber eingeführt (vgl. Gruber 1965, nach Rauh 1988, 221ff.). Zur Charakterisierung dieser Rolle zieht er je nach semantischer Verbklasse unterschiedliche Aspekte heran: Im Hinblick auf Bewegungsverben gilt THEME als NP, die den Gegenstand, der sich bewegt, ausdrückt. Bei Verben, die zur Lokalisierung gebraucht werden, ist THEME dagegen der Gegenstand, über dessen Lokalisierung Angaben gemacht werden. Schwarz definiert die Rolle THEMA/OBJEKT allgemein als „belebtes oder unbelebtes Objekt einer Handlung oder eines Zustands“ (2004, 71). In dem Satz “Bill rolled the ball down the hill intentionally.” hat “the ball” die Rolle THEME inne, “Bill” die Rolle AGENS (vgl. Rauh 1988, 233f.).<sup>42</sup> Allerdings ließe sich “the ball” auch durch die Rolle BETROFFENES oder PATIENS beschreiben. Die Rolle THEME scheint demnach kein Alleinstellungsmerkmal zu besitzen. Dies deckt sich mit der Tatsache, dass in vielen Auflistungen von thematischen Rollen die Rolle THEME nicht enthalten ist (siehe z.B. von Polenz 1985, 170ff.; Leinfellner 1992, 240ff.). Aufgrund der unklaren Definition der Rolle THEME lässt sich die Beziehung zwischen THEME und Thema maximal als vage semantisch-ontologischer Zusammenhang beschreiben: Die Rolle THEME ist nach der Definition Grubers das Objekt, das von der Semantik des Verbs affiziert wird, das Thema ist der Gegenstand, über den in der Prädikation etwas ausgesagt wird. Beide Konzepte sind demnach semantisch eng auf das Verb bezogen, wobei die Rolle THEME durch das Verb determiniert wird, das Thema aus Sicht grammatischer Thema-Konzeptionen dagegen den Bezugspunkt für die durch das Verb gesteuerte Satzaussage bildet, also nicht semantisch vom Verb abhängt, sondern eher als Kopf der Satzmitteilung gelten kann. Abschließend ist festzuhalten, dass THEME (als thematische Rolle) und Thema allerdings eher geringe gemeinsame Bedeutungsanteile haben und deshalb als voneinander unabhängige Termini mit divergierenden Bedeutungen aufgefasst werden sollten – die Verwendung des Terminus THEME zur Bezeichnung der oben beschriebenen thematischen Rolle ist allerdings im Kontext der terminologischen und konzeptuellen Unklarheiten, die es bezüglich des Themas gibt, als wenig hilfreich anzusehen.

#### 2.1.4 Fazit

Die grammatischen Thema-Konzeptionen zeichnen sich vor allem durch einen Mangel einer eindeutigen und distinkten Definition aus und versuchen stattdessen, sich dem Konzept des Themas über sekundäre Merkmale (z.B. Satzposition, grammatische Rolle, thematische Rolle) anzunähern. Der eigentliche Status des Themas als grammatisches Konzept bleibt dagegen schwammig – steuert das Konzept die syntaktische Struktur regelhaft oder ist es ein (mehr oder weniger schwaches) kommunikatives Mittel zur Gewichtung der Informationen im Satz? Wenn unter Thema letztlich nur das Element in der Funktion des grammatischen Subjekts, des AGENS oder des Elements an Satzerstposition verstanden

---

<sup>42</sup> Thema des Satzes nach der Konzeption von Halliday wäre dagegen „Bill“ (zu Hallidays Konzeption siehe Kapitel 2.2.1.1).

wird, stellt sich die Frage, was überhaupt der Mehrwert eines eigenständigen Konzepts Thema ist.

Bei den Ansätzen, in denen das Thema als Element in Satzerstposition gesehen wird, ist weiterhin problematisch, welches Element überhaupt das Erste im Satz ist – dies ist vor allem bei gesprochener Sprache schwierig zu bestimmen (z.B. aufgrund von Wiederholungen, Verbesserungen, nachgeschobenen Satzteilen). Aber auch für geschriebene Texte gibt es bisher keinen Konsens darüber, welches Element „thematisch“ ist und ob z.B. Konnektoren, die in Satzerstposition stehen, ebenfalls als Thema gelten sollen, oder nicht. Ist nur die erste Konstituente im Satz „thematisch“, oder gleich alles, was vor dem finiten Verb steht? Hallidays Antwort eines „multiplen Themas“ ist meiner Ansicht nach in Hinsicht auf eine befriedigende Definition des Konzepts Thema nicht weiterführend, da so unterschiedliche Wortarten wie Konjunktionen, Adverbien, Substantive gleichermaßen als „Theme“ bezeichnet werden – auch wenn Halliday Phrasen in der syntaktischen Funktion des Subjekts, Komplements oder eines bestimmten Typs von Adjunkt noch einmal gesondert als „Topical Theme“ auszeichnet. Für die Operationalisierung der Analyse des Themas ist diese sehr strikt grammatische Thema-Bestimmung dagegen sehr gut geeignet (siehe Schwarz et al. 2008). In den Ansätzen zum Thema als Element an Satzerstposition fehlt darüber hinaus eine differenzierte Darstellung des Zusammenhangs zwischen dem Thema und den verschiedenen Typen von Satzerstpositionen (Vorfeld, Vor-Vorfeld bzw. Linksherausstellung, Spaltsatz etc.) und deren unterschiedlicher kommunikativer Funktion.<sup>43</sup>

## 2.2 Thema als semantisches Konzept

Semantische Thema-Konzeptionen haben eine lange Tradition, sowohl im Hinblick auf die Definition des Satz- als auch Textthemas. Sie lassen sich in zwei Gruppen unterteilen:

1. Thema als Gegenstand, über den etwas gesagt wird
  - a) Thema als Mitteilungsgegenstand
  - b) Thema als fortlaufend konstituierter zentraler Gegenstand des Textes
2. Thema als abstrakter Kern des Textinhalts

Der erste Typ von Thema-Konzept geht zurück auf die bereits in Kapitel 1.1.2 genannten Definitionen von Mathesius (Einführung und Definition von „Thema“ und „Rhema“, vgl. Mathesius 1971 [1929]) und Paul (Einführung und Definition des „psychologischen Subjekts“, vgl. Paul 1898). Das Thema ist demnach das, worüber der Sprecher etwas sagt, das Rhema das, was darüber gesagt wird. Die Thema-Konzepte, die das Thema auf Basis dieser Definitionen als satz- bzw. äußerungsbezogene Kategorie auffassen, werden in Kapitel 2.2.1.1 ausführlicher behandelt; Ansätze, in denen dieses Thema-Konzept vom Satz auf den Text über-

---

<sup>43</sup> Zur weiteren Kritik an grammatischen Thema-Konzeptionen siehe auch Hoffmann (2000, 345) und Gómez-González (2001, 52ff.).

tragen wird, in Kapitel 2.2.1.2. Die satz- und textbezogenen „Thema als Gegenstand“-Konzepte weisen klare Gemeinsamkeiten auf – eine strikte Trennung der Konzepte Satz- und Textthema ist deshalb meines Erachtens nicht sinnvoll, da so interessante theoretische und analytische Zugänge zum Konzept Thema verschlossen würden.

Die zweite Gruppe semantischer Themadefinitionen lässt sich subsumieren unter dem Stichwort „Thema als abstrakter Kern des Textinhalts“. Es wird davon ausgegangen, dass sich aus einem Thema als abstraktem semantischem oder „begrifflichem“ Kern heraus verschiedene Varianten von Texten entwickeln lassen. Diese sind dann zwar an der Textoberfläche unterschiedlich realisiert, teilen aber die gleiche abstrakte Textbasis. Umgekehrt kann ein vorliegender Text auf eine semantische Textbasis, einen semantischen Kern kondensiert werden. Texte werden in diesem Verständnis als hierarchisch geordnete Menge von Propositionen bzw. „Makropropositionen“ verstanden. Das Thema selbst wird als Top-Level-„Makroproposition“ oder „Basisproposition“ aufgefasst, d.h. als „Makroproposition“, die an der Spitze der hierarchisch geordneten Propositionen oder Makropropositionen steht. Eine etwas alternative Konzeption von Thema als „semantischem Kern“ fasst das Thema als „Prädikat-Aktanten-Struktur“ auf, die – analog zur Annahme der Makropropositionstheorie – die semantische Grundlage des Textes darstellt (siehe Kapitel 2.2.2).

## **2.2.1 Thema als Gegenstand, über den etwas gesagt wird**

### **2.2.1.1 Thema als Mitteilungsgegenstand**

Die Definition des Themas als „Mitteilungsgegenstand, über den etwas gesagt wird“, schließt sich an die informationsstrukturellen und grammatischen Thema-Konzepte an, die im Rahmen der Prager Schule und der Systemisch Funktionalen Linguistik entwickelt wurden, und ergänzt diese Konzepte durch eine semantische Charakterisierung des Themas:

- “the Theme (...) is what the clause is going to be about” (Halliday 1985, 39);
- Das Thema ist der Mitteilungsgegenstand („das, worüber etwas mitgeteilt wird“), das Rhema die Satzaussage („das, was darüber mitgeteilt wird“) (Daneš 1970, 591);
- “The topic of a sentence is the thing, which the proposition expressed by the sentence is ABOUT.” (Lambrecht 1994, 118, Hervorhebung im Original)

Auch wenn sich die Definitionen in einigen Punkten unterscheiden – Halliday spricht von “clause”, Daneš von „Satzaussage“ und Lambrecht von der “proposition expressed by the sentence”, es ist die Rede von “Theme”, „Thema“ und “Topic” (im Folgenden: Thema) – überwiegt die Gemeinsamkeit zwischen den Definitionen: Das Thema wird als relationales Konzept aufgefasst, das zum Rest der Mitteilung in einem Zusammenhang steht, der in der Forschung als “Aboutness“-Relation bezeichnet wird (vgl. Lambrecht 1994, 118; Gómez-González 2001, 16). Das Thema selbst wird in diesen relationalen “Aboutness“-Definitionen entweder gar nicht spezifiziert (Halliday) oder nur sehr vage, als

„Mitteilungsgegenstand“ (Daneš) oder “thing” (Lambrecht) – wie dieser Mitteilungsgegenstand grammatisch oder semantisch beschaffen ist, bleibt jedoch weitgehend offen. Gómez-González konstatieren deshalb:

Nevertheless, despite its broad appeal, the semantic interpretation faces one significant problem. For one thing, the basis for the identification of ‘aboutness’ has rarely been made explicit, probably because of the inherent vagueness of the notion, and as a result, its specification and that of the categories involved has been highly problematic. (Gómez-González 2001, 16)

Eine interessante Konkretisierung des „Themas als Mitteilungsgegenstand“ liefert Lambrecht. Er bestimmt das Thema als Ausdruck, der eine Entität bezeichnet bzw. für einen Diskursreferenten<sup>44</sup> steht, über den etwas in der Proposition gesagt wird:<sup>45</sup>

Since the topic relation is a relation between a REFERENT and a PROPOSITION, it is natural that the term ‘topic’ should be understood as designating the ENTITY which the proposition is about, i.e. the discourse referent itself about which information is being conveyed in a proposition. (Lambrecht 1994, 127, Hervorhebung im Original)

Diese Definition des Themas als Diskursreferent oder Diskursentität stellt eine wichtige Brücke zwischen satz- und textbezogenen Thema-Konzepten dar. Sie ist einerseits bezogen auf eine einzelne Proposition, andererseits führt sie aber mittels des Terminus „Diskursreferent“ über die einzelne Proposition hinaus auf den Diskurs als Ganzen. Wie im folgenden Kapitel gezeigt wird, ist eine Möglichkeit der Bestimmung des Themas, die Menge der Diskursreferenten als Basis für die Ermittlung des Diskursthemas zu betrachten, da davon auszugehen ist, dass der zentrale, fortlaufend verwendete Diskursreferent das Text- bzw. Diskursthema reflektiert.

### 2.2.1.2 Thema als fortlaufend konstituierter zentraler Gegenstand des Textes

Ausgehend von den Arbeiten zum Satzthema, die das Thema als Gegenstand bestimmen, über den im Satz etwas mitgeteilt wird, wird das Thema in der Folge definiert als Gegenstand, über den im Text etwas gesagt wird. Diese Ausweitung des „Thema als Gegenstand“-Konzepts auf den Text beginnt mit dem Ansatz der „thematischen Progression“ von Daneš (1970; 1976). Er überträgt die auf den Satz bezogene „Thema-Rhema“-Gliederung der Prager Schule auf den Text und dessen Gliederung. Dabei geht Daneš von der Annahme aus, dass es bei den Übergängen zwischen Sätzen thematisch-strukturelle Regelmäßigkeiten gibt, dass feste Typen „thematischer Progression“ existieren:

<sup>44</sup> Diskursreferenten stellen Einheiten der textinternen Diskurswelt dar, auf die im Diskurs (bzw. Text) mittels eines sprachlichen Ausdrucks Bezug genommen wird. Ein Diskursreferent kann nach Lambrecht sowohl eine Entität als auch eine Proposition sein, d.h. ein Diskursreferent kann sowohl nominal als auch propositional realisiert sein. Syntaktisch sind Diskursreferenten Argumente (NPs, Pronomen, subordinierte Sätze etc.), nicht jedoch Prädikate, da diese keine Diskursreferenten bezeichnen (vgl. Lambrecht 1994, 74f.).

<sup>45</sup> Lambrecht unterscheidet zwischen “topic” (auch “topic referent”) und “topic expression” (auch “topic phrase”): “Topic” ist die Entität, auf die mittels eines linguistischen Ausdrucks (der “topic expression”) verwiesen wird (vgl. Lambrecht 1994, 128).

- a) *einfache lineare Progression*: Rhema des Vordersatzes wird Thema des Folgesatzes;
- b) *Progression mit durchlaufendem Thema*: Vordersatz und Folgesatz haben dasselbe Thema; zusätzlich kann jeweils ein neues Thema hinzukommen;
- c) *Progression mit abgeleitetem Thema*: Themen sind von einem Hyperthema abgeleitet, zu dem sie in einer semantisch-pragmatischen Beziehung stehen;
- d) *Entwicklung eines gespaltenen Themas*: Themen der Folgesätze bezeichnen Teilmengen des Themas oder Rhemas des Vordersatzes;
- e) *Progression mit thematischem Sprung*: wie a) oder c), nur erfolgt der Übergang unter Überspringung eines verbindenden Zwischengliedes.

Die Relationen zwischen zwei Themen aufeinander folgender Sätze oder Segmente sind Daneš zufolge semantischer Natur, sie beruhen darauf, dass

die als Thema ausgewählte Bedeutungskomponente ( $K_2$ ) zu einer anderen im vorhergehenden Textabschnitt beinhaltenen Bedeutungskomponente ( $K_1$ ) in einer spezifischen semantischen Beziehung steht; diese Beziehung ist entweder eine echte Identitätsrelation oder eine Relation der semantischen Verwandtschaft, die es ermöglicht, dass Sprecher und Hörer die Komponente  $K_2$  aus dem Kontext heraus schon als ‚bekannt‘ oder ‚gegeben‘ betrachten können. Man könnte sagen, dass  $K_2$  durch  $K_1$  angedeutet wird, dass zwischen  $K_2$  und  $K_1$  die Beziehung einer spezifizierten (direkten oder vermittelten) semantischen Implikation oder Inferenz besteht. (Daneš 1976, 36)

Gegen das Modell der thematischen Progression wurden vor allem in der germanistischen Linguistik zahlreiche Einwände hervorgebracht (siehe Lötscher 1987; Heinemann und Viehweger 1991; Brinker 1997; Schröder 2003). Unter anderem wurde kritisiert, dass das Modell keine Differenzierung von Satz- und Textthema biete und unklar sei, wie man von der lokalen Textprogression zur globalen Themenentwicklung und vom Satzthema zum Textthema komme (wobei Daneš durch die Konzeption eines „Hyperthemas“ zumindest theoretisch implizite Hauptthemen zulässt).

Hoffmann (1997; 2000) orientiert sich ebenfalls an der satzbezogenen „Thema-Rhema“-Konzeption und weitet sie, wie Daneš, auf den Text und Diskurs aus. Er definiert Thema als „kommunikativ-konstituierten Gegenstand oder Sachverhalt, über den in einem Diskurs oder Text(-abschnitt) fortlaufend etwas gesagt wird“ (Hoffmann 1997, 509). Mit dieser Definition ordnet er seine Konzeption zwar in einen kommunikativ-pragmatischen Rahmen ein, allerdings stehen im Zentrum des Ansatzes die semantischen und referenziellen Beziehungen zwischen Themen; Hoffmanns Thema-Konzeption wird deshalb in dieser Arbeit unter das Kapitel „Thema als semantisches Konzept“ subsumiert.<sup>46</sup>

---

<sup>46</sup> Wie bereits deutlich gemacht wurde, sind Querklassifizierungen nicht zu vermeiden, und die Einordnung eines Themenkonzepts als grammatisch oder semantisch bedeutet nicht, dass es nicht gleichzeitig funktional, handlungstheoretisch oder kognitiv ausgerichtet sein kann.

Wie Daneš beschreibt Hoffmann verschiedene Möglichkeiten der thematischen Anknüpfung zwischen Sätzen. Auf der obersten Ebene seiner „Typen thematischer Organisation“ unterscheidet er Themenfortführung (= konstantes Thema) und Themenentwicklung. Während erstere vor allem durch Rekurrenz realisiert wird, handelt es sich bei der Themenentwicklung „um einen unselbständigen Akt, mit dem von einem Thema  $Th_1$  zu Themen  $Th_2 \dots Th_n$  übergegangen wird, die sich von ihm unterscheiden, aber mit ihm verknüpft sind oder wenigstens in einen gemeinsamen Rahmen einzuordnen sind“ (Hoffmann 2000, 353).

In den Arbeiten von Daneš und Hoffmann wird die „Thema-Rhema“-Gliederung zwar auf den Text übertragen, dennoch werden in erster Linie Satzthemen (und -rhemen) und deren Beziehungen zueinander untersucht. Aufgrund dieser Fokussierung auf Satzthemen stellt sich die Frage, inwiefern mit dieser Analysemethode tatsächlich Aussagen über das Thema des Textes getroffen werden können. Zwar wird in Daneš „thematischer Progression mit abgeleitetem Thema“ und Hoffmanns „Themenkomposition“ die Existenz impliziter Hauptthemen („Hyperthemen“) präsupponiert, es wird aber in keinem Progressions- bzw. Organisationstyp eine differenzierte Beschreibung dieser Zusammenhänge zwischen Hyperthema und Subthemen geleistet. Sowohl bei Hoffmann als auch bei Daneš müssen Thema und Rhema explizit im Text vorkommende Objekte sein, um in die Analyse der „thematischen Progression“ bzw. der „thematischen Organisation“ einbezogen zu werden. Obwohl per definitionem die thematischen Beziehungen untersucht werden sollen, sind es auf diese Weise vor allem die anaphorischen Beziehungen zwischen syntaktischen Positionen im Satz, die in der Analyse berücksichtigt werden. Genau diese Fokussierung auf anaphorische Beziehungen zwischen Sätzen hat zur Folge, dass zwischen der Thema-Rhema-Analyse und der Analyse der Wiederaufnahmestruktur kein wesentlicher Unterschied festzustellen ist (vgl. Brinker 1997, 51).

Die Ansätze von Daneš und Hoffmann leisten deshalb zwar einen Beitrag zur Untersuchung der Kohärenz von Texten, indem sie Abfolgen von Sätzen im Hinblick auf ihre referenzielle Kohärenz untersuchen. Beide Ansätze sind jedoch beschränkt auf die Analyse der lokalen thematischen Zusammenhänge auf der Textoberfläche. Die Untersuchung der globalen thematischen Zusammenhänge, über die beispielsweise der Differenzierung von Haupt- und Nebenthemen eines Textes nähergekommen werden kann, wird durch diese referenziell orientierten Ansätze dagegen nicht geleistet.

Givón (1983a; 1983b; 1992) verfolgt einen ähnlichen Ansatz wie Daneš und Hoffmann, definiert das Thema jedoch nicht als Satzkonstituente, sondern als Diskursreferent, der in einem Absatz oder Text in kontinuierlicher Weise wiederaufgenommen wird.

At the level of a single event/state, ‘topic’ – ‘what is talked about’ or ‘what is important’ – is meaningless. (...) Thematic coherence across a multi-clause chain means continuity (‘recurrence’) of the sub-elements of coherence, chief among which are the referents/topics. Coherent discourse is thus characterized by *equi-topic clause chains*. And ‘topic’ is a relevant functional notion only at the discourse level, minimally at the chain or paragraph level. Put plainly and in operational terms, the topic is only ‘talked about’ or ‘important’ if it remains ‘talked about’ or ‘important’ through a number of successive clauses. (Givón 1992, 11f., Hervorhebung im Original)

Zentral für Givóns Ansatz ist, dass “Topic continuity” mit “Topic availability” verknüpft wird, d.h. dass Diskursreferenten thematisch sind, weil sie mental besonders gut zugänglich für den Rezipienten sind.<sup>47</sup> Faktoren, die die Zugänglichkeit eines Themas beeinflussen, sind “referential distance (‘look-back’)” (Abstand zwischen aktueller und vorausgegangener Erwähnung eines Diskursreferenten in Sätzen), “potential interference (‘ambiguity’)” (Vorkommen konkurrierender Diskursreferenten) und “persistence (‘decay’)” (Grad der Wichtigkeit eines Themas im Diskurs). Er konkretisiert dies durch eine Skala, auf der grammatische Mittel im Hinblick auf ihre Bedeutung für die Realisierung von Themen angeordnet sind; von besonderer Relevanz sind seiner Ansicht nach dabei insbesondere Null-Anapher, unbetonte und betonte Pronomen sowie rechtsdislozierte definite NPs (vgl. Givón 1983a, 13ff.). Thematische Diskursreferenten sind darüber hinaus häufig als grammatisches Subjekt realisiert (vgl. Givón 1983a, 8). Trotz der expliziten Ablehnung von Satzthemen basiert Givóns Ansatz also ebenso wie der von Daneš und Hoffmann auf der Analyse der referenziellen und grammatischen Struktur. Der wichtigste Unterschied seines Ansatzes gegenüber Daneš Arbeit zur „Thematischen Progression“ ist jedoch, dass Givón thematische Kontinuität auch als kognitive bzw. mentale Kontinuität versteht, die in Form bestimmter grammatischer Mittel sprachlich realisiert bzw. signalisiert wird.<sup>48</sup>

Einen weiteren Ansatz zur Verknüpfung aufeinander folgender Sätze stellt die Centering-Theorie dar (Grosz et al. 1995). Die Centering-Theorie ist wie Givóns Ansatz zur “Topic Continuity” ein referenzieller Ansatz zur Erklärung der lokalen Textkohärenz; Gegenstandsbereich der Theorie sind die Interaktionen zwischen lokaler Kohärenz und der Wahl referierender Ausdrücke. Die zentrale Aussage der Theorie ist, dass jede Äußerung in einem Diskurs eine Menge an “forward-looking centers” und exakt ein “backward-looking center” enthält, das nach Auffassung von Oberlander (vgl. 2004, 217f.) das Thema des Satzes darstellt. Forward-Looking Center<sup>49</sup> stellen potentielle Foki bzw. Aufmerksamkeitszentren dar, d.h. Diskursentitäten, die für den gesamten Diskurs relevant sind, während Backward-Looking Center dem aktuellen Diskursfokus entsprechen; es handelt sich um Entitäten, die im Zentrum einer einzelnen Äußerung des Diskurses stehen (vgl. Grosz et al. 1995, 205). Grosz et al.s Ansatz stellt demnach eine Brücke zur Beschreibung der Strukturierung von Sätzen und Texten dar.

<sup>47</sup> “The cognitive correlates of continuity or recurrence is mental accessibility or its converse, mental effort.” (Givón 1992, 7).

<sup>48</sup> Eine weiterführende Arbeit, in der die Vorstellung der thematischen Kontinuität auf mündliche Diskurse übertragen wird, ist die von Dik (1997). Gemeinsam ist dieser Arbeit mit der von Givón (1983a), dass davon ausgegangen wird, dass thematische Kontinuität in erster Linie durch grammatische Mittel signalisiert wird: “Topic maintenance is achieved by a variety of means: different forms of anaphorical reference; syntactic parallelism; switch reference mechanisms; obviation. These different means (...) can be interpreted as signalling that the speaker is still talking about the same things.” (Dik 1997, 437 – im Original als Aufzählungsliste layoutiert).

<sup>49</sup> Die Menge der Forward-Looking Center lässt sich Grosz et al. (1995) zufolge durch ein Ranking ordnen. Das bedeutendste bzw. am höchsten eingeordnete Forward-Looking Center besitzt die größte Wahrscheinlichkeit, zum Backward-Looking Center des Folgesatzes zu werden.

Das Backward-Looking Center bzw. Thema hat die Funktion, die Äußerung mit einem der Forward-Looking Center des Vorgängersatzes zu verbinden. Die Relation zwischen diesen beiden Centern kann dabei auf verschiedene Weise realisiert sein – Grundlage ist jeweils der referenzielle Bezug zwischen den Centern. Grosz et al. (1995, 210) unterscheiden drei Typen von Übergangsrelationen zwischen Äußerungen:

1. “Center Continuation”: Das Backward-Looking Center bzw. Thema der aktuellen Äußerung wird auch das der nachfolgenden Äußerung; die Entität hat die höchste Bedeutung unter allen potentiellen Forward-Looking Centern der zweiten Äußerung, deshalb wird es vermutlich auch in der nächsten Äußerung das Backward-Looking Center bleiben.
2. “Center Retaining”: Das Backward-Looking Center bzw. Thema der aktuellen Äußerung wird auch das der nachfolgenden Äußerung; die Entität hat allerdings nicht die höchste Bedeutung unter allen potentiellen Forward-Looking Centern der zweiten Äußerung; es ist deshalb unwahrscheinlich, dass es in der nächsten Äußerung weiter Backward-Looking Center ist.
3. “Center Shifting”: Das Backward-Looking Center bzw. Thema zwischen zwei aufeinanderfolgenden Äußerungen ist unterschiedlich.

Linguistische Mechanismen, die die Etablierung und Fortführung einer Entität als Backward-Looking Center bzw. Thema begünstigen, sind z.B. Pronominalisierung und Subjektposition. Auch die grammatische Rolle einer Entität stellt eine wichtige Determinante beim Ranking der Forward-Looking Center und somit bei der Auswahl eines Forward-Looking Center als nächstes Backward-Looking Center dar: SUBJECT > OBJECT > OTHER. Vermutlich spielen auch die Wortstellung (Linksherausstellung), Subordination und lexikalische Semantik eine Rolle beim Ranking (vgl. Grosz et al. 1995, 212ff.).

Die Centering-Theorie mit ihrem Begriffspaar Forward- und Backward-Looking Center wird im Gegensatz zu den zuvor genannten Begriffspaaren, die einen Beitrag zur Strukturierung des Satzes leisten, nicht auf einen Satz bzw. eine einzelne Äußerung angewendet, sondern auf den Satz *im Kontext größerer Diskurseinheiten*. Im Zentrum steht dabei die Frage, wie kleinere Textsegmente zu kohärenten größeren Segmenten verknüpft werden. Um diese Frage zu beantworten, untersuchen Grosz et al. vor allem die referenziellen Übergänge zwischen *mehreren* aufeinander folgenden Sätzen.<sup>50</sup> Eine ähnliche Perspektive wird auch in

<sup>50</sup> Zur Bewertung der Kohärenz von Übergängen zwischen Äußerungen und in der Summe dessen auch der globalen Kohärenz des Diskurses schlägt Karamanis (2003, in: Oberlander 2004, 217) ein Verfahren vor, das die in der Centering-Theorie vorgeschlagenen Übergänge zwischen Äußerungen (Continue, Retain, Smooth-Shift und Rough-Shift) mit Punkten bewertet und die addierten Punkte in ein Verhältnis mit der Anzahl der Backward-Looking Center setzt, wobei auch die Übergänge insgesamt gezählt werden.

Beispiel: “I met someone yesterday. His name is John. He likes Mary. Mary is a professor. A professor came to my house yesterday. Her name is Chris. She teaches chemistry. Chemistry is the science of the future, my Dad always says.” → 7 Übergänge, 4 “Continues”, 2 “Smooth-Shifts”, 1 “No-CB” [“No-CB” = kein “backward-looking center” vorhanden] → score: 5.0/7 = 71% (vgl. Oberlander 2004, 217)



den Ansätzen zur thematischen Progression (Daneš 1970) und thematischen Organisation (Hoffmann 1997) eingenommen, wobei hier in erster Linie die Übergänge von je zwei Satzthemen in den Blick genommen werden.

Allen bisher besprochenen Arbeiten gemeinsam ist die Auffassung, dass das Thema ein Mittel zur Verknüpfung von Sätzen bzw. Diskurstteilen ist. Das Thema fungiert dabei in zweifacher Weise als Verknüpfungsmittel: Seine Funktion ist "connecting back and linking in to the previous discourse, maintaining a coherent point of view" (Brown und Yule 1983, 133). Givón (1992, 10) beschreibt diese zweifache Form der Verknüpfung als anaphorischen (Anknüpfung an Vorhergehendes) und kataphorischen Aspekt (Verweis auf Folgendes) des Themas, oder als "referential accessibility" und "thematic importance". Der Fokus dieser Arbeiten (sowie der Arbeiten von Daneš und Hoffmann) liegt also vorwiegend auf dem Aspekt der linearen Verknüpfungsfunktion des Themas, die auf der Textoberfläche verortet ist.

Neben den Arbeiten zur thematischen Kontinuität (Givón), thematischen Progression (Daneš) oder thematischen Organisation (Hoffmann) finden sich zahlreiche weitere Arbeiten, in denen das Thema als Gegenstand, über den im Text etwas gesagt wird, definiert wird (z.B. Bayer 1980; Fritz 1982; Lundquist 1989; Jürgens 1995; Brinker 1997; Schröder 2003; Dürr und Schlobinski 2006).<sup>51</sup> Ich werde im Folgenden nur noch auf drei dieser Ansätze eingehen (Bayer 1980; Fritz 1982, 1994; Jahr 1996, 2000), um einerseits exemplarisch die Probleme einer solchen Thema-Konzeption, andererseits die Bandbreite der Ausformungen dieser Konzeption darzustellen.

---

Das Beispiel wird demnach mit 71% als relativ kohärent beurteilt. Neben dieser Analyse der referenziellen Verbindungen zwischen aufeinander folgenden Segmenten zur Ermittlung der globalen Diskurskohärenz schlägt Oberlander in Anlehnung an Givón (1995) vor, Verbindungen auch nach räumlicher oder aspektbasierter Kontinuität zu analysieren. Eine solche Erweiterung ist meiner Ansicht nach auch dringend nötig, da referenzielle Bezüge zur Erklärung der Kohärenz eines Diskurssegments nicht ausreichen:

Beispiel (aus Asher (2004a, 181)): "(1) My car broke down. (2) Then the sun set (3) and I knew I was in trouble." → 3 Übergänge, 3 "No-CB's" = score: 0%

Nach Oberlanders bzw. Karamanis Ansatz ist dieses Segment nicht kohärent. Dies entspricht aber keineswegs den Tatsachen; jeder Leser wird den Ausschnitt ohne Probleme verstehen und als kohärent beurteilen. Woran jedoch liegt dies? Asher (2004a, 181f.) führt die Kohärenz dieses Beispiels auf das Vorliegen der Diskursrelation "Narration" zurück, wobei die Verbindung zwischen den Elementen der "Narration" durch "Causation" gestiftet wird: "then" signalisiert eine temporale Abfolge zwischen Segment (1) und (2), (1) und (2) wiederum können auf Basis des Weltwissens als Ursache für Segment (3) interpretiert werden. Das kurze Beispiel zeigt deshalb deutlich, dass das Satzthema, verstanden als referenzielle Entität, nicht ausreichend zur Erklärung der Kohärenz einer Folge von Diskurssegmenten ist. Auch umgekehrt lässt sich allerdings zeigen, dass ein Segment nicht unbedingt kohärent ist, nur weil es referenziell verknüpft ist, z.B.: <<Paul ist gestern nach Ankara geflogen. Ankara ist die Hauptstadt der Türkei. Die Türkei ist ein schönes Land. Ein Land kann zu Fuß, per Bahn, Auto oder Pferd bereist werden.>>

<sup>51</sup> Von diesen Arbeiten gehen zwei wie Daneš, Hoffmann und Givón auf die lineare Verknüpfung von Themen bzw. die thematische Kontinuität im Text ein: Dürr/Schlobinski (2006, S.252) beschreiben den „Topikfaden“ des Textes als Wiederaufnahmerelation zwischen den nominalen Satztopiks. Für Dik (1997, 422) ist wie für Givón thematische Kontinuität eine Folge dessen, dass "the topical entity will be referred to again and again by anaphorical means through a whole stretch of clauses".

Bayer (1980) versteht das Thema als referenzielles Konzept, d.h. als Gegenstand der Diskurswelt, auf den mittels eines sprachlichen Ausdrucks Bezug genommen wird. Er unterscheidet dabei nicht zwischen Satz- und Textthema, sondern betont im Gegenteil die ontologischen Gemeinsamkeiten, die es seiner Ansicht nach rechtfertigen, von einem einzigen Konzept Thema auszugehen, das sowohl satz- als auch textbezogen sein kann. Sein Thema-Konzept ist unabhängig davon, „ob sich ein kohärenter Diskurs in einem einzigen Satz oder in einer Folge von Sätzen manifestiert“ (Bayer 1980, 215). Bayer geht davon aus, dass es nicht das eine Thema eines Satzes oder einer Satzfolge gibt, sondern dass sich aus einem Satz oder Text multiple Themen ableiten lassen, die jeweils unterschiedlich deutlich ausgeprägt sind: „Man kann sagen, dass ein gegebener Satz oder Diskurs zumindest über all das ist, wofür es darin einen denotierenden Ausdruck gibt.“ (Bayer 1980, 216). Er gibt weiterhin zu bedenken, dass Äußerungen nicht „absolut“, sondern nur „bedingt“, über ein Thema sind, d.h. eine Äußerung ist „in höherem Maße“ über Thema A als über Thema B, es ist aber nicht „ausschließlich über A“ und auch nicht „überhaupt nicht über B“. Diese Annahme erklärt auch, warum Rezipienten teilweise divergierende Angaben darüber machen, was das konkrete Thema eines von ihnen rezipierten Textes ist. Das bedeutet, dass es in einem Text so viele potentielle Themen wie denotierende Ausdrücke (bzw. Diskursreferenten) gibt, was für die Ermittlung des Textthemas natürlich hoch problematisch ist. Zwar schlägt Bayer ein Verfahren zur schrittweisen Ermittlung des Themas aus der Menge aller nominalen Diskursreferenten vor, allerdings ist dieses sehr an der Textoberfläche verhaftet und ausschließlich auf die Bestimmung referenzidentischer Diskursentitäten ausgerichtet, so dass z.B. implizite Themabehandlungen (siehe z.B. Fritz 1982; 1994 - s.u.) bei der Ableitung des Themas nicht berücksichtigt werden.

Eine etwas andere Ausformung der Definition von „Thema als Gegenstand, über den etwas gesagt wird“, findet sich bei Fritz (1982, 211). Im Gegensatz zu Bayer bezieht Fritz die Möglichkeit mit ein, dass ein Thema auch über „indem-Zusammenhänge“ behandelt werden kann, dass also z.B. in einem Dialog das Thema „Energiesparen“ behandelt wird, *indem* über „Ölpreise“, „Heizungstypen“, „Atomkraftwerke“ etc. geredet wird, das eigentlich Thema „Energiesparen“ jedoch implizit bleibt. Offensichtlich drücken sich in den Ansätzen von Fritz und Bayer also divergente Auffassungen vom Thema als „Gegenstand“ aus. Während Bayer das Thema als Teil der Menge der im Text enthaltenen denotierenden Ausdrücke versteht, hat Fritz einen offeneren Gegenstands- und Thema-Begriff, insofern er die Möglichkeit einbezieht, dass ein Thema nicht sprachlich explizit im Text enthalten ist. In späteren Arbeiten führt er diese Auffassung noch weiter, indem er darlegt, „dass das Thema nicht irgendwie objektiv in den Dialogäußerungen enthalten ist, sondern dass es ein Aspekt des Verständnisses dieser Äußerungen ist.“ (Fritz 1994, 193).<sup>52</sup> Im Gegensatz zu allen bisher beschriebenen

---

<sup>52</sup> Eine handlungstheoretische Konkretisierung von Fritz Themenkonzeption findet sich in Schröders Definition des Themas als „Gegenstand, auf den mit einer Texthandlung Bezug genommen wird, auf den sich die Teil-Handlungen gemeinsam beziehen und dessen Teilaspekte in den zugehörigen Teil-Handlungen behandelt werden“ (Schröder 2003, 92). Das Textthema wird hier nicht als der *wichtigste* Gegenstand des Textes betrachtet (also nicht als der rekurrenteste Dis-

Arbeiten wird das Thema hier zum ersten Mal nicht mehr an einer expliziten sprachlichen Konstituente festgemacht. Sowohl Fritz als auch Bayer weisen außerdem darauf hin, dass ein Thema möglicherweise nicht „absolut“ zu bestimmen ist, insofern Themen graduell und nicht eindeutig distinkt sind: „Oft ist es uns unmöglich zu entscheiden, wo ein Thema aufhört und das andere anfängt.“ (Fritz 1982, 218).

Während in den bisherigen Ansätzen das Thema des Textes als Gegenstand definiert wurde, über den fortlaufend etwas gesagt wird (z.B. über *indem*-Zusammenhänge), nimmt Jahrs Ansatz (1996; 2000) in mehrfacher Weise eine Sonderstellung in diesem Kapitel ein. Ihre Konzeption des Themas ist zum einen eher ein Nebenprodukt der Beschäftigung mit explikativen Texten und als solches speziell auf diese bezogen, d.h. sie hat nicht den Anspruch, ein allgemeines Thema-Konzept zu entwickeln. Ihr Ansatz wird in diesem Kapitel dennoch aufgeführt, da er einen interessanten Übergang zwischen den Ansätzen darstellt, die das Thema über eine Analyse der Textoberfläche zu bestimmen versuchen, und den Ansätzen, die das Thema als abstrakten semantischen Kern auffassen, der nur über eine Analyse der Texttieferstruktur zu ermitteln ist. Jahr stellt die Hypothese auf, dass in explikativen Texten Fachtermini als „Schlüsselbegriffe“ fungieren und dass der zentrale Schlüsselbegriff das Thema des Textes repräsentiert (Jahr 1996, 128). Dieses das Thema repräsentierende Fachwort ist in explikativen Texten das „Explanans“. Von ihm ausgehend wird der Text thematisch entfaltet. Auch das „Explanandum“ besteht aus Fachwörtern bzw. Fachtermini. Diese stellen „das semantische Konzentrat eines mehr oder weniger umfangreichen Wissensbereichs“ dar (Jahr 2000, 392). Während das Explanans als Hauptthema beschrieben werden kann, entsprechen die zum Explanandum zu rechnenden Fachwörter Teilthemen, die mit dem Explanans (oder Textthema) in einer inhaltlichen Verknüpfung stehen, insofern sie semantisch verwandte „Merkmalbündel“, „Schlüsselbegriffe des Explanans“ oder „semantische Komponenten“ des Explanans sind, die sich aus dessen Valenz<sup>53</sup> ergeben (vgl. Jahr 2000, 393f.). Der Zusammenhang zwischen den Fachtermini (ob Explanandum oder Explanans) bzw. der thematische Textzusammenhang leitet sich dabei nach Jahr direkt aus dem „inneren logischen Zusammenhang der außersprachlichen Sachverhaltsstruktur“ ab; allerdings gilt dies nur für wissensvermittelnde Texte, da deren Intention die Vermittlung außersprachlicher Sachverhalte ist und hier die inhaltlich-propositionale Textstrukturierung dominiert (Jahr 2000, 388).

---

kursreferent), sondern als der Gegenstand, auf den alle zugehörigen Teil-Handlungen bzw. Teil-Themen sich gemeinsam beziehen. Ein solcher Gegenstand kann sowohl ein Ereignis, ein Sachverhalt, eine Äußerung, eine Frage als auch eine Person, ein Ort o.ä. sein. Im Gegensatz zu Daneš, Hoffmann und Givón verbinden jedoch weder Schröder noch Fritz mit dieser Bestimmung des Themas eine bestimmte sprachliche Realisierung, z.B. als Nominal- oder Präpositionalphrase.

<sup>53</sup> Diese Valenz wird als textbezogen und über das Konzept der semantischen Rollen näher bestimmbar beschrieben (vgl. Jahr 2000, 393f.).

### 2.2.2 Thema als abstrakter Kern des Textinhalts

Während in den vorigen Kapiteln vor allem lineare Themenentwicklungen (zwischen „Themen als Gegenständen“) im Fokus standen, wird in diesem Kapitel die hierarchische Textstruktur in den Blick genommen und das Thema als abstrakter semantischer Kern verstanden, der an der Spitze dieser hierarchischen Struktur, von der ausgehend der Text entfaltet wird, zu verorten ist. Die auf dieser Annahme basierenden semantisch orientierten Arbeiten zum „Thema als Kern des Textinhalts“ wurden hauptsächlich zwischen 1973 und 1983 im Kontext der von den Vertretern der Generativen Grammatik vorgeschlagenen Unterscheidung von Oberflächen- und Tiefenstruktur entwickelt. Das Konzept Thema wird in diesen Arbeiten als Teil einer impliziten Textbasis bzw. der Texttiefenstruktur betrachtet. Durch Entfaltungsprozesse wird aus einem thematischen Kern der Text abgeleitet, und umgekehrt kann durch Kondensation des Textes das thematische Substrat ermittelt werden. Der thematische Kern wird dabei entweder als „Makro“- oder „Basisproposition“ (van Dijk 1980; Metzelin und Jaksche 1983, s.u.) oder als Sachverhaltskomplex aus Prädikaten und Aktanten (Dressler 1973; Agricola 1976, 1979; Wulff 1979, s.u.) definiert. Interessant ist vor allem die Tatsache, dass z.B. nach van Dijk „einer bestimmten Makrostruktur im Prinzip eine unendlich große Zahl ‚konkreter‘ Texte ‚zugrunde liegen‘ kann. Eine Makrostruktur definiert eine Textmenge, nämlich alle Texte, die dieselbe *globale* Bedeutung haben.“ (van Dijk 1980, 49, Hervorhebung im Original). Bei der Ermittlung des thematischen Kerns eines Textes wird demnach von der spezifischen Komposition des Textes abstrahiert und stattdessen der Inhalt in seiner allgemeinsten Form in den Fokus genommen. Wie die Analyse dieses impliziten semantischen Themas operationalisiert werden kann, ist bis heute nicht zufrieden stellend geklärt worden. Zwei der wenigen ausgearbeiteten Analysemodelle zu diesem Thema-Konzept finden sich bei Agricola (1976; 1979) und van Dijk (1980).

Agricola definiert das Thema eines Textes als

begrifflichen Kern im Sinne der konzentrierten Abstraktion des gesamten Textinhaltes (...) in Form der verbal ausgedrückten, aber auch logisch-semantisch repräsentierbaren Struktur eines Sachverhaltskomplexes aus Prädikaten und Aktanten (Handlungsrollen). (Agricola 1976, 15)

Dieser begriffliche Kern lässt sich seiner Ansicht nach mittels bestimmter Analyseoperationen „in objektiven Schritten“ aus dem Oberflächentext extrahieren (Agricola 1979, 8). Das detailliert ausgearbeitete Verfahren, das bezogen ist auf Erzähltexte und Zeitungstexte, setzt sich dabei aus aufeinander aufbauenden „Stufen der Themenerschließung“ zusammen:

Zunächst müssen einige Vorverarbeitungen vorgenommen werden (1. Stufe), z.B. die „Normalisierung“ von Sätzen bzw. „Textemen“<sup>54</sup> und die Umwandlung in Prädikationen bzw. „semantische Primitivstrukturen“. Aus einem Satz wie „Viele Besucher besichtigen den original erhaltenen Thronsessel“ wird durch Umfor-

---

<sup>54</sup> Agricola (1979, 32) bezeichnet Sätze auch als „Texteme“, um deutlich zu machen, dass es sich dabei um Grundbausteine des Textes handelt.

mung die Prädikationenfolge „Besucher besichtigen Thronsessel. | Besucher sind viele. | Thronsessel ist erhalten. | Erhaltung ist original.“ (vgl. Agricola 1979, 46), die wiederum als Primitivstruktur  $\langle A a_1 B \rangle$ ,  $\langle A b_2 \rangle$ ,  $\langle B b_3 \rangle$ ,  $\langle C b_4 \rangle$  ausgedrückt werden könnte. Durch diese Notation der Prädikationen und semantischen Primitivstrukturen als Folgen aus nominalen und verbalen Elementen findet eine Kategorisierung der nominalen und verbalen Elemente statt (als Element A vs. B etc.), durch die bereits auf dieser Stufe die unterschiedlichen Diskursreferenten des Textes bzw. Aktanten<sup>55</sup> (Elemente aus der Menge {A, B, C, ...}) sowie die Relationen (Prädikate) zwischen diesen Aktanten (Elemente aus der Menge {a, b, c, ...}) identifiziert werden können.

Auf der zweiten Stufe werden Isotopiestruktur bzw. Isotopieketten<sup>56</sup> des Textes ermittelt. Isotopieketten, die deutlich mehr Kettenglieder haben als andere, werden als „Haupt-Isotopieketten“ bezeichnet, ihre Kettenglieder als „Textaktanten“. In einem nächsten Schritt werden die Querverbindungen zwischen den Isotopieketten über die Analyse der Abhängigkeiten zwischen den Kettengliedern ermittelt, um so Erkenntnisse zur isotopischen Verflechtung zweier Texteme zu gewinnen. Je mehr Verknüpfungen ein Textem zu anderen Textemen aufweist, desto wichtiger ist es für den Gesamttext. Texteme, die nur Verbindungen zu ein oder zwei anderen Textemen aufweisen, sind dagegen für den Gesamttext unwichtiger.<sup>57</sup>

Gegenstand der dritten Analysestufe ist die Ermittlung der „Konstellations- und Konnektorenstruktur“. Zur Bestimmung der Konstellationsstruktur werden die Prädikat-Textaktant-Verbindungen über Reduktions- und Abstraktionsprozesse zu einem „Netz der generellen, ständig wiederkehrenden Grundrelationen zwischen den wesentlichen, das Thema prägenden, durchgängigen Begriffen (Textaktanten)“ komprimiert (Agricola 1979, 54). Daneben werden die „logischen, quasilogischen, kausalen, temporalen u.ä. Verhältnisse von mehr oder weniger deutlich hierarchischer Art, wie sie zwischen den vom Text beschriebenen Sach-

<sup>55</sup> Aktanten sind die „Hauptpartizipanten“ der Sachverhalte, die im Ablauf wiederkehrenden Objekte. Textaktanten sind dagegen die Glieder der „Haupt-Isotopieketten“ (Isotopieketten, die deutlich mehr Kettenglieder haben als andere), die wichtigsten Inhaltselemente (vgl. Agricola 1979, 41).

<sup>56</sup> Der Begriff „Isotopie“ wurde zunächst von Greimas eingeführt (vgl. Heinemann 2000a). Isotopieketten sind Ketten von Lexemen desselben Bedeutungsbereichs, die Agricola zufolge realisiert werden in Form von „Rekurrenz semantischer Merkmale bei Sememrepetition oder bei Äquivalenz im weitesten Sinne (Synonymie, Hyper-Hyponymie-Beziehung, Antonymie, Paraphrasierung, funktionale semantische Äquivalenz, d.h. ad hoc vollzogene Einordnung der Isotopiepartner unter einen gemeinsamen, einheitlichen Gesichtspunkt), sowie als mittelbarer Ausdruck solcher Relationen durch Pro-Formen, ferner als Ketten von Zeitstrukturangaben, von Koreferenzbezügen zur Identifizierung gleicher und zur Differenzierung unterschiedlicher Denotate, und schließlich als wechselnde Abfolge der kommunikativ erforderlichen Thema-Rhema-Progression (Ausdruck der funktionalen Satzperspektive).“ (Agricola 1979, 33).

<sup>57</sup> In ähnlicher Weise untersuchen Hasan (1984) und Hoey (1991) die Interaktionen zwischen lexikalischen Ketten (siehe Kapitel 2.6.3.2). Durch eine differenzierte Analyse des Grades der Verflechtung zwischen Sätzen versucht Hoey wie Agricola zwischen marginalen und zentralen Sätzen zu unterscheiden. Während der semantische Ansatz von Agricola in dieser Arbeit nicht weiter verfolgt wird, werden in Kapitel 4 stattdessen lexikalische Ketten und Interaktionen zwischen lexikalischen Ketten als Reflexionen der thematischen Struktur aufgefasst.

verhaltenskomplexen bestehen“ analysiert, um so die Konnektorenstruktur zu bestimmen (Agricola 1979, 54). Diese Sachverhaltsverknüpfungen werden nur zum Teil sprachlich signalisiert (z.B. durch Konjunktionen oder Adverbien). Häufig setzt ihre Bestimmung aber Interpretationsprozesse voraus. Das Ergebnis der beiden Analysen sind zwei Graphen, die aus Knoten und Kanten bestehen: Der Graph der Konstellationsstruktur verknüpft Textaktanten über Prädikate<sup>58</sup>, der Graph der Konnektorenstruktur Texteme über Konnektoren. In einem letzten Schritt müssen diese beiden Strukturen nun unifiziert werden. Wie dieser Prozess aber genau ablaufen soll, bleibt offen.<sup>59</sup>

Agricolas Analysemodell wurde in der Forschung zum Thema selten in dem Umfang gewürdigt, wie es das meiner Ansicht nach verdient hätte. Häufig wurden nur die theoretischen Grundlagen seiner Arbeit rekapituliert und diese dann der gleichen Kritik unterworfen wie andere semantische Thema-Konzepte. Eine ausführliche Auseinandersetzung mit Agricolas Vorschlag zur Operationalisierung der Themenanalyse findet dagegen meines Wissens nicht statt.<sup>60</sup> Auch wenn Agricolas Modell insbesondere im Hinblick auf die Integration der vielen aufeinander aufbauenden und sich teilweise sogar überschneidenden bzw. doppelnden Einzelanalysen in eine Gesamtanalyse Schwächen aufweist (so ist z.B. die Unifizierung von Konnektoren- und Konstellationsstruktur zu wenig konkret ausgeführt), erscheint es mir dennoch als eines der wenigen Analysemodelle zum Thema, die sich um eine Operationalisierung der Analyse bemühen.

Ein zweites, ausführlich ausgearbeitetes Verfahren zur Analyse des Themas als implizitem semantischem Textkern, das zugleich noch häufiger rezipiert (und kritisiert<sup>61</sup>) wurde als das Verfahren von Agricola, stammt von van Dijk (1980, 45ff.). Sein Verfahren basiert auf der Anwendung so genannter „Makroregeln“ – „Auslassen“, „Selektieren“, „Generalisieren“, „Konstruieren oder Integrieren“ –, mittels derer aus dem Text eine Menge von Makropropositionen abgeleitet wird, die die globale Textbedeutung repräsentiert und gleichzeitig als Zusammenfas-

---

<sup>58</sup> Agricola deutet an, dass diese Beziehung auch mittels des Konzepts der semantische Kasusrollen beschrieben werden kann, da auf diese Weise auch die Abhängigkeiten der Aktanten zu einem gemeinsamen Relator (der mehrere Argumentstellen eröffnet) erklärt werden könne (vgl. Agricola 1979, 57f.).

<sup>59</sup> Die Konstellationsstruktur weist Ähnlichkeiten mit der Sachverhaltsstruktur auf, wie sie in Kapitel 3.2.1 vorgestellt wird, während die Konnektorenstruktur Parallelen zur rhetorischen Struktur bzw. relationalen Textstruktur aufweist (vgl. Kapitel 2.6.3.4). Während in dieser Arbeit nicht weiter mit Agricolas Begrifflichkeiten operiert wird, gehen die von ihm bearbeiteten sprachlichen Phänomene in die in Kapitel 4 beschriebene Operationalisierung der Analyse des Themas mit ein.

<sup>60</sup> Ein typisches Beispiel für die Auseinandersetzung mit Agricola findet sich in Brinker (1997, 55f., Fußnote 64): „Von daher ergeben sich grundsätzliche Bedenken gegen den Versuch Agricolas (1979), ein Modell zur Analyse des Textinhalts zu konstruieren, das es ermöglichen soll, „in objektiven Schritten von der Oberflächenstruktur“ eines Textes zu dessen thematischem Kern („Informationskern“) zu gelangen (...). (...) Agricola entwickelt einen komplizierten formalen Apparat zur modellhaften Beschreibung dieser Prozesse [der stufenweisen Erschließung des Themas; M.B.]. Wir können darauf in unserem Zusammenhang nicht weiter eingehen.“

<sup>61</sup> Zur Kritik an van Dijks Modell der Makrostruktur sowie der Methode der Makroregeln siehe (Gülich et al. 1977; Brown und Yule 1983; Hellwig 1984a; Lötscher 1987; Brinker 1997; Vater 2001; Gansel und Jürgens 2002; Schröder 2003)

sung des Inhalts fungiert. Das Thema wird in diesem Modell als die hierarchisch am Höchsten stehende Makroproposition definiert.

Die Makrostruktur eines Textes stellt die Ebene der Entfaltung des Themas dar. Themen werden van Dijk zufolge in Form von Makropropositionen entfaltet, die sich aus „Sachverhalten, die größere Einheiten bilden“, konstituieren. Bei einem Thema „X fährt zum Wintersport“ könnten die konstituierenden Teile des Sachverhaltskomplexes beispielsweise die „Komponenten der Reise wie Ankunft und Aufenthalt“, „vorbereitende Handlungen“, „Konsequenzen“ etc. sein (van Dijk 1980, 55). Der Zusammenhang zwischen den Teilen des Sachverhaltskomplexes kann dabei funktional oder inhaltlich motiviert sein – die Teile des Sachverhalts gehören zu einem gemeinsamen kognitiven Schema bzw. „Wissensrahmen“ (van Dijk 1980, 184). Die Makrostruktur (oder thematische Struktur) wird dementsprechend als „Abbildung bestimmter Strukturen in der Wirklichkeit“ (van Dijk 1980, 152) beschrieben, ihr Verstehen setzt das Verstehen der Welt voraus (vgl. van Dijk 1980, 185).<sup>62</sup>

Van Dijk demonstriert die Anwendung der Makroregeln anhand von kurzen Beispielsequenzen und anhand eines kurzen Beispieltextes. Selbst bei diesen sehr einfachen Analysen zeigt sich die große Schwierigkeit des vorgeschlagenen Verfahrens: Die Makroregeln können unterschiedlich angewandt werden, sie haben eher den Status von Leitprinzipien für die Textzusammenfassung, so dass mehrere, konkurrierende Makrostrukturen und Themenangaben aus der Anwendung der Regeln ableitbar sind (vgl. van Dijk 1985, 75). Van Dijk räumt dieses Problem selbst ein und beschreibt die Analyse deshalb als „lediglich semi-formal“. Er führt dies darauf zurück, dass jeder Leser andere kognitive Einstellungen, Interessen und Wissensbestände habe, so dass die ermittelten Themen gar nicht intersubjektiv identisch sein könnten, dass also das Ziel einer eindeutigen Themenbestimmung unerreichbar sei (vgl. van Dijk 1980, 54f.). Diese Auffassung ist keineswegs singulär, sondern wird in vielen kommunikativ-pragmatisch orientierten Arbeiten vertreten, kollidiert in van Dijks Arbeit jedoch mit dem starken Stellenwert, den er den semantischen Transformationsregeln zuweist und deren Anwendung er durch mehrere Beispielanalysen demonstriert. Van Dijks Verfahren lässt sich meiner Ansicht nach lediglich als Versuch lesen, eine regelhafte Grundlage für die Aufgabe der Textzusammenfassung zu liefern. Im Hinblick auf die zu große Offenheit und Interpretierbarkeit der Regeln erweist sich dieser Versuch jedoch als unzureichend. Warum seine Arbeit dennoch bis heute so populär ist und in jeder „Einführung in die Textlinguistik“ behandelt wird, liegt sicherlich an der intuitiven Plausibilität des Verfahrens. Um eine operationalisierte Methode handelt es sich dabei jedoch nicht.

Sowohl van Dijks als auch Agricolas Analysemodell erscheinen jedoch noch aus einem weiteren Grund als unzureichend. In beiden Arbeiten geht es darum, durch Analysen den zentralen inhaltlichen Kern des Textes zu ermitteln und die semantische Struktur zu kondensieren auf das inhaltlich Wichtige. Was inhalt-

---

<sup>62</sup> Im Hinblick auf den Zusammenhang von Makrostruktur und kognitiver Sachverhaltsstruktur lässt sich van Dijks Ansatz auch als kognitiv fundiert beschreiben.

lich wichtig ist, hängt jedoch nicht zuletzt von weiteren Faktoren als nur der semantischen Struktur ab, z.B. von der Funktion des Textes bzw. der Intention des Produzenten, dem Kontext, den Interessen und Wissensbeständen des Rezipienten etc. Eine Analyse, die diese Faktoren nicht berücksichtigt, kann keine hinreichend begründete Unterscheidung der *kommunikativ-inhaltlich* zentralen von weniger zentralen Propositionen oder Segmenten des Textes treffen.

### 2.2.3 Fazit

Die semantischen Thema-Konzepte unterscheiden sich zum einen im Hinblick auf die Reichweite ihres Thema-Begriffs: satz- vs. textbezogene Thema-Konzepte sowie Thema-Definitionen, die keine Differenzierung zwischen Satz- und Textthema vornehmen, sondern diese als ontologisch gleichartige Konzepte auffassen. Zum anderen lassen sich die Definitionen inhaltlich kategorisieren in Definitionen, in denen das Thema als Gegenstand definiert wird (dabei überwiegt die Bestimmung des „Gegenstands“ als Diskursreferent), und Definitionen, in denen das Thema als semantischer Kern gefasst wird. Einen dritten Unterscheidungspunkt stellt die Definition der thematischen Struktur dar: In einigen Arbeiten wurde diese als lineare, sequenzielle Struktur beschrieben, in anderen als hierarchische.

Sowohl die Ansätze zum Thema als Gegenstand als auch die zum Thema als semantischem Kern weisen eine intuitive Plausibilität auf, die sich aus der Nähe zum alltagssprachlichen Themenverständnis ableitet.<sup>63</sup> Allerdings weisen diese Arbeiten einige Mängel auf: Die Definition des Themas ist in den meisten Arbeiten zu unscharf, um z.B. eine eindeutige Unterscheidung von thematischen vs. nicht-thematischen Gegenständen bzw. Makropropositionen bzw. Prädikat-Aktanten-Strukturen zu erlauben. Zur eindeutigen Eingrenzung des Themas im Satz oder Text sind diese Definitionen nicht ausreichend. In den Ansätzen zum Thema als Gegenstand auf Satzebene werden deshalb weitere nicht semantische Merkmale – in erster Linie grammatische (Satzposition, grammatische Rolle) und teilweise kognitive Merkmale („Bekanntheit“ – siehe Kapitel 2.4) – zur Bestimmung des Themas verwendet. Die Arbeiten zur „thematischen Progression“, „thematischen Organisation“ oder „Topic Continuity“ basieren auf satzbezogenen Thema-Definitionen, da die lineare thematische Entwicklung aufeinander folgender Satzthemen betrachtet wird – insofern unterliegen sie der gleichen Problematik wie die semantischen satzbezogenen Thema-Konzepte. Andere Auffassungen vom Thema, in denen es als zentraler Gegenstand, über den etwas im Text gesagt wird, definiert wird, haben ebenfalls mit der Frage zu kämpfen, wie die thematischen von den nicht-thematischen Gegenständen unterschieden

---

<sup>63</sup> Typische Verwendungen von Thema, wie sie in Kapitel 1 genannt wurden, sind z.B. „über ein Thema reden“, „eine Debatte über ein Thema führen“. In diesen Verwendungsweisen steht der „Aboutness“-Aspekt des Themas im Vordergrund: Das Thema wird als Gegenstand verstanden, über den etwas gesagt wird. Auf der anderen Seite finden sich aber auch Verwendungsweisen von Thema, die die Konzeption vom „Thema als semantischem Kern“ stützen, insofern das Thema als inhaltliches Zentrum eines Diskurses festgelegt wird: „ein Vortrag zum Thema X“, „Informationen zum Thema X“.



werden können – und ob thematische Gegenstände in Form von Diskursreferenten im Text selbst überhaupt explizit vorkommen müssen oder nicht, d.h. ob ein Thema als Gegenstand eines Diskurses nicht auch implizit behandelt werden kann. In den Arbeiten, die das Thema als expliziten Textgegenstand auffassen, wird häufig das Kriterium der Rekurrenz (von Diskursreferenten, Lexemen oder Semen) angesetzt, um zentrale Themen zu bestimmen. Thematisität wird hier demnach vor allem als referenzielle, lexikalische oder semantische Rekurrenz betrachtet. Der kommunikativen Funktion, die das Thema für den Diskurs hat<sup>64</sup>, wird dagegen in kaum einer der semantischen Thema-Konzeptionen Rechnung getragen.

Auch im Hinblick auf die für die Themenanalyse entwickelte Methodik erweisen sich die semantischen Ansätze als problematisch. Die Probleme reichen von der mangelnden Operationalisierbarkeit der Analyse des Themas (z.B. bei der Ermittlung der Makropropositionen) bis zur ungelösten Aufgabe, wie ausgehend von der Analyse der Satzthemen das Textthema zu bestimmen sein kann (vor allem in den Ansätzen in der Tradition der thematischen Progression).

### 2.3 Thema als kommunikativ-pragmatisches Konzept

“It is speakers and writers who have topics, not texts.” Dieses Zitat von Brown/Yule (1983, 68) gibt die grundlegende Annahme aller in diesem Abschnitt behandelten Ansätze wider. Das Thema eines Textes lässt sich demnach nur als sprachlicher Akt des Thematisierens adäquat beschreiben und analysieren. Der Akt des Thematisierens durch einen Sprecher<sup>65</sup> (oder alternativ: der kollaborativen Themenentwicklung mehrerer Sprecher) wurde zunächst vor allem in Arbeiten zur Konversations- bzw. Dialoganalyse in den Blick genommen (Keenan und Schieffelin 1976; Brown und Yule 1983; Yule und Mathis 1992; Fritz 1994), später aber auch auf Texte übertragen (Hellwig 1984a; Rothkegel 1984b; Lötscher 1987; van Kuppevelt 1995; Hoffmann 1997; Rothkegel 1997). Eine relativ große Gruppe von Ansätzen setzt das Thema gleich mit einer impliziten Frage, die sich der Sprecher stellt und mit einer Äußerung, einem Text oder Textabschnitt zu beantworten sucht, oder beschreibt das Thema als Problemstellung, die der Produzent durch den Text bearbeitet. Das Thema wird also als Ausgangspunkt der Textproduktion betrachtet. Eine weitere große Gruppe kommunikativ-pragmatischer Textthema-Konzepte nimmt vor allem das Thematisieren als sprachliche Handlung in den Blick. Ziel dieser Forschungsarbeiten ist u.a. die Ermittlung von Regeln für thematische Übergänge bzw. Themenwechsel. Darüber hinaus wird die Funktion des Themas für das Verstehen von Texten unter-

---

<sup>64</sup> Wie noch gezeigt werden wird und wie auch aus den Hauptverwendungsweisen von Thema ersichtlich ist, ist das Thema auch ein zentrales gesprächs- und textstrukturierendes Mittel: „ein Thema diskutieren“, „sich zu einem Thema äußern“, „ein Thema behandeln“, „ein Thema wechseln“, „vom Thema abschweifen“, „ein Thema fallen lassen“ etc.

<sup>65</sup> In diesem Kapitel soll auch von „Sprechern“ und „Hörern“ die Rede sein (und nicht nur von „Produzent“ und „Rezipient“), da die meisten der hier dargestellten Arbeiten aus der Dialog- oder Gesprächsforschung stammen.

sucht – es wird als „Verstehenshorizont“ (Fritz 1994, 193) bzw. „Einordnungsinstanz für das Verstehen“ (Hoffmann 1997, 513) aufgefasst. Hoffmann geht beispielsweise davon aus, dass „Themen das Grundgerüst der inhaltlichen Organisation von Texten und Diskursen [bilden; M.B.], an dem sich Verstehensprozesse in elementarer Weise festmachen“ (Hoffmann 1997, 509). In ähnlicher Weise spricht auch Schröder (2003, 101) davon, dass das Textthema eine besondere Rolle beim Textverstehen spielt, insofern es während des gesamten Rezeptionsprozesses im aktuellen Speicher präsent zu sein scheint.

Als Konzept auf Satzebene wird dem Thema bzw. dem Thematisieren meistens die Funktion der Strukturierung und Perspektivierung von Informationen zugeschrieben, d.h. es wird als Mittel betrachtet, über das der Sprecher dem Rezipienten signalisiert, welche Elemente des Satzes Thema bzw. Topic oder welche Rhema bzw. Comment sind.<sup>66</sup> Ansätze, in denen das Satzthema als kommunikativ-pragmatisches Konzept begriffen wird, wurden zum Teil schon in den vorherigen Kapiteln behandelt, insofern die Thema-Konzeptionen in der Tradition der Prager Schule und der Systemisch Funktionalen Linguistik grundsätzlich als funktional orientierte Ansätze zu verstehen sind. In diesem Kapitel wird deshalb nur noch ein weiterer Ansatz zum Satzthema vorgestellt (das „Modell der kommunikativen Dynamik“, vgl. Firbas 1992), ansonsten beschäftige ich mich ausschließlich mit Themen auf Text- bzw. Diskursebene.

### 2.3.1 Thema als das kommunikativ am wenigsten dynamische Element des Satzes

Die Theorie der „kommunikativen Dynamik“ („communicative dynamism“ – CD) betont die unterschiedliche kommunikative Dynamik oder Wichtigkeit („communicative importance“)<sup>67</sup> der Elemente des Satzes für den weiteren Verlauf des Diskurses. Diese kommunikative Wichtigkeit bestimmt sich durch den Beitrag, den ein Element zur Erreichung eines kommunikativen Ziels leistet. Sie lässt sich auch als der relative kommunikative Informationswert eines Elements beschreiben.<sup>68</sup>

In the act of communication some elements are more or less dynamic. This induces me to speak of communicative dynamism (CD), a phenomenon constantly displayed by linguistic elements in the act of communication. It is an inherent quality of communication and manifests itself in constant development towards the attainment of a communicative goal; in other words, towards the fulfilment of a communicative purpose. Participating in this development, a linguistic element assumes some position in it and in accordance with this position displays a degree of communicative dynamism. (Firbas 1992, 7)

---

<sup>66</sup> In diesem Kontext spielen Konzepte wie Vordergrund/ Hintergrund, Topik/ Kommentar, Fokus/ Hintergrund, Haupt-/ Nebeninformation eine Rolle. Die Thematisierung dieser Zusammenhänge erfolgt in Kapitel 2.6.3.1.

<sup>67</sup> Firbas verwendet die Begriffe „communicative importance“ und „communicative dynamism“ synonym. Beide Begriffe beziehen sich auf den Status eines Elements im Hinblick auf die Erfüllung des kommunikativen Zwecks (vgl. Firbas 1992, 105).

<sup>68</sup> Der Grad der kommunikativen Dynamik wird als relativer Informationswert für die weitere Entwicklung des Diskurses definiert (vgl. Firbas 1992, 105).

Im Gegensatz zur Thema-Rhema-Forschung wird hier davon ausgegangen, dass ein Satz nicht eine Zweiteilung in thematische und rhematische Elemente aufweist, sondern dass sich Elemente graduell im Hinblick auf ihre kommunikative Dynamik unterscheiden: Ein Satz besitzt nicht nur ein Thema, sondern es kann mehrere thematische (und rhematische) Elemente aufweisen. Darüber hinaus handelt es sich bei CD um eine relative Eigenschaft von Elementen – sie ist bezogen auf den Grad der CD aller anderen Elemente. Firbas geht davon aus, dass der CD-Grad im Satz kontinuierlich ansteigt und bei dem letzten Satzelement am höchsten ist. Durch diese graduelle Steigerung der kommunikativen Dynamik erfolgt eine Perspektivierung des Satzes, dessen Zentrum das letzte, kommunikativ dynamischste Satzelement bildet.

Zur Ermittlung des Grades der kommunikativen Dynamik von Elementen müssen verschiedene Analyseaspekte einbezogen werden: „Isoliert betrachtet, können die einzelnen sprachlichen Mittel den Grad an kommunikativer Dynamik eines Elements nicht angeben.“ (Gülich et al. 1977, 65). Stattdessen handelt es sich um ein Zusammenspiel vieler Faktoren (“interplay of factors”, vgl. Firbas 1992, 106). Ein wichtiger Faktor ist die Kontextabhängigkeit bzw. -unabhängigkeit (vergleichbar mit einer engen Auffassung der Konzepte „bekannt“/„neu“, siehe Kapitel 2.4). Als kommunikativ besonders dynamische Elemente gelten nach Firbas diejenigen, die sich nicht auf eine bereits eingeführte Information beziehen, d.h. kontextunabhängige, neue Informationen. Im unmittelbar vorausgehenden Diskurskontext genannte Informationen sind dagegen kommunikativ weniger dynamisch (vgl. Firbas 1992, 6). Ein weiteres Kriterium für die Bestimmung der kommunikativen Dynamik von Elementen ist die Wortstellung. Firbas (1992, 8f.) nimmt an, dass das Element mit der höchsten Dynamik am Ende des Satzes steht, da es von dieser Stelle aus die Perspektivierung der Kommunikation lenkt. Ein dritter Faktor, der Aufschluss über den Grad der kommunikativen Dynamik gibt, ist ein semantischer. Elementen kann mit Blick auf ihre semantische Eigenschaft bzw. ihre Rolle<sup>69</sup> im Satz ein relativer CD-Grad zugewiesen werden: Elemente, die einen niedrigen CD-Grad haben, drücken das „Setting“ des Satzes aus (z.B. Adverbien), während den höchsten CD-Grad die Elemente aufweisen, die die Perspektive der Kommunikation ausrichten auf a) das „Phänomen“, das durch das Subjekt präsentiert wird, oder b) die „Qualität“, die diesem Phänomen zugeschrieben wird (vgl. Firbas 1992, 66f.). Diese perspektivierenden Elemente sind nach Firbas die Verben.

Das Thema wird von Firbas (1992, 72) als das Element mit der geringsten kommunikativen Dynamik definiert, oder anders gesagt: Der Status Thema ist eine Folge der Eigenschaft eines Elements, einen niedrigen Grad an kommunikativer Dynamik zu besitzen. Insofern als Elemente mit einer geringen kommunikativen Dynamik diejenigen am Anfang des Satzes sowie diejenigen, die sich auf bekannte Informationen beziehen, verstanden werden, weist die Auffassung vom „The-

---

<sup>69</sup> Firbas (1992, 41) ist bei der Beschreibung des semantischen Faktors (“semantic factor”) sehr ungenau und fasst darunter sowohl semantische Eigenschaften als auch semantische Relationen eines Elements: “What it actually covers is (...) the semantic character of a linguistic element, as well as the character of its semantic relations”.

ma als Element mit der geringsten kommunikativen Dynamik“ deutliche Korrespondenzen mit bereits behandelten Thema-Konzepten in der Tradition der Prager Schule auf.<sup>70</sup> Im Gegensatz zu diesen Ansätzen wird Thematizität bei Firbas allerdings nicht als bipolares Konzept aufgefasst, sondern als graduelle und relative Eigenschaft, die in unterschiedlichen Ausprägungsgraden mehreren Elementen im Satz zukommen kann. Problematisch wird diese Auffassung dann, wenn man über die Satzthemen auch zu den Textthemen gelangen möchte – im Gegensatz zu den bipolaren Auffassungen von Thema- bzw. Rhematizität erscheint eine Ausweitung dieses Thema-Konzepts vom Satz auf den Text schwer vorstellbar.

### 2.3.2 Thema als Fragliches

Seit den 70er Jahren wird das Thema immer wieder auch als implizite „Frage“ oder „Fragliches“ definiert, die/das der Produzent durch den Text zu beantworten bzw. bearbeiten versucht (siehe Keenan und Schieffelin 1976; Hellwig 1984a, b; von Stutterheim 1994; van Kuppevelt 1995; Büring 1997). Prägend für diesen Forschungsansatz sind die ethnomethodologischen Untersuchungen von Keenan/Schieffelin (1976) zur thematischen Analyse gesprochener Dialoge zwischen Kindern und Eltern. Sie unterscheiden zwei Hauptformen thematischer Organisation (vgl. Keenan und Schieffelin 1976, 342):

- A. Kontinuierliches Diskursthema – hierunter fallen “Collaborating Discourse Topic” (vergleichbar mit Daneš Progression mit durchlaufendem Thema) und “Incorporating Discourse Topic” (vergleichbar mit Daneš einfacher linearer Progression);
- B. Diskontinuierliches Diskursthema – hierzu gehören “Re-introducing Discourse Topic” (Wiederaufnahme eines bereits behandelten Themas) und “Introducing Discourse Topic” (Neueinführung eines Themas).

Im Unterschied zu Daneš allerdings erklären sie die thematische Organisation von Diskursen als Ergebnis der Zusammenarbeit der Dialogpartner: Der Hörer muss sich die Frage stellen: “Why did the speaker say what he did?”; “What is the speaker informing me of?” (Keenan und Schieffelin 1976, 343). Der Beitrag des Sprechers kann dabei als Antwort auf eine implizite Frage (“question of immediate concern”) aufgefasst werden. Nur wenn dem Hörer klar ist, welche Frage der Sprecher mit seiner Äußerung zu beantworten sucht, wird der thematische Zusammenhang mit dem vorangegangenen Diskurs oder dem Diskurskontext erkennbar. Das Diskursthema wird in diesem Kontext wie folgt definiert:

The discourse topic is based on the question of immediate concern. It is the proposition or set of propositions that the question of immediate concern presupposes. (...) The discourse topic is the primary presupposition of this question. (Keenan und Schieffelin 1976, 344).

---

<sup>70</sup> Allerdings hat Firbas einen relativeren Thema-Begriff: das kontextgebundene Element ist zwar immer thematisch, umgekehrt kann ein kontextunabhängiges Element aber sowohl nicht-thematisch als auch thematisch sein (vgl. Firbas 1992, 72).

Das Diskursthema wird hier also als die Proposition definiert, die den präsupponierten Teil einer impliziten Frage darstellt, z.B. wäre die Präsupposition (d.h. das Diskursthema) der Frage ("question of immediate concern") „Was brauchen wir für das Backen des Kuchens?“ „Wir brauchen etwas für das Backen des Kuchens“. Das Thema selbst wird demnach zwar nicht als Frage definiert, leitet sich aber direkt aus dieser ab.

In den folgenden Jahren wurden weitere Variationen dieser Thema-Konzeption entwickelt, wobei das Thema entweder in der Tradition von Keenan/Schieffelin als präsupponierter Antwortteil auf eine implizite Frage aufgefasst wird (van Kuppevelt 1993, 1995), oder als die Frage selbst, die der Sprecher mit einer Äußerung (oder dem Text als Ganzem) zu beantworten sucht (Hellwig 1984a, b).

Hellwig (1984a, 65) begründet die Gleichsetzung von Thema und Fragestellung mit der Funktion beider Konzepte, „sequenzstiftend“ zu sein und so den Aufbau des Textes zu leiten. Deutlicher als Keenan/Schieffelin formuliert Hellwig ein Modell des Textzusammenhangs, das darauf beruht, dass ein Produzent sich beim Verfassen eines Textes mögliche Fragen eines Lesers vor Augen führt – und zwar an jeder Stelle des Textes, d.h. kontextgebunden – und versucht, Antworten auf diese Fragen zu geben. Das Thema als das „Fragliche“ ist also zum einen rückgebunden an den vorangegangenen Diskurs, zum anderen steuert es den Inhalt der nächsten Äußerung bzw. eines folgenden, inhaltlich kohärenten Abschnitts. Das Thema als „Fragliches“ wird dabei als sequenzstiftendes Mittel betrachtet, das Äußerungen oder größere Abschnitte eines Textes zusammenhält. Der Zusammenhang zwischen den einzelnen Fragen, d.h. die thematische Struktur, wird nicht als hierarchische Struktur betrachtet, sondern als sequenzielle.

Man sieht, die thematische Struktur eines Textes verträgt sich nicht mit dem Konstituentenstrukturprinzip. Sie besteht nicht darin, wie man vielleicht annehmen könnte, dass das Gesamthema des Textes in Unterthemen zu zerlegen ist, diese wiederum in Teilthemen, und so fort (...). Der springende Punkt ist, dass sich nachgeordnete Themen nicht ohne Rückgriff auf vorgängige Ausführungen einordnen lassen (...). (Hellwig 1984a, 68)

Unklar an Hellwigs Modell ist, wie groß die Abschnitte sind, die als Antwort auf eine Frage fungieren, bzw. andersherum, wie feinkörnig die Fragen sind: Werden Fragen so formuliert, dass nur der nächste Satz als Antwort auf sie zu verstehen ist, oder sollen Fragen für einen ganzen Abschnitt formuliert werden? Und wenn letzteres der Fall ist, wie gelangt man in möglichst objektiven Schritten zur Formulierung einer solchen „Makro-Frage“?

Ein stärker formalisiertes Modell zum Konzept „Thema als Fragliches“, das deutlicher zwischen Teilthemen und Diskursthemen unterscheidet und auf ihren Zusammenhang eingeht, findet sich bei van Kuppevelt (1993; 1995). Wie Hellwig betrachtet van Kuppevelt die Textstruktur – ebenso wie die thematische Struktur – als Ergebnis von impliziten Frage-Antwort-Abfolgen. Allerdings bestimmt er diese Struktur anders als Hellwig als hierarchisch und setzt voraus, dass die hierarchische Textstruktur mit der Hierarchie der Themen im Text korrespondiert. In diesem hierarchischen Thema-Modell werden Sub-Themen ("Subtopics") als Teile von Themen ("Topics") und diese wiederum als Teile von Diskursthemen

(“Discourse Topics”) verstanden (vgl. van Kuppevelt 1995, 140). Alle Typen von Themen werden als von einer impliziten Frage determiniert aufgefasst. Für Satzthemen gilt, dass sie den Teil der Antwort darstellen, der durch die Frage prä-supponiert wird; der Teil der Antwort, der die eigentliche Antwort enthält, wird dagegen als “Comment” bezeichnet. Jeder Satz enthält demnach eine Topic-Comment-Struktur, die durch die implizite Frage vorgegeben wird. Dieser Zusammenhang von Frage und Topic-Comment-Gliederung wird bereits von Daneš (1970, 592) und Sgall et al. (1973, 49ff.) beschrieben, die zur Bestimmung von Thema und Rhema bzw. Topic und Comment den sogenannten „Fragetest“ verwenden: Der an die Frage gebundene Teil der Antwort (“bound segment”) gilt als das Thema bzw. Topic des Satzes, der nicht in der Frage enthaltene Antwortteil bzw. die Ergänzung der Frage das Rhema bzw. Comment. Im Gegensatz zu den Arbeiten zur Funktionalen Satzperspektive versucht van Kuppevelt jedoch ein Modell zur Themenanalyse zu entwickeln, das sowohl Satz- als auch Diskursthemen umfasst. In diesem Modell sind nicht nur Satzthemen (Sub-Topics) durch implizite Fragen determiniert, sondern auch die Menge aller im Diskurs behandelten Themen (Topics) – erstere durch “Subtopic-Constituting Sub-questions”, letztere durch “Topic-Constituting Questions” (van Kuppevelt 1995, 110). Insofern Sub-Topic-Constituting Questions kontextuell gebunden sind, sind sie abhängig vom vorangegangenen Diskurs (dies entspricht dem Status von Fragen bei Hellwig). Topic-Constituting Questions sind dagegen unabhängig vom Diskurskontext und unabhängig von anderen Fragen. Ihre einzige Beschränkung ist die, dass sie sich von dem sogenannten “Feeder” ableiten lassen müssen. Ein Feeder ist ein linguistisches oder nicht linguistisches Ereignis, das am Anfang eines Diskurses oder Diskurssegments steht. Ein nicht linguistischer Feeder kann beispielsweise ein Ereignis wie ein Geräusch sein, auf das ein Diskursteilnehmer mit einer Äußerung reagiert; ein linguistischer Feeder ist z.B. ein thematisch und kontextuell ungebundener Einleitungssatz (oder auch die Überschrift) eines Textes oder eines Dialogs, von der ausgehend sich der Text in Frage-Antwort-Sequenzen entwickelt (vgl. van Kuppevelt 1995, 119). Ein Diskursthema (Discourse Topic) wird in diesem Kontext wie folgt definiert:

A discourse topic  $DT_i$  is defined by the set of all topics  $T_p$  that are constituted as the result of one and the same feeder  $F_i$  ( $DT_i = \{ T_p \mid T_p \text{ constituted a.r.o. } F_i \}$ ). As such  $DT_i$  is a set of main, higher order topics usually hierarchically comprising lower topics. (van Kuppevelt 1995, 137, Hervorhebung im Original)

Van Kuppevelt geht davon aus, dass sich nicht alle Texte auf ein einzelnes Diskursthema zurückführen lassen, von dem alle anderen Subthemen abgeleitet werden, sondern unterscheidet zwischen gebundenen (“bound”) und freien (“free”, “unbound”, “spontaneous”) Diskursen. Nur gebundene Diskurse werden demnach durch ein einzelnes Diskursthema determiniert, freie Diskurse dagegen können mehrere Diskursthemen aufweisen.

Van Kuppevelts Ansatz weist dem Thema eine zweifache Funktion zu: die Funktion der sequenziellen Verknüpfung von Äußerungen in Diskursen und die Funktion der hierarchischen Verknüpfung von größeren Diskurseinheiten, wobei sowohl thematische Verknüpfungen zwischen aufeinander folgenden Äußerungen als auch thematische Verknüpfungen zwischen größeren Diskurseinheiten als

Frage-Antwort-Verknüpfung erklärt werden – bei der sequenziellen Verknüpfung sind die Fragen jedoch an den Diskurskontext bzw. die vorangegangene Äußerung gebunden, während Fragen, die größere Diskurseinheiten verknüpfen und hierarchisch höher stehend sind, kontextuell frei sind. Mit diesem Modell gelingt es van Kuppevelt zum einen, Satz- und Textthemen als ontologisch ähnliche Konzepte zu bestimmen. Zum anderen entwickelt er ein plausibles Textmodell, das sowohl eine Erklärung des linearen (oder lokalen) Zusammenhangs von Diskursen als auch des hierarchischen (oder globalen) Zusammenhangs bietet. Wie sich dieses theoretische Thema-Konzept operationalisieren bzw. in einem Analyseverfahren umsetzen lassen könnte, wird in van Kuppevelts Arbeiten leider nicht dargelegt.

Ein weiterer Ansatz, bei dem der Aufbau des Textes als vom Thema als Frage bzw. „Quaestio“ gesteuert aufgefasst wird, ist der von Stutterheim (1994; 1997) bzw. Klein/von Stutterheim (1992). Klein und von Stutterheim gehen davon aus, „dass jeder kohärente Text als Antwort auf eine einleitende Frage, eine Quaestio, zu verstehen ist“ (von Stutterheim 1994, 254). Durch die Quaestio wird ein „Referenzrahmen“ vorgegeben, dessen „Lücken“ mit der Antwort (dem Rest des Textes) „zu füllen“ sind; mit der Quaestio sind demnach Vorgaben für den Textaufbau verbunden (von Stutterheim 1994, 254). Diese Vorgaben beziehen sich auf die Gliederung des Textes in Haupt- und Nebenstruktur, Topik-Fokus-Struktur und auf die „Muster der referenziellen Bewegung“ (vgl. Klein und von Stutterheim 1992, 90). Über bestimmte Muster der referenziellen Bewegung werden Informationen zu einem durch die Quaestio aktivierten Sachverhalt entfaltet, die sechs „Referenzbereichen“ (die als „konzeptuelle Kategorien“ oder Konzeptbereiche verstanden werden, vgl. Kohlmann 1997, 41) zuzuordnen sind: Zeitreferenz, Raumreferenz, Referenz auf Umstände, Personenreferenz, Prädikate, Modalität (vgl. Klein und von Stutterheim 1992, 81ff.).<sup>71</sup>

Insofern die Quaestio als leitende Frage betrachtet wird, die die Textproduktion und den lokalen und globalen Informationsaufbau steuert, weist dieser Ansatz einige Parallelen, zu dem Ansatz von van Kuppevelt (1993; 1995) auf. Funktional scheint die Quaestio eine ähnliche Funktion zu haben wie van Kuppevelts Feeder: Aus beiden können die Inhalte des Textes abgeleitet werden bzw. beide machen Vorgaben für den Textaufbau. Während aus der Quaestio verschiedene Referenz- bzw. Konzeptbereiche abgeleitet werden können, die im Text gefüllt werden können, lassen sich aus dem Feeder die Topic-Constituting Questions ableiten, die in ihrer Gesamtheit das Diskurstopik definieren. Eine weitere Parallele zwischen den Ansätzen betrifft die Unterscheidung von Haupt- und Nebenthemen. Van Kuppevelt unterscheidet in diesem Zusammenhang zwischen den an den lokalen Äußerungskontext gebundenen (und nicht auf den Feeder bezogenen) Themen und Themenfragen und den kontextuell freien, jedoch an den Feeder gebundenen Themen und Themenfragen (erstere sind Nebenthemen,

---

<sup>71</sup> Die zentrale Rolle der „Konzeptbereiche“ bei Klein und von Stutterheim macht deutlich, dass dieser Ansatz stark kognitiv ausgerichtet ist. Insofern die „Quaestio“ als Fragliches allerdings eine noch prominentere Rolle für diesen Ansatz spielt, wird er in der vorliegenden Arbeit unter den kommunikativ-pragmatischen Thema-Konzepten aufgeführt.

letztere Hauptthemen). Klein und von Stutterheim gehen in ähnlicher Weise davon aus, dass es Äußerungen gibt, die sich direkt auf die Quaestio beziehen, die die Hauptstruktur des Textes bilden, und Äußerungen, die zusätzliche, an den Äußerungskontext gebundene Informationen geben (d.h. die kontextuell gebunden, aber nicht an die Quaestio gebunden sind) (vgl. Klein und von Stutterheim 1992, 73). Über die Differenzierung von Hauptstruktur und Nebenstrukturen des Textes werden demnach auch Haupt- und Nebenthemen unterschieden.

Eine Abwandlung des Konzepts „Thema als Frage“ ist die Definition des Themas als „Problemstellung“ bzw. „mangelhaftes Objekt“, dessen Mangel durch den Text behoben werden soll (vgl. Lötscher 1987, 84). Auch wenn Lötscher sein Thema-Konzept bewusst von allen bisherigen Konzeptionen abheben möchte und damit das Ziel verfolgt, eine für alle Texttypen gültige Definition des Themas zu liefern, sind doch deutliche Gemeinsamkeiten zu tradierten Konzeptionen erkennbar: Lötschers Thema steuert wie das „Thema als Frage“ den inhaltlichen Aufbau des Textes und leitet die „Füllung“ der durch die implizite Frage bzw. „Problemstellung“ aufgezeigten inhaltlichen Leerstellen des Textes. Wie Keenan/Schieffelin das Thema als im Hinblick auf die Frage „präsupponiert“ definieren, beschreibt auch Lötscher das Thema als bekannt im Hinblick auf den Mangel des Objekts, insofern dieser Mangel sowohl für Sprecher als auch Hörer offensichtlich bzw. bekannt sein müsse (vgl. Lötscher 1987, 85).<sup>72</sup> Auch die in Lötschers Arbeit betonte Verzahnung von thematischer und funktionaler Struktur („die funktionale Struktur [ist] mit der thematischen Struktur isomorph“, Lötscher 1987, 200), die sich in bestimmten „Themenverknüpfungsverfahren“ niederschlägt, z.B. bei der Themaeführung über „Problematisierung“ (Textfunktion „Begründen“) oder bei der Themaeführung zur „Inhaltsergänzung“ (Textfunktion „Erläutern“) (vgl. Lötscher 1987, 200ff.), ist bereits in Keenans/Schieffelins (1976) Arbeit angelegt: Sie weisen darauf hin, dass für die Bestimmung des Diskursthemas (sowie des thematischen Zusammenhangs) die Analyse der Funktion einer Äußerung notwendig ist: „In determining the discourse topic of an utterance, it is useful to determine the purpose or reason behind each utterance. Why did the speaker say what he did?“ (Keenan und Schieffelin 1976, 345).

Lötscher hat den Anspruch, das Thema als textspezifische Beschreibungskategorie ernst zu nehmen. Eine Analyse von Satzthemen und deren Abfolge sei dem-

---

<sup>72</sup> Die Begriffe Präsupposition und Bekanntheit wurden in der Forschung zum Satzthema häufig in einer ähnlichen Verwendungsweise benutzt. Während sich der Aspekt der Bekanntheit auf den kognitiven Status eines Diskursreferenten im Gedächtnis von Sprecher und Hörer bezieht, wird das Konzept Präsupposition in der Forschung zum Satzthema entweder als pragmatischer Begriff verwendet, der den nicht-kontroversen Teil des gemeinsamen Hintergrundwissens bezeichnet, als semantischer Begriff (siehe Strawson 1950) oder als logischer Begriff (wie bei Keenan/Schieffelin); Präsupposition in diesem Sinne ist eine implizite logische Aussage, die aus einer explizit gemachten sprachlichen Aussagen abzuleiten ist, wobei gilt, dass die Aussage A und die Aussage „Nicht-A“ (d.h. die Negation von A) die gleiche Präsupposition haben. Präsuppositionen stellen dabei Voraussetzungen für das Verstehen bzw. Gelingen von Äußerungen dar (zum Begriff „Präsupposition“ siehe auch Seuren (1991)).



entsprechend nicht ausreichend. Dennoch unterteilt er in seinen Beispielanalysen häufig doch Texte in Sätze und untersucht diese dann im Hinblick auf ihr Thema bzw. ihren thematischen Zusammenhang zu adjazenten Sätzen. Deshalb ist zu konstatieren, dass Lötscher seinen eigenen Anspruch nicht erfüllt und die textkonstituierende Funktion des Themas zu wenig beleuchtet.

### 2.3.3 Thema als Mittel zur Strukturierung von Diskursen: Thematische Textmuster

In den im vorherigen Abschnitt behandelten Ansätzen zum „Thema als Frage“ wurde die Funktion des Themas im Kontext seiner Bedeutung für die Textproduktion betrachtet: Ein Thema stellt dieser Konzeption zufolge eine implizite Frage oder Problemstellung dar, die der Sprecher mit dem Text zu beantworten bzw. zu lösen versucht. Themen als implizite Fragen stellen dabei Anknüpfungspunkte zwischen den einzelnen Äußerungen dar und sorgen so für sequenzielle Kohärenz (siehe den im vorigen Kapitel dargestellten Ansatz von Hellwig 1984a, b). Die sequenzielle thematische Organisation von Diskursen wurde darüber hinaus im Kontext der Arbeiten zur thematischen Organisation (Hoffmann 1997) und thematischen Progression (Daneš 1970) besprochen (siehe Kapitel 2.2.1.2). Zum anderen sind Themen als implizite Fragen aber auch Ausgangspunkt für die hierarchische Organisation von Diskursen, insofern die verschiedenen Diskurstteile aus einem übergeordneten gemeinsamen Textthema (bzw. der Quaestio) abgeleitet werden (siehe die im vorigen Kapitel dargestellten Ansätze von van Kuppevelt (1993; 1995) und Klein und von Stutterheim (1992; 1994; 1997)). Ein weiterer Ansatz zur kommunikativ gesteuerten, hierarchisch-thematischen Strukturierung von Diskursen, der in der Forschung breit rezipiert wurde, beschreibt thematische Textmuster als „Grundformen thematischer Entfaltung“ (Brinker 2000, 170). Diese sollen im Folgenden ausführlich dargestellt werden.<sup>73</sup>

Brinkers „Grundformen thematischer Entfaltung“ basieren auf den Werlichen Texttypen bzw. „texttypischen thematischen Textbasen“ – deskriptive, narrative, expositorische, argumentative und instruktive Textbasen (vgl. Werlich 1975, 30ff.). Grundformen thematischer Entfaltung, die von Heinemann/Viehweger (1991, 237ff.) auch als „komplexe Strategiemuster“ bezeichnet werden, sind „Deskription“ (siehe Heinemann 2000c), „Narration“ (siehe Gülich 2000), „Explikation“ (siehe Jahr 2000) und „Argumentation“ (siehe Eggs 2000). Diese thematischen Vertextungsmuster stellen „Grundentscheidungen für eine pragmatisch-adressatenspezifisch bezogene Ausformung von Kommunikationsgegenständen in einem Text“ dar (Gansel und Jürgens 2002, 138); sie machen

---

<sup>73</sup> Brinker geht davon aus, dass das Thema der „Kern des Textinhalts ist, wobei der Terminus ‚Textinhalt‘ den auf einen oder mehrere Gegenstände (d.h. Personen, Sachverhalte, Ereignisse, Handlungen, Vorstellungen usw.) bezogenen *Gedankengang* eines Textes bezeichnet“ (Brinker 1997, 55, Hervorhebung von M.B.). Obwohl es sich bei diesem Ansatz somit um einen Vertreter der in Kapitel 2.2.2 dargestellten semantischen Thema-Konzepte („Thema als abstrakter Kern des Textinhalts“) handelt, können die „Grundformen thematischer Entfaltung“ jedoch auch als durch das Thema gesteuerte kommunikative Textmuster aufgefasst werden. Sie werden deshalb im Kontext der kommunikativ-pragmatischen Thema-Konzepte dargestellt.

kommunikative Vorgaben darüber, wie ein Text thematisch strukturiert werden kann. Einschränkend ist allerdings festzuhalten, dass diese Formen in der Realität selten isoliert vorkommen, sondern sich häufig mischen. Den Grundformen thematischer Entfaltung werden jeweils bestimmte Typen von Themen (z.B. „Thema als regelhaft dargestellter Vorgang“ bei deskriptiver Themenentfaltung) sowie typische sprachliche Mittel<sup>74</sup> zugeordnet (vgl. Brinker 1997, 63) (siehe Tabelle 2).

---

<sup>74</sup> Da in kommunikativ-pragmatisch orientierten Ansätzen häufig ein hermeneutischer Zugang zur Bestimmung von Themen gewählt wird, werden diese sprachlichen Mittel jedoch gleichzeitig mit einiger Skepsis bezüglich ihrer Eindeutigkeit als „Marker“ gesehen.

<i>Typen thematischer Entfaltung</i>	<i>Thematische Kategorien und linguistische Merkmale</i>
<p><u>Deskriptive Themenentfaltung:</u> Das Thema wird in seinen Komponenten dargestellt und in Raum und Zeit eingeordnet. Die wichtigsten thematischen Kategorien sind Spezifizierung und Situierung.</p> <p><i>Typische Textsorten:</i> Informative Texte (Nachricht, Bericht, Lexikonartikel, wissenschaftliche Abhandlung); instruktive Texte (Bedienungsanleitung, Kochrezept); normative Texte (Gesetz, Vertrag, Vereinbarung, Testament)</p>	<p><i>Variante A:</i> Das Thema bezeichnet einen einmaligen Vorgang, z.B. ein historisches Ereignis (Bericht). Der thematische Textaufbau orientiert sich am zeitlichen Ablauf des berichteten Geschehens. <i>Grammatische Mittel:</i> Vergangenheitstempora, Temporal- und Lokalbestimmungen</p> <p><i>Variante B:</i> Das Thema bezeichnet einen als regelhaft (generalisierbar, wiederholbar) dargestellten Vorgang (Beschreibung). Das Thema wird in seine wesentlichen Teilvorgänge untergliedert. <i>Grammatische Mittel:</i> Handlungsverben, Infinitiv (instruktive Textfunktion, vor allem in Bedienungsanleitungen), Vorgangspassiv (ebenfalls in instruktiver Textfunktion), Wegfall des Artikels und Koordination der infinitivischen Fügungen</p> <p><i>Variante C:</i> Das Thema bezeichnet ein Lebewesen oder einen Gegenstand (Beschreibung). Die Themenentfaltung vollzieht sich nach der Teil-Ganzes oder Enthaltenseins-Relation; weitere Möglichkeit: Beschreibung des zentralen Diskursreferenten. <i>Grammatische Mittel:</i> durchgehende Wiederaufnahmestruktur</p>
<p><u>Explikative Themenentfaltung:</u> Thema = Explanandum (Sachverhalt, der im Text erklärt werden soll)</p> <p><i>Typische Textsorten:</i> Lehrbuch, wiss. Texte, die Wissen erweitern sollen</p>	<p><i>Thematische Kategorien:</i> a) Explanandum: Sachverhalt, der im Text erklärt werden soll; b) Explanans: soll den Sachverhalt erklären und besteht aus zwei Teilen: Anfangs- oder Randbedingungen und allg. Gesetzmäßigkeiten.</p> <p><i>Grammatische Mittel:</i> Dominanz von Konjunktionen, Adverbien und Präpositionen, die Kausalbeziehungen im weitesten Sinn (Grund, Ursache, Bedingung, Folge) signalisieren.</p>
<p><u>Argumentative Themenentfaltung:</u> Thema = strittige Behauptung bzw. These</p> <p><i>Typische Textsorten:</i> Appellative Texte (z.B. Kommentare), bestimmte Typen normativer Texte (z.B. Gerichtsentscheidungen), bestimmte Typen informativer Texte (z.B. Rezensionen)</p>	<p><i>Thematische Kategorien:</i> (strittige) Behauptung bzw. <i>These</i> (Textthema), <i>Argumente</i> für die These, <i>Schlussregel</i>, die die Zulässigkeit der Argumente rechtfertigt, <i>Stützung</i>, um die Zulässigkeit der Schlussregel zu beweisen, <i>Modaloperator</i> (Angabe der Wahrscheinlichkeit der These), Umstände, die die Gültigkeit der Schlussregel einschränken (<i>Ausnahmeregel</i>), <i>Einbettung</i> (ordnet These und Argumente in Kontext ein), <i>Wertbasis</i></p> <p><i>Grammatische Mittel:</i> Subordination von Sätzen, kausale, konditionale, konsekutive und adversative Satzverknüpfung.</p>

Tabelle 2: Grundformen thematischer Entfaltung nach Brinker (1997)

Grundlage für die Beschreibung der Themenentwicklung sind demnach drei Aspekte: Der Thementyp bzw. die thematische Kategorie, die Relationen zwischen Themen sowie die grammatischen Mittel zur Signalisierung dieser Relationen. Die Relationen zwischen Textthema und Teilthemen werden zunächst allgemein als „logisch-semantische Relationen“ bezeichnet. Im Hinblick auf die Formen thematischer Entfaltung spezifiziert Brinker dies und nennt z.B. „Teil-Ganzes-Relation“ (deskriptive Entfaltung), „Explanans-Explanandum“ (explikative Entfaltung) oder „These-Argument-Schlussregel“ (argumentative Entfaltung). Darüber hinaus beschreibt er in einem Beispiel (Zeitungsnachricht zum Thema „Wohnungsbrand“) das Entfaltung-Schema „Bekämpfung-Ursachen-Folgen“.

Die Entfaltungsmuster basieren demnach auf sehr unterschiedlichen Typen von Relationen und unterschiedlichen Typen von Themen: semantische Relationen zwischen Konzepten (Teil-Ganzes-Relationen), Kohärenzrelationen zwischen Propositionen (Ursache, Folge) sowie generische Relationen zwischen Textteilen (These-Argument-Schlussregel). Eine theoretische Begründung für diese disparate Zusammenstellung fehlt; ob sich Teilthemen aus dem Textthema selbst semantisch ableiten lassen, oder inwiefern es sich um generisch motivierte Typen handelt, wird nicht zufrieden stellend beantwortet. Auch die Frage nach dem Zusammenhang der Entfaltungsmuster mit Kohärenzrelationen bleibt offen.

Eine alternative, auf Werlichs Texttypen aufbauende Modellierung thematischer Vertextungsmuster in Texten stellt die im Kontext der Systemisch Funktionalen Linguistik entwickelte Methode der thematischen Entwicklung („method of development“) dar (siehe Ghadessy 1995; Ramm und Villiger 1995; Villiger 1996; Lavid 1997, 2000). In diesem Ansatz wird – ähnlich wie in Brinkers Ansatz zur thematischen Entfaltung – davon ausgegangen, dass es einen Zusammenhang zwischen Texttyp, Verknüpfungsstrategie, Inhalt des Texts und der Methode der thematischen Entwicklung (bezogen auf Satzthemen) gibt.

Theme selection in the clause most clearly correlates with *the thematic progression* of the text. And, as [Fries 81] shows for English texts, different kinds of thematic progression correlate with different genres and are closely related to thematic content, i.e., the subject matter or domain of the text. There is a close relation between thematic content and the method of development of a text which comprises general organizations such as spatial, temporal, general to specific, object to attribute, object to parts or compare and contrast. (Ramm und Villiger 1995, 7, Hervorhebung im Original)

Für jeden Texttyp (bei Lavid: Exposition, Deskription, Narration, Instruktion, Argumentation) können demnach typische Satzthementypen und Typen thematischer Progression beschrieben werden:

Texttyp (Text Type)	Inhalt (Subject Matter)	Verknüpfungs- strategie (Chaining Strategy)	Themen- wahl (Theme <sup>75</sup> Selection)	Beispiel (Sample)
Expositorisch	Klassenkonzepte, generische Konzepte	Charakterisie- rung	Topical	Eintrag in eine Enzy- klopädie
Deskriptiv	Räumliche Bezüge	Räumlich	Ort	Reiseführer
Narrativ	Ereignisse	Zeitlich	Zeit	Stadtgeschichte
	Personen	Teilnehmer	Topical	Biographien
Instruktiv	Schritte einer Proze- dur	Sequenziell	Zeit Prozess	Rezepte Bedienungsanleitun- gen
Argumentativ	Fakten und Ideen	Argument und Gegenargument	Verknüpfung	Zeitungsartikel

Tabelle 3: Korrelationen zwischen Themen und Themenverknüpfungen  
(nach Lavid 2000)

Dieser Ansatz stellt eine Ergänzung der Vertextungsmuster dar, insofern neben den typischen Formen der Satzverknüpfung auch Inhalte und Satzthemenkategorien herausgearbeitet werden. Ghadessy (1995) geht darüber hinaus davon aus, dass auch die grammatischen und lexikalisch-semantischen Eigenschaften von Themen je nach Texttyp unterschiedlich sind und dass die Häufigkeit bestimmter Thementypen mit den für einen Texttyp typischen Strukturelementen korreliert. In Sportberichterstattungen treten beispielsweise überwiegend Eigennamen als (Satz-)Themen auf („animate Themes“).

Während in den bisher behandelten Ansätzen eine Differenzierung der Formen der Themenentwicklung anhand der grundlegenden *Texttypen* Narration, Deskription, Explikation, Argumentation vorgenommen wird, betrachtet Schank (1977a) die Funktion des Themas als ein die „Interaktion“ organisierendes Konzept *textsortenspezifisch* differenziert. Ihm zufolge ist das Thema nur in Textsorten, bei denen es ein „Großthema“ gibt (z.B. Beratungsgespräch, Diskussion), das grundlegende Strukturprinzip für die gesamte Interaktion. In Textsorten, in denen dies nicht der Fall ist (z.B. Small Talk, Interview), wird die Einheit der Interaktion dagegen durch eine bestimmte Intention konstituiert (vgl. Schank 1977a, 236f.; siehe auch Fritz 1994, 194). Wenn ein Text nur ein Großthema aufweist, bleibt dieses nach Schank durchgehend im Fokus der Interaktion der Kommunizierenden. Sämtliche Subthemen müssen deshalb einen Bezug zum Großthema aufweisen, ein Themenwechsel ist dementsprechend nur bei Subthemen möglich, nicht beim Großthema (vgl. Schank 1977a, 236f.).

Neben der Textsorte haben darüber hinaus auch allgemeine kommunikative „Konversationsregeln und Assoziationsregeln“ einen Einfluss auf die thematische Textstruktur (Schank 1977b): „The notion of topic cannot really be ade-

<sup>75</sup> Das Thema wird hier im Anschluss an Halliday grammatisch definiert als „point of departure of the message“, das durch Satzterstposition realisiert wird (Villiger 1996; Lavid 2000).

quately defined without the addition of a set of conversational rules that constitute the use of that notion of topic" (Schank 1977b, 423). Die Konversationsregeln geben an, welche inhaltlichen Anchlüsse an eine Äußerung zugelassen sind (z.B. kann eine Äußerung zum Thema „Handlung“ durch eine Äußerung zum „Effekt“ dieser Handlung auf den Hörer ergänzt werden). Zusätzlich müssen spezielle thematische Regeln berücksichtigt werden, die sowohl beschreiben, was thematisiert werden kann als auch wie thematisiert werden kann und wie Themenwechsel bzw. thematische Übergänge vorgenommen werden können: "The rules for topics, then, are really the rules for topic shifts."; "the rules for topic shift are also the rules for what can follow what in a conversation." (Schank 1977b, 428). Die zentrale thematische Regel besagt beispielsweise, dass eine Äußerung einerseits an das eingeführte Thema anschließen muss, dass sie andererseits aber auch „potentielle neue Themen“ einführen muss, die in der folgenden Äußerung aufgenommen und zu einem neuen Thema verfestigt werden können ("sticking and shifting", vgl. Schank 1977b, 424). "The potential topic comes from the initial conceptualization and from the rules for appropriate topic shifts" (Schank 1977b, 424). Das neue Thema kann dabei z.B. aus Elementen des reduzierten alten Themas ("Reduced Old Topic") sowie neuen Elementen konstruiert werden.

Diese Zweigliedrigkeit entspricht der von Brown/Yule (1983, 87f.) beschriebenen, die wie Schank darauf hinweisen, dass eine Äußerung zum einen an das alte Thema anschließen muss ("conversational topic"), dass Sprecher aber zum anderen auch neue, für sie persönlich bedeutsame Themen einführen ("personal topic" oder "speaker's topic"). Ein Sprecher macht demnach immer zwei Typen von thematischen Beiträgen: Beiträge, die sich in den etablierten thematischen Rahmen einbetten lassen, und Beiträge, die ein persönlich relevantes Thema hervorheben, das aber möglicherweise im weiteren Verlauf des Gesprächs nicht mehr vertieft bzw. vom Gesprächspartner nicht aufgegriffen wird. Das führt dazu, dass bei der Themenentwicklung der Übergang von Themen häufig nicht abrupt erfolgt, sondern fließend: „Oft ist es uns unmöglich zu entscheiden, wo ein Thema aufhört und das andere anfängt.“ (Fritz 1982, 218).

Um zu verdeutlichen, wie mögliche Antworten auf eine Äußerung bzw. Äußerungsfolge beschaffen sein müssen, bringt Schank (1977b, 426) folgendes Beispiel. Adäquate Antworten auf die Äußerungsfolge

(1) "John bought a red car in Baltimore yesterday."

(2) "You mean he's not going to buy my car?"

könnten demnach sein:

(3) "Well, John needed a car in a hurry."

(4) "No, he didn't like your car."

Äußerung (3) würde an das reduzierte alte Thema aus (1) anschließen, das perspektivisch auf das in (2) eingeführte neue Thema ausgerichtet wird, während in (4) der Themenübergang aus (2) bestätigt wird und "my/your car" als neues Thema etabliert wird. Bei beiden Antworten gibt es nach Schank mehrere weite-

re sich anschließende Verzweigungsmöglichkeiten. So könnte das Thema im weiteren Verlauf auf ein „Supertopic“ („input conceptualization and the response to that conceptualization contain concepts in the topic that share the same superset“) oder „Metatopic“ („a metatopic is a comment that can be inferred from the interaction of two conceptualizations“) gelenkt werden (vgl. Schank 1977b, 426). Mit der Äußerung

(5) „It is always difficult to sell a car. Have you tried the Gazette want ads?“

würde beispielsweise das Superthema „Buying/Selling Cars“ als globales Thema des Dialogs etabliert. Adäquate Antworten auf einen Sprecherbeitrag müssen nach Schank also entweder „Supertopic“, „Metatopic“ oder „Reduced Old Topic“ + „New Topic“ sein.

Neben den durch Vertextungsmuster, Textsortenzugehörigkeit sowie allgemeine Konversations- und Assoziationsregeln gesteuerten Beschränkungen wirken möglicherweise auch semantische oder ontologische Vorgaben auf die Entwicklung eines Themas im Text ein. So geht z.B. Fritz (1982, 208) davon aus, dass der Typus eines Themas Auswirkungen auf die „Züge“ in einer Kommunikation hat und dass „verschiedene Gegenstände (...) ganz verschiedenartige Züge in der jeweiligen Kommunikation zulassen“. Fritz führt diese Annahme nicht weiter aus und geht auch nicht weiter auf mögliche kognitiv-ontologische Grundlagen dieser Annahme ein. Schröder (2003), der sich auf Fritz bezieht, geht ebenfalls davon aus, dass

das Thema (...) unter verschiedenen Gesichtspunkten behandelt [wird], die sich aus der ‚Valenz‘ des Textgegenstands ergeben. Gemeint sind die typischen Ergänzungszüge, die sich an eine Gegenstandseinführung anschließen können. Greifbar sind sie in den Fragen, die in bezug auf einen Gegenstand gestellt werden können, oder in den typischen Möglichkeiten der inhaltlichen Spezifizierung, die ein Gegenstand von Mitteilungshandlungen eröffnet. Typische Aspekte bei ereignisbezogener Berichterstattung sind also beispielsweise der Ereignisverlauf, die Vorgeschichte, Ergebnisse, Folgen und Konsequenzen, ein Ausblick auf den erwarteten Fortgang, Hintergründe, Begleitumstände oder Reaktionen. (Schröder 2003, 85).

Schröder nimmt – ebenso wie Fritz und Brinker (s.o.) – an, dass der Typus des Textgegenstands, auf den eine sprachliche Handlung bezogen ist, Auswirkungen auf die „Züge“ in einer Kommunikation hat. Er geht in seinem Ansatz davon aus, dass sich thematische und funktionale Struktur ergänzen und zusammen die Handlungsstruktur von Texten bilden. Themen werden dabei entwickelt, indem sie in Aspekte, Gegenstände oder Verlaufskomponenten zerlegt werden; diese stellen jeweils einen Gegenstand einer sprachlichen Teilhandlung dar.<sup>76</sup> Die thematischen Bezüge konstituieren gleichzeitig den Zusammenhang auf der Handlungsebene, da das Thema den Bezugspunkt der sprachlichen Handlungen bildet, indem es in Form von Teilthemen in sprachlichen Teilhandlungen behandelt wird (vgl. Schröder 2003, 90). Offen bleibt allerdings, wie die Relationen zwischen Thema und Teilaspekt bzw. Teilereignis genau beschaffen sind bzw. begründet werden und inwiefern diese Relationen auf kognitiven bzw. on-

<sup>76</sup> Schröder nennt als Beispiel für eine Aspektzerlegung: „Thema: Anschläge, zerlegt in Aspekt: Schäden, Aspekt: Täter“ (Schröder 2003, 86).

tologischen Zusammenhängen beruhen. Diese Unklarheit hat ihre Ursache auch in der Offenheit der Definition des Themas bei Schröder, der das Thema als „Gegenstand, auf den mit einer sprachlichen Texthandlung Bezug genommen wird“ (Schröder 2003, 92), charakterisiert, jedoch nicht weiter darlegt, wie dieser „Gegenstand“ ontologisch bestimmt ist und wie die Relationen zwischen den Teil-Aspekten bzw. Teil-Gegenständen beschaffen sind. In der beispielhaften Kategorisierung der Teilthemen in ereignisbezogenen Nachrichten – in Vorgeschichte, Ergebnisse, Folgen und Konsequenzen etc. (Schröder 2003, 85) – deutet sich an, dass die Teilthementypen Schröders eher texttypspezifische Kategorien darstellen, wie sie z.B. von van Dijk (1985) beschrieben werden. Wenn dies der Fall ist, stellt sich allerdings die Frage, inwiefern die Terminologie „thematische Entwicklung“ das beschriebene Phänomen passend charakterisiert, oder ob es nicht vielmehr angebrachter wäre, von texttypspezifischen Mustern zu sprechen.

Sowohl *was* als auch *wie* etwas thematisiert werden kann, unterliegt demnach bestimmten Beschränkungen. Themen werden – zumindest in bestimmten Textsorten – in regelhafter und demnach erwartbarer Form entwickelt und strukturieren so den Diskurs. Themenentwicklungen können deshalb auch musterhaft aufgefasst werden. Der Grad der Normativität variiert allerdings zwischen den verschiedenen Gesprächs- bzw. Textsorten.<sup>77</sup>

#### 2.3.4 Fazit

Allen in diesem Kapitel besprochenen Ansätzen ist die Annahme gemeinsam, dass die Beschreibung des Konzepts Thema bzw. der sprachlichen Handlung des Thematisierens eingebettet werden muss in ein kommunikationstheoretisches Modell, d.h. ein Modell, in dem Äußerungen in Texten oder Dialogen als Produkte von Sprechern zu betrachten sind, die im Kontext ihrer Produktions- und Rezeptionsbedingungen zu beschreiben sind. Das Thema wird dabei als Mittel zur Strukturierung von Diskursen gesehen, das den Akt der Textproduktion leitet.

In den kommunikativ-pragmatischen Thema-Konzeptionen wird das Thema dabei nicht mehr als eine Eigenschaft des Textes bzw. Dialogs an sich definiert. Stattdessen wird das Thematisieren als intentionale Handlung aufgefasst, die konstituierend für die Herstellung von Texten oder Dialogen ist. Über diese Funktion des Themas bzw. Thematisierens erklärt sich auch die zentrale Bedeutung dieses Konzepts für den Textzusammenhalt bzw. die Textkohärenz: Es legt dem Sprecher Beschränkungen auf, was er in einem Text sagen kann und wie er thematische Übergänge gestalten kann, und es produziert inhaltliche Erwartbarkeit auf Seiten des Hörers und ermöglicht so die Konstruktion eines Verstehens-

---

<sup>77</sup> In „institutionell gebundenen Gesprächssorten“ (z.B. Lehr- und Lernkommunikationen oder Interviews) werden Fritz (1994, 194) zufolge thematische Zusammenhänge mehr oder weniger nach Vorgabe abgearbeitet. „Kreative Gespräche“ dagegen weisen einen solchen festen thematischen Zusammenhang nicht auf. Schank (1977a, 235) weist darüber hinaus darauf hin, dass bestimmte Themenklassen in bestimmten Textsorten „regelmäßig erwartbar“ sind (z.B. in Textsorten wie Beratungsgespräch oder Diskussion). Wie Fritz geht er allerdings ebenfalls davon aus, dass es andere Textsorten gibt (z.B. „small talk“), in denen solche regelhaften Zusammenhänge nicht existieren.



horizonts für den Text (siehe auch Kapitel 3.4.2).

Im Rahmen dieses Perspektivwechsels vom Text als Produkt zum Text als Prozess bzw. zur Einbeziehung von Textproduktion und -rezeption ist es nach Brown/Yule (1983, 74ff.) unbedingt erforderlich, den Diskurskontext in die thematischen Analyse einzubeziehen und zu ermitteln, wie der kontextuelle Rahmen beschaffen ist, innerhalb dessen das Thema entwickelt wird. Dieser Rahmen enthält u.a. sämtliche relevanten (aktivierten) Informationen zum Diskurskontext (Zeit, Ort, Eigenschaften der Gesprächspartner, Verhältnis der Gesprächspartner), Informationen über die Sprecherbeiträge, d.h. stichwortartige Zusammenfassungen des Inhalts, oder sogar ausführliche Transkriptionen der Sprecherbeiträge. Alles zusammen konstituiert schließlich den „thematischen Rahmen“ („topic framework“), der die Grundlage für die Beurteilung der Frage bildet, ob eine Äußerung thematisch relevant ist.

Part of the process of analysing discourse in terms of ‘topic’ is an attempt to make explicit the basis for our intuitive ability to recognise why what is said is appropriate in a particular discourse fragment. (Brown und Yule 1983, 78)

Obwohl Brown/Yule mittels der Angaben zum Topic Framework das Ziel haben, einen objektiveren Analyserahmen zu schaffen, bleibt ihre Analysemethode, die sie beispielhaft an einem Dialogausschnitt demonstrieren, intuitiv und interpretativ. Generell wird in kommunikativ-pragmatisch orientierten Arbeiten häufig die Position vertreten, dass eine „objektive“, „mechanische“ oder gar „automatische“ Themenbestimmung nicht möglich sei (siehe z.B. Fritz 1982; 1994). Möglich ist es allerdings, die Grundlagen, d.h. die Wissensbestände und Indizien, auf die sich Analysierende bei der hermeneutischen Ermittlung des Themas stützen, offen zu legen. Folgende Aspekte sind Fritz zufolge zentral für die Analyse des Themas (siehe Fritz 1994, 193):

- „Ausdrücke, die im Dialogtext besonders häufig oder an prominenter Stelle vorkommen bzw. die syntaktisch oder intonatorisch besonders hervorgehoben sind“;
- „Wissen über die für diese Dialogform charakteristischen Muster der Themenbehandlung“;
- „Folgerungsbeziehungen zwischen den im Text ausgedrückten Propositionen und übergeordneten Propositionen“;
- „Wissen über relevante thematische Zusammenhänge“;
- „Wissen über aktuelle Interessen, Probleme, Fragestellungen der Dialogpartner“.

Interessant an dieser Aufstellung ist die Tatsache, dass Faktoren, die in grammatischen oder semantischen Thema-Konzeptionen zur Bestimmung des Themas eingeführt wurden, in Fritz kommunikativ-pragmatischem Ansatz ebenfalls als Indizien für die Themenanalyse verwendet werden (syntaktische Hervorhebung, „topic continuity“ – häufig wiederaufgenommene Gegenstände, Relationen zwischen Propositionen). Gleichzeitig wird dieser „bottom-up“-orientierte Ansatz der Analyse jedoch ergänzt durch „top-down“-Strategien: dem Wissen über ge-

nerische Strukturen, Formen der Themenentwicklung, Wissen über relevante thematische Zusammenhänge etc. Diese kombinierte Herangehensweise ist meiner Ansicht nach sehr fruchtbar und wird deshalb in Kapitel 4 dieser Arbeit ausführlicher ausgearbeitet werden. Gleichzeitig soll dabei aber gezeigt werden, dass auf dieser Basis und mit Hilfe computerlinguistischer Methoden und Formalismen ein operationalisierbares und maschinelles Verfahren zur Themenanalyse entwickelt werden kann.

## 2.4 Kognitiv orientierte Thema-Konzepte

Wie in den kommunikativ-pragmatischen Ansätzen wird auch in den kognitiv orientierten Ansätzen zum Thema davon ausgegangen, dass nicht Texte Themen haben, sondern Sprecher und Hörer bzw. Produzenten und Rezipienten. Während kommunikativ-pragmatische Arbeiten allerdings einen Fokus auf die Interaktionen von Sprechern und Hörern sowie die Bedeutung der Kommunikationssituation legen, stellen kognitive Ansätze die kognitiven Verarbeitungsprozesse und die Strukturierung der mentalen Wissensbestände in den Mittelpunkt. In einigen Fällen ist die Abgrenzung allerdings schwierig.<sup>78</sup> Auch mit semantischen Ansätzen weisen einige der in diesem Kapitel behandelten kognitiven Arbeiten Überschneidungen bzw. Gemeinsamkeiten auf: Die Charakterisierung des Themas als „bekannte Information“ ist prinzipiell eher als semantische Themabeschreibung zu verstehen, allerdings werden zur genauen Bestimmung dessen, was eigentlich als „bekannt“ gelten kann, häufig kognitive Erklärungen hinzugezogen und „Bekanntheit“ auch als mentale „Aktiviertheit“ definiert, so dass die entsprechenden Thema-Konzepte in die Ebene der Kognition eingeordnet werden.

Eine rein kognitive Definition des Konzepts Thema findet sich meines Wissens nach nur bei de Beaugrande/Dressler, die das Thema (bzw. Topik) als zentralen Knoten eines konzeptuellen Netzes auffassen, also als Konzept mit der größten Verbindungsdichte zu anderen Konzepten. „Konzept“ wird dabei definiert als „Konstellation von Wissen (...), welches mit mehr oder weniger Einheitlichkeit und Konsistenz aktiviert oder wieder ins Bewusstsein gerufen werden kann“ (de Beaugrande und Dressler 1981, 89).<sup>79</sup> Über die von den sogenannten „Steuerungsmittelpunkten“ („strategisch günstige Punkte für Zugriff und Verarbeitung“) gesteuerten Aktivierung von Konzepten baut der Rezipient „Sinn-Kontinuität“ in einem Text auf. Bedauerlicherweise findet sich in der Arbeit von de Beaugrande/Dressler keine genauere Ausarbeitung des Konzepts Thema; insbesondere Unterschiede oder Gemeinsamkeiten mit dem Konzept des „Steue-

---

<sup>78</sup> Einen solchen Grenzfall stellt z.B. van Dijks Definition des Themas dar. Er definiert das Thema als kognitive Einheit (“themes are cognitive units”) und begründet dies damit, dass es eine sprecher- und hörerezentrierte Kategorie sei und als solche das repräsentiere, was Sprecher und Hörer in Abhängigkeit von ihren Wissensbeständen und Annahmen als relevanten Kern des Textes „verstehen“ bzw. konstruieren würden (vgl. van Dijk 1985, 76).

<sup>79</sup> Mit „Wissen“ werden „kognitive Inhalte aller Art“ bezeichnet (de Beaugrande und Dressler 1981, 89).

ungsmittelpunkts“ bleiben offen. Im Folgenden werden die kognitiven Ansätze in den Blick genommen, in denen das Thema als bekannte Information bzw. aktivierte Diskursentität definiert wird.

Das Merkmal „Bekanntheit“ ist sowohl semantisch bzw. referenziell (Bekanntheit als Rekurrenz, s.u.) als auch kognitiv (Bekanntheit als mentaler Aktivierungsgrad eines Diskursreferenten, s.u.) als auch kommunikativ-pragmatisch (Bekanntheit als Ergebnis der intentionalen Aufmerksamkeitssteuerung, siehe Grosz und Sidner 1986) erklärbar, d.h. es handelt sich um ein schwieriges Analysekriterium, da unter dem Begriff „bekannt“ sehr verschiedene Phänomene verstanden werden.

Traditionell wurde „Bekanntheit“ vor allem in Ansätzen in der Tradition der Prager Schule (Daneš 1970; Mathesius 1971 [1929]; Sgall et al. 1973; Firbas 1992) mit Thematizität assoziiert und als bipolares Merkmal „bekannt vs. neu“ der Thema-Rhema-Strukturierung im Satz zugeordnet.<sup>80</sup> Die grundlegende Annahme ist, dass im Satz sowohl neue als auch bereits bekannte Informationen mitgeteilt werden. Das Thema als bekannte Information wird hier also dem Rhema als neuer Information gegenübergestellt. Wie „Bekanntheit“ genau bestimmt werden kann, bleibt allerdings unklar: Mathesius (1971 [1929], 7) beschreibt „bekannt“ als „auf der Hand liegende Dinge, von denen der Sprecher ausgeht“ – wie diese Art von „Bekanntheit“ in der sprachlichen Analyse ermittelt werden kann, ist vollkommen offen. Daneš (1970, 592) kategorisiert „bekannt“ als kontextuellen Aspekt der Äußerung (d.h. als auf die vorhergehende Äußerungen bezogen), geht ansonsten aber nicht weiter darauf ein, wie „Bekanntheit“ genau bestimmt werden kann.

„Bekanntheit“ als kontextuelles Merkmal lässt sich ebenfalls unterschiedlich interpretieren – Prince zufolge gibt es drei unterschiedliche (kontextuelle) „Bekanntheits“-Konzepte: Givenness als „Predictability/Recoverability“ (Halliday 1967), Givenness als „Saliency“ (Chafe 1976) und Givenness als „Shared Knowledge“, wobei sie letztgenannten Ansatz als den auch für die beiden erstgenannten Konzepte grundlegenden bestimmt (vgl. Prince 1981, 226-232). Auf der Basis der beschriebenen Konzeptionen von Bekanntheit entwickelt sie ein Modell zur Differenzierung verschiedener Bekanntheitstypen (siehe Abbildung 2). Sie unterscheidet dabei zwischen neuen Diskursentitäten (wobei sie nur NPs als Diskursentitäten betrachtet), inferierbaren und evozierten Diskursentitäten. Bekanntheit wird demnach als Eigenschaft einer Diskursentität aufgefasst, die sich zwar zum einen aus dem linguistischen und extralinguistischem Kontext („textually Evoked“, „situationally Evoked“) ableiten lässt, die aber zum anderen auch das Ergebnis von kognitiven Inferenzprozessen der Rezipienten ist.

---

<sup>80</sup> Im Gegensatz zur Prager Schule unterscheidet Halliday zwischen den Konzeptpaaren „Thema/Rhema“ und „Bekannt/Neu“. Der Aspekt „Bekanntheit“ wird bei ihm nicht zur Bestimmung des Themas verwendet, sondern erhält einen unabhängigen Status im System der Informationsstruktur. Fries bezeichnet den Ansatz Hallidays deshalb als „splitting approach“, den der Funktionalen Satzperspektive als „combining approach“ (Fries 1995, 1).

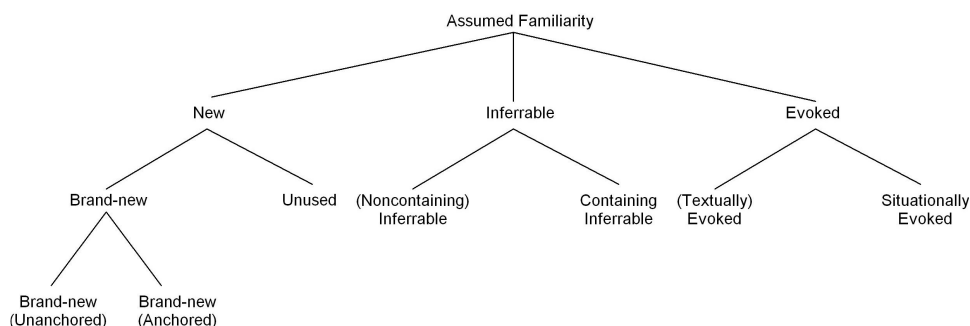


Abbildung 2: Taxonomie für das Konzeptpaar Given-New (vgl. Prince 1981, 237)

In anderen Ansätzen wird die Differenzierung von neuer vs. bekannter Information noch deutlicher an den kognitiven Status einer Diskursentität im mentalen Modell von Produzent und Rezipient gebunden. Hier wird nicht mehr davon geredet, dass eine Diskursentität „neu“ oder „bekannt“ sei, sondern „aktiviert“ vs. „nicht aktiviert“. Ein graduelles Modell zur Beschreibung des Status der Aktiviertheit von Diskursentitäten entwickelt Lambrecht (1994, 109) (siehe Abbildung 3). Lambrecht unterscheidet, ausgehend von Chafes (1987, 25ff.) Beschreibung von drei Aktivierungsstadien von Diskursentitäten im mentalen Modell von Produzent und Rezipient inaktive, aktive/gegebene und zugängliche („accessible“) Entitäten, wobei letztere entweder textuell, situativ oder inferentiell erschlossen werden können.

Das Thema wird dabei mit der aktiven Diskursentität assoziiert, allerdings nimmt Lambrecht im Gegensatz zu den Vertretern der Funktionalen Satzperspektive keine bipolare Zuordnung vor (Thema = aktiviert, Rhema = inaktiv), sondern eine graduelle: Auf einer Skala wird die ‚Eignung‘ der verschiedenen Aktivierungsstadien von Diskursentitäten, das Thema des Satzes zu repräsentieren, bewertet. Aktive Diskursentitäten werden dabei als akzeptabelste Satzthemen aufgefasst, neue, nicht verankerte Diskursentitäten als am wenigsten akzeptable (siehe Abbildung 4). Lambrecht begründet dies damit, dass Sätze, in denen das Thema am besten zugänglich bzw. aktiviert sei, am leichtesten kognitiv zu verarbeiten seien. Solche *aktiven* Diskursentitäten werden im Text durch spezifische grammatische Merkmale signalisiert:<sup>81</sup>

The cognitive category ‘activeness’ thus has grammatical correlates in PROSODY (phonological attenuation) and MORPHOLOGY (pronominal, inflectional, or zero coding). It also has correlates in syntax, but these are not as easy to demonstrate. (Lambrecht 1994, 95)

<sup>81</sup> Dies gilt nicht für zugängliche Diskursentitäten („accessible“): “As for the cognitive category accessible, it has no direct phonological or morphological correlates, though it may have indirect correlates in syntax. Accessible referents may be coded either like inactive or like active ones” (Lambrecht 1994, 107)

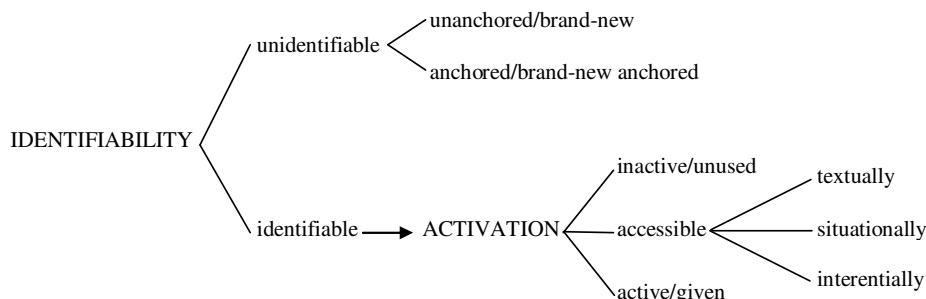


Abbildung 3: "System of identifiability and activation" (Lambrecht 1994, 109)

Aufgrund der typischerweise vorliegenden Korrelation von Thema und aktivierter Diskursentität lässt sich demnach auch das Thema selbst über diese grammatischen Eigenschaften bestimmen.

Dennoch muss festgehalten werden, dass die Bewertung von Satzkonstituenten als bekannt, aktiviert oder zugänglich nicht nur aus der Analyse des Textes selbst erschließbar ist<sup>82</sup>, sondern dass darüber hinaus auch die Einbeziehung und Analyse des situativen Kontextes sowie der Annahmen und Wissensbestände von Produzent und Rezipient nötig ist. Die Notwendigkeit einer solchen über den Text hinausweisenden Analyse von „Bekanntheit“ erschwert aber gleichzeitig die Operationalisierung der Analyse des Themas. Auf die Schwierigkeiten der Bestimmung des Themas über das Merkmal „bekannt“ weisen auch Schlo-

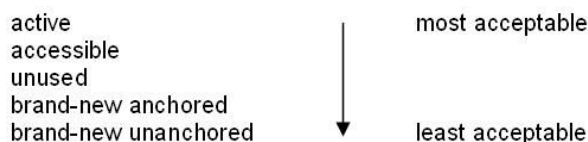


Abbildung 4: The Topic Acceptability Scale (Lambrecht 1994, 165)

<sup>82</sup> Allerdings gibt es durchaus Arbeiten, in denen versucht wird, bekannte vs. neue Diskursreferenten anhand ihrer grammatischen Merkmale voneinander zu unterscheiden, z.B. durch das Merkmal Definitheit bzw. Indefinitheit (z.B. bei Daneš 1970, 592). Die Aussagekraft dieses Merkmals ist jedoch zweifelhaft. So belegen Reinhart (1982, 12) und Lambrecht (1994, 105) anhand von Beispielen, dass auch indefinite NPs Thema des Satzes sein können. Während es für Lambrecht trotz dessen eine starke, wenn auch nicht absolute, Korrelation zwischen Indefinitheit und neuer Information gibt, ist beispielsweise Lernerz (1977, 52) der Ansicht, dass es „keinen zwingenden Zusammenhang zwischen Artikelwahl und Thematizität oder Rhematizität der entsprechenden NP“ gibt.

binski/Schütze-Coburn (1992, 94ff.) hin: In einem Satz sind häufig mehrere Diskursreferenten „bekannt“. Es kann aber nur eine Konstituente Thema des Satzes sein. Wenn man zusätzlich „inferierbare“ oder „zugängliche“ Diskursreferenten als grundsätzlich thematisierbar auffasst (wie z.B. Givón), potenziert sich das Problem.<sup>83</sup>

This obvious fact, [sic] that not only topics may represent old information, [sic] indicates at least that our knowledge of what counts as old information is not sufficient to explain how we identify the topic of a given sentence, e.g. how we choose between the two equally 'old' candidates for topichood. (Reinhart 1982, 19)

Zeitungsschlagzeilen oder allgemein Überschriften enthalten wiederum nur „neue Informationen“ – haben sie also kein Thema? Die Definition des Themas als „bekannte Information“ wird deshalb in verschiedenen Arbeiten als äußerst problematisch kritisiert und abgelehnt (siehe z.B. Reinhart 1982; Schlobinski und Schütze-Coburn 1992; Hoffmann 2000).

Dennoch ist das Kriterium der „Bekanntheit“ meiner Ansicht nach zentral für die ontologische Bestimmung des Konzepts Thema. Insbesondere die Frage nach dem Thema eines Textes lässt sich nur unter Bezugnahme auf diesen Aspekt beschreiben. In vielen Definitionen wird das Thema bestimmt als zentraler Gegenstand, Informationskern oder inhaltlicher Einordnungsrahmen, auf den sich alle Informationen im Text beziehen und aus dem sich der Zusammenhang zwischen den Informationen des Textes erklären bzw. ableiten lässt. Führt man diese Charakterisierung einen Schritt weiter, bedeutet dies, dass das Thema nicht nur als der mental am stärksten aktivierte und am besten zugängliche Gegenstand des Satzes, sondern auch des Textes bestimmt werden kann (siehe auch das Konzept von de Beaugrande und Dressler 1981: Thema als konzeptueller Kern des Textes). Das Merkmal der „Bekanntheit“ ist deshalb ein zentrales Verbindungsglied zwischen den Konzepten Satz- und Textthema.

## 2.5 Das Konzept Thema in computerlinguistischen Arbeiten

Das Konzept Thema spielt eine wichtige Rolle in den computerlinguistischen Anwendungen des Textzusammenfassens, des Abstracting, der Informationsextraktion (bzw. Topic Extraction), des Text Mining, der Textsegmentierung und der Textkategorisierung. In diesen Anwendungen geht es zunächst darum, Themen zu identifizieren. In einigen Arbeiten wird zusätzlich eine thematische Segmentierung des Textes vorgenommen, d.h. thematisch kohärente Passagen des Textes bestimmt, für die dann ein Thema benannt werden kann. Dieser Schritt ist jedoch nicht immer nötig (z.B. für die Informationsextraktion oder das automatische Textzusammenfassen); er wird deshalb häufig unabhängig von der Themenidentifizierung untersucht. In einem nächsten Schritt werden Themen interpretiert, d.h. zu Gruppen geordnet, die durch Themen (bzw. Konzepte) hö-

---

<sup>83</sup> Des Weiteren stellt sich die Frage, wie lange ein Diskursreferent „bekannt“ oder „inferierbar“ ist und ob sich dies z.B. an der Anzahl der Sätze zwischen Antezedent und Anapher messen lässt.

herer Ordnung bestimmt sind (vgl. Hovy und Lin 1999; Ferret und Grau 2002a).

Für die thematische Segmentierung von Texten spielt die Charakterisierung von Themen häufig eher eine untergeordnete Rolle, es geht primär darum, für übergeordnete Zwecke wie das Diskursparsing oder die Hypertextualisierung thematische Segmente zu bilden. Wie das Thema in diesen Segmenten benannt werden kann, ist zunächst nebensächlich. Deshalb werden im Kontext der thematischen Segmentierung häufig statistische Verfahren verwendet, in denen die semantische Ähnlichkeit von Textsegmenten berechnet wird (z.B. Text Tiling oder Latent Semantic Analysis). Beim Text Tiling wird der Text in gleich große Blöcke eingeteilt. Diese werden dann auf Basis von Wortwiederholungen und der Gewichtung von Wörtern miteinander verglichen, um so thematische Brüche zu ermitteln und den Text in thematisch kohärente Passagen einteilen zu können (Hearst 1994, 1997). Latent Semantic Analysis ist ein wesentlich komplexeres Verfahren, bei dem mittels mathematischer Analysen von Merkmalen die semantische Ähnlichkeit von Textpassagen berechnet wird (Landauer et al. 1998). In diesen Arbeiten wird vielfach stillschweigend vorausgesetzt, dass es zwischen einer semantischen und einer thematischen Analyse keine großen Unterschiede gibt. Ursache ist eine fehlende theoretische Auseinandersetzung mit dem Thema. Die statistischen Verfahren zur thematischen Segmentierung sind deshalb nicht weiterführend, wenn es um die Beschreibung computerlinguistischer Thema-Konzepte geht.

Ein alternativer, wortbasierter Ansatz zur thematischen Segmentierung beruht auf der Verteilung lexikalischer Ketten ("Lexical Chains") im Text (siehe auch Kapitel 3.3.1.3). Lexikalische Ketten bestehen aus lexikalisch-semantisch verwandten Wörtern, die eine thematische Einheit des Textes umspannen; "these lexical chains are a direct result of units of text being 'about the same thing'" (Morris und Hirst 1991, 21) – lexikalische Ketten können demnach zur Identifizierung von thematischen Einheiten im Text genutzt werden; im Gegensatz zu den o.g. Verfahren sind sie jedoch nicht nur zur thematischen Segmentierung verwendbar, sondern sie tragen auch zur Charakterisierung der Themen eines Textes bei: Jede Kette repräsentiert ein Thema oder Subthema; Textthemen werden häufig durch starke lexikalische Ketten repräsentiert, untergeordnete Themen durch weniger starke (Barzilay und Elhadad 1999, 14). Sie werden deshalb sowohl für die thematische Segmentierung als auch die Identifizierung und Benennung von Themen genutzt.<sup>84</sup>

---

<sup>84</sup> Je nach Algorithmus werden zur thematischen Segmentierung entweder Start- und Endpunkte der Ketten analysiert, um so Aufschluss über thematische Brüche zu erhalten (Morris und Hirst 1991; Stokes 2003): Zur Berechnung der Stärke von Segmentgrenzen wird die Anzahl der lexikalischen Ketten, die in Paragraph  $n$  enden, und die Anzahl der lexikalischen Ketten, die in Paragraph  $n+1$  beginnen, addiert. Auf diese Weise ergibt sich für alle adjazenten Paragraphen ein Wert zur Stärke ihrer Grenze; liegen die einzelnen Werte über dem Mittelwert, kann man davon ausgehen, dass es sich um eine semantische Segmentgrenze (sS) handelt. Alternativ können statt der Start- und Endpunkte auch komplette Ketten semantisch miteinander verglichen werden. Dazu wird die Wichtigkeit jeder Kette in jedem Absatz des Textes berechnet. Die so ermittelten "chain density"-Vektoren werden anschließend mittels eines Ähnlichkeitskoeffizienten

Für die Identifizierung von Themen kommen darüber hinaus verschiedene weitere Verfahren zur Anwendung: Einfache statistische Verfahren, die auf der Saliienz von Wörtern im Text beruhen (Luhn 1958); erweiterte statistische Verfahren, bei denen eine große Bandbreite von Oberflächenmerkmalen und lexikalischen Merkmalen Berücksichtigung findet, z.B. die Position von Wörtern oder Sätzen im Text (Edmundson 1969; Lin und Hovy 1997), Cue Phrases (Edmundson 1969) oder lexikalische Kohäsion (Barzilay und Elhadad 1999; Hovy und Lin 1999; Boguraev und Neff 2000b; Dahl 2000); wissensbasierte Verfahren, die mit Hilfe terminologischen oder ontologischen Domänenwissens die Themen des Textes auf die Domänenstruktur abbilden (Hovy und Lin 1999; Elhadad und McKeown 2001; Saggion und Lapalme 2002; Farzindar und Lapalme 2004); statistische und symbolische Verfahren, die auf Merkmalen der Diskursstruktur basieren. Solche Diskursstruktur-basierten Verfahren wurden in den letzten zehn Jahren vor allem zur Themenidentifizierung im Kontext des automatischen Textzusammenfassens betrachtet. Ihr wichtigster Vorteil ist, dass Themenwörter bzw. thematisch zentrale Sätze kontextualisiert und in die globale Textstruktur eingeordnet werden und so die Möglichkeit zu einer strukturierten und flexiblen Präsentation der thematisch zentralen Segmente des Textes gegeben ist (z.B. indem in der Textzusammenfassung die argumentative Struktur des Textes bewahrt und, wenn gewünscht, sogar hervorgehoben wird). Neben der genrespezifischen Texttypstruktur (Teufel und Moens 1999, 2002; Farzindar und Lapalme 2004) wurde vor allem die relationale Diskursstruktur bzw. rhetorische Diskursstruktur (Marcu 1997, 1999, 2000; Thione et al. 2004; Cristea et al. 2005) für die Themenidentifizierung genutzt.

Bei der Interpretation von Themen geht es darum, Generalisierungen von Themen vorzunehmen. Für diesen Schritt sind externe Wissensquellen notwendig, da z.B. nach gemeinsamen semantischen Hyperonymen gesucht wird oder nach übergeordneten Super-Konzepten (Hovy und Lin 1999; Saggion und Lapalme 2000; Andrews und Rajman 2004). Bei den genannten Ansätzen werden lexikalisch-semantische Wissensquellen (z.B. WordNet) bzw. Konzeptwörterbücher (oder Ontologien) hinzugezogen, um so Relationen zwischen Wörtern bzw. Konzepten zu ermitteln. Auf Basis dieser Relationen werden schließlich generalisierende Konzepte als Themenangaben erstellt. D.h. ein Satz wie „Sie kauften Stifte, Hefte, Radiergummis, Büroklammern.“ könnte generalisiert werden zu „Sie kauften Schreibwaren.“ Lin/Hovy (2000, 1) beschreiben die dem generalisierenden Konzept („Schreibwaren“) zu Grunde liegenden Konzepte („Stift“, „Heft“, „Radiergummi“ etc.) als „Topic Signatures“: „Topic signatures can be used to identify the presence of a complex concept – a concept that consists of several related components in fixed relationships.“ Indem im Text solche Topic Signatures bestimmt werden, können in einem zweiten Schritt die generelleren Konzepte als Repräsentationen des Themas instanziiert werden.

Explizite Themadefinitionen finden sich nur in sehr wenigen der genannten Arbeiten. In der Arbeit von Moens/Busser (2001) wird das Textthema beispielsweise

---

miteinander verglichen, um so die semantische Ähnlichkeit aller Absätze zu ermitteln (Green 1997).



se als “kernel of the content”, als inhaltlicher Textkern im Sinne von van Dijks Makroproposition definiert. Eine sehr allgemeine Definition von Thema als Gegenstand verwendet Lin (1995): “A topic is a particular subject that we write about or discuss.” Allan et al. (1998, 194) dagegen bestimmen das Thema in sehr spezifischer Weise als “event”, als “some unique thing that happens at some point in time”, allerdings hat das damit zu tun, dass sie “News Stories” analysieren, die ein Ereignis thematisieren. Die Definition ist demnach an das Genre gebunden. Nomoto/Matsumoto (1996, 102) bestimmen das Thema dagegen sehr simplifizierend und zweckgebunden. Sie gehen davon aus, dass das Thema über die Menge der Terme aus dem Titel charakterisiert werden kann. Ein derart definiertes Thema kann relativ einfach automatisch identifiziert werden: “What is involved in identifying a topic is a task of locating an instance of a title term” (Nomoto und Matsumoto 1996, 102). Nomoto/Matsumoto verkennen allerdings, dass der Titel neben der Angabe des Themas auch andere Funktionen haben kann und dass im Titel beispielsweise auch Gattungsangaben stehen können (vgl. Kapitel 3.3.1.4).

Trotz der größtenteils fehlenden Explikation des Konzepts Thema lässt sich aus den Darstellungen jedoch entnehmen, dass das Thema sehr häufig als Wort oder (kognitives) Konzept aufgefasst wird, das sich über folgende Merkmale bestimmen lässt:

- **Oberflächenmerkmale:** Thematische Ausdrücke lassen sich über Merkmale wie Position im Satz, Absatz und Text, Definitheit sowie durch lexikalische Cues bestimmen (Lin und Hovy 1997; Hovy und Lin 1999; Moens und De Busser 2001).
- **Häufigkeit/Salienz:** Themen sind NPs, die in einem lokalen Segment am salientesten sind (lokale Themen), bzw. NPs, die in mehreren Segmenten am salientesten sind (globale Themen) und auf die sich die lokalen Themen beziehen (Hernandez und Grau 2003); Themen können über häufig vorkommende Inhaltswörter ermittelt werden (Kupiec et al. 1995; Lawrie und Croft 2000; Lawrie et al. 2001); Themen sind häufig vorkommende Konzepte (oder Super-Konzepte) (Lin 1995; Lin und Hovy 2000).
- **Rekurrenz/Kookurrenz:** Themen sind im Text durch Gruppen von “key entities” präsent und können über Gruppen von miteinander verwandten Entitäten (d.h. Cluster von Schlüsselentitäten) beschrieben werden (Clifton et al. 2004); Themen sind charakterisierbar durch Gruppen semantisch oder lexikalisch verwandter, häufig vorkommender Deskriptoren bzw. Inhaltswörter (Ferret et al. 1998; Ferret 2002; Ferret und Grau 2002b).
- **Konzepthierarchie:** Konzeptuell ähnliche Wörter lassen sich als ein Konzept repräsentieren, verwandte Konzepte lassen sich als Super-Konzepte repräsentieren – die konzeptuelle Struktur wird genutzt, um Themenwörter/Inhaltswörter/Schlüsselwörter zu interpretieren und zu generalisieren, um so übergeordnete Textthemen zu ermitteln (Hovy und Lin 1999; Saggion und Lapalme 2000; Andrews und Rajman 2004).

Aus den Ausführungen in diesem Kapitel sollte deutlich geworden sein, dass die Thema-Konzepte in computerlinguistischen Arbeiten im Vergleich zu linguistischen Thema-Konzepten rein zweckorientiert sind und nicht auf einer theoretischen

schen Auseinandersetzung mit dem Konzept Thema beruhen. Eine theoretische Fundierung des Konzepts Thema findet sich in keiner der rezipierten computerlinguistischen Arbeiten, ebenso wenig, wie eine Diskussion um die „richtige“ Thema-Konzeption. Es wird stillschweigend vorausgesetzt, dass der Terminus Thema genügend etabliert ist, so dass eine explizite Definition unnötig ist. In der Mehrzahl der Arbeiten wird das Konzept Thema deshalb entweder gar nicht oder nur über die Aufzählung seiner Merkmale beschrieben. Eine kritische Sichtung der computerlinguistischen Arbeiten zur Themenidentifikation und -interpretation wirft deshalb die Frage auf, was an den identifizierten Ausdrücken und Konzepten eigentlich „thematisch“ ist. Diese Frage gewinnt an Schärfe, weitet man sie auf andere Verfahren und Bereiche, wie z.B. die Keyword-/Schlüsselwort-Identifikation bzw. -Extraktion, aus. Im Vergleich von Schlüsselwortidentifikation und Themenidentifikation, speziell jener computerlinguistischen Ansätze, in denen Themen als Cluster von Inhaltswörtern oder Schlüsselentitäten aufgefasst werden, d.h. als sprachliche Einheiten des Textes, sind nur marginale Unterschiede erkennbar, sowohl was die verwendeten Methoden als auch die Definitionen anbelangt: Schlüsselwörter/-phrasen werden beispielsweise aufgefasst als “simplex nouns or noun phrases (NPs) that represent the key ideas of the document” (Kim und Kan 2009, 9) oder als “essential content of a document” (Rose et al. 2010, 3). Insofern auch Themen als zentrale Ideen, semantische oder kognitive Ausgangspunkte bzw. Kerne aufgefasst werden, die wie Schlüsselwörter zur Textzusammenfassung genutzt werden können (“Keyphrases can serve as a representative summary of the document”, Kim und Kan 2009, 9), sind die Übergänge zwischen Themenidentifikation und Schlüsselwortidentifikation fließend (für einen Überblick über zentrale Verfahren zur Schlüsselwortidentifikation siehe z.B. Heyer et al. 2006, 99ff u. 278ff.).

Trotz der fehlenden Entwicklung eigener Thema-Konzepte findet sich erstaunlicherweise nur sehr selten ein Verweis auf linguistische Thema-Konzepte, obwohl diese ja in allen Variationen vorliegen. Diese fehlende Verzahnung von Textlinguistik und Computerlinguistik führt einerseits dazu, dass computerlinguistische Arbeiten zum Thema häufig wenig theoretisch fundiert sind, in ihrer Anwendung – der automatischen Themenanalyse – jedoch sehr ausgearbeitet und komplex sind. Andererseits führt dies aber auch dazu, dass in weiten Teilen der Linguistik angenommen wird, dass das Thema einer automatischen Analyse nicht zugänglich sei. In Kapitel 3.3 dieser Arbeit sollen die theoretischen linguistischen Fundierungen und praktischen computerlinguistischen Anwendungen deshalb miteinander verknüpft werden.

## Theoretische Grundlagen

In diesem Kapitel sollen die theoretischen Grundlagen für das in der vorliegenden Arbeit entwickelte integrative Thema-Konzept ausgearbeitet werden. Das Kapitel ist in drei Abschnitte unterteilt: Der erste Abschnitt (Kapitel 2.6) beschäftigt sich mit dem Konzept „Text“, der zweite (Kapitel 2.7) mit den kognitiven Grundlagen der Textproduktion und -rezeption und der dritte Abschnitt (Kapitel 2.8) wird sich mit den kommunikativen Prozessen der Textproduktion und -rezeption selbst auseinandersetzen. In der Zusammenfassung (Kapitel 2.9) wird schließlich ein komplexes Modell der Interaktionen zwischen den Ebenen der Kommunikation und Kognition (insbesondere den Wissensbeständen der Kommunikanten) vorgestellt.

In der vorliegenden Arbeit wird davon ausgegangen, dass Produzenten und Rezipienten in der interaktiven und dynamischen Auseinandersetzung mit ihren sowohl individuellen als auch gemeinsamen Wissensbeständen Texte produzieren bzw. rezipieren. Texte werden dabei aufgefasst als komplex strukturierte Äußerungseinheiten, mit denen eine sprachliche Handlung vollzogen wird. Sowohl die Äußerungseinheiten selbst als auch Verknüpfungen von Äußerungseinheiten können aus mehreren Perspektiven bzw. auf mehreren Ebenen beschrieben bzw. erklärt werden: Auf der Ebene der Informationsstruktur, der lexikalischen, referenziellen und relationalen Struktur sowie – im Hinblick auf globale Zusammenhänge von Äußerungseinheiten – auf der Ebene der generischen Texttypstruktur sowie der logischen bzw. visuellen Textstruktur. Äußerungseinheiten bzw. Elemente von Äußerungseinheiten können darüber hinaus als indexikalische Zeichen fungieren.<sup>85</sup> Sie können verwendet werden, um auf eine konventionalisierte kommunikative Situation Bezug zu nehmen. Umgekehrt evozieren sie durch ihr Vorkommen im Text eben diese Situation und unterstützen auf diese Weise den Prozess der Konstruktion von Kohärenz durch den Rezipienten. Die Äußerungseinheiten (sowie die (potentiellen) indexikalischen Zeichen) können deshalb als eine basale Grundlage der Kommunikation aufgefasst werden, die Wissensbestände der Kommunikanten stellen die zweite Grundlage dar. Im Akt der Kommunikation interagieren Wissen und Text; zentrale Schnittstelle zwischen diesen beiden Aspekten ist dabei das Thema bzw. die textorganisierende Handlung des Thematisierens.

---

<sup>85</sup> Der Begriff des indexikalischen Zeichens wird hier nicht im Sinne der traditionell als Indices aufgefassten sprachlichen Zeichen (wie z.B. deiktisch gebrauchte oder bestimmte referenziell gebrauchte Wörter wie Eigennamen) verwendet (vgl. Nöth 2000, 186), sondern im Anschluss an Antos (2009, 408) verstanden als „wahrnehmbares (oder ein im Hinblick auf begründete Erwartungen fehlendes) Phänomen (...) in einem mündlichen oder schriftlichen Kommunikationsakt, das als ein bedeutsames Zeichen nahegelegt oder interpretiert wird und aus dem verständnisrelevante Schlüsse vorzugsweise auf der Grundlage von Kontiguitäts- (Nachbarschafts-) Beziehungen vom Initianten und/oder von Rezipienten gezogen werden können.“

## 2.6 Texttheorie

### 2.6.1 Texte als komplexe sprachliche Handlungen

Texte stellen sprachliche Einheiten dar, mit denen kommuniziert wird. Sie werden in dieser Arbeit als komplex strukturierte Äußerungseinheiten aufgefasst, mit denen eine Texthandlung vollzogen wird (vgl. Schröder 2003, 32f.).<sup>86</sup> Diese Texthandlung stellt eine komplexe sprachliche Handlung dar, die sich aus mehreren aufeinander folgenden sprachlichen Teilhandlungen zusammensetzt. Sowohl der Zusammenhang von Handlung und Äußerung als auch der zwischen komplexer sprachlicher Handlung und Teilhandlung kann dabei als „indem“-Zusammenhang aufgefasst werden: Eine komplexe Handlung wird vollzogen, indem eine Folge von Teilhandlungen vollzogen wird (zwischen diesen aufeinander folgenden Handlungen besteht wiederum ein „und-dann“-Zusammenhang), indem eine Folge von Äußerungseinheiten realisiert wird (die wiederum realisiert werden, indem spezifische sprachliche Mittel eingesetzt werden) (vgl. Fritz 1994, 183ff.; Gloning 1994b, 117; Schröder 2003, 33ff.). Sprachliche Handlungen sind einerseits als geschlossene Einheiten zu betrachten – sie sind funktionale und thematische Einheiten –, andererseits lassen sie sich aber auch in Teilhandlungen bzw. Teilkonstituenten zerlegen. Diese Teilhandlungen, in die eine komplexe sprachliche Handlung aufgesplittet werden kann, sind wiederum funktional und thematisch aufeinander bezogen, d.h. sie sind durch thematische und funktionale Beziehungen verknüpft (vgl. Schröder 2003, 7ff. u. 41ff.).

Texte weisen also einerseits eine Konstituentenstruktur auf, insofern sich komplexe sprachliche Handlungen in Teilhandlungen zerlegen lassen. Darüber hinaus sind Texte aber auch durch relationale Verknüpfungen zwischen sprachlichen Teilhandlungen strukturiert (funktionale und thematische Relationen). Diese relationalen Verknüpfungen zwischen Teilhandlungen werden realisiert, indem Äußerungseinheiten oder Elemente der Äußerungseinheiten miteinander verknüpft bzw. aufeinander bezogen werden; auf der lokalen Ebene durch die Informationsstruktur, lexikalische, referenzielle und relationale Struktur, auf der globalen Ebene durch die generische Texttypstruktur sowie die logische und visuelle Textstruktur (siehe das Beschreibungsmodell der Textstruktur, Abbildung 5, S. 80). Die Äußerungseinheiten sind demnach die basalen Einheiten der Kommunikation (Text als Produkt): Zum einen signalisiert der Produzent durch sie, welche sprachlichen Handlungen er vollziehen will und in welchem Zusammenhang diese Handlungen zueinander stehen; zum anderen kann der Rezipient die Äußerungseinheiten aber auch als indexikalische Zeichen (im Sinne von Antos 2009, siehe auch Kapitel 3.1.2.2) interpretieren, die ihm beim

---

<sup>86</sup> In der vorliegenden Arbeit wird die Aussage „Texte sind komplex strukturierte Äußerungseinheiten, mit denen eine sprachliche Handlung vollzogen wird“ aus Gründen der Vereinfachung auf der Ausdrucksebene oft verkürzt zu „Texte als komplexe sprachliche Handlungen“ – dieser komplexe Ausdruck wird in der vorliegenden Arbeit also referenzidentisch zu o.g. Aussage bzw. Definition gebraucht.

Aufbau eines kohärenten mentalen Modells bzw. Situationsmodells des Textes helfen.

Zum Zusammenhang von Thema und Texthandlung lässt sich festhalten, dass das Thema als Gegenstand aufgefasst werden kann, auf den mit einer Handlung Bezug genommen wird (vgl. Schröder 2003, 78f.). Das Thema hat dabei folgende strukturelle Funktionen: 1. Das Thema konstituiert die Einheit des Textes, da das Thema der Gegenstand ist, auf den sich der Text als Ganzes bezieht. 2. Über das Thema werden die Teilhandlungen des Textes verknüpft. 3. In den Teilhandlungen werden jeweils spezifische Teilthemen behandelt. Diese Teilthemen werden im Folgenden als Aspekte des Textthemas bezeichnet. 4. Das Thema des Textes wird über seine Aspekte behandelt. Wie zwischen sprachlicher Handlung und Teilhandlung besteht also auch zwischen Thema und Aspekt ein „indem-Zusammenhang“: Das Thema wird behandelt, indem im Text seine Aspekte behandelt werden, d.h. als Gegenstände instanziiert werden, auf die jeweils mit einer sprachlichen Handlung Bezug genommen wird. Von zentraler Bedeutung für die Einheit des Textes ist, dass sprachliche Handlungen global und lokal thematisch verknüpft sind: Die Texthandlung insgesamt bezieht sich dabei auf ein übergeordnetes, alle Teilhandlungen verbindendes Thema, die Teilhandlungen beziehen sich jeweils auf einen Aspekt dieses Themas.

In den folgenden Kapiteln wird die These von Texten als Äußerungsfolgen, mit denen komplexe sprachliche Handlungen vollzogen werden, komplementiert durch die These von Texten als komplexen indexikalischen Zeichen und Texten als auf mehreren Ebenen strukturierten Äußerungseinheiten. Obwohl es aufgrund der gewählten Terminologie zunächst so scheinen mag, handelt es bei diesen drei Auffassungen meiner Ansicht nach keineswegs um unvereinbare handlungstheoretische, semiotische und textstrukturalistische Positionen. Handlungstheoretisch gesehen, werden Texte bzw. Äußerungen bewusst dazu verwendet, bestimmte sprachliche Handlungen und Teilhandlungen zu vollziehen. Dieser Aspekt der Intentionalität wird in der semiotischen Auffassung vom Text als komplexem indexikalischem Zeichen in gewisser Weise negiert – indexikalische Zeichen müssen nicht unbedingt intentional verwendet werden; ein Rezipient kann eine sprachliche Einheit als indexikalisches Zeichen interpretieren, ohne dass der Produzent dies beim Verfassen des Textes bewusst geplant hat. Die potentiell als indexikalische Zeichen interpretierbaren sprachlichen und visuellen Phänomene des Textes sind vielmehr eine implizite und nicht-intentionale Folge dessen, dass Produzenten sich in der Regel bemühen, einen verstehbaren, kohärenten Text zu produzieren, indem sie komplexe Äußerungseinheiten produzieren, mit denen bestimmte (intentionale) sprachliche Handlungen vollzogen werden. Zwischen den sprachlichen Einheiten auf der Ebene der Äußerungsstruktur bestehen dabei in der Regel Kontiguitätsbeziehungen (auch wenn diese nicht zwingend sind zur Konstruktion von Kohärenz), die Auslöser sein können für das Interpretieren sprachlicher Einheiten und Strukturen als indexikalische Zeichen, mit denen auf ein bestimmtes Schema bzw. ein bestimmtes Situationsmodell Bezug genommen wird. Texte bzw. komplexe Äuße-

rungseinheiten sind damit als sowohl intentionale als auch emergente<sup>87</sup> Gebilde aufzufassen.

Allen drei Charakterisierungen des Textes ist gemeinsam, dass für sie die kommunikative Situation ein zentraler Bezugspunkt ist. Sprachliche Handlungen sind bezogen auf kommunikative Erfahrungen und, damit einhergehend, auf konventionalisierte kommunikative Situationen. Insofern sprachliche Handlungen durch Äußerungseinheiten realisiert werden, wird der Bezug auf konventionalisierte kommunikative Situationen von Textproduzenten hergestellt durch die auf Ebene der Äußerungseinheiten wahrnehmbaren – d.h. auf der Textoberfläche sichtbaren – sprachlichen und visuellen Phänomene, die von Rezipienten wiederum als indexikalische Zeichen für eine bestimmte kommunikative Situation interpretiert werden können (aber nicht müssen). Diese Annahmen sollen in den folgenden Kapiteln im Detail erläutert werden.

## 2.6.2 Texte als komplexe indexikalische Zeichen

### 2.6.2.1 Zum Begriff der „kommunikativen Situation“

Sowohl Texte als komplexe sprachliche Handlungen als auch stereotypische Wissensbestände der Mitglieder einer Kommunikationsgemeinschaft lassen sich auf *kommunikative Situationen* rückbeziehen. Der Ausdruck „kommunikative Situation“ meint in diesem Zusammenhang mehr als nur die äußere „Kommunikationssituation“. Er bezieht sich vielmehr auf einen kommunikativen Diskurs zu einem bestimmten Thementyp, der auf eine spezifische Art und Weise sprachlich gestaltet ist (sowohl mikro- als auch makrostrukturell) und der an einen bestimmten Kommunikationskontext gebunden ist. Kommunikative Situationen lassen sich über komplexe sprachliche Handlungsmuster beschreiben:<sup>88</sup> „jemand [ein Kunde] bestellt über [das Telefon | das Internet] etwas [eine Ware] bei jemandem [einem Vertreter eines Unternehmens]“, „jemand unterhält sich mit jemand anderem über [das Wetter | ein Fußballspiel]“, „jemand [ein Unternehmer] hält zu einem bestimmten Anlass [zum Firmenjubiläum | zum Börsengang] eine Rede“, „jemand [ein Spieler | ein Lehrer] bringt jemandem [einem Neuling |

---

<sup>87</sup> „Im ‚klassischen‘ Sinne bedeutet Emergenz die Entstehung neuer Seinsschichten (...), die in *keiner* Weise aus den Eigenschaften einer darunter liegenden Ebene ableitbar, erklärbar oder voraussagbar sind. (...) In einer modernen Version spricht man von Emergenz, wenn durch mikroskopische Wechselwirkung auf einer makroskopischen Ebene eine neue Qualität entsteht, die nicht aus den Eigenschaften der Komponenten herleitbar (kausal erklärbar, formal ableitbar) ist, die aber dennoch allein in der Wechselwirkung der Komponenten besteht.“ (Krohn und Küppers 1992, 389, Hervorhebung im Original)

<sup>88</sup> Der Begriff der Situation reicht allerdings über den der sprachlichen Handlung hinaus – kommunikative Situationen stellen komplexe kommunikative Erfahrungen dar, die nicht nur durch spezifische Konfigurationen sprachlicher Handlungen gekennzeichnet sind, sondern auch durch spezifische sprachliche Mittel, spezifische Thementypen sowie einen bestimmten Kommunikationskontext, und die aufgrund ihrer Konventionalität und Routinisierung in das individuelle wie kollektive Sprachwissen bzw. Sprachverwendungswissen eingegangen sind. Der Begriff der kommunikativen Situation stellt darüber hinaus – wie im weiteren Verlauf der Arbeit noch gezeigt werden wird – einen zentralen Bezugspunkt zur Beschreibung der Interaktionen von Thema, Text und Wissen dar.

einem Schüler] etwas [ein Spiel | einen Unterrichtsinhalt] bei“, „der Meteorologe präsentiert die Wettervorhersage“.<sup>89</sup>

Durch wiederholte Erfahrungen mit und in kommunikativen Situationen lernen Kommunikanten Typen kommunikativer Situationen mit jeweils bestimmten Eigenschaften zu unterscheiden.

Die Fähigkeit, einen Wetterbericht, eine Festrede oder einen wissenschaftlichen Text zu identifizieren, beruht auf kommunikativen Erfahrungen mit Texten der betreffenden Domänen und Textsorten. (Steinhoff 2007, 100).

Wie Feilke (2003b) durch ein empirisches Textexperiment zeigt, können die meisten Muttersprachler auf Basis einer Auswahl von idiomatisch geprägten-Ausdrücken (z.B. „Durchzug von“, „weitgehend trocken“ oder „am Markt bestehen“, „vorbildliche Zusammenarbeit“, „Die Konkurrenz schläft nicht“, „erhebe mein Glas“) aus Texten verschiedener Textsorten erkennen, um welche Textsorte es sich handelt, und worum es im Text geht. Diese im Hinblick auf die Bezugnahme auf einen bestimmten Typus einer kommunikativen Situation konventionalisierten Ausdrücke sind dabei nicht auf Basis der absoluten Frequenz ihrer Verwendung als konventionell anzusehen, sondern auf Basis der relativen Frequenz ihrer Verwendung in einer bestimmten Textsorte (vgl. Bubenhofer 2009, 65), d.h. dass die stabile Verwendung eines Ausdrucks im Kontext einer bestimmten kommunikativen Situation für die Konventionalisierung bzw. Prägung dieses Ausdrucks eine zentrale Rolle spielt. Die verschiedenen Typen kommunikativer Situationen mit ihren jeweiligen Eigenschaften haben sich demnach durch stabile bzw. sich im Laufe der Zeit stabilisierende Routinen des Sprachgebrauchs, verstanden als „Routinen der Verständigung“, herausgebildet und sind Grundlage des konventionalen kommunikativen Wissens der Sprach- und Kulturgemeinschaft (vgl. Feilke 2003b, 209f.).

Da Menschen ständig neue Erfahrungen machen und diese dann die abgespeicherten Erfahrungen erweitern, aktualisieren und neu strukturieren, ist das Wissen über kommunikative Situationen dynamisch (vgl. van Dijk 2008, 69). Die Eigenschaften, die eine bestimmte kommunikative Situation ausmachen, sind demnach ebenfalls dynamisch. Trotz dieser Dynamik ist davon auszugehen, dass Kommunikanten in einer kommunikativen Situation in der Regel durch die Wahl ihrer sprachlichen Mittel auf vorangegangene Situationen Bezug nehmen, die zur aktuellen Situation in einer Relation der Äquivalenz stehen. Diese „Rückbezüglichkeit im Ausdrucksverhalten“ stellt eine notwendige Bedingung für die Verständigung zwischen den Kommunikanten dar (Feilke und Schmidt 1995, 289). Aktuelle und vorangegangene Situation lassen sich dabei aufgrund ihrer

---

<sup>89</sup> In eckigen Klammern sind Spezifizierungen der kommunikativen Situation angegeben – kommunikative Situationen unterscheiden sich im Hinblick auf die Spezifität von sozialen Rollen und Umständen. Einige Situationen sind im Hinblick auf die sozialen Rollen unspezifiziert (jemand unterhält sich mit jemand anderem über etwas), andere sind teilweise spezifiziert (jemand, der die Rolle „Kunde“ einnimmt bzw. in dieser Rolle vom Kommunikationspartner wahrgenommen wird, bestellt etwas) oder komplett spezifiziert: der Meteorologe (oder präziser: jemand, dem von den Rezipienten die Rolle des meteorologischen Experten zugeschrieben wird) präsentiert die Wettervorhersage (und nicht etwa die Nachrichten).

Äquivalenz als Mitglieder oder Varianten eines bestimmten Typs einer kommunikativen Situation charakterisieren. Typen (oder besser: Prototypen) kommunikativer Situationen zeichnen sich durch bestimmte Kerneigenschaften bzw. Merkmale aus, wobei kein Merkmal für sich als hinreichend oder notwendig für die kommunikative Situation betrachtet werden kann. Erst über das Zusammenspiel der verschiedenen Merkmale können Kommunikanten eine aktuelle kommunikative Situation als Ausdruck der Bezugnahme auf einen bestimmten Typ von kommunikativer Situation interpretieren (vgl. z.B. Auer 1986, 26). Dabei ist allerdings zu berücksichtigen, dass – unbewusst oder bewusst, z.B. als Spiel mit der Sprache – eine aktuelle kommunikative Situation mehr oder weniger stark von der prototypischen Situation abweichen kann.

Typen kommunikativer Situationen lassen sich auf folgenden Ebenen beschreiben:

1. Sprachebene: Auf der sprachlichen Ebene unterscheiden sich Typen kommunikativer Situationen durch ihre lexikalischen und semantischen Eigenschaften wie den Wortschatz (z.B. Verwendung von Fachwörtern vs. Allgemeinwörtern), lexikalische Explizitheit (z.B. nominale Rekurrenz vs. pronominale Verkettung), lexikalische und semantische Varianz (z.B. unterschiedliche Wortfelder), Kollokationen<sup>90</sup> (alltagssprachliche vs. fachsprachliche vs. domänenspezifische), idiomatische Muster (z.B. Sprichwörter vs. keine Sprichwörter, Redewendungen) sowie durch ihre morphologischen und syntaktischen Eigenschaften (z.B. Nominalstil vs. Verbalstil, Passivstil vs. Aktivstil).
2. Handlungs-/Interaktionsebene: Für jeden Typ einer kommunikativen Situation sind bestimmte Konfigurationen und Sequenzen sprachlicher Handlungen und Funktionen (z.B. Informieren, Kommentieren, Referieren, Argumentieren), Muster rhetorischer Relationen bzw. Kohärenzrelationen (z.B. ELABORATION, CAUSE), Arten der Verknüpfung von sprachlichen Handlungen (z.B. explizit über metasprachliche Formulierungen, grammatische Konnektoren, semantische, lexikalische oder referenzielle Relationen vs. implizit, d.h. schema- oder inferenzbasiert), Kommunikationsstile (z.B. formell vs. informell, konzeptionell mündlich vs. schriftlich, Verwendung rhetorischer Mittel wie Metaphern, Ironie etc. oder nicht) sowie texttypspezifische Textmuster<sup>91</sup> als typisch anzusehen.

---

<sup>90</sup> „Kollokationen sind als Zeichen bestimmt, und zwar als Zeichen relativ zu typischen Textkontexten.“ (Feilke 2003b, 214)

<sup>91</sup> Der Texttyp Kochrezept nimmt beispielsweise auf die kommunikative Situation Bezug „jemand [in der Rolle des Experten] instruiert jemanden [in der Rolle des Neulings], wie eine bestimmte Speise zuzubereiten ist“. Es ist anzunehmen, dass sich durch die ständige Wiederholung dieses Typus einer kommunikativen Situation bestimmte Muster herausgebildet haben, die sich für diese Situation (bzw. diese kommunikative Aufgabe) als besonders geeignet erwiesen haben. Für den Texttyp „Kochrezept“ hat es sich in den letzten Jahrzehnten etabliert, zunächst die Zutaten in Listenform anzugeben und anschließend die für die Zubereitung der Speise nötigen Handlungsschritte (als Instruktionen) in Form eines Fließtextes anzugeben.



3. Sachebene: Inhaltlich sind kommunikative Situationen durch das Vorkommen bestimmter Entitäten und Relationen zwischen diesen Entitäten gekennzeichnet (Sachverhaltskomplexe). Diese Entitäten lassen sich in Anlehnung an Kasusrollen kategorisieren in Zeit, Ort, Handelnder, Betroffener, Ziel, Instrument etc. Insofern in diesem Kontext nicht die durch das Verb induzierten Kasusrollen gemeint sind, sondern Rollen, die sich aus der kommunikativen Situation selbst ableiten, wird hierfür im Folgenden der Begriff der „Textrolle“ verwendet.<sup>92</sup>
4. Ebene des Kontextes:
  - a. außersprachliche Faktoren wie Zeit, Ort und soziokultureller Hintergrund; in Gesprächen außerdem Eigenschaften der Gesprächsteilnehmer, deren Verhältnis sowie die Rollen, die sie im Diskurs einnehmen bzw. die ihnen zugeschrieben werden,<sup>93</sup> darüber hinaus aktuelle Interessen, Probleme und Fragestellungen der Dialogpartner, in geschriebenen Texten dagegen beispielsweise die Zielgruppe, an die der Text gerichtet ist.
  - b. Diskurskontext, z.B. Art der medialen Vermittlung, Verhältnis zu anderen Texten zum gleichen Thema (kontroverse oder gleiche Meinung bzw. Argumentation) oder Stellung im Diskurs (Text, der einen sprachlichen Diskurs einleitet, oder Text, der einen Diskurs beendet, etc.).

Durch kommunikative Erfahrungen wird Wissen über kommunikative Situationen und ihre Eigenschaften erworben, kategorisiert und kognitiv abgespeichert. Zu diesem Wissen gehört auch das Wissen darüber, welche sprachlichen Mittel in einer bestimmten Situation als konventionell angemessen aufzufassen sind und welche Typen von Inhalten bzw. Textrollen in dieser Situation eine Rolle spielen können. Es wird davon ausgegangen, dass sich Kommunikanten in aktuellen kommunikativen Situationen auf dieses abgespeicherte Wissen über kommunikative Situationen beziehen: Sie kommunizieren auf Basis der Annahme, dass sie Erfahrungen zu Typen von kommunikativen Situationen teilen. Auf diese Erfahrungen zu einem bestimmten Typus einer kommunikativen Situation kann mittels sprachlicher Mittel Bezug genommen werden. Voraussetzung hierfür ist, dass diese sprachlichen Mittel konventionell geprägt sind und routinisiert verwendet werden können, d.h. dass sie ebenfalls zum Wissen über eine kom-

---

<sup>92</sup> Textrollen sind vergleichbar mit Kasusrollen. Im Gegensatz zu Kasusrollen sind Textrollen aber nicht auf die Ebene des Satzes, sondern des Textes bezogen. Das Verhältnis von Text- und Kasusrollen ist ähnlich wie das zwischen Konzept- vs. Verbframes – während sich die Verbframes und Kasusrollen aus der Valenz des Verbs ergeben, stellen Konzeptframes und Textrollen prinzipiell sprachunabhängige Konzeptualisierungen dar, die auf der Valenz einer kommunikativen Situation basieren und den Status von Argumentstellen im Sinne der Bestandteile/Argumente einer konzeptuellen Sachverhaltsstruktur haben, auf die diese kommunikative Situation bezogen ist (eine ähnliche, jedoch nicht identische Bestimmung von Textrollen findet sich in Rothkegel 1997, 466).

<sup>93</sup> Feilke weist darauf hin, dass Rollen in einem bestimmten Kontext nicht einfach nur eingenommen werden, sondern auch Ergebnis eines „konstruktiven, situationsdefinierenden Vorgangs“ sind, d.h. von den Kommunizierenden selbst durch ihre sprachlichen Handlungen konstruiert werden (Feilke 2003a, 50).

munikative Situation gehören (s.o.).

Die Annahme, dass sich Kommunikanten in aktuellen kommunikativen Situationen auf abgespeicherte kommunikative Situationen beziehen, bietet beispielsweise eine Erklärung dafür, warum in Texten Auslassungen nicht die Ausnahme, sondern die Regel darstellen (vgl. auch Grice (1979) Maxime der Quantität: Sei informativ – aber nicht informativer als nötig!). Der berühmte indirekte Sprechakt „Es zieht!“ wird unzweideutig als Aufforderung verstanden, das Fenster zu schließen, ohne dass diese Aufforderung selbst explizit gemacht werden muss.<sup>94</sup> Auch Hintergrundinformationen, die inferiert werden können, werden in Texten in der Regel im Sinne der Ökonomie des Sprachgebrauchs zugunsten wichtigerer Informationen weggelassen.

(...) each day, facing the same daily communicative goals and conditions, people activate the same routine contexts, allowing them to focus on what is currently unique, important and relevant (...) or what is problematic or troublesome in the communicative event. (van Dijk 2008, 69)

Aufgrund der Erfahrungen mit kommunikativen Situationen, die als konventionales Wissen abgespeichert sind, können Auslassungen diverser Art von Rezipienten problemlos ergänzt werden.

Produzenten verfassen Texte demnach in einer bestimmten Art und Weise, um damit auf eine bestimmte kommunikative Situation Bezug zu nehmen, während umgekehrt Rezipienten die Erwartung haben, dass in einem Text auf eine ihnen bekannte kommunikative Situation (bzw. einen Typus von Situation) Bezug genommen wird. Sprache wird nach dieser Auffassung also immer in Bezug zu einer kommunikativen Situation mit Personen, Handlungen, Ereignissen erlebt und verwendet. In Bezug auf diese Situation (bei Halliday: „scenario“) erst erhält Sprache Bedeutung (vgl. Halliday 1978, 28).

### 2.6.2.2 Indexikalische Zeichen

Sprachliche Ausdrücke (als Teil linguistischer Strukturen) können als „Erinnerungsmittel“ (Wegener 1991 [1885], 100), „contextualization cues“ (Gumperz 1982, 131) bzw. „Kontextualisierungshinweise“ (Auer 1986, 24) oder Zeichen mit indexikalischem Charakter (Ortner und Sitta 2003, 23; Steinhoff 2007, 99) interpretiert werden, die zur Bezugnahme auf (kommunikative) Situationen<sup>95</sup> verwendet werden bzw. durch die bekannte Situationen angedeutet werden. Gemeinsam ist allen drei Begriffen, dass eine Relation zwischen einem Oberflächenmerkmal (sprachliche Ausdrücke, aber auch nicht-sprachliche Faktoren der Kommunikationssituation, z.B. visuelle Eigenschaften) und einem Schema (ei-

<sup>94</sup> Ein anderes Beispiel für eine Auslassung ist: „Ihr Baby war krank. Anna fuhr zum Arzt.“ In der Regel werden Rezipienten problemlos verstehen, dass Anna *wegen* der Krankheit des Babys zum Arzt fährt, und dass sich nicht Anna vom Arzt untersuchen lassen will, sondern dass das *Baby* untersucht werden soll.

<sup>95</sup> Im Unterschied zu soziolinguistischen oder sozio-kognitiven Theorien wird der Begriff der Situation in der vorliegenden Arbeit etwas enger gefasst, insofern „Situation“ als *kommunikative Situation* verstanden wird, d.h. als sprachliche Situation, und nicht als Situation im Sinne von außersprachlicher Kontext.

nem Gebrauchs- oder Ereignisschema, das eine kommunikative Situation repräsentiert) angenommen wird, die auf dem Prinzip des *pars pro toto* (*pars* = Oberflächenmerkmal; *toto* = Schema) basiert (vgl. z.B. Feilke 2003a, 51; Steinhoff 2007, 99). Diese Relation stellt jedoch keine stabile Relation dar, sondern wird erst durch den Kommunikanten selbst – unter Rekurs auf sein abgespeichertes Wissen – konstituiert. Die genannten Begriffe basieren demnach nicht auf einem „aliquid stat pro aliquo“-Zeichenbegriff (vgl. Nöth 2000, 137), sondern auf einem triadischen Zeichenbegriff, bei der die Relation zwischen Zeichen und Objekt als indirekte Relation gesehen wird. Ein Zeichen wird nicht per se als Index, Symptom oder Ikon für ein Objekt betrachtet, vielmehr wird die Relation zwischen Zeichen und Objekt (bzw. kommunikativer Situation) erst durch den Zeichenbenutzer selbst hergestellt, indem dieser etwas als Zeichen für etwas (ein Objekt, eine kommunikative Situation) interpretiert (vgl. Nöth 2000, 139 u. 185).

Insofern alle drei genannten Begriffe („Erinnerungsmittel“, „Kontextualisierungshinweis“, „Zeichen mit indexikalischem Charakter“ bzw. „indexikalisches Zeichen“) zur Beschreibung des gleichen Phänomens verwendet werden – Kommunikanten nehmen mittels sprachlicher und außersprachlicher Mittel auf bekannte kommunikative Situationen Bezug – weisen diese Begriffe deutliche Bedeutungsüberschneidungen auf. Der Begriff des indexikalischen Zeichens ist allerdings weiter zu fassen als der des Kontextualisierungshinweises oder des Erinnerungsmittels: Nicht jedes indexikalische Zeichen stellt einen Kontextualisierungshinweis im Sinne von Gumperz und Auer oder ein Erinnerungsmittel im Sinne von Wegener dar, aber Kontextualisierungshinweise und Erinnerungsmittel sind immer auch indexikalische Zeichen, insofern sie „*pars pro toto*“ verschiedene Schemata *indizieren* (vgl. Steinhoff 2007, 99). Beim Prozess der Kontextualisierung wird eine zeichenhafte Beziehung zwischen einem Oberflächenmerkmal und einem kognitiven Schema etabliert, wobei diese Zeichenbeziehung von der traditionellen Beziehung von *signifiant* und *signifié* abweicht, insofern durch einen Kontextualisierungshinweis keine semantische Bedeutungsbeziehung etabliert wird – ein Kontextualisierungshinweis trägt an sich keine Bedeutung –, sondern ein kognitives Schema indiziert wird (vgl. Gumperz 1982, 131; Auer 1986, 25). Kontextualisierungshinweise steuern demnach die Verständigung, indem sie einen Verständigungskontext mit dem hierzu gehörenden Hintergrundwissen aktivieren.

Von zentraler Bedeutung sind hierbei zwei Aspekte: Zum einen ist davon auszugehen, dass nicht nur die semantische Bedeutung eines Ausdrucks oder das kognitive Konzept, auf das mit einem sprachlichen Ausdruck Bezug genommen wird, als Kontextualisierungshinweis fungiert, sondern dass vor allem die Ausdrucksoberfläche selbst einen bestimmten Kontext bzw. einen konventionalisierten Verwendungszusammenhang indiziert und damit ein bestimmtes Hintergrundwissen aktiviert, insofern die Oberfläche selbst „Signalkraft“ hat (von verschiedenen Formen von Begrüßungsfloskeln, über den Gebrauch von Dialekt vs. Standardsprache bis hin zu grammatischen und morphologischen Strukturen

sowie Textstrukturen) (vgl. Feilke und Linke 2009, 6ff.).<sup>96</sup>

„Das Verstehen von sprachlichen Handlungen und Texten hängt wesentlich ab von ihrer Formulierung.“ Und: „Der Ausdruck, das heißt die Oberfläche in ihrer Prägung, ist Erkennungszeichen und Voraussetzung gelingender Verständigung“ (Feilke 2003b, 218).

Zum zweiten ist von Bedeutung, dass die Interpretation eines Ausdrucks als Kontextualisierungshinweis abhängig davon ist, dass er überhaupt als ein auf ein bekanntes Schema Bezugnehmender Ausdruck erkannt wird, dass er im Hinblick auf die Situation als erwartbar betrachtet werden kann.

Interpretation (...) requires first of all judgements of expectedness and then a search for an interpretation that makes sense in terms of what we know from past experience and what we have perceived. (Gumperz 1982, 170).

Dies setzt eine gewisse Regelmäßigkeit und Konventionalität in der Relation zwischen Verwendungssituation und sprachlichem Ausdruck voraus. Erst durch diese Regelmäßigkeit kann aus den Routinen des Sprachgebrauchs ein äquivalentes Sprachgebrauchswissen entstehen. Grundlage für die Interpretation von Kontextualisierungshinweisen ist demnach, dass Kommunikanten ein gemeinsames Wissen darüber besitzen, mit welchen sprachlichen und außersprachlichen Merkmalen in der Regel auf einen bestimmten Kontext, ein bestimmtes kognitives Schema Bezug genommen wird (Produktionsperspektive), und umgekehrt auf welche konventionale Situation mit einem spezifischen sprachlichen Ausdruck bzw. mit spezifischen sprachlichen Strukturen Bezug genommen wird (Rezeptionsperspektive). Erst durch Konventionalität (Sprachgesellschaft) und Gebrauchswissen (Individuum), d.h. durch gemeinsames Wissen, werden Kommunikanten in die Lage versetzt, Sprache als Erinnerungsmittel zu verstehen und sprachliche Ausdrücke als Kontextualisierungshinweise für einen bestimmten konventionalisierten Kontext zu interpretieren. Trotzdem soll an dieser Stelle noch einmal betont werden, dass Kontextualisierungshinweise keine vom Kommunikationskontext losgelöste, inhärente Bedeutung tragen und vom Produzenten auch nicht unbedingt bewusst, d.h. intentional als Zeichen zur Signalisierung einer Situation gebraucht werden. Indizien bzw. Kontextualisierungshinweise müssen nicht in einer Verursacherrelation oder Mittel-Zweck-Beziehung stehen – erst durch die Interpretation eines Kommunikanten wird etwas zum Indiz:

Indem der Beobachter *etwas* (einen wahrnehmbaren Aspekt der Text-Performanz) *als etwas* (Zeichen) identifiziert, kann das beobachtete Objekt für den Beobachter zu einem Indiz für etwas werden. (Antos 2009, 421, Hervorhebungen im Original)

In der „Lexical Priming“-Theorie (Hoey 2004) wird angenommen, dass Ausdrücke in Verbindung zu anderen Ausdrücken stehen, weil sie durch ihren konven-

---

<sup>96</sup> Allerdings ist festzuhalten, dass Kontextualisierungshinweise und indexikalische Zeichen die Symbolizität sprachlicher Zeichen durchaus nutzen, d.h. nicht nur die Zeichenoberfläche bzw. Formseite des Zeichens für die Bezugnahme auf einen bestimmten Kontext/einen Verwendungszusammenhang/eine kommunikative Situation verwenden, sondern auch den arbiträren, konventionalisierten Zusammenhang zwischen Zeichenform und Bedeutung, d.h. das Zeichen als Symbol.

tionalisierten Gebrauch in der Kommunikation von den Kommunikanten als miteinander verbunden aufgefasst werden. Während in der Semantik die ‐Priming‐-Effekte<sup>97</sup> von Wörtern ausschließlich auf ihre semantische Verwandtschaft zurückgeführt werden, geht Hoey davon aus, dass ‐Priming‐-Effekte nicht nur durch semantische oder phonologische Verwandtschaft entstehen, sondern dass Wörter durch ihren Gebrauch in der Kommunikation assoziiert werden bzw. ‐geprimed‐ sind mit Wörtern, die in Form von Kollokationen gemeinsam mit ihnen auftreten:

(...) each lexical item is primed for collocational use. By primed, I mean that as the word is learnt through encounters with it in speech and writing, it is loaded with the cumulative effects of those encounters such that it is part of our knowledge of the word that it co-occurs with other words. (Hoey 2004, 386)

(...) learning a lexical item entails learning what it occurs with and what grammar it tends to have. (Hoey 2004, 387)

Kommunikanten lernen demnach Wörter in ihren spezifischen Verwendungskontexten kennen, was bedeutet, dass sie sowohl die Kollokationen eines Wortes in verschiedenen Kontexten erlernen als auch deren semantische Assoziationen (z.B. Wortfelder) und grammatische Eigenschaften. Diese zu einem Wort abgespeicherten Verbindungen und Eigenschaften können sich, je nach Verwendungskontext des Wortes bzw. je nach Domäne und Textsorte, in denen das Wort verwendet wird, unterscheiden (vgl. Hoey 2004, 386f.).

Wie bereits weiter oben angesprochen, können Kontextualisierungshinweise als indexikalische Zeichen aufgefasst werden, aber nicht unbedingt umgekehrt. In der vorliegenden Arbeit wird der ursprünglich auf Peirce zurückgehende Begriff des ‐indexikalischen Zeichens‐ in einem weiten Sinne verwendet. Ein Indiz oder indexikalisches Zeichen wird im Anschluss an Antos (2009, 408f.) verstanden als

wahrnehmbares (oder ein im Hinblick auf begründete Erwartungen fehlendes) Phänomen (...) in einem mündlichen oder schriftlichen Kommunikationsakt, das als ein bedeutsames Zeichen nahegelegt oder interpretiert wird und aus dem verständnisrelevante Schlüsse vorzugsweise auf der Grundlage von Kontiguitäts- (Nachbarschafts-)Beziehungen vom Initianten und/oder von Rezipienten gezogen werden können. (Antos 2009, 408).<sup>98</sup>

Indexikalität ist damit eine Eigenschaft aller wahrnehmbaren Phänomene. Wie bei den Kontextualisierungshinweisen ist auch für das Erkennen von Indizien eine gewisse Regelmäßigkeit oder Konventionalität der Kontiguitätsbeziehungen und damit die kognitive Verfügbarkeit dieser Beziehungen nötig – als ‐individuelles Wissen und Können einerseits, vor allem aber ein Nutzen und Unterstellen

<sup>97</sup> Durch experimentelle Untersuchungen konnte gezeigt werden, dass Wörter, die in einer semantischen (oder phonologischen!) Relation zueinander stehen, offenbar auch im Gedächtnis nah beieinander abgespeichert sind. Durch das Perzipieren eines Wortes werden deshalb gleichzeitig die in der Nähe abgespeicherten Wörter mitaktiviert – dieser Effekt wird als Priming bezeichnet (vgl. Schwarz und Chur 2004, 74f.)

<sup>98</sup> Der Begriff des indexikalischen Zeichens wird hier also nicht im Sinne der traditionell als Indices aufgefassten sprachlichen Zeichen (wie z.B. deiktisch gebrauchte oder bestimmte referenziell gebrauchte Wörter wie Eigennamen) verwendet (vgl. Nöth 2000, 186).

von kollektivem Wissen andererseits“ (Antos 2009, 422). Kontiguität wird dabei verstanden als

Assoziationsprinzip, das auf der gehäuften Wahrnehmung von zeitlichen, räumlichen, kausalen und anderen (z.B. pars pro toto) Nachbarschaftsbeziehungen („Zusammenbefindlichkeiten“) beruht (Antos 2009, 408).

Während Feilke (2003b, 214) und Hoey (2004, 386) den Aspekt der konventionalisierten und damit auf kommunikativen Erfahrungen basierenden Zusammenhänge von Wörtern betonen, spezifiziert Antos diese Zusammenhänge und führt hierfür die Relation der Kontiguität ein. Insofern diese Relation auf der *Wahrnehmung* assoziativ verknüpfter Phänomene beruht, handelt es sich ebenfalls um eine Relation, die auf kommunikativen Erfahrungen basiert und nicht etwa um eine lexikalische oder semantische Relation. Um die Bedeutung der individuellen und kollektiven kommunikativen Erfahrungen für die Interpretation indexikalischer Zeichen noch mehr hervorzuheben, soll an dieser Stelle Antos Definition des indexikalischen Zeichens etwas abgewandelt werden: Das indexikalische Zeichen wird bestimmt als wahrnehmbares (oder im Hinblick auf begründete Erwartungen fehlendes) Phänomen, das im Kontext anderer wahrnehmbarer Phänomene auftritt, die auf Grundlage der kommunikativen Erfahrungen mit Situationen und des auf diesen Erfahrungen basierenden Sprachgebrauchswissens vom Produzenten und/oder Rezipienten als assoziierte Phänomene erkannt werden, und das im Zusammenspiel mit diesen anderen Phänomenen vom Produzenten und/oder Rezipienten als ein bedeutsames Zeichen interpretiert wird, indem es zur Bezugnahme/zur Ziehung von Schlüssen auf die kommunikative Situation genutzt wird.

### 2.6.2.3 Zur Semiose einfacher und komplexer indexikalischer Zeichen

Alle wahrnehmbaren – und möglicherweise auch bestimmte fehlende – Phänomene können als indexikalisches Zeichen interpretiert werden, um auf die kommunikative Situation zu schließen. Nicht alles Wahrnehmbare wird in der konkreten Kommunikationssituation allerdings tatsächlich als Indiz semiotisiert, d.h. als Zeichen genutzt – dies ist allein schon durch die Textproduktion wie -rezeption leitende Prinzip der Ökonomie geboten. Nach Antos stellt sich deshalb die Frage, was als Indiz wahrgenommen werden soll, und was nicht, und warum (vgl. Antos 2009, 422f.).

Indexikalische Zeichen sind inhärent vage; erst durch ihr Zusammenspiel bzw. die Konkurrenz mit anderen indexikalischen Zeichen können sie von einem Rezipienten zu einem verständnisrelevanten Schluss genutzt werden. Indexikalische Zeichen werden demnach erst in der Verbindung mit anderen indexikalischen Zeichen zu einem komplexen Indiz für eine bestimmte kommunikative Situation. Im Folgenden sollen deshalb folgende Fragen beantwortet werden: 1. Welche Arten indexikalischer Zeichen können unterschieden werden? 2. Wie unterscheiden sich diese Zeichen im Hinblick auf ihre Bedeutung für die Bezugnahme auf eine bestimmte kommunikative Situation (wie wird aus einer Menge (potentieller) einfacher indexikalischer Zeichen ein komplexes indexikalisches Zeichen)?

Insofern es sich bei den potentiellen indexikalischen Zeichen um *wahrnehmbare* Phänomene der aktuellen Kommunikationssituation handelt, müssen sie im Text auf Ebene der Textoberfläche sichtbar sein. Wie im folgenden Kapitel näher dargestellt, können Texte als auf mehreren Ebenen strukturierte Äußerungseinheiten aufgefasst werden. In der vorliegenden Arbeit werden Äußerungseinheiten als materiale Seite des Textes aufgefasst, die den Input für eine bottom-up-verlaufende Textverarbeitung liefert. Dieser Ebene des Textes als Produkt (Text als materialer Ausgangspunkt für die manuelle sowie automatische Textverarbeitung) stehen die Ebenen der Kognition und Kommunikation als Ebenen des Textes im Gebrauch gegenüber. Der Text als Produkt weist sowohl eine lokale als auch globale Strukturierung auf. Auf der lokalen Ebene wird unterschieden zwischen der Informationsstruktur, lexikalischen, referenziellen und relationalen Struktur, auf der globalen Ebene zwischen der generischen Texttypstruktur und der visuellen sowie logischen Textstruktur (siehe Kapitel 2.6.3). Indexikalische Zeichen lassen sich nun analog zu diesen Ebenen unterscheiden in informationsstrukturelle, lexikalische, referenzielle, relationale, texttypspezifische, visuelle sowie logische Zeichen. Diese Arten indexikalischer Zeichen beziehen sich damit in erster Linie auf die Eigenschaften der kommunikativen Situation auf der Sprach- und Handlungsebene (siehe Kapitel 2.6.2.1), indirekt auch auf die inhaltliche Ebene, insofern die genannten Ebenen der Äußerungsstruktur auch Indizien zur Bestimmung des Themas enthalten. Da sich die folgenden Kapitel intensiv mit den genannten Ebenen der Äußerungsstruktur von Texten beschäftigen werden, soll es an dieser Stelle bei diesen wenigen Bemerkungen zu den Arten indexikalischer Zeichen bleiben.

Der zweite Fragenkomplex zur unterschiedlichen Bedeutung indexikalischer Zeichen für die Bezugnahme auf kommunikative Situationen, dem Zusammenspiel einfacher indexikalischer Zeichen und der Frage danach, wie aus einer Menge (potentieller) einfacher indexikalischer Zeichen ein komplexes indexikalisches Zeichen wird, ist deutlich schwieriger zu beantworten als die erste Frage. Es ist zunächst einmal davon auszugehen, dass kein wahrnehmbares Phänomen für sich allein notwendig und hinreichend für die Bezugnahme auf eine Situation ist. Erst im Zusammenspiel mit anderen wahrnehmbaren Phänomenen wird ein einzelnes und für sich genommen *einfaches indexikalisches Zeichen* zu einem verständnisrelevanten, bedeutungskonstituierenden, *komplexen indexikalischen Zeichen*. Mit der Unterscheidung von einfachen und komplexen indexikalischen Zeichen wird der Tatsache Rechnung getragen, dass wahrnehmbare Phänomene im Text zunächst immer als vereinzelte Phänomene wahrgenommen werden (der physikalisch-optische Prozess der Rezeption erfolgt sequenziell), und dass erst sukzessive, im Fortschreiten der Rezeption auch andere, im gleichen Verwendungskontext stehende sprachliche oder visuelle Phänomene in den Blick geraten und miteinander „verrechnet“<sup>99</sup> werden:

---

<sup>99</sup> Der Ausdruck „verrechnen“ meint ein „Sinnproduzieren, sprechend, schreibend und/oder verstehend, das sowohl holistisch wie elementaristisch ist; das gleichzeitig das Ganze und die Teile aufeinander bezieht.“ (Ortner und Sitta 2003, 23).

Doch wie die Zeichnung mehr ist als die Summe der Punkte, so ist auch die Äußerung / der Text mehr als eine Folge von Morphemen oder Wörtern oder auch als eine Folge von Sätzen – arrangiert nach den Gesetzmäßigkeiten einer Textsorte. Die Leistungen der Elemente addieren sich nicht zu einer Gesamtleistung, sondern: aus der Kombination der Elemente ergeben sich neue Leistungen, die von den Teilen (als Teilen) nicht erbracht werden können. (Ortner und Sitta 2003, 17)

Einfache indexikalische Zeichen sind demnach alle wahrnehmbaren Phänomene, die ein Kommunikant als potentiell bedeutungskonstituierend erkennt, während komplexe indexikalische Zeichen tatsächlich bedeutungskonstituierend sind. Das Konzept der komplexen indexikalischen Zeichen resultiert aus der Annahme, dass Kommunikanten die Menge der einfachen indexikalischen Zeichen in eine verständnisrelevante Ordnung bringen müssen, d.h. selektieren, gewichten, Zusammenhänge herstellen müssen, um aus den unterschiedlichen potentiell bedeutsamen indexikalischen Zeichen überhaupt einen sinnvollen Bezug auf eine spezifische kommunikative Situation konstruieren zu können. Dabei werden diejenigen indexikalischen Zeichen, für die Kommunikanten erkennen, dass sie in einer Kontiguitätsbeziehungen stehen, die in einer bestimmten kommunikativen Situation regelmäßig gemeinsam verwendet werden, privilegiert behandelt und in der Interpretation ihres Zusammenspiels – durch ihre Ordnung – für einen bedeutungskonstituierenden Schluss genutzt. Die im vorigen Kapitel eingeführte Definition indexikalischer Zeichen bezieht sich demnach auf komplexe indexikalische Zeichen. Meiner Ansicht nach ist es jedoch nötig zur Kenntnis zu nehmen – und dies auch terminologisch zu reflektieren –, dass ein komplexes indexikalisches Zeichen immer Folge eines Vorverarbeitungsprozesses ist und nie am Beginn dieses Prozesses steht.

Um als komplexes indexikalisches Zeichen verwendet zu werden, muss etwas sprachlich oder visuell Wahrnehmbares in der individuellen wie kollektiven Kommunikationsgeschichte also in einem gemeinsamen kommunikativen Kontext verwendet worden sein und auf Basis dieser konventionellen Verwendung in das Sprachgebrauchswissen übergegangen sein – und zwar entweder als Klasse oder als Entität, als Type oder Token, als abstraktes strukturelles oder visuelles Muster oder als konkrete Instanz eines solchen Musters. Die Interpretation als komplexes indexikalisches Zeichen, das auf eine kommunikative Situation verweist, beruht demnach vor allem auf der Konventionalität und damit Erwartbarkeit einer Menge wahrnehmbarer sprachlicher oder visueller Phänomene im Kontext einer bestimmten kommunikativen Situation.

An diesem Punkt stellt sich darüber hinaus die Frage, ob sich indexikalische Zeichen in ihrer Bedeutung für die Bezugnahme auf eine kommunikative Situation unterscheiden, ob es also eine Hierarchie indexikalischer Zeichen gibt. Insofern Rezipienten bereits nach den ersten konsumierten Äußerungseinheiten Hypothesen aufbauen, worum es im Text geht, an welche bekannte kommunikative Situation der Text anschließbar ist, ist davon auszugehen, dass es eine auf der Sequenz der sprachlichen Einheiten basierende Hierarchie gibt, d.h. dass die ersten Äußerungseinheiten (zu denen u.a. der Titel des Textes zählt) eine größere Bedeutung für die Aktivierung eines Schemas zu einer kommunikativen Situation aufweisen, als Äußerungseinheiten, die in der Mitte des Textes vorkommen.



Darüber hinaus ist zu vermuten, dass die visuelle Erscheinung des Textes ebenfalls zu den ersten wahrgenommenen Phänomenen des Textes zu rechnen ist und insofern ebenfalls zentral ist für den kommunikativen Anschluss des Textes.

Im Folgenden sollen die Ebenen der Äußerungsstruktur des Textes, die als Grundlage für die Differenzierung der verschiedenen Arten indexikalischer Zeichen dienen, näher erläutert werden.

### **2.6.3 Texte als auf mehreren Ebenen strukturierte Äußerungseinheiten**

Textlinguisten sind sich darüber einig, dass Texte auf mehreren Ebenen strukturiert sind. Gansel/Jürgens schreiben hierzu:

Wie ein Sprechakt aus unterschiedlichen simultanen Teilakten besteht (Lokution, Proposition, Illokution, Perlokution), lässt sich auch der Text als eine Struktur kennzeichnen, die sich aus verschiedenen Ebenen aufbaut. Damit wird der Erkenntnis der Textlinguistik Rechnung getragen, dass die Struktur eines Textes nicht allein unter dem Aspekt einer verknüpften Folge von Sätzen zu betrachten ist, sondern der Text vielmehr eine vielschichtige Struktur oder Organisationsform komplexen Wissens darstellt, die die Textwelt konstituiert. (Gansel und Jürgens 2002, 116)

Ein theoretisch fundiertes Beschreibungsmodell, das die Interaktion der Ebenen der Textstruktur erklärt, fehlt nach Ansicht von Stede (2007) jedoch bis heute. Zwar wurde in vielen Arbeiten auf Zusammenhänge zwischen einzelnen dieser Ebenen eingegangen (siehe z.B. Motsch 1996; Brinker 2000; Schröder 2003; Adamzik 2004b); sowohl über die Art der Ebenen, ihre Bezeichnung als auch ihr Zusammenspiel gibt es dabei allerdings sehr unterschiedliche Auffassungen.

In der Anfangsphase der Textlinguistik wurde zumeist davon ausgegangen, dass es zwei Beschreibungsdimensionen von Texten gibt: die textinterne und die textexterne Ebene, wobei zur ersten Textgrammatik, Textsemantik, und Textthematik gehören<sup>100</sup>, zur letzten die Textpragmatik (siehe z.B. Dressler 1973; Brinker 1980); in späteren Arbeiten wurden diese beiden Grundebenen weiter ausdifferenziert, beispielsweise in die Ebenen „formal-grammatisch“, „inhaltlich-thematisch“, „funktional“ und „situativ“ (vgl. Adamzik 2004b, 53) oder „funktional-illokutiv“ (Handlungsstruktur), „inhaltlich-propositional“ (Inhaltsstruktur), „sprachlich-ausdrucksseitig“ (Nussbaumer 1991, in: Jahr 2000, 389). Eine stärker pragmatisch ausgerichtete Differenzierung findet sich bei Grosz/Sidner (1986), die eine intentionale Ebene („intentional structure“, „structure of purposes“), eine Ebene der Aufmerksamkeitssteuerung („attentional state“, „state of focus of attention“) und eine linguistische Ebene („structure of the sequence of utterances“, „aggregation of utterances“) unterscheiden.

Ein komplexes Modell der Textstruktur, das „separierbare Teilgebiete ausweist und deren Zusammenwirken bei der Textstrukturierung beschreibt“ wurde von Motsch (1996, 3) entwickelt. Motsch geht davon aus, dass die Textstruktur zum einen Ergebnis der Umsetzung grammatischen Wissens ist (Modul „Gramma-

<sup>100</sup> Bei Dressler (1973) werden Textsemantik und Textthematik der Textgrammatik zugeordnet.

tik“: Lautstruktur, Lexikon und Syntax) und dass sie zum anderen auf der Anwendung eines „konzeptuellen Systems“ basiert, das die Grundlage für die „Einheitenbildung und Kombinatorik“ in Texten bildet und über das lokale und globale Textstrukturen erzeugt werden können (Motsch 1996, 5).

Alle Modelle der Textstruktur sind eng bezogen auf den Aspekt der Textualität, sie versuchen die Frage zu beantworten, wann ein Text ein Text ist, und warum. Eine prägende und bis heute immer wieder diskutierte Arbeit zu den Kriterien der Textualität ist die von de Beaugrande/Dressler (1981), in der ein Text als kommunikatives Gebilde definiert wird, das sieben Textualitätskriterien erfüllt: Kohäsion, Kohärenz, Intentionalität, Akzeptabilität, Informativität, Situationalität und Intertextualität. Die Annahme, dass alle sieben Kriterien erfüllt sein müssen, damit ein Text zum Text wird, wurde vielfach zurückgewiesen. Vater schreibt beispielsweise, dass nur Kohärenz „das dominierende Textualitätskriterium“ sei und dass „auch wenn alle anderen (...) postulierten Kriterien nicht erfüllt sind“, es sich trotzdem um einen Text handeln kann; er weist gleichzeitig darauf hin, „dass das Thema weitgehend Kohärenz-Beziehungen im Text bestimmt“, dass also letztlich die Themaorientierung das zentrale Textualitätsmerkmal sei (Vater 2001, 54).<sup>101</sup> Wie sowohl Vater (2001) als auch Adamzik (2004b, 50) betonen, ist es allerdings generell fraglich, ob man Texte und Nicht-Texte deutlich voneinander abgrenzen kann. Die Aussage, dass alle Textualitätskriterien erfüllt sein müssen, um von einem Text sprechen zu können, scheint deshalb problematisch – wobei diese Problematik auch damit zu tun hat, dass es immer noch keine befriedigende und allgemein anerkannte Definition des Textes selbst gibt (siehe z.B. die „Vorbemerkungen zum Definitionsproblem“ bei Adamzik 2004b, 31ff.). Nach Adamziks Ansicht ist es deshalb sinnvoll, sich vor einer Definition des Textbegriffs zunächst über die Beschreibungsdimensionen von Texten klar zu werden. Tabelle 4 enthält eine Übersicht über einige der in den letzten Jahrzehnten entwickelten Kataloge von Beschreibungsebenen:

---

<sup>101</sup> Andere Autoren dagegen bestimmen das Kriterium der Intentionalität (bzw. Funktionalität) als das dominierende Textmerkmal (vgl. Gansel und Jürgens 2002, 31).

<i>Grundunterscheidung</i>	<i>Dressler 1972</i>	<i>Beaugrande/ Dressler 1981</i>	<i>Heinemann 2000</i>	<i>Schröder 2003</i>	<i>Rehm 2005</i>	<i>Stede 2007</i>
textintern	Textgrammatik  -semantik -thematik	Kohäsion	formalgrammatisch	Äußerungsform	Grammatische Ebene	referenzielle Struktur temporale Struktur
		Kohärenz Informativität	inhaltlichthematisch	Handlungsstruktur:  Thematische Beziehungen  Funktionale Beziehungen	Semantische Ebene	thematische Struktur
textextern	Textpragmatik	Intentionalität	funktional		Thematische Beziehungen  Funktionale Beziehungen	Pragmatische kommunikative Ebene
			Akzeptabilität	situativ		Argumentative rhetorische Ebene
		Situationalität			kommunikativ-interaktionale Ebene	
		Intertextualität		Kognitive Ebene Mediale und konzeptionelle Ebene Prototypische Ebene		

Tabelle 4: Kataloge von Beschreibungsdimensionen

Im Folgenden soll ein eigenes Beschreibungsmodell der Textstruktur eingeführt werden (siehe Abbildung 5), mit dessen Hilfe einerseits die in Kapitel 2 eingeführten Thema-Konzepte eingeordnet werden sollen, und das andererseits die Grundlage für die Auseinandersetzung mit der Frage bilden soll, welche Rolle das Thema auf den verschiedenen Ebenen der Textstruktur spielt. Im Gegensatz zu anderen Ebenen-basierten Modellen der Textstruktur (siehe z.B. Stede 2007), wird hier davon ausgegangen, dass das Thema keine eigenständige Strukturebene konstituiert, sondern stattdessen auf mehreren Ebenen der Textstruktur durch die diesen Ebenen eigenen Mittel reflektiert wird. Das im Folgenden eingeführte Modell beschreibt die Ebenen, auf denen die sprachlichen und visuellen Mittel angesiedelt sind, die Kommunikanten nutzen, um Funktion und Bedeutung von Texten zu signalisieren bzw. zu verstehen.

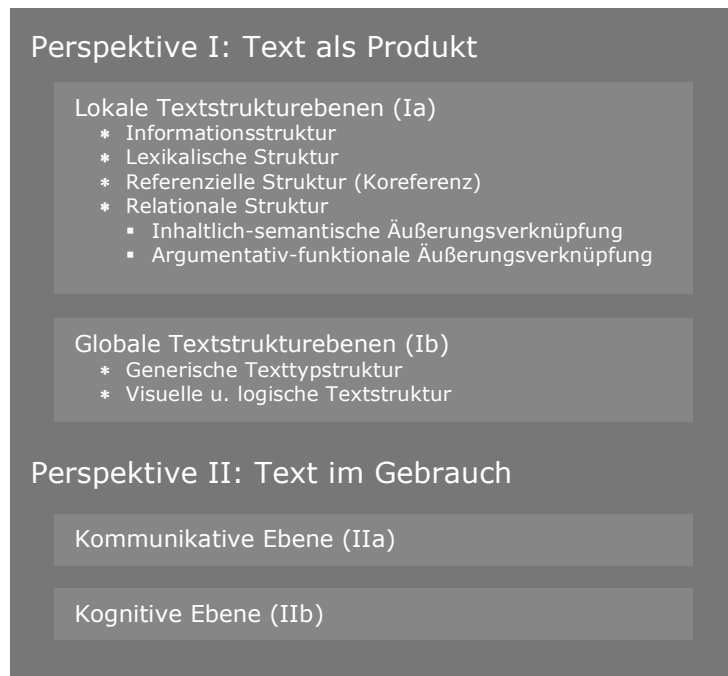


Abbildung 5: Beschreibungsmodell der Textstruktur

Texte können aus zwei Hauptperspektiven beschrieben werden: Sie können sowohl als Produkte betrachtet werden, die lokal und global strukturiert sind (Ebene I), als auch als Bestandteil einer Gebrauchssituation gesehen werden, die sowohl kommunikativ als auch kognitiv bestimmt ist (Ebene II). Auf der lokalen Textstrukturebene (Ia) können vier Arten von Textstrukturen unterschieden werden: Die Informationsstruktur, die lexikalische, die referenzielle<sup>102</sup> und die relationale Struktur.

Die globale Textstruktur (Ib) kann beschrieben werden über die generische Texttypstruktur<sup>103</sup> sowie die visuelle und logische Textstruktur.<sup>104</sup> Diese globalen

<sup>102</sup> Die referenzielle Struktur wird hier im Sinne von Stede (2007, 52) als Koreferenz-Struktur verstanden: Sie wird konstituiert durch die Wiederaufnahme von im Text bereits eingeführten Gegenständen, d.h. durch die Relationen zwischen zwei oder mehreren sprachlichen Ausdrücken, die sich aufeinander beziehen. Die „Summe dieser Beziehungen“ kann nach Stede (2007, 52) als die referenzielle Struktur des Textes aufgefasst werden.

<sup>103</sup> Die generische Texttypstruktur beschreibt die Einteilung des Textes in texttypspezifische Kategorien und deren Abfolgen; konzeptionell stellt sie am ehesten eine Verbindung aus Super- und Makrostruktur im Sinne von van Dijk (1980, 128ff.) dar, insofern sie sowohl als konventionelle als auch inhaltlich charakterisierte genre-spezifische Textstruktur beschrieben werden kann. Mit dem Begriff der Texttypstruktur wird nicht auf das gesamte Merkmalbündel Bezug genom-

Ebenen sind durch ihre kategoriale Prototypikalität bzw. Musterhaftigkeit bestimmt, d.h. ein Text weist – je nach Textsorte – eine mehr oder weniger musterhafte globale Textstruktur auf, die sowohl durch das Vorhandensein einer Menge bestimmter globaler Texttypkategorien und Einheiten der logischen Textstruktur geprägt ist als auch durch eine mehr oder weniger musterhafte Abfolge dieser Kategorien und Einheiten. Die zweite Hauptperspektive auf den Text wird als „Text im Gebrauch“ umschrieben (Ebene II). Im Gegensatz zu den Beschreibungsebenen Ia und Ib geht die Ebene II über die Beschreibung des eigentlichen Textes hinaus. Sie umfasst zum einen die Textproduktion und -rezeption, zum anderen die kognitiven Prozesse, die hierbei ablaufen, sowie die Wissensbestände, die bei der Textproduktion und -rezeption eine Rolle spielen.

Die sprachlichen und visuellen Mittel auf den lokalen und globalen Textstrukturebenen werden in diesem Modell weniger als Eigenschaften von Texten betrachtet, sondern als kommunikative Mittel – sie stellen einerseits Mittel dar, mit denen der Textproduzent signalisieren kann, wie er den Text verstanden haben möchte, andererseits dienen sie dem Rezipienten als Indizien für den Aufbau eines Textverständnisses. Die Strukturierungsebenen dienen in der vorliegenden Arbeit vor allem dazu, diese kommunikativen Mittel zur Konstituierung (Produktionsperspektive) bzw. zum Aufbau (Rezeptionsperspektive) von Kohärenz so differenziert zu beschreiben, wie es für ein formales Modell zur maschinellen Themenidentifikation nötig ist.

Gleichzeitig wird der Versuch unternommen, die linguistische Modellbildung an Modelle aus der Computerlinguistik anschließbar zu machen. In der Computerlinguistik und Texttechnologie werden Texte seit einigen Jahren vermehrt in Form von Multi-Ebenen-Annotationen repräsentiert (siehe z.B. Witt 2004a; Witt 2004b; Dipper, Götze, Küssner et al. 2007; Chiarcos et al. 2008; Goecke et al. 2010). Bei diesen Multi-Ebenen-Annotationen werden formale Analysen auf mehreren Ebenen der Textstruktur durch verschiedene, voneinander unabhängige Annotationsschichten repräsentiert (siehe Kap. 3.3.3). Jede Annotationsebene repräsentiert dabei eine eigene Sicht auf den Text. Der Vorteil von solchen Multi-Ebenen-Annotationen ist, dass Texte im Hinblick auf statistische Auffälligkeiten und Zusammenhänge zwischen verschiedenen Analyse- bzw. Annotationsebenen ausgewertet werden können (z.B. Korrelationen zwischen Einheiten verschiedener Ebenen), ohne dass diese Zusammenhänge zuvor als Hypothesen existent sein müssen. Es kann sich beispielsweise durch statistische Auswertungen zeigen, dass Merkmale einer Ebene A als Cues für die Identifizierung eines

---

men, das für einen Texttyp oder eine Textsorte konstitutiv ist; unter die Beschreibungsebene der generischen Texttypstruktur fallen in der vorliegenden Arbeit ausschließlich die texttypspezifischen Kategorien und deren Abfolgen (siehe auch Langer et al. 2004a, b; Bärenfänger et al. 2006; Bärenfänger et al. 2010; Lobin et al. 2010).

<sup>104</sup> Die relationale Struktur wird konstituiert durch das Geflecht der Kohärenzrelationen zwischen Äußerungen. In der vorliegenden Arbeit wird zur Beschreibung der relationalen Struktur auf die Rhetorical Structure Theory (RST, Mann und Thompson 1988) rekurriert. Die relationale Struktur ist jedoch auch Gegenstand anderer Theorien wie z.B. der Segmented Discourse Representation Theory (SDRT, Asher und Lascarides 2003) und ist deshalb nicht zwingend an die RST gebunden.

sprachlichen Phänomens auf der Ebene B genutzt werden können. Ein Ebenen-basiertes Textmodell und seine formale Repräsentation durch Multi-Ebenen-Annotationen bietet demnach die Möglichkeit, auch emergente Textphänomene – wie z.B. das Thema eines Textes – durch statistische Auswertungen zu ermitteln. Durch die Verrechnung von Merkmalen verschiedener sprachlicher Ebenen lassen sich demnach neue Erkenntnisse über den Text gewinnen, die bei einem weniger offenen oder weniger differenzierten Modell möglicherweise nicht ermittelt werden könnten.<sup>105</sup> Ein Ebenen-basiertes Beschreibungsmodell der Textstruktur bietet demnach einen Mehrwert gegenüber integrativen Textmodellen (vgl. Stede 2007, 185ff.).

Die vorgeschlagenen lokalen und globalen Ebenen (Ia und Ib) stellen ein Minimalinventar dar, das durch weitere, eher texttypspezifischere Ebenen erweitert werden könnte (für narrative Texte beispielsweise eine temporale Struktur, vgl. Stede 2007, 182). Sie spiegeln unterschiedliche Perspektiven auf den Text wider und ermöglichen so eine differenzierte und zunächst isolierte Beschreibung diverser sprachlicher und außersprachlicher Phänomene. Dass die einzelnen Ebenen in vielfältiger Weise miteinander interagieren, wird dabei keineswegs negiert. Es wird vielmehr angenommen, dass die Interaktionen zwischen den diversen sprachlichen und außersprachlichen Phänomenen des Textes so komplex und vielfältig sind, dass sie nur durch statistische Auswertungen von auf mehreren Ebenen-annotierten Textkorpora wirklich fundiert erfasst werden können (siehe auch Stede 2007, 183). Auch die Frage nach einer Hierarchie der Ebenen lässt sich meiner Ansicht nach nicht rein hermeneutisch ermitteln, sondern bedarf korpusgestützter Untersuchungen. Da zur Zeit noch keine umfassend, d.h. auf allen genannten Beschreibungsebenen annotierten Korpora (für verschiedene Texttypen und Textsorten) existieren, lässt sich die Frage nach den Interdependenzen zwischen den Ebenen der Textstruktur in der vorliegenden Arbeit nicht beantworten. Im Folgenden sollen die in Abbildung 5 genannten Strukturebenen im Detail erläutert werden.

### 2.6.3.1 Informationsstruktur

Der Begriff der Informationsstruktur<sup>106</sup> bezieht sich auf die pragmatische Gliederung von Äußerungen sowie die pragmatisch motivierte Verknüpfung aufeinander folgender Äußerungen. Im Zentrum steht hierbei der Zusammenhang zwischen der linguistischen Form von Äußerungen und den kognitiven Annahmen

---

<sup>105</sup> Ein Beispiel für die Verwendung eines solchen Ebenen-basierten Textmodells in einem computerlinguistischen Szenario stellt der in Gießen entwickelte Diskursparser dar (vgl. Lobin et al. 2011). Hier wurde ein Verfahren entwickelt, bei dem auf Basis von Multi-Ebenen-Annotationen (Annotationen auf den Ebenen der logischen Dokumentstruktur, der morphologischen und syntaktischen Struktur, der lexikalischen Kohäsion (lexikalische Ketten), der anaphorischen Struktur, der Texttypstruktur sowie der lexikalischen Diskursmarker) sowie externer Wissensquellen (GermaNet, Diskursmarkerlexikon, Relationstaxonomie, Texttypstrukturontologie) (= Input) automatisch eine neue Annotationsebene (relationale Diskursstruktur im Sinne der Rhetorical Structure Theory) produziert wird (vgl. Lobin et al. 2010; Lobin et al. 2011).

<sup>106</sup> Einen Überblick über die verschiedenen Theorien und Forschungsansätze zur Informationsstruktur gibt Lambrecht (1994, 2ff.)

von Textproduzenten und Rezipienten (Intentionen, aktiviertes Wissen etc.). Grundlegende Einheiten der Informationsstruktur sind Informationseinheiten wie Thema, Rhema, Topic, Comment, Fokus, Hintergrund etc. Diese Einheiten werden verwendet, um einerseits die pragmatische Struktur von Äußerungen, andererseits pragmatische Verknüpfungen zwischen Äußerungen zu beschreiben. Je nach Forschungsrichtung wird die Informationsstruktur deshalb auch als Thema-Rhema-Gliederung, Fokus-Hintergrund-Gliederung (siehe z.B. Altmann 1993) oder Topik-Kommentar-Gliederung (siehe z.B. Gundel 1977) bezeichnet. Da die Begriffe Fokus, Rhema und Comment zum Teil als Synonyme aufgefasst werden (siehe Reis 1977, 212), werden in der Literatur sogar noch weitere Gliederungsebenen genannt, z.B. Topik-Fokus-Gliederung (vgl. Gundel und Fretheim 2004).<sup>107</sup> Allen Gliederungsebenen gemeinsam ist, dass es sich im Gegensatz zur Subjekt-Prädikat-Gliederung um pragmatisch motivierte Strukturierungsebenen handelt. Der Produzent hat die Freiheit, aus der Menge aller möglichen Wortstellungen diejenige auszuwählen, die seinem Zweck am Besten dient, die die Informationen im Satz in einer bestimmten, intentionalen Weise perspektiviert.<sup>108</sup>

In Arbeiten in der Tradition der Funktionalen Satzperspektive werden die Thema-Rhema- und Topik-Kommentar-Gliederung (z.B. bei Daneš 1970) und die Gliederungsebenen Thema-Rhema und Hintergrund-Fokus (z.B. in Firbas Modell zur kommunikativen Dynamik, siehe Firbas 1992) bzw. die diesen Gliederungen zugeschriebenen Eigenschaften theoretisch und analytisch integrativ behandelt (vgl. Fries 1995, 1). Die Position, dass sich Thema-Rhema und Topik-Kommentar unterscheiden, wird dagegen von Halliday (1967; 1985) vertreten. Die Abgrenzung der Gliederungsebenen voneinander ist also nicht eindeutig. Kriterien, die zur Bestimmung des Themas genannt werden, finden sich auch bei der des Topiks, z.B. Erstposition oder Intonation<sup>109</sup> (vgl. Schröder 2003, 160).

Einen interessanten Vorschlag zur Differenzierung der Gliederungsebenen Topik-Kommentar, Thema-Rhema und Hintergrund-Fokus liefert Molnár (1993, 164). Mittels des Bühlerschen Organonmodells (Bühler 1978 [1965], 28), in dem drei Funktionen des sprachlichen Zeichens unterschieden werden (*Appell* → empfängerbezogen; *Darstellung* → gegenstandsbezogen; *Ausdruck* → senderbe-

<sup>107</sup> Eine weitere Variante stellt die Unterscheidung von Haupt- und Nebengedanken und deren Gliederung in Haupt- und Nebensatz dar. Diese wird von Brandt (1996, 212) als „Reliefgebung“ bezeichnet und als globale Informationsstrukturierung aufgefasst (im Gegensatz zur lokalen Informationsstrukturierung durch die Thema-Rhema- und Fokus-Hintergrund-Gliederung).

<sup>108</sup> Warum sich die verschiedenen Möglichkeiten der Strukturierung von Informationen im Satz funktional unterscheiden bzw. warum sie zu unterschiedlichen sprachlichen Repräsentationen bei der Sprachrezeption führen, hat mit sprachpsychologischen und kognitiven Prozessen der Sprachverarbeitung zu tun. Diese der Informationsstruktur zugrunde liegenden Prozesse werden in der Forschungsliteratur jedoch sehr selten untersucht – eine Ausnahme bildet z.B. Scherner (1984, 82ff. u. 178f.; 2000, 193).

<sup>109</sup> Ein bisher noch nicht angesprochenes Mittel zur Bestimmung von Thema und Rhema ist die Intonation. In der Regel gilt das Element mit dem Satzakkzent als Rhema, das unakzentuierte Satzglied als Thema. Umstritten ist die Behandlung des Satzglieds mit kontrastivem Akzent. In einigen Arbeiten wird dieses ebenfalls als thematisch betrachtet und als „Kontrasttopik“ bezeichnet, in anderen Arbeiten nicht (vgl. Molnár 1993, 160).

zogen), unterscheidet Molnár drei Ebenen der kommunikativen Strukturierung:

- |                        |   |                     |
|------------------------|---|---------------------|
| 1. <i>Darstellung:</i> | → | Topik – Kommentar   |
| 2. <i>Empfänger:</i>   | → | Thema – Rhema       |
| 3. <i>Sender:</i>      | → | Hintergrund – Fokus |

Die erste Ebene ist bezogen auf den Inhalt: Die Topik-Kommentar-Gliederung beruht auf einer „Aboutness“-Relation – Topik wird als Gegenstand aufgefasst, über den im Kommentar etwas gesagt wird (siehe die Thema-Konzeption in Kapitel 2.2.1.1). Die zweite Ebene wird als empfängerbezogene Ebene aufgefasst. Die Kategorien Thema und Rhema werden auf den Kenntnisstand des Empfängers bezogen und mit dem Status „bekannt“ vs. „neu“ assoziiert (siehe die Thema-Konzeption in Kapitel 2.4). Die senderbezogene dritte Ebene dagegen basiert auf dem Kriterium der „Relevanz“, die der Sender den Informationen im Satz zuweist (siehe das Thema-Konzept in Kapitel 2.3.1). Für die kommunikative Strukturierung bedeutet dies, dass Sätze sowohl eine spezifische Gewichtung und Perspektive im Hinblick auf den Inhaltsaspekt (Topik-Kommentar) als auch den Sender- bzw. Empfängeraspekt (Hintergrund-Fokus, Thema-Rhema) tragen.

Zusammenfassend charakterisieren demnach folgende zentrale Aspekte das Konzept Thema im Hinblick auf seine Verortung auf der Ebene der Informationsstruktur: a) die Gewichtungs- bzw. Perspektivierungsfunktion (im Hinblick auf die Relevanz bzw. Aufmerksamkeitssteuerung), b) die „Aboutness“-Funktion (das Thema ist der Mitteilungsgegenstand, über den etwas gesagt wird), und c) die Funktion der Anknüpfung (Thema als Bekanntes, Wiederaufgenommenes). Die Anknüpfungsfunktion des Themas besteht allerdings nicht nur auf Ebene der Informationsstruktur, sondern auch auf Ebene der referenziellen Struktur – hier liegen demnach im Hinblick auf das Thema funktionale Überschneidungen zwischen den Strukturebenen vor.

Durch die Gliederung auf der Ebene der Informationsstruktur wird die Aufmerksamkeit des Rezipienten auf ein konkretes Thema gelenkt bzw. andersherum nutzt der Produzent eine spezifische Satzgliedstellung, um damit eine bestimmte Gewichtung und Perspektivierung hin auf ein konkretes Thema bzw. Rhema zu bewirken. Aus Produzentsicht ist das Thema demnach Ausgangspunkt und Quelle einer spezifischen Informationsstrukturierung, aus Rezipientensicht ist das Thema Ergebnis dieser Informationsstruktur – der Rezipient „liest“ die Satzstruktur als Ausdruck der Intention des Produzenten, eine bestimmte Gewichtung und Perspektivierung vorzunehmen, in dessen Mittelpunkt ein konkretes Thema steht. Informationsstrukturelle Kohärenz und thematische Kohärenz interagieren demnach miteinander (zum Zusammenhang von Thema und grammatischer Struktur sowie Informationsstruktur siehe Kapitel 3.3.1.1).



### 2.6.3.2 Lexikalische Struktur

Die lexikalische Struktur wird konstituiert durch die lexikalischen Beziehungen zwischen Wortpaaren oder Wortketten, die eine Verkettung des Textes bzw. den Eindruck von Geschlossenheit bewirken. Traditionell wird diese Strukturebene auch mittels des Begriffs der „lexikalischen Kohäsion“ beschrieben. Grundlegend für die Forschung zur lexikalischen Kohäsion ist die Arbeit von Halliday/Hasan (1976), in der lexikalische Kohäsion als eine von fünf Möglichkeiten (Referenz, Substitution, Ellipsen, Lexik, Konjunktion) der Konstituierung von Kohäsion im Text eingeführt wird.<sup>110</sup> Lexikalische Kohäsion ergibt sich nach Halliday/Hasan durch die Existenz der lexikalischen Relationen Reiteration<sup>111</sup> und Kollokation<sup>112</sup>. In Hasan (1984) wird die Relation der Kollokation aufgrund der Schwierigkeiten, diese objektiv zu ermitteln – “the notion of collocation proved problematic” (1984, 195) –, nicht mehr zu den Arten lexikalischer Kohäsion gezählt. Sie unterscheidet stattdessen “general” (Wiederholung, Synonymie, Antonymie, Hyponymie, Meronymie) vs. “instantial” (Äquivalenz, Eigennamen, Vergleiche) “lexical cohesion”. Cruse (1986) schließlich erweitert die lexikalischen Relationen, die lexikalische Kohäsion bewirken, um die Relationen Kongruenz, kognitive Synonymie, Kompatibilität, Inkompatibilität u.a.

Durch lexikalische Kohäsion können nicht nur zwei lexikalische Einheiten miteinander verknüpft werden (zwischen diesen existiert demnach eine Verknüpfung (“tie”), vgl. Halliday und Hasan 1976, 3), sondern multiple Einheiten, so dass kohäsive bzw. lexikalische Ketten (“chains”) entstehen (Hasan 1984, 188; Hoey 1991, 10; Morris und Hirst 1991, 23).<sup>113</sup>

<sup>110</sup> Wobei nach Hoey (1991, 9) fast 50% aller kohäsiven Verknüpfungen in Texten lexikalisch sind (32% referenziell), d.h. dass lexikalische und referenzielle Mittel die frequentesten kohäsiven Mittel darstellen und über 80% der kohäsiven Verknüpfungen repräsentieren.

<sup>111</sup> Reiteration bezeichnet die Verknüpfung zweier sprachlicher Einheiten, insofern sie sich auf denselben Diskursreferenten beziehen (z.B. durch Wiederholung, durch einen allgemeinen Ausdruck, Synonymie oder Superordination). Das Phänomen der Reiteration kann sowohl der lexikalischen als auch referenziellen Textstrukturebene zugeordnet werden. Reiteration auf der lexikalischen Ebene basiert allerdings auf lexikalisch-semantischer Verwandtschaft, während Reiteration auf der referenziellen Ebene sich auf den Zusammenhang zweier Ausdrücke im Text bezieht und durch den Text selbst konstituiert wird.

<sup>112</sup> Der Begriff der Kollokation wird in der Linguistik unterschiedlich definiert. Kollokation kann zum einen als Spezialfall der Kookurrenz aufgefasst werden. Kookurrenz wiederum ist als das gemeinsame, statistisch signifikante Auftreten zweier oder mehrerer Wörter in einem bestimmten Kontext (z.B. „gestern Abend“) zu verstehen. Wenn diese Kookurrenz im pragmatischen Sinne typisch bzw. gebräuchlich ist, kann auch von Kollokation gesprochen werden (z.B. „Zähne putzen“ vs. „Zähne waschen“) (vgl. Bubenhofer 2009, 111ff.). Zum anderen kann Kollokation aber auch als das gemeinsame Auftreten von Wörtern, die sachlich oder assoziativ zusammengehören und die häufig in ähnlichen lexikalisch-semantischen Kontexten auftreten, betrachtet werden (vgl. Adamzik 2004a, 79). Halliday und Hasan beziehen sich auf Kollokation im zweiten Sinne und fassen darunter Assoziationspaare wie “laugh – joke”, “ill – doctor”, “boy – girl”, “king – crown” etc.

<sup>113</sup> Der Ansatz, lexikalische Kohäsion über lexikalische Ketten zu erklären, ist vor allem in computerlinguistischen und sprachtechnologischen Ansätzen weit verbreitet und es wurde – aufbauend auf den Algorithmus von Morris/Hirst (1991) – eine Vielzahl von Methoden zur maschinellen Identifizierung und Verarbeitung lexikalischer Ketten entwickelt.

Oftentimes, lexical cohesion occurs not simply between pairs of words but over a succession of a number of nearby related words spanning a topical unit of the text. These sequences of related words will be called lexical chains. (...) Lexical chains tend to delineate portions of text that have a strong unity of meaning. (Morris und Hirst 1991, 23)

Zentral für die Textorganisation sind zwei Aspekte: 1. Lexikalische Ketten umspannen Äußerungseinheiten, die eine thematische Einheit bilden, d.h. das Ende einer oder mehrerer Ketten kann einen thematischen Wechsel signalisieren (siehe z.B. Jobbins und Evett 1998; Boguraev und Neff 2000a; Stokes 2003). 2. Lexikalische Ketten interagieren, indem Mitglieder einer Kette in einer funktionalen Relation zu zwei oder mehr Mitgliedern einer anderen Kette stehen. Diese Eigenschaft, multiple Relationen eingehen zu können, unterscheidet lexikalische Kohäsion von anderen Formen der Kohäsion:

Lexical cohesion is the only type of cohesion that regularly forms multiple relationships (though occasionally reference does so too). If this is taken into account, lexical cohesion becomes the dominant mode of creating texture. In other words, the study of the greater part of cohesion is the study of lexis, and the study of cohesion in text is to a considerable degree the study of patterns of lexis in text. (Hoey 1991, 10)

Die Interaktionen zwischen Ketten machen nach Hasan die kohäsive Harmonie ("cohesive harmony") des Textes aus. Zur Beschreibung dieser Interaktionen können sowohl klassische (systematische) lexikalisch-semantische Relationen verwendet werden, als auch nicht-klassische bzw. nicht-systematische (Morris und Hirst 1991, 2004).<sup>114</sup> Es ist davon auszugehen, dass die nicht-systematischen, nicht-klassischen lexikalischen Relationen einerseits zwar schwierig zu klassifizieren und benennen sind – wobei eine empirische Studie von Morris/Hirst (2005) zeigte, dass immerhin 60%-70% der Leser bei der Ermittlung von verwandten (auch assoziativ verwandten) Wortgruppen übereinstimmen –, dass sie andererseits aber einen wichtigen Beitrag zur Identifikation lexikalisch-semantischer Ähnlichkeit liefern.<sup>115</sup>

Die nicht-systematischen Relationen werden in der vorliegenden Arbeit in vier Gruppen unterteilt: 1. Semantische Rollen – diese werden insofern als nicht-systematisch aufgefasst, als die Menge der semantischen Rollen prinzipiell offen ist –, 2. Mitglieder einer logischen Reihe oder Menge, 3. Kollokationen (verstanden als Assoziationspaare oder -ketten) oder Bridging-Relationen, d.h. u.a. schema- und inferenzbasierte Relationen (vgl. Clark 1977). Darüber hinaus kann

<sup>114</sup> Morris/Hirst (1991, 21ff.) definieren fünf Klassen von Verwandtschaft, die die Grundlage zur Bildung lexikalischer Ketten sind: 1. "Reiteration with identity of reference", 2. "Reiteration without identity of reference", 3. "Reiteration by means of superordinate", 4. "Systematic semantic relation (systematically classifiable)", 5. "Nonsystematic semantic relation (not systematically classifiable)" – diese letzte Kategorie ist – wie die Beispiele in Morris/Hirst zeigen – der Relation der Kollokation bei Halliday/Hasan ähnlich. Für die Ermittlung dieser Relationen ist nach Morris/Hirst einerseits ein Thesaurus nötig, andererseits situatives Wissen, Wissen über Eigennamen sowie Wissen über Relationen zwischen Wörtern, die ein ähnliches Gefühl ausdrücken (vgl. Morris und Hirst 1991, 34).

<sup>115</sup> Auch in computerlinguistischen Ansätzen zur Ermittlung der semantischen Ähnlichkeit von Wörtern oder kompletten Texten scheint es in letzter Zeit verstärkt zu einer Integration nicht-systematischer Relationen zu kommen, so dass eine Operationalisierung und damit maschinelle Klassifikation in greifbare Nähe rückt (siehe z.B. Zesch und Gurevych 2009).

auch partielle lexikalische (bzw. morphologische) Rekurrenz (z.B. „Bahnschlag“ – „Bahnstrecke“ – „Bahnhof“) dazu beitragen, dass zwei Wörter als lexikalisch miteinander verknüpft betrachtet werden – diese Verwandtschaftsrelation stellt allerdings die unspezifischste dar und kann deshalb auch zusätzlich zu einer der anderen lexikalisch-semanticen Relationen vorliegen (z.B. bei vollständiger Rekurrenz oder bei einer generischen Relation). Zur Bildung lexikalischer Ketten wird in der vorliegenden Arbeit auf das folgende Relationsinventar zurückgegriffen:

1. Klassische lexikalisch-semanticen Relationen:
  - a. Identität (z.B. Peter – Peter)
  - b. Synonymie (z.B. Lift – Aufzug)
  - c. Paraphrase (z.B. der Mann – der Kerl mit dem blauen Hemd)
  - d. Hyponymie (z.B. Dackel – Hund)
  - e. Hyperonymie (z.B. Hund – Dackel)
  - f. Meronymie (z.B. Finger – Hand)
  - g. Holonymie (z.B. Hand – Finger)
  - h. Menge-Teilmenge (z.B. fünf Mädchen – die eine)
  - i. Generische Relation (z.B. Kraftwagen – Personenkraftwagen)
  - j. Antonymie (z.B. Hitze – Kälte)
2. Nicht-systematische lexikalisch-semanticen Relationen:
  - a. Semanticen Rollen (nach de Beaugrande und Dressler 1981, 101f.; von Polenz 1985, 170ff.)
    - i. AGENS
    - ii. PATIENS
    - iii. BENEFAKTIV
    - iv. AFFIZIERTES OBJEKT
    - v. INSTRUMENT
    - vi. UMSTAND
    - vii. URSACHE
    - viii. FOLGE
    - ix. ...
  - b. Mitglieder einer logischen Reihe oder Menge (z.B. „Hessen, Thüringen, Niedersachsen“, aber auch „hübsch, niedlich, süß“ oder „gelb, rot, blau“)
  - c. Kollokationen (z.B. „Student, Professor, Universität“, aber auch „Gewalt, zerstören“) oder Bridging-Relationen (schema- oder inferenzbasierte Relationen)
3. Partielle lexikalische Rekurrenz

Die Einträge einer lexikalischen Kette können als Cluster von Themenwörtern aufgefasst werden, d.h. als Gruppen von Wörtern, die sich auf ein Thema beziehen, d.h. jede Kette repräsentiert ein Thema oder Subthema (vgl. Barzilay und

Elhadad 1999, 14).<sup>116</sup>

Das Thema kann demnach sowohl als Quelle (Produzentsicht) als auch als Ergebnis der lexikalischen Strukturierung aufgefasst werden (Rezipientensicht): Ein Produzent greift auf Gruppen verwandter Wörter (Gruppen lexikalischer Ketten) zurück, um ein Thema und seine Aspekte sprachlich zu realisieren. Je nachdem, ob es sich um ein Haupt- oder Nebenthema handelt, wird eine spezifische Gruppe von Wörtern bzw. Cluster von Themenwörtern (eine lexikalische Kette) im gesamten Text (globales Thema) oder nur in einem Abschnitt des Textes (lokales Thema) verwendet – lexikalische Ketten sind deshalb als eine Konsequenz der Orientierung am Thema als Sachverhaltskomplex aufzufassen. Der Rezipient wiederum nutzt die lexikalische Struktur, um das Thema des Textes zu ermitteln – besonders häufig und durchgehende verwendete Gruppen verwandter Wörter (lexikalische Ketten) können dabei als Indiz für das Textthema interpretiert werden (zum Zusammenhang von Thema und lexikalischen Ketten siehe auch Kapitel 3.3.1.3).

### 2.6.3.3 Referenzielle Struktur

Die referenzielle Struktur wird in der vorliegenden Arbeit im Sinne von Stede (2007, 52) als Koreferenz-Ebene aufgefasst: Sie wird konstituiert durch die Wiederaufnahme von im Text bereits eingeführten Diskursentitäten, d.h. durch die Relationen zwischen zwei oder mehreren sprachlichen Ausdrücken, die sich auf dieselbe Diskursentität beziehen. Die „Summe dieser Beziehungen“ kann nach Stede (2007, 52) als die referenzielle Struktur des Textes aufgefasst werden. Ausdrücke, die auf eine im Text bereits referenzierte Diskursentität Bezug nehmen, d.h. Koreferenz herstellen, werden als Anaphern bezeichnet; Ausdrücke, auf die in koreferenter Weise Bezug genommen wird, als Antezedenten.

Bei der Beschreibung von Koreferenzphänomenen sind zwei Ebenen zu differenzieren: die linguistische Ebene und die konzeptuelle Ebene. Auf der linguistischen Ebene existiert ein sprachlicher Ausdruck  $A_1$ , der auf eine Diskursentität  $D_1$  auf der konzeptuellen Ebene verweist. Dieser Ausdruck kann eine einfache oder komplexe Nominalphrase sein, ein Eigenname, eine Proposition oder ein Cluster aus Satzsequenzen unterschiedlicher Länge. Wenn sich ein im Text nachfolgender zweiter sprachlicher Ausdruck  $A_2$  ebenfalls auf die mit dem Ausdruck  $A_1$  referenzierte Diskursentität  $D_1$  bezieht, sind die beiden Ausdrücke korefe-

---

<sup>116</sup> Das semantisch orientierte Thema-Konzept von Agricola (1976; 1979) versucht in ähnlicher Weise über das Konzept der (semantisch definierten) Isotopieketten, die – ähnlich wie die lexikalischen Ketten – aus durch semantische Relationen verknüpften Einträgen bestehen, die die Themenstränge des Textes repräsentieren, Aussagen über das Thema bzw. die thematische Struktur zu machen (siehe Kapitel 2.2.2). Das Konzept der Isotopieketten unterscheidet sich von dem der lexikalischen Ketten allerdings insofern, als dass lexikalische Ketten auf Basis der Wortbedeutung gebildet werden, während sich das Isotopiekonzept auf semantische Merkmale von Wortbedeutungen bezieht und dementsprechend unterhalb der Wortebene angesiedelt ist (vgl. Heinemann 2000a, 56).

rent.<sup>117</sup> Die Ausdrücke  $A_1$  und  $A_2$  können syntaktisch und lexikalisch divergieren, z.B. kann  $A_1$  als NP realisiert sein („der Mann“) und  $A_2$  als Pronomen („er“). Der Ausdruck  $A_2$  fungiert in dieser Relation als Anapher, die sich auf eine durch den Bezugsausdruck – den Antezedenten  $A_1$  – bereits eingeführte Diskursentität bezieht. Koreferenz basiert demnach auf einer Text-Welt-Relation.<sup>118</sup>

Neben den koreferenten Anapher-Antezedent-Relationen existieren darüber hinaus auch Anapher-Antezedent-Relationen, die auf Kospezifikation basieren. Bei der Kospezifikation wird eine Diskursentität  $D_1$  durch eine zweite Diskursentität  $D_2$ , die zur ersten Entität in einer semantischen oder ontologischen Beziehung steht, näher spezifiziert. Die Ausdrücke, die auf diese Diskursentitäten Bezug nehmen, sind demnach nicht koreferent, sondern sie dienen der Kospezifikation zweier semantisch oder konzeptuell verwandter Diskursentitäten (vgl. Goecke, Holler et al. 2008). Während zwei Ausdrücke im Text nur dann in der Relation der Koreferenz stehen, wenn sie auf dieselbe Diskursentität in der Welt verweisen, genügt für Kospezifikation, dass ein Ausdruck einen vorangegangenen sprachlich wieder aufgreift. Kospezifikation basiert demnach auf einer Text-Text-Relation zwischen zwei Ausdrücken.

Anapher-Antezedent-Verknüpfungen können einerseits syntaktisch beschrieben werden, andererseits semantisch. Syntaktisch können Anaphern realisiert sein als Pronomen (Personal-, Demonstrativ-, Relativ-, Possessivpronomen) oder Pronominaladverbien, Eigennamen, rekurrente einfache oder komplexe Nominalphrasen, Ellipsen, Ereignisverben oder Cluster aus Satzsequenzen unterschiedlicher Länge. Semantisch stehen Anaphern zu ihren Antezedenten in der Relation der Identität, Synonymie (z.B. Lift – Aufzug), Paraphrase (z.B. der Mann – der Kerl mit dem blauen Hemd), Hyperonymie (z.B. Hund – Dackel), Hyponymie (z.B. Dackel – Hund), Meronymie (z.B. Finger – Hand), Holonymie (z.B. Hand – Finger), Menge-Teilmenge (z.B. fünf Mädchen – die eine), in einer generischen Relation (z.B. Kraftwagen – Personenkraftwagen) oder Bridging-Relation (Lehrer – Schule – Klassenzimmer – Tafel), d.h. u.a. schema- und inferenzbasierte Relationen (vgl. Clark 1977). Insbesondere im Fall der Bridging-Relationen spricht man auch von indirekter Anaphorik: Der Antezedent zu einer Anapher ist in diesem Fall nicht explizit im Text enthalten, kann aber dennoch auf Grundlage des Sprachwissens oder des Kontextes erschlossen werden.

<sup>117</sup> Seltener gibt es den Fall, dass ein Ausdruck auf etwas im Text Bezug nimmt, was erst im Folgenden eingeführt wird – diese vorwegnehmende Bezugnahme wird als kataphorischer Verweis bezeichnet.

<sup>118</sup> Insofern es verschiedene Typen von Referenz gibt (siehe Kapitel 2.7.2) – Ereignisreferenz (ein Ereignisverb, z.B. „kaufen“ wird zur Bezugnahme auf ein Ereignis verwendet), Situationsreferenz (ein Satz wird zur Bezugnahme auf eine Situation verwendet), Dingreferenz (in erster Linie durch Nominalphrasen und Pronomina), Ortsreferenz, Zeitreferenz (durch Tempora des Verbs, Temporaladverbiale und Temporalsätze), Eigenschaftsreferenz (durch Adjektive), Modalitätsreferenz (durch Modalverben oder Modaladjektive), Quantitätsreferenz – können auch verschiedene Typen von Diskursentitäten (z.B. Ereignisreferent) und Typen von anaphorischen Ausdrücken und Bezugsausdrücken (Antezedenten) unterschieden werden (neben den nominalen und pronominalen auch Ereignisverben oder Cluster aus Satzsequenzen).

Insofern Koreferenz auch einhergehen kann mit dem Vorliegen einer lexikalisch-semanticen Relation (z.B. Hyponymie: „Mein Hund ist ein Dackel.“) zwischen den koreferenten Ausdrücken, überschneidet sich die referenzielle Ebene zum Teil mit der lexikalischen Ebene. Im Unterschied zu lexikalisch-semanticen Relationen sind referenzielle bzw. anaphorische Relationen allerdings gebunden an das Verstehen des konkreten Textes, d.h. sie existieren nicht unabhängig von diesem, wie das bei lexikalisch-semanticen Relationen der Fall ist (diese können z.B. in einem Lexikon aufgelistet werden). Im Fall der Ereignisreferenz gibt es darüber hinaus Fälle, in denen die Relation zwischen Anapher und Antezedent nicht auf einer lexikalisch-semanticen Relation basiert, z.B. wenn die Anapher als verbloser, elliptischer Satz realisiert ist: „Ich habe mir ein neues Auto gekauft! Und Martin eine Stereoanlage.“ (siehe Stede 2007, 54).

Insofern referenzielle Kontinuität als Marker für thematische Kontinuität aufzufassen ist (siehe Kapitel 2.2.1.2), kann eine Analyse der im Text referenzierten Diskursentitäten zur Unterscheidung von zentralen bzw. globalen Diskursentitäten (als kontinuierlich wiederaufgenommenen Themen) und nebensächlichen bzw. lokalen Diskursentitäten (als den in einem begrenzten Textabschnitt behandelten Themen) beitragen. Die durchgehende Orientierung auf eine Menge globaler Diskursentitäten, über die im Text etwas gesagt wird, kann also als durchgehende Orientierung auf eine spezifische Gruppe zentraler bzw. globaler Themen interpretiert werden. Themazentrierung kann deshalb sowohl als Quelle (Produzentensicht) als auch als Resultat (Rezipientensicht) der referenziellen Strukturierung des Textes aufgefasst werden – globale, durchgehend wiederaufgenommene Diskursentitäten dienen der Signalisierung des globalen Themas bzw. globale Diskursentitäten können als indexikalische Zeichen für das Thema betrachtet werden. Koreferenz stellt demnach ein zentrales Mittel zur Bestimmung des Themas dar (zum Zusammenhang von Thema und referenzieller Struktur siehe Kapitel 0).

#### 2.6.3.4 Relationale Struktur

Die relationale Textstruktur beschreibt die Verknüpfung von Äußerungen (bzw. Diskurssegmenten) zu größeren Einheiten; diese Verknüpfungen werden dabei als Kohärenz- oder Diskursrelationen bezeichnet (vgl. Redeker 2000; Bärenfänger et al. 2010, 82ff.).<sup>119</sup> Sie konstituieren lokale Textkohärenz (vgl. Stede 2007, 24). Kohärenzrelationen bestehen dabei nicht nur zwischen benachbarten Sätzen, sondern auch zwischen größeren Einheiten, die selbst aus miteinander durch Kohärenzrelationen verknüpften Äußerungseinheiten bzw. Diskurssegmenten bestehen (z.B. zwischen Absätzen). In dem hier vorgeschlagenen Textbe-

---

<sup>119</sup> Die Begriffe „Kohärenzrelation“ und „Diskursrelation“ werden in der vorliegenden Arbeit synonym verwendet. Welche und wie viele Kohärenzrelationen existieren bzw. nötig sind, darüber herrschen unterschiedliche Auffassungen (vgl. Hovy und Maier 1995). Hovy/Maier sprechen von 400 verschiedenen Kohärenzrelationen, die in der Forschung von der Linguistik über die Computerlinguistik bis zur Künstlichen Intelligenz vorgeschlagen wurden. Sie selber entwickeln eine Klassifikation mit ca. 70 Relationen, organisieren diese jedoch hierarchisch in einer Taxonomie, so dass sowohl sehr spezifische als auch sehr allgemeine Relationen zur Analyse der Textstruktur verwendet werden können.

schreibungsmodell werden zwei Typen von Kohärenzrelationen unterschieden: inhaltliche (“ideational”/“subject-matter”) und funktionale (“interpersonal”/“presentational”).<sup>120</sup> Zu den inhaltlichen Kohärenzrelationen gehören z.B. ELABORATION, CIRCUMSTANCE, CAUSE, RESULT, CONDITION, COMPARATIVE etc., zu den funktionalen INTERPRETATION, ENABLEMENT, ANTI-THESIS, EXHORTATION, CONCESSION, JUSTIFICATION etc. (vgl. Hovy und Maier 1995, 14).

Die Frage, ob ein Diskurssegment jeweils nur durch eine Relation an ein anderes Segment gebunden ist, d.h. ob Kohärenz entweder funktional oder inhaltlich erzeugt wird, oder ob ein Diskurssegment auf mehreren Ebenen relational an andere Segmente angeschlossen ist, d.h. ob gleichzeitig sowohl funktionale als auch inhaltliche Kohärenz vorliegt, wird in den verschiedenen Theorien, die sich mit der relationalen Diskursstruktur beschäftigen, unterschiedlich beantwortet. In der Rhetorical Structure Theory (RST, Mann und Thompson 1988) wird davon ausgegangen, dass zwar multiple Analysen von Texten möglich sind und dass dementsprechend mehrere Relationen zwischen Segmenten vorliegen können, dass man sich aber in jeder RST-Analyse für eine einzige Relation zwischen zwei Segmenten entscheiden muss, dass eine Relation die vorherrschende sei.<sup>121</sup> Gleichzeitig ist eine Grundannahme der RST, dass immer nur adjazente Segmente durch eine Relation verbunden werden können (adjazent in der hierarchischen Struktur, nicht adjazent auf der Textoberfläche). Die relationale Struktur wird deshalb als Baumstruktur repräsentiert. Dieser Auffassung widersprechen andere Theorien zur relationalen Diskursstruktur, wie z.B. die Segmented Discourse Representation Theory (SDRT, siehe Asher und Lascarides 2003). In SDRT werden Baumstrukturen als unzureichend betrachtet, die relationale Struktur wird stattdessen als Graph modelliert.

Thematische Kohärenz und relationale Kohärenz bedingen sich wechselseitig: Voraussetzung für thematische Kohärenz ist u.a. das Vorliegen von thematischen Zusammenhängen zwischen Diskurssegmenten. Diese thematischen Zusammenhänge können realisiert sein als inhaltliche oder funktionale Kohärenzrelationen (ein Thema wie ein Ereignis wird im Text näher ausgeführt, indem über die Ursachen dieses Ereignisses gesprochen wird oder eine Begründung gegeben wird, wie es zu dem Ereignis kommen konnte etc.). Auf der anderen Seite ist relationale Kohärenz abhängig davon, dass zwei Diskurssegmente auch thematisch zusammenhängen – zwei Diskurssegmente sind nur dann durch Kohärenz-

<sup>120</sup> Diese Unterscheidung und deren Benennung basiert auf der zur Kategorisierung der Relationen diverser Forscher (z.B. Hobbs 1979; Mann und Thompson 1988) entwickelten Relations-Taxonomie von Hovy/Maier (1995), die wiederum auf der von Halliday vorgeschlagenen Differenzierung linguistischer Phänomene in die drei Metafunktionen “ideational”, “interpersonal” und “textual”, die “three aspects of the meaning of the clause”, basiert (Halliday 1985, 53).

<sup>121</sup> Die Problematik dieses Paradigmas und die dem entgegenstehende Notwendigkeit multipler Relationen sprechen z.B. Moore/Pollack (1992) oder Knott et al. (2001) an. Auch die Annotationspraxis zeigt, dass die Annahme einer einzigen Relation problematisch ist: Annotatoren haben häufig Probleme bei der Auswahl einer Relation, weil oft mehrere Relationen gleichermaßen passend sind – wie Mann/Thompson (1988, S.245ff.) schreiben, können Relationen selten mit Bestimmtheit (“certainty”) ausgewählt werden, sondern die Auswahl basiert stattdessen eher auf einem “plausibility judgement”, auch weil es für viele RST-Relationen keine eindeutigen sprachlichen Indizien gibt.

relationen verknüpfbar, wenn zwischen den beiden Segmenten auch ein thematischer Bezug (der als inhaltlicher oder funktionaler Bezug realisierbar ist) hergestellt werden kann (zum Zusammenhang von Thema und relationaler Struktur siehe Kapitel 3.3.2.1).

### 2.6.3.5 Visuelle und logische Textstruktur

Texte sind – im Gegensatz zu Dialogen – dadurch beschränkt, dass sie immer an ein bestimmtes Trägermedium gebunden sind, um visuell sichtbar werden zu können (z.B. ein Buch, ein Computerbildschirm oder ein Handy). Sie besitzen also eine visuelle Realität. Diese Dimension von Texten lässt sich als visuelle Textstruktur<sup>122</sup> beschreiben: Texte sind zusammengesetzt aus geometrischen Einheiten, insofern sie, geometrisch gesehen, aus Rechtecken bestehen, die abgrenzbare Textblöcke bilden: Überschriften, Absätze, Fußnoten, Aufzählungslisten, Tabellen, darüber hinaus aber auch Grafiken, Bilder etc. Die grundlegenden Einheiten dieser geometrischen Struktur sind Linien und Vierecke, die durch ihre Kombination und Verschachtelung einfache oder komplexe Textobjekte formen (einfache Textobjekte wären z.B. Wörter – diese sind durch Leerzeichen und Zeilenabstand von den sie umgebenden Textobjekten getrennt). Diese geometrischen Textobjekte können einerseits in ihren Relationen zueinander beschrieben werden, z.B. durch Kategorien wie „Umschließung“ (rekursive Blockbildung, d.h. Verschachtelung von Textobjekten), Ähnlichkeit (im Hinblick auf visuelle Merkmale), Bezugnahme (aufgrund von physikalischer Nähe). Andererseits können sie durch Attribute wie Schriftart, -farbe, -größe, -schnitt, Laufweite und Wortabstände, d.h. mikrotypographische Eigenschaften, näher bestimmt werden (vgl. Bateman et al. 2001, 414ff.; Power et al. 2003, 213; Bärenfänger et al. 2010, 86ff.; Lobin et al. 2010).

Die visuelle Textstruktur trägt zum Bedeutungsaufbau mit bei:

Just as all speech has prosody (even if it is a monotone), so too do all texts have layout (even if it is simple wrapped format, in a single face and font, and makes rudimentary use of white space). And just as prosody undoubtedly contributes to the meaning of utterances, so too does a text's graphical presentation contribute to its meaning (Power et al. 2003, 211).

Ein sprachlicher Ausdruck, der visuell als Überschrift hervorgehoben ist, hat eine andere Funktion als derselbe Ausdruck, der z.B. in einer Fußnote steht. Diese, durch die Visualisierung kodierten funktionalen Unterschiede wiederum führen zu einer Veränderung der Bedeutung bzw. des kommunikativen Sinns dieser Ausdrücke. Allerdings muss einschränkend festgehalten werden, dass visuelle Mittel

kein konventionalisiertes Mittel für eine über die Semantisierung hinausgehende Bedeutungskonstitution (Deppermann 2002) dar[stellen; M.B.]: Ob und welche ‚Bedeutung‘ ihnen zuge-

<sup>122</sup> In dieser Arbeit wird der eher neutrale Begriff der „visuellen Textstruktur“ verwendet und nicht die extralinguistisch geprägten Begriffe „typographische Struktur“ oder „Layoutstruktur“. Letztere können als Synonyme aufgefasst werden (vgl. Bateman et al. 2001, 416), insofern beide Begriffe zur Beschreibung der komplexen Gestaltungsprozesse von gedruckten oder elektronischen Medienprodukten verwendet werden.



sprochen oder unterstellt wird, hängt damit offensichtlich ganz entscheidend vom (vielleicht sehr asymmetrisch verteilten) Wissen und Können der Kommunikationspartner und deren Beobachter ab. (Antos 2009, 407f.)

Visuelle Strukturen haben zwar „indexikalisches Potential“ (Feilke und Linke 2009, 15), aber ob Kommunikanten dieses Potential nutzen und visuelle Mittel als indexikalische Zeichen interpretieren, ist offen bzw. hängt auch davon ab, ob Textproduzenten visuelle Gestaltungsmittel stringent und konsistent einsetzen (vgl. Hagemann 2007, 77). Wenn dies der Fall ist, leisten die visuellen Mittel einen Beitrag zur Steuerung des Textverstehens. Nach Hagemann (2007, 80) ist eine wichtige Leistung der (konsistent) eingesetzten visuellen Mittel die der Etablierung einer „kategorialen Ebenenbildung im propositionalen Gesamtzusammenhang“: Durch eine spezifische visuelle Gestaltung von Textteilen wird dem Rezipienten signalisiert, welche Textteile bzw. Einheiten welcher kategorialen Ebene zuzuordnen sind (Überschriften, Definitionen, Zitate, Beispiele, Hyperlinks etc.). Die systematische, konsistente visuelle Gestaltung eines Textes lässt demnach zu, dass Rezipienten von der formal-strukturellen bzw. visuellen Gleichartigkeit von Texteinheiten auf eine spezifische kategoriale Zusammengehörigkeit schließen und die visuellen Mittel als indexikalische Zeichen für eine bestimmte Kategorie deuten können (vgl. Hagemann 2007, 82).

Die logische Textstruktur stellt eine Abstraktion der visuellen Textstruktur dar und beschreibt die Strukturierung eines Textes in Komponenten wie Überschrift, Kapitel, Absatz, Tabelle, Grafik, Fußnote, Inhaltsverzeichnis etc. Diese Kategorien weisen jeweils unterschiedliche Funktionen und inhaltliche Eigenschaften auf (z.B. haben Überschriften und Inhaltsverzeichnis die Funktion, eine Orientierung auf Kommendes zu geben, Grafiken dienen der Visualisierung von etwas sprachlich Vermitteltem, Tabellen ordnen zwei oder mehr Parameter einander zu, Absätze signalisieren eine inhaltliche oder funktionale Einheit). Wie die generische Texttypstruktur ist auch die logische Textstruktur als globales Muster aufzufassen, das die Textproduktion und -rezeption leitet: Eine Überschrift ist im Text nie an letzter Stelle zu finden, ein Kapitel umfasst immer mindestens einen Absatz und absatzwertige Objekte (z.B. Grafiken) etc. Die musterhaften Kategorien und deren Abfolgen basieren auf Konventionen, die Teil des Wissens der Sprachgemeinschaft sind. Diese können in Form einer logischen Text- bzw. Dokumentgrammatik beschrieben werden, insofern die logische Textstruktur konstituentiell aufgebaut ist: Der Text als logisch größte Einheit besteht aus untergeordneten Einheiten (z.B. Kapiteln), die wiederum aus kleineren Einheiten (z.B. Absätzen, Grafiken) zusammengesetzt sind (vgl. Bärenfänger et al. 2006; Lobin et al. 2010). Textteile stehen dementsprechend in einer Teil-Ganzes-Relation – sie können grafisch und formal als Bäume dargestellt werden (vgl. z.B. Lobin 2001, 37ff.; Lobin 2003, 52ff.). Dokumentgrammatiken können dabei formal modelliert werden durch DTD's (Document Type Definition) oder Schemasprachen wie XML Schema (W3C – World Wide Web Consortium 2004b) oder RELAX NG (OASIS – Organization for the Advancement of Structured Information Standards 2001) (vgl. Lobin 2004, 27ff.).

Logische und visuelle Textstruktur interagieren miteinander, insofern visuelle Objekte funktional bestimmte logische Texteinheiten erkennbar machen helfen.

Darüber hinaus interagiert die visuelle Struktur mit der generischen Texttypstruktur, insofern spezifische Konfigurationen auf der Ebene der visuellen Textstruktur als „visuelle Erkennungscode“ (Stein 2003, 148) für bestimmte Texttypen fungieren (so sind allein vom visuellen Eindruck her Zeitungstexte von wissenschaftlichen Artikeln oder Kochrezepten zu unterscheiden). Diese visuellen Konfigurationen sind, ebenso wie die Texttypstruktur selbst, als musterhaft beschreibbar (vgl. Sandig 1997, 36). Textinstanzen sind deshalb an ihrer visuellen Prototypikalität zu messen und unterscheiden sich im Grad ihrer visuellen „Textmuster“-Kohärenz.

### 2.6.3.6 Generische Texttypstruktur

Die generische Texttypstruktur beschreibt die Einteilung des Textes in texttypspezifische Kategorien und deren Abfolgen. Das Konzept der Texttypstrukturkategorien interagiert dabei mit dem Konzept der Textrollen (siehe Fußnote 92, S. 74), insofern Texttypstrukturkategorien als Textrollen aufgefasst werden können, die für einen bestimmten Texttyp prototypisch sind. Über Texttypstrukturkategorien werden demnach aus der Menge der Textrollen, die in einem Text instanziiert sein können, bestimmte Textrollen als prototypische Rollen hervorgehoben. Texttypen unterscheiden sich demnach durch ihre prototypische Konfiguration von Textrollen, d.h. durch eine spezifische Kombination von Textrollen. Erst durch diese spezifische Kombination bzw. Konfiguration wird eine Menge von Textrollen zum prototypischen Muster eines Texttyps. Einige Texttypen lassen sich über eher allgemeine, an Kasusrollen angelehnte Textrollen beschreiben, andere nur über sehr spezifische Textrollen (z.B. die Texttypstrukturkategorien zum Texttyp „wissenschaftlicher Artikel“<sup>123</sup>). Texttypstrukturkategorien – also auch Textrollen – können dabei auch als unterspezifizierte Themen (bzw. Aspekte des Themas) bzw. als abstrakte Thementypen aufgefasst werden.

Auch Stede (2007, 77) geht davon aus, dass es einen Zusammenhang zwischen der generischen Texttypstruktur und der globalen Abfolge von Themen gibt. Die globale thematische Struktur ist seiner Ansicht nach aus einem Inventar von texttypspezifischen „Inhaltszonen“ (zum Begriff „Inhaltszonen“ siehe auch Teufel 1999) zusammengesetzt, die in dem von ihm untersuchten Texttyp „Filmrezension“ auf der Grundlage von Konventionen in einer spezifischen Weise linear angeordnet sind, z.B. *Titel, Regie, Datum der Rezension, Copyright-Vermerk, Beschreibung, Meinung* – wobei die beiden letztgenannten weiter untergliedert werden (in verschiedene Subtypen, die sich durch das, *was* beschrieben bzw. kommentiert wird, unterscheiden). Die genannten Inhaltszonen lassen sich – im Gegensatz zu den Inhaltszonen von Zeitungsnachrichten<sup>124</sup> oder auch wissen-

---

<sup>123</sup> Die generische Texttypstruktur wird auch in Langer et al. (2004a, 49) als globale thematische Textstruktur aufgefasst. Für den Texttyp „wissenschaftlicher Zeitschriftenartikel“ werden die texttypspezifischen Thementypen („Topic Types“) *Background, Problem, Research Topic, Framework, Method, Answers* etc. beschrieben.

<sup>124</sup> Zeitungsnachrichten weisen typischerweise die folgende Konfiguration von Textrollen auf: *Ereignis* („Was?“), *Person* („Wer?“), *Ort* („Wo?“), *Zeit* („Wann?“), *Ursache/Zweck* („Warum?“), *Mittel* („Wie?“) (vgl. Weischenberg 1990, 70; Schröder 2003, 202) sowie *Hintergrund, verbale Reaktionen, vorangegangene Ereignisse, Konsequenzen* (vgl. van Dijk 1985, 87).

schaftlichen Artikeln – nur schwer als abstrakte Thementypen interpretieren. Dies liegt jedoch meiner Ansicht nach vorwiegend daran, dass die genannten Inhaltszonen zum Teil besser als Einheiten der logischen bzw. visuellen Textstruktur aufgefasst werden sollten (z.B. *Titel, Datum, Copyright* – diese Inhaltszonen sind eher als visuell-ausdrucksseitig markierte Einheiten beschreibbar). Andere Kategorien wie *Beschreibung* und *Meinung* bedürfen, um als abstrakte Thementypen aufgefasst werden zu können, einer Spezifizierung durch Subtypen, die weniger auf das „Wie“ der Darstellung („beschreiben“, „bewerten“) als vielmehr das „Was“ bezogen sind (was Stede zufolge auch getan wurde, wobei er diese Subtypen nicht nennt).

Zusammenfassend ist festzuhalten, dass bestimmte Texttypen und Textsorten im Hinblick auf erwartbare Thementypen konventionell geprägt sind. Schank schreibt hierzu:

Für den inhaltlichen Aspekt gilt, dass Themen und Themenklassen zum Teil eine ganz bestimmte Distribution aufweisen. Diese lässt sich mit Hilfe des Textsortenbegriffs beschreiben. In bestimmten Typen von sozialen Situationen sind bestimmte Themenklassen regelhaft erwartbar, so etwa die Thematisierung persönlicher Lebensprobleme in der Institution Beratungsgespräch, aus der die Textsorte Beratungsgespräch resultiert. Andere Situationstypen sind nicht so eng auf bestimmte Themenklassen festgelegt. In einer Diskussion können grundsätzlich alle Themenklassen behandelt werden, so auch im small talk. (Schank 1977a, 235)

Die aus Konventionen resultierenden Vorgaben über mögliche Thementypen (und deren Abfolge) gelten demnach nicht für alle Textsorten – wie z.B. Fritz (1994) und Schank (1977a) darlegen, unterliegen manche Textsorten mehr Festlegungen als andere. Nach Fritz sind z.B. institutionell gebundene Textsorten thematisch beschränkter als z.B. kreative Gespräche. Was thematisiert werden darf, ist also abhängig von der Textsorte bzw. dem Texttyp (zum Zusammenhang von Thema und Textmusterwissen und Thema und generischer Texttypstruktur siehe Kapitel 3.2.2 und 3.3.2.2).

## **2.7 Kognitive Grundlagen der Textproduktion und -rezeption**

### **2.7.1 Wissensbestände beim Kommunizieren**

Textproduktion und -rezeption werden u.a. beschränkt durch Wissen. Dieses Wissen wird in der linguistischen Forschung in unterschiedlicher Weise kategorisiert und in unterschiedlichen Detailgraden beschrieben. Im Folgenden sollen vier konkurrierende Einteilungen von Wissen vorgestellt werden.

Wissen kann zunächst ganz allgemein in deklaratives, prozedurales und episodisches Wissen unterschieden werden (1). Das deklarative Wissen umfasst Wissen über die Welt, das prozedurale beschreibt Handlungswissen und Strategien, das episodische Wissen stellt Wissen über persönlich Erlebtes dar. Eine zweite Möglichkeit der Differenzierung von Wissensbeständen (2) unterscheidet „Wozu-Wissen“ (Konzeptionswissen), „Was-Wissen“ (Realisierungswissen) und „Wie-

Wissen“ (Routinewissen). Zum Konzeptionswissen gehört einerseits das Wissen über Kommunikationsnormen bzw. -maximen, andererseits Weltwissen. Das Realisierungswissen beschreibt linguistisches Makro- und Mikrostrukturwissen (Textsortenwissen, Textordnungsmuster, syntaktisches und lexikalisches Wissen etc.), während das Routinewissen genuin schriftsprachliche Routinen umfasst, z.B. Schreibmotorik, Interpunktion, literale Routinen (vgl. Feilke und Augst 1989, 301ff.). Dieses Wissensmodell bezieht sich primär auf den Textproduktionsprozess, d.h. es beschreibt Wissen, das für das Verfassen von Texten relevant ist.

Eine dritte Variante (3) der Beschreibung von Wissen unterscheidet die Wissenskomponenten Sprachwissen, enzyklopädisches bzw. Sachwissen und Handlungs- bzw. Interaktionswissen (vgl. Viehweger 1989, 259f.; Heinemann und Viehweger 1991, 93f.) – dieses Modell bezieht sich sowohl auf die Produktions- als auch Rezeptionsperspektive. Das Sprachwissen umfasst Wissen über phonologische, syntaktische und semantische Regeln, aber auch Wissen über den Gebrauch der Sprache zur Strukturierung komplexerer sprachlicher Einheiten wie Texte, z.B. Wissen um die Verknüpfung von Äußerungen, Wissen um die Verteilung von Informationen im Satz und Text, Wissen um die Etablierung und Beibehaltung eines Themas im Text, Wissen, wie Propositionen kommuniziert werden und wie diese Propositionen als sprachliche Äußerungen interpretiert werden können. Das enzyklopädische Wissen bzw. Sachwissen bezieht sich auf Wissen um Sachverhalte in der Natur und Gesellschaft, Wissen um Handlungen, die Ereignissen zu Grunde liegen, d.h. sowohl gemeinschaftlich geteiltes Wissen als auch spezifisches Wissen, z.B. Fachwissen. Zum Interaktionswissen gehört das Wissen über kommunikative Normen bzw. Konversationsmaximen, Illokutionswissen (Wissen um die Ziele und Intentionen, die mit dem Vollzug eines Sprechaktes verbunden sein können), metakommunikatives Wissen (Wissen um die den Text bzw. die Rede organisierenden sprachlichen Handlungen und Sprechakte) und Wissen über globale Textstrukturen bzw. Textsorten (Wissen um Textmuster sowie Wissen um typische Sequenzierungen funktionaler sprachlicher Handlungen). Diese Wissensarten bzw. „Kenntnissysteme“ stellen keine statischen Wissensressourcen dar. Zu jedem Kenntnissystem gehört vielmehr auch prozedurales Wissen – dieses gewährleistet, dass die Kenntnissysteme im Prozess der Textproduktion (und -rezeption) im Hinblick auf die globalen Intentionen, die Handlungssituation und -beteiligten adäquat aktualisiert und angewendet werden (vgl. Heinemann und Viehweger 1991, 112f.).

Alternativ (4) kann Wissen über Zeichensysteme, Wissen über Außersprachliches und Wissen über die Kommunikationssituation unterschieden werden (vgl. Adamzik 2004a, 23ff.). Zum Wissen über Zeichensysteme gehört Wissen über nicht-sprachliche Zeichensysteme (z.B. Gesten, Mimik), Sprachsysteme (Grammatik und Lexik), Varietäten (z.B. Besonderheiten der Fachsprache), „geläufige Formen sprachlichen Handelns“ (die Kenntnis der Kommunikationsgewohnheiten, z.B. der typische Beginn von Telefongesprächen, die Sortierung von Lexika), Kenntnis von Einzeltexten wie Kinderreime, Liedtexte, Werbesprüche, Comicserien, Kultfilme, literarische oder biblische Texte, da diese als Teil des Gedächtnisses der Sprachgemeinschaft betrachtet werden können und als solche Voraus-

setzung für das Verstehen neuer Texte sind, die sich auf diese früheren Texte beziehen (z.B. „Hopp, hopp, hopp, Gysi, lauf Galopp“) (vgl. Adamzik 2004a, 23ff.). Außersprachliches Wissen kann in allgemein menschliches (Wissen, das sich auf Erfahrungen zurückführen lässt, die alle Menschen machen, z.B. Wissen über den eigenen Körper), kulturspezifisches Wissen, Wissen aus dem historisch-sozialen Kontext, Wissen aus gruppenspezifischen Erfahrungen und Wissen aus individuellen Erfahrungen unterteilt werden. Zum Wissen über die Kommunikationssituation gehört Wissen über den Kommunikationspartner, den (sozialen) Ort, die zeitlichen Umstände, den Situationstyp. Adamzik geht davon aus, dass sich die Wissensbestände eines Individuums kontinuierlich wandeln, auch innerhalb der aktuellen Kommunikationssituation, da z.B. in der Kommunikation neues Wissen (sowohl sprachlicher als auch außersprachlicher Art als auch Wissen über die Kommunikationssituation, z.B. den Kommunikationspartner) konstituiert werden kann, das dann in den Wissensbestand eingeht bzw. diesen aktualisiert (z.B. wird ein bisher unbekanntes Wort aus dem Kontext erschlossen; Vorstellungen, die man vom Kommunikationspartner hatte, werden revidiert). Im Gegensatz zu Heinemann/Viehweger (s.o.) und Viehweger (s.o.) fasst Adamzik die „Kommunikationssituation“ dennoch als Wissensbereich auf. Sämtliche textexternen Faktoren, die eine Rolle beim Produzieren und Verstehen des Textes spielen, lassen sich so einheitlich als Wissen modellieren und dem Text als ‚Datenbasis‘ gegenüberstellen.

In dieser Arbeit werden folgende Wissenskomponenten unterschieden: Sprachwissen, Sachwissen, Handlungs- bzw. Interaktionswissen sowie Wissen über die Kommunikationssituation. Wie Wissen kognitiv repräsentiert ist, darüber gibt es unterschiedliche Auffassungen. Zur Näherung an dieses Forschungsfeld soll im Folgenden zunächst ein Überblick über kognitive Schemata und Frames gegeben werden, der die Grundlagen für die formale Modellierung von Wissen in einem System zur maschinellen Themenidentifikation legt, bevor in Kapitel 2.7.3 schließlich eine Theorie zum Zusammenhang von Kognition, Textproduktion und Thema vorgestellt wird.

### **2.7.2 Kognitive Schemata und Frames**

Der Begriff des Schemas wurde 1932 von Bartlett eingeführt und definiert als „strukturierte Wissensbereiche im LZG [Langzeitgedächtnis; M.B.]“ (vgl. Schwarz 1996, 91). Diese Wissensstrukturen bestehen aus miteinander verknüpften, mental repräsentierten Konzepten. Konzepte stellen aus kognitionswissenschaftlicher Sicht mentale Einheiten dar, die entweder Klassen von Entitäten repräsentieren, z.B. die Klasse der „PFLANZEN“ (Type-Konzepte oder Konzeptklassen), oder einzelne Entitäten, z.B. „der Farn auf meinem Schreibtisch“ (Token-Konzepte oder Individualkonzepte); Konzepte werden auch als Grundbausteine des menschlichen Wissens aufgefasst (siehe Reimer 1991, 17ff.; Jürgens 1995, 250; Schwarz und Chur 2004, 24f.). Konzepte sind einerseits durch spezifische Merkmale determiniert, andererseits durch Beziehungen zu

anderen Konzepten (räumliche, zeitliche oder kausale Beziehungen etc.).<sup>125</sup>

Konzepte sind im Gedächtnis nicht isoliert abgespeichert, sondern sie sind durch verschiedene Relationen mit anderen Konzepten verknüpft. Die interkonzeptuellen Beziehungen stellen kognitive Strukturen dar, die Zusammenhänge eines Realitätsbereichs repräsentieren. (Schwarz 1996, 91)

Nach Jackendoff lassen sich Konzepte in bestimmte ontologische Gruppen einteilen, zu denen mindestens “[THING], [PLACE], [DIRECTION], [ACTION], [EVENT], [MANNER], and [AMOUNT]” sowie “[SMELL], [TIME]” gehören (Jackendoff 1988, 87, Hervorhebung im Original). Diese Kategorien beziehen sich auf das vom Textproduzenten konstruierte Modell der Welt, und lassen sich deshalb nicht auf physikalische oder metaphysische Gegebenheiten zurückführen, sondern auf psychologische. In ähnlicher Weise unterscheiden de Beaugrande/Dressler (1981, 101f.) „Primär“- und „Sekundärkonzepte“: „OBJEKTE“, „SITUATIONEN“, „EREIGNISSE“, „HANDLUNGEN“ (Primärkonzepte); „ZUSTAND“, „HANDLUNGSGEGENSTAND“, „ZEIT“, „QUANTITÄT“, „APPERZEPTION“ etc. (Sekundärkonzepte). Sekundärkonzepte entsprechen zum größten Teil Textrollen (siehe Fußnote 92); sie stellen keine unabhängigen Konzepte dar, sondern sind bezogen auf die Primärkonzepte.

Sowohl die Konzepte selbst als auch die Zusammenhänge zwischen ihnen werden durch Erfahrungen (sowohl sprachlich als auch motorisch oder visuell) erworben und mental als kognitive Schemata abgespeichert.

Unter einem *Schema* versteht man ein Modell für eine Gedächtnisstruktur, das nicht allein Assoziationen zwischen Begriffen berücksichtigt, sondern dem Phänomen *stereotypischer Erinnerungsmuster* besonders Rechnung trägt. (Reimer 1991, 159, Hervorhebung im Original)

Diese „stereotypischen Erinnerungsmuster“ werden im Lauf des Lebens erworben, stellen jedoch keine geschlossenen Muster dar, sondern können durch neue Erfahrungen immer wieder modifiziert werden. Schemata sind demnach variabel und flexibel, so dass auch Informationen verarbeitet werden können, die von den stereotypischen (bzw. prototypischen) Mustern abweichen können.<sup>126</sup> Van Dijk weist darauf hin, dass die Struktur eines Schemas insbesondere davon ab-

<sup>125</sup> Solche Relationen sind z.B. innerbegriffliche Relationen wie Über- und Unterordnung, Kontrastrelation, Qualitätsmerkmal, komparative Relation, und zwischenbegriffliche Relationen wie Handlungsträgerrelation, Aktorrelation, Lokalisationsrelation, Instrumentrelation, Objektrelation, Finalitätsrelation (vgl. Heinemann und Viehweger 1991, 69).

<sup>126</sup> In der vorliegenden Arbeit wird der Begriff des Stereotyps im Sinne von Putnam (1990 [1975], 64ff.) gebraucht. Er wird definiert als konventionale und sozial verbindliche Vorstellung über einen Prototyp (wobei der Begriff der „Vorstellung“ wiederum als Idee oder auch mentale Repräsentation zu verstehen ist, vgl. Herrmanns 2002, 292), die die Menge der Eigenschaften oder „Kernfakten“, die diesen Prototyp ausmachen, berücksichtigt. Ein Prototyp wird dabei als typischer Referent im Hinblick auf eine bestimmte Kategorie verstanden. Zu den stereotypischen Eigenschaften des Wortes „Wasser“ gehören z.B. farblos, transparent, geschmacklos, durstlöschend. Diese stereotypischen Eigenschaften resultieren dabei aus einer Alltagstheorie (d.h. sie müssen aus wissenschaftlicher Sicht nicht unbedingt zutreffend sein) bzw. aus den im Alltag erworbenen Erinnerungsmustern bzw. dem kollektiven Wissen über einen Gegenstandsbereich oder Sachverhalt (vgl. Konecny 2001, 157ff.; Kosta 2009, 831f.). Stereotype stellen demnach stereotypisches Wissen bzw. kollektives Weltwissen dar, das in Schemata oder Frames repräsentiert werden kann.

hängt, auf welche Art und Weise wir etwas wahrgenommen und verstanden haben – dies bestimmt, „in welcher ‚Schublade‘, auf welcher ‚Ebene‘ und innerhalb welcher umfassenderen Struktur diese Information im Gedächtnis behalten wird“ (vgl. van Dijk 1980, 169). Die Bestimmung von Schemata als stereotypischen Erinnerungsmustern bedeutet, dass mit der Aktivierung eines Schemas auch eine bestimmte Erwartungshaltung aktiviert wird: Ein Rezipient erwartet nach der Aktivierung eines spezifischen Schemas das Vorkommen typischer Objekte und Ereignisse (bzw. sprachlicher Ausdrücke, die zur Referenz auf diese gebraucht werden) und interpretiert Objekte und Ereignisse seiner Umwelt schemaabhängig. Reimer (1991, 159) bringt als Beispiel die Beobachtung eines schwarzen, krähengroßen Vogels – hat man das Schema „Hochgebirge“ aktiviert, wird der Vogel eher als Alpendohle gesehen, ist das Schema „Meeresküste“ aktiviert, würde der Vogel dagegen als Kormoran interpretiert.<sup>127</sup>

Es ist davon auszugehen, dass das Wissen über Konzepte und deren Relationen – die Grundlage kognitiver Schemata – interindividuell leicht divergent ist. Dies liegt daran, dass Konzepte vage sind und prototypische Einheiten darstellen (vgl. Schwarz 1996, 90; Vater 2005, 58). Je nach den Erfahrungen, die ein Mensch gemacht hat, können die Konzepte deshalb interindividuell unterschiedlich abgespeichert sein, die Zuordnung von Token-Konzepten zu Type-Konzepten graduell unterschiedlich sein. Allerdings muss davon ausgegangen werden, dass der Kern eines Konzeptes (das das Konzept „determinierende Wissen“, vgl. de Beaugrande und Dressler 1981, 89ff.) innerhalb einer sozio-kulturell konstituierten Sprechergemeinschaft identisch ist, weil die Erfahrungen der Mitglieder dieser Gemeinschaft sehr ähnlich sind.

To the extent that our conceptualizations are alike, it is because they are constrained by the way our minds have evolved to deal with the world, in particular to construe the world in terms of ‘ordinary objects’ and the like. On the other hand, when we deal with topics like religion and semantic theory, which are not as constrained by evolution, there really is a possibility of radical difference in conceptualization, with accompanying loss of mutual understanding. (Jackendoff 1998, 216)

Konzepte von Spezial-Domänen unterscheiden sich interindividuell stärker voneinander als Konzepte, die mit den alltäglichen Gegenständen der Welt zu tun haben. Das erklärt auch, dass Fachtexte von Experten anders verstanden werden als von Laien, da diese keine fachspezifischen Konzeptualisierungen der referenzierten Entitäten besitzen – andererseits können Fachtermini auch unter Experten mit unterschiedlichen Konzeptualisierungen belegt sein; dies zeigt sich z.B. deutlich an der Vielzahl der Konzepte, die mit dem Terminus Thema verknüpft sind (siehe Kapitel 2). Unterschiedliche Konzeptualisierungen führen dazu, dass Texte interindividuell unterschiedlich verstanden werden können. Die mentalen Modelle des Textes, die der Textproduzent und die verschiedenen Rezipienten des Textes entwickeln, können demnach niemals hundertprozentig identisch

<sup>127</sup> In ähnlicher Weise interpretiert auch van Dijk die Funktion von Frames bzw. „Rahmen“ als notwendig für die „korrekte Interpretation unterschiedlichster sozialer Ereignisse, für die adäquate eigene Teilnahme an solchen Ereignissen und im allgemeinen für die Sinnggebung unseres eigenen Verhaltens und dessen der anderen“ (van Dijk 1980, 169).

sein (vgl. van Dijk 2008, 59f.).

Auf Konzepte kann sprachlich durch den Akt des „Referierens“ Bezug genommen werden. Eine relativ neutrale Definition von „Referenz“ liefern Schwarz/Chur:

Referenz ist die Relation, die in einer bestimmten Situation zwischen sprachlichen Ausdrücken und Gegenständen besteht, bzw. etabliert wird. Nicht Wörter per se referieren, sondern Sprecher referieren mit der Hilfe von Wörtern, und Hörer etablieren Referenz im Verstehensprozess. (Schwarz und Chur 2004, 83)

Zum Konzept „Referenz“ gibt es allerdings sehr unterschiedliche Auffassungen, z.B. darüber, mit welchen Ausdrücken referiert werden kann (nur mit Nominalphrasen – definit und indefinit?<sup>128</sup> – und Präpositionalphrasen, oder auch mit Verbalphrasen, Adverbialphrasen oder sogar Sätzen?) und worauf referiert wird (auf Gegenstände in der Welt oder einer möglichen Welt oder auf Konzepte im Modell der Welt des Sprechers?).<sup>129</sup> In dieser Arbeit wird der Standpunkt von Jackendoff vertreten: Laut Jackendoff reicht es nicht aus, dass eine Entität in der Welt existiert, um auf sie referieren zu können, sondern sie muss auch konzeptualisiert und in das vom Sprecher imaginierte Modell der Welt eingeordnet werden. Umgekehrt kann auch auf Entitäten in der imaginierten Modellwelt referiert werden, die nicht in der realen Welt existieren.

We cannot refer to an entity at all without forming a mental representation that enters the entity in question into the world as we currently conceptualize it. (...) Therefore an entity's being in the real world is not a sufficient condition for being able to refer to it. (Jackendoff 1998, 214)

Mit sprachlichen Ausdrücken wird nach Ansicht von Jackendoff demnach auf Entitäten in der konzeptualisierten Welt des Sprechers referiert. Solche Entitäten können auch abstrakte Einheiten wie Propositionen, Mengen, Zahlen und Prädikate sein, die nicht notwendigerweise einer Einheit in der realen Welt entsprechen müssen. Dementsprechend können nicht nur Nominalphrasen, sondern auch andere Phrasen wie Präpositionalphrasen, Adjektivphrasen oder sogar Sätze referenziell gebraucht werden. Auf Basis dieser Annahmen unterscheidet Vater (2005, 70ff.) folgende Referenzbereiche: Situationsreferenz (ein Satz kann z.B. auf eine Situation referieren), Dingreferenz (in erster Linie durch Nominalphrasen und Pronomina), Ortsreferenz, Zeitreferenz (durch Tempora des Verbs, Temporaladverbiale und Temporalsätze), Eigenschaftsreferenz (durch Adjektive), Modalitätsreferenz (durch Modalverben oder Modaladjektive), Quantitätsreferenz.

In den 70er Jahren des 20. Jahrhunderts wurde das Konzept des Schemas von Kognitionswissenschaftlern wie Rumelhart und Kintsch im Detail ausgearbeitet.

---

<sup>128</sup> Karttunen (2003) zufolge können sowohl definite als auch indefinite NPs Diskursreferenten etablieren – letztere allerdings nur unter spezifischen Bedingungen, nämlich dann, wenn im weiteren Textverlauf ein koreferentes Pronomen oder eine koreferente definite NP auftaucht. Diskursreferenten sind Individuen, auf die im Diskurs mittels unterschiedlicher sprachlicher Ausdrücke Bezug genommen werden kann. Die Instanzen eines Diskursreferenten sind demnach koreferent.

<sup>129</sup> Einen umfassenden Einblick in aktuelle Forschungsfragen, -richtungen und -ergebnisse gibt Vater (2005).



Dabei wurden mehrere alternative Konzepte vorgeschlagen, die zwar alle als schematheoretische Begriffe einen Bezug zum ursprünglichen Schema-Begriff aufweisen, aber dennoch jeweils andere Modellierungen des Schema-Konzepts darstellen. Zu den bekanntesten dieser Begriffe gehören „Frames“ (Minsky, Charniak, Fillmore), „Szenarios“ (Sanford/Garrod), „Skripts“ (Schank, Schank/Abelson, van Dijk) und „mentale Modelle“ (Johnson-Laird) (siehe de Beaugrande und Dressler 1981, 95f.; Schwarz 1996, 91f.; Heinemann 2000b, 91). Wie Konerding (1993, 21f.) deutlich macht, finden sich zwischen diesen Arbeiten und Definitionen viele Widersprüche und Unklarheiten. Da der Begriff des „Frames“ am Evaluiertesten zu sein scheint, wird sich die vorliegende Arbeit auf dieses Konzept konzentrieren.

Van Dijk (1980, 169) zufolge stellen Frames bzw. Rahmen bestimmte „Organisationsformen für das konventionell festgelegte Wissen, das wir von der ‚Welt‘ besitzen“, dar. Bei diesem Wissen handelt es sich um „Regelmäßigkeiten, Konventionen, Normen, Personen, Rollen, Funktionen, Handlungen u.ä., die in sozialen Situationen eine Rolle spielen“. In ähnlicher Weise definiert auch Minsky Frames als sozial und situativ gebundenes Wissen: „A frame is a data-structure for representing a stereotyped situation like being in a certain kind of living room or going to a child’s birthday party. Attached to each frame are several kinds of information.“ (Minsky 1980, 1). Ein Frame repräsentiert demnach Informationen zu einer spezifischen sozialen Situation. Frames stellen damit ein Bindeglied zwischen der in Kapitel 2.6.2 entworfenen Texttheorie dar, der zufolge sprachliche Handlungen bezogen sind auf konventionalisierte kommunikative Situationen (die gleichzeitig soziale Situationen im Sinne von Minsky und van Dijk darstellen), auf die Bezug genommen wird durch die auf Ebene der Äußerungseinheiten wahrnehmbaren und interpretierbaren indexikalischen Zeichen. Wie dargestellt, bestehen zwischen diesen indexikalischen Zeichen (bzw. den sprachlichen Ausdrücken und Strukturen, die als indexikalische Zeichen interpretiert werden können) Nachbarschafts- bzw. Assoziationsbeziehungen, die als äquivalent zu Frame-Relationen aufzufassen sind. Das in Frames strukturierte und repräsentierte Wissen kann demnach auch als Wissen über *kommunikative* Situationen und deren Eigenschaften (inklusive der zur Bezugnahme auf diese Situation verwendeten, konventionalisierten sprachlichen Mittel, die die Funktion indexikalischer Zeichen haben) aufgefasst werden.

Das in Frames repräsentierte Wissen bzw. die repräsentierten Informationen bestehen, wie bei den kognitiven Schemata, aus Knoten und Relationen zwischen diesen Knoten und sind sowohl netzartig (durch Querbezüge zwischen Konzepten) als auch hierarchisch strukturiert, insofern ein Frame aus „top level“- und „low level“-Informationen besteht.<sup>130</sup>

---

<sup>130</sup> Knoten (bei van Dijk spezifischer: Konzepte) und Relationen zwischen diesen Knoten sind im Gedächtnis hierarchisch gegliedert sind. Diese hierarchische Strukturierung ist Voraussetzung für die effektive Arbeit des Gedächtnisses, da auf diese Weise Eigenschaften höherer Ebenen an untere Ebenen vererbt werden können – das Wissen über Frauen im Allgemeinen (+ menschlich, + potentielle Mutter, + sichtbar, etc.) kann übertragen werden auf „Petra“ (vgl. van Dijk 1980, 168).

We can think of a frame as a network of nodes and relations. The 'top levels' of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals – 'slots' that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet. (...) Simple conditions are specified by markers that might require a terminal assignment to be a person, an object of sufficient value, or a pointer to a sub-frame of a certain type. More complex conditions can specify relations among the things assigned to several terminals. (Minsky 1980, 1)

Während die oberen Ebenen eines Frames also allgemeine Informationen enthalten, weisen die unteren Ebenen Informationslots auf, die in einer aktuellen kommunikativen Situation mit konkreten Informationen gefüllt werden müssen. Dabei stellt sich die Frage, ob die "Filler" der "Slots" der unteren Ebenen mental repräsentiert sind oder erst bei der Aktivierung des Frames in einer spezifischen Situation durch Informationen aus dem Kontext gefüllt und mental repräsentiert werden.<sup>131</sup> Da die "Terminals" der unteren Ebenen in erster Linie über ihre Bedingungen definiert sind, ist eher davon auszugehen, dass die Filler erst in einem spezifischen Kontext instanziiert werden.<sup>132</sup> Andererseits gibt Minsky an, dass "a frame's terminals are normally already filled with 'default' assignments" und dass diese Standardbelegungen ("default assignments") oder Variablen ("variables") in einer spezifischen Situation einfach durch passendere Einheiten ersetzt werden können (Minsky 1980, 2). Diese Aussage würde deshalb eher stützen, dass alle Knoten in einem Frame mental repräsentiert sind, dass jedoch die Knoten der oberen Ebenen im Hinblick auf ihre Werte festgelegt sind, während die Werte der Knoten der unteren Ebenen variabel sind, wobei diese Variabilität beschränkt ist durch einfache oder komplexe Bedingungen, die erfüllt werden müssen. Denkbar wäre deshalb, dass die Terminals der unteren Ebenen Token- bzw. Individual-Konzepten entsprechen, die über Bedingungen als Teil einer bestimmten Klasse bzw. eines Type-Konzepts definiert werden.

Nach Minsky können Terminals auch als Fragen aufgefasst werden, die eine bestimmte soziale Situation zu strukturieren erlauben. So ist beispielsweise mit einer „Geburtstagsparty“ verbunden, dass man dem Gastgeber G ein Geschenk P kauft, das die Bedingung erfüllt, dass der Gastgeber dieses Geschenk mag (Slot = „Wird G P mögen?“). Durch die Aktion „Kaufen“ wird wiederum ein Sub-Frame aktiviert, bei dem die folgenden Slots gefüllt werden müssen: „Wo kann ich P kaufen?“, „Wie komme ich an das Geld, mit dem ich P kaufen will?“ Zum Hauptframe „Geburtstagsparty“ gehören außerdem Slots wie: „Was ziehe ich an?“ etc. Indem Slots der unteren Ebenen als Fragen definiert werden, lassen sich Bedingungen formulieren, die ein Filler erfüllen muss, ohne die möglichen Filler zu eng einzuschränken (wie das z.B. bei einer Aufzählung der Fall wäre).

---

<sup>131</sup> Nach Konerding entspricht die obere Ebene eines Frames dem semantischen Wissen, das propositional repräsentiert ist und Informationen über die „konstitutiven Bestandteile typischer Situationen“ darstellt, während die untere Ebene episodisches Wissen enthält (vgl. Konerding 1993, 26).

<sup>132</sup> Dies deckt sich mit der Beschreibung von Slots und Fillern bei Schwarz (1996, 92), der zu Folge Slots „konzeptuelle Einheiten“ darstellen und Filler „konkrete Werte“, die erst im Verstehensprozess besetzt werden.

Minskys Verdienst ist – unabhängig von seiner eher intuitiven Bestimmungen des Framebegriffs und der Bestandteile von Frames – insbesondere die Anbindung der Frametheorie an die Linguistik. Er geht zum einen davon aus, dass die durch Frames repräsentierten Wissensinhalte durch sprachliche Ausdrücke evokiert werden. Zum anderen integriert er die das Wissen über eine soziale Situation repräsentierenden Frames in ein linguistisch komplexeres Modell von Frames, das Wissen über Texttypstrukturen und Vertextungsmuster sowie syntaktisches und semantisches Wissen repräsentiert:

The key words and ideas of a discourse evoke substantial thematic or scenario structures, drawn from memory with rich default assumptions.

In any event, the individual statements of a discourse lead to temporary representations – which seem to correspond to what contemporary linguists call ‘deep structures’ – which are then quickly rearranged or consumed in elaborating the growing scenario representation. In order of ‘scale’ among the ingredients of such a structure there might be these kinds of levels:

Surface Syntactic Frames – Mainly verb and noun structures. Prepositional and word-order indicator conventions.

Surface Semantic Frames – Action-centered meanings of words. Qualifiers and relations concerning participants, instruments, trajectories and strategies, goals, consequences and side-effects.

Thematic Frames – Scenarios concerned with topics, activities, portraits, setting. Outstanding problems and strategies commonly connected with topics.

Narrative Frames – Skeleton forms for typical stories, explanations, and arguments. Conventions about foci, protagonists, plot forms, development, etc., designed to help a listener construct a new, instantiated Thematic Frame in his own mind. (Minsky 1980, 16)

Wie diese verschiedenen Wissensstrukturen miteinander interagieren, wird von Minsky leider nicht weiter ausgeführt. Hervorzuheben ist, dass er nicht davon ausgeht, dass Frames lediglich außersprachliches Wissen repräsentieren, sondern dass Frames gleichermaßen geeignet sind, sprachliches Wissen zu repräsentieren, das, wie das außersprachliche Wissen, zum Aufbau eines Situationsmodells, Szenarios oder mentalen Modells des Textes beiträgt.

Auf die Möglichkeit der Anwendung des Framekonzepts in der Linguistik verweist auch Fillmore (2001; 2006 [1982]). Bereits Anfang der 70er Jahre spricht er sich wiederholt für die Berücksichtigung von Frames in der Linguistik aus, insbesondere im Bereich der lexikalischen Semantik und des Textverstehens. Unter dem Terminus “Frame Semantics” beschäftigt sich Fillmore mit dem Zusammenhang von Frames und Wortbedeutungen. Zunächst nutzt er Frames (spezifischer: oberflächenbasierte syntaktische Frames), um Klassen von Wörtern in paradigmatischen Beziehungen zu modellieren. Diesen syntaktischen Frame-Ansatz weitet er aus, indem er im Rahmen der “Case Grammar” (Kasusgrammatik“) zusätzlich semantische Rollen (bzw. Kasusrollen, siehe Kapitel 2.2.1.3) und semantische Klassen (z.B. Bewegung, Wahrnehmung, Verursachung etc.) zur Gruppierung von Verben in Frames nutzt. Trotz dieser syntaktischen und semantischen Ausrichtung verknüpft Fillmore die “Case Frames” mit der Vorstellung abstrakter Szenen oder Situationen und verbindet so semantische mit kognitiven Ideen: “To understand the semantic structure of the verb it was necessary to understand the properties of such schematized scenes” (Fillmore 2006 [1982], 377). In den 70er Jahren beschäftigt sich Fillmore intensiver mit kognitiven Strukturen, um Domänen von Vokabular semantisch zu beschreiben.

Zu diesem Zweck geht er wieder von abstrakten Szenen aus, für die teilnehmende Personen, Handlungen und Rahmensituationen angegeben werden können, denen wiederum bestimmte Verben zugeordnet werden (z.B. Verben der Domäne „Recht“: „anklagen“, „beschuldigen“ etc.). Ein Frame strukturiert demnach Wörter bzw. Wortbedeutungen<sup>133</sup>:

The frame structures the word-meanings, and (...) the word 'evokes the frame.

The structures I have mentioned so far can be thought of as motivating the categories speakers wish to bring into play when describing situations. (Fillmore 2006 [1982], 378).

Textproduzenten verwenden demnach bestimmte Wörter, um einen bestimmten Frame zu signalisieren. Für die Rezeption hat dies zur Folge, dass durch das Wortmaterial des Textes eben dieser Frame vom Rezipienten aktiviert und zum Aufbau der Textwelt genutzt wird (Fillmore 2006 [1982], 382f.). Neben den Wörtern aktivieren aber auch grammatische Konstruktionen ein bestimmtes Szenario, einen bestimmten Frame – die Zeitform Präteritum evoziert beispielsweise, dass die geschilderte Situation in der Vergangenheit liegt und abgeschlossen ist.

Neben diesen lexikalische und grammatische Einheiten und Strukturen abbildenden Frames existieren nach Fillmore auch interaktionale Frames (“interactional frames”), die schematisch die kommunikative Situation, in der ein Text produziert wird, repräsentieren (“we conceptualize what is going on between the speaker and the hearer, or between the author and the reader”); zu den Faktoren, die die kommunikative Situation strukturieren, gehören nach Fillmore u.a. das Wissen über Konversationsprinzipien und Sprechaktroutinen sowie Wissen über Textmuster und Typen thematischer Entwicklung (Fillmore 2006 [1982], 379).

Sowohl Fillmore als auch Minsky gehen also davon aus, dass Frames nicht nur außersprachliches Wissen repräsentieren können, sondern auch sprachliches bzw. interaktionales Wissen. Beide sind sich darüber hinaus einig, dass Frames durch sprachliches Material evoziert werden. Der große Unterschied zwischen Fillmore und Minsky besteht darin, dass Fillmore keine Trennung von Weltwissen und sprachlichem Bedeutungswissen vornimmt und dass die Bestandteile von Frames als linguistische Einheiten aufgefasst werden können.

Frames im Sinne von Minsky und insbesondere Fillmore sind demnach hervorragend geeignet als Bindeglied zwischen der Ebene der Kognition und Kommunikation: Wie in Kapitel 2.6.2 dargestellt, wird angenommen, dass in Texten auf kommunikative Erfahrungen bzw. konventionalisierte kommunikative Situationen Bezug genommen wird, und dass diese Bezugnahme an der Textoberfläche durch bestimmte wahrnehmbare Phänomene, die als indexikalische Zeichen für eine kommunikative Situation interpretiert werden können, realisiert wird. Fill-

---

<sup>133</sup> Die Bedeutung eines Wortes wird nach Fillmore immer vor einem (prototypischen) Hintergrundwissen zu den Institutionen, Praktiken und Gebräuchen einer Gesellschaft definiert – um ein Wort wie „Wochenende“ verstehen zu können, muss man z.B. wissen, dass unsere kalendrische Woche sieben Tage umfasst und dass zwei dieser Tage als freie Zeit für das Privatleben gedacht sind, an denen nicht gearbeitet wird.

more spricht in diesem Zusammenhang davon, dass Textproduzenten einen Frame auf eine aktuelle kommunikative Situation anwenden und dies dem Rezipienten signalisieren, indem sie Wörter oder grammatische Strukturen verwenden, die vom Rezipienten mit diesem Frame assoziiert werden: “a speaker ‘applies’ a frame to a situation, and shows that he intends this frame to be applied by using words recognized as grounded in such a frame.” (Fillmore 2006 [1982], 382). Wörter und grammatische Strukturen evozieren nach Fillmore demnach Frames und damit zugleich kommunikative Situationen. Dieses sprachliche Material hat Fillmore zufolge eine Indexfunktion, insofern es als Indiz für einen bestimmten Frame interpretiert werden kann:

(...) we have cases in which the lexical and grammatical material observable in the text ‘evokes’ the relevant frames in the mind of the interpreter by virtue of the fact that these lexical forms or these grammatical structures or categories exist as indices of these frames. (Fillmore 2006 [1982], 385)

Über das Konzept der kognitiven Frames können demnach Interaktionen zwischen der Ebene der Kognition und der Kommunikation modelliert werden.

Im folgenden Kapitel soll ein etwas anderer Frame-Ansatz, der Frames als Listen von Themen beschreibt, eingeführt werden. Diese Frame-Konzeption baut auf Minsky auf, ist jedoch noch näher an linguistischen und insbesondere lexikologischen Forschungen orientiert und baut insofern ebenfalls eine Brücke zwischen den Ebenen der Kognition und Kommunikation, die insbesondere für die Frage nach der Rolle des Themas auf diesen Ebenen von Interesse ist.

### 2.7.3 Kognition und Textproduktion: Frames als Listen von Themen

Zur Modellierung des Zusammenhangs von Thema, Sachwissen und den kommunikativen Prozessen der Textproduktion und -rezeption wird in dieser Arbeit auf das – auf Minsky aufbauende – linguistische Framekonzept<sup>134</sup> von Konerding (1993, 143) zurückgegriffen. Dieses Framekonzept basiert auf der Annahme, dass Frames modelliert werden können als Listen von Fragen zu stereotypischen Situationen. Diese Fragen haben dabei den Status von Themen, an denen sich der Textproduzent bei der Textproduktion orientiert und die er für die Kommunikation eines Sachverhaltskomplexes nutzt. Die Annahme von impliziten Fragehandlungen als Gerüst für die Textproduktion ist zwar nicht neu (siehe Kapitel 2.2.3.2),<sup>135</sup> allerdings wird in diesem Ansatz zum ersten Mal ein Zusam-

<sup>134</sup> „In Übereinstimmung mit Minsky ist ein Frame als ein sprachlicher Text zu rekonstruieren, der eine Menge (epistemisch fungierender) Fragen bereitstellt.“ (Konerding 1993, 161).

<sup>135</sup> In den Arbeiten von Hellwig und van Kuppevelt (z.B. Hellwig 1984a; van Kuppevelt 1995) werden Themen als „Fragliches“ definiert, die in einem Text oder Textabschnitt beantwortet werden. Das Thema wirkt auf diese Weise als produktives Prinzip auf die Textproduktion ein; der Zusammenhang mehrerer Fragen (oder Themen) konstituiert dabei das Grundgerüst des Textes und strukturiert ihn in linearer (Hellwig) bzw. hierarchischer (van Kuppevelt) Weise. „Ein Textproduktionsplan stellt also ein Gerüst von interdependenten Themen bereit, das eine Sequenz von Handlungen vom Typ des Fragens, des Suchens, Findens und Vergleichens sowie approximierenden Verbindens unter Bezug auf Texturen verschiedener Konzepte leitet. Insbesondere zeigt es sich jetzt, daß nicht Makropropositionen im Sinne von Kintsch und van Dijk konstitutiv

menhang zwischen Fragehandlungen als Textproduktionshandlungen und der Repräsentation von Wissen in Frames<sup>136</sup> hergestellt. Themen als Fragen besitzen dieser Auffassung zufolge eine handlungsanweisende Funktion. Typische textkonstitutive Fragehandlungen sind z.B. „Warum ist das geschehen?“, „Wie geht es nun weiter?“, „Was ist das?“. Korrespondierende Handlungsanweisungen für die erste Frage wären z.B.: „Stelle fest, warum das geschehen ist. Stelle die Gründe/Ursachen für das Geschehnis fest. Stelle die Regel fest, die dem Geschehnis zugrunde liegt. Suche/Finde die Gründe, die Regel.“ (vgl. Konerding 1993, 145). Über diese Fragen und Handlungsanweisungen wird nach Konerding „stillschweigendes Wissen“ aktiviert – über die Rekonstruktion der Fragen und der darauf basierenden Handlungsanweisungen lässt sich dieses stillschweigende Wissen erschließen. Der Zusammenhang zwischen Frames und Fragen bzw. Handlungsanweisungen als Textgerüsten lässt sich auf Basis dieser Annahmen wie folgt formulieren: Frames repräsentieren Fragen bzw. Handlungsanweisungen zu einer stereotypischen Situation; diese Fragen eröffnen gleichzeitig Themen. Insgesamt stellen Frames aufgrund dieser Merkmale Textgerüste dar:

Die Frames werden (...) als Listen ‚strategisch entscheidender‘ Fragen (i.S.v. Hintikka 1985:252) konzipiert. Sie steuern als Themen nach Hellwig (1984) eine thematisch-progressive Textproduktion im Sinne einer dynamischen (Re)Präsentation bzw. Rekonstruktion oder Modellierung des stillschweigenden Wissens zu einer Sache, die typischerweise mittels eines Substantivs (...) benannt und klassifiziert ist. (Konerding 1993, 172)

Als grundlegendes Problem dieser Framekonzeption stellt sich dabei die systematische, wissenschaftlich fundierte Bestimmung der für einen Frame grundlegenden Fragen dar – Hellwig zufolge entzieht sich das, was sinnvolle Fragen sein können, der Objektivität: „Welche Fragestellungen ein Leser in einem Text beantwortet sieht, hängt nicht zuletzt von seinem Interesse und seiner Aufmerksamkeit ab.“ (Hellwig 1984a, 70).<sup>137</sup> Konerding geht im Gegensatz zu Hellwig davon aus, dass man hoffen kann, „zumindest das *stereotypische* Wissen über *typische* Fragestellungen erschließen zu können“ (Konerding 1993, 150, Hervorhebung im Original). Stereotypisches Wissen lässt sich seiner Ansicht nach aus sprachlichen Strukturen ableiten, insofern die „Nennlexik der wesentliche Ausdruck kulturell identifizierter, kontrollierter und damit der stereotypischen Texturtypen/Konzepte ist“ (Konerding 1993, 150). Sprachliche Strukturen spiegeln demnach kognitive Strukturen wider und können deshalb zur Erschließung des stereotypischen Wissens genutzt werden, und zwar mittels eines empirisch erstellten Verbthesaurus mit Gruppen von Verbschemata, die auch als Prädikato-

---

für einen Produktionsplan sind (vgl. Eigler u.a. 1990:39ff.), sondern Themen im Sinne von Hellwig.“ (Konerding 1993, 148).

<sup>136</sup> Allerdings bezieht sich Konerding im Gegensatz zu Minsky und Fillmore ausschließlich auf nominales Wissen, nicht jedoch auf syntaktisches oder interaktionales Wissen (syntaktischer Frame, Wissen über Vertextungsmuster/narrativer Frame etc.).

<sup>137</sup> Diese Problematik zeigt sich z.B. bei Rothkegel (1997, 464f.), die ohne weitere Begründung zwischen einfachen Themafragen („Wer oder was ist X?“), erweiterten einfachen Themafragen („Was ist X in einer bestimmten Zeit?“) und komplexen Themafragen („Was oder wie muss X (etwas) tun, um y zu erreichen?“) unterscheidet, und gleichzeitig darauf hinweist, dass der Themafragetyp festlegt, wie ein Text beginnen muss und wie er fortgesetzt werden kann, d.h. dass die Form der Themafrage die thematische Textstruktur beeinflusst.

renschemata (z.B. „sich fortbewegen – jd 1 zu Fuß von ort 3 nach ort 3“<sup>138</sup>) fungieren können, insofern die Verbschemata bzw. die allgemeineren Verbmodelle (= Gruppen von Verbschemata zu einem spezifischen Bereich) als Material für die Konstruktion von typischen Fragen (den Slots des Frames) genutzt werden können.

Die Verbschemata (...) fungieren als Präsuppositionen von Fragen bzw. Anwendungsbedingungen für Fragetypen: die durch die syntaktischen Valenzslots der Verbschemata geforderten Ergänzungen sowie die fakultativen Angaben werden traditionellerweise über Fragen bestimmt. (Konerding 1993, 168)

Die Verbschemata des Verbthesaurus können nach Konerding direkt in Fragen transformiert werden. Für das Verbschema „seinen Ursprung haben etw 1 in etw 3“, das dem Substantiv „Gegenstand“ zugeordnet werden kann, lässt sich das Prädikationsschema „Ein Gegenstand / hat seinen Ursprung / in etw 3 / an Ort / zur Zeit / unter den Umständen ...; auf die Art und Weise ... / aus dem Grund .../“ bestimmen, das wiederum zur Bestimmung von Fragen (Ergänzungsfragen und Angabenfragen) genutzt werden kann, die den Frame „Gegenstand“ konstituieren: „Worin hat der Gegenstand seinen Ursprung?“, „Unter welchen besonderen Bedingungen (wann, wo)?“, „Auf welche Art und Weise (wie)?“, „Aus welchem Grund (warum)?“ (vgl. Konerding 1993, 171).<sup>139</sup> In ähnlicher Weise geht auch Jahr (1996) davon aus, dass die Valenzstruktur eine Erklärung für semantische und kognitive Zusammenhänge bietet und „valenzbedingte Aktanten“ den Textzusammenhang determinieren, insofern sie über logisch-semantische Relationen (bei Konerding: Fragen, die sich aus den Ergänzungen und Angaben aus dem Verbschema ergeben) mit anderen Aktanten verknüpft sind, z.B. „Wurzelschäden“ als URSACHE („Warum“) für „fehlende Mykorrhiza“, „Feinwurzelverluste“ als Explikation der ART UND WEISE („Wie“) der „Wurzelschäden“, „Störung der Nährstoff- und Wasserversorgung“ als FOLGE von „Wurzelschäden“. Nach Jahr gehören diese „Bedeutungseinheiten“ thematisch zusammen und bilden „kognitive thematische Untereinheiten eines komplexen Sachverhalts“, wobei der dominierende valenzbedingte Aktant als „Schlüsselbegriff“ fungiert (in diesem Fall: „Wurzelschäden“) (vgl. Jahr 1996, 127).

Neben der Abstraktion von Verbschemata durch die Bildung von Bedeutungsklassen, die wiederum zu Verbmodellen (z.B. Bewegungsmodell) zusammengefasst werden können, lassen sich auch die Substantive einer Sprache zu Ähnlichkeitsgruppen zusammenfassen und durch Hyperonyme reduzieren, z.B. „Ge-

<sup>138</sup> Der von Konerding verwendete Verbthesaurus wurde von Ballmer/Brennenstuhl (1986) erstellt. Die Abkürzung „jd 1“ steht für „jemand im Nominativ“, die Zahl „3“ signalisiert den Dativ. Das Beispiel stellt eine gemeinsame Paraphrase bzw. Implikation mehrerer Schemata zu den Verben „schleichen“, „zuckeln“, „gehen“, „trotten“, „marschieren“, „rennen“ und „sprinten“ dar (= Verben einer Bedeutungsklasse). Die abstrakte Paraphrasierung dient dabei der Etikettierung einer Bedeutungsklasse oder Kategorie. Diese Bedeutungsklassen oder Kategorien können wiederum in Ähnlichkeitsgruppen sortiert und in einem Modell angeordnet werden (z.B. einem Modell der Bewegungsverben). Insgesamt unterscheiden Ballmer/Brennenstuhl vierzehn verschiedene Gruppen von Modellen, z.B. „Bewirken“ („Bewirkungsmodell“), „Kontrolliertes Eigentum“ („Haben-Modell“, „Kauf-und-Verkaufen Modell“) (vgl. Konerding 1993, 154f.).

<sup>139</sup> Das Prädikationsschema lässt sich in Ergänzungen (etw 1 = „Gegenstand“, etw 3) und (fakultative) Angaben (Ort, Zeit, Art und Weise, Grund/Zweck) aufteilen.

genstand“, „Organismus“, „Person/Aktant“, „Ereignis“, „Handlung/ Interaktion/ Kommunikation“ – insgesamt unterscheidet Konearding zehn primäre und drei sekundäre Typen von Substantiven. Sowohl die Verb- als auch die Substantivgruppen müssen empirisch aus der Sprache bzw. großen Textkorpora ermittelt werden. Im Gegensatz zu Ontologien ist die Erstellung dieser Typologien sprachbasiert und nicht philosophisch begründet.

Die Konstruktion von Substantivtyp-Frames orientiert sich also zum einen an Substantivtypen, zum anderen an den syntaktischen Valenzstellen von Verb-schemata: Für einen Substantivtyp (bzw. die Token eines Typs) wird ein Frame konstruiert; die Slots dieses Frames werden aus den für diesen Substantivtypen wichtigen Bestandteilen des Prädikatorenschematas gebildet:<sup>140</sup> Durch diese Bindung der Slots von Frames an Verb- bzw. Prädikatorenschemata lassen sich Frames bestimmen, ohne über einen semantischen Tiefenkasus spekulieren zu müssen. Die Konstruktion von Frames kann demnach vor allem auf das Lexikon einer Sprache zurückgeführt werden.

Konearding unterscheidet in seiner Arbeit „Matrixframes“ und „Superframes“. Matrixframes stellen elementare Substantivtyp-Frames wie „Gegenstand“, „Organismus“, „Ereignis“ dar. Sie können auf Basis existierender Verbthesauri und großer Korpora mittels Gewichtung- und Paraphrasierungsprozessen systematisch konstruiert werden. Superframes dagegen sind komplexe Frames, die im Verlauf der thematischen Progression eines Textes aus elementaren Matrixframes entwickelt werden. Superframes sind damit vergleichbar mit klassischen kognitiven Schemata oder Skripten, die zusammenhängendes Wissen modellieren, z.B. den „Besuch eines Restaurants“. Die Bestimmung der elementaren Bestandteile eines Superframes stellt sich allerdings meines Erachtens als schwierig dar. Konearding weist selbst darauf hin, dass in Texten sowohl „kontrollierte Progressionen“ (thematische Progressionen bzw. Frageprogressionen ohne thematischen Wechsel oder mit bewusstem thematischem Wechsel) als auch „unkontrollierte Progressionen“ (Progressionen mit beliebigen thematischen Wechseln im Sinne freier Assoziation) auftreten. Was einen Superframe konstituiert (welche thematischen Progressionen sinnvoll zusammenhängen), ist damit aber abhängig von dem Wissen darüber, was in einem sinnvollen Zusammenhang steht (was eine sinnvolle oder kontrollierte Progression ist) – Voraussetzung für die Konstruktion von Superframes ist so das Wissen über *mögliche* Superframes, d.h. die Erstellung von Frames setzt das Wissen über die Gestalt eben dieser Frames voraus.

Das Interessante an Koneardings Arbeit ist vor allem die Umsetzung eines sowohl in theoretischer (valenzbasierte Fundierung von Frames) als auch praktischer Hinsicht (korpusbasierte Konstruktion von Frames) konsequent linguistischen Frameansatzes. Für die vorliegende Arbeit von besonderer Bedeutung ist dabei

---

<sup>140</sup> Was wichtig ist, ergibt sich zum einen aus dem Verbthesaurus und den dort dargestellten Ähnlichkeitsgruppen von Prädikatorenschemata, zum anderen aus der Analyse großer Textkorpora, durch die bestimmt werden kann, ob und wie häufig ein bestimmtes Substantiv mit einem bestimmten Prädikatorenschemata verwendet wird, sowie zum dritten aus einem manuellen Prozess der Gewichtung und Paraphrasierung.



vor allem seine Koppelung des Framekonzepts an kommunikativ-pragmatische Ansätze zum Thema als Fraglichem. Durch die Auffassung von Frames als Liste von Fragen bzw. Themen<sup>141</sup> wird einerseits der Zusammenhang von Thema und Wissen aus einer linguistischen, und nicht kognitiven Perspektive dargestellt. Andererseits werden Frames als Grundlage für den Textproduktionsprozess begriffen, insofern sie das stereotypische Wissen in Form von zu thematisierenden Fragen modellieren, das der Textproduzent typischerweise über die Vertextung eines Sachverhalts entweder explizit oder implizit behandelt bzw. die er im Textverlauf typischerweise beantwortet.<sup>142</sup>

## 2.8 Zum Kommunikationsprozess: Textproduktion und -rezeption

### 2.8.1 Textproduktion

Es existiert eine Vielzahl an Textproduktionsmodellen, die z.B. nach ihrer fachlichen Verortung (Linguistik – u.a. Textlinguistik, Psycholinguistik, Patholinguistik, Computerlinguistik – sowie Psychologie), ihrer Struktur (sequenzielle vs. rekursive vs. ebenenbasierte Modelle) oder ihrem Anwendungsbereich („Schreibprozesse beim Übersetzen und in der Fremdsprache“, „Textproduktion und Computer“, rhetorische Schreibforschung, „muttersprachliche Schreibforschung“ etc.) kategorisiert werden können (vgl. Antos 1989, 15ff.; Eigler et al. 1990, 5ff.; Molitor-Lübbert 1996, 1005).<sup>143</sup>

Antos (2000, 105) zufolge lassen sich drei zeitliche Abschnitte der Textproduktionsforschung identifizieren:

- Die didaktische Phase (bis etwas 1980), bei der die Textproduktion als Produktion von syntaktisch, semantisch und thematisch kohärenten Texten verstanden und mit dem Ziel untersucht wird, das Schreiben zu fördern und zu verbessern – diese Phase ist nach Antos stark von der Schreibforschung beeinflusst;
- die kognitive Phase (ab 1980), bei der die Textproduktion in Bezug zu kognitiven Prozessen (z.B. Informationsverarbeitung) gesetzt und Schreiben als kognitives Problemlösen aufgefasst wird, und

<sup>141</sup> Hellwig (1984a), auf dessen Theorie zum Textzusammenhang Koneerdings Ansatz basiert, definiert Themen als das Fragliche, von daher ist es gerechtfertigt von Frames als Liste von Fragen *oder* Themen zu sprechen.

<sup>142</sup> Hellwig (1984a, 70) geht allerdings davon aus, dass „Textstruktur und Wissensstruktur (...) nie isomorph“ sind, da die Textstruktur eine „pragmatische Dimension“ habe, die der Wissensstruktur fehle. Das stereotypische Wissen, wie es von Konearding modelliert wird, kann dementsprechend nur Grundlage für die Textstruktur sein, aber nie mit dieser identisch sein.

<sup>143</sup> Zu den bekanntesten dieser Modelle gehören das rekursive kognitive Modell des Textproduzierens von Hayes/Flower (1980), das Modell parallelen Verarbeitens von de Beaugrande (1984) und das entwicklungspsychologische Modell der Entwicklung von Schreibstrategien von Bereiter (1980) (vgl. Eigler et al. 1990, 5ff.; Wrobel 1995, 11ff.).

- die sozio-kognitive Phase (ab 1990), bei der Textproduzieren als „sprachliches, kommunikatives und sozio-kulturelles Handeln im Kontext von Diskursgemeinschaften“ bestimmt wird, d.h. dass hier die Rahmenbedingungen des Schreibens und die sprachlichen, stilistischen und rhetorischen Konventionen und Erwartungen in den Blick genommen werden. In den Arbeiten dieser Phase wird davon ausgegangen, dass kognitive und kommunikativ-soziale Prozesse im Akt der Textproduktion interagieren.

Obwohl es Jakobs (1997, 3) zufolge beim Textproduzieren um weit mehr als um Schreiben geht – es umfasst beispielsweise auch die visuelle Gestaltung des Textes –, wird im Folgenden nicht zwischen Ansätzen zum Schreiben und Textproduzieren differenziert, da im Hinblick auf die Handlung des Thematisierens als Mittel der Selektion und der Strukturierung von Texten beide Richtungen (die Schreib- und die Textproduktionsforschung) gleichermaßen relevant sind.

In der aktuellen Forschung („sozio-kognitive Phase“) wird das Schreiben bzw. Textproduzieren als „komplexer Prozeß, für den kognitive, sprachliche und sozial-kommunikative Faktoren gleichermaßen eine Rolle spielen“ aufgefasst (Wrobel 1995, 21). Diese Faktoren, die gleichzeitig Bedingungen für erfolgreiches Textproduzieren darstellen, sind beispielsweise der kommunikative Rahmen (z.B. Medium, Adressaten), Konversationsmaximen bzw. Kooperationsprinzipien (z.B. „Sei klar!“, „Sei relevant!“), Ziele, persönliche Faktoren sowie Wissen (z.B. Musterwissen oder Weltwissen) und Sprech- bzw. Schreibfertigkeiten des Textproduzenten (beispielsweise stilistische, fachsprachliche oder rhetorische Fähigkeiten), d.h. ein prozedurales Wissen im Sinne einer Text- und Schreibkompetenz (vgl. Antos 1989, 6f.).<sup>144</sup> Ein zentraler Bestandteil dieses prozeduralen Wissens sind nach Feilke (2010, 2f.) die literalen Prozeduren. Diese stellen stabile Schreibroutinen dar, die die Vermittlung zwischen dem Produkt- und dem Prozessaspekt der Sprache steuern. Literale Prozeduren basieren auf der Kenntnis konventionalisierter bzw. typisierter Text- und Ausdrucksmuster, d.h. sie weisen eine ausdrucksseitige Musterhaftigkeit auf. Insofern literale Prozeduren funktional bezogen sind auf rekurrente kommunikative Aufgaben spielt die Kenntnis dieser kommunikativen Aufgabe bzw. des Gebrauchszusammenhangs und der je spezifischen Gebrauchskonfiguration einer literalen Prozedur (z.B. Domäne, Textsorte) für das Erlernen dieser Prozedur allerdings eine größere Rolle als das mechanische Einüben des mit der jeweiligen Prozedur verknüpften Ausdrucksmusters allein. Durch die typisierte Verbindung von ausdrucksseitiger Musterhaftigkeit und Gebrauchsmuster hat das Vorkommen eines spezifischen Musters in einem Text – wie die in Kapitel 2.6.2.2 beschriebenen Kontextualisierungshinweise oder indexikalischen Zeichen – das Potential, ein bestimmtes Gebrauchschema (bzw. in der Begrifflichkeit der vorliegenden Arbeit: einen Typus einer kommunikativen Situation, siehe Kapitel 2.6.2.1) zu evozieren (Rezi-

---

<sup>144</sup> Winter (1992, 10ff.) fasst die kommunikativen Faktoren etwas genereller als situative Bedingungen auf und subsumiert unter diese – neben den von Antos (1989, 6f.) genannten Faktoren – auch das Thema, den Anlass und die Funktion der Textproduktion, die dem Produzenten zur Verfügung stehenden Materialien bzw. externen Informationen (z.B. Texte, Notizen, Grafiken etc.) sowie die Kontinuität oder Diskontinuität der Textproduktion.

pientsicht), während umgekehrt der Textproduzent zur Bezugnahme auf einen solchen Typus einer kommunikativen Situation (bzw. ein Gebrauchschema) auf eine spezifische, ausdrucksseitig typisierte literale Prozedur zurückgreifen kann. Literale Prozeduren haben demnach ein „Kontextualisierungspotential“ (Feilke 2010, 4) bzw. sie stellen potentielle indexikalische Zeichen dar.

In der Linguistik besteht über das für die Textproduktion und -rezeption relevante Wissen prinzipiell ein Konsens, allerdings werden teilweise unterschiedliche Benennungen von Wissensarten verwendet und Wissensarten in unterschiedlichen Detailgraden differenziert. Wie bereits in Kapitel 2.7.1 dargestellt, werden in dieser Arbeit die Wissensbestände Sprachwissen, Sachwissen, Handlungswissen und Wissen über die Kommunikationssituation differenziert.

Einigkeit besteht in der Forschung nicht nur über die Relevanz des Wissens für die Textproduktion, sondern auch darüber, dass Schreiben als Problemlösen aufzufassen ist – aktuell wird darunter vor allem das problemlösende kommunikative Handeln verstanden, weniger das problemlösende kognitive Verarbeiten (vgl. Wrobel 1995; Feilke 1996). Mit den Begriffen des Handelns und Problemlösens ist eine dynamische, prozedurale Perspektive auf das Schreiben verbunden. Schreiben bzw. Textproduzieren lässt sich demnach als rekursive Abfolge von Phasen oder „Abfolge von Problemlöseschritten“ modellieren (vgl. Feilke 1996, 1180). Zentrale Phasen – wobei diese Phasen nicht nacheinander ablaufen, sondern miteinander interagieren (vgl. Wrobel 1995, 28; Heinemann und Heinemann 2002, 182) – oder „Handlungsstadien“ (Wrobel 1995, 32) bei der Textproduktion sind das Planen (sowie das Ordnen), das Formulieren (bzw. Versprachlichen), das Inskribieren und das Revidieren (vgl. Wrobel 1995, 27f.; Gansel und Jürgens 2002, 132ff.):<sup>145</sup>

---

<sup>145</sup> In einem sehr ähnlichen Modell von Heinemann/Heinemann (2002, 180ff.) wird der Planungsphase ein „Stimulus“ (eine Kommunikationsaufgabe) und die „Kognizierung von Situation und Partner“ vorangestellt, die Phase des Formulierens und Inskribierens wird zusammengezogen zu einer Phase der „Translation“ oder „Transformation des mentalen Planungs-Konstrukts in sprachliche Signalketten“ und die Phase des Revidierens wird umbenannt in „Kontrollphase“ bzw. „Überprüfungsphase“. Diese drei Phasen (exkl. Stimulus) basieren wiederum auf dem Modell des Textproduzierens von Hayes/Flower (1980, 12ff.), die den Schreibprozess in die Subprozesse „Planning“, „Translating“, „Reviewing“ untergliedern. Chenoweth/Hayes (2001, 84) gehen dagegen von vier Prozessen aus, die sie als „proposer“, „translator“, „reviewer“ und „transcriber“ bezeichnen. Diese Prozesse werden dabei nicht als unidirektionale, sondern interaktive Prozesse aufgefasst.

In der *Planungsphase* werden aus der kommunikativen Aufgabe<sup>146</sup> bzw. der Problemstellung Ziele, Inhalte und sprachliche Handlungen des Textes entworfen bzw. grob geplant. Der Produzent entscheidet, was er mitteilen will, dazu muss er zunächst einmal sein Wissen über das Thema aktivieren und aktualisieren (vgl. Winter 1992, 32). Qualität und Quantität des aktivierten Wissens hängen zum einen vom Thema ab, zum anderen von den individuellen Wissensschemata des Produzenten, ihrer Komplexität und Dichte.

Wissen ist aber nicht nur Ausgangspunkt für die inhaltliche Gestaltung des Textes und wird im Text geordnet und linearisiert, sondern es wird auch durch den Text selbst verändert. „Textproduzieren wird so zu einem Bestandteil des Denkens.“ (Winter 1992, 59). Die Auseinandersetzung mit einem Thema im Akt der Textproduktion führt dazu, dass sich das Wissen, das der Textproduzent von diesem Thema hat, verändert, oder anders gesagt: Durch das Planen, Ordnen und Versprachlichen von Wissen beim Produzieren des Textes wird das Wissen des Textproduzenten aktualisiert, erweitert und strukturiert. Diese Funktion des Textproduzierens wird als Stufe des epistemischen Textproduzierens aufgefasst (vgl. z.B. Winter 1992, 59).

Aus der Menge des aktivierten Wissens bzw. der „thematischen Subkomplexe“ (vgl. Wrobel 1995, 27) wird nun eine Auswahl getroffen (thematische Selektion) und in eine den Zielen des Autors am Besten dienende spezifische Abfolge gebracht bzw. linearisiert (thematische Strukturierung). Diese Planungsprozesse lassen sich auf Ebene der Kognition als Konzeptualisierung (Wissensaktivierung), Selektion (aus der Menge des „Sagbaren“ wird eine Auswahl getroffen), Linearisierung (Ordnung des Wissens) und Encodierung (Übersetzung mentaler Einheiten in „vorsprachliche Signale“) beschreiben (vgl. Heinemann und Heinemann 2002, 181). Die Planungsaktivitäten insgesamt können als „Aktivitäten der Makroplanung“ bezeichnet werden, über die „globale thematische Strukturen des geplanten Textes festgelegt werden“ (Wrobel 1995, 27). Heinemann/Heinemann (2002, 181) sprechen in ähnlicher Weise davon, dass das Ergebnis dieser Phase ein „Gesamtplan“ im Sinne der „Entfaltung“ des Textthemas sei – wobei dieser Plan beim Formulieren durchaus wieder revidiert werden kann. Darüber hinaus können allerdings nicht nur Ideen oder kognitive Struktu-

---

<sup>146</sup> Die kommunikative Aufgabe, die die Selektion, Linearisierung und Versprachlichung von Informationen im Prozess der Textproduktion steuert, wird in einigen Arbeiten auch als „Quaestio“ oder Textfrage definiert, als Frage, auf die der Text eine Antwort liefert (vgl. Klein und von Stutterheim 1992, 69ff.; von Stutterheim 1994, 253ff.; Kohlmann 1997, 41ff.; von Stutterheim 1997, 15ff.) – diese Textfrage weist enge Bezüge zu dem in Kapitel 2.3.2 dargestellten Konzept des „Themas als Fraglichem“ auf (siehe Keenan und Schieffelin 1976; Hellwig 1984a, b; van Kuppevelt 1995; Büring 1997). Das Thema als Fragliches (oder die Quaestio als Textfrage) steuert dieser Auffassung zufolge den Textaufbau durch die durch eine implizite Frage initiierte Verknüpfung von aufeinander folgenden Äußerungen. Im Quaestio-Ansatz wird darüber hinaus davon ausgegangen, dass die Frage einerseits einen Referenzrahmen vorgibt – dieser basiert auf einem Sachverhalt bzw. einem Thema –, der durch den Text gefüllt werden muss (Inhaltsaspekt), andererseits aber auch Vorgaben über die Struktur des Textes macht, insofern die Quaestio bzw. Frage die Topik-Fokus-Struktur des Textes vorgibt, indem den inhaltlichen Komponenten ein unterschiedlicher Informationsstatus zugewiesen wird – die Informationen werden perspektivisch gewichtet.

ren Ausgangspunkt für die Textplanung sein, sondern auch literale Prozeduren bzw. Textroutinen, d.h. ausdrucksseitig typisierte Muster. Feilke (2010, 15ff.) weist darauf hin, dass beispielsweise Dreierschemata in der Wissenschaft als literale Prozedur aufzufassen sind, derer sich Wissenschaftler zur Erkenntnis oder zur Unterscheidung von Phänomenen bedienen. Die Dreiergliederung eines Sachverhalts ergibt sich demnach nicht aus diesem selbst, d.h. Ausgangspunkt für den Textplan ist nicht eine außersprachliche Wissensstruktur, sondern vielmehr ein bestimmtes Textmuster.

Bezogen auf die Ebenen der Textstruktur muss der Produzent bei der Planung des Textes folgende Bedingungen erfüllen (vgl. Rickheit et al. 2002, 79f.): Die Äußerung muss der Kommunikationssituation angepasst werden, indem sowohl die Rahmenbedingungen der Kommunikation als auch der intendierte Rezipientenkreis angemessen berücksichtigt werden (pragmatische Ebene). Die Äußerung muss darüber hinaus inhaltlich hinsichtlich des übergeordneten Themas und hinsichtlich der einzelnen Propositionen geplant werden (thematische Strukturierung), hierzu gehören Themenwahl (thematische Selektion), Argumentationsstrategien, Referenzherstellung und Konzeptaktivierung (semantische Ebene).

Im Hinblick auf die Ordnung bzw. Linearisierung des selektierten Wissens bzw. der thematischen Subkomplexe muss der Produzent

verschiedenartige Formen nicht nur der Linearisierung, sondern auch der inhaltlich-logischen und textmustergerichten Organisation seines Wissens einsetzen und dabei entscheiden, welches Wissen nach welchen Prinzipien linearisiert werden kann. Dies erfordert strategische Entscheidungen. (Wrobel 1995, 139f.)

Die vom Produzenten gewählte Ordnung kann u.a. auf der intrinsischen bzw. natürlichen Ordnung des Gegenstandes beruhen (z.B. chronologische Ordnung von Ereignissen), dies wäre die einfachste Form der Linearisierung (vgl. van Dijk und Kintsch 1983, 275f.; Herrmann und Hoppe-Graff 1989, 158). Produzenten nutzen in diesem Fall sachlogische Strukturen oder Frames als Grundlage für Textstrukturen – diese Strukturen erfüllen „heuristische und textorganisierende Funktionen im Schreiben“ (Feilke 1996, 1187). Andere Anordnungsmuster orientieren sich an Vertextungsmustern (z.B. Narration, Deskription, Argumentation), Konventionen des Texttyps (siehe Kapitel 2.6.3.6), Mustern zur Gestaltung von Schrift-Texten (z.B. durch das explizite metasprachliche oder visuelle Gliedern des Textes), aber auch an pragmatischen und kontextuellen Bedingungen der Kommunikationssituation. Diese Formen der Linearisierung können als Muster beschrieben werden, die zum Wissen des Produzenten gehören (Musterwissen als Teil des Handlungs- sowie des Sprachgebrauchswissens), als literale Prozeduren (s.o.) oder als Strategien, insofern der Produzent sich für die Realisierung bestimmter Muster entscheiden muss, d.h. „strategische Entscheidungen“ treffen muss. Strategien sind nach Heinemann/Viehweger

das Resultat einer Kette von – in der Regel bewusst ablaufenden – Auswahl- und Entscheidungsoperationen, durch die Lösungsschritte und Mittel markiert werden zur Durchsetzung kommunikativer Ziele. (Heinemann und Viehweger 1991, 214)

Oben genannte Muster lassen sich dementsprechend auch als „Strategien als individuelle Realisierungen einfacher Musterkomplexionen“, „Strategien auf der Basis ganzheitlicher Vertextungsmuster“ oder „Strategien zur Gestaltung umfangreicherer Schrift-Texte“ beschreiben (Heinemann und Heinemann 2002, 186ff.).

In der *Phase des Formulierens* oder der Herstellung der sprachlichen Oberflächenstruktur werden die in der Planungsphase fokussierten Inhalte, Ziele und Handlungen linearisiert, selektiert und segmentiert, d.h.

im Enkodierungsprozess realisieren Textproduzenten ihre Perspektiven auf die darzustellenden Sachverhalte in Abhängigkeit von der Funktion des Textes, seiner Adressaten, sozialen, räumlichen und zeitlichen Rahmenbedingungen und vom Medium. Enzyklopädisches Wissen muss in diesem Sinne selektiert und linearisiert und unter dem Aspekt der perspektivierten Verbalisierbarkeit organisiert werden. (Gansel und Jürgens 2002, 134).

Anders gesagt: Die thematischen Subkomplexe müssen in eine den Zielen des Produzenten am Besten dienende Ordnung gebracht werden (thematische Strukturierung). Der Prozess des Formulierens beinhaltet aus kognitiver Sicht die Versprachlichung kognitiver Inhalte – die Strukturierung von Wissen erscheint dabei auf Ebene des Textes als Thematisierung (vgl. Rothkegel 1989, 5) – und aus kommunikativ-pragmatischer Sicht als sprachliche Problemlösungshandlung. Bei letzterer sind u.a. Text- und Formulierungsmuster (als literale Prozeduren, d.h. als Routinen, die auf der Kenntnis ausdrucksseitig typisierte Muster beruhen, die mit einem bestimmten Gebrauchsschema bzw. einer kommunikativen Situation verknüpft sind bzw. auf diese verweisen, vgl. Feilke 2010, 2f.) zu berücksichtigen. Das Ziel dieser Handlung besteht darin, einen kohärenten Text zu erzeugen – hierbei ist es vordringlich, eine kohärente thematische und funktionale Mikro- und Makrostruktur des Textes zu erzeugen, oder anders gesagt: Eine kohärente Handlungsstruktur, die auf die Entfaltung eines Themas bezogen ist und über einfache und komplexe sprachliche Handlungen realisiert wird.

Der Textproduzent muss aus der Menge der möglichen zu verbalisierenden Informationen diejenigen auswählen, die relevant sind. Das Thema hat also die inhärente Funktion, eine Selektion von Informationen oder Gegenständen zu initiieren (vgl. Rothkegel 1984b, 66f.).<sup>147</sup> Thematisieren und Selektieren werden von Rothkegel deshalb auch nicht als zwei aufeinander bezogene Prozesse aufgefasst, sondern als ein ineinander greifender Prozess.

Thematische Texthandlungen beziehen sich auf die Strukturierung der propositionalen Komponente des Textes. Sie sorgen für die Verknüpfung von Wissen über Sachverhalte und Ereignisse der Umwelt mit der Textwelt. Hierbei gilt die Vorstellung daß ausgewählte kognitive Inhalte durch Einbettung in eine kohärente Textstruktur spezifiziert und linearisiert werden. Dieser Inhalt wird dadurch zum Thema eines Textes. Die komplexe Sprachhandlung, mit der dies geschieht, soll THEMATISIEREN heißen. (Rothkegel 1984b, 65, Hervorhebung im Original)

---

<sup>147</sup> Dies ist nur eine von mehreren Funktionen, die Rothkegel dem Thema zuschreibt. Weitere Funktionen sind „Konzentration“, „Erinnerbarkeit“, „Vereinfachung“, „Erzeugungsregel“ (Rothkegel 1984b, 66).

Durch die sprachliche Handlung des Thematisierens werden also kognitive Inhalte in propositionale Inhalte transformiert. Diese Propositionen werden dabei gleichzeitig linearisiert, d.h. in eine Textstruktur überführt. Das Thema (bzw. der Akt des Thematisierens) hat demnach eine strukturierende Funktion im Hinblick auf den Inhalt des Textes. Basis für die Entfaltung und Strukturierung des Inhalts (der propositionalen Struktur) bzw. des Themas (der thematischen Struktur) sind dabei kognitive Schemata (siehe auch Rothkegel 1989, 5; 1997, 461). Die thematische Struktur eines Textes ist demnach abhängig von der kognitiven Struktur des thematisierten Sachverhalts (siehe hierzu Kapitel 3.2.1).

An der Überführung eines Plans in eine Textstruktur<sup>148</sup> sind nach Rothkegel zum einen die Faktoren Kommunizierende, Kontext, Intertext beteiligt, zum anderen Prozesse wie „Konzentration, Erinnerbarkeit, Selektion und Vereinfachung“ sowie eine Komponente, die Rothkegel als „Erzeugungsregel“ bezeichnet (Rothkegel 1984b, 66). Der Faktor „Konzentration“ bezieht sich darauf, dass alle Kommunikationsteilnehmer ein „gemeinsames Thema haben, das im Fokus der Interaktion steht“ und die Aufmerksamkeit auf einen zentralen Punkt ausrichtet. Mit „Erinnerbarkeit“ ist gemeint, dass an vorher Gesagtes angeknüpft werden kann – ein Thema wird in einer bestimmten Weise entwickelt (entweder als Themafortführung oder als Themavariation etc.) und schließt dabei an das im Text oder Dialog vorangehende (intratextueller Bezug) oder aber an eines der im gesamten Diskurs<sup>149</sup> bereits behandelten Themen an (intertextueller Bezug). „Selektion“ ist, wie bereits mehrfach angesprochen, als zentraler Prozess des Thematisierens aufzufassen: Bestimmte Gegenstände, über die etwas gesagt wird, werden ausgewählt, andere Gegenstände müssen außer Acht bleiben, weil sie keinen thematischen Bezug aufweisen bzw. weil der Sprecher diesen Bezug nicht herstellen will oder kann. Welche Gegenstände oder Teilthemen zu einem Thema gehören, ist dabei zumindest teilweise durch das Sachwissen vorstrukturiert. Der Faktor „Vereinfachung“ meint, dass der Detailgrad, mit dem über ein Thema geredet wird, variabel ist – ein Thema kann oberflächlich oder intensiv behandelt werden. Die „Erzeugungsregel“ schließlich kombiniert die genannten Prozesse und beschreibt die Funktion des Thematisierens als Erzeugungsregel für den Text oder Dialog (vgl. Rothkegel 1984b, 66f.).

Bezogen auf die sprachlichen Verarbeitungsprozesse muss der Produzent in der Phase des Formulierens darüber hinaus geeignete Wörter zur Übermittlung der Inhalte finden, ihnen grammatische Funktionen zuordnen und sie adäquat anordnen (morphosyntaktische Strukturierung, vgl. Rickheit et al. 2002, 79f.).<sup>150</sup> Ebenso wie das Planen stellt auch das Formulieren einen Prozess dar, bei dem

---

<sup>148</sup> Wobei noch einmal darauf hingewiesen werden soll, dass in manchen Fällen die Textstruktur (als ausdrucksseitig typisiertes Muster) auch vor dem Plan vorhanden sein kann (s.o.).

<sup>149</sup> Mit Diskurs ist an dieser Stelle der gesellschaftliche Diskurs (als sprachliche Interaktion) über einen Themenkomplex gemeint, der in Form mehrerer sprachlich verfertigter Produkte geführt wird.

<sup>150</sup> Bestimmte Wörter oder Phrasen sind als lexikogrammatistische Muster mit einem bestimmten Schema bzw. einem Typus einer kommunikativen Situation verknüpft, so dass in diesen Fällen durch das aktuelle Schema bzw. die Situation die Verwendung dieser Wörter oder Phrasen zur Bezugnahme auf eben diese Situation konventionalisiert ist.

### der Produzent retrospektiv und prospektiv vorgehen muss:

Jede Äußerung steht im Kontext des Bisherigen, ändert den Kontext, bildet neuen Kontext für weitere Äußerungen, verändert also den Stand des Themas und der Handlungen in Richtung auf das Gesamtziel. (...) Die Formulierungen sind ständig anhand der komplexen Vorgaben und evtl. individueller Ziele reflexiv zu bewerten. (Sandig 1997, 38)

Die dritte Phase der Textproduktion ist das *Inskribieren*, d.h. der eigentliche motorische Schreibprozess: Die Äußerung muss motorisch realisiert und in ein physikalisches Produkt überführt werden. Die vierte Phase stellt das *Revidieren* dar, das sowohl die Oberflächenstruktur als auch die zugrundeliegende Sachverhaltsstruktur des Textes betreffen kann.

Das Thema spielt nach diesem Modell des Schreibprozesses also eine zentrale Rolle in der Phase der Planung des Textes, insofern es die Aktivierung bzw. Aktualisierung, Selektion und Organisation von Wissen steuert. Darüber hinaus wirkt es über den in der Planungsphase entwickelten „Gesamtplan“ zur Entfaltung bzw. Entwicklung des Themas auf die Versprachlichung des Wissens ein. Diese beiden Aspekte finden sich auch in de Beaugrande/Dresslers Modell der Textproduktion (1981, 42ff.), in dem fünf Phasen der Textproduktion beschrieben werden: die Phase der Planung, der Ideation, der Entwicklung, des Ausdrucks und der grammatischen Synthese (wobei diese Phasen als lineare Phasen aufgefasst werden, während nach dem aktuellen Forschungsstand von einer Rekursion der Phasen auszugehen ist). Interessant an diesem Modell ist allerdings die Art der expliziten Bezugnahme auf das Thema bzw. auf die thematische Organisation in zwei dieser fünf Phasen: In der Phase der Ideation erwächst nach diesem Modell aus einer Idee ein Thema, das als gedanklicher Rahmen für das mögliche zu aktivierende Wissen dient (vgl. Eigler et al. 1990, 12). Dieser vom Textthema aufgespannte Rahmen beschränkt demnach das, was gesagt werden kann (bzw. zunächst einmal das, was gedacht werden soll) – es stellt den Rahmen für die Auswahl der möglichen Subthemen dar. Diese Subthemen werden schließlich in der Entwicklungsphase ausgearbeitet und in Relation zueinander gesetzt (Themenentwicklung). Diese Rolle des Themas als Mittel zur Selektion und Organisation des Textes wird in Kapitel 3.4 noch ausführlicher beleuchtet.

### 2.8.2 Textrezeption

Die Textrezeption wurde in den letzten Jahrzehnten vor allem in kognitiv oder psycholinguistisch orientierten Forschungsarbeiten untersucht. Im Vordergrund steht vor allem die Frage, wie Rezipienten Texte verstehen und eine Textbedeutung aufbauen. „Einen Text verstehen“ heißt, dass ein Rezipient in der Lage ist, eine kohärente, in sich stimmige Repräsentation des im Text dargestellten Sachverhalts zu bilden (vgl. Lerch 1991, 32). Nach Auffassung von Rickheit und Strohner lassen sich folgende Ebenen des Textverstehens unterscheiden (vgl. Rickheit und Strohner 1993, 70; Rickheit et al. 2002, 93ff.):<sup>151</sup>

---

<sup>151</sup> Rickheit/Strohner (1993) betonen dabei, dass die Ebenen miteinander interagieren.



1. *Perzeptuelle Ebene*: Der Text wird hinsichtlich seiner sensorischen Merkmale erkannt und verarbeitet, hierzu gehören z.B. das Erkennen der Sprache, der Wörter und der Satzarten.
2. *Syntaktische Ebene*: Wortarten und morphologische Merkmale von Wörtern werden erkannt und Mehrdeutigkeiten auf der syntaktischen Ebene aufgelöst.
3. *Semantische Ebene*: Diese Ebene umfasst a) die *konzeptuelle Verarbeitung*, bei der für jedes Wort das damit verbundene Konzept aktiviert wird; gibt es mehrere mögliche Konzepte, findet eine Disambiguierung statt (auf Basis des konzeptuellen Kontexts, des Schemawissens oder der Kommunikationssituation); b) die *referenzielle Verarbeitung*: Zwischen Wortkonzept und Objekten der realen oder fiktiven Textwelt wird eine Beziehung hergestellt; durch experimentelle Studien konnte gezeigt werden, dass nur solche Objekte für Referenzen zur Verfügung stehen, die dem aktuellen kognitiven Schema angehören (vgl. Rickheit et al. 2002, 101); der sprachliche Kontext ist demnach von großer Bedeutung für die Verarbeitung von Referenzen, ebenso wie die Kommunikationssituation; c) *Sinnverarbeitung*: Die semantischen bzw. konzeptuellen Relationen zwischen den Konzepten bzw. Referenzobjekten werden erkannt, Prädikat-Argument-Strukturen bzw. Propositionen miteinander verknüpft, d.h. Beziehungen zwischen Propositionen hergestellt.
4. *Pragmatische Ebene*: Während auf der semantischen Ebene vor allem die propositionale, referenzielle und konzeptuelle Verarbeitung im Fokus steht, geht es auf dieser Ebene um eine Verarbeitung der Illokutionen und ihrer Relationen (nach der Begrifflichkeit der vorliegenden Arbeit: der sprachlichen Handlungen). Dabei wird auch die kommunikative Situation als Einflussfaktor wahrgenommen; hier spielen situative und kulturelle Einflüsse ebenso eine Rolle wie individuelle Faktoren (z.B. das Vorwissen des Rezipienten, Interessen sowie emotionale Befindlichkeiten) eine wichtige Rolle bei der Textverarbeitung; wie Rickheit et al. (2002, 106) feststellen, sind diese Faktoren bisher jedoch weder empirisch noch konzeptuell in befriedigender Weise erforscht worden, auch aufgrund methodischer Probleme.

In textlinguistischen Arbeiten wird die Modellierung des Rezeptionsprozesses häufig ersetzt durch die Beschreibung der für das Textverstehen nötigen Wissensbestände sowie durch Untersuchungen zur Verständlichkeit von Texten – es werden weniger Aussagen getroffen über die mit der Textrezeption verbundenen kognitive Prozesse, die experimentell nur schwer erforschbar und belegbar sind, sondern das Rezipieren und Verstehen des Textes wird unter dem Blickwinkel der Frage nach der Konstruktion von Kohärenz und Verständlichkeit gesehen.<sup>152</sup> Wie wird Kohärenz konstruiert? Was ist kohärent, was ist nicht kohärent? Was ist verständlich, was nicht? Und welche Arten von Wissen spielen bei der Bewer-

---

<sup>152</sup> Kohärenz wird dabei als „Bewertungsprädikat [betrachtet], durch das im Ergebnis der [sowohl text- als auch wissensgeleiteten; M.B.] Textinterpretation einer Äußerungsfolge die Eigenschaft ‚kohärent‘ bzw. ‚nicht kohärent‘ zugeschrieben wird“ (Heinemann und Viehweger 1991, 120).

tung, ob etwas kohärent oder verständlich ist, eine Rolle? Die Wissensarten, die beim Textproduzieren und -rezipieren eine Rolle spielen, wurden bereits in Kapitel 2.7.1 genannt (Sprachwissen, Sachwissen, Handlungswissen und Wissen über die Kommunikationssituation). Sie lassen sich allerdings nur bedingt einzelnen Ebenen der Textverarbeitung zuordnen: Das Sprachwissen ist Voraussetzung sowohl für die syntaktische als auch die semantische Verarbeitung; Sachwissen spielt am ehesten auf der semantischen Ebene eine Rolle, muss aber auch zur pragmatischen Verarbeitung hinzugezogen werden; Handlungswissen und Wissen über die Kommunikationssituation sind Voraussetzung für die pragmatische Verarbeitung von Texten.

Wie die einzelnen Prozesse beim Textverstehen genau ablaufen, d.h. in welcher Art und Weise die sprachlichen Ebenen verarbeitet werden, welche Arten von Interaktionen zwischen den Ebenen sowie den vier Arten von Wissensbeständen genau ablaufen, darüber gibt es unterschiedliche Vorstellungen, die sich in verschiedenen Textverstehensmodellen niederschlagen. Im Folgenden sollen einige dieser Modelle exemplarisch dargestellt werden.

Modelle zum Textverstehen lassen sich in vier Gruppen ordnen: Propositionale, prozedurale, schematheoretische sowie mentale Textverstehensmodelle (vgl. Christmann 2000, 115). In den seit Mitte der 70er Jahre entwickelten *propositionalen Modellen* wird davon ausgegangen, dass Propositionen die zentralen Basiseinheiten des Textverstehens darstellen, die im Verstehensprozess miteinander verknüpft und hierarchisch organisiert werden. Empirische Untersuchungen haben gezeigt, dass hierarchiehöhere Propositionen (d.h. z.B. Makropropositionen) von Rezipienten besser behalten werden als niedrigere, und stark vernetzte besser als wenig vernetzte (vgl. Christmann 2000, 116). Das deutet darauf hin, dass es sich bei den hierarchiehöheren und stark vernetzten Propositionen um die bedeutsamen und zentralen Propositionen handelt. Diese Erkenntnis findet sich in analoger Form in einigen semantischen Forschungsansätzen zum Textthema, in denen davon ausgegangen wird, dass das Textthema die oberste Makroproposition (siehe Kapitel 2.2.2) des Textes ist oder der zentrale Knoten in einem konzeptuellen Netz. Die propositionalen Modelle sind eng auf den Text bezogen. Problematisch ist bei diesen Modellen zum einen die wenig objektive Ermittlung der hierarchiehöheren Makropropositionen, zum anderen die mangelnde Berücksichtigung des Wissens. Diese Bezugnahme auf Wissen (oder Schemata bzw. Typen kommunikativer Situationen) ist jedoch Voraussetzung zur Erklärung vieler empirisch zu beobachtender Phänomene beim Textverstehen, wie z.B. dem, dass Rezipienten sich an etwas erinnern, das gar nicht explizit im Text stand, aber unter Rekurs auf Wissen aus diesem erschließbar ist.

Auf einem propositionalen Verstehensmodell baut das *Modell der zyklischen Textverarbeitung* von Kintsch/van Dijk (1978) auf. Dieses prozedurale Modell geht davon aus, dass die Propositionen eines Textes in aufeinander folgenden Zyklen verarbeitet werden. In jedem Zyklus befindet sich nur eine bestimmte Menge von Propositionen im Arbeitsgedächtnis und wird unter Rekurs auf den vorherigen Zyklus (und die dort eingelesenen Propositionen) zur Kohärenzbildung genutzt; diese Propositionen stehen nur solange im Arbeitsgedächtnis zur Verfügung, wie sie zur Bildung einer komplexeren kohärenten Bedeutung ge-

braucht werden. Bei dieser Verarbeitung mit dem Zweck der Bildung kohärenter Propositions- bzw. Bedeutungskomplexe können Probleme auftreten, insofern es in einem Zyklus unter Umständen nicht gelingt, Kohärenz herzustellen. In diesem Fall gehen Kintsch/van Dijk davon aus, dass diese Kohärenzlücken durch Inferenzen oder „Reinstatement“ geschlossen werden können. Inferenzen (Erschließungsprozesse unter Nutzung von Wissen) müssen demnach nur dann gezogen werden, wenn in dem vorherigen Zyklus keine Proposition zur Kohärenzherstellung genutzt werden kann. „Reinstatement“ stellt die zweite Möglichkeit der Schließung einer Kohärenzlücke dar. Hier wird eine bereits verarbeitete Proposition reaktiviert, um einen Anknüpfungspunkt für den aktuellen Zyklus zu bilden (vgl. Rickheit und Strohner 1993, 73f.; Christmann 2000, 117). In diesem Ansatz wird davon ausgegangen, dass das Textverstehen im Normalfall ohne den Rekurs auf Wissen und ohne das Ziehen von Inferenzen abläuft und dass diese Strategien nur im ‚Notfall‘ angewendet werden. Dieses Modell ist deshalb sehr textnah und präferiert eine bottom-up-Strategie der Textverarbeitung; allerdings steht es dadurch im Gegensatz zu kommunikativ-pragmatisch oder kognitiv orientierten linguistischen Arbeiten, in denen die Berücksichtigung der Einbettung eines Textes in eine spezifische Kommunikationssituation oder die Berücksichtigung des Vorwissens eines Rezipienten als Voraussetzung für die adäquate Ermittlung der Kohärenz eines Textes betrachtet wird.

Ab Mitte der 70er Jahre wurden neben den propositionalen Modellen vermehrt *schematheoretische Modelle* zum Textverstehen entwickelt, in denen im Gegensatz zu o.g. Modell der Fokus auf die Rolle der top-down- (und nicht bottom-up) Strategien gelegt und Wissen als Einflussfaktor für das Textverstehen begriffen wird. In empirischen Untersuchungen konnte gezeigt werden, dass schemarelevante Texteinheiten besser behalten werden als schemairrelevante (vgl. Christmann 2000, 118). Ein Verstoß gegen die Maxime der Relevanz (im Sinne von inhaltlicher/thematischer Relevanz) scheint sich demnach empirisch als schlechtere Behaltensleistung belegen zu lassen (zum Zusammenhang von Thema und Relevanz siehe Kapitel 3.4.1).

Als Reaktion sowohl auf die propositionalen als auch schematheoretischen Modelle wurden seit den 80er Jahren solche Textverstehensmodelle in den Fokus genommen, in denen den Interaktionen zwischen Text und Rezipient mehr Gewicht zugemessen wird und auch nichtpropositionale Aspekte der Kommunikation erfasst werden. Diese Modelle basieren auf der Annahme, dass Rezipienten beim Textverstehen ein mentales Modell des im Text repräsentierten Sachverhalts aufbauen, das jedoch über den Text selbst hinausgehen kann (vgl. Rickheit und Strohner 1993, 83). Ein prägender Ansatz in diesem Kontext ist der von Johnson-Laird (1988) zur *Theorie mentaler Modelle*. Die Theorie mentaler Modelle geht davon aus, dass bei der Textverarbeitung Text und auf Schemata basierendes Wissen interagieren und gemeinsam in die Repräsentation des kommunizierten Sachverhalts einfließen. Der Text bzw. die Wörter und Sätze des Textes werden als Auslöser für den Aufbau eines mentalen Modells aufgefasst, wobei dieser Aufbau durch semantisches und syntaktisches Wissen, Wissen über den Kontext, Domänen- bzw. Weltwissen und kommunikatives Wissen (z.B. um Diskurskonventionen) gesteuert wird. Der Konstruktionsprozess selbst erfolgt

schrittweise, und das bis zum aktuellen Satz konstruierte mentale Modell wird mit jedem neu rezipierten Satz gegebenenfalls revidiert (vgl. Johnson-Laird 1988, 114f.). Ein Text wird dann als kohärent aufgefasst, wenn es einem Rezipienten möglich ist, ein mentales Modell für den gesamten Text zu konstruieren. Im Gegensatz zu den rein schematheoretischen Modellen bezieht Johnson-Laird auch die propositionale Ebene in die Repräsentation des Textverständnisses mit ein, insofern die im Text kommunizierten Sachverhalte sprachnah repräsentiert werden und als Ausgangspunkt für die Konstruktion des mentalen Modells dienen.

According to this theory [die Theorie mentaler Modelle; M.B.], the initial representation of an utterance, which is close to its linguistic form, is used to construct a model of the state of affairs that is described, requested, or questioned. The process is guided by a knowledge of the contribution to truth conditions made by words in the utterance, by a knowledge of how to combine meanings according to syntax (probably on a rule-by-rule basis), by a knowledge of the context, which in part is represented in the existing model of the situation, and by general knowledge of the domain and the conventions of discourse. (Johnson-Laird 1988, 110)

Das mentale Modell selbst stellt eine „analoge“ Repräsentation dar, in der der im Text kommunizierte Sachverhalt direkt abgebildet wird (vgl. Schnotz 1994, 158), und zwar unterstützt durch Wahrnehmungs-, Imaginations- und Erinnerungsprozesse (vgl. Johnson-Laird 1988, 115).

Ein alternatives Textverstehensmodell, das nach Ansicht von Christmann (2000, 119) die „elaborierteste Modellierung des Textverarbeitungsprozesses im Rahmen mentaler Modelle“ darstellt, ist das *Strategie- oder Situationsmodell* von van Dijk/Kintsch (1983). Wie Johnson-Laird gehen auch van Dijk/Kintsch davon aus, dass Rezipienten beim Textverstehensprozess eine mentale Repräsentation der Sachverhalte konstruieren, über die im Text geredet wird, und zwar indem sie die im Text angesprochenen Sachverhalte als Instanzen einer bestimmten Klasse interpretieren (z.B. als Unfall), die in einem kognitiven Situationsmodell modelliert ist. Bei der Interpretation des Textes und dem dazugehörigen Ziehen von Inferenzen werden demnach nicht mehr Propositionen inferiert, sondern Teile des Situationsmodells. Das mentale Modell des Textes repräsentiert nicht mehr die Textbedeutung allein, sondern vor allem die Sachverhalte, über die es im Text geht. Van Dijk/Kintsch nehmen an, dass ein Text dann kohärent ist, wenn Rezipienten ein mentales Modell der kommunizierten Sachverhalte konstruieren können und wenn sie die einzelnen Ereignisse und Fakten, die diesen Sachverhalt ausmachen, miteinander verknüpfen können, z.B. durch zeitliche oder kausale Relationen (vgl. auch van Dijk 2008, 58).

Sicherlich auch aufgrund der Probleme, die beim Versuch, aussagekräftige empirische Einblicke in die kognitiven Prozesse des Textverstehens zu gewinnen, auftreten, ist keines der Textverstehensmodelle allgemein akzeptiert. Allerdings können mehrere Annahmen als relativ gesichert gelten – wobei sich Unterschiede zwischen verschiedenen Forschungsrichtungen im Hinblick auf Details ergeben –, von denen auch in dieser Arbeit ausgegangen wird: Textrezeption ist nicht einfach eine Umkehrung der Textproduktion. Das bedeutet zum einen, dass die Textrezeption nicht passiv, sondern konstruktiv ist – Rezipienten nehmen nicht einfach den Text in der vom Produzenten intendierten Bedeutung auf, dekodier-

ren nicht einfach ein komplexes sprachliches Zeichen (den Text), sondern konstruieren im Prozess des Textverstehens ein mentales Korrelat des Textes, das mit dem des Produzenten übereinstimmen kann, aber nicht muss; Rezipienten können demnach unterschiedliche Auffassungen von der Bedeutung eines Textes gewinnen, diesen unterschiedlich interpretieren.

Zum Zweiten muss festgehalten werden, dass die Textrezeption sowohl text- als auch wissensgeleitet verläuft: top-down- und bottom-up-Prozesse, Wissen und Text interagieren miteinander (vgl. Schwarz 2001, 19). Bei der top-down-Verarbeitung interpretiert der Rezipient den Text auf Basis seiner Wissensbestände (sowohl Sachwissen, Sprachwissen, Handlungswissen als auch Wissen über die Kommunikationssituation). Er bildet Hypothesen darüber aus, wie der Text auf Basis dieses Wissens zu verstehen ist (z.B. vor dem Hintergrund einer spezifischen Kommunikationssituation). Bei der bottom-up-Verarbeitung werden Texte auf Basis ihrer sprachlichen Struktur interpretiert, Rezipienten müssen die Bedeutung und Funktion von Morphemen, Wörtern, Sätzen und komplexeren Textstrukturen verstehen bzw. pars-pro-toto auf einen bestimmten Frame bzw. einen Typus einer kommunikativen Situation beziehen.

Neben der skizzierten Interaktion von Text und Wissen ist eine dritte Grundannahme der Textverstehensforschung, dass neben dem Text selbst sowie den Wissensbeständen des Rezipienten auch die Erwartungen des Rezipienten den Textverstehensprozess beeinflussen. Diese Erwartungen basieren zunächst auf der Bewertung der Kommunikationssituation und richten sich auf „erwartbare Veränderungen der Situation und den zu erwartenden Text/die Textsorte“ (Heinemann und Heinemann 2002, 168); sie grenzen den Textverstehensprozess ein und fokussieren ihn in Hinblick auf die erwartete Situation, den erwarteten Text/die Textinterpretation (vgl. Heinemann und Viehweger 1991, 114; Heinemann und Heinemann 2002, 167f.). Neben den auf der Situation selbst basierenden Erwartungen weckt aber auch der Text selbst Erwartungen – von Dijk/Kintsch zufolge dienen insbesondere die ersten verstandenen Makropropositionen des Textes als top-down-Orientierung für nachfolgende Sätze – sie sind Ausgangspunkt für die Ausbildung von Erwartungen, die sich auf den weiteren Textverarbeitungsprozess auswirken (vgl. van Dijk und Kintsch 1983, 196).

In dieser Arbeit wird darüber hinaus davon ausgegangen, dass beim Textverstehen Wissen über *kommunikative* Situationen<sup>153</sup> bzw. Wissen über Typen kommunikativer Situationen inferiert wird und in den Aufbau eines mentalen Korrelats des Textes mit einfließt. Dabei spielt auch das Wissen über den Gebrauch von Wörtern und Wortfeldern, aber auch Wissen über den Gebrauch von grammatischen Strukturen (z.B. Tempusgebrauch, syntaktische Komplexität) in bestimmten kommunikativen Situationen eine Rolle, insofern Rezipienten erst über das Textmaterial bzw. die (potentiellen) indexikalischen Zeichen einen Bezug zu einer bestimmten Situation herstellen. Wissen über Typen von Situationen sowie Wissen über den spezifischen Gebrauch der Sprache in einer bestimmten Situation kann dabei repräsentiert werden in Form von Frames (als

---

<sup>153</sup> Zum Begriff der „kommunikativen Situation“ siehe Kapitel 2.6.2.1.

stereotypische Erinnerungsmuster zu einer bestimmten kommunikativen Situation).

## 2.9 Zusammenfassung

In diesem Kapitel wurden einige theoretische Festlegungen getroffen und Klärungen vorgenommen. Dabei wurde zunächst diskutiert, was in der vorliegenden Arbeit unter dem Begriff „Text“ verstanden wird, bevor im Folgenden die Prozesse der Textproduktion und -rezeption sowie die kognitiven Grundlagen dieser Prozesse thematisiert wurden (für eine Übersicht über die Einbettung des Texts als Produkt in die Prozesse der Textproduktion und -rezeption siehe Abbildung 6, insbesondere Abschnitt C). In Kapitel 2.6 wurde dargestellt, dass mit Texten einerseits komplexe sprachliche Handlungen vollzogen werden, indem bestimmte sprachliche Teilhandlungen vollzogen werden. Während sich die Texthandlung insgesamt auf ein Textthema bezieht, beziehen sich die Teilhandlungen jeweils auf spezifische Aspekte des Themas. Sprachliche Handlungen werden realisiert durch Äußerungseinheiten, deren Elemente wiederum potentielle indexikalische Zeichen sein können. Der Text hat deshalb auch die Funktion eines komplexen indexikalischen Zeichens. Mit indexikalischen Zeichen nehmen Kommunikanten Bezug auf kollektive kommunikative Erfahrungen bzw. stereotypische kommunikative Situationen. In der vorliegenden Arbeit wird dabei angenommen, dass einfache indexikalische Zeichen miteinander interagieren und dass durch die Interpretation des Zusammenspiels dieser einfachen indexikalischen Zeichen Kommunikanten komplexe indexikalische Zeichen konstruieren (können), die für einen bedeutungskonstituierenden Schluss genutzt werden können. Insofern die Prämisse lautet, dass Texte, die auf Verständigung abzielen, kohärent sind, d.h. den Aufbau eines kohärenten mentalen Modells bzw. Situationsmodells unterstützen, ist auch davon auszugehen, dass Texte als holistische komplexe indexikalische Zeichen aufgefasst werden können, die die Bezugnahme auf und/oder Konstruktion eines kohärenten kommunikativen Situationsmodells ermöglichen. Neben dieser holistischen Sicht auf Texte gilt allerdings gleichzeitig, dass Texte auch aus elementaren Einheiten aufgebaut und auf mehreren Ebenen strukturiert sind – der Ebene der Informationsstruktur, der lexikalischen, referenziellen, relationalen Struktur, der generischen Texttypstruktur sowie der visuellen und logischen Textstruktur. Diese sprachlichen Ebenen interagieren dabei sowohl untereinander als auch mit den Ebenen der Kommunikation und Kognition (Text in Gebrauch), und zwar insbesondere über das Thema des Textes, wie in Kapitel 3 noch dargelegt werden soll.

Im weiteren Verlauf des Kapitels (2.7) wurden die kognitiven Grundlagen der Textproduktion und -rezeption eingeführt. Dabei wurden zunächst die Wissensbestände identifiziert, die für die Produktion und Rezeption von Texten relevant sind (siehe Abbildung 6, Abschnitt A). Als zentrale Wissenskomponenten wurden das Sachwissen, das Sprachwissen, das Interaktions- bzw. Handlungswissen sowie das Wissen über die Kommunikationssituation identifiziert. Daran anschließend wurde dargestellt, in welcher Weise dieses Wissen repräsentiert sein kann. Hier wurden vor allem kognitive Schemata und Frames näher betrachtet –

diese können als stereotypische Erinnerungsmuster aufgefasst werden, oder noch spezifischer: als Wissen über stereotypische kommunikative Situationen. Dieses Wissen ist dabei in Form von Konzepten (bzw. prototypischen Einheiten) und Relationen zwischen diesen Konzepten/Einheiten organisiert. Mit der Aktivierung eines bestimmten kognitiven Schemas bzw. Frames, z.B. durch einen sprachlichen Ausdruck in einem Text, der vom Kommunikanten als indexikalische Zeichen für diesen Frame interpretiert wird, wird gleichzeitig eine Erwartungshaltung aktiviert: Der Rezipient erwartet – aufgrund der Vernetzung der Konzepte in dem aktuell aktivierten kognitiven Frame – das Vorkommen typischer Objekte und Ereignisse sowie auf der Ebene der Sprache: der für diese Sprachverwendungssituation typischen sprachlichen Mittel. Nicht nur das lexikalische Material hat dabei das indexikalische Potential, einen Frame bzw. eine bestimmte Situation zu evozieren, sondern auch grammatische Strukturen (z.B. Zeitformen).

Insofern kognitive Schemata bzw. Frames sowie mentale Modelle auch als Repräsentationen stereotypischer Situationen aufgefasst (vgl. Minsky 1980; van Dijk 1980; van Dijk und Kintsch 1983; Konecny 1993; Fillmore 2006 [1982], siehe auch Kapitel 3.2.3 und 3.2.4) und sprachliche Ausdrücke und Strukturen als indexikalische Zeichen verstanden werden können, die zur Bezugnahme auf diese Schemata gebraucht und als solche Bezugnehmenden Zeichen interpretiert werden können (vgl. Auer 1986, 25; Ortner und Sitta 2003, 22; Fillmore 2006 [1982], 385; Steinhoff 2007, 99; Antos 2009, siehe auch Kapitel 3.1.2 und 3.2.3), stellt der Begriff der „kommunikativen Situation“ ein Bindeglied zwischen den Ebenen der Kommunikation, Kognition sowie dem Text dar. Aus dieser Perspektive nehmen Texte als komplexe indexikalische Zeichen über kognitive Schemata bzw. Frames Bezug auf einen Typus einer kommunikativen Situation mit bestimmten Eigenschaften. Diese Eigenschaften ihrerseits lassen sich verschiedenen Typen von Wissen zuordnen (Sach-, Sprach-, Handlungswissen sowie Wissen über die Kommunikationssituation), das, insofern es sowohl konventionalisiert als auch dynamisch ist, zugleich einen Ausgangs- als auch Endpunkt des Kommunikationsprozesses darstellt: Es stellt eine Grundlage für das top-down-orientierte Produzieren und Verstehen von Texten dar, gleichzeitig kann es durch die Produktions- und Rezeptionsprozesse selbst aktualisiert, erweitert und neu strukturiert werden.

Im Anschluss an die Darstellung der kognitiven Grundlagen wurde der Kommunikationsprozess insgesamt in den Blick genommen (Kapitel 2.8), d.h. die Textproduktion und -rezeption. Im Hinblick auf die Textproduktion wurde in diesem Kapitel argumentiert, dass die aus den miteinander interagierenden Phasen des Planens, Formulierens, Inskribierens und Revidierens bestehende Textproduktionshandlung in inhaltlicher Hinsicht in weiten Teilen (insbesondere den Phasen des Planens und Formulierens) durch das Thema gesteuert wird (siehe Abbildung 6, Abschnitt B): Das Thema, das der Produzent auswählt, über das er etwas schreiben möchte, bestimmt einerseits, welches Sachwissen aktiviert wird und als gedanklicher Rahmen für die inhaltliche Gestaltung des Textes dient. Das Thema hat in diesem Zusammenhang demnach die Funktion der Selektion, es beschränkt das, was gesagt werden soll. Darüber hinaus wirkt das Thema aber

auch auf die Strukturierung des Textes ein, insofern eine kohärente thematische Mikro- und Makrostruktur des Textes entworfen werden muss, oder korrekter: eine kohärente Handlungsstruktur, die auf die Entfaltung eines Themas und seiner Aspekte bezogen ist. Diese Aspekte müssen dabei durch die sprachlichen Teilhandlungen in einen Zusammenhang zueinander bzw. zum Textthema gebracht werden.

Zum Prozess der Textrezeption bzw. des Textverstehens gibt es sehr unterschiedliche Theorien, von denen einige der relevantesten und für bestimmte historische Forschungsepochen repräsentativsten kurz in Kapitel 2.8.2 vorgestellt wurden. In der vorliegenden Arbeit wird die Ansicht vertreten, dass Grundlage für die Textrezeption bzw. das Textverstehen einerseits das Textmaterial ist, andererseits die Wissensbestände des Rezipienten (und Produzenten). Auf Basis dieser beiden Bestandteile bildet der Textrezipient in einem zyklischen Verarbeitungsprozess Erwartungen über das Thema und die Funktion der sprachlichen Teilhandlungen des Textes aus, die wiederum in die zyklische Konstruktion eines mentalen Korrelats des Textes münden (siehe Abbildung 6, Abschnitt D). Ausgangspunkt für die Bildung der initialen Erwartungen (iE) sind einerseits situative Präsignale, andererseits textuelle Präsignale (eine zentrale Rolle spielen hier insbesondere der Titel und die ersten rezipierten Äußerungseinheiten bzw. indexikalischen Zeichen). Diese initialen Erwartungen werden im weiteren Verlauf des Rezeptionsprozesses in der Interaktion von Text und Wissen (bottom-up- und top-down-Prozesse) schrittweise aktualisiert (aE) und gegebenenfalls revidiert (rE). Das Thema (bzw. die Erwartungen an das Thema und die thematische Organisation) bildet demnach den Verstehenshorizont für den Text (es leitet das Verstehen des Textes), im Hinblick auf den die sprachlichen Teilhandlungen interpretiert und miteinander verknüpft werden.

Die Auseinandersetzung mit einem Thema sowohl bei der Textproduktion als auch -rezeption führt dazu, dass sich das Wissen, das Produzent und Rezipient von diesem Thema haben, verändern kann: Es kann erweitert, aktualisiert und neu strukturiert werden. Wissen ist demnach nicht nur Grundlage der Textproduktion und -rezeption, sondern kann durch diese Prozesse selbst verändert werden – im Hinblick auf das Sachwissen kann dies z.B. bedeuten, dass der Rezipient neues Wissen über die Welt erwirbt (z.B. neues deklaratives Wissen); sprachliches Wissen kann erweitert werden, wenn z.B. für den Rezipienten neue Lexeme oder Bedeutungsvarianten im Text vorkommen; durch neue Textsorten oder Textmuster kann das Handlungswissen erweitert oder aktualisiert werden; ebenso kann auch das Wissen über die Kommunikationssituation durch den Text selbst verändert werden.



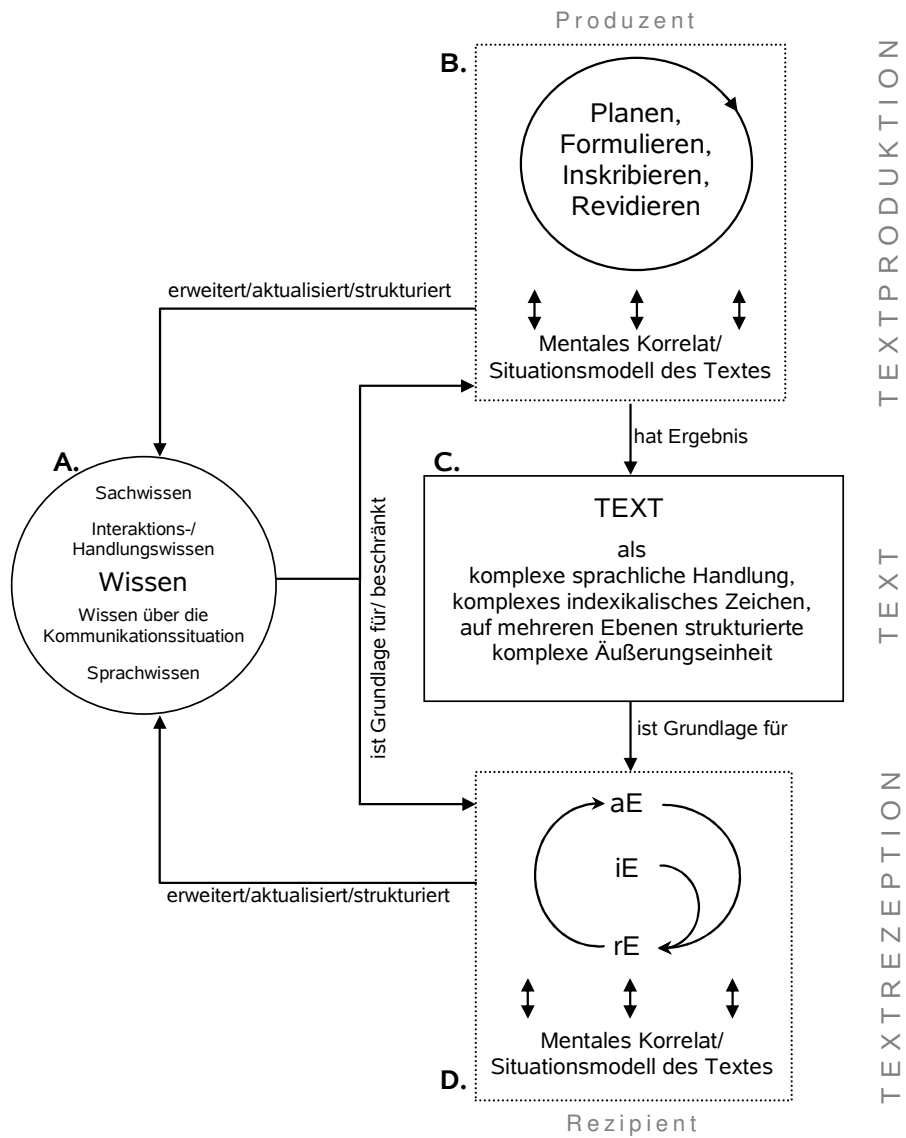


Abbildung 6: Zum Zusammenspiel von Textproduktion/-rezeption, Wissen und Text

### 3 Ein Ebenen-basiertes Thema-Modell

In diesem Kapitel soll ein integratives Thema-Konzept entwickelt werden, das die Diversität der existierenden Thema-Konzepte zu erklären versucht, indem es diese anhand des in Kapitel 3 entwickelten Modells der Textstruktur den Beschreibungsebenen des Textes zuordnet. Auf diese Weise sollen die in den diversen Forschungsansätzen zum Thema beschriebenen unterschiedlichen Arten sprachlicher Manifestationen und begrifflicher bzw. theoretischer Modellierungen des Themas im Text strukturell differenziert werden. Diese Differenzierung dient als Ausgangspunkt für ein integratives, Ebenen-basiertes Thema-Modell, bei dem das Thema als Konzept aufgefasst wird, das auf drei zentralen Ebenen wirkt: der kommunikativen, der kognitiven und der textinternen Ebene. Nachdem in Kapitel 3.1.1 zunächst ein Überblick über das Ebenen-basierte Thema-Modell gegeben wird und in Kapitel 3.1.2 die in Kapitel 2 dargestellten Thema-Konzepte in dieses Ebenen-basierte Thema-Modell eingeordnet werden, soll in den danach folgenden drei Unterkapiteln ausführlich der Zusammenhang von Thema, Text und Wissen erörtert werden. In Kapitel 3.2 steht dabei der Zusammenhang von Thema und Wissen (kognitive Ebene des Themas) im Zentrum; Kapitel 3.3 beschäftigt sich mit den Reflexionen des Themas im Text, d.h. den Manifestationen des Themas auf der Textoberfläche (materiale, textinterne Ebene des Themas); in Kapitel 3.4 geht es um die Rolle des Themas für die Textproduktion und -rezeption (kommunikative Ebene des Themas).

Darüber hinaus soll in diesem Kapitel ein auf dem theoretischen Thema-Modell basierendes, operationalisierbares und automatisierbares Analyseverfahren für das Thema beschrieben und exemplarisch angewendet werden. Dieses Verfahren basiert einerseits auf der Analyse und formalen Repräsentation der Reflexionen des Themas im Text, andererseits auf der formalen Modellierung von Wissen (zu den verfügbaren maschinell verarbeitbaren Wissensressourcen siehe Kapitel 3.2.5). Die Reflexionen des Themas werden dabei als XML-basierte Multi-Ebenen-Annotationen realisiert, so dass Zusammenhänge zwischen den Reflexionen verschiedener Ebenen statistisch ermittelbar werden (zur „Verrechnung“<sup>154</sup> der Reflexionen des Themas siehe Kapitel 3.3.3). Die zu jeder Beschreibungsebene des Textes vorgestellten Analysemethoden werden jeweils ergänzt durch Darstellungen der aktuellen computerlinguistischen Verfahren, mit deren Hilfe die Analysen der Reflexionen des Themas automatisiert werden können. Kapitel 4 endet schließlich mit der Darstellung der Architektur eines Systems zur maschinellen Themenidentifizierung, das auf den in der vorliegenden Arbeit entwickelten theoretischen Annahmen und Modellen basiert (siehe Kapitel 3.4.3).

---

<sup>154</sup> Der Ausdruck „verrechnen“ bzw. „Verrechnung“ meint ein „Sinnproduzieren, sprechend, schreibend und/oder verstehend, das sowohl holistisch wie elementaristisch ist; das gleichzeitig das Ganze und die Teile aufeinander bezieht.“ (Ortner und Sitta 2003, 23).

## 3.1 Einleitung

### 3.1.1 Einführung in das Ebenen-basierte Thema-Modell

Das Thema ist zum einen – wie auch die Intention<sup>155</sup>, mit der ein Text verfasst wird, bzw. die Funktion, die ein Text erfüllen soll – einer der zentralen Faktoren, die den Textaufbau (Produktionsperspektive) und das Verstehen des Textes (Rezeptionsperspektive) steuern. Texte sollten – um verständlich zu sein – eine thematische Einheit bilden, d.h. auf ein Thema hin orientiert sein, indem in den einzelnen Äußerungseinheiten des Textes Aspekte behandelt werden, die thematisch relevant<sup>156</sup> sind (*Maxime der Relation oder Relevanz*<sup>157</sup>). Dem Textproduzenten sind durch dieses auf der Maxime der Relevanz basierende *Prinzip der Orientierung auf ein Thema* also inhaltliche Beschränkungen darüber auferlegt, was er in einem Text sagen kann, welche Teil- oder Nebenthemen er aufbringen kann. Nicht nur der Text als Ganzes muss auf ein Thema hin orientiert sein, sondern auch jede Äußerungseinheit sollte dabei in ihrem Kontext thematisch relevant sein – zwischen den aufeinander folgenden Äußerungseinheiten sollten thematische Zusammenhänge bestehen. Lokale und globale thematische Relevanz sind demnach Voraussetzung für den thematisch kohärenten Aufbau des Textes und damit die kohärente *inhaltlich-thematische Organisation* des Textes.

Aus Sicht des Produzenten kann das Thematisieren als textorganisierende Handlung aufgefasst werden, das einen zentralen Bestandteil der Textproduktion darstellt. Das Prinzip der Orientierung auf ein Thema und die damit einhergehende inhaltlich-thematische Organisation des Textes wirkt jedoch nicht nur auf die Textproduktion ein, sondern auch auf die Rezeption: Es evoziert eine Erwartungshaltung des Textrezipienten im Hinblick auf die behandelten Themen (bzw. Inhalte). Der Textrezipient erwartet, dass der Textproduzent sich thematisch

---

<sup>155</sup> Zum Zusammenhang von Thema und Intention stellen Heinemann/ Heinemann (2002, 180) fest, dass Text-Thema und Intention gleichermaßen aus dem „Stimulus“ für das Verfassen des Textes (der „Kommunikationsaufgabe“) hervorgehen. Auch Brinker (1997, 60) betont, dass im Kontext der Textproduktion Thema und Intention interagieren, insofern die „Themenentfaltung wesentlich durch kommunikative und situative Faktoren (wie Kommunikationsintention und Kommunikationszweck (...) etc.) gesteuert wird“. Es scheint sinnvoll, davon auszugehen, dass nicht alle Texttypen in gleicher Weise durch diese Faktoren gesteuert sind. Während bei sachverhaltsbezogenen Texten das Thema der zentrale Faktor für den Aufbau des Textes zu sein scheint (vgl. Lötscher 1987, 121), sieht dies bei Face-to-Face-Kommunikation sicherlich anders aus (vgl. Heinemann und Heinemann 2002, 179). Bei der Beschreibung der Prozesse und Prinzipien für die Textproduktion und -rezeption ist deshalb grundsätzlich zu unterscheiden zwischen verschiedenen Texttypen. Wie bereits in Kapitel 2 angesprochen, befasste ich mich in dieser Arbeit ausschließlich mit konzeptionell und medial schriftlichen Texten, insbesondere expositorischen Texten.

<sup>156</sup> Was von Produzent und Rezipient als thematisch relevant aufgefasst wird, hängt von diversen Faktoren ab: dem situativem Kontext, dem enzyklopädischem und episodischem Hintergrundwissen, dem sprachlichen Wissen (z.B. Wissen über die Struktur einer Textsorte), dem Textkontext (Stichwort: Intertextualität) etc.

<sup>157</sup> Die Konversationsmaxime der Relevanz wurde ursprünglich von Grice (1979, 249) entwickelt („Maxime der Relation“: „Sei relevant!“) und später z.B. von Wilson/Sperber (2004) im Rahmen einer kognitiven Theorie weiterentwickelt (siehe Kapitel 3.4.1).

relevant äußert, d.h. sich an das zu Beginn des Textes aktivierte Textthema hält. Dementsprechend versucht der Textrezipient im Akt des Textverstehens alle Äußerungen in den evozierten *thematischen Rahmen* (s.u.) einzubinden und thematische Kohärenz zu konstruieren, indem er nach lokalen oder globalen thematischen Bezügen zwischen den Äußerungseinheiten sucht. Die *Orientierung auf ein Thema* ist demnach ein *zentrales Organisationsprinzip* sowohl für die Produktion als auch die Rezeption von Texten.

Die thematisch gesteuerte Organisation des Textes wird beschränkt durch Wissen (Ebene der Kognition): Neben dem Einfluss des Sprachwissens (z.B. des Wissens um die Verwendung sprachlicher Ausdrücke in bestimmten Gebrauchssituationen) ist im Zusammenhang mit dem Thematisieren vor allem das Handlungs- bzw. Interaktionswissen (insbesondere das Wissen über Textklassen und Textmuster) sowie das Sachwissen von entscheidender Bedeutung für die inhaltlich-thematische Organisation des Textes. Die generische Texttypstruktur sowie die visuelle und logische Textstruktur beschränken die möglichen Thementypen und deren Abfolge sowie die Art der Signalisierung von Themen (z.B. durch Titel oder Überschriften) und Themenwechseln. Die prototypische bzw. musterhafte globale Texttypstruktur kann dabei auch als unterspezifizierte *Thementypenstruktur* aufgefasst werden (siehe Kapitel 2.6.3.6 sowie Kapitel 3.3.2.2). Auf der anderen Seite beeinflusst auch das Sachwissen, welche Gegenstände und Sachverhalte als Aspekte eines Themas im Text thematisiert werden können. Wie bereits oben angesprochen, muss das, was zu einem Thema gesagt wird, über welche Aspekte ein Thema behandelt werden kann, von Kommunikanten als relevant erkannt werden, zwischen den sprachlichen Teilhandlungen des Textes müssen sich thematische Zusammenhänge konstruieren lassen. Das aber setzt voraus, dass sich die Teilhandlungen in einen gemeinsamen *thematischen Rahmen* einbinden lassen. Dieser thematische Rahmen beinhaltet alle denkbaren Zusammenhänge zwischen Themen bzw. Inhalten; insofern es um „denkbare Zusammenhänge“ geht, ist er einerseits abhängig von den in Form von kognitiven Schemata repräsentierbaren *Sachverhaltskomplexen* (siehe Kapitel 3.2.1). Diese bilden somit die Basis für die potentiell konstruierbaren thematischen Zusammenhänge des Textes. Beim Thematisieren als textorganisierender Handlung wird Sachwissen demnach strukturiert und sprachlich repräsentiert (vgl. Rothkegel 1989, 5). Kognitive Strukturen werden jedoch nicht eins zu eins in Textstrukturen überführt, sie werden im Text vielmehr durch sprachliche Explikation aktualisiert, erweitert, variiert oder neu konstruiert.<sup>158</sup> Wenn in einem Text neue Gegenstände als Aspekte des Themas eingeführt werden, d.h. neue Verknüpfungen im Sachbereich erzeugt werden, muss der Zusammenhang zum Textthema oder zu einem bereits eingeführten Aspekt des Textthemas allerdings sprachlich expliziert werden (z.B. durch kohäsive oder visuelle Mittel). Dies setzt wiederum voraus, dass Produzenten im Vertextungsprozess Hypothesen über das Wissen der potentiellen Rezipienten aufstellen und abschätzen können, an welchen Stellen des Textes Zusammenhänge zwischen Sachverhalten bzw. Äuße-

---

<sup>158</sup> Dies gilt insbesondere für explikative Texte, deren Funktion ja gerade die Darstellung und Erläuterung von Informationen ist, die dem Rezipienten noch unbekannt sein können.

rungseinheiten implizit oder explizit ausgedrückt werden können. Ein einführendes Lehrbuch zur Semantik muss beispielsweise wesentlich expliziter sein als ein Nachrichtentext zum Thema „Flugzeugabsturz“, über das Rezipienten ein größeres vorgängiges Wissen besitzen, insofern sie beispielsweise wissen, dass „Flugroute“, „Absturzstelle“, „Passagiere“, „Wetter“, „Reaktionen von Politikern“ Aspekte des Themas darstellen und in einem direkten thematischen Zusammenhang zueinander stehen.

Der Text spiegelt demnach auch die Annahmen des Textproduzenten über das Sachwissen der Rezipienten wider, insofern die in den Teilhandlungen bzw. Äußerungseinheiten behandelten Aspekte des Themas entweder implizit oder explizit verknüpft sind und so entweder implizit oder explizit als Aspekte des Themas instanziiert werden: Wenn ein Gegenstand implizit als Aspekt des Themas (Teilthema) instanziiert wird, muss der Rezipient zum Verstehen des Zusammenhangs von Thema und Aspekt auf kognitive Schemata bzw. stereotypisches Wissen zurückgreifen (zum Begriff des Stereotyps und des stereotypischen Wissens siehe Fußnote 126). Wenn ein Gegenstand explizit als Aspekt des Themas (Teilthema) instanziiert wird, kann der Rezipient zum Verstehen des Zusammenhangs von Textthema und Aspekt auf das sprachliche Material zurückgreifen. Das Verstehen des Textthemas kann erleichtert werden, wenn Aspekte des Themas sowohl kognitiv plausibel sind – dies ist der Fall, wenn der Zusammenhang von Teilthema und Textthema nicht im Widerspruch zu stereotypischem Sachwissen steht, d.h. Wissen, das man als Produzent beim Rezipienten als bekannt voraussetzen kann – als auch sprachlich explizit mit einem anderen Teilthema (horizontaler Zusammenhang) oder dem Textthema (vertikaler Zusammenhang) verknüpft werden können.

Der Text erweitert demnach das vorhandene kognitive Wissen, so dass zu den „denkbaren Zusammenhängen“ nicht nur die abgespeicherten Sachverhaltskomplexe gehören, sondern auch die durch Sprache konstruierbaren Sachverhaltskomplexe.<sup>159</sup> Der Begriff des „thematischen Rahmens“ ist damit ein rein gedankliches Konstrukt, das niemals extensional bestimmt werden kann. Der im Text versprochene Sachverhaltskomplex stellt einen Ausschnitt des thematischen Rahmens dar und wird im Folgenden als *thematisierter Sachverhaltskomplex* bezeichnet, insofern das Thema als zentrales Konzept aufgefasst werden kann, das im Zentrum dieses im Text behandelten Sachverhaltskomplexes steht.

Die Behandlung eines *Sachverhaltskomplexes* wird auf der Ebene der Äußerungseinheiten des Textes *realisiert* durch die sprachliche Bezugnahme auf *Dis-*

---

<sup>159</sup> Hellwig (1984a, 56) bemerkt hierzu folgendes: „Ein Teil des Zurechtfindens in der Welt besteht darin, die aktuellen Gegenstände und Vorgänge den vertrauten Stereotypen zuzuordnen. Dies gilt auch für die Welt, die ein Text vermittelt. Angesichts der Ausschließlichkeit, die gegenwärtig ‚frames‘, ‚scenes‘, ‚scripts‘ usw. in den textlinguistischen Ansätzen der künstlichen Intelligenz beanspruchen, ist zu betonen, daß es sich wirklich nur um einen Teil (...) handelt. Der Wert der Sprache für den menschlichen Geist liegt nicht zuletzt darin, daß man sich mit ihrer Hilfe über die Änderung von ‚frames‘ und über den Wandel von Prototypen verständigen kann. (...) Ich vermute daher, daß in den Texten nur das, was man implizit läßt, an das Stereotypische gebunden ist, während das, was explizit gesagt werden kann, prinzipiell frei ist.“

*kursentitäten* bzw. Gegenstände<sup>160</sup>, und durch die Verknüpfung dieser Diskursentitäten durch semantische bzw. konzeptuelle Relationen sowie durch die Verknüpfung der auf diese Diskursentitäten bezogenen sprachlichen Handlungen durch *Diskurs- bzw. Kohärenzrelationen* (zum Zusammenhang von Kohärenzrelationen und thematischen Zusammenhängen siehe Kapitel 3.3.2.1.). Insofern Texttypstrukturkategorien als unterspezifizierte Thementypen aufzufassen sind, werden Aspekte des Themas im Text als Diskursentitäten in verschiedenen texttypspezifischen Kategorien instantiiert. Über die Texttypstruktur können Diskursentitäten demnach kategorial klassifiziert werden (siehe Kapitel 3.3.2.2). Zwischen den verschiedenen kategorialen Einheiten der Texttypstruktur können darüber hinaus ebenfalls Diskurs- bzw. Kohärenzrelationen ermittelt werden (vgl. Bärenfänger et al. 2010, 93), die somit implizit zur Beschreibung der thematischen Zusammenhänge zwischen Diskursentitäten genutzt werden können. Über die Strukturierung des Textes auf den Ebenen der Informationsstruktur, lexikalischen, referenziellen und relationalen Struktur werden die den Sachverhaltskomplex konstituierenden Aspekte des Themas als zentrale Diskursentitäten realisiert bzw. als *zentrale Gegenstände* des Textes instantiiert (siehe Kapitel 3.3.1). Ob es sich bei einem in einer Äußerungseinheit behandelten Gegenstand um einen Aspekt des Textthemas bzw. einen Teil eines Sachverhaltskomplexes handelt – bzw. dass die Referenzierung *verschiedener* Gegenstände im Text die Behandlung *eines* Themas und seiner Aspekte realisiert – kann entweder unter Hinzuziehung des Sachwissens und damit des Wissens über den Zusammenhang von Gegenständen, Ereignissen, Situationen erschlossen werden, d.h. unter Rekurs auf kognitive Schemata, oder durch das explizite sprachliche Verknüpfen von Teilhandlungen, die einen Gegenstand behandeln (z.B. durch Konjunktionen wie „weil“ – kann eine CAUSE-Relation signalisieren).

Neben der Instantiierung der Teile des Sachverhaltskomplexes durch Diskursentitäten wird das Thema im Text darüber hinaus *reflektiert* durch verschiedene metasprachliche, visuelle und lexikalische Mittel: Metasprachliche Phrasen (z.B. „Was +th betrifft“, „Im Hinblick auf +th“, „Zum Beispiel +th“, „Im Zusammenhang mit +th X +th' Y“, „Erstens/zweitens/... X +th Y“, siehe Hoffmann 1997, 524f.) dienen beispielsweise dazu, die thematische Organisation des Textes zu signalisieren, d.h. z.B. Themenwechsel zu markieren. Über bestimmte konventionalisierte visuelle Mittel, wie die Hervorhebung einer Einheit des Textes in Form einer Überschrift, können Einheiten des Textes als thematisch besonders relevant hervorgehoben werden – sie werden *markiert*.<sup>161</sup> Eine Teilmenge der Inhaltswörter des Textes wird dazu verwendet, um über ein spezifisches Thema zu reden (vgl. Kapitel 3.3.1.3); sie werden zur Bezugnahme auf ein bestimmtes Schema gebraucht (vgl. Kapitel 2.6.2.2) und reflektieren dieses Thema demnach

<sup>160</sup> Gegenstand wird synonym zu Diskursentität und Diskursreferent gebraucht; er kann sowohl eine Situation, ein Ereignis als auch ein konkreter oder abstrakter Entität sein und sowohl nominal als auch propositional realisiert sein.

<sup>161</sup> Allerdings dienen nicht alle Überschriften der Signalisierung des Themas. Überschriften können vielmehr verschiedene Funktionen haben. Die Arten von Überschriften und ihre Funktionen können dabei zwischen Instanzen verschiedener Texttypen divergieren (vgl. Kapitel 3.3.1.4).

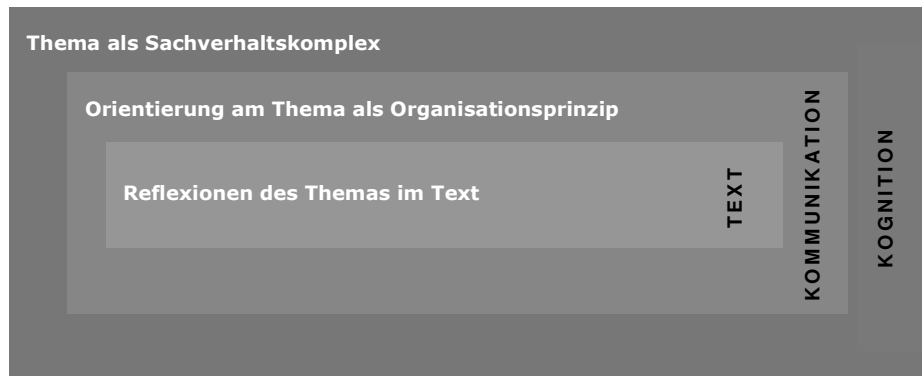


Abbildung 7: Ebenen des Themas

auf der Textebene.<sup>162</sup>

Zusammenfassend kann festgehalten werden, dass sich das Thema als Konzept beschreiben lässt, das auf drei miteinander interagierenden Ebenen wirkt (siehe Abbildung 7):

1. Das Thema mit seinen Aspekten ist aufzufassen als der in einer Kommunikation behandelte Sachverhaltskomplex, der auf stereotypischen Sachverhaltskomplexen basiert und die Funktion eines kognitiven Einordnungsrahmens hat (→ kognitive Ebene);
2. das Thema bzw. die Orientierung am Thema stellt ein Organisationsprinzip dar, das die Textproduktion und -rezeption steuert; das Thema hat auf dieser Ebene die Funktion eines kommunikativen Einordnungsrahmens (→ kommunikative Ebene);
3. das Thema wird im Text durch sprachliche und visuelle Mittel reflektiert und realisiert<sup>163</sup>; einen besonderen Stellenwert hat dabei die Menge zentra-

<sup>162</sup> Gloning (1994a) weist darauf hin, dass die Fähigkeit, über bestimmte Themen zu reden, unter anderem auf der Beherrschung der betreffenden thematischen Wortschätze basiert. Umgekehrt kann demnach der Wortschatz als Indiz für das Reden über ein bestimmtes Thema gesehen werden.

<sup>163</sup> In der vorliegenden Arbeit wird einerseits davon gesprochen, dass ein Thema im Text reflektiert wird, andererseits, dass ein Thema im Text realisiert (oder auch instanziiert) wird. Die Ausdrücke „realisiert“ und „instanziiert“ werden dazu verwendet, um auf das Thema und seine Aspekte (als Teile einer kognitiven Sachverhaltsstruktur) Bezug zu nehmen und darzustellen, dass diese Teile des kognitiven Sachverhaltskomplexes als Diskursentitäten, über die etwas gesagt wird, im Text realisiert bzw. instanziiert werden (d.h. eine Diskursentität im Text ist eine Instanz eines Konzepts einer kognitiven Sachverhaltsstruktur). Daneben wird das Thema durch verschiedene sprachliche und visuelle Mittel reflektiert (z.B. lexikalisch, durch nicht-nominale Inhaltswörter). Eine sprachliche Einheit, die das Thema realisiert, reflektiert dieses dabei gleichzeitig, nicht aber umgekehrt.

ler Gegenstände (bzw. Diskursentitäten), über die in den Äußerungseinheiten des Textes etwas gesagt wird (→ informationsstrukturelle, lexikalische und referenzielle Ebene); die Aspekte des Themas werden dabei durch lexikalisch-semantische Relationen (→ lexikalische Ebene), Koreferenzrelationen (→ referenzielle Ebene) oder Diskursrelationen zwischen den sprachlichen Teilhandlungen miteinander verknüpft (→relationale Ebene); die Aspekte des Themas können darüber hinaus in verschiedene Texttypstrukturkategorien sowie logische Textstrukturkategorien eingeordnet werden; sie stehen demnach in einem bestimmten kategorialen Zusammenhang (→ generische und logische Text(typ)struktur); darüber hinaus wird das Thema und die thematische Organisation des Textes durch metasprachliche Ausdrücke, visuelle Mittel und den verwendeten Wortschatz (speziell Inhaltswörter) im Text reflektiert – diese Mittel können deshalb ebenso wie die Diskursentitäten als Indizien zur Bestimmung des Themas als Sachverhaltskomplex genutzt werden.

Aus einer kognitiven Perspektive gesehen, stellt das Thema eine Wissensstruktur zu einem bestimmten Ereignis, einer Situation oder einem Gegenstand dar (Thema als Sachverhaltskomplex und kognitiver Einordnungsrahmen). Im Kommunikationsprozess fungiert das Thema als kommunikativer Einordnungsrahmen, an dem sich Textproduzenten und -rezipienten beim Produzieren bzw. Verstehen des Textes orientieren können; diese Orientierung am Thema wirkt dabei als Organisationsprinzip auf die Textproduktion und -rezeption ein. Auf der Ebene des Textes als Produkt wird das Thema durch sprachliche und visuelle Mittel reflektiert (Thema als Texteigenschaft) und durch die Menge zentraler Gegenstände, über die in den Äußerungseinheiten etwas gesagt wird, realisiert. Diese Gegenstände sind thematisch zusammenhängend, insofern sie Teil des im Text thematisierten Sachverhaltskomplexes sind, der auf stereotypischen Sachverhaltskomplexen basiert, jedoch nicht identisch mit diesen ist.

In den Kapiteln 3.2, 3.3 und 3.4 sollen die drei oben genannten Ebenen des Themas auf die Beschreibungsebenen „Kognition“ (als Voraussetzung und Beschränkung der Textproduktion und -rezeption), „Text“ (als Produkt der Textproduktion und -rezeption) und „Kommunikation“ (als Prozess der Textproduktion und -rezeption) bezogen und dadurch näher beleuchtet werden.

Im Folgenden sollen jedoch zunächst die in Kapitel 2 dargestellten Thema-Konzepte in das Ebenen-basierte Thema-Modell eingeordnet werden und eine erste, hypothesenartige Differenzierung der Rolle des Themas auf den unterschiedlichen Beschreibungsebenen der Textstruktur vorgenommen werden, die in den nachfolgenden Kapiteln noch ausführlicher ausgearbeitet werden wird. Durch die Zuordnung der in Kapitel 2 dargestellten Thema-Konzepte zu den verschiedenen Ebenen des Themas sowie der Textstruktur werden diese Konzepte zugleich in einen integrativen Rahmen gebracht, der es ermöglicht, die Rolle dieser Konzepte bei der Beschreibung des komplexen Phänomens Thema gleichzeitig zu differenzieren als auch – im Rahmen einer ganzheitlichen Perspektive – miteinander zu verknüpfen, und sie auf diese Weise als unterschiedliche Facetten eines komplexen Konzepts zu begreifen.



### 3.1.2 Einordnung der zentralen Thema-Konzepte aus Kapitel 2 in die drei Ebenen des integrativen Thema-Modells

In diesem Kapitel sollen die in Kapitel 2 behandelten Thema-Konzepte den Ebenen des integrativen Thema-Modells zugeordnet werden – eine erste Übersicht hierüber gibt Tabelle 5.<sup>164</sup>

Ebenen des Themas	Beschreibungsebenen des Textes	Thema-Konzepte
Thema als Menge zentraler Gegenstände, über die etwas gesagt wird  Eigenschaften auf der Textoberfläche als Reflexionen des Themas im Text	Informationsstruktur	<ul style="list-style-type: none"> <li>- Thema als Mitteilungsgegenstand („aboutness“) (Kapitel 2.2.1.1)</li> <li>- Thema als bekannte Information (Kapitel 2.4) bzw. als die kommunikativ am wenigsten dynamische Information des Satzes (Kapitel 2.3.1)</li> <li>- Thema als Element in Satzerstposition (Kapitel 2.1.1) bzw. als Ausgangspunkt der Mitteilung (Kapitel 2.2.1.1)</li> </ul>
	Lexikalische Struktur	<ul style="list-style-type: none"> <li>- Thema als Gruppe häufig vorkommender Inhaltswörter (Kapitel 2.5)</li> <li>- Thema als Cluster von Schlüsselentitäten (Kapitel 2.5)</li> </ul>
	Referenzielle Struktur	<ul style="list-style-type: none"> <li>- Thema als zentraler, salienter Gegenstand des Textes (Kapitel 2.2.1.2 und 2.5)</li> </ul>

<sup>164</sup> Anmerkungen zur Tabelle: Bei den der Texttypstruktur und der visuellen Struktur zugeordneten „Thema-Konzepten“ handelt es sich nicht um Thema-Konzepte im engeren Sinne, sondern vielmehr um Aussagen zum Zusammenhang dieser Strukturen mit dem Thema. Gleichzeitig wird hier, im Gegensatz zu allen übrigen in der Tabelle aufgeführten „Thema-Konzepten“ nicht auf die bisherige Forschung, wie sie in Kapitel 2 dargestellt wurde, rekurriert. Stattdessen werden hier eigene Annahmen ergänzt, um darzulegen, dass sich das Thema auch auf den beiden genannten Strukturebenen widerspiegelt.

Die semantischen Thema-Konzeptionen, die das Thema als abstrakten semantischen Kern des Textes (als Basis- oder Makroproposition oder Prädikat-Aktanten-Struktur) definieren, das den semantischen Ausgangspunkt für die inhaltliche bzw. semantische Entfaltung des Textes darstellt (siehe Kapitel 2.2.2), weisen dem Thema eine ähnliche Funktion zu, wie die kognitiven Konzeptionen. Statt von der Annahme einer semantischen Basis- oder Makroproposition wird in kognitiven Arbeiten allerdings davon ausgegangen, dass die inhaltliche Basis des Textes ein kognitives Schema bzw. ein Sachverhaltskomplex ist, der als Grundlage für die thematische Organisation des Textes dient. Während in den semantischen Thema-Konzeptionen das Thema als Proposition, d.h. als zweistellige Einheit aus Referenz und Prädikation aufgefasst wird (d.h. als Gegenstand plus Aussage), wird in den kognitiven Konzeptionen angenommen, dass das Thema im Text repräsentiert wird über die den Sachverhaltskomplex konstituierenden und im Text behandelten Diskursentitäten, über die Relationen zwischen diesen Entitäten sowie die Eigenschaften, die diesen Entitäten im Text zugewiesen werden. Die Prädikation oder Aussage über eine Diskursentität wird demgegenüber nicht grundsätzlich als Teil des Themas aufgefasst. Die Konzeption vom Thema als semantischem Kern bzw. als Makroproposition unterscheidet sich demnach nur im Grad ihrer verbalen Spezifität von der kognitiven Thema-Konzeption, nicht aber im Hinblick auf die Funktion, die dem Thema zugewiesen wird. Insofern in dieser Arbeit die kognitive Ebene von Texten im Fokus steht und das Konzept der Texttiefenstruktur, auf der das Thema als semantischer Kern anzusiedeln ist, demgegenüber keine Rolle spielt, sollen die semantischen Thema-Konzeptionen im weiteren Verlauf der vorliegenden Arbeit ausgeklammert und der Fokus stattdessen auf die kognitiven Konzeptionen gelegt werden.

	Relationale Struktur	<ul style="list-style-type: none"> <li>- Thema als Redegegenstand, der im Text in spezifischer Weise entwickelt oder organisiert wird (Themenentwicklung /thematische Organisation/thematische Progression, siehe Kapitel 2.2.1.2)</li> <li>- Thema als semantischer Kern, der im Text in spezifischer Weise entfaltet wird (Grundformen thematischer Entfaltung, siehe Kapitel 2.3.3)</li> </ul>
	Generische Texttypstruktur	- Aspekte des Themas als im Text instanziierte unterspezifizierte Texttypstrukturkategorien (siehe Kapitel 3.2.2)
	Visuelle und logische Struktur	- Titel als Reflexion des Themas im Text (siehe Kapitel 3.3.1.4)
Orientierung am Thema als Organisationsprinzip	Kommunikative Ebene	<ul style="list-style-type: none"> <li>- Thema als Fragliches oder Problemstellung (Kapitel 2.3.2)</li> <li>- Thema als Mittel zur Strukturierung von Diskursen (Kapitel 2.3.3)</li> <li>- Thema als „Verstehenshorizont“ bzw. „Einordnungsinstanz“ (Kapitel 2.3)</li> </ul>
Thema als Sachverhaltskomplex	Kognitive Ebene	<ul style="list-style-type: none"> <li>- Thema als kognitiv aktivierte Diskursentität (Kapitel 2.4)</li> <li>- Thema als Kern eines konzeptuellen Netzes/Schemas (Kapitel 2.4 und 2.5)</li> <li>- Thema als semantischer Kern/Makro-/Basisproposition (Kapitel 2.2.2)</li> </ul>

Tabelle 5: Zum Zusammenhang der Thema-Konzepte mit den in der vorliegenden Arbeit definierten Ebenen des Themas sowie den Beschreibungsebenen des Textes

In allen drei Fällen wirkt das Thema als Quelle und Ergebnis einer in erster Linie bottom-up, durch die Textoberfläche bzw. die ausdrucksseitige Gestaltung des Textes initiierten Kohärenzkonstruktion. Für den Textproduzenten ist das Thema der Ausgangspunkt für die ausdrucksseitige Gestaltung und Strukturierung des Satzes – ein Satz muss sich ausdrucksseitig in den Äußerungskontext einfügen und in diesem (thematischen) Kontext kohärent sein (z.B. indem er referenziell oder lexikalisch kohärent ist). Für den Textrezipienten wiederum ist das Thema Ergebnis der Interpretation der ausdrucksseitigen Gestaltung des Textes: Er kann die Satzgliedreihenfolge als indexikalisches Zeichen für eine (thematische) Gewichtung und Perspektivierung interpretieren; er kann das lexikalische Material in Bezug zueinander setzen und als Ausdruck bzw. indexikalisches Zeichen für das Reden über ein bestimmtes Thema verstehen; er kann die referenzielle Wiederaufnahme auf die Orientierung des Textes auf ein (oder mehrere) zentrale Gegenstände bzw. Themen zurückführen. Sowohl bei der Produktion als auch Rezeption werden dabei Merkmale der Textoberfläche von den Kommunikanten dazu genutzt, das Thema zu erschließen.

Die lexikalische und referenzielle Ebene sind daneben auch Quellen für die Konstruktion inhaltlich-thematischer Kohärenz<sup>165</sup>: Durch die Verwendung verwandter Inhaltswörter (thematischer Wortschatz) und durch die Wiederaufnahme zentraler Diskursentitäten bewirkt der Produzent die Orientierung des Textes auf ein (oder mehrere) zentrale Themen. Das Thema lässt sich aus Produzentenperspektive als Quelle dieser inhaltlichen Kohärenz beschreiben, aus Rezipientenperspektive als Ergebnis des Verstehensprozesses, der sich an der lexikalischen und referenziellen Ebene festmacht. Zur inhaltlichen Kohärenz des Textes trägt darüber hinaus die relationale Ebene durch die inhaltliche Verknüpfung von zwei oder mehreren Äußerungen bei, die so als miteinander verknüpfte Aspekte des Themas eingeführt werden. Solche Aspekte können z.B. Angaben zu sätzlicher Details zu einer Situation sein (ELABORATION), Angaben einer Ursache (CAUSE) oder Folge (CONSEQUENCE). Daneben können Äußerungen auch argumentativ verknüpft sein, durch Relationen wie Antithese (ANTITHESIS), Vorbereitung (PREPARATION) oder Stützung (SUPPORT).<sup>166</sup> Die relationalen Zusammenhänge können deshalb als Quelle (Produktionssicht) und Ergebnis (Rezeptionssicht) für die Konstruktion inhaltlicher sowie argumentativer Kohärenz aufgefasst werden, die Grundlage sowohl für das Signalisieren bzw. Etablieren als auch für das Verstehen der Zusammenhänge zwischen Aspekten des Themas ist.

Auf der Ebene der Kommunikation dient das Thema als Mittel zur Organisation und Strukturierung von Dialogen und Texten (siehe Kapitel 2.2.1.2 und 2.3.3); das Thema kann in diesem Zusammenhang als „Fragliches“ oder als „Problemstellung“ beschrieben werden (siehe Kapitel 2.3.2), die in einer Äußerungsfolge bearbeitet wird. Die Funktion des Thematisierens besteht in der Beantwortung dieses Fraglichen bzw. im Bearbeiten der Problemstellung. Produzenten müssen sich demnach bei der Texterstellung an diesem Fraglichen bzw. dieser Problemstellung abarbeiten. Rezipienten haben wiederum die Erwartung, dass sich Produzenten bei der Texterstellung an einem globalen Thema orientieren. Die auf Basis der ersten Rezeptionsschritte aufgebaute Hypothese über das Thema des Textes dient Rezipienten deshalb als „Verstehenshorizont“ (siehe Kapitel 2.3) und Grundlage für den Verstehensprozess – dessen Ergebnis wiederum das Verstehen des Themas und seiner thematischen Organisation ist. Das Thematisieren als Mittel zur Organisation des Textaufbaus ist damit sowohl Quelle als

<sup>165</sup> Obwohl in diesem Abschnitt von verschiedenen Arten von Kohärenz gesprochen wird, die durch bestimmte sprachliche Mittel auf jeweils spezifischen Ebenen der Textstruktur konstituiert werden, soll dies keineswegs bedeuten, dass bestimmte Typen sprachlicher Mittel oder bestimmte Ebenen der Textstruktur immer nur eine Art von Kohärenz konstituieren. Eine sprachliche Einheit kann sehr gut sowohl einen Beitrag zur Bildung von ausdrucksseitiger, inhaltlicher als auch funktional-argumentativer Kohärenz leisten. Eine getrennte Beschreibung (sowohl der sprachlichen Mittel auf den verschiedenen Ebenen der Textstruktur als auch der Arten der Kohärenz) bietet allerdings im Gegensatz zu einer integrierten Beschreibung den Vorteil, dass die beschriebenen Merkmale auf diese Weise deutlicher sichtbar werden und durch computerlinguistische Verfahren einzeln ermittelt werden können – was wiederum Grundlage für eine maschinelle Identifikation des Themas ist. Das in einem nächsten Schritt die verschiedenen Merkmale wieder im Hinblick auf ihre Interaktionen untersucht und ausgewertet werden müssen, bleibt davon unbenommen.

<sup>166</sup> Zu den Kategorien ELABORATION, CAUSE, CONSEQUENCE etc. siehe Kap. 2.6.3.4 sowie 3.3.2.1.

auch Ergebnis kommunikativ-interaktionaler Kohärenz.

Im Zusammenhang mit der Ebene der Kognition stellt das Thema einerseits eine „kognitiv aktivierte Diskursentität“ dar, die im Zentrum der mentalen Aufmerksamkeit des Rezipienten liegt, andererseits den „Kern eines konzeptuellen Netzes“, das im Text entfaltet wird (siehe Kapitel 2.4). Thematische Kohärenz ist in diesem Zusammenhang abhängig von kognitiver Kohärenz und wird geprägt durch kommunikative, kognitive und sprachliche Konventionen: *Was* man zu einem Thema sagen *kann*, ist abhängig von dem Wissen, das man über einen Themenkomplex (verstanden als Sachverhaltskomplex, kognitives Schema oder konzeptuelles Netz, deren kleinste Einheit die aktivierten Diskursentitäten sind) hat; was man sagen *sollte* und darf, ist dagegen geprägt von den Konventionen des Texttyps sowie von kommunikativen Gegebenheiten; *wie* man etwas sagen kann, ist sowohl abhängig von der Struktur des Sachwissens als auch vom Sprach- und Handlungswissen (z.B. Wissen über den Gebrauch von Wörtern in einer bestimmten kommunikativen Situation oder Wissen über Texttypkonventionen).

Die existierenden Thema-Konzepte sind einerseits strukturell zu differenzieren: Sie beschreiben Phänomene auf der Ebene der Informationsstruktur, der lexikalischen, referenziellen und relationalen Struktur, der generischen Texttypstruktur sowie der visuellen bzw. logischen Textstruktur (textinterne Ebenen) sowie auf der Ebene der Kommunikation und Kognition. Aus diesen strukturellen Unterschieden resultieren auch funktionale Unterschiede im Hinblick auf die Konstruktion von Kohärenz; das Thema fungiert in den unterschiedlichen Konzeptionen als Quelle und/oder Ergebnis ausdrucksseitiger, inhaltlicher, argumentativer, kommunikativ-interaktionaler und kognitiver Kohärenz.

### 3.2 Zum Zusammenhang von Thema und Wissen

Die thematische Struktur eines Textes beruht zum einen darauf, dass der Text bzw. die sprachliche Texthandlung auf ein übergeordnetes Textthema bezogen ist, zum anderen darauf, dass zwischen den sprachlichen Teilhandlungen des Textes ein thematischer Zusammenhang besteht: die Gegenstände, auf die die sprachlichen Teilhandlungen bezogen sind, müssen sich in einen *thematischen Rahmen* einbinden lassen. Die thematischen Zusammenhänge zwischen diesen Gegenständen bzw. Aspekten können im Text sprachlich explizit oder implizit realisiert sein. Für sprachlich implizit realisierte Zusammenhänge gilt, dass sie auf dem in kognitiven Schemata oder Frames repräsentierbaren Sachwissen zu einem Thema bzw. den Sachverhaltskomplexen, die zum Thema eines Textes erhoben werden, beruhen sollten, um für einen Rezipienten verständlich sein zu können. Was in einem Text zum Thema gesagt werden kann, basiert demnach u.a. auf dem Wissen, das ein Produzent über den thematisierten Sachverhaltskomplex besitzt; auch das, was ein Rezipient beim Textverstehen als thematisch relevant versteht, ist abhängig von diesem Wissen (top-down-Verarbeitung). Insofern stereotypische Sachverhaltskomplexe durch *kognitive Schemata* oder

*Frames* repräsentiert werden können, ist davon auszugehen, dass sich Produzenten und Rezipienten beim Schreiben und Verstehen eines Textes an diesen Schemata oder Frames orientieren.

Im Folgenden soll zunächst die Rolle des Sachwissens und das Zusammenspiel von Sachverhaltskomplex und thematischen Zusammenhängen zwischen den Gegenständen, auf die die sprachlichen Teilhandlungen des Textes bezogen sind, beleuchtet werden (3.2.1). Daran anschließend werden Zusammenhänge zwischen Textmusterwissen und Thema (3.2.2), Sprachwissen und Thema (3.2.3) sowie Wissen über die Kommunikationssituation und Thema (3.2.4) erörtert.

### 3.2.1 Thema und Sachwissen: Thema als Sachverhaltskomplex

Ein Sachverhaltskomplex repräsentiert ein stereotypisches Erinnerungsmuster zu einer bestimmten Situation, einem Ereignis oder Gegenstand. Er besteht aus einer Menge von miteinander verknüpften einfachen Sachverhalten  $2-n$ , die Teile dieser Situation repräsentieren. Einfache Sachverhalte bestehen dabei aus einer Diskursentität und einem Merkmal oder einer Aussage, die dieser Entität zugewiesen wird. Ein Sachverhalt kann sprachlich realisiert werden in Form einfacher oder komplexer Diskurssegmente. Im Folgenden wird unterschieden zwischen a) *stereotypischen Sachverhaltskomplexen* und b) *thematisierten Sachverhaltskomplexen*. Bei den stereotypischen Sachverhaltskomplexen handelt es sich um die mental abgespeicherten stereotypischen Erfahrungen, die ein Textproduzent/Rezipient zu einem bestimmten Realitätsbereich<sup>167</sup> bzw. Sachverhaltskomplex gemacht hat. Diese können in kognitiven Frames repräsentiert werden als stereotypisches Wissen über ein Ereignis, eine Situation, Eigenschaften eines Gegenstands, seinen Zweck und seine Beziehungen zu anderen Gegenständen sowie als Wissen über typische Handlungen, die mit einer Situation oder einem Gegenstand verbunden sind (siehe Kapitel 2.7.2 und 2.7.3).

Thematisierte Sachverhaltskomplexe bezeichnen diejenigen Sachverhaltskomplexe, die in einem Text zum Thema erhoben werden: Die sprachliche Texthandlung insgesamt bezieht sich dabei auf ein übergeordnetes Textthema (den Sachverhaltskomplex), während in den sprachlichen Teilhandlungen des Textes jeweils ein Aspekt des Themas behandelt wird (Teile des Sachverhaltskomplexes). Stereotypische Sachverhaltskomplexe werden beim Thematisieren in der Regel durch den Sprecher individuell modifiziert, indem beispielsweise Teile des Sachverhaltskomplexes weggelassen oder eingefügt werden – diese Modifizierungen sind bereits als Leerstellen im Frame angelegt; sie fungieren dabei als „Platzhalter für spezifischere Daten“ (Schnotz 1994, 61). Auch kommunikative und sprachliche Prinzipien und Muster wirken auf die Realisierung von Sachverhaltskomplexen in Texten ein (siehe Kapitel 2.8.1), so dass die Produktion von Texten als Zusammenspiel von Wissensstrukturen und sprachlichen und kommunikativen Prinzipien zu sehen ist (vgl. Rothkegel 1984a).

---

<sup>167</sup> „Realität“ schließt auch die „Realität“ in fiktiven oder möglichen Welten ein (siehe z.B. Kapitel 2.7.2)

Aufgrund der Fülle der im Prozess der Textproduktion vorgenommenen Aktualisierungen, Erweiterungen und Umstrukturierungen stereotypischer Sachverhaltsstrukturen ist anzunehmen, dass Texte Sachverhaltskomplexe nicht nur wiedergeben, sondern selbst erschaffen (vgl. Antos 1997, 55). Durch Sprache werden neue, bisher unbekannte Zusammenhänge zwischen Teilen des Sachverhaltskomplexes expliziert. Allerdings nehmen auch solche Texte, die neues Wissen konstituieren, auf einer grundlegenden Verständnisebene auf allgemeine kognitive Frames Bezug, insofern nie alle Sachverhalte, die für das Verständnis eines Satzes oder Textes nötig sind, sprachlich expliziert werden, sondern stattdessen vom Textrezipienten auf Basis seines Weltwissens inferriert werden müssen (vgl. Schwarz 1996, 159).

Thematisierter Sachverhaltskomplex und Frame bzw. stereotypischer Sachverhaltskomplex sind demnach nicht identisch, sie weisen allerdings eine Schnittmenge auf: Ein Frame stellt die Repräsentation der Menge der möglichen Teilthemen bzw. fraglichen Aspekte zu einem *stereotypischen Sachverhaltskomplex* bzw. stereotypischem Situationsmodell dar, während die *thematisierte Sachverhaltsstruktur* die im Text realisierten Aspekte des Themas darstellt, die einerseits aus dem stereotypischen Sachverhaltskomplex bzw. dem Frame abgeleitet sind, andererseits aber auch im Text selbst neu konstituiert werden können. Frames als Listen von Fragen bzw. Handlungsanweisungen zu einer stereotypischen Situation (vgl. Kapitel 2.7.3) werden durch den Textproduzenten im Prozess der Textproduktion abgearbeitet, indem er die einzelnen Fragen in den Teilhandlungen des Textes zu beantworten sucht. In der Regel bezieht sich jede Teilhandlung auf eine spezifische Frage, allerdings kann eine Frage auch in mehreren aufeinander folgenden Teilhandlungen behandelt werden. Die Teilhandlungen beziehen sich demnach auf Teile des Frames, während die Texthandlung auf den Frame insgesamt bezogen ist – wobei davon auszugehen ist, dass sich ein Text auch auf mehrere Frames beziehen kann – über jeden Slot eines Frames kann theoretisch ein neuer Frame eingebunden werden, insofern ein Slot als Kopf eines neuen Frames definiert werden kann. Der Zusammenhang von Thema bzw. thematisiertem Sachverhaltskomplex und Frame soll im Folgenden an zwei Beispieltexten illustriert werden. Bei Beispiel 1 handelt es sich um eine Zeitungsnachricht mit dem Titel „Gewalt gegen Castor nimmt zu“ (Text siehe Anhang, 6.1). Beispiel 2 ist ein Enzyklopädie-Eintrag zum Stichwort „Hering“ (Text siehe Anhang, 6.2).

<b>Frame: „Anschlag“ (Instanz der Klasse „Handlung“)</b>	
<b>Slots</b>	<b>Filler</b>
Welche Motive gibt es für die Handlung (den Anschlag)?	Protest gegen Atomtransport
Was hat die Handlung (der Anschlag) zur Voraussetzung bzw. durch was wird die Handlung ausgelöst?	Castor-Transport
Welcher angestrebte Zielzustand soll welches Bedürfnis erfüllen?	Atomtransport-Blockade
Welche wesentlichen Mitspieler fungieren in der Handlung (dem Anschlag)? (→ könnte weiter differenziert werden)	a) AGENS: Atomkraftgegner/Autonome/Einzeltäter b) Verfassungsschutz, Bundeskriminalamt/ „die Behörde in Wiesbaden“ c) Grüne
Was (welche Mittel) benutzt der Aktant bei der Handlung (dem Anschlag)?	Gewalt
Wie ist die Handlung (der Anschlag) realisiert?	Zerstörung von Oberleitungen von Bahnstrecken mit Wurfankern
Welche Folgen zieht die Handlung (der Anschlag) nach sich?	Sachschäden Zugverspätungen („Züge verspäteten sich“) Personenschäden (neg.) („niemand wurde verletzt“)

#### Analyse 1: Thematisierter Sachverhaltskomplex in Beispiel 1

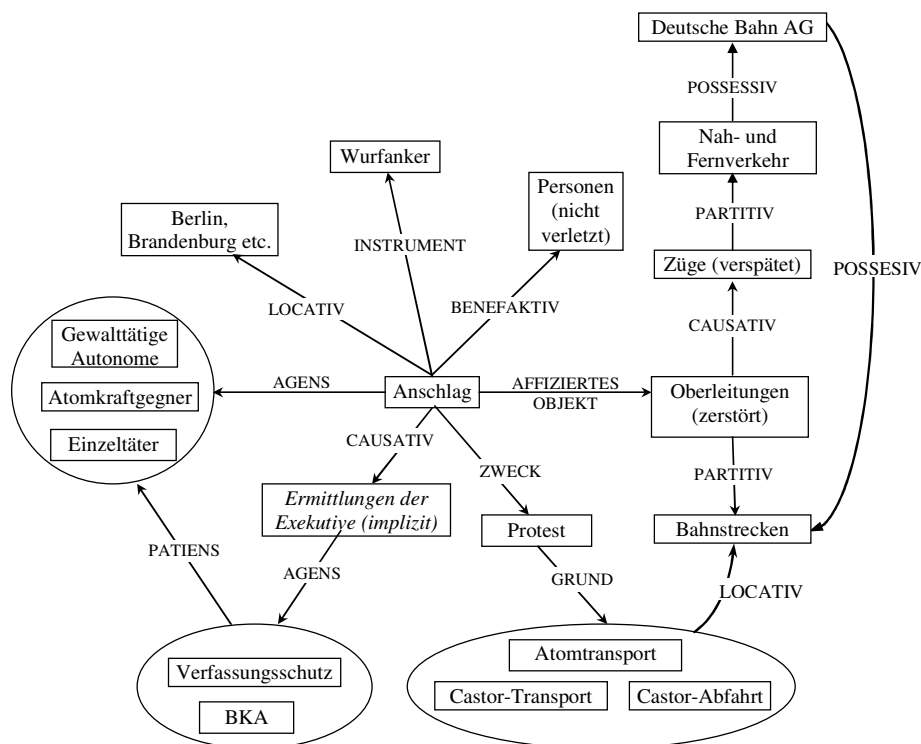
Die Slots dieses Frames repräsentieren abstraktes Wissen über die Teile des Sachverhaltskomplexes „Anschlag“ (bzw. „Handlung“), z.B. Wissen, dass es für die Handlung ein Motiv gibt, dass sie bestimmte Voraussetzungen bzw. Auslöser hat, dass durch den angestrebten Zielzustand der Handlung ein Bedürfnis erfüllt werden soll, dass es in der Handlung bestimmte Mitspieler gibt, dass sie in einer bestimmten Weise realisiert ist, dass sie Folgen nach sich zieht etc. Die Filler des Frames stellen dagegen die konkreten, im Text instanziierten Teile des Sachverhaltskomplexes dar: Das Motiv für die Handlung („Anschlag“) ist „Protest“, Voraussetzung bzw. Auslöser der Handlung ist der bevorstehende „Castor-Transport“, etc. (siehe Analyse 1).<sup>168</sup>

Insofern sich die Diskursentitäten aus Beispiel 1 zum überwiegenden Teil als Bestandteile des Frames „Anschlag“ identifizieren lassen (im Text durch die Ausdrücke „Anschläge“ (3x), „Bahnanschläge“, „die Aktionen“ instanziiert) – die nicht damit zusammenhängenden Entitäten gehen aus den Texttypstrukturkategorien hervor (z.B. Orts- und Zeitangaben, siehe Analyse 14, Kapitel 3.2.2) –, kann der Frame mit dem Kopf „Anschlag“ als Ausgangspunkt der thematischen Textstrukturierung betrachtet werden. Insofern der Text auf der Instanzierung und Behandlung eines einzigen Frames basiert, kann er als monothematisch aufgefasst werden. Der Kopf des Frames („Anschlag“) stellt dabei das singuläre

<sup>168</sup> Neben den hier genannten Slots wird der kognitive Matrixframe „Handlung“ (bzw. der Frame „Anschlag“) durch eine Vielzahl weiterer Slots definiert, die jedoch für den hier behandelten Text keine Rolle spielen (siehe Konerding 1993, 341f.).

Textthema dar, das über verschiedene Aspekte behandelt wird. Diese Aspekte werden dabei durch die Slots (als Fragen, die beantwortet werden wollen) initiiert und an das Textthema angebunden.

Eine alternative, netzartige Repräsentation des im Text thematisierten Sachverhaltskomplexes zeigt Analyse 2. Hier sind die Bezüge (beschriftete Pfeile) zwischen den Teilen des Sachverhaltes bzw. den Aspekten des Themas (eckige Kästen) deutlicher erkennbar als bei der Frame-Repräsentation. Auch das Zentrum des Sachverhaltskomplexes (der Kopf des Frames/das Textthema) kann durch diese Darstellung noch deutlicher herausgestellt werden, insofern es der Teil des Sachverhaltskomplexes ist, von dem die meisten Pfeile (Relationen) zu anderen Teilen gehen. Relationen zwischen Teilen des Sachverhaltskomplexes sind in Analyse 2 als Textrollen (siehe Kapitel 2.6.2.1, Fußnote 92) repräsentiert.<sup>169</sup>



Analyse 2: Zusammenhänge zwischen den thematisierten Teilen des Sachverhaltskomplexes in Beispiel 1

<sup>169</sup> Für die maschinelle Verarbeitung wäre eine noch stärker formalisierte Repräsentation nötig (wie z.B. bei dem semantischen Formalismus MultiNet, siehe Helbig 2006), bei der u.a. ein geschlossenes Inventar von Relationen (Slots) und Konzepten (Fillern) Verwendung findet und zwischen Klassen und Instanzen differenziert wird. Auch eine solche Repräsentation wäre aller-



Im Zentrum des thematisierten Sachverhaltskomplexes des zweiten Beispieltex-tes steht die Diskursentität „Hering“. Das Charakteristische dieses Textes ist, dass die thematisierte Sachverhaltsstruktur sehr eng an die konzeptuelle Ord-nung des Konzepts „Hering“ angelehnt ist: Im Text werden mehrere semantisch bzw. konzeptuell über- und untergeordnete Konzepte als Diskursentitäten in-stantiiert, über die etwas gesagt wird. Die Aspekte der zentralen Diskursentitä-ten werden dabei über die Matrix-Frames „Gesamtheit“ sowie „Gegenstand – Konkretum – Nat. Art. – Diskontinuativum“ initiiert. Die diese Matrix-Frames konstituierenden Slots werden im Text demnach mehrfach instantiiert, so dass die Frame-Repräsentation des in Beispiel 2 thematisierten Sachverhaltskomple-xes in mehrere Sub-Frames zu semantisch bzw. konzeptuell untergeordneten Gegenständen bzw. Diskursentitäten untergliedert werden kann (siehe Analyse 3).

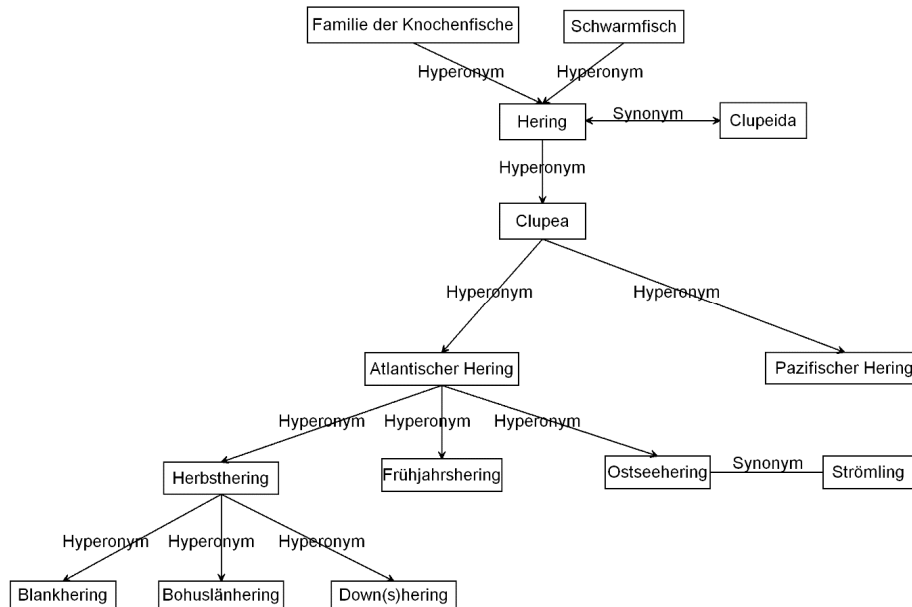
<b>Frame: „Hering“ (Instanz der Matrixframes „Gegenstand – Konkretum – Nat. Art. – Diskontinuativum“ sowie „Gesamtheit“)<sup>170</sup></b>		
<b>Frame</b>	<b>Slots</b>	<b>Filler</b>
Gesamtheit „He- ring“ (als Gat- tung)	Unter welchen Namen ist die Ge- samtheit bekannt?	Heringe, Clupeida
	Von welchem (allgemeineren) Typ ist die Gesamtheit?	Familie der Knochenfische
	Wie ist die Gesamtheit klassifiziert?	Ordnung der Heringsartigen
	Aus welchen Bestandteilen, Ele- menten besteht die Gesamtheit?	180 Arten Gattung Clupea Pazifischer Hering (Clupea pallasii) Atlantischer Hering (Clupea harengus)
Gegenstand „He- ring“ (als Nat. Art.)	Wo ist der Gegenstand typischer- weise zu finden?	in tropischen Meeresgebieten
	Welches Maß misst der Gegen- stand?	50 cm lang
	Welche Charakteristika hat der Gegenstand?	Schwarmfisch
	Welche weiteren Eigenschaften hat der Gegenstand?	Ernährt sich von Plankton
Gegenstand „At- lantischer He- ring“ (als Teil von „Hering“ als „Gattung“)	Welche Formeigenschaften hat der Gegenstand?	graugrüner Rücken, Seiten und Bauch silbrig
	Wo ist der Gegenstand typischer- weise zu finden?	Atlantik
	Welche Charakteristika hat der Gegenstand?	riesige Schwärme
	Welche Bedeutung hat der Gegens-	Wirtschaftlich wichtig

dings problemlos in eine Framerepräsentation zurück zu übersetzen, in der die Slots als Fragen formuliert sind, die sich wie die semantischen Kasusrollen als semantischen Relationen aus der Verbalenz ableiten lassen.

<sup>170</sup> Zu den Slots der Matrixframes „Gegenstand – Konkretum – Nat. Art – Diskontinuativum“ und „Gesamtheit“ siehe Konerding (1993, 303ff. und 360ff.).

	tand für den Menschen?	
	Was hat der Gegenstand als Teil? Wie ist der Gegenstand klassifiziert (Nachbarbegriffe)?	Herbsthering Frühjahrshering Ostseehering
	Auf welche Art und Weise entsteht der Gegenstand?	Weibchen legt rund 20 000 bis 70 000 Eier ab, die auf den Meeresgrund sinken und dort an Steinen, Algen oder Ähnlichem festkleben. Die schlüpfenden Larven steigen an die Oberfläche
	Welche typischen Existenzphasen durchläuft der Gegenstand?	Ei Larve Jugendentwicklung (im Küstenbe- reich) Junghering (Alter: 2-3 Jahr; Größe: 20cm; wandert von Küste ab) Geschlechtsreife (im Alter zwischen 3- 7 Jahren) Tod (nach ca. 20 Jahren)
Gegenstand „Herbsthering“ (als Teil der Klas- se/Gattung „At- lantischer He- ring“)	Welche Namen hat der Gegen- stand?	Blankhering Down(s)hering Bohuslänhering
	Unter welchen Bedingungen ent- steht der Gegenstand?	Blankhering: als Laich im Spätjahr vor der schottischen Nordseeküste Down(s)hering: als Laich in der südli- chen Nordsee Bohuslänhering: als Laich in schwedi- schen und norwegischen Gewässern
Gegenstand „Frühjahrshering“ (als Teil der Klas- se/Gattung „At- lantischer He- ring“)	Auf welche Art und Weise entsteht der Gegenstand?	Als Laich im Frühjahr in den norwegi- schen Fjorden
	Welche typischen Existenzphasen durchläuft der Gegenstand?	1. laicht im Frühjahr ab; 2. Nahrungs- wanderungen bis zur Ostküste Islands; 3. kehrt im Spätjahr zur norwegischen Küste zurück
Gegenstand „Ost- seehering“ (als Teil der Klas- se/Gattung „At- lantischer He- ring“)	Welche Namen hat der Gegen- stand?	Ostseehering Strömling
	Welches Maß misst der Gegen- stand?	20 cm lang
	Wo ist der Gegenstand typischer- weise zu finden?	Ostsee

Analyse 3: Thematisierter Sachverhaltskomplex in Beispiel 2



Analyse 4: Zusammenhänge zwischen den thematisierten Teilen des Sachverhaltskomplexes in Beispiel 2

Insofern die semantischen bzw. konzeptuellen Relationen zwischen Konzepten bzw. Diskursentitäten in Beispieltext 2 die Basis des thematisierten Sachverhaltskomplexes darstellen, bietet es sich für diesen Text an, die Zusammenhänge zwischen den zentralen Diskursentitäten nicht über Textrollen zu modellieren, sondern über semantische Relationen. Eine Repräsentation der für den thematisierten Sachverhaltskomplex zentralen Diskursentitäten, über die im Text etwas gesagt wird, könnte demnach als semantischer Baum visualisiert werden (siehe Analyse 4).

Die Auffassung von Frames als Menge der möglichen Teilthemen zu einem Sachverhaltskomplex (der den Frame selbst benennt und kategorisiert) bietet auch eine Erklärung dafür, warum in einem Text etwas thematisiert werden kann (z.B. „Anschlag“), indem über etwas anderes geredet wird (z.B. „Zugverspätung“, „Protest“, „Bundeskriminalamt“) (vgl. auch Fritz 1982). Diese implizite Themenbehandlung funktioniert dann, wenn in einem Text mehrere Aspekte des Themas behandelt werden und so der Frame als Ganzes aktiviert wird. Der Frame stellt damit die Einordnungsinstanz für die Themen des Textes dar – ein Rezipient *versteht*, dass über die Thematisierung eines Sachverhalts  $x$  ein Thema  $th$  behandelt wird, wenn er den Frame bzw. den zugrunde liegenden Sachverhaltskomplex rekonstruieren kann. „Aus dem vorgängigen Wissen über gewisse Zusammenhänge im Objektbereich rührt die Intuition des gemeinsamen Merkmals [des Zusammenhangs; M.B.] von Redeteilen her“ (Hellwig 1984a, 57). Der

framebasierte Zusammenhang zwischen den Aspekten eines Sachverhaltskomplexes (den Teilthemen bzw. Aspekten des Themas, wie sie im Frame modelliert sind) muss einem Rezipienten also bekannt sein, damit er die Aspekte des Themas auf ein implizites Textthema beziehen kann.

Neben dem durch den Frame bzw. das übergeordnete Thema gestifteten Zusammenhang zwischen den Aspekten des Themas (vertikaler thematischer Zusammenhang), können die Aspekte des Themas aber auch untereinander verknüpft sein, und es können im Text Themen behandelt werden, die nur 2. oder 3. Grades auf das Textthema bezogen sind, d.h. die wiederum als Aspekte eines Aspekts des Textthemas aufgefasst werden können (horizontaler thematischer Zusammenhang) (siehe Kapitel 3.3.2.1).

Der enge Zusammenhang von Wissensstruktur und thematisierter Sachverhaltsstruktur, d.h. der sprachlich realisierten thematischen Sachstruktur, wird besonders deutlich bei explikativen Texten: „Für wissensvermittelnde Texte ist (...) charakteristisch, dass ihnen normalerweise eine außersprachliche Sachverhaltsstruktur zugrunde liegt.“ (Jahr 2000, 388). Wie bereits in Kapitel 2.8.1 angesprochen, werden kognitive Strukturen nicht direkt in linguistischen Strukturen abgebildet, sie werden im Prozess der Versprachlichung stattdessen durch kommunikative und sprachliche Filter verändert. Dennoch gibt es – nicht nur in explikativen Texten – eine Schnittmenge zwischen Frame und thematisierter Sachverhaltsstruktur. Für das Verstehen des Textes sind dabei im Hinblick auf das Zusammenspiel von Frame und thematisierter Sachverhaltsstruktur vor allem zwei Faktoren relevant: 1. Der Grad der Übereinstimmung der aktivierten Frames (bzw. des Wissens über Sachverhalte und deren Zusammenhänge) von Textproduzent und -rezipient, und 2. der Grad der sprachlichen Explizitheit bei der Ausführung und Verknüpfung von Sachverhalten.

Zur maschinellen Ermittlung der im Text thematisierten Sachverhaltsstruktur kann auf diverse computerlinguistische Ansätze zurückgegriffen werden. Besonders relevant sind dabei die Ansätze zur Ermittlung der semantischen Verwandtschaft von Wörtern oder Dokumenten, insofern es hierbei um die Identifizierung der semantisch zentralen Wörter eines Textes oder einer Dokumentmenge und ihre Relation zu anderen semantisch zentralen Wörtern geht, d.h. um die Bestimmung der zentralen semantischen Knoten und Kanten. Diese Ansätze lassen sich in zwei Gruppen unterteilen: Flache oder statistische (d.h. in gewisser Weise wissensfreie) vs. wissensbasierte Verfahren (wobei in konkreten Anwendungen häufiger beide Ansätze kombiniert werden). Zu den statistischen Verfahren zur Ermittlung semantischer Verwandtschaften von Wörtern gehört das Vektorraum-Modell von Salton et al. (1996; 1999). Ziel dieses Verfahrens ist die Ermittlung von Ähnlichkeiten von Dokumenten, aber auch von Absätzen innerhalb eines Dokumentes. Bei diesem Verfahren wird auf Basis der mathematisch-statistischen Analyse eines großen domänengebundenen Korpus (bzw. einer Dokumentsammlung) eine Term/Dokument-Matrix erzeugt, in der die Anzahl

der Vorkommen jedes Terms<sup>171</sup> in jedem Dokument festgehalten wird. Jedes Dokument oder auch jeder Absatz kann demnach durch einen Vektor von gewichteten Termen charakterisiert werden, wobei durch die Gewichtung ausgedrückt wird, wie wichtig ein Term für ein Dokument oder einen Absatz ist. Die Gewichtung der Terme basiert auf der Häufigkeit, mit der ein Term in einem einzelnen Dokument im Verhältnis zu einer domänengebundenen Dokumentensammlung insgesamt vorkommt und wird häufig mit Hilfe der  $tf \cdot idf$ -Formel berechnet. Der  $tf \cdot idf$ -Wert gibt das Verhältnis zwischen der Termfrequenz ( $tf_i$ ) eines Terms  $t$  und seiner inversen Dokumentenfrequenz ( $idf_i$ ) an. Es wird davon ausgegangen, dass ein Term umso signifikanter und relevanter für einen Text  $d$  ist, je weniger er auch in anderen Texten vorkommt und je häufiger er im Vergleich zu anderen Texten in  $d$  vorkommt (vgl. Bieler und Dipper 2008; Neumann 2010, 588ff.).

Die Dokumente der Sammlung sowie die für ein Dokument relevanten Terme werden schließlich in einer Term/Dokument-Matrix erfasst. Sowohl Terme als auch Dokumente können dabei als Vektoren kodiert werden; ein Termvektor repräsentiert dabei das (gewichtete) Vorkommen eines Termes in allen Dokumenten; jedes Dokument entspricht einem Merkmalsvektor, der durch seine gewichteten Terme charakterisiert ist. Alle Vektoren (bzw. über die Vektoren: Dokumente) können auf dieser Basis paarweise auf ihre Ähnlichkeit hin überprüft werden. Das dargestellte Verfahren kann demnach einen Ähnlichkeitsgrad für Term- und Dokumentcluster ermitteln – worauf sich diese Ähnlichkeit, linguistisch gesehen, gründet, ob auf lexikalisch-semantische oder schemabasierte Relationen kann dagegen nicht angegeben werden.

Die Ähnlichkeiten zwischen Dokumenten lassen sich mittels einer Graphenstruktur abbilden, in der die Dokumente als Knoten repräsentiert sind und die Verbindungen (Kanten) zwischen den Knoten eine hinreichend große Ähnlichkeit der verknüpften Knoten charakterisieren. Durch die Beschriftung der Kanten kann darüber hinaus der Grad der Ähnlichkeit zweier Knoten angegeben werden (siehe Abbildung 8). Der Graph kann als Variante des in diesem Kapitel dargestellten Modells des thematisierten Sachverhaltskomplexes aufgefasst werden, insofern die Terme, über die im Text etwas gesagt wird, auf einer konzeptuellen bzw. semantischen Ebene in Bezug gesetzt werden und auf diese Weise eine Struktur aus Knoten und Kanten entsteht, wie sie auch für kognitive Schemata, Frames und Sachverhaltskomplexe grundlegend ist.

---

<sup>171</sup> Terme sind in diesem Kontext in der Regel Wörter (Wortformen), Phrasen (Mehrwort-Verbindungen) oder Stammformen von Wörtern (vgl. Bieler und Dipper 2008).

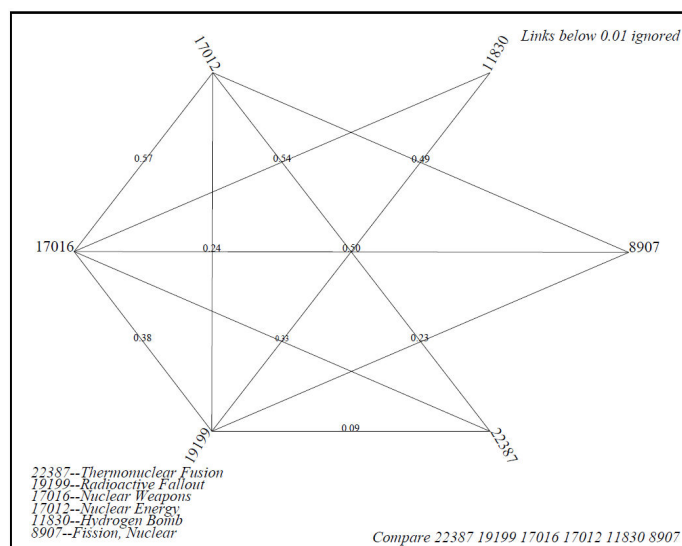


Abbildung 8: Gewichtete Relationen zwischen verschiedenen Texten auf Basis der berechneten statistischen Vektor-Ähnlichkeit (aus Salton et al. 1996, Figure 1)

Salton et al. (1996) beschreiben, wie das Verfahren auch für die Berechnung der thematischen Ähnlichkeit von Absätzen angewendet werden kann und wie diese dann dazu verwendet werden können, einerseits die thematische Organisation eines Textes (insbesondere thematische Verknüpfungen oder thematische Brüche zwischen Absätzen) zu analysieren und andererseits das zentrale Textthema zu identifizieren. Auf Basis der Berechnung der thematisch homogenen Textsegmente (Absätze, die eine hohe Ähnlichkeit im Hinblick auf ihre Vektoren aufweisen) können demnach zentrale und spezielle Textthemen ermittelt werden (Abbildung 9 zeigt eine auf Basis des Vektorraum-Modells konstruierte graphische Repräsentation der Themen eines Textes, über die auch das zentrale Thema eines einzelnen Textes identifizierbar wird).

Ein ähnliches, ebenfalls vektorbasiertes Verfahren zur Ermittlung von semantischen Ähnlichkeiten von Dokumenten, aber auch Termen oder Absätzen innerhalb eines Textes, das allerdings in seiner Komplexität noch über das Vektorraum-Modell von Salton et al. hinausgeht, ist die Latent Semantic Analysis (Landauer et al. 1998) oder auch das Latent Semantic Indexing (Deerwester et al. 1990). Diese Verfahren basieren auf der Annahme, dass es unter der Oberfläche von Texten bzw. Dokumenten eine „latente“ semantische Struktur gibt.<sup>172</sup>

<sup>172</sup> Das Verfahren kann nicht auf einzelne Texte angewendet werden, sondern operiert auf Dokumentmengen; Wörter und Dokumente werden dabei gemeinsam in einem Vektorraum darge-

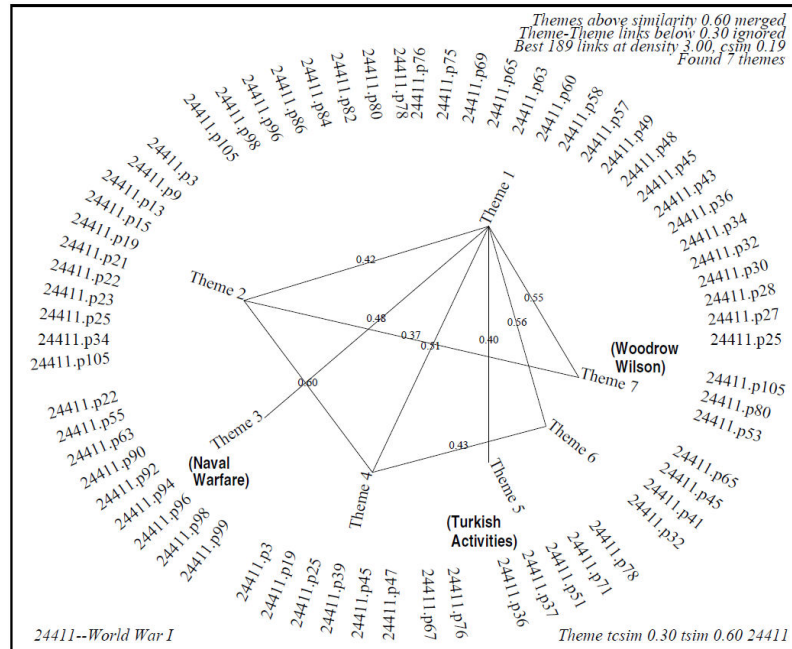


Abbildung 9: Graph der Themen eines Textes, ihrer Erstreckung (Absätze, in der Graphik repräsentiert in Form der den Themen am äußeren Rand zugeordneten Ziffern, z.B. 14411 (=Dokument) .p99 (=Absatz 99 des Dokuments)) und ihrer Ähnlichkeiten (aus Salton et al. 1996, Figure 8)

Ziel ist es, diese latente semantische Struktur offenzulegen, indem die assoziativen Beziehungen von Wörtern und Dokumenten ermittelt werden. Auf Basis eines Vektorraummodells wird zunächst eine Termfrequenz-Matrix erzeugt. Im Gegensatz zu dem Verfahren von Salton et al. wird diese Matrix anschließend weiterbearbeitet, indem sie durch Singulärwertzerlegung und Dimensionsreduktion kondensiert und generalisiert wird. Die Generalisierung führt dazu, dass semantisch ähnliche Wörter (Zeilen der Matrix) und Dokumente (Spalten der Matrix) als geometrische Nachbarn repräsentiert werden. In dem resultierenden Vektorraum liegen damit semantisch verwandte Wörter und Dokumente nahe beieinander; über den Kosinus von Zeilenvektoren in der Matrix kann im Folgenden, wie in dem Vektorraum-Modell von Salton et al., die Ähnlichkeit von Wörtern und Texten berechnet werden (vgl. Landauer et al. 1998; Mehler 2004, 101f.; Klein-Berning 2007, 220f.).

stellt, so dass ein direkter Ähnlichkeitsvergleich zwischen allen Objekten im Raum ermöglicht wird.

Neben den hier dargestellten Verfahren zur Ermittlung semantischer Ähnlichkeit existiert eine Vielzahl weiterer rein mathematischer Verfahren, auf die in diesem Kontext jedoch nicht weiter eingegangen werden soll (einen Überblick über quantitative Methoden in der Computerlinguistik, die zum großen Teil wissensfrei arbeiten bzw. ohne Wissensressourcen auskommen, gibt Mehler 2004).

Wissensbasierte Verfahren benötigen im Gegensatz zu den dargestellten rein mathematischen, wissensfreien Verfahren externe Wissensressourcen für die Ermittlung der lexikalisch-semantischen Verwandtschaft von Wörtern. Dabei kann zwischen klassischen Wissensressourcen wie Thesauri, Wortnetzen oder Ontologien und Wissensressourcen, die aus Korpora bzw. aus dem World Wide Web (speziell: Wikipedia) abgeleitet sind, unterschieden werden (einen Überblick über existierende Wissensressourcen gibt Kapitel 3.2.5). Verfahren, die klassische Wissensressourcen zur Ermittlung der semantischen Verwandtschaft von Wörtern verwenden (siehe z.B. Barzilay und Elhadad 1999; Silber und McCoy 2002; Teich und Fankhauser 2003), sind in der Regel von vorne herein in der Vollständigkeit ihrer Analysen limitiert, insofern einerseits nur ein begrenztes Inventar an Wörtern oder Termen in diesen Ressourcen enthalten ist (dies gilt insbesondere für das Deutsche) und andererseits kaum assoziative, schemabasierte Verknüpfungen zwischen Konzepten repräsentiert sind. Neuere Arbeiten zur maschinellen Ermittlung semantischer Verknüpfungen von Wörtern oder Konzepten versuchen deshalb, diese Limitierungen durch die Ersetzung oder Ergänzung klassischer Wissensressourcen durch neue Ressourcen wie Wikipedia zu überwinden (siehe z.B. Gabrilovich und Markovitch 2007; Zesch 2010). Bei einem Vergleich der Performanz von 18 verschiedenen Algorithmen zur Ermittlung semantischer Verwandtschaft und Ähnlichkeit stellen Cramer et al. (2011) allerdings fest, dass "the performance of all measures – as compared to human estimates – is still mediocre" (zu den in der Studie verglichenen Verfahren gehören auch die meisten der zuvor genannten Verfahren, sowohl wissensfreie als auch wissensbasierte).

### **3.2.2 Thema und Textmusterwissen: Globale und lokale thematische Muster**

Wie ein Frame sprachlich realisiert wird, hängt auch von der Intention des Produzenten, der Textsorte sowie der gewählten Vertextungsform ab (siehe Kapitel 2.8.1). Im Folgenden werde ich mich deshalb mit der Frage beschäftigen, wie sich das Handlungs- bzw. Interaktionswissen auf die Versprachlichung der Aspekte des Themas im Text auswirkt. Im Hinblick auf das Thematisieren als Selektieren (welche Aspekte des Themas können in einem Text thematisiert werden) und das Thematisieren als Strukturieren (wie werden die Aspekte des Themas im Text organisiert und verknüpft)<sup>173</sup> ist in diesem Kontext insbesondere der Bereich des Wissens über globale texttypspezifische Textmuster hervorzuheben. Texttypspezifische Muster, Schemata oder auch „Superstrukturen“ (vgl. van

---

<sup>173</sup> Zum Thematisieren als Selektieren und Strukturieren siehe Kapitel 3.4.1.



Dijk 1985, 69) prägen die globale Ordnung des Textes:<sup>174</sup>

Eine Superstruktur ist eine Art abstraktes Schema, das die globale Ordnung eines Textes festlegt und das aus einer Reihe von Kategorien [die als Thementypen interpretiert werden können, M.B.] besteht, deren Kombinationsmöglichkeiten auf konventionellen Regeln beruhen. (van Dijk 1980, 131)

Die Entwicklung von globalen Textmustern kann als Reflex auf die Bezugnahme auf eine bestimmte Klasse einer kommunikativen Situation aufgefasst werden: Textmuster können verstanden werden als konventionalisierte Routinen der Textproduktion, die sich zur Bezugnahme auf diese Situation als besonders geeignet erwiesen haben und die deshalb in das Interaktionswissen der Sprachgemeinschaft eingegangen sind. Globale Textmuster, die für das Thematisieren (Textproduktion) und das Verstehen des Themas (Textrezeption) eine besondere Rolle spielen, sind die inhaltsseitig geprägten texttypspezifischen Textmuster. Diese beziehen sich einerseits auf die globale Einteilung eines Textes in Inhaltszonen bzw. Texttypstrukturkategorien wie *Vorarbeiten*, *Methode*, *Schlussfolgerung* (Texttypstrukturkategorien wissenschaftlicher Zeitschriftenartikel) oder *Ereignis*, *Personen*, *Zeit*, *Ort* (Texttypstrukturkategorien von Zeitungsnachrichten) (siehe Kapitel 2.6.3.6) (*kategoriale thematische Muster*), andererseits auf die globale Entfaltung des Themas in einem bestimmten Texttyp – wobei in diesem Kontext die Texttypen der Narration, Deskription, Argumentation, Exposition, Explikation, Instruktion etc. gemeint sind (siehe Kapitel 2.3.3) (*sequenzielle thematische Muster*).<sup>175</sup>

Das Wissen über *kategoriale thematische Textmuster* prägt die Auswahl der Aspekte eines konkreten Themas, über die in einem Text etwas gesagt wird, bzw. die Erwartungen eines Rezipienten an die Aspekte des Themas, über die möglicherweise etwas gesagt werden wird. In einem Polizeibericht zu einem Autounfall wären typische Texttypstrukturkategorien beispielsweise: *Unfallhergang/Ereignis*, *beteiligte Personen (Agens, Patiens)*, *beteiligte Fahrzeuge (Instrument, affiziertes Objekt)*, *Zeugen (evtl. Experiens)*, *Ursachen (Causativ)*, *Folgen*, *Zeit (Temporativ)*, *Ort (Locativ)* etc.<sup>176</sup> In einer Narration zu einem Au-

<sup>174</sup> Zum Zusammenhang von Superstruktur-Kategorien und Themen schreibt van Dijk (1985, 89): "themes are the contents of news category slots". Superstruktur-Kategorien können demnach als abstrakte Thementypen aufgefasst werden, die im Text durch konkrete Inhalte (Themen) instanziiert werden.

<sup>175</sup> Neben den globalen texttypspezifischen Textmustern existieren noch eine Reihe weiterer globaler Textmuster, die ebenfalls Gegenstand des Textmusterwissens sind, z.B. wichtige Informationen am Textanfang vs. wichtige Informationen am Textende, Themen werden nacheinander eingeführt und entfaltet vs. alle Themen werden gemeinsam eingeführt und dann nacheinander entfaltet. Für Zeitungsnachrichten gilt beispielsweise, dass zunächst alle Themen präsentiert werden (vorzugsweise schon im einleitenden „Lead“-Satz), und zwar geordnet nach ihrer Wichtigkeit, und dass erst nach der Einführung aller Themen diese der Reihenfolge nach im weiteren Verlauf des Textes spezifiziert bzw. elaboriert werden. In anderen Texttypen ist die thematische Strukturierung dagegen genau andersherum: Zunächst wird nur ein Thema eingeführt und dieses dann ausführlicher behandelt. Erst wenn die Behandlung des ersten Themas abgeschlossen ist, wird das nächste Thema eingeführt und im Detail thematisiert, d.h. in Form der Behandlung seiner Aspekte besprochen (vgl. van Dijk 1985, 78ff.).

<sup>176</sup> In runden Klammern sind mit den genannten Textrollen vergleichbare Kasusrollen von von Polenz (1985, 170ff.) angegeben – allerdings existieren nicht zu allen Textrollen äquivalente Ka-

tounfall wären teilweise gleiche, teilweise aber auch andere Texttypstrukturkategorien erwartbar: *Unfallhergang, beteiligte Personen, Ursachen, Folgen, Zeit, Ort*, aber auch *Gefühle, Gedanken* usw. Auch Aspekte wie *äußere Umstände* („Es war so ein herrlicher, sonniger Frühlingstag“; „Ich hörte gerade Dave Matthews“) oder *persönliche Umstände* („Ich fühlte mich krank“) können in so einem Kontext relevant sein und Gegenstand der Thematisierung werden. In einem Polizeibericht sind die gleichen Umstände dagegen nicht relevant bzw. nur unter ganz bestimmten Voraussetzungen relevant – wenn nämlich die *äußeren Umstände* („Es herrschte Glatteis.“) oder die *persönlichen Umstände* („Der Unfallverursacher war alkoholisiert.“) *Ursachen* für den Unfall darstellen und damit zur Klärung der Schuldfrage beitragen. Eine Texttypstrukturkategorie *Umstand* ist demnach in einem Polizeibericht nur als Realisierung der Texttypstrukturkategorie *Ursache* denkbar, aber nicht als eigenständige Texttypstrukturkategorie.

Die für einen Text relevanten inhaltlichen Aspekte können einerseits beschrieben werden mit Hilfe von Texttypstrukturschemata, d.h. Konfigurationen von Texttypstrukturkategorien (siehe Kapitel 2.6.3.6), und andererseits mit Hilfe von Sachverhaltsframes, d.h. den Teilen eines Sachverhaltskomplexes (siehe Kapitel 2.7.3 und 3.2.1). Während die Texttypstrukturkategorien abstrakte Thementypen repräsentieren, die für einen bestimmten Texttyp konstitutiv sind und deshalb in der Regel in einem Text dieses Typs mehr oder weniger vollständig realisiert werden, stellen die Teile des Sachverhaltskomplexes texttypunabhängige Aspekte dar, die aufgrund der potentiell unendlichen Menge assoziativ verknüpfter Entitäten in einem konkreten Text nie in aller Gänze instanziiert werden können. Sowohl die in einem konkreten Text realisierten Teile des Sachverhaltskomplexes als auch die im Text instanziierten Texttypstrukturkategorien können in diesem Text als Textrollen (siehe Fußnote 92, S. 74) beschrieben werden. Texttypspezifische und sachspezifische Textrollen können dabei Überschneidungen aufweisen. Zeitungsnachrichten beispielsweise sind auf der Ebene der Texttypstruktur durch Textrollen wie *Ereignis, Personen, Zeit, Ort, Ursache, Mittel, Folgen* charakterisiert. Der im Text thematisierte Sachverhaltskomplex zu dem konkreten Ereignis (z.B. „Anschlag“, siehe Beispiel 1) kann ebenfalls Slots aufweisen, die durch Textrollen wie *Ursache, Mittel, Folgen* beschrieben werden können. Die beiden Typen von Textrollen sind in der Regel jedoch nie deckungsgleich, da sie orthogonal zueinander liegen – ein konkreter Text stellt eine Instanz eines bestimmten Texttyps dar; unabhängig davon kann in diesem Text auf einen Sachverhaltskomplex Bezug genommen werden.

Die *sequenziellen thematischen Textmuster* (siehe Kapitel 2.3.3) beziehen sich im Gegensatz zu den kategorialen thematischen Mustern nicht auf das „Was“, sondern das „Wie“ der thematischen Entwicklung im Text, d.h. die Formen der thematischen Entfaltung und Verknüpfung von Textteilen. Sequenzielle Formen thematischer Entfaltung werden bei Brinker (1997, 63) differenziert in deskriptive, explikative und argumentative thematische Muster. Argumentative themati-

---

susrollen. Neben Kasusrollen können darüber hinaus auch die sog. „Sekundärkonzepte“ bei de Beaugrande/Dressler (1981, 101f.) als Typen von Textrollen aufgefasst werden.

sche Muster beispielsweise sind orientiert an argumentativen Mustern der Reihung von Argumenten, Gegenargumenten, Stützungen, Schlussregeln, Modaloperatoren etc. Deskriptive thematische Muster lassen sich in drei Varianten untergliedern:

- a) Das Thema bezeichnet einen einmaligen Vorgang, z.B. ein historisches Ereignis (Bericht). Der thematische Textaufbau orientiert sich am zeitlichen Ablauf des berichteten Geschehens (narrativ-deskriptive thematische Muster).
- b) Das Thema bezeichnet einen als regelhaft (generalisierbar, wiederholbar) dargestellten Vorgang (Beschreibung). Das Thema wird in seine wesentlichen Teilvorgänge untergliedert (instruktiv-deskriptive thematische Muster).
- c) Das Thema bezeichnet ein Lebewesen oder einen Gegenstand (Beschreibung). Die Themenentfaltung vollzieht sich nach der Teil-Ganzes oder Enthaltenseins-Relation; weitere Möglichkeit: Beschreibung des zentralen Diskursreferenten.

Neben den beiden beschriebenen Typen globaler thematischer Textmuster existieren auch *lokale thematische Textmuster*, d.h. Möglichkeiten der thematischen Verknüpfung von einzelnen Äußerungen. Zu den bekanntesten lokalen thematischen Textmustern, die die thematische Verknüpfung von Äußerungen beschreiben, gehören die Muster thematischer Progression bzw. Organisation, wie sie in Kapitel 2.2.1.2 ausführlich beschrieben wurden, z.B. Themaufbau, Themaentwicklung (Themaableitung, Themakomposition), Themawechsel. Die lokalen thematischen Textmuster sind in der Regel ausdrucksseitig markiert (siehe auch Kapitel 3.3.1.3) – Themaufbau kann beispielsweise durch Eigennamen, definite Kennzeichnungen oder Anaphern realisiert werden oder durch sprachliche Mittel, die die Fortführung einer Sequenz signalisieren, z.B. „Erstens“, „Zweitens“, „Drittens“; Neuthematisierung oder Themawechsel dagegen können durch formelhafte Thematisierungsausdrücke signalisiert werden, z.B. „Zum Thema X“, „Was X betrifft“ (vgl. Hoffmann 1997, 513ff.).

Abschließend kann festgehalten werden, dass das gemeinsame Wissen über lokale und globale thematische Textmuster Voraussetzung sowohl für adäquates Thematisieren als auch das Verstehen der thematischen Zusammenhänge im Text ist (einen Überblick über maschinell verarbeitbare Wissensressourcen (u.a. zu Textmustern) gibt Kapitel 3.2.5).

### **3.2.3 Thema und Sprachwissen: sprachliche Möglichkeiten der Signalisierung, Etablierung und Strukturierung des Themas**

Auch das Sprachwissen spielt bei der adäquaten Etablierung und Strukturierung der Aspekte des Themas im Text eine wichtige Rolle (zum Sprachwissen allg., siehe Kapitel 2.7.1). Relevant ist dabei vor allem der Teil des Sprachwissens, der sich auf die sprachlichen Möglichkeiten der Signalisierung des Themas (Koreferenz, lexikalische Rekurrenz etc.), seiner Etablierung und Beibehaltung sowie der Signalisierung von Themenwechseln bezieht. Für den Produzenten ist dieses Wissen nötig, um einen thematisch kohärenten Text verfassen zu können, für

den Rezipienten ist dieses Wissen eine wichtige (aber nicht die einzige) Quelle für die Konstruktion von thematischer Kohärenz, insofern er auf Basis seines auf das sprachliche Material angewendeten Sprachwissens Erwartungen darüber ausbildet, worum es im Text geht bzw. wie der Text thematisch organisiert ist.

Das Sprachwissen interagiert dabei auch mit dem Sachwissen sowie dem Wissen über kommunikative Situationen: Wissen über den Zusammenhang von Sachverhalten und semantisch-lexikalisches Wissen über die Kookkurrenz von Ausdrücken hängen insofern zusammen, als beide durch typische Erfahrungen mit bzw. in kommunikativen Situationen geprägt sind: Zwei Sachverhalte stehen in einer Relation (AGENS der HANDLUNG „Anschlag“ ist der „Täter“), insofern sie Teil eines stereotypischen Erinnerungsmusters bzw. eines Situationsmodells sind; zwei Ausdrücke treten in einer Textumgebung gemeinsam auf („Jemand (ein ‚Täter‘) verübt einen Anschlag.“ → „verüben“, „Täter“, „Anschlag“), insofern sie zu einem gemeinsamen stereotypischen kommunikativen Gebrauchskontext gehören. Das Wortmaterial eines Textes kann demnach als komplexes indexikalisches Zeichen aufgefasst werden, das das Potential zur Aktivierung einer bestimmten kommunikativen Situation bzw. eines Frames besitzt (vgl. Kapitel 2.6.2.2 sowie Kapitel 2.7.2).

Das sprachliche Material kann nicht nur zum indexikalischen Zeichen für eine bestimmte kommunikative Situation werden, sondern auch zum indexikalischen Zeichen für ein Thema – das sprachliche Material kann als Menge der Reflexionen des Themas im Text betrachtet werden. Reflexionen des Themas im Text finden sich auf mehreren Ebenen der Textstruktur (siehe Kapitel 3.3.1) – dementsprechend ist auch das Sprachwissen, das für die Produktion und Rezeption dieser Reflexionen des Themas nötig ist, zu differenzieren in Wissen über die Funktion von grammatischen, lexikalischen, referenziellen, relationalen sprachlichen Aspekten und Aspekten der visuellen Textoberfläche für die Signalisierung des Themas. Im Bereich der Grammatik bzw. Informationsstruktur lässt sich hier z.B. Wissen über die Funktion der Satzgliedstellung (bzw. Thema-Rhema-Strukturen), der Satzfunktionen (grammatisches Subjekt als potentiellles Satzthema) sowie der grammatischen Theta-Rollen (AGENS als potentiellles Satzthema) unterscheiden. Daneben spielt das Wissen über die Funktion lexikalischer und referenzieller Zusammenhänge (koreferente bzw. rekurrente Diskursentitäten als potentielle Textthemen), Wissen über die Funktion von Kohärenzrelationen (zur Verknüpfung von Aspekten des Themas), Wissen über die Bedeutung und den Gebrauch von Wörtern im Kontext bestimmter Themenkomplexe und Wissen darüber, welche thematischen Zusammenhänge zwischen diesen Diskursentitäten bestehen, Wissen über die Bedeutung und Funktion bestimmter metasprachlicher Phrasen zur Signalisierung der thematischen Organisation sowie Wissen über die Funktion von Titeln sowie anderer visueller Einheiten der Textstruktur für die Signalisierung des Themas (zu existierenden maschinell verarbeitbaren Ressourcen im Bereich des Sprachwissens siehe Kapitel 3.2.5).

Diesen verschiedenen Formen des auf das Thema bezogenen Sprachwissens ist gemeinsam, dass es sich um Wissen über den Gebrauch der Sprache bei der thematischen Organisation von Texten handelt, das durch kommunikative Er-

fahrungen erworben wird und sich durch diese Erfahrungen weiterentwickelt und verändert. Voraussetzung für die Entwicklung eines solchen Wissens ist eine gewisse Regelmäßigkeit und Konventionalität in der Relation zwischen dem sprachlichem Material und der Signalisierung und Etablierung des Themas im Text. Erst durch diese Regelmäßigkeit kann aus den Routinen des Sprachgebrauchs ein äquivalentes thematisches Sprachgebrauchswissen entstehen. Grundlage für die Interpretation der Reflexionen des Themas im Text ist demnach, dass Kommunikanten Wissen darüber besitzen, mit welchen sprachlichen und außersprachlichen Merkmalen in der Regel auf ein bestimmtes Thema Bezug genommen wird, wie dieses signalisiert, etabliert oder gewechselt wird. Erst durch Konventionalität (Sprachgesellschaft) und Gebrauchswissen (Individuum) werden Kommunikanten in die Lage versetzt, das sprachliche Material als indexikalisches Zeichen für ein bestimmtes Thema zu interpretieren.<sup>177</sup>

In diesem Sinne handelt es sich bei der Fähigkeit, dieses thematische Sprachgebrauchswissen in einer aktuellen Situation anzuwenden, um eine „Common sense-Kompetenz“ im Sinne von Feilke (1993), eine Fähigkeit also, Sprache in typischer, d.h. der kommunikativen Erfahrung folgender Weise zu verwenden. Diese Kompetenz ermöglicht es beispielsweise Kommunikanten, idiomatisch geprägte Ausdrücke (wie z.B. „das junge Glück“, „das freudige Ereignis“, „der junge Vater“) zu verstehen und auf ein bestimmtes Gebrauchs- oder Ereignisschema zu beziehen, und zwar unabhängig von textgrammatischen, lexikalischen, referenziellen oder semantischen Merkmalen sowie unabhängig vom situativen Kontext (vgl. Feilke 1993, 15). Insofern in dieser Arbeit Schemata und Frames bzw. Sachverhaltskomplexe – verstanden als stereotypische Erinnerungsmuster, d.h. als Erfahrungsschemata und damit im weitesten Sinne auch als Ereignisschemata – zugleich als potentielle Themenkomplexe aufgefasst werden (siehe Kapitel 3.2.1), ist die Common sense-Kompetenz damit als die grundlegende Kompetenz für das Produzieren und Verstehen thematisch kohärenter Texte zu begreifen, die auf diversen Aspekten des Sprachverwendungswissens operiert: Sie ermöglicht es Kommunikanten, sprachliche Ausdrücke als indexikalische Zeichen zu verstehen, die auf einen bestimmten Themenkomplex bzw. ein typisiertes Themenwissen bezogen sind.

### 3.2.4 Thema und Wissen über die Kommunikationssituation

Wie bereits in Kapitel 2.7.1 angesprochen, ist das Wissen über die Kommunikationssituation hochgradig dynamisch, da es in der Kommunikationssituation selbst konstruiert und immer wieder aktualisiert und revidiert wird. Im Vergleich zu den anderen bisher angesprochenen Wissensbereichen ist das Wissen über die Kommunikationssituation demnach nicht unabhängig von der Kommunikationssituation selbst verfügbar. Es stellt eine nur temporär gebildete und fragile bzw. veränderliche Wissensquelle für das Produzieren und Rezipieren von Texten dar. Im Folgenden wird „Wissen über die Kommunikationssituation“

---

<sup>177</sup> Welche sprachlichen Strukturen als Reflexionen des Themas eine besondere Rolle spielen, wird in den Kapiteln 3.3.1.1 bis 3.3.1.4 sowie 3.3.2 in aller Ausführlichkeit dargestellt.

deshalb im Sinne von Annahmen oder bestätigten Annahmen über die Kommunikationssituation gebraucht.

Das Wissen über die Kommunikationssituation bzw. Annahmen über die Situation beeinflussen auch das Thematisieren (als Textproduktionshandlung) und das Verstehen der thematischen Organisation des Textes. Faktoren, die in diesem Zusammenhang eine Rolle spielen, sind z.B. Anlass und Umstände der Publikation, Länge des Textes, Art der medialen Vermittlung und Stellung des Textes im Diskurs. In einem längeren Text können beispielsweise mehr Aspekte des Themas zur Sprache kommen (Aspekt der Selektion von Themen) bzw. Aspekte können in einem größeren Umfang und mit mehr Gewicht behandelt werden (Aspekt der Strukturierung und Gewichtung von Themen). Ein Text, der in einem wissenschaftlichen Forschungsband veröffentlicht werden soll, wird thematisch anders organisiert sein als ein Text zum gleichen Thema, der in einer Zeitschrift für Laien platziert ist (in diesem Fall interagieren Kommunikationssituation und Handlungs-/Interaktionswissen). Ein Hypertext zu einem Thema, der auf mehrere Seiten verteilt sein kann, kann darüber hinaus thematisch anders organisiert sein als beispielsweise ein gedruckter Text – thematische Bezüge einer lokalen Seite können durch die Existenz von Links angedeutet werden, die globale thematische Organisation des gesamten Hypertextes durch Sitemaps oder kontextualisierte Navigationsmenüs. Texte, die nicht den Beginn eines Diskurses über ein Thema darstellen (z.B. ein Nachrichtendiskurs zu einem Ereignis, z.B. das Erdbeben auf Haiti), sondern Folgetexte sind, werden bestimmte, in vorhergehenden Texten bereits abgehandelte Aspekte des Themas weglassen oder in einem geringeren Umfang behandeln (z.B. Ort, Zeit, Opferzahlen) und dafür neue Aspekte einführen bzw. diesen Aspekten größeres Gewicht geben (z.B. Hilfsaktionen, Stellungnahmen von Politikern). Die Stellung des Textes im Diskurs ist damit entscheidend sowohl für die Selektion als auch die Gewichtung von Themen.

Neben diesen äußeren Faktoren der Kommunikationssituation beeinflussen darüber hinaus auch Annahmen (als Vorstufe von Wissen) über das Sach- und Sprachwissen, die Interessen, Fragestellungen sowie die sozialen Rollen der Kommunikanten die Produktion und Rezeption von Texten. Diese Annahmen wirken sich sowohl auf die Art und Weise der thematischen Organisation als auch die Wahl der Aspekte des Themas aus.

### **3.2.5 Wissensressourcen für die maschinelle Themenidentifikation**

Wissen wird von den Mitgliedern einer Kommunikationsgemeinschaft in kommunikativen Situationen erworben und durch Lernprozesse im Gedächtnis verankert. Es ist demnach zugleich individuelles als auch kollektives Wissen. Individuelles und kollektives Wissen können dabei divergieren, insofern das individuelle Wissen der Mitglieder der Kommunikationsgemeinschaft interindividuell divergieren kann. Dies hängt mit den unterschiedlichen kommunikativen Erfahrungen von Kommunikanten zusammen. Digitale Wissensressourcen (in der Computerlinguistik auch "Language Resources") stellen demnach Abstraktionen des gemeinsamen Wissens einer Kommunikationsgemeinschaft dar. Um diese

Aufgabe zu erfüllen, werden sie entweder korpusbasiert entwickelt oder von Gruppen linguistischer Experten einer Kommunikationsgemeinschaft. In beiden Varianten wird versucht, das gemeinsame Wissen der Kommunikationsgemeinschaft zu modellieren – dass dies grundsätzlich auf Grenzen stoßen muss, ist klar. Deutlich wird dies z.B. an der Frage nach der Dynamik von digitalen Wissensressourcen. Wie dargelegt, ist Wissen dynamisch (vgl. van Dijk 2008, 69) – es kann sich sowohl kurzfristig, innerhalb einer Kommunikationssituation, als auch mittel- und langfristig, im Laufe der Zeit, verändern – diese Änderungen können nicht nur das individuelle, sondern auch das kollektive Wissen betreffen: Durch neue Kommunikationsformen und -routinen (Sprach- und Textmusterwissen), durch neue technische Entwicklungen oder Erkenntnisse (Sachwissen) verändert sich auch das gemeinsame Wissen der Kommunikationsgemeinschaft. Um diese Dynamik von Wissen adäquat zu repräsentieren, müssten digitale Wissensressourcen demnach dynamisch erweitert und angepasst werden; teilweise ist dies auch der Fall, allerdings nicht immer. Digitale Wissensressourcen stellen damit „Hilfskonstrukte“ dar, über die die genannten Wissensbestände, die für die Identifikation des Themas von Belang sind, letztlich mehr oder weniger gut bzw. schlecht simuliert werden.

Für die maschinelle Themenidentifikation kann auf eine Reihe existierender digitaler Wissensressourcen zurückgegriffen werden. Allerdings sind nicht alle vier in den vorigen Kapiteln genannten zentralen Wissensbestände in gleichem Maße verfügbar (siehe Abbildung 10) Die am besten ausgebauten Wissensressourcen beziehen sich auf das syntaktische Wissen (in Form von Grammatiken) und das lexikalisch-semantische Wissen (d.h. einen Teil des Sprachwissens) sowie das konzeptuelle Wissen – wobei dieses nur zum Teil mit dem in Kapitel 3.2.1 dargestellten Sachwissen übereinstimmt. Sowohl Sprachgebrauchswissen als auch Interaktions- und Handlungswissen, wie z.B. das Textmusterwissen, ist dagegen nur sehr eingeschränkt (oder indirekt, über den in großen Korpora fixierten Sprachgebrauch) und nur für bestimmte Texttypen in Form von Schemata oder Lexika verfügbar. Wissen über die Kommunikationssituation schließlich hat einen gänzlich anderen Status als die übrigen Wissensbestände, da es ausschließlich dynamisch, während der Kommunikationssituation selbst, erzeugt wird und so noch schlechter als die anderen Wissensbestände in Form einer sta-

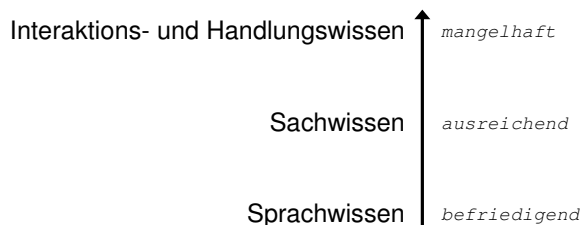


Abbildung 10: Verfügbarkeit der verschiedenen Wissensbestände/ Wissensressourcen für die maschinelle Themenidentifikation

tischen Wissensressource modellierbar ist. Im Folgenden soll ein Überblick über die für diese Wissensbereiche verfügbaren Wissensressourcen gegeben werden.

Im Hinblick auf das Wissen über den Zusammenhang von Konzepten bzw. Entitäten, das für die Identifikation der Teile des in einem konkreten Text thematisierten Sachverhaltskomplexes nötig ist, stehen sowohl konzeptuelle Wissensressourcen (siehe Carstensen 2010) als auch lexikalisch-semantiche Wissensressourcen zur Verfügung (siehe Kunze 2010). In den konzeptuellen Wissensressourcen werden Entitäten als Konzepte repräsentiert und Verbindungen zwischen diesen Konzepten als konzeptuelle Relationen. Hier wird demnach nicht mehr rein sprachliches Wissen repräsentiert, sondern in erster Linie nicht-sprachliches, konzeptuelles Wissen, d.h. Wissen darüber, welche Entitäten bzw. Konzepte in der Welt existieren und welche Zusammenhänge es zwischen diesen Entitäten gibt (vgl. Carstensen 2010, 532). Wissensressourcen, die konzeptuelles Wissen repräsentieren, d.h. Wissen über den Zusammenhang von Entitäten, werden in der Forschung zur Künstlichen Intelligenz heute meist in Form von Ontologien modelliert.<sup>178</sup> Ontologien stellen eine „formale, sprachunabhängige Beschreibung einer intersubjektiven Sicht der Welt“ dar (Carstensen 2010, 539). Aktuell stehen bereits mehrere große Ontologien wie SUMO (Suggested Upper Merged Ontology)<sup>179</sup> oder Cyc bzw. OpenCyc<sup>180</sup> sowie eine Vielzahl kleinerer domänenspezifischer Ontologien zur maschinellen Themenidentifikation zur Verfügung. Die meisten der verfügbaren Ontologien sind allerdings englischsprachig kodiert, so dass ihre Verwendung für die maschinelle Themenidentifikation in deutschsprachigen Texten die Abbildung deutscher Terme auf die englischen Bezeichner voraussetzt. Das Sachwissen, das für die Identifizierung der Zusammenhänge zwischen den Teilen des in einem konkreten Text thematisierten Sachverhaltskomplexes nötig ist, wird allerdings, in eingeschränkter Form, auch durch lexikalisch-semantiche Wissensressourcen modelliert.

Diese lexikalischen Ressourcen können differenziert werden in direkt verfügbare Wissensressourcen wie GermaNet<sup>181</sup> – für das Englische: Wordnet<sup>182</sup> – und indi-

---

<sup>178</sup> Eine Alternative zur Repräsentation von Konzepten und deren Relationen stellen Frames dar (siehe Kapitel 2.7.2) dar. Als elektronische Ressource steht für das Englische das Berkeley FrameNet zur Verfügung, das auf der Frame-Theorie von Charles Fillmore (Johnson und Fillmore 2000; Fillmore 2006 [1982]) basiert. Es umfasst über 150.000 annotierte Sätze, 11.600 lexikalische Einheiten, 960 semantische Frames (siehe <http://framenet.icsi.berkeley.edu>; Stand zuletzt abgerufen am 05.08.2010). Für das Deutsche existieren zwei framebasierte Lexika: das German FrameNet (<http://www.laits.utexas.edu/gframenet/>) sowie SALSA (<http://www.coli.uni-saarland.de/projects/salsa/>); zur Nutzung von FrameNet für das Deutsche siehe (Burchardt et al. 2009). Im Gegensatz zu Wortnetzen wie GermaNet werden Frames (wie das Berkeley FrameNet) in der Regel komplett korpusbasiert erstellt (vgl. Kunze 2010, 512); eine ausführliche Darstellung der Vorgehensweise bei der Extraktion und Annotation von Konzeptframes aus dem World Wide Web findet sich bei Lönneker (2003).

<sup>179</sup> SUMO: <http://ontologyportal.org>.

<sup>180</sup> Cyc bzw. OpenCyc enthält mehrere 100.000 Begriffe und mehrere Millionen Aussagen, die diese Begriffe verknüpfen. Nach eigenen Angaben handelt es sich bei Cyc bzw. OpenCyc um “the world’s largest and most complete general knowledge base and commonsense reasoning engine” (<http://www.cyc.com/cyc/openccyc>; zuletzt abgerufen am 05.08.2010).

<sup>181</sup> GermaNet 5.3: 76.709 Wörter, 61.659 Synsets (siehe <http://www.sfs.uni-tuebingen.de/lsd/>; Stand zuletzt abgerufen am 05.08.2010).



rekt verfügbare Wissensressourcen wie z.B. Wikipedia<sup>183</sup> oder Wiktionary<sup>184</sup>. GermaNet stellt ein lexikalisch-semantisches Wortnetz dar. Im Gegensatz zu den prinzipiell sprachunabhängigen Ontologien<sup>185</sup> bilden Wortnetze natürlichsprachliche Hierarchien ab – auch wenn sie, wie Ontologien, u.a. Konzepte und konzeptuelle Relationen modellieren. In GermaNet werden Wörter als Konzeptknoten repräsentiert; zentrales Repräsentationskonzept ist das Synset, das die Synonymmenge eines bestimmten Konzeptes (z.B. „fleißig, emsig, tüchtig“) umfasst. Die Konzeptknoten werden durch lexikalische und konzeptuelle Relationen verknüpft; lexikalische Relationen bestehen zwischen Wörtern eines Synsets (z.B. Synonymie, Antonymie), konzeptuelle Relationen zwischen Konzeptknoten, d.h. sie gelten für alle Mitglieder eines Synsets (Beispiel: Hyperonymie, Hyponymie, Meronymie, Implikation, Kausation) (vgl. Kunze 2002; Lemnitzer und Kunze 2007 [Kapitel 6: Lexikalische und ontologische Ressourcen]; Kunze 2010). Für domänenspezifische Texte, die über das allgemeinsprachliche Vokabular der deutschen Sprache hinausgehen, z.B. wissenschaftliche Artikel, ist leider oft zu konstatieren, dass längst nicht alle Lexeme in GermaNet enthalten sind – für das SemDok-Korpus deutschsprachiger wissenschaftlicher Linguistik-Artikel wurde z.B. ermittelt, dass 30.74% aller Nomen-Token und 59.17% aller Nomen-Types nicht in GermaNet 5.0 enthalten sind (vgl. Bärenfänger, Hilbert et al. 2008, 24).

Seit einigen Jahren wird verstärkt auf Wikipedia oder Wiktionary zurückgegriffen, wenn es um die computerlinguistische Modellierung von Wissen geht. Dabei werden verschiedene Bereiche von Wikipedia als Quelle verwendet: Artikel, Artikel-Links, Kategorien und Disambiguierungsseiten (Zesch 2010, 15f.). Diese Bereiche tragen verschiedene Arten lexikalisch-semantischer Informationen: Der erste Absatz eines Wikipedia-Artikels enthält eine Definition; im Volltext des Artikels werden verwandte Begriffe genannt; Links in Artikeln sowie zwischen Artikeln deuten in der Regel auf eine semantische Relation hin; Link-Label können als Quelle für Synonyme, Schreibvarianten oder verwandte Begriffe aufgefasst werden; das Wikipedia-Kategoriensystem stellt eine Art Thesaurus oder Taxonomie dar (zur Verwendung des Wikipedia-Kategoriensystems zur maschinellen Entwicklung einer Taxonomie siehe Ponzetto und Strube 2007b, a); Disambiguierungsseiten repräsentieren Polyseme.

In einer umfangreichen Evaluationsstudie vergleichen Zesch/Gurevych (2009) bzw. Zesch (2010) die Qualität (Präzision und Vollständigkeit) von Wikipedia als Wissensressource (“wisdom of crowds”) vs. Wordnet bzw. GermaNet (“wisdom of linguists”) im Hinblick auf Korrelationen mit manuellen Bewertungen der semantischen Verwandtschaft von Konzepten. Dabei kommen sie zu dem Ergebnis, dass “‘wisdom of crowds’ based knowledge sources are not generally supe-

<sup>182</sup> WordNet 3.0: 155.287 Wörter, 117.659 Synsets, 84.586 lexikalische Einheiten (siehe <http://wordnet.princeton.edu/>; Stand zuletzt abgerufen am 05.08.2010).

<sup>183</sup> Wikipedia: 1.102.706 Artikel in deutscher Sprache (siehe <http://www.wikipedia.org>; Stand zuletzt abgerufen am 05.08.2010).

<sup>184</sup> Wiktionary: 114.131 deutschsprachige Einträge (siehe <http://de.wiktionary.org>; Stand zuletzt abgerufen am 05.08.2010).

<sup>185</sup> Ontologien sind insofern sprachunabhängig, als sie nicht-sprachliches Wissen repräsentieren; dennoch sind Ontologien natürlich in einer bestimmten Sprache kodiert.

rior to the ‘wisdom of linguists’ based knowledge sources” (Zesch und Gurevych 2009, 55). Zesch (2010) zeigt allerdings, dass kollaborativ konstruierte semantische Ressourcen wie Wikipedia linguistisch entwickelte semantische Ressourcen im Kontext der maschinellen Sprachverarbeitung durchaus ersetzen können: Für das Englische, für das ein großes lexikalisch-semantisches Wortnetz wie WordNet existiert, sind die Ergebnisse von Wikipedia und WordNet demnach vergleichbar gut; für das Deutsche, mit einem wesentlich kleineren Wortnetz wie GermaNet, kann die Performanz sprachverarbeitender Verfahren durch die Nutzung von Wikipedia dagegen gesteigert werden (vgl. Zesch 2010, 75). Es erscheint deshalb sinnvoll, beide Typen von Wissensressourcen miteinander zu verbinden. Dies kann zum einen durch eine inhaltliche Anreicherung einer Wissensressource mit einer anderen Ressource geschehen. Zum anderen können Ressourcen auch in ein Austausch-Format konvertiert werden, welches die Entitäten und Relationen dieser Ressourcen in uniformer Weise repräsentiert (vgl. Zesch 2010, 21ff.).

Neben dem Sachwissen wurde in den vorigen Kapiteln u.a. auch die besondere Bedeutung des Textmusterwissens für die Themenidentifikation dargestellt, und zwar insbesondere des Wissens über lokale und globale thematische Textmuster (kategoriale und sequenzielle Muster). Beide Typen von Mustern können entweder über XML-Dokumentgrammatiken (sequenzielle thematische Muster) oder über OWL-Ontologien<sup>186</sup> (kategoriale thematische Muster) in maschinenlesbarer Form verfügbar gemacht werden (zu den Themen „XML“ und „Dokumentgrammatiken“ siehe Kapitel 3.3.3). Da es sich bei den globalen Textmustern um texttypspezifische Muster handelt, müssen entsprechende Wissensressourcen ebenfalls texttypspezifisch differenziert sein. Bislang ist die Verfügbarkeit dieser Arten von Ressourcen eher mangelhaft; auch wenn in computerlinguistischen Projekten möglicherweise texttypspezifische Ontologien oder Dokumentgrammatiken entwickelt werden, stehen diese deshalb noch lange nicht der Forschungsgemeinde zur Verfügung – sei es auch nur aufgrund infrastruktureller Hindernisse. Ein Beispiel für eine solche Ressource, die das Wissen über globale kategoriale thematische Textmuster repräsentiert, stellt das in Gießen entwickelte Texttypstrukturschema zum Texttyp „wissenschaftlicher Artikel“ dar (siehe Langer et al. 2004a). Dieses Schema besteht aus insgesamt 120 Texttypstrukturkategorien, aus dieser Version wurde durch die Abbildung von Subtypen auf übergeordnete Typen (auf Basis der Baumstruktur) ein reduziertes Schema mit 23 Texttypstrukturkategorien generiert (siehe Abbildung 11). In der formalen Repräsentation in Form eines XML Schemas sind die Texttypstrukturkategorien als Elemente und untergeordnete Texttypstrukturkategorien als eingeschachtelte XML-Elemente definiert. Für eine effiziente und weniger stark auf kanonischen Abfolgen basierende Annotierung von Textinstanzen wurde darüber hinaus mittels XSLT<sup>187</sup> eine flache Schema-Version abgeleitet (vgl. Langer et al. 2004a, b). In analoger Weise könnten auch für die globalen sequenziellen thematischen

---

<sup>186</sup> OWL: Web Ontology Language (W3C – World Wide Web Consortium 2004a).

<sup>187</sup> XSL: Extensible Stylesheet Language; XSLT: XSL Transformations (W3C – World Wide Web Consortium 2007c).

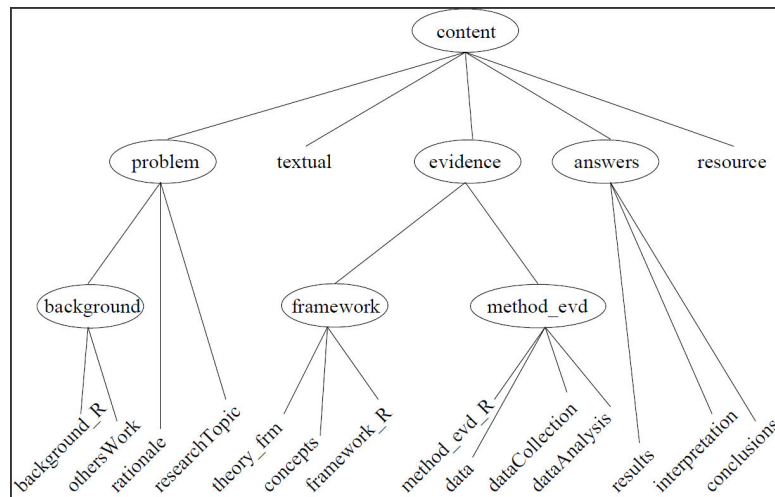


Abbildung 11: Texttypstrukturschema für wissenschaftliche Artikel

Textmuster und die lokalen thematischen Muster Dokumentgrammatiken bzw. XML Schema-basierte Repräsentationen entwickelt werden.

Im Hinblick auf die Bedeutung des Sprachwissens für das Thematisieren bzw. das Verstehen des Themas bzw. der thematischen Zusammenhänge sind in Kapitel 3.2.3 verschiedene Bereiche des Sprachwissens als besonders relevant identifiziert worden:

Bezugsebenen des Sprachwissens	Sprachliche und visuelle Mittel, die eine Tendenz haben, als Thema-Marker zu fungieren
Informationsstruktur	grammat. Subjekt Kasusrolle AGENS Satzerstposition Definitheit
Lexikalische Struktur	Gruppe zentraler Inhaltswörter zentrale lexikalische Kette mit den meisten Querbezügen zu anderen Ketten durch lexikalische Thematisierungsausdrücke eingeleitete Ausdrücke
Referenzielle Struktur	koreferenter Ausdruck salientester Diskursreferent
Visuelle/logische Struktur	Titel Überschriften erste Position im Satz, Absatz, Text
Relationale Struktur	Diskursmarker als Marker für thematische Verknüpfungen von Äußerungen

Tabelle 6: Wissen über die Möglichkeiten der Reflexion des Themas im Text

Das Wissen über potentielle sprachliche Thema-Marker kann in Form eines XML-basierten digitalen Thema-Marker-Lexikons repräsentiert werden. Eine derartige Wissensressource existiert zurzeit noch nicht. Dennoch kann auf Vorarbeiten aufgebaut werden. Von besonderer Relevanz sind dabei Arbeiten, die sich mit der Modellierung digitaler maschinenlesbarer Diskursmarkerlexika beschäftigen. Ein existierendes Lexikon für das Deutsche ist DiMLex (Stede und Umbach 1998). Ein weiteres, speziell auf RST zugeschnittenes Diskursmarkerlexikon stellt die Gießener DM-Liste dar, die allerdings auf das SemDok-Korpus deutschsprachiger linguistischer wissenschaftlicher Artikel zugeschnitten ist (vgl. Lobin et al. 2010, 113ff.). Dieses in XML kodierte Diskursmarkerlexikon umfasst zurzeit 103 lexikalische Diskursmarker, die für die maschinelle Identifikation und Zuweisung von Diskursrelationen im Sinne der RST genutzt werden können.

Diskursmarker stellen Elemente dar, die als indexikalische Zeichen (im Sinne von Antos 2009, 408) (vgl. Kapitel 2.6.2) für eine spezifische RST-Relation (oder eine geschlossene Menge von RST-Relationen) interpretiert werden können. Potentiell können neben den lexikalischen Diskursmarkern auch andere Elemente des Textes als Diskursmarker aufgefasst werden, z.B. syntaktische, logisch-visuelle Elemente sowie Elemente auf der Ebene der logischen Dokumentstruktur (insbesondere Elemente wie Überschriften, Listen, Tabellen, Grafiken) sowie der Texttypstruktur, z.B. besteht zwischen einem "Evidence"-Textabschnitt und einem "Problem"-Textabschnitt in der Regel die RST-Relation MEANS (vgl. Bärenfänger et al. 2006; Bärenfänger et al. 2010, 93). In der DM-Liste sind allerdings ausschließlich lexikalische Diskursmarker enthalten, während die abstrakteren Diskursmarker als Constraints in die Formulierung von sog. Reduce-Regeln, über die RST-Relationen zwischen Diskurssegmenten maschinell identifiziert werden, einfließen (zur Form und Funktion von Reduce-Regeln im Kontext des Diskursparsings siehe Lobin et al. 2011).

```

<dm id="c41" typ="lexical">
  <cue>
    <text>zwar</text>
    <lemma pos="ADV">zwar</lemma>
    <position>
      <vorfeld>+</vorfeld>
      <mittelfeld>+</mittelfeld>
    </position>
  </cue>
  <filter>
    <hypothese relname="Concession">
      <bezugseds>
        <word>
          <text>aber</text>
          <lemma pos="CC">aber</lemma>
        </word>
      </bezugseds>
      <bezugseds>
        <word>
          <text>jedoch</text>
          <lemma pos="ADV">jedoch</lemma>
        </word>
      </bezugseds>
    </hypothese>
  </filter>
  <rels default="Concession">
    <relation score="1" relname="Concession" skopus="eds+" typ="s"
      beds- richtung="r"/>
  </rels>
</dm>

```

Abbildung 12: Eintrag im Diskursmarkerlexikon zum Diskursmarker „zwar“

Abbildung 12 zeigt einen Eintrag des Diskursmarkerlexikons zu „zwar“, Beispiel-Annotation 1 einen Ausschnitt einer auf RST-basierenden Diskurs-Annotation (zum Thema „XML-Annotationen“ siehe Kapitel 3.3.3); Tabelle 7 die RST-Relationsdefinition zu CONCESSION.

```

<hypo relname="concession" id="i217">
<s><dm lemma="zwar" pos="ADV">Zwar</dm> wurde in der Fremdsprachenerwerbs-
forschung im Zusammenhang mit der noticing-Hypothese die Rolle der auf den
Input gerichteten Aufmerksamkeit untersucht. </s>
<n>Die Funktion der lernerseitigen Aufmerksamkeit für den Output im L2-
Erwerb blieb bisher <dm lemma="jedoch" pos="ADV">jedoch</dm>weitgehend
unberücksichtigt. </n>
</hypo>

```

Beispiel-Annotation 1: CONCESSION-Relation zwischen Satellit und Nukleus

relation name:	CONCESSION
constraints on N:	W (the writer) has positive regard for the situation presented in N
constraints on S:	W is not claiming that the situation presented in S doesn't hold
constraints on the N + S combination:	W acknowledges a potential or apparent incompatibility between the situations presented in N and S; W regards the situations presented in N and S as compatible; recognizing that the compatibility between the situations presented in N and S increases R's positive regard for the situation presented in N
the effect:	R's (the reader's) positive regard for the situation presented in N is increased
locus of the effect:	N and S
discourse markers:	in S: obwohl, obschon, obgleich, obzwar, zwar [although, though ...] in N: dennoch, doch, trotzdem [however, anyhow, nevertheless ...] in S: obwohl – in N: so in S: zwar – in N: jedoch

Tabelle 7: Definition der Relation Concession  
(Mann und Thompson 1988, 254f.)

In der Gießener DM-Liste ist – nach dem Vorbild von DiMLEX (s.o.) – nicht nur festgehalten, dass das Adverb „zwar“ Diskursmarker („Cue“) für eine CONCESSION-Relation sein kann, sondern auch, an welcher Position es vorkommen muss, um als ein solcher Diskursmarker fungieren zu können (Position: Vorfeld + , Mittelfeld +). Im Filter dieses Lexikoneintrags wird außerdem angegeben, welche Wörter in der Regel in Verbindung mit „zwar“ im Bezugssegment vorkommen, wenn tatsächlich eine CONCESSION-Relation vorliegt („aber“, „jedoch“, „doch“). Mit Bezugssegment ist dabei das Segment gemeint, das über CONCESSION mit dem Segment, in dem das Adverb „zwar“ vorkommt, angeschlossen wird. Im Lexikon ist darüber hinaus auch festgehalten, wie hoch die Wahrscheinlichkeit ist, dass „zwar“ eine CONCESSION-Relation signalisiert („score=1“ = hohe Wahrscheinlichkeit), welche Segment-Typen über CONCESSION verknüpft werden („eds“ = elementare Diskurssegmente), in welchem Diskurssegment (Nukleus oder Satellit) der Diskursmarker „zwar“ vorkommt („s“ = Satellit) und in welcher Richtung (links oder rechts) das Bezugssegment (in diesem Fall: der Nukleus) liegt. Beispiel-Annotation 1 repräsentiert ein Diskurssegmentpaar, das über eine hypotaktische Relation („hypo“) vom Typ CONCESSION verknüpft ist. In den beiden Teilen des Diskurssegmentpaars (dem Satelliten, „s“, und dem Nukleus, „n“) ist darüber hinaus jeweils ein Wort markiert („zwar“, „jedoch“), das als Diskursmarker („dm“) für die CONCESSION-Relation fungiert.

Für ein auf der theoretischen Thema-Konzeption der vorliegenden Arbeit basierendes System zur maschinellen Themenidentifikation wäre es sinnvoll, ein analoges Thema-Marker-Lexikon zu entwickeln; wobei hier hervorzuheben ist, dass nur ein kleiner Teil der Reflexionen des Themas im Text überhaupt in Form solcher expliziter Thema-Marker erfasst werden könnte; in erster Linie könnte dies für die formelhaften Thematisierungsausdrücke möglich sein (diese stellen allerdings auch besonders wichtige Mittel zur Signalisierung der thematischen Organisation des Textes dar). Ein weit größerer Teil des Wissens über die mehr oder weniger typischen sprachlichen und visuellen Mittel, die als Reflexionen des Themas im Text interpretiert werden können, müsste dagegen in Form von

Reduce-Rules (in einem Parsing-System) oder auch in Form von Merkmalen (in einem Klassifikationssystem) modelliert werden.

Ein Beispiel für einen möglichen Eintrag in einem Thema-Marker-Lexikon zeigt Abbildung 13, ein Beispiel für einen annotierten Textausschnitt Beispiel-Annotation 2 und ein Beispiel für eine Relationsdefinition Tabelle 8.

```
<tm id="a4" typ="lexical">
  <cue>
    <text>weiterhin</text>
    <lemma pos="ADV">weiterhin</lemma>
    <position>
      <vorfeld>+</vorfeld>
      <mittelfeld>+</mittelfeld>
    </position>
  </cue>
  <rels default="Themafortführung">
    <relation score="0.3" relname="Themareihung"/>
    <relation score="0.8" relname="Themafortführung"/>
  </rels>
</tm>
```

Abbildung 13: Eintrag im Thema-Markerlexikon zum Thema-Marker „weiterhin“

```
<rel name="Themareihung" id="i217">
<segment>Der Frühjahrshering laicht im Frühjahr in den norwegischen Fjorden ab und führt anschließend Nahrungswanderungen bis zur Ostküste Islands durch, um im Spätjahr zur norwegischen Küste zurückzukehren.</segment>
<segment><tm lemma="weiterhin" pos="ADV">Weiterhin unterscheidet man den bis 20 cm langen Strömling (Ostseehering), der in der Ostsee lebt.</tm>
</segment>
</rel>
```

Beispiel-Annotation 2: THEMAREIHUNG zwischen zwei Segmenten (Textausschnitt aus Beispieltext 2)

Relationstyp:	THEMAFORTFÜHRUNG
Relationsname:	THEMAREIHUNG
Bedingungen:	Der Textproduzent führt das Thema aus Segment m in Segment n fort; dies wird realisiert durch die Angabe eines Aspekts des Themas, der zur gleichen Klasse gehört wie der Aspekt des Themas, über den in Segment n etwas gesagt wird.
Thema-Marker:	in Segment m: „erstens“; in Segment n: „zweitens“ in Segment n: „weiterhin“, „daneben“, „auch“ etc.

Tabelle 8: Definition der Relation THEMAREIHUNG

Im Gegensatz zum Sach-, Textmuster- und Sprachwissen lässt sich das Wissen über die Kommunikationssituation nicht als mehr oder weniger statische Wissensressource modellieren. Hier müssten stattdessen auf Grundlage von maschinellen Klassifikationsverfahren adäquate Modellierungen und Repräsentationen dynamisch entwickelt werden. In diesem Zusammenhang kann z.B. auf Verfahren zur maschinellen Gewinnung von Metadaten Bezug genommen werden, durch die beispielsweise Angaben zum Publikationsmedium bzw. dem technischen Format, der Länge des Textes, dem Autor o.ä. extrahiert werden könnten. Zur Ermittlung des Autors wurden außerdem spezielle Verfahren zur Authorship Attribution entwickelt. Neben den Merkmalen der Kommunikationssituation, die über eine Analyse des Texts selbst identifiziert werden können, sollten auch Informationen zur Stellung eines Textes im Diskurskontext bzw. einer Textreihe zu demselben Thema ermittelt werden. Hierzu kann – zumindest für digitale Texte wie Online-Zeitungsnachrichten – auf Topic Tracking-Verfahren (siehe z.B. Allan et al. 1998; Cieri 2000; Makkonen et al. 2004; Stokes 2004) oder Meme Tracking-Verfahren (Leskovec et al. 2009) zurückgegriffen werden. Die Ermittlung der Kommunikationssituation ist darüber hinaus eng an die automatische Genre-Identifikation gekoppelt, so dass auch Verfahren aus diesem Kontext für die Modellierung von Wissen relevant sind (für einen Überblick über Verfahren zur maschinellen Genre-Identifikation siehe Rehm 2005, 632ff.).

Zur maschinellen Themenidentifikation werden bereits seit Mitte der 90er Jahre Wissensressourcen wie Wortnetze oder Thesauri eingesetzt; zu nennen sind hier vor allem die Arbeiten von Lin (1995; 1997; 1998) bzw. Lin/Hovy (1999; 2000), die WordNet zur Repräsentation und Generalisierung von Konzepten verwenden.<sup>188</sup> Auch Wissen über die Bedeutung von Positionen im Satz, Absatz bzw. Text sowie Wissen über die Bedeutung von Titeln wurden in verschiedenen Verfahren zur Themenidentifikation genutzt (siehe Kapitel 2.5). Allerdings gibt es keine Arbeiten, in denen alle in den vorigen Kapiteln genannten Wissensbestände in ein wissensbasiertes oder heuristisches System zur maschinellen Themenidentifikation integriert werden. Insbesondere der Einbezug von Thema-Marker-Lexika oder Texttypstrukturschemata, die bei der Identifizierung von Texttypstrukturkategorien bzw. unterspezifizierten Themen helfen könnten, stellt meiner Ansicht nach eine Chance zur Verbesserung der Qualität der maschinellen

---

<sup>188</sup> Wikipedia wird beispielsweise in den Arbeiten von Coursey/Mihalcea (2009) und Coursey/Mihalcea/Moen (2009) zur Themenidentifikation verwendet. Diese Arbeiten gehen allerdings über die Identifizierung von Themen in Texten hinaus. Sie entwickeln einen wissensbasierten Ansatz zur Identifizierung von Themen, die zu einem gegebenen Input-Dokument relevant sind. Zunächst wird aus Wikipedia ein Wissensgraph extrahiert, der auf den in Wikipedia vorkommenden enzyklopädischen Konzepten (Entitäten in Wikipedia-Artikeln und dem Wikipedia-Kategoriensystem) basiert. In einem zweiten Schritt werden für jedes Input-Dokument die zentralen enzyklopädischen Konzepte im Text ermittelt – eine solche Ermittlung der zentralen Konzepte eines Textes stellt allerdings eher eine reine Keyword-Extraktion dar, wie sie z.B. auch von Zesch (2010, 77ff.) beschrieben wird – und es werden Links zwischen diesen Konzepten und dem Wikipedia-Wissensgraph hergestellt. In einem dritten Schritt werden schließlich alle Knoten des Graphen im Hinblick auf ihre Relevanz für das Input-Dokument in gerichteter Form bewertet.



Themenidentifikation dar, und zwar insbesondere eine Verbesserung der Identifikation der thematischen Zusammenhänge zwischen den im Text thematisierten Aspekten.

### 3.3 Reflexionen und Realisationen des Themas im Text

Das Verstehen des Textes, und damit auch das Verstehen des Themas des Textes, basiert nicht nur auf dem Wissen über Sprache, Interaktionen und Sachverhalte, wie es teilweise in Schemata, Frames, Wortnetzen, Ontologien, Dokumentgrammatiken oder Lexika modelliert werden kann (top-down), sondern auch, bzw. vor allem, auf dem Text, d.h. dem sprachlichen Material, selbst (bottom-up). Texte sind, wie bereits dargestellt, auf mehreren Ebenen strukturiert: Auf der Ebene der Informationsstruktur, der lexikalischen, referenziellen und relationalen Textstruktur, der generischen Texttypstruktur sowie der visuellen und logischen Textstruktur. Das Thema des Textes wird dabei auf jeder dieser Strukturebenen in unterschiedlicher Weise realisiert bzw. reflektiert. Die sprachlichen Reflexionen des Themas dienen aus Sicht des Produzenten dazu, dem Rezipienten das Thema des Textes zu signalisieren – aus dieser Perspektive heraus können Texte als „geordnete Anweisungsmengen“ (Schermer 1984, 83), sprachliches Material als „mental processing instructions“ (Givón 1992, 139) oder „Signale“ (Schröder 2003, 103) aufgefasst werden. Aus der Perspektive des Rezipienten betrachtet, dienen die sprachlichen Reflexionen des Themas im Text als „Cues“ (Givón 1995, 91; Kintsch 1995) oder Indizien bzw. indexikalische Zeichen (im Sinne von Antos 2009, 408) (vgl. Kapitel 2.6.2), die das Verstehen des Textes bzw. der thematischen Zusammenhänge im Text leiten.

Im Text realisiert wird das Thema als Sachverhaltskomplex durch die Menge der über den Sachverhaltskomplex in einem Zusammenhang stehenden Diskursentitäten. Zusammenhänge zwischen Teilen des Sachverhaltskomplexes werden im Text dabei durch Diskursrelationen reflektiert. Während die Menge der thematisch zentralen Diskursentitäten als direkte Realisierungen bzw. Instanzierungen des Themas im Text betrachtet werden können, die auf den thematisierten Sachverhaltskomplex bezogen werden können, finden sich im Text darüber hinaus eine Reihe weiterer Reflexionen des Themas, die als Folge des Thematisierens als einer textorganisierenden Handlung aufgefasst werden können und die vom Rezipienten wiederum als Ausdruck der thematischen Textorganisation interpretiert und zum Verstehen des Themas als Sachverhaltskomplex genutzt werden können.

Erst durch die Menge der als Thema-Marker fungierenden sprachlichen und visuellen Merkmale des Textes, ihre „Verrechnung“ miteinander sowie mit den in Kapitel 3.2 behandelten Wissensbeständen kann das Thema als Sachverhaltskomplex für einen Rezipienten (in einem konstruktiven und dynamischen Verstehensprozess) sichtbar werden – auch wenn unbestritten ist, dass Rezipienten bereits auf Basis eines einzigen Merkmals (z.B. eines bestimmten Phraseologismus im Titel des Textes) Hypothesen über das Thema des Textes aufstellen; um aber ein vollständiges mentales Modell der thematischen Zusammenhänge des

Textes aufbauen zu können, bedarf es der „Verrechnung“ mehrerer Merkmale. Damit aber kann das Thema als emergente<sup>189</sup> Größe des Textes aufgefasst werden. Für die maschinelle Identifikation bietet es sich somit an, alle potentiellen indexikalischen Zeichen für das Thema zunächst getrennt voneinander zu erfassen und in einem darauf folgenden Schritt miteinander zu „verrechnen“ (siehe hierzu Kapitel 3.3.3).

Im Folgenden sollen die sprachlichen Mittel näher betrachtet werden, die eine herausragende Rolle als potentielle indexikalische Zeichen bzw. Indizien zur Identifikation des Themas spielen.

### 3.3.1 Thema als zentraler Gegenstand

Der Zusammenhang zwischen dem Thema als zentralem Gegenstand und dem sprachlichen Material, insbesondere den referierend gebrauchten Ausdrücken, ist ein indirekter bzw. vermittelter – die Menge der referierenden Ausdrücke im Text bezieht sich auf eine Vielzahl an Diskursentitäten, die alle als Gegenstände, über die im Text etwas gesagt wird, aufgefasst werden können. Durch den Rekurs auf kognitive Frames können diese Diskursentitäten in eine relationale Ordnung gebracht werden, in der die Relationen des Themas des Textes und seiner Aspekte widerspiegelt sind. Diese Ordnung von Diskursentitäten beruht darauf, dass Produzenten in dem in dieser Arbeit fokussierten Texttyp in der Regel einen Sachverhaltskomplex thematisieren, der einen Bezug zu einem oder mehreren stereotypischen Sachverhaltskomplexen aufweist, die in Form kognitiver Frames mental abgespeichert sind, d.h. als gemeinsames Wissen verfügbar sind (Produktionsperspektive). Rezipienten wiederum versuchen, die im Text behandelte Menge von Gegenständen vor dem Hintergrund der insbesondere durch Nomen evozierten<sup>190</sup> kognitiven Frames zu ordnen und als Ausdruck der Behandlung eines übergeordneten Textthemas zu verstehen (siehe Kapitel 2.8.1 und 2.7). Diskursentitäten können demnach auf Teile des Sachverhaltskomplexes oder Frames abgebildet bzw. in diesen Frame neu integriert werden. Die Auffassung des Themas als Sachverhaltskomplex, der auf einem abgespeicherten stereotypischen Sachverhaltskomplex basiert, ist demnach eine notwendige Voraussetzung dafür, dass die im Text behandelten Gegenstände als Aspekte *eines* Textthemas erkannt werden.

Der Zusammenhang von Thema, referierenden Ausdrücken und dem in einem Frame repräsentierten Sachverhaltskomplex ist allerdings nicht hinreichend: Erstens sind nicht alle Individuen der Menge der Diskursentitäten *Aspekte* des Textthemas. Wie bereits Bayer (1980, 218) konstatiert, sind „Äußerungen (...) möglicherweise nicht absolut, sondern nur bedingt über ein gegebenes Thema“. Zweitens sind Thema und kognitiver Sachverhaltskomplex nicht deckungsgleich (aus diesem Grund hatte ich zwischen stereotypischem und thematisiertem Sachverhaltskomplex eine Unterscheidung getroffen), insofern ein Text einen stereotypischen Sachverhaltskomplex erweitern, aktualisieren und umstrukturiert

---

<sup>189</sup> Zur Definition und Bedeutung von „Emergenz“ siehe Fußnote 87, S. 71.

<sup>190</sup> Frames werden Fillmore (2001) zufolge durch Verben und *Nomen* evoziert.

rieren kann. Diese Erweiterungen und Aktualisierungen müssen – so die Hypothese – allerdings sprachlich signalisiert werden. Das bedeutet, dass der Zusammenhang von Teilthema (bzw. Aspekt) und Textthema entweder auf einem stereotypischen Zusammenhang beruht, oder aber explizit sprachlich markiert ist. Diese sprachliche Explizierung kann z.B. durch lexikalische und referenzielle Kohäsion realisiert sein, aber auch durch metasprachliche Phrasen. Im Folgenden sollen die zentralen Markierungsmöglichkeiten des Themas als Gegenstand genauer betrachtet werden.

### 3.3.1.1 Grammatische und informationsstrukturelle Markierung des Themas

Bei der Darstellung der informationsstrukturellen Ansätze zum Thema wurde deutlich, dass das Thema im Satz vor allem auf Basis grammatischer Eigenschaften identifiziert (nicht definiert!) wird. Mittel, die in diversen Ansätzen zur Bestimmung des Satzthemas verwendet werden, sind Satzerstposition, grammatisches Subjekt, Kasusrolle AGENS, definite Nominalphrase, Pronomen, und Linksversetzung („Was X betrifft, ...“) (siehe Kapitel 2.1 sowie Givón 1983b, 347ff.; Stede 2007, 84ff.). Die ersten drei Merkmale sind allerdings weder hinreichende noch notwendige Bedingungen für Thematizität – das grammatische Subjekt ist nach Welke (2002, 126) allerdings in 60% der Fälle thematisch, so dass es hier zumindest eine Tendenz zu geben scheint, das, worüber im Satz geredet wird bzw. worauf sich die Teilhandlung bezieht, auch grammatisch als Subjekt zu realisieren. Auch die Merkmale Definitheit oder Pronominalität stellen keine eindeutigen Kriterien zur Bestimmung des Themas dar. Aufgrund ihres Status als zentrale Gegenstände des Textes, die auch auf der sprachlichen Oberfläche rekurrent sind und so den Status „+bekannt“ aufweisen, ist es allerdings plausibel, dass Themen eher als definite NPs oder PPs oder Pronomen realisiert sind, seltener als indefinite Phrasen (Ausnahme: wenn ein Thema neu eingeführt wird).<sup>191</sup> Linksversetzung ist im Gegensatz zu den bisher genannten grammatischen Merkmalen tatsächlich als hinreichende Bedingung für Thematizität aufzufassen, nicht jedoch als notwendige Bedingung – allerdings kommt diese syntaktische Konstruktion eher selten vor, so dass für die Mehrzahl der Sätze keine zuverlässigen grammatischen Markierungen vorhanden sind, sondern lediglich Indizien zum Verstehen des Themas zur Verfügung stehen.

Neben den genannten grammatischen Markierungen wird das Thema in informationsstrukturellen Ansätzen darüber hinaus mit Merkmalen wie „Bekanntheit“ und „Aboutness“ assoziiert. Wie in den Kapiteln 2.2.1.1 und 2.4.1 herausgearbeitet wurde, handelt es sich bei beiden um sehr vage Analyse Kriterien, deren Definition oder auch nur Beschreibung schwierig ist, da es hierzu verschiedene, sich teilweise widersprechende Positionen gibt. Trotz dieser Schwierigkeit ist das Kriterium der „Bekanntheit“ zentral für die Markierung des Themas und sollte bei der Analyse des Themas deshalb unbedingt berücksichtigt werden (siehe

---

<sup>191</sup> Gegen eine solche Auffassung sprechen sich allerdings Reinhart (1982, 12), Lambrecht (1994, 105) und Lernerz (1977, 52) aus, siehe auch Kapitel 2.4.1.

meine Anmerkungen am Ende von Kapitel 2.4.1). Da das kognitive oder kommunikativ-pragmatische Konzept „Bekanntheit“ sprachlich allerdings mittelbar sichtbar wird durch Definitheit, Koreferenz, insbesondere Pronominalität, sowie lexikalische Rekurrenz, ist es meiner Ansicht nach nicht nötig, bei der informationsstrukturellen Analyse zusätzlich nach einem Merkmal „+bekannt“ zu fahnden.

Auch das Kriterium der „Aboutness“ ist sicherlich von zentraler Bedeutung bei der Bestimmung des Themas: Ein Text bzw. eine sprachliche Einheit ist „über ein Thema“ bzw. „zum Thema X“, es wird etwas „über ein Thema gesagt“ (siehe auch Kapitel 1.1.2 zu zentralen Verwendungsweisen des Wortes Thema). Aber auch für dieses Merkmal gilt, dass es vage ist und schwer zu operationalisieren. In dieser Arbeit werden aus diesen Gründen nur die folgenden maschinell ermittelbaren grammatischen bzw. informationsstrukturellen Merkmale als Marker für Thematizität verwendet: Nominale oder pronominale Ausdrücke in Satzerstposition<sup>192</sup> – hier verstanden als Vorfeld oder Vor-Vorfeld bzw. Linksherausstellung –, Ausdrücke in der Funktion des grammatischen Subjekts, Ausdrücke in der Kasusrolle AGENS. Pronominalität wird dagegen auf der Ebene der Koreferenz behandelt.

Im Hinblick auf die maschinelle Identifizierung der genannten Merkmale kann auf eine Vielzahl computerlinguistischer Tokenisierungs-, Tagging- und Parsing-Verfahren zurückgegriffen werden. Insofern nur nominale oder pronominale Ausdrücke als informationsstrukturelle und grammatische Thema-Marker Berücksichtigung finden sollen, müssen Texte bzw. Sätze zunächst im Hinblick auf ihre morphosyntaktischen Kategorien hin analysiert werden (Part-of-Speech Tagging/POS-Tagging, darüber hinaus evtl. Parsing), was wiederum – als Vorverarbeitungsschritt – einer Segmentierung und Tokenisierung des Textes bedarf. Hierbei wird der Text in seine Bestandteile, d.h. in einzelne Worteinheiten und Sätze zerlegt (auch eine Segmentierung in größere Einheiten der Textstruktur wie Kapitel, Überschrift, Fußnote, Absätze etc. ist möglich und z.B. für die Ermittlung logischer und visueller Einheiten der Textstruktur von Interesse, siehe Kapitel 3.3.1.4). Innerhalb von Sätzen werden Wörter, Zahlen, Satzzeichen, Anführungsstriche, Klammern u.ä. als Token aufgefasst. Dieser Verarbeitungsschritt ist relativ unproblematisch. Schwierigkeiten, welche Einheiten als ein Token oder zwei Token gelten sollen, sind lediglich für eine eingeschränkte Liste von Fällen wie z.B. kontrahierte Formen (z.B. „am“), Ordinalzahlen („1.“), Mehrwortlexeme (z.B. „New York“) oder idiomatische Phrasen (z.B. „für alle Fälle Fitz“) zu erwarten (vgl. Lemnitzer und Zinsmeister 2006, 64ff.). Beim an die Tokenisierung anschließenden POS-Tagging werden den Token des Textes Wortarten-Informationen zugewiesen. Jedem Wort wird dabei ein Label oder Tag aus einem spezifizierten Tagset zugewiesen. Das gebräuchlichste Tagset für das Deutsche ist das Stuttgart-Tübingen Tagset (STTS), das 54 Wortartenlabel umfasst, wobei die größte Gruppe die Pronomen ausmachen.<sup>193</sup> Tagger (insbe-

<sup>192</sup> Die Satzerstposition kann auch als visuelle oder logisch-strukturelle Position beschrieben werden (siehe hierzu auch Kapitel 3.3.1.4).

<sup>193</sup> STTS (Stuttgart-Tübingen Tagset): <http://www.sfs.uni-tuebingen.de/Elwis/stts/stts.html>.

sondere die am weitesten verbreiteten Hidden Markov Model Tagger/HMM Tagger) sind mittlerweile sehr weit entwickelt und erreichen eine hohe Präzision, so dass selbst die semantische Disambiguierung kein großes Problem mehr darstellt und beim POS-Tagging eine Genauigkeit von 95% bis 98% erzielt werden kann (vgl. Schmid 2008, 547). Als Tagger für das Deutsche stehen beispielsweise der TreeTagger (probabilistischer Parser, basierend auf Entscheidungsbäumen), der TnT-Tagger (stochastischer Tagger, basierend auf einem Markov-Modell) und der Brill-Tagger (transformationsbasierter Tagger) zur Verfügung, die alle eine durchschnittliche Genauigkeit von mind. 96% erreichen (vgl. Hagenbruch 2010, 274f.).

Auch das Parsing syntaktischer Satzstrukturen und Satzfunktionen stellt einen Kernbereich des Natural Language Processing dar (zu den Grundlagen des Parsings siehe Langer 2001). Für den Zweck der Identifizierung grammatischer und informationsstruktureller Thema-Marker ist es dabei unerheblich, auf welcher grammatischen Theorie ein Parser basiert, so dass hier sowohl dependenzielle als auch konstituentielle Parser verwendet werden können. Zu den bekanntesten Parsern für das Deutsche gehören der Berkeley Parser, Stanford Parser und der BitPar Parser, wobei nur letztgenannter auch syntaktische Funktionen ausgibt (zu den Besonderheiten beim konstituentiellen vs. dependenziellen Parsing deutscher Sätze siehe Kübler 2008). Zur Ermittlung des Elements an Satzerstposition bzw. im Vorfeld des Satzes kann auf topologische Parser zurückgegriffen werden (siehe Becker und Frank 2002; Cheung und Penn 2009). Diese Parser erzielen – je nach Parametern und je nach topologischem Feld – eine Präzision und Vollständigkeit von 90% bis 99%, wobei die besten Ergebnisse für die linke Satzklammer erreicht werden; für das Vorfeld, das für die Ermittlung grammatischer und informationsstruktureller Thema-Marker von besonderem Interesse ist, werden mit einem  $F_1$ -Maß<sup>194</sup> von 97.2% (Berkeley Parser) ebenfalls gute Ergebnisse erzielt. Beispiel-Annotation 3 zeigt einen Ausschnitt des (transformierten) Outputs des kommerziellen Syntaxparsers „Machine Syntax“ der Firma Connexor Oy für Beispieltext 1.

Kasusrollen (wie AGENS) können mit Hilfe flacher semantischer Parsingverfahren (shallow semantic parsing) automatisch ermittelt werden. Dabei werden die Argumente des Prädikats über ein Klassifikationsverfahren automatisch mit Labels semantischer Kasusrollen versehen (für ein modulares Verfahren, bei dem existierende NLP Verfahren und Tools in einen Black Box-Ansatz integriert werden, siehe Erk und Padó 2006; als Beispiel für ein unüberwachtes, probabilistisches Verfahren, siehe Lang und Lapata 2010). Insbesondere die Präzision dieser Verfahren ist relativ gut (0.76 beim semantischen Parser von Erk/Padó), während die Vollständigkeit – insbesondere für das Deutsche – leider relativ schlecht ist (0.49 bei Erk/Padó), was auch damit zusammenhängt, dass es im Hinblick auf semantische Rollen keine ausreichend großen Trainingsdaten (annotierte Korpora) für das Deutsche gibt.

---

<sup>194</sup> Beim  $F_1$ -Maß werden die Maße für Präzision und Vollständigkeit in einem einzigen Maß kombiniert (vgl. Dörre et al. 2001, 434); beim hier genannten  $F_1$ -Maß sind die Maße für Präzision und Recall gleichgewichtig.

Auch für die automatische Identifikation der Informationsstruktur im Sinne der Topik-Fokus- und/oder Thema-Rhema-Struktur (siehe Kapitel 2.6.3.1) kann auf computerlinguistische Arbeiten und Anwendungen zurückgegriffen werden, sowohl für die manuelle als auch die automatische Analyse der Informationsstruktur. Dabei gibt es unterschiedliche Auffassungen darüber, aus welchen Einheiten die Informationsstruktur aufgebaut ist, so dass hieraus auch unterschiedliche Annotationsschemata und Herangehensweisen resultieren: In einigen Ansätzen werden traditionell Topik-Fokus und/oder Thema-Rhema analysiert, in anderen Ansätzen Typen von Diskursreferenten, die bezüglich ihrer informationsstrukturellen Eigenschaften unterschieden werden; teilweise werden beide Ansätze miteinander kombiniert (siehe Baumann et al. 2004; Postolache et al. 2005; Dipper, Götze und Skopeteas 2007). Arbeiten wie die von Postolache et al. zeigen, dass die Kategorien Topik und Fokus in einem maschinellen Lernverfahren relativ gut gelernt werden können, so dass diese Kategorien im von Postolache et al. verwendeten Korpus mit einer Akkuratheit von über 90% identifiziert werden konnten.

```

<sentence id="w31">
  <token id="w32">
    <text>Gewalt</text>
    <lemma>gewalt</lemma>
    <depend head="w35">subj</depend>
    <tags>
      <syntax>@NH</syntax>
      <morpho>N FEM SG NOM</morpho>
    </tags>
  </token>
  <token id="w33">
    <text>gegen</text>
    <lemma>gegen</lemma>
    <depend head="w34">pm</depend>
    <tags>
      <syntax>@PREMARK</syntax>
      <morpho>PREP</morpho>
    </tags>
  </token>
  <token id="w34">
    <text>Castor</text>
    <lemma>castor</lemma>
    <depend head="w32">mod</depend>
    <tags>
      <syntax>@NH</syntax>
      <morpho>N MSC SG NOM</morpho>
    </tags>
  </token>
  <token id="w35">
    <text>nimmt</text>
    <lemma>nehmen</lemma>
    <depend head="w31">main</depend>
    <tags>
      <syntax>@MAIN</syntax>
      <morpho>V IND PRES SG P3</morpho>
    </tags>
  </token>
  <token id="w36">
    <text>zu</text>
    <lemma>zu</lemma>
    <depend head="w35">phr</depend>
    <tags>
      <syntax>@ADVL</syntax>
      <morpho>VPART</morpho>
    </tags>
  </token>
</sentence>

```

### Beispiel-Annotation 3: POS-Annotation und syntaktische Annotation eines Ausschnitts von Beispieltext 1

Noch einmal festzuhalten ist, dass die genannten informationsstrukturellen und grammatischen Thema-Marker nur als bedingte Marker aufgefasst werden, d.h. weder als notwendige noch hinreichende Kriterien für Thematizität. Eine

sprachliche Einheit, bei der einer der genannten Marker für Thematizität zutrifft, hat demnach nur eine bestimmte statistische Tendenz, tatsächlich einen Aspekt des Themas zu realisieren bzw. reflektieren. Im Hinblick auf eine statistische Gewichtung und Hierarchisierung von sprachlichen Einheiten des Textes auf Basis der quantitativen Verteilung von Thema-Markern auf den diversen Ebenen der Textstruktur (repräsentiert als Multi-Ebenen-Annotation) ist es dennoch von sowohl theoretischer als auch praktischer Relevanz, die genannten Merkmale in die – auf der Auswertung der Multi-Ebenen-Annotationen basierenden – Identifizierung der Aspekte des Themas einzubeziehen.

Im Folgenden sollen für zwei Texte beispielhaft die grammatischen bzw. informationsstrukturellen Reflexionen des Themas ermittelt werden. Aufgrund der teilweise sehr aufwändigen Adaption existierender computerlinguistischer Verfahren und der bloß illustrierenden Funktion der Beispielanalysen werden diese (und sämtliche weiteren Analysen in Kapitel 4) manuell durchgeführt. Wie gezeigt wurde, stehen aber für alle Merkmale auf dieser Ebene auch Verfahren zur maschinellen Analyse zur Verfügung.

Satz	Satzerstposition	Subjekt	Kasusrolle Agens
(1)	Gewalt gegen Castor	Gewalt	
(2)	Anschläge auf Bahnstrecken	Grüne	Grüne
(3)			
(4)	Mit einer Welle von Bahnanschlägen	der gewaltsame Protest	der gewaltsame Protest
(5)	In Brandenburg, Berlin, Hessen, Niedersachsen und Schleswig-Holstein	Atomkraftgegner	Atomkraftgegner
(6)	Durch die insgesamt neun Anschläge	erheblicher Sachschaden	
(7)		niemand	
(8)	Zum Teil	Züge des Nah- und Fernverkehrs	
(9)	Der Verfassungsschutz	Der Verfassungsschutz	Der Verfassungsschutz
(10)	Das Bundeskriminalamt	Das Bundeskriminalamt	
(11)	Die Behörde in Wiesbaden	Die Behörde in Wiesbaden	

Analyse 5: Grammatische bzw. informationsstrukturelle Reflexionen des Themas in Beispiel 1

Analyse 5 ist für sich betrachtet wenig aussagekräftig; ihre Kategorien können aber als ein Klassifikationsmerkmal unter vielen in die maschinelle Identifikation des Themas und seiner Aspekte einfließen. Ob sich dadurch tatsächlich eine qualitative Verbesserung der Ergebnisse erzielen lässt, muss allerdings sorgfältig geprüft werden. Möglicherweise ist es sinnvoll, die extrahierten Ausdrücke dieser Ebene als Grundlage für eine sich anschließende semantisch-lexikalische Analyse zu nutzen – auf diese Weise könnten z.B. übergeordnete Hyperonyme ermittelt werden, über die mehrere nominale Ausdrücke zusammengefasst werden könnten (z.B. könnten „Verfassungsschutz“, „Bundeskriminalamt“ und „Be-



hörde in Wiesbaden“ zusammengefasst werden zu „Staatliche Behörden“). Insofern der Ausdruck „Anschläge“ („Anschläge“, „Bahnanschlänge“, „Anschläge auf Bahnstrecken“) quantitativ gesehen innerhalb der Kategorie „Satzersposition“ am Häufigsten vorkommt, könnte dieser Entität – zumindest vorläufig – der Status „+Thema“ zugeschrieben werden.

In Beispieltext 2 wird die Gattung „Hering“ über die Satzersposition sowie die Besetzung des grammatischen Subjekts als zentrale thematische Entität des Textausschnitts etabliert (siehe Analyse 6). Die Analyse von Beispieltext 1 zeigt, dass eine solch eindeutige informationsstrukturelle Gewichtung, durch die eine Entität als zentrales thematisches Satzelement hervorgehoben wird, allerdings keineswegs die Regel ist: Hier lässt sich im Gegensatz zu Beispiel 2 keine eindeutig zentrale thematische Entität auf Ebene der Informationsstruktur feststellen.

Satz	Satzersposition	Subjekt	Kasusrolle Agens
(1)	Heringe, Clupeidae, Familie der Knochenfische (Ordnung der Heringsartigen) mit rund 180 Arten		
(2)	Die bis 50 cm langen Heringe	Die bis 50 cm langen Heringe	
(3)	Die wirtschaftlich wichtigste Gattung Clupea	Die wirtschaftlich wichtigste Gattung Clupea	
(4)	Er [→ der Atlantische Hering]	Er	
(5)		der Herbsthering	der Herbsthering
(6)	Der Frühjahrshering	Der Frühjahrshering	Der Frühjahrshering
(7)		man	
(8)	Ein Weibchen des Atlantischen Herings	Ein Weibchen des Atlantischen Herings	Ein Weibchen des Atlantischen Herings
(9)	Die schlüpfenden Larven	Die schlüpfenden Larven	Die schlüpfenden Larven
(10)	sie [→ die Larven]	sie	
(11)		die mit zwei bis drei Jahren etwa 20 cm langen Jungheringe	die mit zwei bis drei Jahren etwa 20 cm langen Jungheringe
(12)	Die Geschlechtsreife		
(13)	die Lebensdauer	die Lebensdauer	
(14)	Einige Heringsarten	Einige Heringsarten	

Analyse 6: Grammatische bzw. informationsstrukturelle Reflexionen des Themas in Beispiel 2

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE text SYSTEM "file:/D:/Dissertation/Material-
Kapitel4/Schemata/mb-schemata/de_inf.dtd">
<text src="Heringe.txt">

<s><de_inf agens="null" definite="null" sentence_first="yes" sub-
ject="yes">Heringe, Clupeidae, Familie der Knochenfische (Ordnung der
Heringsartigen)</de_inf> mit rund 180 Arten, die meist in tropischen
Meeresgebieten leben.</s>

<s> <de_inf agens="no" definite="yes" sentence_first="yes" sub-
ject="yes">Die bis 50 cm langen Heringe</de_inf> sind überwiegend
Schwarmfische, die sich von Plankton ernähren.</s>
```

#### Beispiel-Annotation 4: Einfache informationsstrukturelle Annotation eines Ausschnitts von Beispieltext 2

Die in diesen Beispielen verwendeten informationsstrukturellen bzw. grammatischen Merkmale können, wie dargestellt, maschinell identifiziert und in einem maschinenlesbaren Format repräsentiert werden. Beispiel-Annotation 4 zeigt einen Ausschnitt von Beispieltext 2, der auf Basis einer flachen und sehr einfachen XML-basierten Dokumentgrammatik manuell annotiert wurde. Die Dokumentgrammatik beschränkt sich ausschließlich auf die genannten Merkmale und stellt keinesfalls eine Dokumentgrammatik für die gesamte Informationsstruktur dar, wie beispielsweise das Annotationsschema von Dipper et al. (2007), das eine Vielzahl von sprachlichen Phänomenen auf der Ebene der Informationsstruktur zu beschreiben und repräsentieren erlaubt.

#### 3.3.1.2 Koreferenz: Wiederaufnahme zentraler Diskursentitäten

Koreferenz kann als zentrales Mittel zur Bestimmung des Themas aufgefasst werden, insofern die über die koreferenten Ausdrücke des Textes referenzierten Diskursentitäten als zentrale Gegenstände des Textes etabliert werden, über die etwas gesagt wird. Referenzielle Kontinuität ist deshalb als Marker für thematische Kontinuität aufzufassen (siehe Kapitel 2.2.1.2 sowie Dressler 1973, 40ff.; Givón 1992, 15; Dik 1997, 436f.; Rothkegel 1997; Hoffmann 2000, 511; Stede 2007, 59). Die referenzielle Struktur stellt dabei auch ein verbindendes Element zwischen Satz- und Textthemen dar: Satzthemen zeichnen sich einerseits dadurch aus, dass sie als Mitteilungsgegenstand definiert sind, über den etwas gesagt wird (siehe Kapitel 2.2.1.1), andererseits dadurch, dass sie bereits bekannte Informationen repräsentieren (siehe Kapitel 2.4). Als bekannte Gegenstände stellen sie rekurrente Elemente dar, über die der Anschluss an den vorhergehenden Äußerungskontext erreicht wird. Textthemen wiederum können als Gegenstände beschrieben werden, über die im Text fortlaufend etwas gesagt wird – sie werden in Form von koreferenten Diskursentitäten wiederholt wiederaufgenommen (siehe Kapitel 2.2.1). Aufgrund der Bedeutung der koreferenten Diskursentitäten für die thematische Entwicklung kann „die Wiederaufnahmestruktur als Trägerstruktur für die thematischen Zusammenhänge des Textes“ inter-

pretiert werden (Brinker 2000, 168).

In der vorliegenden Arbeit wird ein enger Koreferenz-Begriff vorausgesetzt (siehe Kapitel 3.1.1.3), so dass sämtliche Formen der Kospezifikation (einschließlich indirekter Anaphorik) ausgeklammert werden. Grund hierfür ist, dass Fälle der Kospezifikation oder indirekten Anaphorik in der Regel durch semantische oder schema-basierte Relationen realisiert sind – diese Arten von Relationen werden allerdings bereits im Rahmen der lexikalischen sowie der kognitiven Ebene analysiert. An dieser Stelle soll dennoch festgehalten werden, dass sowohl koreferente als auch kospezifizierende Anaphern als Thema-Marker aufgefasst werden können. Während koreferente Anaphern in erster Linie der Etablierung eines Themas (als Diskursentität, über die fortlaufend etwas gesagt wird) dienen, können kospezifizierende Anaphern, insbesondere indirekte Anaphern, genutzt werden bei der Einführung eines Teilthemas zu einem gegebenen Thema (vgl. Dik 1997, 437).

Die thematische Wichtigkeit einer durch eine Anapher wiederaufgenommenen Diskursentität hängt davon ab, wieviele Koreferenz-Verbindungen sie insgesamt hat, welche Distanz sie zum vorherigen koreferenten Ausdruck hat, an welcher Position und in welcher grammatischen und semantischen Rolle sie im Satz steht, an welcher Stelle sie im Absatz vorkommt, in welcher syntaktischen und semantischen Form sie als Anapher realisiert ist usw. (vgl. z.B. Levy 1982, 303; Givón 1983a; Ariel 2001, 37). Einerseits hat eine Diskursentität, die wiederholt wiederaufgenommen wird, eine größere Wahrscheinlichkeit, vom Rezipienten als Thema interpretiert zu werden als eine Diskursentität, die nicht wiederaufgenommen wird. Ariel (2001, 34) weist allerdings darauf hin, dass man zwischen globalen und lokalen Themen unterscheiden muss: koreferente Ausdrücke, die einen großen Abstand aufweisen, sind demnach häufig Ausdruck für ein globales Thema. Dies kann damit begründet werden, dass globale Themen ein hohes Maß an (mentaler) Zugänglichkeit aufweisen. Goecke/Witt (2006) weisen in diesem Zusammenhang darauf hin, dass sich durch Korpusauswertungen belegen lässt, dass Diskursentitäten, die im Abstract eines wissenschaftlichen Artikels vorkommen, über den gesamten Text hinweg als Antezedenten zugänglich bleiben, während Diskursentitäten, die in einer Fußnote o.ä. eingeführt wurden, eine geringere Wahrscheinlichkeit haben, als Antezedent für eine Anapher im Haupttext zu fungieren.

In einem Text existieren Koreferenzrelationen nicht nur zwischen jeweils zwei Ausdrücken, sondern zwischen Mengen von Ausdrücken: Zentrale Diskursentitäten werden in einem Text wiederholt wiederaufgenommen, so dass sich Ketten von Anapher-Antezedent-Relationen ergeben („referenzielle Ketten“, vgl. Stede 2007, 56ff.). Dabei kann entweder davon ausgegangen werden, dass sich eine Anapher immer auf den im Text zuerst vorkommenden koreferenten Ausdruck bezieht oder auf den unmittelbar vorangehenden Ausdruck. Da sowohl der Abstand zwischen als auch die Form von Anapher und Antezedent relevant sind für die Dichte und Bedeutung einer referenziellen Kette, wird hier die letztgenannte Auffassung vertreten. Über die referenziellen Ketten lassen sich lokale und globale Diskursentitäten des Textes ermitteln: Diskursentitäten, die nur in einem begrenzten Abschnitt des Textes referenziert werden, stellen lokale Entitäten

dar; Diskursentitäten, die durchgehend wiederaufgenommen werden (auch mit einem größeren Abstand zwischen den einzelnen Wiederaufnahmen, s.o.), globale. Mit der als referenzielle Ketten repräsentierbaren durchgehenden Orientierung auf eine begrenzte Gruppe von globalen Entitäten ist dabei die Orientierung auf eine spezifische Gruppe zentraler Themen verbunden. Themazentrierung ist demnach sowohl Quelle (Produzentensicht) als auch Resultat (Rezipientensicht) der referenziellen Struktur.

Aus den Ketten lassen sich „Rückschlüsse auf die thematische Gliederung des Textes“ ziehen: „Die Positionen, an denen referenzielle Ketten beginnen und andere enden, deuten auf eine Themenverschiebung im Text hin.“ (Stede 2007, 59). Darüber hinaus können Querbezüge zwischen Ketten als Indiz für Relationen zwischen Teilthemen interpretiert werden. Dies gilt insbesondere für indirekte Anaphorik oder Kospezifikation. In dieser Arbeit werden deshalb referenzielle Ketten als Reflexionen des Themas interpretiert, wobei davon ausgegangen wird, dass die Mitglieder einer referenziellen Kette in einer Relation der Koreferenz (nicht Kospezifikation) stehen.

Im Folgenden werden die referenziellen Ketten in den beiden Beispieltexten 1 und 2 dargestellt. Schon an diesen beiden Beispielen lässt sich zeigen, dass sich Texte in der Anzahl und Dichte ihrer referenziellen Ketten enorm unterscheiden können. Es gibt sowohl Texte, mit sehr vielen referenziellen Ketten, als auch Texte, mit sehr wenigen. Beide Arten von Texten können aber in gleichem Maße als thematisch kohärent empfunden werden. Wie die Reflexionen des Themas auf der Ebene der Informationsstruktur stellen Reflexionen des Themas auf der Ebene der referenziellen Struktur demnach keine notwendigen und hinreichenden Thema-Marker dar.

- (a) Anschläge auf Bahnstrecken (S2)
- (b) eine Welle von Bahnanschlägen (S4) [realisiert als Paraphrase zu (a)]
- (c) die insgesamt neun Anschläge (S6) [realisiert als Identität zu (a)]
- (d) die Aktionen (S9) [realisiert als Hyperonymie zu (a), (b), (c)]
- (e) die bundesweiten Anschläge (S11) [realisiert als Identität zu (a) und (c)]
- (a) Atomtransport (S2) [Relation der Hyperonymie zu (a)]
- (b) Castor-Transport (S4) [Relation der Synonymie zu (b)]
- (a) Atomkraftgegner (S5)
- (b) gewalttätige Autonome (S9) [realisiert als Paraphrase zu (a)]
- (c) Einzeltäter (S10) [realisiert als Paraphrase zu (b)]

#### Analyse 7: Referenzielle Ketten in Beispieltext 1

Die zentrale referenzielle Kette stellt in diesem Text die Kette 1 dar, sowohl von ihrer Dichte (Kettenglieder finden sich in 5 von 11 Äußerungseinheiten) als auch ihrer Textabdeckung bzw. Länge (sie wird zunächst in Äußerungseinheit 2, zuletzt in Einheit 11 wiederaufgenommen). Die im Zentrum dieser Kette stehende Diskursentität „Anschläge“ bzw. „Bahnanschläge“ stellt demnach die zentrale, fortlaufend wiederaufgenommene Diskursentität des Textes dar. Sie erhält damit

den Status „+Thema“, d.h. sie wird als Ausdruck des Hauptthemas des Textes betrachtet. An Beispieltext 1 wird übrigens deutlich, dass eine Koreferenz-Relation zwar als lexikalisch-semantische Relation realisiert sein kann, im Gegensatz zu dieser allerdings an das Verstehen des Textes gebunden ist: „gewalttätige Autonome“ wird in Beispieltext 1 als Paraphrase zu „Atomkraftgegner“ verwendet; eine solche Relation zwischen diesen Entitäten existiert allerdings in keinem Lexikon.

- (a) Heringe (S1)
- (b) Clupeidae (S1) [realisiert als Synonymie zu (a)]
- (c) Die Heringe (S2) [realisiert als Identität (a)]
- (a) der pazifische Hering (S3)
- (b) Clupea pallasii (S3) [realisiert als Synonymie zu (a)]
- (a) der Atlantische Hering (S3)
- (b) Clupea harengus (S3) [realisiert als Synonymie zu (a)]
- (c) er (S4) [realisiert als Identität zu (a)]
- (a) der Herbsthering (S5)
- (b) der Blankhering (S5) [realisiert als Synonymie zu (a)]
- (c) der Down(s)hering (S5) [realisiert als Synonymie zu (a)]
- (d) der Bohuslänhering (S5) [realisiert als Synonymie zu (a)]
- (e) der Frühjahrshering (S6) [realisiert als Synonymie zu (a)]
- (a) der Strömmling (S7)
- (b) Ostseehering (S7) [realisiert als Synonymie zu (a)]
- (a) die schlüpfenden Larven (S8)
- (b) sie (S10) [realisiert als Identität zu (a)]

#### Analyse 8: Referenzielle Ketten in Beispieltext 2

Beispieltext 2 zeichnet sich dadurch aus, dass die Koreferenz zwischen Ausdrücken in allen referenziellen Ketten über die Relationen der Synonymie und Identität realisiert wird. Darüber hinaus fällt auf, dass zwischen den Ketten sehr enge lexikalisch-semantische Relationen (Hyponymie bzw. Hyperonymie) bestehen, – d.h. fast alle Ketten stehen über mindestens ein Kettenmitglied in einer Kospezifikationsrelation zu einem Mitglied einer anderen Kette. Im Hinblick auf das Thema bedeutet dies, dass die sprachlichen Teilhandlungen des Textes durchgehend auf semantisch eng aufeinander bezogene Diskursentitäten bezogen sind. Auffällig ist auch, dass es keine Kette gibt, die aufgrund ihrer Länge und insbesondere ihrer globalen Erstreckung als dominante Kette aufzufassen wäre. Der analysierte Auszug von Beispieltext 2 lässt demnach auf der Ebene der Koreferenz keine eindeutigen Schlüsse auf das zentrale Thema des Textes zu. Erst wenn man die lexikalisch-semantische Ebene mit hinzuzieht, werden die engen Bezüge zwischen den Koreferenzketten deutlich. Dies zeigt, dass eine einzelne Textstrukturebene nicht hinreichend ist und stattdessen mit anderen Ebenen verrechnet werden muss, um Schlussfolgerungen im Hinblick auf das Thema des Textes ziehen zu können.

Grundlage für eine maschinelle Ermittlung referenzieller Ketten ist die automatische Identifizierung der Anapher-Antezedent-Verbindungen des Textes. Diese

Aufgabe fällt unter einen klassischen Forschungsgegenstand der Computerlinguistik, die Anaphernresolution oder auch Koreferenzresolution (einen umfassenden Überblick über zentrale Verfahren und Systeme gibt Strube 2010). Ansätze zur Anaphern- bzw. Koreferenzresolution können bezüglich ihrer Methodik voneinander unterschieden werden (heuristische/regelbasierte, statistische Methoden, unüberwachte und überwachte Lernverfahren), bezüglich des Anapherntyps, der verarbeitet werden kann (pronominale, NP- und Ereignisanaphern) sowie bezüglich der Art der Textstrukturannotation, auf deren Basis Anaphernresolution betrieben wird (hierarchische oder relationale Diskursstruktur vs. lineare Textstruktur) – Cristea et al. (1999) konnten in einer empirischen Evaluation zeigen, dass Anaphernresolutionsansätze, die auf der hierarchischen Diskursstruktur operieren, im Gegensatz zu denen, die auf der linearen Textstruktur operieren, zu signifikant besseren Ergebnissen führen. Anaphernresolutionsansätze unterscheiden sich darüber hinaus auch bezüglich der Sprache, die verarbeitet werden kann (sprachunabhängig vs. sprachabhängig); für das Deutsche existieren dabei deutlich weniger Arbeiten als für das Englische (vgl. Strube 2010, 407).

In der Regel werden bei der maschinellen Ermittlung referenzieller (wie auch lexikalischer) Ketten zunächst einmal verschiedene Vorverarbeitungen des Input-Textes durchgeführt (siehe auch Kapitel 3.3.1.1). In einem ersten Schritt muss der Text segmentiert bzw. tokenisiert und lemmatisiert werden, um so Ein- und Mehrwortlexeme zu differenzieren. Darüber hinaus muss ein POS-Tagging stattfinden, um Inhaltswörter von Funktionswörtern zu differenzieren; nur erstgenannte werden in die Identifizierung referenzieller Ketten einbezogen.

In den letzten Jahrzehnten wurde der Fokus der Forschung zur Anaphernresolution überwiegend auf die Resolution von pronominalen Anaphern gelegt (siehe Lappin und Leass 1994; Ge et al. 1998; für einen Überblick siehe Mitkov 2001): Lappin/Leass erzielten bei der Anaphernresolution mit einem heuristischen Modell eine Akkuratheit von 86% bzw. 88% für Pronomen. Ihr Ansatz basiert dabei auf Salienzfaktoren, die von der syntaktischen Struktur abgeleitet sind. Zu diesen Faktoren gehören u.a. die grammatische Funktion des Antezedenten, die Distanz zwischen Anapher und Antezedent, der Parallelismus grammatischer Funktionen von Anapher und Antezedent, die Anzahl der Elemente in einer anaphorischen Kette. Auch Ge et al.s statistischer Ansatz ist in der Lage, pronominale Anaphern mit einer Akkuratheit von 84% aufzulösen. In einem anhand deutschsprachiger Zeitungsnachrichten evaluierten Verfahren berichten Hinrichs et al. (2005) von einem  $F_1$ -Maß von 73.4% bei der Auflösung pronominaler Anaphern. Für diesen Typ von Anapher (sowie für Eigennamen) werden demnach akzeptable bis gute Ergebnisse erzielt (vgl. Strube 2010, 406).

Zur Auflösung von NP-Anaphern, Ereignisanaphern oder Bridging-Koreferenz<sup>195</sup> zwischen Anapher und Antezedent werden häufig maschinelle Lernverfahren (siehe z.B. Aone und Bennett 1995) eingesetzt, bei denen auch Wissen mit einbezogen wird. Als lexikalisch-semantische Wissensressourcen werden dabei entweder Ressourcen wie GermaNet genutzt (siehe Goecke, Stührenberg et al. 2008) oder große Korpora bzw. Teile des World Wide Webs (z.B. Wikipedia, siehe z.B. Ponzetto und Strube 2007b). Sowohl die heuristischen Anaphernresolutionssysteme als auch maschinelle Lernverfahren zur Resolution von NP-Anaphern und Ereignisanaphern weisen allerdings in der Regel eine Akkuratheit auf, die deutlich unter der für pronominale Anaphern oder Eigennamen liegt. Versley (2006), der ein Constraint-basiertes maschinelles Lernverfahren entwickelt und auf ein Korpus deutscher Zeitungsnachrichten anwendet, erzielt mit seinem Ansatz beispielsweise  $F_1$ -Maße von 87.5% für Eigennamen gegenüber nur 46.9% für definite Nominalphrasen.

Neben der Identifizierung von Anapher-Antezedent-Verbindungen bzw. der Identifizierung des Antezedenten einer Anapher wurden in den letzten Jahren erstmals auch Ansätze zur Identifizierung referenzieller Ketten bzw. Cluster koreferenter Nomen entwickelt (siehe z.B. Culotta et al. 2007; Poon und Domingos 2008). In einem Vergleich der paarweisen Koreferenzresolution mit der Cluster- bzw. Mengen-bildenden Koreferenzresolution zeigen Culotta et al., dass paarweise Resolutionssysteme zu schlechteren Ergebnissen führen als Clusterbildende ( $F_1$ -Maß für Koreferenz-Cluster = 79.3%;  $F_1$ -Maß für Koreferenz-Paar = 72.5%). Auch das Verfahren von Poon/Domingos kommt für die Clusterbildende Koreferenzresolution zu einem  $F_1$ -Maß von 79.2% im Gegensatz zu 59.9% für die paarweise Resolution (wobei die zur Evaluation verwendeten Korpora zwischen beiden Ansätzen differieren).

Abschließend muss festgehalten werden, dass Anaphern- bzw. Koreferenzresolutionssysteme nur schwer vergleichbar sind, da sie zum einen auf Basis verschiedener Korpora getestet und evaluiert werden (und sich schon für ein einzelnes System große Unterschiede bzgl. der Akkuratheit bei der Auflösung von Koreferenz in verschiedenen Korpora zeigen, siehe z.B. Versley et al. 2008). Zum anderen sind die Evaluationsmaße, die in den einzelnen Ansätzen verwendet werden, ebenfalls unterschiedlich. Darüber hinaus unterscheiden sich die Systeme in dem, was sie als Vorverarbeitungsschritte voraussetzen, und in dem, was sie zu verarbeiten in der Lage sind. Vergleiche zwischen Systemen sind deshalb schwierig, die o.g.  $F_1$ -Maße nur bedingt aussagekräftig bezüglich der Leistungsfähigkeit eines Systems. Obwohl insbesondere im Bereich der maschinellen Lernverfahren in den letzten Jahren eine Vielzahl neuer Verfahren entwickelt wurde, die zu einer Neubelebung der Forschungssituation geführt haben, liegen für das Deutsche noch keine befriedigenden, für einen Anwender nutzbaren Ko-

---

<sup>195</sup> Für Bridging-Koreferenz berichten Poesio et al. (2004) von 74.2% bzw. 75.2% korrekten Klassifikationen von Bridging-Instanzen in dem englischsprachigen GNOME-Korpus. Anmerkung: Bridging wird in diesem Ansatz eingeschränkt auf die in WordNet repräsentierten Bridging-Klassen der Meronymie, Hyponymie und Synonymie; d.h. dieser Bridging-Begriff ist nicht identisch mit dem in Kapitel 2.6.3.3 angesprochenen Bridging-Begriff von Clark (1977).

referenzresolutionssysteme vor.<sup>196</sup> Aus diesem Grund können auch in diesem Kapitel keine maschinell generierten Ergebnisse für die beiden Beispieltex-te verwendet werden.

Für die manuelle Annotation von Koreferenz liegen mehrere Annotationssche-mata vor. Annotationsschemata für Koreferenz sind teilweise Teil komplexerer Annotationsschemata (dies ist z.B. bei dem Annotationsschema der Text Encod-ing Initiative, TEI, der Fall), oft handelt es sich aber um eigenständige Annota-tionsschemata (siehe z.B. Goecke, Holler et al. 2008). Zu differenzieren sind Annotationsschemata zum einen dadurch, ob sie Standoff- oder Inline-Annotationen modellieren. Während bei Standoff-Annotationen die Annotatio-nen zu Anaphern, Antezedenten und den sie verknüpfenden Koreferenz-Relationen vom eigentlichen Text selbst getrennt gehalten werden, stehen Inli-ne-Annotationen im Text selbst, d.h. an der Stelle des Textes, wo eine Anapher vorkommt und über eine anaphorische Relation auf einen Antezedenten verweist (zum Thema “Standoff-” und “Inline-Annotation” siehe Kapitel 3.3.3). Zum an-deren unterscheiden sich Schemata in der Frage, was als Koreferenz annotiert wird, ob nur Koreferenz im engen Sinne oder auch Kospezifikation oder indirekte Anaphorik (zu den Problemen von Koreferenz-Annotationsschemata siehe van Deemter und Kibble 2001).

### 3.3.1.3 Thematischer Wortschatz, lexikalische Ketten und metasprach-liche lexikalische Thema-Marker

Das lexikalische Material eines Textes, d.h. die Menge der Wörter und Phrasen im Text, stellt eine wichtige Basis für die Bestimmung des Themas dar. Man kann drei Typen lexikalischer Thema-Marker unterscheiden: a) Metasprachliche Ausdrücke oder Phrasen zur Markierung der thematischen Organisation (insbe-sondere Markierungen thematischer Wechsel), b) Inhaltswörter, die sich aus der kommunikativen Bezugnahme auf einen bestimmten Texttyp ergeben (bzw. der Bezugnahme auf Texttypstrukturkategorien als abstrakte Thementypen), c) In-haltswörter, die dazu verwendet werden, um über ein spezifisches Thema zu sprechen.

Metasprachliche Ausdrücke oder Phrasen zur Markierung der thematischen Or-ganisation (a) können differenziert werden in texttypunabhängige und texttyp-spezifische (zur Diskussion der Textsortenspezifik von “topic-shift markers” sie-he Brown und Yule 1983, 94ff., insbesondere 100; für texttypunabhängige „Thematisierungsformeln“ siehe Hoffmann 1997, 524). Lexikalische Mittel zur Markierung von Themenwechseln sind z.B. bestimmte satzinitiale Adverbien wie „ansonsten“, „übrigens“ (können einen Themenübergang signalisieren) sowie explizite (mehr oder weniger formelhafte) lexikalisierte Phrasen, die entweder

---

<sup>196</sup> Für das kommerzielle Anaphernresolutionssystem ROSANA (“robust syntax-based interpretation of anaphoric expressions”, vgl. Stuckardt 2001) finden sich nur Evaluationsergebnisse für einen englischsprachigen Korpus, nicht aber für einen deutschsprachigen – aus diesem Grund können keine Aussagen über die Qualität von ROSANA für das Deutsche getroffen werden und ROSANA wird – obwohl es von Anwendern direkt genutzt werden kann – bei dieser Aussage außen vor gelassen.



einen Themenwechsel, die Einführung eines neuen Themas oder die Wiedereinführung eines bereits eingeführten Themas anzeigen (z.B. „im Hinblick auf“). Lexikalische Mittel zur Markierung von Themenwechseln können außerdem Ausdrücke sein, die einen Orts-, Zeit- oder Personenwechsel anzeigen (vgl. van Dijk und Kintsch 1983, 204; Pufahl 1992, 222). Weitere Beispiele für metasprachliche lexikalische Markierungen der thematischen Organisation von Dialogen gibt Hoffmann (1997, 513ff.); im Folgenden sind einige explizite „Thematisierungsformeln“ aufgeführt (siehe Hoffmann 1997, 524f.):

**Neuthematisierung:**

„(Zum) Thema/Punkt (+th)“ „Was th anbelangt/betrifft“, „Im (Hin-) blick auf +th“, „Um (nun) auf +th zu (sprechen zu) kommen“, „Da ist/wäre noch +th“

**Dethematisierung:**

„Dabei fällt mir ein“, „Nebenbei (bemerkt)“, „im Übrigen“, „Ansonsten“

Diese und weitere Lexeme und lexikalisierte Phrasen könnten in einem speziellen Thema-Marker Lexikon erfasst werden, um so bei der automatischen Verarbeitung von Texten genutzt werden zu können (vgl. Kapitel 3.2.5).

Texttypspezifische Inhaltswörter (b) können ebenfalls in einem Lexikon erfasst werden; ein erster Ansatz für die systematische Erfassung dieses Typus von Thema-Markern findet sich bei Brinker (1997). Wie bereits in Kapitel 3.2.2 angesprochen, können Texttypstrukturkategorien als abstrakte Thementypen aufgefasst werden, insofern ist das Wortmaterial, das verwendet wird, um auf diese Kategorien Bezug zu nehmen, als Reflexion dieses abstrakten Thementyps im Text zu sehen. In wissenschaftlichen Zeitschriftenartikeln beispielsweise würden sich texttypspezifische Inhaltswörter demnach auf die texttypspezifischen Kategorien „theoretischer Rahmen“, „Methode“, „Daten“, „Ergebnisse“ u.a. (siehe Kapitel 2.6.3.6) beziehen. Die Menge der texttypspezifischen Inhaltswörter bildet dabei den speziellen (texttypspezifischen) Wortschatz – für den genannten Texttyp wäre dies der allgemeinwissenschaftliche Wortschatz bzw. das allgemeine Vokabular wissenschaftlichen Handelns, insbesondere das logisch-methodologische Vokabular (siehe Strauß und Zifonun 1985, 84f.). Zu diesem Vokabular gehören beispielsweise folgende Wörter:

Parameter, Versuch, Auswertung, Charakterisierung, Kritik, Methode, Untersuchung, Schlussfolgerung, Summary, Ziel, Arbeit, Studie, Ansatz, Bestimmung, Analyse, Interpretation, Vergleich, Typisierung, Überblick, Frage, Problem, Grundbegriff, Methode, Verifikation, Prinzip, Test, Diskussion, Hypothese, Experiment, Versuch, Quelle, Fragestellung, Probe, Entnahme, Sammlung, Daten, Material, Transformation, Markierung, Reaktion, Analyse  
(Auszug aus einem Vokabular wissenschaftlichen Handelns, erstellt von Harald Längen im Rahmen der Arbeit im LOEWE-Teilprojekt „Die Ordnung von Wissen in Texten. Von Dokumentstrukturen zu Wissensordnungen“, Justus-Liebig-Universität Gießen 2009)

Wie bereits angesprochen, können diese Wörter als Reflexionen der abstrakten Thementypen aufgefasst werden. Allerdings bleibt festzuhalten, dass texttypspezifische Wörter prinzipiell auch dazu verwendet werden können, um auf ein konkretes Thema Bezug zu nehmen – eine Differenzierung der Reflexionen abstrakter vs. konkreter Themen ist deshalb für jeden Text konkret vorzunehmen, so dass im Gegensatz zu den metasprachlichen Thema-Markern ein automatischer Abgleich von Lexikon und Text mit dem Ziel einer automatischen Bestim-

mung der texttypspezifischen lexikalischen Thema-Marker nur eingeschränkt möglich ist.

Der dritte Typ lexikalischer Thema-Marker (c) ist der für die Identifizierung des konkreten Themas als Sachverhaltskomplex zentrale. Die Menge der Inhaltswörter dieses Typs repräsentiert den thematischen Wortschatz eines Textes. Grundlegend für den Zusammenhang von Thema und lexikalischer Struktur ist dabei die Annahme, dass Themen in Texten durch ein spezifisches Wortmaterial (konkreter: spezifische Inhaltswörter)<sup>197</sup> ausgeformt werden. Die für einen Text spezifischen thematisch relevanten Inhaltswörter können maschinell mittels computerlinguistischer Methoden zur Berechnung der Termrelevanz<sup>198</sup> von den thematisch nicht relevanten Inhaltswörtern, den texttypspezifischen Inhaltswörtern sowie den Funktionswörtern differenziert werden. Am verbreitetsten ist in diesem Kontext die  $tf \cdot idf$ -Formel (siehe Kapitel 3.2.1). Die Forschungsrichtung, die sich mit der Identifizierung thematisch relevanter Inhaltswörter bzw. domänenspezifischer Informationen beschäftigt, ist das Information Retrieval und darauf aufbauend die Informationsextraktion, speziell die Ansätze zur Term-, Keyword- bzw. Keyphrase-Extraktion, zur Topic-Extraktion sowie – im weitesten Sinne – Ansätze zur Concept-Extraktion. Prinzipiell kommen alle Methoden aus diesen Forschungs- und Anwendungsbereichen für die maschinelle Identifizierung der domänenspezifischen Inhaltswörter in Frage.

Da davon auszugehen ist, dass die relevanten domänenspezifischen Inhaltswörter, d.h. die das Thema reflektierenden Wörter, d.h. die Wörter, die dazu verwendet werden, um über ein bestimmtes Thema zu sprechen, in einer Verwandtschaftsrelation zueinander stehen, werden im Folgenden nur solche linguistischen und computerlinguistischen Ansätze behandelt, in denen es um die Modellierung und maschinelle Identifizierung von Clustern verwandter Wörter oder Terme geht. Die linguistischen Modelle können dabei differenziert werden in zwei Typen: Modelle, in denen eher der Aspekt der semantischen Verwandtschaft hervorgehoben und von „Isotopie“ (vgl. z.B. Greimas 1971; Dressler 1973; Agricola 1976) oder „Wortfeldern“ (Lutzeier 1995, 103ff.; Adamzik 2004a, 69ff.; Schwarz und Chur 2004, 60ff.) gesprochen wird (in der Computerlinguistik ist auch die Rede von semantischer Ähnlichkeit oder semantischer Verwandtschaft, siehe Kapitel 3.2.1), und Modelle, in denen der Aspekt der lexikalischen Verwandtschaft betont wird, bei dem davon ausgegangen wird, dass Themen durch „lexical chains“ (Halliday und Hasan 1976; Hasan 1984; Morris und Hirst 1991), „lexical patterns“ (Wikberg 1990, 231f.) oder „patterns of lexis“ (Hoey 1991) beschrieben werden können.

Der Ansatz, lexikalische Kohäsion über lexikalische Ketten zu erklären, ist vor allem in der computerlinguistischen Forschung weit verbreitet und es wurde – aufbauend auf dem Algorithmus von Morris/Hirst (1991) – eine Vielzahl von Methoden zur maschinellen Identifizierung und Verarbeitung lexikalischer Ket-

---

<sup>197</sup> Besonders deutlich ist dies nach Jahr (1996, 125ff.; 2000, 393) bei Fachwörtern, die als „maximal kondensierte Texte“ und damit als „Makroproposition der höchsten Hierarchiestufe“ aufgefasst werden können.

<sup>198</sup> „Terme“ sind in diesem Kontext Wörter (als Token) oder auch Stammformen von Wörtern.

ten entwickelt. Lexikalische Ketten werden in der Computerlinguistik z.B. zur Ermittlung thematisch homogener Textblöcke verwendet, insofern Brüche zwischen lexikalischen Ketten als Themenwechsel (“topic shifts”) interpretiert werden. Das Konzept der lexikalischen Ketten wird in diesem Zusammenhang zur *automatischen* thematischen Segmentierung von Texten eingesetzt (siehe z.B. Jobbins und Evett 1998; Boguraev und Neff 2000a; Stokes 2003). Lexikalische Ketten werden andererseits auch zur Identifizierung von Themen (Topic Identification)<sup>199</sup>, Extraktion von Themen (Topic Extraction), zum Topic Tracking oder zur automatischen Zusammenfassung von Texten genutzt (siehe z.B. Barzilay und Elhadad 1999; Hovy und Lin 1999; Dahl 2000; Stokes 2004; Ercan und Cicekli 2008). Die Einträge einer Kette werden dabei als Cluster von Themenwörtern aufgefasst, d.h. Gruppen von Wörtern, die sich auf ein Thema beziehen.<sup>200</sup> Barzilay/Elhadad (1999, 14) gehen beispielsweise davon aus, dass jede Kette ein Thema oder Subthema repräsentiert. In den Ketten werden ausschließlich Inhaltswörter (keine Funktions- oder Strukturwörter) berücksichtigt, da nur diese in direkter Form vom Thema abhängig sind (vgl. Adamzik 2004b, 152).<sup>201</sup>

Ketten unterscheiden sich nach Ansicht von Morris/Hirst (1991, 32), Barzilay/Elhadad (1999, 14) und anderen im Hinblick auf ihre Wichtigkeit für die thematische Beschreibung des Textes. Diese Wichtigkeit ist demnach abhängig von der „Stärke“ der Kette, die wiederum abhängig ist von den drei Faktoren “Reiteration” (Häufigkeit der Wiederholung von Einträgen), “Density” (Anzahl der Einträge), “Length” (Länge der Kette). Es kann davon ausgegangen werden, dass starke Ketten globale Textthemen signalisieren, weniger starke lokale Aspekte des Themas: “The dominant lexical chains (stretching over many sentences with many words participating) indicate the topic of a text.” (Teich und Fankhauser 2003, 141). Darüber hinaus können lexikalische Ketten – wie die referenziellen Ketten auch – bei der Differenzierung von lokalen und globalen Themen helfen. Ketten, die sich nur über einen bestimmten Teil des Textes erstrecken, deuten auf lokale Themen hin, während Ketten, die einen großen Teil des Textes abdecken, eher für globale Themen stehen. Zwischen lexikalischen Ketten gibt es, wie auch zwischen referenziellen, Interaktionen und Querbezüge. Diese können ebenfalls als Signale zur Bestimmung der thematischen Entwick-

<sup>199</sup> Lexikalische Interpretationen des Themas werden vorwiegend in computerlinguistischen Ansätzen vertreten (siehe Kapitel 2.5), da in diesen Arbeiten das Ziel verfolgt wird, die im Text behandelten Themen automatisch zu identifizieren oder zu extrahieren. Lexikalische Thema-Auffassungen sind deshalb in diesem Kontext sehr beliebt, weil sie einer Analyse der Textoberfläche zugänglich sind.

<sup>200</sup> Barzilay/Elhadad (1999, 14) gehen davon aus, dass die Einträge einer Kette dasselbe Konzept repräsentieren.

<sup>201</sup> Das semantisch orientierte Thema-Konzept von Agricola (1976; 1979) versucht in ähnlicher Weise über das Konzept der (semantisch definierten) Isotopieketten, die – ähnlich wie die lexikalischen Ketten – aus durch semantische Relationen verknüpften Einträgen bestehen, die die Themenstränge des Textes repräsentieren, Aussagen über das Thema bzw. die thematische Struktur zu machen (siehe Kapitel 2.2.2). Das Konzept der Isotopieketten unterscheidet sich von dem der lexikalischen Ketten allerdings insofern, als lexikalische Ketten auf Basis der Wortbedeutung gebildet werden, während sich das Isotopiekonzept auf semantische Merkmale von Wortbedeutungen bezieht und dementsprechend unterhalb der Wortebene angesiedelt ist (vgl. Heinemann 2000a, 56).

lung im Text aufgefasst werden. Grundlage hierfür ist die Vergleichbarkeit der Struktur lexikalischer Ketten und der thematischen Entwicklung:<sup>202</sup>

I think that cohesion, particularly lexical cohesion, does contribute importantly to a text's organization: (...) the patterning (...) relates to the complex ways in which topics may interrelate in their development. (Hoey 1991, 13f.)

Lexikalische Interaktionen zwischen Sätzen lassen sich in Form eines Netzes oder einer Matrix repräsentieren, in der für alle Satzpaare des Textes sämtliche lexikalischen Verbindungen zwischen lexikalischen Einheiten markiert sind (siehe Kapitel 3.1.1.2). Hoey (1991, 91ff.) zeigt für einen Beispieltext, dass 70% der Satzpaare keine signifikante Verknüpfung aufweisen.<sup>203</sup> Sätze, die keine oder wenige Bezüge ("repetition links") zu anderen Sätzen aufweisen, sind seiner Ansicht nach eher unwichtig ("marginal") für den Text, während Sätze mit vielen Bezügen zu anderen Sätzen zentral ("central") sind.<sup>204</sup> Zentrale Sätze wiederum haben eine höhere Wahrscheinlichkeit, gleichzeitig auch thematisch wichtig zu sein:

We might expect those sentences that are germane to the development of the theme(s) of a text (that is, in our terms, central) to make a number of connections with other sentences, while those sentences that contribute less to the development of its theme(s) (that is, marginal) we might expect to show fewer signs of connection with the rest of the text. (Hoey 1991, 43).

In dieser Arbeit werden lexikalische Ketten als Reflexionen des Themas betrachtet. Für die Analyse lexikalischer Ketten wird dabei auf die Ansätze von Hoey (1991) und Morris/Hirst (1991) Bezug genommen, da diese bereits in zahlreichen computerlinguistischen Anwendungen operationalisiert wurden (Barzilay und Elhadad 1999; Harabagiu 1999; Silber und McCoy 2002; Teich und Fankhauser 2003; Stokes 2004; Ercan und Cicekli 2008; Karamuftuoglu und Vechtomova 2008). Statt lexikalische Interaktionen zwischen *Sätzen* werden allerdings Interaktionen zwischen lexikalischen *Ketten* untersucht (in Anlehnung an die Vorgehensweise von Agricola (1976; 1979)).

Für die Bildung von Ketten soll folgendes gelten: In den Ketten werden ausschließlich Inhaltswörter (keine Funktions- oder Strukturwörter) berücksichtigt, da nur diese in direkter Form vom Thema abhängig sind (vgl. Adamzik 2004b, 152). Kettenglieder sind einzelne Wörter, nicht Phrasen, d.h. nur monolexikalische, keine polylexikalischen Einheiten.<sup>205</sup> Zwei aufeinanderfolgende Kettenmit-

<sup>202</sup> Auf den Zusammenhang zwischen lexikalischer Kohäsion und thematischer Entwicklung weisen auch Hasan/Halliday (1976, 288) und Hasan (1984, 216) hin.

<sup>203</sup> Was eine signifikante Verknüpfung darstellt, wird quantitativ nach der Anzahl der Verbindungen zwischen einem Satzpaar bewertet – dies muss für jeden Text in Abhängigkeit von seiner Länge sowie seiner lexikalischen Dichte neu bestimmt werden. Im Beispiel von Hoey sind zwei Sätze signifikant verknüpft, wenn diese Sätze mehr als drei lexikalische Verbindungen aufweisen.

<sup>204</sup> In ähnlicher Weise geht auch Hasan (1984) davon aus, dass diejenigen sprachlichen Einheiten, die mit Gliedern anderer Ketten interagieren, den Status von "central tokens" haben. Durch die Prozentzahl der "central tokens" im Verhältnis zur Gesamtzahl der Tokens kann ihrer Ansicht nach der Grad der kohäsiven Harmonie des Textes ermittelt werden (Hasan 1984, 216ff.).

<sup>205</sup> Die Einschränkung auf monolexikalische Einheiten hat damit zu tun, dass Kettenmitglieder in einer lexikalisch-semantischen Relation stehen müssen und dass zur Bestimmung lexikalischer

glieder dürfen in kürzeren Texten (< 30 Sätze) maximal 5 Sätze voneinander entfernt sein; ist der Abstand größer, wird eine neue Kette eröffnet.<sup>206</sup> Ketten mit weniger als drei Mitgliedern werden in der Analyse nicht berücksichtigt. Für längere Texte erhöhen sich diese Maße. Ein Wort wird einer Kette hinzugefügt, wenn vorgenannte Bedingungen erfüllt sind und es zu einem Mitglied der Kette in einer der in Kapitel 2.6.3.2 genannten lexikalisch-semantischen Relationen steht; zur Bestimmung lexikalisch-semantischer Beziehungen sollte dabei Bezug auf einen Thesaurus o.ä. genommen werden.<sup>207</sup> Ein dieser Anweisung entsprechender detaillierter Algorithmus zur Bestimmung lexikalischer Ketten findet sich in Morris/Hirst (1991, 34).

Für den Beispieltext 1 wurde eine Analyse lexikalischer Ketten simuliert, d.h. die dargestellten Analyseergebnisse stellen ein mögliches Ergebnis einer maschinellen Analyse dar (Analyse 9).<sup>208</sup> Für Text 1 können fünf lexikalische Ketten ermittelt werden. Die Lexeme in Kette 1a stehen in einer Relation der Identität bzw. der partiellen Identität; sowohl Kette 1b als auch Kette 2 bestehen aus mehreren Gruppen von Lexemen, die in nicht-systematischen lexikalisch-semantischen Relationen zueinander stehen, die auf den semantischen Rollen AGENS, UMSTAND, FOLGE (Kette 1b) bzw. EREIGNIS, UMSTAND, ZIEL (Kette 2) beruhen. Kette 3 umfasst zwei Gruppen, die jeweils Mitglieder einer Menge enthalten, wobei die Lexeme der ersten Gruppe („Hannover“, „Gorleben“) zu einem Lexem der zweiten Gruppe in einer Teil-Ganzes-Relation stehen („Niedersachsen“). Die Lexeme der Kette 4 weisen untereinander verschiedene semantische Relationen (z.B. ist „Bahnstrecke“ ein Holonym zu „Oberleitung“) oder auch nicht-systematische Relationen auf – alle Mitglieder der Kette stellen dabei Bezeichnungen von Teilen des Frames „Bahn“ dar. Zwischen den Mitgliedern der Kette 5 besteht eine Relation der Hyperonymie („Behörde“ – „Bundeskriminalamt“, „Behörde“ – „Verfassungsschutz“) bzw. der Kohyponymie („Bundeskriminalamt“, „Verfassungsschutz“).

---

Ketten häufig ein Thesaurus oder eine semantische Wissensquelle wie z.B. WordNet bzw. GermaNet verwendet wird, in denen in der Regel keine Mehrwortlexeme enthalten sind (siehe z.B. Morris und Hirst 1991; Barzilay und Elhadad 1999; Dahl 2000; Teich und Fankhauser 2003). Die Einbeziehung von Mehrwortlexemen, insbesondere Phraseologismen, in die Analyse der lexikalischen Struktur muss deshalb unabhängig von der Analyse lexikalischer Ketten erfolgen und stellt damit ein eigenständiges und in dieser Arbeit nicht behandeltes Problem im Kontext der maschinellen Identifikation des Themas dar.

<sup>206</sup> Die Werte dieser Parameter (ab wann ist ein Text kurz, ab wann lang; wieviel Sätze dürfen Kettenmitglieder voneinander entfernt sein) sind bisher nicht im Hinblick auf ihre Brauchbarkeit validiert worden. Eine empirische Evaluation kann deshalb möglicherweise zu einer Änderung der Werte führen. Auch die Textlänge spielt eine Rolle dabei, welcher Abstand zwischen Kettenmitgliedern als sinnvoll zu erachten ist.

<sup>207</sup> Für deutschsprachige Texte, insbesondere wissenschaftliche Artikel, ist leider oft zu konstatieren, dass längst nicht alle Lexeme in lexikalisch-semantischen Wissensquellen enthalten sind (siehe auch Kapitel 3.2.5).

<sup>208</sup> In den folgenden Analysen stehen Ausdrücke in Kapitälchen für Konzepte, die den Kopf oder die Slots eines Frames darstellen. In Klammern hinter den Ausdrücken sind die Satznummern mit „S“ abgekürzt.

**Lexikalische Kette (1a):**

Anschläge (S2, S6, S11), Bahnanschläge (S4)

[Lexeme zur Bezeichnung des EREIGNISSES Anschlag]

**Lexikalische Kette (1b):**

Gewalt (S1), gewaltsam (S4), gewalttätig (S9) | zerstört (S5), Sachschaden (S6), verletzt (S7) | Einzeltäter (S10),

[Lexeme zur Bezeichnung von UMSTÄNDEN | FOLGEN | und VERURSACHERN des EREIGNISSES Anschlag]

**Lexikalische Kette (2):**

Castor (S1), Atomtransport (S2), Castor-Transport (S4) | Atomtransport-Blockade (S2), atomares Zwischenlager (S4)

[Lexeme zur Bezeichnung des BEWEGUNGSEREIGNISSES Anschlag | der direkten BEGLEITUMSTÄNDE | sowie des ZIELS DER BEWEGUNG]

**Lexikalische Kette (3):**

Hannover (S3), Gorleben (S4) | Brandenburg (S5), Berlin (S5), Hessen (S5), Niedersachsen (S5), Schleswig-Holstein (S5)

[Lexeme zur Bezeichnung von GEOPOLITISCHEN ORTEN und REGIONEN IN DEUTSCHLAND]

**Lexikalische Kette (4):**

Bahnstrecke (S3, S5), Oberleitung (S5), Deutsche Bahn AG (S6), Züge (S8), Nahverkehr (S8), Fernverkehr (S8)

[Lexeme zur Bezeichnung von verschiedenen Teilen des Frames BAHN]

**Lexikalische Kette (5):**

Verfassungsschutz (S9), Bundeskriminalamt (S10), Behörde (S11)

[Lexeme zur Bezeichnung von INSTITUTIONEN, die die Aufgabe der Beobachtung und Bewertung krimineller Ereignisse haben]

## Analyse 9: Lexikalische Ketten in Beispieltext 1

Die zentrale lexikalische Kette ist aufgrund ihrer Länge (von Satz 1 bis Satz 11), der Anzahl ihrer Einträge (10) sowie der (partiellen oder vollständigen) Reiteration ihrer Kettenglieder („[anschläge]“ (Satz 2, 4, 6, 11), „[gewalt]“ (Satz 1, 4, 9)) die Kette 1 (1a plus 1b). Diese erhält den Status „+Thema“, d.h. sie wird als Reflexion des Hauptthemas des Textes betrachtet (während die übrigen Ketten als Reflexionen von Aspekten dieses Hauptthemas aufgefasst werden). Interaktionen zwischen den Ketten können in Form einer Matrix abgebildet werden, die für Beispieltext 1 deutlich belegt, dass Kette 1 die meisten Bezüge zu anderen Ketten aufweist und deshalb das lexikalische Zentrum des Textes darstellt (siehe Analyse 10).

	(1a)			
(1b)	Gewalt [EIGENSCHAFT von] → Anschläge  Sachschaden/zerstört/verletzt [FOLGE von] → Anschläge	(1b)		
(2)	Bahn(anschläge) [INSTRUMENT für] → (Castor)-Transport		(2)	
(3)	Anschläge → [ORT] Brandenburg, Berlin, Hessen, Niedersachsen, Schleswig-Holstein, bundesweit			(3)
(4)	Bahn(anschläge) [das lexikalische Morphem „Bahn“ in der Lesart „Fortbewegungsmittel“ ist SYNONYM zu] → Züge  Bahn(anschläge) [partielle Rekurrenz] → Bahnstrecken, Deutsche Bahn AG		Züge [INSTRUMENT für] → Castor-Transport/Atomtransport	(4)
(5)	Anschläge [hat KONSEQUENZ Ermittlung des AGENS] Verfassungsschutz, Bundeskriminalamt	zerstört/verletzt/Sachschaden [hat KONSEQUENZ Ermittlung des AGENS] → Bundeskriminalamt		bundesweit [partielle lexikalische Rekurrenz] → Bundeskriminalamt

Analyse 10: Interaktionen zwischen den lexikalischen Ketten in Beispieltext 1

Für Beispieltext 2 kann – aufgrund der Länge, der Anzahl der Mitglieder sowie der Häufigkeit der Wiederholung von Kettenmitgliedern – Kette 1 als zentrale lexikalische Kette bestimmt werden, d.h. als Reflexion des globalen Textthemas (siehe Analyse 11). Bei Kette 2 handelt es sich wie bei Kette 1 um eine globale Kette, die Ketten 3 bis 5 stellen dagegen lokal begrenzte Ketten dar, die nur für einen Teil des Textes relevant sind – sie stellen lokale Aspekte des Textthemas dar.

**Lexikalische Kette (1):**

Heringe (S1, S2), Knochenfische (S1), Heringsartige (S1); Schwarmfische [Schwärme (S3)] (S2); Pazifischer Hering (S3), Atlantischer Hering (S3, S8), Herbsthering (S5), Blankhering (S5), Down(s)hering (S5), Bohuslänhering (S5); Frühjahrshering (S6), Ostseehering (S7)  
[Lexeme zur Bezeichnung von ARTEN VON FISCHEN]

**Lexikalische Kette (2a):**

Meeresgebiete (S1, S3); Küste (S3, S6, S11), Pazifik (S3); Nordseeküste (S5), Nordsee (S5), Gewässer (S5); Fjorde (S6), Ostküste (S6); Ostsee (S7), Meeresgrund (S8), Küstenbereich (S10), Flüsse (S14)  
[Lexeme zur Bezeichnung verschiedener Arten von GEWÄSSERN und an diese Gewässer angrenzende Bereiche (LANDSCHAFTSARTEN)]

**Lexikalische Kette (2b):**

tropisch (S1); nördlich, Europa (S3); schottisch (S5), südlich (S5), schwedisch (S5); norwegisch (S6), Island (S6)  
[Lexeme zur Bezeichnung von REGIONEN (sowohl relationale Angaben als auch konkrete regionale Punkte), die nicht Gewässer sind]

**Lexikalische Kette (3):**

Familie (S1), Ordnung (S1), Heringsartige (S1), Arten (S1); Gattung (S3)  
[Lexeme zur Bezeichnung von DIMENSIONEN zur ontologischen Einordnung von Lebewesen]

**Lexikalische Kette (4):**

Laichzeit (S4), Laichplätze (S4), Laichbedingungen (S4); laichen (S5, S6, S14); Eier (S8), Larven (S9), Geschlechtsreife (S12)  
[Lexeme zur Bezeichnung von GEGENSTÄNDEN, EREIGNISSEN, SITUATIONEN, ZEITEN oder ORTEN, die mit der FORTPFLANZUNG VON FISCHEN zu tun haben]

**Lexikalische Kette (5):**

Jugend(entwicklung) (S10), Jahre (S11, S12, S13), Alter (S12), Lebensdauer (S13)  
[Lexeme zur Bezeichnung von ZEITSPANNEN, ZEITABSCHNITTEN oder ZEITPUNKTEN, die LEBEWESEN betreffen]

## Analyse 11: Lexikalische Ketten in Beispieltext 2

Sowohl im Hinblick auf die Bildung der lexikalischen Ketten als auch im Hinblick auf die Analyse der Interaktionen zwischen den Ketten muss einschränkend gesagt werden, dass sich manuelle (von Menschen durchgeführte) Analysen interindividuell unterscheiden, insbesondere aufgrund der Variabilität und Offenheit der nicht-systematischen Relationen. Eine Möglichkeit der Objektivierung und gleichzeitig auch der Automatisierung stellt die Verwendung externer Ressourcen zur Identifizierung der Relationen zwischen Wörtern (oder Konzepten, auf die diese Bezug nehmen) dar. Solche Ressourcen können lexikalisch-semantic Wörterbücher bzw. Wortnetze wie GermaNet sein, Ontologien oder Korpora (z.B. DeReKo<sup>209</sup>), über die z.B. Kollokationen von Wörtern ermittelt werden können. Während bei den lexikalisch-semantic Wissensquellen das

<sup>209</sup> Das DeReKo (Deutsches Referenzkorpus) enthält mit insgesamt 3,9 Milliarden Wörtern die weltweit größte Sammlung deutschsprachiger Korpora (Stand 02.03.2010): [www.ids-mannheim.de/kl/projekte/korpora/](http://www.ids-mannheim.de/kl/projekte/korpora/)



Problem darin besteht, dass nicht alle Wörter oder Relationen zwischen Wörtern erfasst sind (siehe Kapitel 3.2.5), hat die Verwendung von Korpora zur Ermittlung von lexikalisch-semantischen Relationen den Nachteil, dass hierdurch nicht die Art der lexikalisch-semantischen Verwandtschaft zwischen Wörtern bestimmt werden kann, sondern nur Aussagen dazu gemacht werden können, ob zwei Wörter in einem bestimmten Kontext gemeinsam verwendet werden, ob sie kookkurrieren; eine Ausnahme stellt die Verwendung von Wikipedia als Korpus dar; hier können sehr wohl Relationen zwischen Konzepten ermittelt werden (vgl. z.B. Ponzetto und Strube 2007a). Morphologische (partielle) Rekurrenz kann darüber hinaus sehr einfach mittels Pattern-Matching-Verfahren automatisiert ermittelt werden.

Wie bei der maschinellen Ermittlung referenzieller Ketten müssen auch bei der maschinellen Ermittlung lexikalischer Ketten zunächst einmal verschiedene Vorverarbeitungen des Input-Textes durchgeführt werden: Tokenisierung, POS-Tagging, Lemmatisierung (siehe auch Kapitel 3.3.1.1). Im Gegensatz zu den Vorverarbeitungen auf der Ebene der Informationsstruktur ist auf der lexikalischen Ebene darüber hinaus eine Lesartendisambiguierung nötig. Hier geht es darum, einem mehrdeutigen Wort (z.B. „Bank“) den passenden Lesartenindex einer semantischen Ressource wie GermaNet zuzuweisen (und als Lesart entweder „Sitzgelegenheit“ oder „Geldinstitut“ auszuwählen) – die maschinelle Lesartendisambiguierung stellt demnach ein Klassifikationsproblem dar. Zur Lösung dieses Problems wird zunächst der Kontext eines mehrdeutigen Wortes untersucht, d.h. die Kontextwörter werden dazu verwendet, die korrekte Lesart eines Wortes zu ermitteln. Dabei ist zu beachten, dass nicht alle Kontextwörter in gleichem Maße relevant sind – eine zentrale Aufgabe eines automatischen Disambiguierungsverfahrens besteht deshalb darin, aus den Kontextwörtern die besten Indikatorwörter für eine der betrachteten Lesarten des mehrdeutigen Wortes zu ermitteln. Neben den wissensbasierten Disambiguierungsverfahren werden in neueren Arbeiten auch wissensfreie maschinelle Lernverfahren zur Disambiguierung eingesetzt. Nachteil dieser Verfahren ist allerdings, dass in der Regel für jedes zu disambiguierende Wort eigene, von Hand erstellte (disambiguierte) Trainingsdaten verwendet werden müssen (vgl. Ebert et al. 2010, 382ff.). Einen Überblick über die verschiedenen Verfahren und deren Leistungen geben Agirre/Edmonds (2006). Insofern die Akkuratheit der verschiedenen Verfahren und Systeme in hohem Maße abhängt von der Granularität der Lesartendisambiguierung sowie von der Komplexität und Schwierigkeit des Textes, in dem eine Lesartendisambiguierung durchgeführt wird, ist ein direkter Vergleich allerdings schwierig – die in den Evaluationen der einzelnen Verfahren ermittelten Maße für die Akkuratheit der Lesartendisambiguierung liegen zwischen 66% und 77% (vgl. Agirre und Edmonds 2006, 14f.).

In einem nächsten Schritt geht es darum, die lexikalisch-semantische Verwandtschaft zwischen den Wörtern eines Textes zu bestimmen, um so die Mitglieder einer lexikalischen Kette zu identifizieren. Hierzu liegt eine Vielzahl von Ansätzen vor, von rein mathematischen Verfahren, bei denen in der Regel nur der Grad der semantischen Verwandtschaft von Wörtern ermittelt, nicht aber deren lexikalisch-semantische Relation benannt wird, bis hin zu wissensbasierten Ver-

fahren, bei denen unter Rekurs auf lexikalisch-semantische Ressourcen (z.B. GermaNet) oder als solche nutzbare Ressourcen (wie Wikipedia) lexikalische Ketten für einen Text gebildet werden und für jedes Wort einer Kette deren lexikalisch-semantische Relation zu einem anderen Wort der Kette identifiziert wird. Die in diesem Kontext verwendeten computerlinguistischen Verfahren zur Ermittlung semantischer Ähnlichkeiten und Verwandtschaften von Absätzen, Wörtern oder Dokumenten wurden bereits in Kapitel 3.2.1 zusammenfassend dargestellt.

Der letzte Verarbeitungsschritt besteht darin, aus den lexikalischen Ketten Rückschlüsse auf lokale vs. globale und wichtige vs. weniger wichtige Themen sowie auf die thematische Organisation des Textes zu ziehen. Barzilay/Elhadad (1999, 14) zufolge signalisieren Konzepte, die durch eine starke lexikalische Kette repräsentiert sind, das zentrale Thema eines Textes. Sie definieren die Stärke einer Kette dabei in Anlehnung an Morris/Hirst (1991) auf Basis der drei Faktoren "repetition", "density" und "length" (s.o.). Einen empirischen Beleg für diese Hypothese liefern Silber/McCoy (2002, 8f.) – für ein Korpus wissenschaftlicher Artikel sowie wissenschaftlicher Lehrbücher verschiedener Domänen zeigen sie, dass in 79.12% aller Fälle die in starken lexikalischen Ketten enthaltenen Nomen auch in manuell erstellten Zusammenfassungen der Texte vorkommen (Vollständigkeit) und dass in 80.83% aller Fälle Nomen, die in der Zusammenfassung vorkommen, auch in starken lexikalischen Ketten vorkommen (Präzision) – wobei starke Ketten nur einen Anteil von 5% bis 15% aller Ketten eines Textes ausmachen.<sup>210</sup> Ercan/Cicekli (2008, 585) gehen demgegenüber davon aus, dass eine einzelne Kette noch nicht als Reflexion des Themas aufgefasst werden kann und dass stattdessen Cluster starker lexikalischer Ketten verwendet werden müssen. Sie begründen dies damit, dass ein Thema auf der sprachlichen Ebene auch durch Wörter signalisiert werden kann, die nicht verwandt sind, z.B. gilt dies insbesondere für komplexe Themen, die sprachlich durch Themenangaben, die zwei oder mehr Bestandteile aufweisen, repräsentiert werden könnten, z.B. „Ziehung der Lottozahlen am Samstag“, „Terroranschlag in Madrid“, „Unfall auf der B 3 bei Marburg“.

Im Hinblick auf die Verwendung lexikalischer Ketten zur Identifizierung der thematischen Organisation des Textes können drei Varianten unterschieden werden, wobei in allen Fällen davon ausgegangen wird, dass sich thematische Brüche bzw. Themenwechsel in spezifischen Konfigurationen lexikalischer Ketten bzw. lexikalischer Strukturen widerspiegeln:

---

<sup>210</sup> Dabei wird von der Prämisse ausgegangen, dass die Funktion einer Zusammenfassung darin besteht, die zentralen Themen des Ausgangstextes zu nennen; Themen werden dabei sowohl im Volltext als auch in der Zusammenfassung mit Hilfe der gleichen sprachlichen Mittel reflektiert, d.h. Zusammenfassungen selbst werden keinesfalls selbst als Themen betrachtet.

*1. Variante: Beginn und Ende einer lexikalischen Kette signalisieren einen thematischen Bruch*

Morris/Hirst (1991) versuchen einen direkten Zusammenhang zwischen lexikalischen Ketten und strukturellen Einheiten (den Diskurssegmenten) herzustellen, indem sie versuchen, lexikalische Segmente auf strukturelle Segmente direkt abzubilden. Korpusanalysen zeigen allerdings deutlich, dass diese Variante auch aufgrund der in Texten vorkommenden Vielzahl sich überlappender lexikalischer Ketten problematisch ist.

*2. Variante: Beginn und Ende aller lexikalischen Ketten signalisieren einen thematischen Bruch*

Zur Berechnung der Stärke von Segmentgrenzen wird die Anzahl der lexikalischen Ketten, die in Paragraph  $n$  enden, und die Anzahl der lexikalischen Ketten, die in Paragraph  $n+1$  beginnen, addiert. Auf diese Weise ergibt sich für alle adjazenten Paragraphen ein Wert zur Stärke ihrer Grenze; liegen die einzelnen Werte über dem Mittelwert, kann man davon ausgehen, dass es sich um eine semantische Segmentgrenze (sS) handelt (vgl. Stokes 2003). Verbessert werden kann dieses Verfahren durch einen Filter, der alle Segmentgrenzen (sS) entfernt, wenn im Text eine andere Segmentgrenze mit einem höheren Wert „direkt“ vorausgeht oder nachfolgt (der „direkte“ Abstand zwischen sS kann auch größer als 1 gewählt werden).

*3. Variante: Auf Basis lexikalischer Ketten können semantische Ähnlichkeiten zwischen Segmenten ermittelt werden – adjazente, semantisch ähnliche Segmente signalisieren eine Themafortführung.*

Bei dieser Variante wird zunächst die Wichtigkeit bzw. Dichte einer lexikalischen Kette für ein Segment (z.B. einen Absatz) berechnet; dazu wird die Zahl der Ketten-Wörter in diesem Absatz geteilt durch die Zahl der Inhaltswörter in diesem Absatz. Als Ergebnis dieser Berechnungen erhält man eine Matrix mit „chain density vectors“: Für jede lexikalische Kette in jedem Absatz des Textes erhält man einen Wert zur „Chain Density“, außerdem erhält man für alle Absätze die Anzahl aller „Chain Words“ und aller „Content Words“ (siehe Abbildung 14). Im Anschluss wird die semantische Nähe zwischen Absätzen ermittelt, indem die Ähnlichkeit zwischen den „Chain Density Vectors“ mittels eines Ähnlichkeitskoeffizienten – z.B. dem „Dice coefficient of similarity“ – berechnet wird (vgl. Green 1997). Dabei werden alle möglichen Paare von Vektoren verglichen; aus der resultierenden Matrix von Ähnlichkeitswerten kann dann der Mittelwert und die Standardabweichung ermittelt werden und auf dieser Basis die semantische Nähe zweier Absätze.

Chain	Paragraph										
	1	2	3	4	5	6	7	8	9	10	11
1	0.154		0.125	0.100	0.105	0.050	0.105		0.091		0.111
2	0.308	0.161		0.300	0.053	0.100	0.211	0.062	0.091	0.125	
3	0.077										
4							0.105	0.062			
5		0.097	0.125		0.105	0.100	0.053	0.125	0.136	0.125	0.333
6							0.053	0.125			
7	0.077										
8	0.077	0.065			0.053	0.050			0.091		
9		0.032									
10	0.077	0.065	0.125								0.056
11									0.045		
12						0.050			0.045		
13			0.125								
14				0.100							
15		0.032				0.050					
16					0.053		0.158	0.125			0.111
17				0.100							
18						0.050			0.045		
19		0.032	0.125		0.105		0.053	0.125			
20					0.053						0.111
21		0.065							0.045		
22					0.053					0.125	
Chain Words	10	17	5	6	11	9	14	10	13	3	13
Content Words	13	31	8	10	19	20	19	16	22	8	18

Abbildung 14: Matrix der Chain Density Vektoren (aus Green 1997, Table 2)

Die Qualität aller auf Lexical Chaining basierenden Anwendungen ist in hohem Maße abhängig davon, wie gut die lexikalischen Ketten sind – und dies hängt wiederum in erster Linie von der Präzision und Vollständigkeit bei der Identifizierung der lexikalisch-semantischen Verwandtschaftsbeziehungen zwischen den Wörtern eines Textes ab. Empirische Evaluationsstudien aber zeigen, dass sowohl bei der manuellen Annotation von lexikalischen Ketten für deutsche Texte kein zufriedenstellendes Maß an Inter-Annotator-Agreement erreicht wird (vgl. Cramer et al. 2006, 2) als auch bei der maschinellen Annotation – wobei mathematisch-statistische Verfahren wie die Latent Semantic Analysis (siehe Kapitel 3.2.1) zu besseren Ergebnissen führen als wissensbasierte Verfahren, die beispielsweise auf einem lexikalisch-semantischen Wortnetz wie GermaNet oder auch auf Wikipedia basieren (vgl. Cramer et al. 2011, 17). Verbesserungen bei der maschinellen Erstellung lexikalischer Ketten können nach Ansicht von Cramer et al. u.a. durch die Einbeziehung maschineller Lernverfahren erreicht werden sowie durch verstärkte Anstrengungen der Forschungsgemeinschaft im Hinblick auf eine bessere theoretische Definition des Konzepts der semantischen Verwandtschaft, eine Identifizierung der am besten geeigneten Wissensressource (bzw. eine Methode, um mehrere Wissensressourcen zu kombinieren) sowie ein hierauf abgestimmter Algorithmus bzw. eine Familie von Algorithmen. Fazit: Eine zufriedenstellende maschinelle Konstruktion lexikalischer Ketten ist zur Zeit noch nicht möglich. Insbesondere mathematisch-statistische Verfahren wie das Latent Semantic Indexing haben meines Erachtens allerdings ein großes Potential, das möglicherweise durch maschinelle Lernverfahren noch gesteigert werden könnte, so dass davon auszugehen ist, dass in Zukunft qualitativ bessere Verfahren zur maschinellen Konstruktion lexikalischer Ketten existieren werden.

### 3.3.1.4 Visuelle und logische Thema-Marker

In den letzten Jahren ist das visuelle Erscheinungsbild von Texten stärker in den Blickwinkel der Linguistik geraten (siehe z.B. Schröder 1993, 193ff.; Graefen 1997, 163ff.; Stein 2003, 133ff.; Hagemann 2007; Lobin 2009, 53ff.; Spitzmüller 2009). Visuelle Mittel können, wie die in den vorherigen Kapiteln beschriebenen sprachlichen Gestaltungsmittel von Texten auch, den Status von Reflexionen des Themas im Text bzw. von indexikalischen Zeichen für das Thema erhalten. Als indexikalische Zeichen für das Thema fungieren vor allem drei, durch visuelle Merkmale identifizierbare Einheiten der logischen Textstruktur: Titel und (Zwischen-)Überschriften, das aus den Überschriften bzw. der Kapitelstruktur erstellte Inhaltsverzeichnis (falls vorhanden) sowie Absätze. Titel haben die Funktion, den Leser inhaltlich über den Text zu orientieren und ihm dabei zu helfen, sich ein Bild davon zumachen, um was es in einem Text geht, ob er inhaltlich für ihn von Interesse ist (vgl. Hellwig 1984b, 8). Gleichzeitig wecken Titel Erwartungen über den Inhalt bzw. restringieren die Erwartungen des Rezipienten im Hinblick auf den dazugehörigen Text (vgl. Brown und Yule 1983, 139; Schröder 2003, 151).

Hellwig unterscheidet drei Haupttypen von Titeln: 1) die Schlagzeile, die eine zentrale Aussage des Textes vorwegnimmt; 2) der Titel als Beschreibung des „Ko-Textes“<sup>211</sup>, d.h. ein Titel, der u.a. das Thema des Ko-Textes benennt; 3) der Titel als indirekte Charakterisierung des Ko-Textes, z.B. in Form eines Mottos, einer Anspielung o.ä. (Hellwig 1984b, 8ff.). Sowohl der erste Typus von Titel als auch der zweite – eine Schlagzeile, definiert als zentrale Aussage des Textes, kann als oberste Makroproposition oder reduzierteste Paraphrase des Textes interpretiert werden, die wiederum als Textthema aufgefasst werden kann (siehe Wulff 1979; van Dijk 1980) – lassen sich als explizite Themenangaben verstehen, während der dritte Typus nur in indirekter Form auf den Inhalt bzw. das Thema des Textes Bezug nimmt. Diese Titel als Themenangaben können sowohl die grammatische Form einer komplexen NP oder PP, eines Frage- oder Aussagesatzes annehmen. Hellwig (1984b, 9) gibt die folgenden Beispiele:

1. „Der Streit um Schmidts DDR-Reise“
2. „Zum/von/über den Streit um Schmidts DDR-Reise“
3. „Bleibt Schmidts DDR-Reise umstritten?“
4. „Warum Schmidts DDR-Reise umstritten bleibt“

Lässt sich aus dem Gesagten schließen, dass Titel generell als Themenangaben aufzufassen sind? Nicht unbedingt. Die Funktion eines Titels ist nicht für alle Texttypen gleich – nicht immer steht die Informationsfunktion im Zentrum, z.B. bei narrativen Texten. In diesem Fall lässt sich der Titel nicht zwingend als Themenangabe verstehen. Für deskriptive und explikative Texttypen – z.B. wis-

---

<sup>211</sup> Hellwig definiert den Begriff „Ko-Text“ als das „einem Titel zugeordnete aktuelle Stück Sprache“; ein Ko-Text ist „zunächst nur eine äußerlich, u.a. eben durch einen Titel, abgegrenzte Einheit“ (Hellwig 1984b, 5)

senschaftliche Artikel – kann der Titel jedoch sehr wohl als explizite Themenangaben bzw. Themenankündigungen fungieren<sup>212</sup> (vgl. Klimanow 1977, 181; de Beaugrande und Dressler 1981, 166; Fritz 1994, 193), durch die ein bestimmter Frame oder Sachverhaltskomplex aktiviert wird.

Dies gilt allerdings nicht für Überschriften im Text – diese können sehr unterschiedliche Funktionen aufweisen, die darüber hinaus je nach Texttyp divergieren können. Lobin (2009, 91ff.) unterteilt Überschriften auf Basis ihrer Relation zu dem ihnen zugeordneten Textabschnitt in zwei Gruppen mit jeweils unterschiedlichen Subtypen: a) textstrukturelle Relationstypen, b) inhaltliche Relationstypen. Bei den textstrukturellen Relationstypen sorgt die Überschrift für eine funktionale Einordnung des Abschnitts (z.B. zu den typischen Komponenten eines Texttyps; bei wissenschaftlichen Artikeln wäre eine typische textstrukturelle Überschrift beispielsweise „Einleitung“). Ergänzend können hier auch die Überschriften genannt werden, die auf spezifische Ordnungsschemata Bezug nehmen (in wissenschaftlichen Artikeln wären solche Ordnungsschemata beispielsweise die Komponenten wissenschaftlichen Handelns, z.B. Methode, Auswertung, Versuch) oder die auf generische Ordnungsschemata Bezug nehmen (z.B. auf die intrinsische Ordnung von Gegenständen, die chronologische Ordnung von Ereignissen etc.).<sup>213</sup>

Überschriften, die in einer inhaltlichen Relation zum zugeordneten Textabschnitt stehen, stellen dagegen inhaltliche Einordnungsangaben dar. Lobin (2009, 92f.) unterscheidet hier folgende Subtypen: „Darstellung“, „Thematisierung“, „Zuordnung“, „Motto“, „These“, „Frage“. Insofern eine Frage, die in einem Textabschnitt beantwortet wird, im Sinne der in Kapitel 2.3.2 dargestellten Thema-Konzeptionen als Thema aufgefasst werden kann, ebenso wie eine Aussage im Sinne einer Makroproposition, wie sie konstitutiv für die Typen „Motto“ und „These“ ist, können die meisten dieser inhaltlichen Typen als direkte („Thematisierung“, „Darstellung“, „Frage“) oder indirekte Themenangaben („Motto“, „These“) bzw. Themenankündigungen aufgefasst werden.

Neben Titeln und Überschriften sowie dem Inhaltsverzeichnis, welches aus diesen generiert ist, können auch die grafisch abgesetzten Textabsätze als Mittel zur Signalisierung der thematischen Struktur interpretiert werden – Absätze haben den Status thematisch homogener Segmente (vgl. z.B. Longacre 1979, 118; Wulff 1979, 205; Giora 1983, 153; Stein 2003, 110). Dem Absatz kommt der „Status einer grundlegenden Gliederungseinheit für geschriebene Texte zu, die, im Idealfall, die innere Organisation eines Textes mit der äußeren Gestaltung verbindet“ (Stein 2003, 110). Im Hinblick auf den Zusammenhang zwischen Absatzbildung und Themenbearbeitung weist Stein jedoch darauf hin, dass ausführlich behandelte Themen sich häufig auch über mehrere Absätze erstrecken können.

---

<sup>212</sup> Hellwig fasst den Titel nicht nur als Themenankündigung, sogar als den konstituierenden Ausgangspunkt für die Entfaltung eines Themas zum Text auf – „Zwischenstufen auf dem Weg vom Gesamttitel zum Gesamttext“ seien dabei z.B. Gliederungen, Kapiteltitle, Abstracts und Zusammenfassungen (Hellwig 1984b, 19).

<sup>213</sup> Die ergänzend genannten Typen von Überschriften wurden in gemeinsamen Diskussionen von Maja Bärenfänger, Henning Lobin und Harald Lingen entwickelt.

Das bedeutet, dass Absätze nicht automatisch einen Themenwechsel oder einen Übergang zwischen einem Haupt- und Nebenthema signalisieren müssen. Grundsätzlich gilt aber, dass die Wahrscheinlichkeit eines Themenwechsels mitten in einem Absatz eher gering ist, dies gilt insbesondere für Texttypen, die einen konventionell festgelegten Aufbau haben (z.B. wissenschaftliche Textsorten oder Textsorten der Pressekommunikation).

Als visuelles sowie logisch-strukturelles Merkmal kann auch die relative Position einer sprachlichen Einheit im Text, Absatz oder auch Satz betrachtet werden. Auf Satzebene interagiert die visuelle Struktur dabei mit der Informationsstruktur (der Satzerstposition oder auch der Linksherausstellung wird in diesem Kontext eine wichtige Rolle zugewiesen). In analoger Weise könnte man dementsprechend auch auf Absatzebene und Textebene von einer Informationsstruktur sprechen, bei denen ebenfalls die Position eines Elements Aufschluss über seine Funktion oder relative Wichtigkeit geben kann.<sup>214</sup>

In der Computerlinguistik gibt es eine lange Tradition, die Position einer sprachlichen Einheit (als ein Merkmal unter vielen) zur Identifizierung von Themen zu nutzen (siehe z.B. Edmundson 1969; Lin und Hovy 1997).

In the Location method the machine-recognizable clues are certain general characteristics or the corpus provided by the skeletons of documents, i.e. headings and format. The Location method is based on the hypotheses that: (1) sentences occurring under certain headings are positively relevant; and (2) topic sentences tend to occur very early or very late in a document and its paragraphs. (Edmundson 1969, 272)

Die Positionen, die einen Beitrag zur Identifizierung leisten, sind Lin und Hovy (1997, 283; 1999, 20) zufolge teilweise texttypspezifisch, teilweise nicht. Zu den texttypunabhängigen Positionen, die in der Regel eine wichtige Rolle für die Signalisierung bzw. das Verstehen des Themas spielen, gehört z.B. der Titel. In wissenschaftlichen Artikeln ist beispielsweise der vorangestellte Abstract eine texttypspezifische Position, die bei der Identifikation des Themas in prominenter Weise helfen kann; in Zeitungsnachrichten wäre eine solche Position der Lead-Satz (s.u.). Um auch dynamisch, d.h. ohne vorheriges Wissen über die in einem bestimmten Texttyp thematisch relevanten Positionen, diese Positionen für die maschinelle Themenidentifikation nutzen zu können, kann auf die von Lin/Hovy (1997) entwickelte Methode der sog. "Optimal Position Policy" zurückgegriffen werden. Mittels dieser Methode kann für eine Sammlung von Texten automatisch ermittelt werden, welches die Positionen in dieser Textsammlung sind, in denen Sätze vorkommen, die mit einer hohen Wahrscheinlichkeit das Thema des Textes reflektieren.

Zusammenfassend ist festzuhalten, dass demnach eine Reihe visueller Mittel als Reflexionen des Themas aufgefasst werden kann – der Produzent kann Titel und Inhaltsverzeichnis nutzen, um dem Rezipienten zu signalisieren, über was er im Text zu sprechen gedenkt, welche Themen im Text behandelt werden. Im Text

---

<sup>214</sup> Mit der Rolle der Informationsstruktur auf Satz- und Textebene befasst sich seit 2003 der SFB 632 „Informationsstruktur: Die sprachlichen Mittel der Gliederung von Äußerung, Satz und Text“ an der Universität Potsdam (siehe <http://www.sfb632.uni-potsdam.de>).

selbst wiederum können Absätze zur Signalisierung thematischer Einheiten und thematischer Brüche, die unterhalb der Kapitelebene angesiedelt sind, verwendet werden. Der Rezipient kann umgekehrt aufgrund ihrer mehr oder weniger konventionalisierten Verwendung den Titel, die Überschriften, das Inhaltsverzeichnis sowie die Absatzgliederung für das Verstehen der thematischen Organisation nutzen. Im Folgenden sollten diese Annahmen anhand von Beispiel 1 illustriert werden.

# Gewalt gegen Castor nimmt zu

## Anschläge auf Bahnstrecken – Grüne machen bei Atomtransport-Blockade mit

**Hannover (dpa) – Mit einer Welle von Bahnanschlägen hat der gewaltsame Protest gegen den bevorstehenden Castor-Transport ins atomare Zwischenlager nach Gorleben am Dienstag bundesweit einen neuen Höhepunkt erreicht.**

In Brandenburg, Berlin, Hessen, Niedersachsen und Schleswig-Holstein zerstörten vermutlich Atomkraftgegner am frühen Morgen mit Wurfkanern die Oberleitungen von Bahnstrecken. Durch die insgesamt neun Anschläge entstand nach Angaben der Deutsche Bahn AG erheblicher Sachschaden. Verletzt wurde niemand. Zum Teil verspäteten sich Züge des Nah- und Fernverkehrs bis zu drei Stunden.

Der Verfassungsschutz schreibt die Aktionen gewalttätigen Autonomen zu. Das Bundeskriminalamt geht von Einzeltätern

aus. Die Behörde in Wiesbaden hat bisher keine Anhaltspunkte dafür, daß die bundesweiten Anschläge von einer Gruppe koordiniert waren.

Anfang März sollen sechs Castor-Behälter per Bahn ins Zwischenlager Gorleben rollen. Drei stammen aus dem Atomkraftwerk Neckarwestheim (Baden-Württemberg), einer aus dem Reaktor Gundremmingen (Bayern) und zwei Atommüllbehälter aus der französischen Wiederaufbereitungsanlage La Hague. Rund 19 000 Polizisten und Bundesgrenzschutzbeamte werden Ende dieser Woche den Transport begleiten. Nach bisherigen Informationen sollen die Behälter am 28. Februar in Walheim nahe Neckarwestheim per Bahn eintreffen. Von dort soll der rund 700 Tonnen schwere Zug auf dem kürzesten Weg durch Baden-Württemberg, Hessen und Niedersachsen nach Gorleben gehen. Der Zug soll

in Walheim am kommenden Sonntag abends abfahren und am Montag an der Verladestation Dannenberg eintreffen.

Die Parteispitze und Fraktionsmitglieder von Bündnis 90/Die Grünen in Bonn und Niedersachsen kündigten am Dienstag an, sich an den Protesten zu beteiligen. Mit einer Fuhre Mist aus der Region Gorleben protestierte Greenpeace am Kernkraftwerk Gundremmingen gegen den Transport quer durch Deutschland.

Die Gewerkschaft der Polizei (GdP) bezeichnete die Transporte als unsinnig. „Die Atomlobby hat seit 20 Jahren ihre Hausaufgaben nicht gemacht, um für den enormen giftigen Dreck, den sie produziert, Vorsorge zu treffen“, sagte GdP-Vorstandsmitglied Eugen Roth. Die geschätzten Kosten von rund 100 Millionen Mark müsse nun der Steuerzahler tragen. (Tagesthema „Rückfall“ und Region Seite 22)

Abbildung 15: Visuelles Erscheinungsbild von Beispieltext 1

Die Textüberschrift in Beispiel 1 ist fett gedruckt und hat ca. die Schriftgröße 36-40 pt, ist also visuell als besonders wichtig hervorgehoben. Sie dient vor allem der Einführung des globalen Themas des gesamten Diskurses („Gewalt gegen Castor“) und ordnet den Text in den Diskurskontext ein: Über die Wahl des Vorgangsverbs „zunehmen“ wiederum wird signalisiert, dass das im Text thematisierte Ereignis weder gerade begonnen noch geendet hat („zunehmen“ setzt voraus, dass etwas bereits vorhanden ist), so dass mit weiteren Texten zu diesem Ereignis zu rechnen ist. Der Titel ordnet demnach den Text thematisch und zeitlich in die komplexe Diskurssituation ein.

Die Unterüberschrift ist typographisch ebenfalls hervorgehoben (Fettdruck und ca. Schriftgröße 20-22 pt). Inhaltlich geht es um den konkreten bzw. aktuellen Anlass für die Berichterstattung und führt das – im Hinblick auf den Diskurs als Ganzen – lokale Thema des Textes ein: Der erste Teil der Überschrift „Anschläge auf Bahnstrecken“ bezieht sich dabei auf den ersten Teil des Nachrichtentextes, der zweite Teil „Grüne machen bei Atomtransport-Blockade mit“ auf den zweiten Teil (der in den Analysen zu Beispiel 1 nicht behandelt wird).



Der Lead-Satz hebt sich wie Überschrift und Unterüberschrift typographisch vom eigentlichen Nachrichtentext ab (Fettdruck, Absatz mit vergrößertem Zeilenabstand zum nachfolgenden Text). Er stellt einerseits eine Paraphrase der Überschrift und Unterüberschrift dar, andererseits bringt er zusätzliche Details, die Antworten auf die noch fehlenden W-Fragen liefern („Wann?“ – „am Dienstag“; „Wo?“ – „bundesweit“). Insofern in den Überschriften bereits „Gewalt“, „Castor“, „Anschläge“, „Bahnstrecken“ als Aspekte des Themas eingeführt wurden, dient der Lead-Satz der Etablierung und Festigung dieses Themenkomplexes beim Rezipienten.

Neben den bereits erwähnten computerlinguistischen Ansätzen zur maschinellen Identifizierung der für das Thema besonders wichtigen Positionen des Textes soll im Folgenden abschließend noch etwas zur Möglichkeit der automatischen Identifizierung der in diesem Kapitel genannten Einheiten der logischen und visuellen Struktur insgesamt gesagt werden. Die Ansätze, die sich hiermit beschäftigen, werden in der Regel als Document Understanding-Ansätze bezeichnet. Beim Document Understanding geht es darum, die geometrischen und typografischen Merkmale eingescannter Texte so gut wie möglich zu identifizieren und zu kategorisieren (z.B. in Seitentypen wie Deckblatt, Titel, Inhaltsverzeichnis etc.). Dabei wird davon ausgegangen, dass die Layout-Struktur eines Dokuments auch Informationen über die logische Dokumentstruktur in sich trägt, dass eine Analyse der visuellen Textoberfläche Rückschlüsse für die Analyse der logischen Struktur zu ziehen erlaubt.

Systeme zur Dokumentenanalyse können differenziert werden in Systeme, die nur bestimmte Dokumentklassen, und Systeme, die eine breite Gruppe von Dokumentklassen verarbeiten können. Während erstgenannte Systeme vor allem spezifisches Wissen über eine bestimmte Dokumentklasse nutzen, verwenden allgemeine Dokumentenverarbeitungssysteme vor allem generisches Dokumentwissen.

Insofern der Zusammenhang zwischen visuellem Merkmal und logischer Einheit in der Regel konventionell festgelegt ist (z.B. ist ein fettgedrucktes geometrisches Segment von max. zwei Zeilen Länge, das am Anfang eines Dokuments zu finden ist und eine größere Schriftgröße als der folgende Text aufweist, in der Regel der Dokumenttitel) bzw. durch Konsistenz im Gebrauch gekennzeichnet ist (z.B. werden im vorliegenden Dokument längere Zitate typographisch immer gleich kodiert, als eingerückte Textblöcke in der Schriftgröße 10), kann für jede logische Einheit (bzw. jedes Objekt der logischen Dokumentstruktur) eine Reihe von typographischen bzw. geometrischen Merkmalen angegeben werden, über die dieses Objekt mittels statistischer Methoden zu identifizieren ist. Ein Merkmalbündel signalisiert demnach in regelhafter Weise ein bestimmtes logisches Objekt.

Regeln (bzw. das in diesen Regeln ausgedrückte Wissen über Dokumente und logische Objekte) können differenziert werden in spezifische Regeln (spezifisches Wissen) – diese Regeln haben nur für eine bestimmte Dokumentklasse Geltung – und generische Regeln (generisches Wissen) – diese sind für eine breite Klasse von Dokumenten gültig. Regeln (bzw. Wissen) können beispiels-

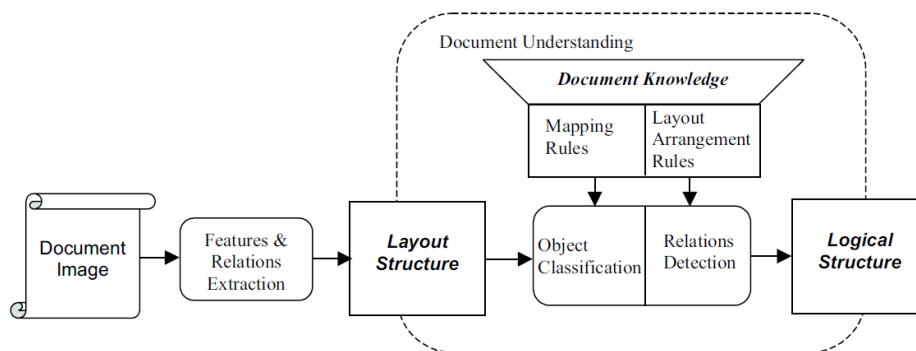


Abbildung 16: Verarbeitungsschritte bei der Dokumentenanalyse und Rolle des Dokumentwissens (aus Aiello et al. 2002, 6, Figure 4)

weise über statistische Methoden, Common-sense Reasoning, maschinelle Lernverfahren oder auch durch manuelle Eingabe erfasst und modelliert werden (vgl. Aiello et al. 2002, 6). In einige Systeme zur maschinellen Dokumentenanalyse fließen dabei nicht nur visuelle bzw. geometrische und typographische Merkmale in die Regeln ein (z.B. Schriftgröße, Schriftstil, Größe des Objekts in Anzahl der Zeilen sowie der Zeichen, Nachbarschaft zu einem anderen Typ von Objekt etc.), sondern auch Textmerkmale (z.B. bestimmte dokumentspezifische Schlüsselwörter wie „Sehr geehrte ...“) sowie relationale Merkmale (z.B. steht der Autorname in der Regel über oder unter dem Titel) (vgl. Klink und Kieninger 2001, 21). Einen Überblick über die typischen Verarbeitungsschritte beim Document Understanding gibt Abbildung 16.

### 3.3.2 Zusammenhänge zwischen Aspekten des Themas

#### 3.3.2.1 Kohärenzrelationen

Da in der vorliegenden Arbeit davon ausgegangen wird, dass Themen Gegenstand sprachlicher Handlungen sind, können zur Identifizierung der thematischen Zusammenhänge zunächst die Relationen zwischen sprachlichen Handlungen, die auf einen Gegenstand bezogen sind, in den Blick genommen werden: Es wird angenommen, dass thematische Zusammenhänge vor allem auf den funktionalen oder inhaltlichen Zusammenhängen<sup>215</sup> zwischen den sprachlichen

<sup>215</sup> Die Frage, ob ein Diskurssegment jeweils nur durch eine Relation an ein anderes Segment gebunden ist, d.h. ob Kohärenz entweder funktional oder inhaltlich erzeugt wird, oder ob ein Diskurssegment auf mehreren Ebenen relational an andere Segmente angeschlossen ist, d.h. ob gleichzeitig sowohl funktionale als auch inhaltliche Kohärenz vorliegt, wird in den verschiedenen Theorien, die sich mit der relationalen Diskursstruktur beschäftigen, unterschiedlich beantwortet. In der Rhetorical Structure Theory (RST, Mann und Thompson 1988) wird davon

Handlungen bzw. Äußerungseinheiten beruhen, die sich auf ein Thema beziehen. Diese Auffassung korrespondiert mit traditionellen Modellen der thematischen Struktur, in denen die thematische Struktur häufig als relationale Struktur beschrieben wird, insofern die thematischen Zusammenhänge zwischen Textteilen als Grundlage der thematischen Struktur aufgefasst werden. Die thematischen Zusammenhänge werden dabei allerdings sehr unterschiedlich gesehen: als sequenzielle Relationen, die vor allem die Entwicklung des Themas von Satz zu Satz betreffen (Daneš 1970; Givón 1983a; Hellwig 1984a; Hoffmann 1997)<sup>216</sup>, als hierarchische Relationen, die auf der Zerlegung eines Hauptthemas in diverse Teilthemen beruhen (Agricola 1979; van Dijk 1980) oder als netz- oder graphenartige Relationen, die durch vielfältige Zusammenhänge zwischen Teilthemen zustande kommen (de Beaugrande und Dressler 1981; Fritz 1982). Allen Ansätzen gemeinsam ist die Annahme, dass die thematische Struktur eines Textes aus Themen besteht, die sich in irgendeiner Form entwickeln bzw. die im Textverlauf entfaltet und z.B. in Subthemen zerlegt werden.<sup>217</sup> Die wichtigsten Elemente der thematischen Struktur sind deshalb die (impliziten oder expliziten) Verknüpfungen zwischen Themen. Werden Themen kohärent miteinander verknüpft, resultiert hieraus der thematische Zusammenhang eines Textes. In der vorliegenden Arbeit wird von drei Haupttypen thematischer Organisation ausgegangen (Beispiel-Analyse zu Beispieltext 1: siehe Abbildung 17<sup>218</sup>):

---

ausgegangen, dass zwar multiple Analysen von Texten möglich sind und dass dementsprechend mehrere Relationen zwischen Segmenten vorliegen können, dass man sich aber in jeder RST-Analyse für eine einzige Relation zwischen zwei Segmenten entscheiden muss, dass eine Relation die vorherrschende sei. Die Problematik dieses Paradigmas und die dem entgegenstehende Notwendigkeit multipler Relationen sprechen z.B. Moore/Pollack (1992) oder Knott et al. (2001) an. Auch die Annotationspraxis zeigt, dass die Annahme einer einzigen Relation problematisch ist: Annotatoren haben häufig Probleme bei der Auswahl einer Relation, weil oft mehrere Relationen gleichermaßen passend sind – Mann/Thompson (1988, S.245ff.) weisen darauf hin, dass Relationen selten mit Bestimmtheit (“certainty”) ausgewählt werden können, sondern dass die Auswahl stattdessen eher auf einem “plausibility judgement” basiert, auch weil es für viele RST-Relationen keine eindeutigen sprachlichen Hinweise gibt.

<sup>216</sup> Zur maschinellen Verarbeitung der thematischen Textstruktur entwickeln Lenz/Storrer (2003) ein XML-Schema zur thematischen Annotation, das auf den Formen der Themenentwicklung bei Hoffmann basiert.

<sup>217</sup> Ob der Begriff Themenentfaltung verwendet wird oder der Begriff Themenentwicklung, hängt von der Sichtweise ab: Bei der Themenentfaltung wird davon ausgegangen, dass sich der Text aus einem semantischen Kern entwickelt. Bei der Themenentwicklung stehen die Relationen zwischen den Themen zweier Sätze im Zentrum, über die die Art der Themenentwicklung charakterisiert werden kann.

<sup>218</sup> Das Beispiel wurde manuell annotiert und dient lediglich der Illustration der vorgeschlagenen Typen thematischer Organisation. Für die Annotation eines Korpus müssten Annotatoren genaue Annotationsguidelines erhalten und im Vorfeld der Annotation geschult werden. Zur Evaluation der Interrater- und Intrarater-Reliabilität müssten darüber hinaus Kappa-Werte erhoben werden.

1. Themafortführung
  - a. Themaidentität
  - b. Themaspezifizierung
  - c. Themailustration (durch ein Beispiel oder eine Grafik)
2. Themenableitung
  - a. Themazerlegung
  - b. Themareihung (Reihung im Hinblick auf ein gemeinsames Hyperthema)
3. Themaergänzung/-assoziation (basiert auf einer funktionalen oder kognitiven Relation zwischen Segmenten)

Vergleicht man diese (sowie andere in der Forschung vorgeschlagene) Formen der Themenentwicklung mit Kohärenzrelationen, stellt man fest, dass die Formen thematischer Entwicklung mehr oder weniger plausibel durch Kohärenzrelationen ersetzbar sind bzw. durch diese modelliert werden könnten. Eine besondere Rolle bei der Modellierung der Formen thematischer Entwicklung durch das Inventar der RST-Diskursrelationen spielt dabei die Relation *ELABORATION*, die für diesen Zweck allerdings in weitere Subtypen differenziert werden muss. Sie gehört zu den "subject-matter"- oder "ideational"-Relationen; Moore/Pollack (1992) verwenden auch den Terminus "informational". Eine *ELABORATION*-Relation besteht zwischen Diskurssegmenten, bei denen der Satellit zusätzliche Details über die Situation oder das Thema präsentiert, das im Nukleus präsentiert wird oder das über diesen inferentiell zugänglich ist:

S [the satellite; M.B.] presents additional detail about the situation or some element of subject matter which is presented in N [the nucleus; M.B.] or inferentially accessible in N. (Mann und Thompson 1988, 273)

One text segment expands on the other by specifying it in greater detail or specifying it in other words. (Hovy und Maier 1995, 8)

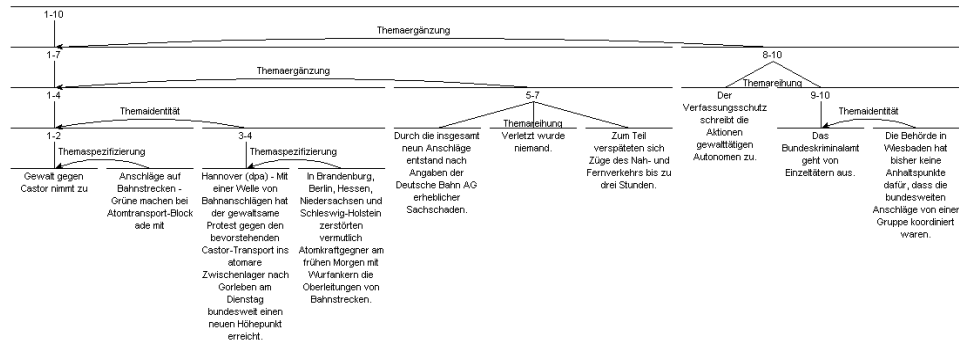


Abbildung 17: Thematische Organisation von Beispieltext 1

Die Relation ist eine der “most prevalent forms of modification of a nucleus” und “extremely common at all levels of the discourse structure” (Carlson und Marcu 2001, 33 u. 53).<sup>219</sup> Da sie prinzipiell zwischen allen aufeinander folgenden Sätzen mit einem gemeinsamen Thema bestehen kann, ist sie als eine Standard-Relation zu betrachten, die in ihrer Anwendung weniger Beschränkungen unterliegt als alle übrigen RST-Relationen. Um diese sehr weit gefasste Relation einzugrenzen bzw. in Form von Subtypen differenzieren zu können, schlagen Mann/Thompson (1988), Hovy/Maier (1995) und Carlson/Marcu (2001) verschiedene Subtypen von ELABORATION vor (siehe Tabelle 9):

<sup>219</sup> Im SemDok-Korpus deutschsprachiger wissenschaftlicher Artikel aus dem Fachgebiet Linguistik stellt ELABORATION die zweithäufigste Relation dar (ca. 25% aller Relationsinstanzen des Korpus, vgl. Lobin et al. 2010, 105).

<b>Mann/Thompson 1988</b>	<b>Hovy/Maier 1995</b>	<b>Carlson/Marcu 2001</b>
Elaboration	Elaboration	Elaboration
	ElabObject	
Object-Attribute	Object-Attribute	Object-Attribute
	Object-Function	
	ElabPart	
Set-Member	Set-Member	Set-Member
Process-Step	Process-Step	Process-Step
Whole-Part	Whole-Part	Whole-Part
	ElabGenerality	
Generalization-Specific	Genl-Specific	General-Specific
Abstract-Instance	Abstr-Instance	
		Additional
		Definition
		Example
	Identification	
Restatement	Restatement	Restatement
Summary	Summary	Summary

Tabelle 9: Übersicht über die in der Forschung vorgeschlagenen Subtypen von ELABORATION

Für die Kategorisierung und Annotation der thematischen Zusammenhänge zwischen Diskurssegmenten ist es sinnvoll, eine enger an den etablierten Formen der thematischen Organisation orientierte Spezifizierung der Subtypen von elaboration vorzunehmen. Eine entsprechende Taxonomie der ELABORATION-Relationen wurde in Bärenfänger, Goecke et al. (2008, 55ff.) und Bärenfänger, Hilbert et al. (2008, 22ff.) präsentiert. Diese Taxonomie weist allerdings Spezifika im Hinblick auf das untersuchte Korpus wissenschaftlicher Artikel auf (siehe auch Lobin et al. 2010, 104ff.); Abbildung 18 zeigt deshalb eine generalisierte

1. ELABORATION-CONTINUATION
  - a. ELABORATION-IDENTITY
  - b. ELABORATION-SPECIFICATION
  - c. ELABORATION-ILLUSTRATION
    - i. ELABORATION-EXAMPLE
    - ii. ELABORATION-DEFINITION
2. ELABORATION-VARIATION
  - a. ELABORATION-DERIVATION/ELABORATION-INTEGRATION
    - i. \*ELABORATION-SET-MEMBER
    - ii. \*ELABORATION-WHOLE-PART
    - iii. \*ELABORATION-CLASS-SUBCLASS
    - iv. \*ELABORATION-CLASS-INSTANCE
    - v. \*ELABORATION-PROCESS-STEP
  - b. ELABORATION-SEQUENCE

Abbildung 18: Subtypen von ELABORATION, die für die Annotation thematischer Zusammenhänge verwendet werden können

und überarbeitete Version dieser Taxonomie.

Die Definitionen der einzelnen Subtypen basieren dabei auf folgenden Annahmen (vgl. Bärenfänger 2008, 27): Zwischen einer im Nukleus referenzierten Diskursentität A und einer im Satelliten referenzierten Diskursentität B oder zwischen einem semantischen Prädikat des Nukleus p und einem semantischen Prädikat des Satelliten q besteht eine semantische oder ontologische Relation. Das Ziel ist nun, die Subtypen von ELABORATION anhand von möglichen Relationen zwischen A und B einerseits und p und q andererseits zu definieren. Bei einer ELABORATION-CONTINUATION-Relation wird das Thema des Nukleus im Satelliten fortgeführt. Eine Diskursentität A wird im folgenden Diskurssegment als Thema fortgeführt, indem es wiederaufgenommen wird. Neben der wiederaufgenommenen Diskursentität A enthält das Diskurssegment ein semantisches Prädikat q, das jedoch von dem semantischen Prädikat p abweichen kann. ELABORATION-CONTINUATION umfasst drei Subtypen:

- a) ELABORATION-IDENTITY: Zwischen Nukleus und Satellit besteht eine thematische bzw. referenzielle Identität, d.h. eine Diskursentität A im Nukleus wird im Satelliten wiederaufgenommen: A = B. Auch propositionale Diskursentitäten (Mikro- und Makropropositionen) können durch eine Ereignisanapher wie „dies“, oder „diese Analyse“ referenziell wieder aufgegriffen werden. Beispiel: *[Das Bundeskriminalamt]<sub>A</sub> geht von Einzeltätern aus. [Die Behörde in Wiesbaden]<sub>B</sub> hat bisher keine Anhaltspunkte dafür, dass die bundesweiten Anschläge von einer Gruppe koordiniert waren.*
- b) ELABORATION-SPECIFICATION: Die Bedeutung einer Diskursentität im Nukleus wird durch den Satelliten erweitert, eingeschränkt oder spezifiziert A = B'. Beispiel: *[Gewalt gegen Castor]<sub>A</sub> nimmt zu. [Anschläge auf Bahnstrecken]<sub>B'</sub> – (...).*
- c) ELABORATION-ILLUSTRATION: Subtypen: ELABORATION-EXAMPLE (im Satelliten wird ein Beispiel zu einer Diskursentität A oder einem semantischen Prädikat p aus dem Nukleus gegeben) und ELABORATION-DEFINITION (im Satelliten wird durch das semantische Prädikat q eine Diskursentität A aus dem Nukleus definiert). Beispiel für ELABORATION-EXAMPLE: *Zum Teil [verspäteten sich]<sub>p1</sub> Züge des Nah- und Fernverkehrs [bis zu drei Stunden]<sub>p2</sub>. Fahrgäste, die von Hamburg nach Berlin wollten, mussten beispielsweise [fast zwei Stunden warten]<sub>q</sub>.*

ELABORATION-VARIATION umfasst alle Relationen zwischen einem Nukleus und einem Satelliten, die auf einem Zerlegungs-, Ableitungs-, Über- oder Unterordnungszusammenhang basieren oder die auf ein gemeinsames Hyperthema bezogen sind. ELABORATION-DERIVATION und ELABORATION-INTEGRATION sind dabei als konverse Relationen aufzufassen. Für ELABORATION-DERIVATION und ELABORATION-INTEGRATION gilt: Im Nukleus wird eine Diskursentität A eingeführt, im Satelliten eine Diskursentität B – zwischen diesen Diskursentitäten besteht eine semantische oder ontologische Relation der Hyperonymie, Hyponymie, Meronymie, Holonymie, Partonymie o.ä. (für ein Beispiel siehe Abbildung 21). ELABORATION-SEQUENCE liegt dann vor, wenn sich die Diskursentitäten A (Nukleus) und B

Formen thematischer Organisation	Semantische Relationen	RST-Diskursrelationen
Themaforthführung/-beibehaltung	Identität Synonymie Paraphrase	ELABORATION-CONTINUATION
Themenableitung	Hyponymie Hyperonymie Holonymie Meronymie Kohyponymie	ELABORATION-VARIATION
Themaergänzung/-assoziation		PROBLEM-SOLUTION BACKGROUND CIRCUMSTANCE CAUSE RESULT CONSEQUENCE PURPOSE CONDITION CONTRAST INTERPRETATION EVALUATION ...

Tabelle 10: Zuordnung der Formen der thematischen Organisation zu semantischen Relationen und RST-Relationen

(Satellit) oder die beiden Propositionen  $p(A)$  und  $q(B)$  auf ein gemeinsames Hyperthema beziehen, d.h. entweder in einer Relation der Kohyponymie stehen ( $\text{kohyponymie}(A, B)$ ) oder dieselbe Textrolle inne haben ( $\text{TEXTROLLE}(p(A)) = \text{Textrolle}(q(B))$ ). (Beispiel: *[Durch die insgesamt neun Anschläge entstand nach Angaben der Deutsche Bahn AG erheblicher Sachschaden.]<sub>p(A)</sub>. [Verletzt wurde niemand.]<sub>q(B)</sub>*).

Tabelle 10 stellt die RST-Diskursrelationen (inkl. der beiden zentralen Subtypen von ELABORATION) und die Formen thematischer Organisation gegenüber. Dabei wird deutlich, dass die thematische Organisation nicht alleine über ELABORATION zu modellieren ist, sondern dass auch andere Diskursrelationen zur Modellierung hinzuzuziehen sind – weshalb es auch überhaupt erst gerechtfertigt ist, für die Annotation der thematischen Organisation auf Diskursrelationen Bezug zu nehmen.

Abgesehen von der Vergleichbarkeit der Relationen stellt sich die Frage, inwiefern es weitere Gemeinsamkeiten zwischen den Strukturen selbst gibt. Im Hinblick auf die thematische Struktur wurde unterschieden zwischen sequenziellen thematischen Zusammenhängen, die im Folgenden als lokale thematische Strukturen aufgefasst werden, und hierarchischen oder netzartigen thematischen Zusammenhängen, die im Folgenden als globale thematische Strukturen aufgefasst werden. Zentral für die lokale thematische Struktur ist die Verknüpfung adjazenter Segmente, wie z.B. in den Modellen von Daneš (1970) oder Hoffmann (1997) dargestellt. Bei der Analyse der relationalen Diskursstruktur nach RST werden ebenfalls adjazente Segmente verknüpft – allerdings wird Adjazenz hier nicht als



Eigenschaft auf der Textoberfläche verstanden, sondern als Eigenschaft der hierarchischen Textstruktur. Das bedeutet, dass ein Segment durch eine Kohärenzrelation nicht nur an das Vorgängersegment angeschlossen werden kann, sondern auch an eine Gruppe von Vorgängersegmenten, sofern diese ein größeres, komplexes Segment bilden (siehe Abbildung 19).

Globale thematische Zusammenhänge werden entweder als hierarchisch oder netzartig beschrieben. Hierarchisch strukturiert ist z.B. die thematische Makrostruktur nach van Dijk (1980), aber auch die relationale Diskursstruktur im Sinne der RST. Beide Ansätze gehen von einer Baumstruktur aus, die sich von einem zentralen Knoten entfaltet – in van Dijks Modell wäre dieser Knoten die Top-Level Makroproposition, für RST der dem Top-Level Nukleus übergeordnete Textknoten. Netzartige oder graphenartige Repräsentationen der globalen thematischen Struktur (z.B. bei de Beaugrande und Dressler 1981; Fritz 1982) lassen sich dagegen nicht durch RST repräsentieren, wohl aber durch andere Ansätze zur Diskursrepräsentation (z.B. SDRT – hier wird die relationale Diskursstruktur als azyklischer Graph aufgefasst, bei dem Diskurssegmente mit mehreren anderen Diskurssegmenten in einer Relation stehen können und bei dem es mehrere Relationen zwischen Diskurssegmenten geben kann; auch überkreuzende Kanten sind in diesem Ansatz erlaubt). Da es allerdings elaboriertere Ansätze zur maschinellen Identifizierung der Diskursstruktur im Sinne der RST als im Sinne von SDRT gibt, wird in der vorliegenden Arbeit aus pragmatischen Erwägungen – trotz der offensichtlichen Schwächen von RST<sup>220</sup> – RST gegenüber SDRT favorisiert. Diese Favorisierung stellt allerdings nur eine Momentaufnahme dar, die sich auf den momentanen Forschungsstand im Bereich des Diskursparsings stützt, und ist keine prinzipielle Entscheidung zugunsten von RST und zuungunsten von SDRT.

---

<sup>220</sup> RST weist mehrere theoretische Schwächen auf: In Kapitel 2.6.3.4 war schon darauf hingewiesen worden, dass RST von einem Baummodell der Textstruktur ausgeht, was aus theoretischer Sicht nicht hinreichend ist. Auch die Kontroverse „einfache vs. multiple Relationen zwischen Diskurssegmenten“ war bereits angesprochen worden: In RST wird ein Diskurssegment nur über eine Relation mit einem anderen Diskurssegment verknüpft, obwohl insbesondere die Analysepraxis zeigt, dass durchaus mehrere Typen von Relationen gleichzeitig vorliegen können. Aus beiden Annahmen – a) Diskursstruktur als hierarchische Baumstruktur, b) zwischen Diskurssegmenten besteht nur eine Diskursrelation – resultieren zwei weitere Schwächen der RST: 1. Diskurssegmente können nicht zwei oder mehr Verknüpfungen zu zwei oder mehr Diskurssegmenten aufweisen; 2. Diskurssegmente können nur mit einem in der hierarchischen Struktur adjazenten (einfachen oder komplexen) Diskurssegment in Verbindung gesetzt werden. Beide Ansichten sind sowohl theoretisch als auch praktisch nicht haltbar.

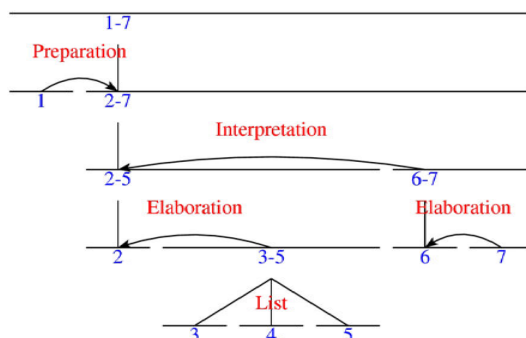
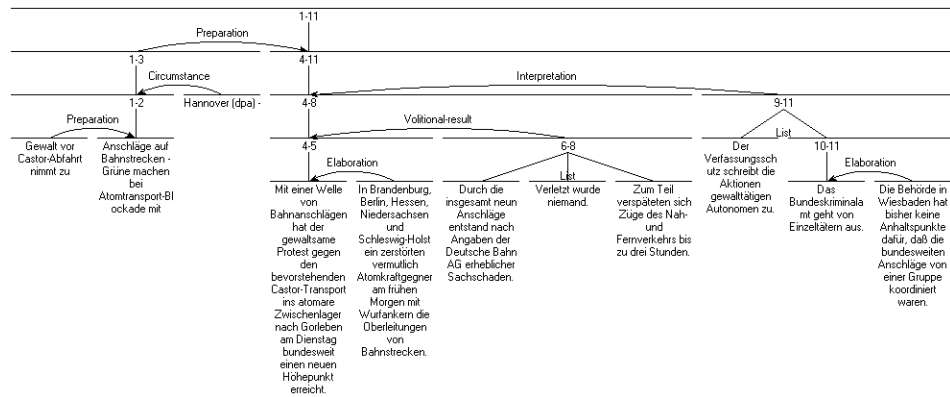


Abbildung 19: Einfacher RST-Baum  
(die Zahlen stehen jeweils für ein Textsegment)

Sowohl die Strukturmodelle, mit denen thematische und relationale Textstruktur beschrieben werden, als auch das Inventar der Relationen selbst weisen also klare Überschneidungen auf. Eine Analyse der Diskurs- oder Kohärenzrelationen kann deshalb wichtige Aufschlüsse über thematische Zusammenhänge geben, insofern thematische und relationale Struktur zumindest teilweise deckungsgleich sind. Die Zusammenhänge zwischen den Aspekten des Themas basieren darauf, dass Rezipienten Kohärenzrelationen konstruieren können bzw. verstehen, wie Aspekte lokal oder global mit anderen Aspekten bzw. dem Textthema zusammenhängen. Unterstützt werden sie hierbei zum Teil durch Diskursmarker wie Konjunktionen, deiktische Mittel, temporale oder räumliche Angaben, referenzielle, lexikalische oder semantische Verknüpfungen oder die visuelle bzw. logische Textstruktur. Generell ist allerdings festzuhalten, dass Kohärenzrelationen – und insbesondere die Subtypen von *ELABORATION* – häufig nicht an der Textoberfläche signalisiert werden und stattdessen vom Rezipienten unter Rekurs auf das Sprach- oder Sachwissen konstruiert werden müssen.

Im Folgenden soll für die beiden Beispieltexte 1 und 2 dargestellt werden, wie solche auf der RST basierenden Analysen relationaler Zusammenhänge aussehen könnten. Aufgrund der angesprochenen Schwächen von RST können diese Analysen keinen Anspruch auf Allgemeingültigkeit haben; sie stellen vielmehr mögliche Analysen dar. Zur Verbesserung der Reliabilität und Qualität der Annotationen müssten zunächst Annotationsguidelines (mit Definitionen und Beispielen) erstellt werden, die dann in einem speziellen Annotationstraining von einer Gruppe von Annotatoren angewendet und so trainiert werden müssten. Um die Reliabilität der Annotationen dieser Annotatoren zu evaluieren, müssten sowohl während als auch im Anschluss an die tatsächlichen Annotationen Inter-rater- und Intraraterreliabilitätsprüfungen durchgeführt und mittels Kappa-Werten gemessen werden.

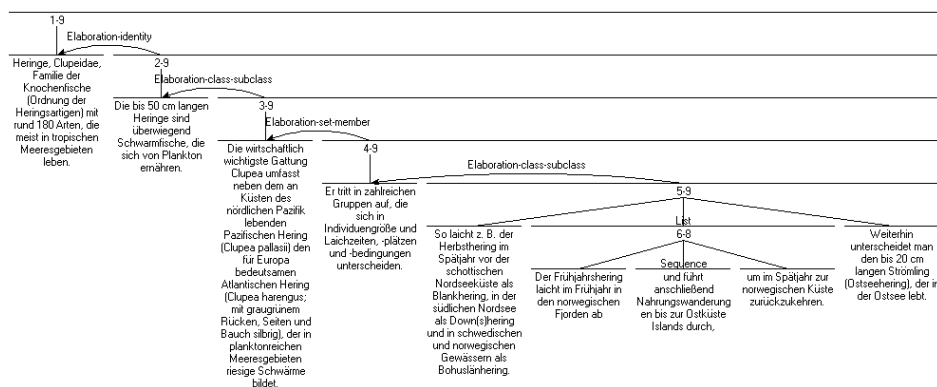


### Analyse 12: Relationale Textstruktur im Sinne der Rhetorical Structure Theory in Beispiel 1

Das Kapitel wird im Anschluss an die Beispielanalysen abschließend noch auf die Ansätze zum Diskursparsing eingehen, d.h. auf die Möglichkeiten der maschinellen Identifikation der Kohärenz- oder Diskursrelationen von Texten.

Die Relationen<sup>221</sup> zwischen den Textsegmenten in Beispiel 1 sind zum Teil funktionaler Natur („presentational relations“), zum Teil inhaltlicher Natur („subject matter relations“) (siehe Analyse 12): Die Relation PREPARATION besteht zwischen zwei Diskurssegmenten, wenn das linke der beiden Segmente die Funktion der motivierenden „Vorbereitung“ des folgenden Segments hat. Eine LIST-Relation besteht zwischen Diskurssegmenten, die in irgendeiner Form miteinander vergleichbar sind – in diesem Fall sind die drei durch LIST verknüpften Diskurssegmente Angaben des Resultats des Anschlags. Sie werden deshalb als zusammengefasstes größeres Diskurssegment über VOLITIONAL-RESULT an die vorherigen Diskurssegmente, in denen Angaben zu den Anschlügen gemacht werden, angebunden. VOLITIONAL-RESULT drückt aus, dass etwas Folge einer Situation oder eines Ereignisses ist. Im Fall des Nachrichtentextes werden die Folgen des Anschlags aufgezählt: „Sachschaden“, „Zugverspätungen“, aber „keine Verletzten“. Ein Diskurssegment stellt eine ELABORATION eines vorhergehenden Segments dar, wenn es zusätzliche Details über die Situation enthält. Elaboriert wird in Segment 4 zum einen das quasi „hyperonymisch“ gebrauchte Adjektiv „bundesweit“, zum anderen die „Bahnanschläge“ – zu diesem Ereignis werden zusätzliche Details geliefert (wer, wie, wo, wann). Die Aussagen des „Verfassungsschutzes“ und des „Bundeskriminalamts“ stellen Interpretationen des Ereignisses dar und werden deshalb über die inhaltliche Relation INTERPRETATION an den Ereignis

<sup>221</sup> Für die vorliegende Arbeit wird das erweiterte Relationsset von Mann/Thompson verwendet (1988, siehe auch <http://www.sfu.ca/rst/01intro/definitions.html>).



### Analyse 13: Reflexionen des Themas auf Ebene der relationalen Struktur (zu Beispiel 2)

niskomplex angeschlossen.

Für Beispiel 2 lässt sich feststellen, dass die Relation *ELABORATION* (speziell: *ELABORATION-CLASS-SUBCLASS*) die dominante Verknüpfungsrelation darstellt (siehe Analyse 13). Der zentrale Nukleus ("top-level nukleus") des gesamten Beispiels ist Segment 1, d.h. der erste Satz des Textes. Dieses Segment wird durch die folgenden Segmente mit zusätzlichen Details angereichert, die wiederum durch nachfolgende Segmente spezifiziert werden.

Über die in der RST-Theorie vorgenommene Differenzierung von Segmenten in Nuklei und Satelliten (siehe Mann und Thompson 1988) werden in der relationalen Analyse nicht nur Aussagen über die Relationen zwischen Diskurssegmenten (d.h. über thematische Zusammenhänge) gemacht, sondern es werden auch die zentralen, nicht weglassbaren Kerne des Textes ermittelt, die nach Ansicht einiger semantischer Thema-Konzeptionen (siehe Kapitel 2.2.2) das Thema des Textes darstellen. Der thematische Kern des Nachrichtentextes (Beispiel 1) würde damit durch die Segmente 4 und 5 reflektiert, wobei wiederum Segment 4 Nukleus zu Segment 5 ist, so dass der Satz

„Mit einer Welle von Bahnanschlägen hat der gewaltsame Protest gegen den bevorstehenden Castor-Transport ins atomare Zwischenlager nach Gorleben am Dienstag bundesweit einen neuen Höhepunkt erreicht.“

als Reflexion des thematischen Kerns des Textes charakterisiert werden könnte. Für Beispiel 1 kann dagegen der erste (elliptische Satz) als thematischer Kern gelten:

„Heringe, Clupeidae, Familie der Knochenfische (Ordnung der Heringsartigen) mit rund 180 Arten, die meist in tropischen Meeresgebieten leben.“

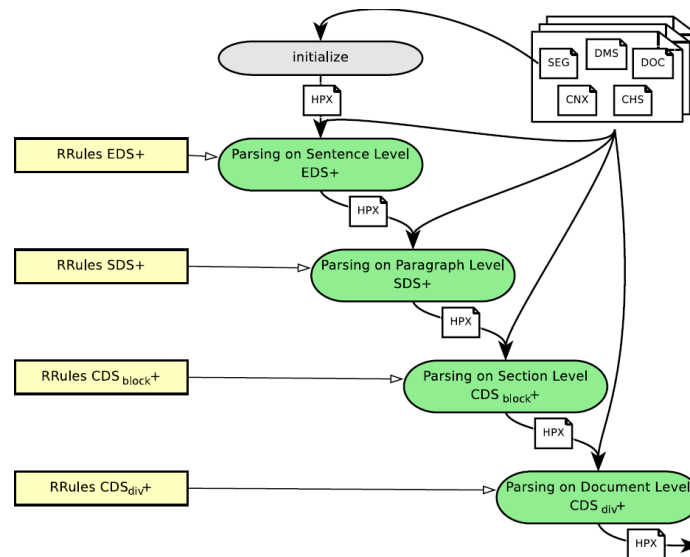


Abbildung 20: Parsing-Kaskade (aus Lobin et al. 2011, Figure 2)

Als Inventar zur Analyse der thematischen Zusammenhänge in einem Text können Kohärenzrelationen bzw. RST-Diskursrelationen insbesondere deshalb gewinnbringend eingesetzt werden, da in der Computerlinguistik eine Reihe von Ansätzen zur maschinellen Identifikation von RST-Diskursrelationen entwickelt wurden (für einen Überblick über verschiedene Ansätze zum Diskursparsing siehe Lobin et al. 2011). Der Forschungsbereich, der sich hiermit beschäftigt, ist der des Diskursparsings. Hierbei gibt es zwei Kernaufgaben: 1. Identifizierung der Diskurssegmente, zwischen denen eine Relation hergestellt werden kann; 2. Identifizierung des Relationstyps zwischen diesen Segmenten. Diese Aufgaben können allerdings nicht nacheinander bearbeitet werden, sondern sie sind miteinander verschränkt.

Beim Diskursparsing nach dem Modell der RST geht es darum, einfache Diskurssegmente zu komplexen Diskurssegmenten zu kombinieren, und zwar soweit, bis schließlich das gesamte Dokument als ein komplexes Diskurssegment repräsentiert werden kann. Bärenfänger et al. (2006; 2010, 88f.) unterscheiden aus diesem Grund Diskurssegmente verschiedener Größe bzw. Diskurssegmente verschiedener Komplexitätsgrade: Elementare Diskurssegmente ("elementary discourse segments", EDS),<sup>222</sup> satzbasierte Diskurssegmente ("sentential dis-

<sup>222</sup> Elementare und satzbasierte Diskurssegmente (EDS und SDS) können von einer externen Analysekomponente in einem Vorverarbeitungsschritt maschinell identifiziert und annotiert werden und dem Diskursparser dementsprechend als Input zur Verfügung gestellt werden (siehe Lünge et al. 2006). CDS müssen dagegen vom Diskursparser selbst konstruiert werden.

course segments”, SDS), komplexe Diskurssegmente auf der sog. Block-Ebene (“complex discourse segments on block level”, CDS<sub>block</sub>), d.h. auf der Ebene von Absätzen und Absatz-ähnlichen Segmenten wie z.B. Tabellen oder Grafiken, komplexe Diskurssegmente vom Typ “division” (Kapitel u.ä., CDS<sub>div</sub>), komplexe Diskurssegmente auf der Dokumentebene (CDS<sub>doc</sub>). Der in Lobin et al. (2011) vorgestellte Diskursparser verarbeitet die verschiedenen Segmentebenen dabei in Form einer Verarbeitungskaskade, bei der das Input-Dokument zunächst auf der EDS-, dann der SDS- und schließlich der CDS-Ebene verarbeitet wird (siehe Abbildung 20).

Für die Identifizierung der Relationen zwischen zwei oder mehr Diskurssegmenten (bei multinuklearen Segmentverknüpfungen können auch mehr als zwei Diskurssegmente über eine einzige Relation miteinander verknüpft werden) muss zunächst – auf Basis existierender RST Relationen (z.B. von Mann und Thompson 1988; Hovy und Maier 1995; Carlson und Marcu 2001) – ein Relationsinventar entwickelt werden, das für den zu verarbeitenden Texttyp geeignet ist. Diese Relationen können in einer Relationstaxonomie hierarchisch geordnet werden, indem sie über hierarchische Über- und Unterordnungsbeziehungen verknüpft werden. Bärenfänger, Hilbert et al. (2008, 17ff.) stellen eine solche Taxonomie vor und repräsentieren diese als OWL-Ontologie, in der neben den normalen RST-Relationen auch stark unterspezifizierte Relationen wie MONO-NUCLEARRELATION enthalten sind sowie unterspezifizierte Superrelationen bzw. künstliche Relationen wie INTERPRETATIONEVALUATION. Der Vorteil einer solchen Ontologie mit unterspezifizierten Relationen ist, dass im Falle konkurrierender oder nicht eindeutiger Diskursmarker, die für das Vorliegen sowohl von INTERPRETATION als auch EVALUATION sprechen, der Diskursparser eine unterspezifizierte Relation zuweisen kann, die beide Hypothesen im Parsing-Prozess kombiniert (vgl. Bärenfänger, Hilbert et al. 2008, 17ff.).

Zum Diskursparsing verwendet der in Gießen entwickelte und in Lobin et al. (2011) vorgestellte Diskursparser drei Input-Komponenten: 1. Wissensressourcen, 2. eigenständige Analysekomponenten und daraus resultierende Annotationsschichten sowie 3. eine Menge deklarativer Regeln. Zu den verwendeten Wissensressourcen gehören sowohl lexikalisch-semantische Ressourcen (das Diskursmarkerlexikon sowie das lexikalisch-semantische Wortnetz GermaNet) als auch konzeptuelle Ressourcen (eine Texttypstrukturontologie (siehe Kapitel 3.2.5) sowie die RST-Relationsontologie (s.o.)). Diese Wissensressourcen werden dabei (mit Ausnahme der RST-Relationsontologie/-taxonomie) nicht vom Diskursparser selbst (d.h. zur Laufzeit des Parsers) verwendet, sondern von externen und vom Parser unabhängigen Analysekomponenten. Diese Komponenten erzeugen in einem Vorverarbeitungsschritt Multi-Ebenen-Annotationen für die unterschiedlichen Ebenen der Textstruktur, und zwar auf der Ebene der logischen Dokumentstruktur, der morphologischen, syntaktischen und anaphorischen Struktur sowie Annotationen, bei denen lexikalische Diskursmarker und elementare sowie satzbasierte Diskurssegmente ausgezeichnet werden (zum Thema „Multi-Ebenen-Annotation“ siehe Kapitel 2.6.3, insbesondere S. 86ff.). Diese unterschiedlichen, als XML-basierte Multi-Ebenen-Annotationen realisierten Annotationsschichten fungieren als direkter Input für den Diskursparser.

Ihre Bedeutung für die Identifizierung der relationalen Diskursstruktur im Sinne der RST wird dabei in einer Menge deklarativer Regeln festgehalten, die als Constraints für den Diskursparser fungieren: Sie leiten den Diskursparser, indem sie Aussagen zu Merkmalen oder Merkmalskombinationen verschiedener Annotationsschichten (d.h. Merkmalen auf verschiedenen Ebenen der Textstruktur) als Constraints für die Kombinierung von Diskurssegmenten sowie für die Zuweisung einer bestimmten Relation nutzen. Die Regeln stellen dabei auf einer abstrakten Ebene „wenn-dann“-Bedingungen dar:

*wenn*

Merkmal x in Diskurssegment a vorhanden  
und  
Merkmal y in Diskurssegment a vorhanden  
und  
Merkmal z in Diskurssegment b vorhanden,

*dann*

liegt mit einer Wahrscheinlichkeit von n die Diskursrelation r zwischen Diskurssegment a und b vor

Die Regeln kodieren das Wissen über die sprachlichen Merkmale, die als potentielle indexikalische Zeichen für eine bestimmte Diskursrelation interpretiert werden können. Es basiert in der Regel auf den Ergebnissen empirischer Korpusanalysen. Formal haben die deklarativen Regeln dabei den Status von Reduce Rules. Sie steuern, gemeinsam mit der RST-Relationstaxonomie/-ontologie, den als bottom-up Chart-Parsing realisierten Parsing-Prozess (vgl. Lobin et al. 2011).

In Bärenfänger et al. (2006; 2010), Bärenfänger, Hilbert et al. (2008) und Bärenfänger, Goecke et al. (2008) wurde insbesondere untersucht, welche sprachlichen Merkmale für die maschinelle Identifizierung der Subtypen von ELABORATION genutzt werden können. Im Gegensatz zu anderen RST-Relationen wird ELABORATION in der Regel nicht durch lexikalische Diskursmarker signalisiert. Für die maschinelle Analyse mussten deshalb andere Merkmale ermittelt werden. Zu diesen gehören die Korrelationen bestimmter Subtypen von ELABORATION mit bestimmten semantischen und anaphorischen Relationen. ELABORATION-DERIVATION und ELABORATION-INTEGRATION beispielsweise korrelieren in der Regel mit einer semantischen Relation wie Hyperonymie, Hyponymie, Holonymie oder Meronymie (ein Beispiel für den Zusammenhang von ELABORATION-DERIVATION mit der semantischen Relation der Holonymie zeigt Abbildung 21). ELABORATION-CONTINUATION wiederum korreliert fast immer mit einer anaphorischen Relation, bei der die Anapher entweder syntaktisch als Pronomen realisiert ist (d.h. zwischen Anapher und Antezedent eine Relation der Koreferenz besteht) oder bei der das Anapher-Antezedent-Paar ein identisches Kopfnomen aufweist. Eine ausführliche Darstellung der Ergebnisse der Korpusstudie und die Möglichkeiten, anaphorische Relationen verschiedenen Typs als Indikatoren für verschiedene Subtypen von ELABORATION zu verwenden, findet sich in Bärenfänger, Goecke et al. (2008, 63ff.).

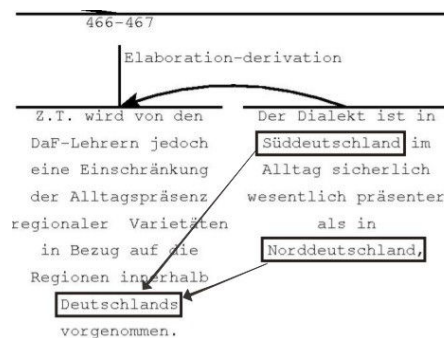


Abbildung 21: Zum Zusammenhang von ELABORATION-DERIVATION und der semantischen Relation der Holonymie (Holonymie als Cue für ELABORATION) (aus Lobin et al. 2010, 108)

Auf Basis der verschiedenen Annotationsschichten (“annotation layer 1-n”) sowie der Menge der Reduce-Regeln (“reduce rule set”) erzeugt der Diskursparser in kaskadierten Verarbeitungsschritten (siehe Abbildung 20) schließlich eine neue Annotationsschicht (“annotation layer n+1”): die Ebene der relationalen Diskursstruktur im Sinne der RST. Eine Übersicht über die Architektur des in Gießen (siehe Lobin et al. 2011) vorgestellten Diskursparsers gibt Abbildung 22.

### 3.3.2.2 Texttypstrukturkategorien

Die globale Texttypstruktur ist durch die mehr oder weniger konventionalisierte Abfolge spezifischer Texttypstrukturkategorien geprägt. Diese Texttypstrukturkategorien lassen sich auch als unterspezifizierte Themen bzw. abstrakte Thementypen des Textes auffassen. Je nach Grad der Konventionalisierung der Kategorien und ihrer Abfolge weist jedes Exemplar eines Texttyps dasselbe unterspezifizierte globale thematische Schema auf. Dieses bildet das strukturelle Skelett des Textes, das dann im Akt der Textproduktion spezifiziert und abgewandelt werden kann. Die Texttypstruktur kann demnach auch als Thementypenstruktur beschrieben werden, die die Themenentwicklung prägt (Produzentsicht). Die Kohärenz auf dieser Ebene basiert auf der möglichst umfassenden Erfüllung der Erwartungen des Rezipienten an prototypische Texttypkategorien bzw. Thementypen und deren Abfolge (Rezipientensicht). Die mit der Herstellung der Texttypstruktur verbundene Kohärenz kann deshalb als unterspezifizierte thematische Kohärenz bezeichnet werden.<sup>223</sup>

<sup>223</sup> Die Auffassung von Texttypstruktur-Schemata als unterspezifizierten Thementypen-Schemata erfüllt die Forderung von Asher (2004b) nach einer neuen Theorie zum Thema: Nach Asher ist ein wichtiges Merkmal von Themen, dass sie im Hinblick auf ihren Inhalt oft unterspezifiziert sind. Deshalb fordert er eine Theorie, “that gives us constraints on what the topic can be rather



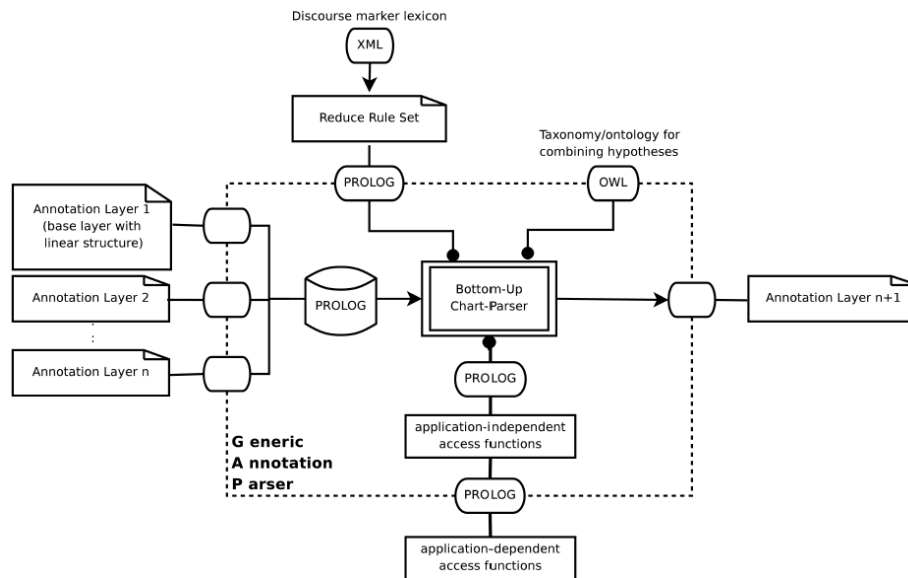


Abbildung 22: Architektur des in Gießen realisierten Diskursparsers  
(aus Lobin et al. 2011, Figure 4)

Für die in dieser Arbeit fokussierte Textklasse der expositorischen Texte ist davon auszugehen, dass die Aspekte des Themas als Instanzen prototypischer Texttypstrukturkategorien bzw. abstrakter Thementypen aufgefasst werden können. Im Hinblick auf den Zusammenhang der Aspekte des Themas können Texttypstrukturschemata demnach dazu verwendet werden, die in einem Text instanziierten Thementypen (als Aspekte des Themas) in einen kategorialen und teilweise sequenziellen Zusammenhang zu stellen. Diese These soll an Beispieltext 1 illustriert werden.

---

than specifying exactly one topic. A constraint based approach is one that allows naturally for underspecification of the topic" (Asher 2004b, 256).

Texttypspezifische Thementypen		Konkrete Themeninstanzen
Was?	EREIGNIS	Anschläge auf Bahnstrecken (2), (4), (6)
Wo?	ORT	bundesweit (4)   in Brandenburg, Berlin, Hessen, Niedersachsen und Schleswig-Holstein (5)
Wann?	ZEIT	am Dienstag (4); am frühen Morgen (5)
Wer?	PERSON	Atomkraftgegner (5)
Warum?	URSACHE/ZWECK	Atomtransport-Blockade (2)
Wie?	MITTEL	Gewalt (1); Zerstörung von Oberleitungen von Bahnstrecken mit Wurfankern (5)
	HINTERGRUND	„Protest gegen den bevorstehenden Castor-Transport ins atomare Zwischenlager nach Gorleben“ (4)
	EINORDNUNG	„hat der gewaltsame Protest (...) bundesweit einen neuen Höhepunkt erreicht“ (4)
	KONSEQUENZEN/FOLGEN	Sachschaden (6) „Verletzt wurde niemand.“ (7) Zugverspätungen (8)
	(VERBALE) REAKTIONEN	„Der Verfassungsschutz schreibt die Aktionen gewalttätigen Autonomen zu.“ (9) „Das Bundeskriminalamt geht von Einzeltätern aus.“ (10)

Analyse 14: Aspekte des Themas als texttypspezifische Thementypen in Beispiel 1

Die Instanzierung der texttypspezifischen Thementypen *Ereignis*, *Ort*, *Zeit*, *Person*, *Ursache/Zweck* und *Mittel* im Nachrichtentext (siehe Kapitel 2.6.3.6) stellt einen zentralen Schritt im Prozess der Textherstellung dar und ist Teil des Thematisierens als sprachlicher Handlung. Für Nachrichtentexte gilt, dass in der Regel bereits im Lead bzw. Vorspann die ersten drei Fragen kurz beantwortet werden und im Folgenden weitere Details nachgeliefert werden, insbesondere zu weiteren Themenklassen wie *Hintergrund*, *Einordnung*, *vorangegangene Ereignisse*, *Konsequenzen*, *(verbale) Reaktionen* (s.o.). Insofern in Nachrichtentexten die „Was?“-Frage (zum Teil auch die „Wer?“-Frage) als zentrale Textfrage gelten kann, muss sie (bzw. ihre Beantwortung) als Hauptthema des Textes aufgefasst werden; die anderen Thementypen repräsentieren dagegen Aspekte des Themas.

Auch Enzyklopädieartikel weisen nach Hoffmann (1988, 165f.) eine typische Thementypenstruktur auf. Zu den grundlegenden Bestandteilen von Enzyklopädieartikeln gehören in der Regel *Stichwort*, *Definition*, *Merkmal 1*, *Merkmal 2-n*. Sowohl Hoffmann als auch Nuopponen (1993, 103ff.) weisen allerdings darauf hin, dass die globale Textstruktur bei Enzyklopädieartikeln stark an der Begriffsstruktur orientiert ist, so dass sich für unterschiedliche Typen von Begriffen auch unterschiedliche Texttypstrukturen bzw. Thementypenstrukturen ergeben. Für Beispieltext 2 lässt sich folgende Thementypenstruktur ermitteln:

Texttypspezifische Thementypen		Konkrete Themeninstanzen
Stichwort Alternative Namen/Synonyme Klassifikatorische Einordnung		Hering Clupeida Familie der Knochenfische
Arten/Unterbegriffe, ihre Eigenschaften und Unterschiede		
	Arten	Pazifischer Hering Atlantischer Hering Ostseehering Weitere 180 Arten
	Alternative Namen/ Synonyme	Pazifischer Hering → Clupea pallasii Atlantischer Hering → Clupea harengus Ostseehering → Strömling
	Form- und Wahrnehmungseigenschaften	mit graugrünem Rücken, Seiten und Bauch silbrig
	Sonstige Eigenschaften (z.B. Laichzeiten und -plätze)	der Herbsthering laicht im Spätjahr vor der schottischen Nordseeküste als Blankhering ab ...
Existenzphasen		Ein Weibchen des Atlantischen Herings legt rund 20 000 bis 70 000 Eier ab, die auf den Meeresgrund sinken und dort an Steinen, Algen oder Ähnlichem festkleben. Die schlüpfenden Larven steigen an die Oberfläche, sie machen ihre Jugendentwicklung im Küstenbereich durch. Erst die mit zwei bis drei Jahren etwa 20 cm langen Jungheringe wandern von der Küste ab. ...
Bedeutung für den Menschen		
	Wirtschaft	Heringe, Sardinen und die ihnen nahe verwandten Sardellenarten sind die wirtschaftlich wichtigsten Objekte der Meeresfischerei. Sie stehen als Planktonfresser an der Basis der Nahrungskette, was ihre hohe produktionsbiologische Leistung bedingt. Der Weltfangertrag des Atlantischen Herings betrug (2002) 1,87 Mio. t. ...
	Geschichtliches	Heringsfischerei wurde schon im 7. Jahrhundert in Europa betrieben. Um 900 war der gesalzene Hering bereits im Handel begehrt. ...

#### Analyse 15: Aspekte des Themas als texttypspezifische Thementypen in Beispiel 2

Bei der Betrachtung verschiedener Exemplare dieser (und auch anderer) Textsorten zeigt sich, dass es eine starke Heterogenität in der globalen Texttypstruktur gibt. Hoffmann zufolge beruht diese Heterogenität auf dem Unterschied der behandelten Gegenstände:

Die Makrostruktur des Fachtextes entspringt und entspricht in hohem Maße der Struktur des in ihm behandelten Gegenstandes, aus der sich eine Art ‚innere Logik‘ ergibt, der sich letztlich auch die Absicht des Textverfassers unterzuordnen hat. (Hoffmann 1988, 167)

Während bisher davon ausgegangen wurde, dass die im Text behandelten Aspekte eines Themas u.a. aus den texttypspezifischen Thementypen abgeleitet werden können, erscheint es demnach andererseits plausibel, dass auch der Typus eines Redegegenstands, d.h. der Thementyp, unabhängig vom Texttyp abstrakte Vorgaben über die möglichen zu thematisierenden Aspekte des Themas macht, so dass auf einer abstrakten Ebene für jeden Thementyp die Menge der potentiellen Teilthemen bzw. Aspekte des Themas beschrieben werden könnte.

Zu diesem Zweck ist es hilfreich, die Gegenstände, über die im Text etwas gesagt werden kann, in Typen zu unterteilen (wie dies z.B. auch bei der Erstellung von Top-Level-Ontologien getan wird). Adamzik schlägt hierzu eine Unterteilung in drei Typen von Entitäten bzw. Themen vor: „Statische Objekte“ (unbelebte Dinge bzw. Gegenstände, Lebewesen, Zustände/Situationen), „dynamische Objekte“ (Ereignisse) und „kognitive Objekte“ (Begriffe, Kategorien, Propositionen The- sen, Theorien) (vgl. Adamzik 2004b, 123). Wie Schank ist auch Adamzik der Ansicht, dass es aufgrund des logisch-ontologischen Status dieser Thementypen möglich ist, die Menge der möglichen Teilthemen einzugrenzen:

Bei einem gegebenen Thementyp [ist] die inhaltliche Ausführung auf einem sehr abstrakten Niveau voraussehbar (...) und (...) es [ist] eine durchaus überschaubare Menge von Inhaltsaspekten (...), die dabei zur Sprache kommt. (Adamzik 2004b, 125)

Bei ihrem Versuch der Beschreibung typischer Themenkonstellationen stützt sich Adamzik zum einen auf die bei de Beaugrande/Dressler (1981) definierten Gruppen von „Primärkonzepten“ (Objekt, Situation, Ereignis, Handlung) und „Sekundärkonzepten“ (z.B. Grund, Ermöglichung, Spezifizierung, Modalität)<sup>224</sup>, zum anderen auf die von Polenz (1985) beschriebenen semantischen Rollen. Mit Hilfe dieses Inventars lässt sich ihrer Ansicht nach für jeden Thementyp bestimmen, wie dieser Thementyp in einem Text inhaltlich ausgeformt werden kann bzw. welche Aspekte eines Themas denkbar sind:

Statische Objekte haben *Eigenschaften*, bestehen aus *Einzelteilen* oder sind selbst *Bestandteil* eines größeren Ganzen. Sie befinden sich zu einer bestimmten *Zeit* an einem bestimmten *Ort* und lassen sich also situieren. Der Beantwortung entsprechender Fragen (Welche Eigenschaften hat das Objekt?; Aus welchen Teilen besteht es?; Wo befindet es sich? usw.) sind Teile eines Textes gewidmet, der nur statische Objekte zum Thema hat. (Adamzik 2004b, 124, Hervorhebungen im Original)

Eine ähnliche Annahme verfolgt bereits Schank (1977b). Ihm zufolge unterliegen die Anschlussmöglichkeiten an eine Äußerung, den Regeln der gedanklichen Assoziation (“rules of associations in thought”):

The rules that determine what one can reasonably say in response to an input are a subset of those rules that determine what one can possibly think in response to an input. The exception to this is that some selection rules must be employed to choose the appropriate association given the conversational circumstances. (Schank 1977b, 429)

---

<sup>224</sup> Primärkonzepte stellen die Knoten eines konzeptuellen Netzes dar, Sekundärkonzepte die Relationen (vgl. de Beaugrande und Dressler 1981, 101f.).

Schank unterteilt die „Input-Äußerungen“ in zwei Klassen: solche Äußerungen, in denen ein „Objekt“ thematisiert wird, auf das man sich in einer Antwort beziehen kann, und solche, bei denen eine „Handlung“ thematisiert wird. Was als Antwort auf den auf einen Objekt fokussierten Input gesagt werden kann, ist nach Schank determiniert durch die Eigenschaften des Objekts: Was als Antwort auf einen Satz gesagt werden kann, der ein Objekt zum Gegenstand hat, ist abgeleitet aus den assoziativen Anschlüssen an dieses Objekt. Gleiches gilt für Sätze, die eine Handlung zum Gegenstand haben. Die Regeln, die dies beschreiben, werden als „syntactic association rules“ bezeichnet. Beispiele für Anschlussmöglichkeiten an einen Objekt-Input sind z.B. *Herkunft* („Where did you get a Y?“), *Funktion* („What do you do with a Y?“), *assoziierte Objekte* (Input: „I just got a Corvette“ – Anschluss: „I once had a Jaguar“). Neben diesen syntaktischen Assoziationsregeln sind allerdings kontextgebundene Auswahlregeln nötig, die vorgeben, was in einer spezifischen Situation ein angemessener Anschluss ist. Diese Regeln sind kulturspezifisch, abhängig von den konkreten Umständen sowie abhängig von dem Wissen, das man über die Person hat, die die Input-Äußerung macht, etc. Mit Hilfe beider Typen von Regeln lassen sich Schank zufolge einige allgemeine Standard-Anschlussmöglichkeiten bestimmen, die in Form von „Conversational/Associational Categories (CACs)“ repräsentiert werden können (vgl. Schank 1977b, 431f.) (siehe Tabelle 11).

Nach Schank beschränkt also der Typ des aktuellen Themas („Objekt“ vs. „Handlung“) die Auswahl möglicher Teilthementypen. Die Beschränkungen wirken sich dabei jedoch nicht nur auf der Ebene des abstrakten Thementyps aus, sondern zeigen sich auch auf der Ebene der konkreten Objekte: So kann bei dem Thema „Kamera“ z.B. über die Funktion, den Preis, mit der Kamera gemachte Fotos etc. geredet werden. Was ein möglicher Aspekt eines Themas oder ein mögliches Anschluss Thema für ein thematisiertes Objekt ist, ergibt sich demnach – wie bereits erwähnt – durch die im kognitiven Schema/Frame gespeicherten Relationen dieses Objekts zu anderen Objekten (siehe auch Kapitel 3.4.1 und 3.2.1).

CAMERA (object)	take pictures	FUNCTION
	picture	RESULT
	middle	PRICE
	nil	ENABLE
	nil	OWNER
	photographer	USER
	easy	AVAILABILITY
BUY (action)	nil <sup>225</sup>	PROBLEM
	shopping	SCRIPT
	nil	ASSOC
	own	RESULT
	nil	ACTOR
	nil	EFFECT
	nil	OBJ
	high	FREQUENCY
	2	UNUSUAL

Tabelle 11: Beispiel für CAC's<sup>226</sup>

Eine eigene, exemplarische Vorstudie, bei der Texte verschiedener Texttypen zum Thema „Waldsterben“ analysiert wurden, um so einen ersten Einblick zu bekommen, ob Thementypkonstellationen unabhängig von Texttypen sind (siehe Analyse 16), stützt die Annahmen von Adamzik (2004b) und Schank (1977b).

<b>Texttyp</b>	<b>Im Text behandelte Aspekte des Themas</b>
Lexikonartikel I (deskriptiver Text)	Entwicklung der Waldschäden/Chronik Merkmale/Symptome Ausmaß der Schäden Dokumentation der Schäden Ursachen Auswirkungen/Folgen Gegenmaßnahmen
Lexikonartikel II (deskriptiver Text)	Merkmale/Symptome Ursachen Auswirkungen/Folgen Ausmaß der Schäden Dokumentation der Schäden
Zeitungsbericht (expositorischer Text)	Ursache/Verursacher Waldsterben in anderen Ländern Auswirkungen/Folgen Bewertung der Politik Gegenmaßnahmen
Pressemeldung I (appellativ-expositorischer Text)	Ausmaß der Schäden Chronik des Waldsterbens Ursache Gegenmaßnahmen

<sup>225</sup> “‘Nil’ indicates that there is no interest of the subject in this category for this object.” (Schank 1977b, 432)

<sup>226</sup> Beispiele entnommen aus Schank (1977, S.433 und S.438).

	Dokumentation der Schäden Bewertung der Politik
Pressemeldung II (appellativ-expositorischer Text)	Bewertung der Politik Gegenmaßnahmen Ausmaß der Schäden Chronik des Waldsterbens Zukunftsprognose über Auswirkungen/Folgen Symptome Ursachen
Offener Brief an Politiker (appellativer Text)	Ausmaß der Schäden Zukunftsprognose über Auswirkungen/Folgen Ursachen Gegenmaßnahmen

Analyse 16: Überblick über die Ergebnisse einer Analyse der Aspekte des Themas von Texten verschiedener Texttypen zum Thema „Waldsterben“<sup>227</sup>

Das Textthema und seine Aspekte werden je nach Texttyp unterschiedlich behandelt. Unterschiede gibt es z.B. hinsichtlich der Sequenzierung der Teilthemen. In Textsorten der Pressekommunikation werden die Teilthemen fast immer vollständig in Titel, Schlagzeile und 1. Absatz eingeführt und dann nacheinander in den folgenden Absätzen spezifiziert. In Lexikonartikeln werden die Teilthemen dagegen nacheinander entfaltet, ebenso wie in Briefen. Unabhängig vom Texttyp werden allerdings in nahezu allen Texten die gleichen Aspekte des Themas „Waldsterben“ behandelt: *Entwicklung der Waldschäden/Chronik, Merkmale/Symptome, Ausmaß der Schäden, Dokumentation der Schäden, Ursachen, Auswirkungen/Folgen, Gegenmaßnahmen*. In den Textsorten der Pressekommunikation wird darüber hinaus das *Verhalten* der Politik/Regierung bewertet und damit als Aspekt, unter dem das Thema behandelt werden kann, eingeführt.

Problematisch für ein solches Vorhersagemodell für Thementypkonstellationen sind Texte, die ein multiples Thema zum Gegenstand haben, oder Texte, in denen Themen durch sprachliche Mittel explizit miteinander verknüpft werden, die keinen Zusammenhang auf der Ebene der kognitiven Sachverhaltsstruktur aufweisen. Die Möglichkeiten solcher expliziten Themenverknüpfungen sind potentiell unbegrenzt. Ein Vorhersagemodell für Thementypkonstellationen kann daher nur als modulares Modell funktionieren, bei dem fortlaufend zu jedem im Text behandelten Aspekt des Themas in Abhängigkeit von seinem Typus wieder neue Vorhersagen über die potentiellen im Text folgenden Thementypen gebildet werden.

Der Forschungsbereich der maschinellen Identifizierung der Texttypstruktur ist zu trennen vom Forschungsbereich der maschinellen Textsorten- oder Genre-Identifikation insgesamt. Die Arbeiten, die sich mit dem Bereich der Texttypstrukturidentifizierung befassen, unterscheiden sich im Hinblick auf ihre Terminologie, ihren Gegenstand, d.h. die untersuchten Texttypen bzw. Textsorten,

<sup>227</sup> Textsammlung „Waldsterben“ – Quellen siehe Kapitel 5.2.

sowie ihre Methodik. Neben dem Begriff der Texttypstruktur (Langer et al. 2004a, b) werden als alternative Begriffe zur Beschreibung der kategorialen globalen Textstruktur und ihrer grundlegenden Einheiten auch die Begriffe generische Inhaltsstruktur (Stede und Suriyawongkul 2010), generische Textstruktur (van Mol und O'Donnell 2004), globale Textstruktur (Kando 1997, 1999), thematische Textstruktur (Farzindar und Lapalme 2004), „Inhaltszonen“ (Stede 2007, 39f. u. 77), „formale Zonen“ (Bieler et al. 2007) oder „argumentative Zonen“ (Teufel 1999; Teufel et al. 1999; Teufel und Moens 2002) verwendet. Den genannten Arbeiten ist gemeinsam, dass sie die für eine bestimmte Textsorte typischen funktionalen, inhaltlichen oder argumentativen Kategorien, die im Text typischerweise an einer bestimmten Stelle vorkommen, in Form von Schemata modellieren und gleichzeitig den Versuch unternehmen, diese Kategorien mittels computerlinguistischer Methoden maschinell zu identifizieren. In den letzten Jahren wurde dabei eine Reihe sehr verschiedener Textsorten in den Blick genommen: wissenschaftliche Artikel (siehe Kando 1997, 1999; Teufel 1999; Teufel et al. 1999; Teufel und Moens 2002; Langer et al. 2004a, b), juristische Texte (Farzindar und Lapalme 2004), insbesondere Gerichtsentscheidungen (Moens 2007; Stede und Kuhn 2009), Filmrezensionen (Bieler et al. 2007; Stede und Suriyawongkul 2010), medizinische Begleitbriefe („medical discharge notices“) (van Mol und O'Donnell 2004), institutionelle und persönliche Homepages (Rehm 2005) u.v.m. Die automatische Identifizierung der kategorialen Einheiten der Texttypstruktur ist dabei als Klassifikationsaufgabe zu betrachten, bei der den einzelnen Einheiten des Textes ein Label (eine Texttypstrukturkategorie, ein Thementyp, eine texttypspezifische Inhaltszone, formale oder argumentative Zone etc.) zugewiesen wird. Hierzu werden in der Regel maschinelle Lernverfahren eingesetzt. Die Güte der Ergebnisse ist dabei in einem hohen Maße vom Korpus und seinen (manuellen und maschinellen) Vorverarbeitungen abhängig.

### 3.3.3 Zur „Verrechnung“ der Reflexionen des Themas

Die verschiedenen Reflexionen des Themas im Text sind – jede für sich genommen – nur sehr schwache und weder hinreichende noch notwendige Marker für das Thema als Sachverhaltskomplex. Ein Rezipient ist deshalb zur Identifizierung des Themas darauf angewiesen, dass er einerseits die einzelnen Reflexionen miteinander „verrechnet“<sup>228</sup> und dass er andererseits einen Bezug zu seinen Wissensbeständen herstellt und dabei die miteinander „verrechneten“ Reflexionen als Mittel zur Aktivierung und Bezugnahme auf ein bestimmtes Situationsmodell bzw. einen Sachverhaltskomplex verwendet. Die Reflexionen des Themas im Text können deshalb auch als indexikalische Zeichen aufgefasst werden, die ein Kommunikant als Anzeichen für die Orientierung des Textes an einem Thema als Sachverhaltskomplex interpretieren kann (aber nicht muss).

Alle wahrnehmbaren – und möglicherweise auch bestimmte fehlende – Phänomene können demnach potentiell als indexikalisches Zeichen interpretiert werden, um auf das Thema eines Textes zu schließen. Allerdings ist hierbei folgen-

---

<sup>228</sup> Zur Bedeutung des Ausdrucks „verrechnen“ siehe Fußnote 154, S. 131.



des zu berücksichtigen:

Doch wie die Zeichnung mehr ist als die Summe der Punkte, so ist auch die Äußerung/der Text mehr als eine Folge von Morphemen oder Wörtern oder auch als eine Folge von Sätzen – arrangiert nach den Gesetzlichkeiten einer Textsorte. Die Leistungen der Elemente addieren sich nicht zu einer Gesamtleistung, sondern: aus der Kombination der Elemente ergeben sich neue Leistungen, die von den Teilen (als Teilen) nicht erbracht werden können. (Ortner und Sitta 2003, 17)

Es ist demnach dringend nötig, die einzelnen Reflexionen des Themas miteinander zu „verrechnen“, um so von den einfachen indexikalischen Zeichen zu den komplexen und bedeutungskonstituierenden indexikalischen Zeichen zu gelangen. Hierzu müssen Kommunikanten die Menge der einfachen indexikalischen Zeichen in eine verständnisrelevante Ordnung bringen, d.h. selektieren, gewichten und Zusammenhänge herstellen, um aus den unterschiedlichen potentiell bedeutsamen indexikalischen Zeichen überhaupt einen sinnvollen Bezug auf das Thema als Sachverhaltskomplex konstruieren zu können. Dabei werden diejenigen indexikalischen Zeichen, für die Kommunikanten erkennen, dass sie in einer Kontiguitätsbeziehungen stehen, die bei der Behandlung eines Themas (als Sachverhaltskomplex) regelmäßig gemeinsam verwendet werden, privilegiert behandelt und in der Interpretation ihres Zusammenspiels für einen bedeutungskonstituierenden Schluss genutzt. Rezipienten bauen bereits nach den ersten konsummierten sprachlichen Einheiten Hypothesen auf, worum es im Text geht, an welche bekannte kommunikative Situation der Text anschließbar ist. Es ist deshalb davon auszugehen, dass es eine auf der Sequenz der sprachlichen Einheiten basierende Hierarchie indexikalischer Zeichen gibt, d.h. dass sowohl die ersten Äußerungseinheiten des Textes (zu denen u.a. der Titel zählt) als auch die ersten Einheiten des Absatzes eine größere Bedeutung für die Aktivierung eines Frames aufweisen, als Äußerungseinheiten, die in der Mitte des Textes oder Absatzes vorkommen. Darüber hinaus ist zu vermuten, dass die visuelle Erscheinung des Textes ebenfalls zu den ersten wahrgenommenen Phänomenen des Textes zu rechnen ist und insofern ebenfalls zentral ist für die Identifizierung des Themas als Sachverhaltskomplex.

Die differenzierte Beschreibung der Reflexionen des Themas auf den verschiedenen Ebenen des Textes scheint für dieses „Verrechnen“ eher kontraproduktiv zu sein – das Thema ist mehr als die Summe lexikalischer oder referenzieller Ketten. Trotzdem ist es meiner Ansicht nach wichtig, die Reflexionen des Themas auf den verschiedenen Textebenen zu differenzieren und für diese unterschiedlichen Ebenen zu ermitteln, welche sprachlichen und außersprachlichen Merkmale überhaupt als potentielle Indices für das Thema (in seinen verschiedenen Facetten) in Frage kommen (dass für die Identifikation des Themas als Sachverhaltskomplex sowohl eine „Verrechnung“ dieser einfachen indexikalischen Zeichen als auch der Einbezug der genannten Wissensbestände nötig ist, bleibt davon unbenommen). Da in den vergangenen Jahrzehnten operationalisierbare Analyseverfahren für das Textthema eher rar gesät und oft heftig umstritten waren, ist dies nach wie vor ein Desiderat der Forschung. Meines Erachtens ist es in diesem Zusammenhang sinnvoll, eine an der Textoberfläche ansetzende Analyse des Themas in den Blick zu nehmen. Dabei darf jedoch keinesfalls

außer Acht gelassen werden, dass das Thema als Sachverhaltskomplex im Text selbst nur reflektiert wird – Ausdrücke, die auf Teile dieses Sachverhaltskomplexes Bezug nehmen, sind deshalb nicht mit dem Thema selbst zu verwechseln.

In der vorliegenden Arbeit wurde der Vorschlag gemacht, die einzelnen Reflexionen des Themas auf den verschiedenen Ebenen der Textstruktur als Multi-Ebenen-Annotationen zu repräsentieren, da durch ihre integrative Auswertung – d.h. durch die Verrechnung von Merkmalen auf verschiedenen Ebenen des Textes (repräsentiert durch Annotationslayer 1-n) – statistische Zusammenhänge zwischen Einheiten auf verschiedenen Annotationsebenen ermittelt und so neue Erkenntnisse über den Text gewonnen werden können. Allerdings ist festzuhalten, dass zur Gewinnung statistisch signifikanter Ergebnisse große Korpora benötigt werden.

=====

### ***Exkurs: Annotation und Markup***

Annotationen (auch Markup genannt) stellen Textauszeichnungen dar, die Metainformationen über den Text enthalten. Sie werden über Tags in einen Text eingefügt; Anfangstags und Endtags markieren dabei einen bestimmten Textabschnitt und fügen diesem Textabschnitt Metainformationen hinzu (Beispiel: `<s>Dies ist ein Satz.</s>`; das Anfangstag `<s>` markiert den Beginn eines Satzes, das Endtag `</s>` dessen Ende). Die formalen Regeln dafür, wie Annotationen bzw. Markup einem Text zugewiesen werden können, werden in allgemeinen Metasprachen zur Spezifikation von spezifischen Markup-Sprachen definiert. Die verbreitetste Meta-Markupsprache ist XML (Extensible Markup Language), eine vereinfachte Version der Standard Generalized Markup Language (SGML)<sup>229</sup>. Beiden Sprachen ist gemeinsam, dass sie die markierten Textteile in einen hierarchischen Zusammenhang bringen und Texte damit sowohl als sequenzielle als auch hierarchische Gebilde zu beschreiben erlauben. Annotierte Texte weisen demnach, formal gesehen, eine Baumstruktur auf – sie können deshalb durch Text- bzw. Dokumentgrammatiken beschrieben werden. Auf XML basierende Markup-Sprachen können dabei durch zwei Typen von Dokumentgrammatiken definiert werden: 1. DTD's (Document Type Definition), 2. Schemas (z.B. XML Schema, RELAX NG) (vgl. Lobin 2003, 51ff.).

Grundlage für die Annotation eines Textes ist demnach die Dokumentgrammatik, da sie vorgibt, welche Elemente in einem Text vorkommen können, welche Teil-Ganzes-Relationen zwischen diesen Elementen bestehen, wie häufig ein Element vorkommen kann, von welchem Datentyp es ist etc. Ein Auszug einer einfachen Dokumentgrammatik könnte wie folgt aussehen:

```
<!ELEMENT text (block+)>
<!ELEMENT block (textsegment | objekt)+>
<!ELEMENT objekt (grafik | foto | tabelle)>
```

---

<sup>229</sup> SGML: Standard Generalized Markup Language (International Organization for Standardization (ISO) 1986).

```
<!ELEMENT textsegment (#PCDATA)>
```

Dieser Dokumentgrammatik zufolge besteht ein Text mindestens aus einem oder beliebig vielen („+“) Blöcken. Ein Block wiederum setzt sich aus mindestens einem oder beliebig vielen („+“) Kombinationen von einem Textsegment oder („|“) einem Objekt zusammen – da der Konnektor „+“ nach der Klammer steht, bezieht er sich auf die gesamte Klammer, so dass die gesamte Klammer mindestens einmal oder beliebig oft vorkommen kann. Ein Objekt wiederum enthält entweder eine Grafik, ein Foto oder („|“) eine Tabelle, ein Textsegment dagegen konkrete Textdaten („#PCDATA“). In der Dokumentgrammatik fehlen weitere Deklarationen für Grafik, Foto und Tabelle. Darüber hinaus könnten Attribute zu den Elementen eingeführt werden, z.B. könnte jedem Text, jedem Textsegment sowie jedem Objekt eine ID (ein eindeutiger Identifikator) zugewiesen werden – dies könnte z.B. durch entsprechende Attributdeklarationen geschehen:

```
<!ELEMENT text (block+)>
<!ATTLIST text Nr ID #REQUIRED>
```

Multi-Ebenen-Annotationen repräsentieren mehrere parallele Auszeichnungen für einen einzigen Text. Diese Auszeichnungen oder Annotationen stellen dabei verschiedene Sichten auf den Text dar, wobei jede Sicht bzw. jede Annotationsebene durch eine eigene Dokumentgrammatik beschrieben werden kann. Für die Integration bzw. die integrative Auswertung dieser verschiedenen Annotationsschichten wurden verschiedene Modelle vorgeschlagen. Dabei lassen sich drei Hauptvarianten unterscheiden (siehe z.B. Witt 2004a; Witt 2004b; Dipper 2005; Dipper, Götze, Küssner et al. 2007; Chiarcos et al. 2008; Goecke et al. 2010):

1. **Inline-Annotationen:** Annotationen aller Schichten werden dem Text direkt zugeordnet, d.h. wie eine normale, singuläre Annotation repräsentiert. Das Problem der in XML nicht vorgesehenen überlappenden Strukturen wird hier entweder durch Verwendung von Milestone-Elementen, durch die Fragmentierung von Elementen oder durch die Verwendung der aus SGML abgeleiteten „concur“-Funktion (in XML: XCONCUR) umgangen (vgl. Hilbert et al. 2005; Dipper, Götze, Küssner et al. 2007; Schonefeld 2007). Nachteil dieser Variante ist, dass die einzelnen Annotationsebenen schlechter getrennt werden können und die Konsistenz der Annotationen nicht gewährleistet ist, so dass statistische Zusammenhänge zwischen Ebenen weniger zuverlässig ermittelbar sind.
2. **Standoff-Annotationen:** Annotationen und Text werden getrennt. Der Zusammenhang von Annotation und Text wird hergestellt, indem in den Annotationen entweder der in einer externen Textdatei abgespeicherte Text (bzw. (Zeichen)bereiche des Textes) referenziert werden (ein Tag <seg> beginnt mit dem 11. Zeichen des Textes und endet mit dem 15. Zeichen) oder indem in den Annotationen die in einer primären Annotationsschicht annotierten und mit einem eindeutigen Identifikationsausdruck (ID) versehenen „markables“ (z.B. Wortformen) referenziert werden (vgl. Dipper 2005; Dipper, Götze, Küssner et al. 2007). Nachteil dieses Verfahrens ist, dass die Komplexität der einzelnen Annotationen schnell zunimmt. Die Annotatio-

nen können dadurch einerseits nur schwer editiert werden, andererseits sind statistische Anfragen an das Korpus sehr verarbeitungs- und damit zeitintensiv und können sogar zu einem Abbruch der Anfrage führen (“query did not terminate due to stack overflow”), wie Dipper, Götze, Küssner et al. (2007, 343f.) bei einem Vergleich der Performanz einer Inline- und einer Standoff-Version zweier Korpora zeigen konnten. Darüber hinaus ist bei der Standoff-Variante ein manuelles Lesen und Auswerten der Annotationen schwer bis unmöglich.

3. **Separate Annotationen:** Für jede Annotationsschicht wird eine separate Inline-Annotation erstellt; jede Annotationsschicht wird mit einer identischen Kopie des Textes in einer eigenen Datei gespeichert. Das Bindeglied zwischen den Annotationsschichten ist dabei der Text, d.h. die Primärdaten selbst. Dieses Verfahren setzt allerdings voraus, dass alle Annotationsschichten primärdatenidentisch sind. Wenn der Text in einer Datei (einer Annotationsschicht) nur geringfügig geändert wird (z.B. indem ein Zeichen hinzugefügt wird), kommt es zu Identitätskonflikten und es kann keine Verbindung zwischen den verschiedenen Annotationsschichten mehr hergestellt werden (vgl. Witt 2004b; Witt et al. 2005; Goecke et al. 2010). Der Vorteil separater Annotationen liegt vor allem in ihrer Unabhängigkeit voneinander: Sie können unabhängig voneinander erstellt werden, es gibt weniger Probleme mit überlappenden Strukturen und alle Annotationsschichten haben den gleichen Stellenwert, d.h. es wird nicht eine Annotationsschicht als primäre Schicht ausgezeichnet. Für Auswertungen können beliebige Annotationsschichten miteinander kombiniert werden. Darüber hinaus ist jede Annotationsschicht problemlos auch von Menschen les- und verstehbar, während dies bei der Standoff-Annotation nicht der Fall ist.

Da die Identifizierung des Themas eines Textes u.a. auf der Identifizierung seiner Reflexionen auf den verschiedenen Beschreibungsebenen des Textes basiert, d.h. mindestens sechs Annotationschichten erstellt werden müssen, erscheint es in unserem Zusammenhang sinnvoll, Multi-Ebenen-Annotationen in Form separater Annotationen zu realisieren. Diese separaten Annotationen müssen für die Auswertung allerdings wieder integrativ betrachtet werden, d.h. es muss ermittelt werden, welche Relationen es zwischen den verschiedenen Ebenen gibt. Ziel der Analyse von Multi-Ebenen-Annotationen ist u.a., statistisch signifikante gemeinsam auftretende Auszeichnungselemente zu ermitteln, z.B. um so Merkmale einer Ebene A (und möglicherweise auch B und C) als Cues für das gleichzeitige Vorliegen eines bestimmten sprachlichen Phänomens auf der Ebene D zu nutzen – in Lobin et al. (2011) wurden beispielsweise Einheiten auf der Ebene der logischen Dokumentstruktur sowie der anaphorischen bzw. referenziellen Struktur als Cues für die automatische Identifizierung von RST-Diskursrelationen (insbesondere der Relation *ELABORATION*) genutzt. Im Kontext der Themaidentifizierung muss es darüber hinaus darum gehen zu ermitteln, wie Reflexionen des Themas auf den verschiedenen Ebenen miteinander interagieren, welche sprachlichen Einheiten auf mehreren Ebenen als Reflexionen des Themas annotiert wurden und das Thema demnach in prominenter Weise reflektieren. Für die Auswertung separater primärdatenidentischer Annotationen

```

node('output', 0, 19, [1], element('pos.xml_s')).
attr('output', 0, 19, [1], 'xml:lang', 'en').
node('output', 0, 19, [1, 1], element('syll.xml_syll')).
node('output', 0, 4, [1, 1, 1], element('pos.xml_np')).
node('output', 0, 4, [1, 1, 1, 1], element('pos.xml_pron')).
node('output', 0, 4, [1, 1, 1, 1, 1], element('syll.xml_s')).
node('output', 5, 18, [1, 1, 2], element('pos.xml_vp')).
node('output', 5, 7, [1, 1, 2, 1], element('pos.xml_v')).
node('output', 5, 7, [1, 1, 2, 1, 1], element('syll.xml_s')).
node('output', 8, 18, [1, 1, 2, 2], element('pos.xml_np')).
node('output', 8, 9, [1, 1, 2, 2, 1], element('pos.xml_det')).
node('output', 8, 9, [1, 1, 2, 2, 1, 1], element('syll.xml_s')).
node('output', 10, 18, [1, 1, 2, 2, 2], element('pos.xml_n')).
node('output', 10, 13, [1, 1, 2, 2, 2, 1], element('syll.xml_s')).
node('output', 13, 18, [1, 1, 2, 2, 2, 2], element('syll.xml_s')).

```

Abbildung 23: Prolog-Faktenbasis mit den Element- und Attribut-Prädikaten aller Annotationsschichten (aus Goecke et al. 2010, 9)

werden hauptsächlich zwei Methoden eingesetzt: 1. Auswertung mittels eigens entwickelter Tools (bisher existiert hierfür meines Wissens nach nur der Ansatz von Witt (2004b; 2005), der auf der Programmiersprache Prolog basiert); 2. Auswertung mittels der Nutzung von XML-Technologien (native XML-Datenbanken und XML-basierte Abfragesprachen wie XQuery<sup>230</sup>).

Zu 1.: Zur Unifikation und Analyse multipler Annotationen greift Witt (2004b; 2005) auf die Programmiersprache Prolog zurück. Die einzelnen Annotationsschichten werden hierbei zunächst in eine gemeinsame abstrakte Prolog-Faktenbasis überführt (umgekehrt kann auch aus einer entsprechenden Prolog-Faktenbasis wieder ein XML-Dokument bzw. eine Menge von XML-Dokumenten generiert werden). Der Primärtext bzw. die Textdaten, die XML-Elemente und XML-Attribute stellen jeweils einzelne Prolog-Fakten dar. XML-Attribute werden dabei als Prolog-Prädikate vom Typ „attr“ repräsentiert, XML-Elemente bzw. -Elementinstanzen, d.h. Teilbäume im XML-Dokument, als Prolog-Prädikate vom Typ „node“; Elementprädikate weisen fünf Argumentstellen auf: a) Annotationsebene, b) Startposition und c) Endposition (im Hinblick auf die Primärdaten), d) Position (im XML-Dokumentbaum) und e) XML-Elementname (siehe Abbildung 23). Auf der Prolog-Faktenbasis können nun mittels der in Prolog verfügbaren Mechanismen zum automatischen Schließen Relationen zwischen diesen XML-Elementinstanzen ermittelt werden. Die möglichen Relationen zwischen Elementinstanzen sind dabei durch das Verhältnis der Start- und Endpositionen zueinander definiert; Elementinstanzen können in einer der folgenden Relationen zueinander stehen: Identität, Inklusion, Startpunktidentität, Endpunktidentität, Adjazenz, Überlappung und Unabhängigkeit.

<sup>230</sup> XQuery: XML Query Language (W3C – World Wide Web Consortium 2007b).

Darüber hinaus existieren neben den dargestellten Relationen, die den Zusammenhang von Elementinstanzen beschreiben, auch Relationen zur Beschreibung der Zusammenhänge von Elementklassen, d.h. allen Instanzen eines Elementtyps. Zwischen zwei Elementklassen besteht allerdings nur dann eine Metarelation, wenn zwischen allen Paaren von Instanzen dieser beiden Elementklassen eine identische Relation besteht. Auf Basis der Relationen und Metarelationen kann auch eine Unifikation von XML-Dokumenten vorgenommen werden, d.h. die Annotationen aus zwei oder mehr Ebenen können in eine neue gemeinsame Annotationsschicht geschrieben werden, die die Elemente aller unifizierten Ebenen enthält (vgl. Bayerl et al. 2003; Witt 2004b; Witt et al. 2005; Goecke et al. 2010).

Zu 2.: Alternativ kann zur Auswertung von separaten Multi-Ebenen-Annotationen auch auf XML-basierte Austauschformate zurückgegriffen werden. Der Vorteil dieser Auswertungsvariante ist, dass dabei auf eine Reihe von XML-Technologien bzw. -Sprachen (XPath<sup>231</sup>, XSLT, XQuery) zurückgegriffen werden kann, die komplexe Anfragemöglichkeiten für die Auswertung von in XML-annotierten Korpora bieten. Die separaten Annotationen werden hierzu zunächst in ein XML-basiertes abstraktes Austauschformat konvertiert (z.B. PAULA – Potsdamer Austauschformat für linguistische Annotation, siehe Chiarcos et al. 2008; ein alternatives Format ist SGF – Sekimo Generic Format – bzw. dessen Nachfolger XStandoff, siehe Stührenberg und Goecke 2008; Stührenberg und Jettka 2009). In SGF bzw. XStandoff wird das Problem (nicht vorgesehener) überlappender Strukturen in ähnlicher Weise gelöst wie in der Prolog-Repräsentation: Sequenzen von Zeichendaten in den Primärdaten werden über ihre Positionen im Text markiert und für den folgenden Prozess als primäre Annotationsschicht verwendet, indem sie jeweils in den multiplen Annotationen referenziert werden (siehe Abbildung 24). Die separaten Annotationen werden hierzu in das SGF-/XStandoff-Format konvertiert und der primären Annotationsschicht hinzugefügt. Bei einem größeren Korpus von Multi-Ebenen-annotierten Texten ist es empfehlenswert, die integrierten Multi-Ebenen-Annotationen schließlich in einer nativen XML-Datenbank zu speichern, um mit Hilfe der XML-basierten Anfragesprache XQuery relativ einfach statistische Auswertungen auf diesem Korpus vornehmen zu können. Multi-Ebenen-Annotationen, die in PAULA repräsentiert sind, können beispielsweise in der speziell für dieses Format entwickelten Datenbank ANNIS gespeichert werden. Anfragen an das annotierte Korpus sind hier allerdings nur über die ANNIS-eigene Anfragesprache möglich, wobei in deren Zentrum reguläre Ausdrücke und relationale Operatoren stehen (siehe Götze und Dipper 2006).

=====

Im Hinblick auf a) das Zusammenspiel der Reflexionen des Themas auf den verschiedenen Ebenen der Textstruktur und b) die Identifizierung der prominenten vs. weniger prominenten Reflexionen (als Indizien für die zentralen vs. weniger zentralen Themen bzw. Aspekte des Themas) ist es von großer Bedeutung, nicht

---

<sup>231</sup> XPath: XML Path Language (W3C – World Wide Web Consortium 2007a).

```

<base:corpus xmlns="http://www.text-technology.de/sekimo"
  xmlns:base="http://www.text-technology.de/sekimo">
  <base:corpusData xml:id="c1" type="text">
    <base:meta>
      <!-- [...] -->
    </base:meta>
    <base:primaryData start="0" end="19" xml:lang="en">
      <base:textualContent>This is a sentence.</textualContent>
    </base:primaryData>
    <base:segments>
      <base:segment xml:id="seg1" type="char" start="0" end="19"/>
      <base:segment xml:id="seg2" type="char" start="0" end="4"/>
      <base:segment xml:id="seg4" type="char" start="5" end="18"/>
      <base:segment xml:id="seg5" type="char" start="5" end="7"/>
      <base:segment xml:id="seg7" type="char" start="8" end="18"/>
      <base:segment xml:id="seg8" type="char" start="8" end="9"/>
      <base:segment xml:id="seg10" type="char" start="10" end="18"/>
    </base:segments>
  </base:corpusData>
</base:corpus>

```

Abbildung 24: Repräsentation der primären Annotationsschicht im SGF-Format (aus Goecke et al. 2010, 14)

nur die vertikalen Relationen zwischen den verschiedenen Annotationsebenen zu untersuchen, sondern auch Kombinationen aus vertikalen und horizontalen Relationen. Während die von Witt vorgelegte Prolog-Implementation in erster Linie eine Analyse vertikaler Relationen erlaubt, ermöglichen sowohl die AN-NIS-Infrastruktur als auch die komplexen Auswertungs- und Anfragemöglichkeiten von XPath, XSLT und XQuery eine Analyse sowohl vertikaler als auch horizontaler Relationen. Die Verwendung einer Datenbank ist darüber hinaus generell als sinnvoll zu erachten, da auch Wissensressourcen, wie z.B. GermaNet 5.0, Thema-Marker-Lexika, Texttypstrukturontologien etc. in einer Datenbank (entweder einer XML-fähigen oder einer XML-nativen) gespeichert werden könnten und so in die statistischen Anfragen und Auswertungen integriert werden könnten, ohne dass hierzu zusätzliche Software-Schnittstellen benötigt würden.

Im Folgenden soll beispielhaft angedeutet werden, welche Auswertungen eines auf den beschriebenen Ebenen der Textstruktur annotierten Korpus nötig sind, um Erkenntnisse zu den beiden o.g. Aufgaben (a) und (b) zu gewinnen.

Ein System zur automatischen Themenidentifizierung, das auf der theoretischen Thema-Konzeption dieser Arbeit basiert, hat mehrere Aufgaben und Teilaufgaben zu lösen:

1. Identifizierung des Themas und seiner Aspekte
  - a. Identifizierung von wichtigen und weniger wichtigen Aspekten des Themas
  - b. Identifizierung der Bedeutung/Gewichtung von Aspekten des Themas
  - c. Identifizierung von lokalen und globalen Aspekten des Themas
2. Identifizierung der Relationen zwischen dem Thema und seinen Aspekten
  - a. Identifizierung der Relationen zwischen dem zentralen Textthema und seinen Aspekten
  - b. Identifizierung der Relationen zwischen den Aspekten des Themas

Die Reflexionen des Themas auf den Ebenen der Textstruktur können in verschiedener Weise miteinander verrechnet werden, um zur Lösung o.g. Aufgaben und Teilaufgaben beizutragen. Reflexionen des Themas können zunächst einmal als indexikalische Zeichen für die *Identifizierung des Themas und seiner Aspekte* (Aufgabe 1) verwendet werden. Es wird angenommen, dass der in einem Text thematisierte Sachverhaltskomplex in erster Linie durch die Menge der Diskursentitäten, über die etwas gesagt wird, reflektiert wird. Da das Thema die Orientierungs- und Einordnungsinstanz beim Produzieren und Rezipieren von Texten darstellt, es also durchgehend mental aktiviert ist, ist es nötig, die Diskursentitäten im Hinblick auf den Grad ihrer mentalen Aktiviertheit und damit ihrer Bedeutung zu gewichten. Dies kann durch eine Verrechnung von Reflexionen verschiedener Textstrukturebenen geschehen: Entitäten auf der Textoberfläche, die auf verschiedenen Ebenen der Textstruktur als Reflexionen des Themas markiert sind, besitzen prinzipiell ein größeres Gewicht, als Entitäten, die nicht oder nur auf einer Ebene als Reflexionen des Themas markiert sind. Um diese Gewichtungen vornehmen zu können, ist demnach eine statistische Auswertung der Multi-Ebenen-Annotationen nötig, die zwischen den Annotationsschichten nach Relationen der Identität, Inklusion, Startpunktidentität, Endpunktidentität, oder Überlappung sucht. Liegt eine solche Relation vor, wird die entsprechende sprachliche Einheit mit einem bestimmten Gewicht bzw. einem Score versehen. Neben diesen, auf vertikalen Korrelationen (vertikal im Hinblick auf die Ebenen der Textstruktur) basierenden Gewichtungen, können Gewichtungen auch im Hinblick auf die horizontale Verteilung im Text vorgenommen werden. Die These ist, dass weniger wichtige Aspekte des Themas eher in jeweils lokal begrenzten Bereichen des Textes reflektiert werden, wichtige Aspekte dagegen durchgehend bzw. in weiten Teilen des Textes. Um solche lokalen von globalen Aspekten des Themas zu differenzieren, muss einerseits für jede Textstrukturebene ermittelt werden, welche Einheiten bzw. welche Diskursentitäten auf der horizontalen Ebene in einer Relation der Identität oder partiellen Identität stehen, also kontinuierlich als Reflexionen des Themas markiert sind. Andererseits müssen zur



Ermittlung der horizontalen Gewichtungen Reflexionen auf verschiedenen Ebenen miteinander verrechnet werden. Da davon ausgegangen wird, dass bestimmte Einheiten der visuellen und logischen Textstruktur eine zentrale Funktion bei der Signalisierung des Themas des Textes spielen (z.B. Titel, Inhaltsverzeichnis, Lead-Satz, bestimmte Positionen im Text, Absatz, Satz), kann in einer initialen Hypothese angenommen werden, dass die Diskursentitäten, die in diesen Einheiten vorkommen, als primäre Thema-Marker fungieren. Diese Hypothese kann auf ihre Korrektheit überprüft werden, indem das Gewicht dieser Diskursentitäten auf anderen Annotationsebenen ermittelt wird. Haben Diskursentitäten, die im Titel, Lead-Satz, im ersten und/oder letzten Absatz des Textes vorkommen, auch auf der Ebene der Informationsstruktur, der referenziellen und lexikalischen Struktur ein entsprechend hohes Gewicht, ist davon auszugehen, dass sie wichtige Aspekte des Themas reflektieren.

Für die *Identifizierung der Relationen zwischen dem Thema und seinen Aspekten* (Aufgabe 2) spielen insbesondere die relationale Struktur und die generische Texttypstruktur eine wichtige Rolle. Einerseits werden Verknüpfungen zwischen Aspekten des Themas über Kohärenzrelationen reflektiert (Beispiel: Ein Diskurssegment B steht in einer CAUSE-Relation zu einem Segment A; mit dem Segment B werden demnach Angaben zur Ursache eines Geschehens oder Vorgangs in A gemacht, um zu signalisieren, dass in B über die Ursachen von A geredet wird, dass B und A über diese Ursachen-Relation thematisch zusammenhängen). Andererseits kann das Thema auch über die als Texttypstrukturkategorien instanziierten Entitäten behandelt werden – diese instanziierten Texttypstrukturkategorien stellen demnach Reflexionen der Aspekte des Themas dar; ihre Relation zum Thema ist schemabasiert. In Zeitungsnachrichten wird ein Thema beispielsweise dadurch behandelt, dass etwas über die beteiligten Personen, den Ort, die Konsequenzen etc. gesagt wird – *Person, Ort, Konsequenz* kategorisieren die entsprechenden Aspekte des Themas und stellen sie darüber in eine spezifische Relation zueinander sowie zum Thema. Es ist davon auszugehen, dass bestimmte Texttypstrukturkategorien durch Einheiten instanziiert werden, die auch auf anderen Ebenen der Textstruktur als Reflexionen des Themas markiert sind – Kategorien wie *Person* oder *Ort* werden durch spezifische lexikalische Einheiten instanziiert, die möglicherweise eigene lokale lexikalische Ketten bilden. Kategorien wie *Konsequenz* werden demgegenüber möglicherweise in Diskurssegmenten realisiert, die in einer CONSEQUENCE-Relation zu einem anderen Segment stehen. Reflexionen des Themas einer Textstrukturebene dienen in diesen Fällen vor allem als Cues für das Vorliegen eines bestimmten Phänomens bzw. einer Einheit (z.B. der Texttypstrukturkategorie KONSEQUENZ) einer anderen Strukturebene.<sup>232</sup> Neben den Kohärenzrelationen und den Texttypstruk-

---

<sup>232</sup> Insbesondere für die Ermittlung von Kohärenzrelationen spielen Ebenen wie die referenzielle oder lexikalische eine wichtige Rolle: Eine referenzielle Relation zwischen zwei Segmenten kann z.B. ein Indiz für das Vorliegen einer ELABORATION-Relation auf der Ebene der relationalen Diskursstruktur sein. Solche Cues bzw. Indizien stellen allerdings in der Regel weder notwendige noch hinreichende Bedingungen hierfür dar. Durch eine Summierung von Cues (wenn z.B. gleichzeitig ein expliziter lexikalischer Thema-Marker vorliegt, der auf eine Themafortführung hindeutet) wächst allerdings die Wahrscheinlichkeit, dass eine solche ELABORATION-Relation

turschemata können auch lexikalisch-semantische oder konzeptuelle Relationen zwischen Entitäten zur Identifizierung von Relationen zwischen Aspekten des Themas genutzt werden. Insbesondere für enzyklopädische Texte scheinen zwischen Aspekten des Themas eher begriffliche oder semantische Relationen (z.B. Hyperonymie, Synonymie) vorzuliegen als Kohärenzrelationen. D.h. die Funktion und das Gewicht, das eine bestimmte Ebene für die Identifizierung von Relationen zwischen Aspekten des Themas spielt, ist u.a. abhängig vom Texttyp. Ein auf der vorliegenden Arbeit basierendes Software-System zur Themenidentifikation sollte demnach je nach Texttyp unterschiedliche Ebenen in die Verrechnung einbeziehen; die Regeln zur Bedeutung bestimmter Kombinationen von Cues und Constraints für die Themenanalyse müssen demnach modular sein und dynamisch ausgewählt werden.

Wie in der vorliegenden Arbeit immer wieder betont wurde, spielen für die Identifizierung des Themas und seiner Aspekte sowie der Relationen zwischen dem Thema und seinen Aspekten neben den Reflexionen des Themas auch Wissensbestände eine wichtige Rolle. Da in diesem Kapitel allerdings die „Verrechnungen“ der Reflexionen des Themas im Zentrum stehen, wird das Zusammenspiel von Top-Down-Constraints (= wissensbasierten Einflüssen und Beschränkungen) und Bottom-Up-Cues (= Reflexionen des Themas auf der Textoberfläche) an dieser Stelle nicht vertieft, sondern es wird auf die abschließende Darstellung der Architektur eines auf der Thema-Konzeption dieser Arbeit basierenden Systems zur maschinellen Themenidentifizierung und der dafür nötigen Komponenten und Prozesse in Kapitel 3.4.3 verwiesen. Abschließend soll festgehalten werden, dass ein Ebenen-basiertes Beschreibungsmodell der Textstruktur, das durch Multi-Ebenen-Annotationen formal repräsentiert und so maschinell auswertbar gemacht wird, für die Analyse und Identifizierung insbesondere emergenter Textphänomene wie dem Thema klare Vorteile gegenüber einem integrativen Textmodell bietet.

### **3.4 Die Rolle des Themas für die Produktion und Rezeption von Texten**

Die Bedeutung des Konzepts Thema für die Produktion und Rezeption von Texten spiegelt sich in einer Vielzahl von Thema-Konzepten wider, angefangen bei den semantischen Thema-Konzepten, bei denen davon ausgegangen wird, dass das Thema den Kern des Textes bildet, der im Produktionsprozess „entfaltet“ wird (siehe Kapitel 2.2.2.4). Heinemann/Viehweg sprechen deshalb von der

---

tatsächlich vorliegt. Wie die Erfahrungen aus Diskursparsing-Projekten gezeigt haben, können für eine Reihe von Diskursrelationen entsprechende Cues ermittelt werden. Einschränkend muss festgehalten werden, dass es manuell nahezu unmöglich zu ermitteln ist, welche Kombinationen und Gewichtungen von Indizien zum besten Ergebnis führen. Stattdessen sind hier umfangreiche Tests und Evaluationen nötig. Eine solche Implementierung und Evaluation steht noch aus, so dass in der vorliegenden Arbeit keine empirisch abgesicherten Aussagen über die besten Kombinationen und Gewichtungen von Indizien getroffen werden können.

außerordentlichen Relevanz, die dem Text-Thema-Begriff in der Praxis der Textproduktion und Textrezeption zukommt. Der Sprecher geht bei der Textgestaltung von einem Text-Thema aus, das dann die Basis für die Programmierung von Texten im Sinne einer Entfaltung des Text-Themas darstellt. Beim Prozess der Textrezeption durch den Hörer wiederum bildet das Text-Thema das Ergebnis des Verstehensprozesses, da der Hörer einen Text erst dann wirklich verstanden hat, wenn er außer zahlreichen Einzelinformationen vor allem das Text-Thema (und die damit verbundene Intention des Sprechers) voll erfasst, d.h. aus den Textinformationen rekonstruiert hat. (Heinemann und Viehweger 1991, 46)

Der Textproduzent nutzt das Thema demnach als Ausgangspunkt und Quelle für den Textaufbau, insofern es den inhaltlichen Kern des Textes darstellt, während das Thema für den Rezipienten zum (inhaltlichen) Ergebnis des Rezeptions- und Verstehensprozesses wird.

Vor dem Hintergrund der Betrachtung von Textproduktion und -rezeption erscheint es angebracht, einen Wechsel der Perspektive vorzunehmen: weg vom *Thema* und hin zum *Thematisieren*, weg vom Text als *Produkt*, hin zu den kommunikativen *Prozessen* des Produzierens und Rezipierens, in dessen Mittelpunkt die Verständigung der Kommunikanten mit Hilfe eines Textes steht. Hierbei interessiert vor allem, welche Faktoren bei der Produktion thematisch kohärenter Texte eine Rolle spielen. In diesem Zusammenhang müssen *die textorganisierenden Teilhandlungen des Thematisierens*, nämlich das Thematisieren als Mittel der Informationsselektion und das Thematisieren als Mittel der Informationsstrukturierung näher betrachtet werden. Auch bei der Diskussion der Rolle des Themas für den Rezeptions- bzw. Verstehensprozess soll der Blick verstärkt auf dem kommunikativen Prozess liegen. Ausgangspunkt hierbei ist die These, dass das Thema die Rolle einer „Einordnungsinstanz für das Verstehen“ (Hoffmann 1997, 514) spielt. Der Rezipient ordnet die Einzelinformationen des Textes in einen verbindenden, über den einzelnen Satz hinausreichenden *thematischen Rahmen* ein, um auf diese Weise thematische Kohärenz aufzubauen.

Fritz beschreibt diese doppelte Funktion des Themas für die Textproduktion einerseits und die Textproduktion andererseits wie folgt:

Aus einer anderen Perspektive sieht man den Status von Themen, wenn man Themaformulierungen in der Funktion als Themenvorgabe oder Themenankündigung betrachtet (z.B. Tagesordnung eines Gremiums oder Ankündigung einer Geschichte zu einem bestimmten Thema). Hier dient das Thema in zweierlei Hinsicht als Orientierung. Den jeweiligen Sprechern auferlegt die Vorgabe die Verpflichtung, thematisch relevante Beiträge zu machen bzw. die thematische Relevanz ihrer Beiträge zu verteidigen. Den jeweils Zuhörenden gibt das Thema einen Verstehenshorizont und das Recht, thematische Relevanz einzuklagen. (Fritz 1994, 193)

Welche Faktoren bei diesem Zusammenspiel von Thema und Textproduzieren und -rezipieren zu berücksichtigen sind, soll im Folgenden genauer erläutert werden.

### 3.4.1 Thematisieren als textorganisierende Handlung

Das Thema spielt eine zentrale Rolle in der Phase der Planung des Textes, insofern es die Aktivierung, Selektion und Organisation von Wissen steuert. Die Aufgabe der Selektion betrifft nicht nur das globale Textthema, sondern auch die

lokalen Themen der einzelnen Äußerungen (die Aspekte des Themas). Was in Form lokaler Themen thematisiert werden darf, ist in erster Linie abhängig davon, was *relevant* ist. Die Aufgabe der Selektion von Themen bedeutet demnach, aus der Menge der potentiellen Themen diejenigen auszuwählen, die für den Text und die Kommunikationssituation relevant sind. Sich thematisch zu äußern, heißt demnach also auch immer, sich relevant zu äußern. Dies hat damit zu tun, dass von Teilnehmern einer Kommunikation, den Kommunikanten, erwartet wird, kommunikativ kooperativ zu sein. Dazu gehört u.a., dass sie relevante Kommunikationsbeiträge machen („Maxime der Relation“, siehe Grice 1979, 249 u. 251) – das, was vom Textproduzenten gesagt werden darf, muss relevant sein. Was als relevant verstanden wird, ist abhängig von verschiedenen Faktoren: Wie Brown/Yule (1983, 84) schreiben, ist thematische Relevanz kontextabhängig, insofern es gemessen wird an dem aktivierten thematischen Rahmen, d.h. an den durch die bisher gemachten Äußerungen evozierten Themen – eine Äußerung ist dann relevant, wenn sich diese in den aktivierten thematischen Rahmen einbinden lässt. Voraussetzung für akzeptables Thematisieren ist demnach die Anpassung des aktuellen Beitrags an den Stand des Textes und der bisher thematisierten Inhalte.<sup>233</sup> Auch die Kommunikationssituation hat einen Einfluss darauf, was thematisiert werden kann bzw. was thematisch relevant ist. Was in einer Situation als angemessen und relevant beurteilt wird, ist in einer anderen Situation noch lange nicht ebenso relevant. Ein wissenschaftlicher Aufsatz kann beispielsweise bei einer Tagung angenommen werden, weil er als relevant erachtet wird, während er bei der Vorjahrestagung abgelehnt wurde, weil diese unter einem anderen thematischen Schwerpunkt stand. Auch die Zielgruppe, an die ein Text gerichtet ist, entscheidet mit darüber, ob etwas relevant ist: In einem wissenschaftlichen Vortrag gegenüber Fachkollegen wäre ein einführender Vortragsteil zu den Grundlagen des Fachgebietes irrelevant oder zumindest ungewöhnlich; der gleiche Vortrag kann jedoch als relevant beurteilt werden, richtet er sich z.B. an Erstsemesterstudierende. Die Fähigkeit zur Beurteilung thematischer Relevanz hängt darüber hinaus in nicht geringem Maße davon ab, wie das inhaltliche Hintergrundwissen des Rezipienten beschaffen ist. Einem Mathematiker wird es wahrscheinlich schwerer fallen, die thematische Relevanz der Teile eines Buches zur Textlinguistik zu beurteilen, als einem Linguisten. Fritz (1994, 197) weist darauf hin, dass es dementsprechend divergierende Relevanzbewertungen geben kann und dass Relevanz generell graduell sei.

Relevanz lässt sich auf verschiedenen Ebenen feststellen. Man unterscheidet propositionale, sequenzielle, funktionale, kommunikative, sprachliche sowie thematische Relevanz. Eine Äußerung kann z.B. thematisch relevant sein, aber nicht dem aktuellen Dialogstand entsprechen, d.h. sequenziell nicht relevant sein. Vergleicht man diese Auffassung von Relevanz mit derjenigen bei Brown/Yule (s.o.), stellt man fest, dass Brown/Yule von einem ganzheitlicheren Relevanzbegriff ausgehen, insofern sie unter Relevanz sowohl sequenzielle als

---

<sup>233</sup> „Ich erwarte vom Beitrag des Partners, daß er dazu paßt, was an dem jeweiligen Punkt der Interaktion gerade vonnöten ist; wenn ich gerade die Zutaten für einen Kuchen verrühre, möchte ich kein gutes Buch und nicht einmal einen Topflappen gereicht bekommen (obwohl das zu einem späteren Zeitpunkt passend sein mag).“ (Grice 1979, 251)

auch thematische Relevanz fassen. Ich möchte mich dieser Sichtweise anschließen und gehe davon aus, dass eine Äußerung in einer spezifischen Kommunikationssituation nur dann relevant sein kann, wenn sie alle o.g. Aspekte erfüllt; ist sie im aktuellen Diskurskontext nicht thematisch relevant, kann sie auch nicht sequenziell relevant sein, und umgekehrt kann sie nicht sequenziell relevant sein, wenn sie nicht gleichzeitig thematisch relevant ist (vgl. auch Sanford und Moxey 1995, 162).<sup>234</sup>

Relevanz wurde bisher als Voraussetzung und Beschränkung für den Akt des Thematisierens betrachtet. Ein Gegenstand darf vom Textproduzenten nur thematisiert werden, wenn dieser auch relevant ist (inhaltlich und im gegebenen Äußerungskontext). Was bedeutet dies nun konkret? Wenn wir davon ausgehen, dass sich ein Textproduzent an die Maxime der Relevanz hält und im besten Fall nur das thematisiert, was auch relevant ist, muss daraus die Schlussfolgerung gezogen werden, dass zum Thematisieren notwendigerweise auch das Selektieren von Informationen gehört: Der Textproduzent muss aus der Menge der möglichen zu verbalisierenden Informationen diejenigen auswählen, die relevant sind. Diese Selektion wird gesteuert von verschiedenen Faktoren – der Intention, mit der der Text verfasst wird, der Kommunikationssituation (bzw. dem Wissen über diese), dem Text selbst bzw. dem aktuellen Textkontext sowie der Stellung des Textes im Diskurs. Insbesondere in expositorischen Texten ist aber davon auszugehen, dass das Thema, über das der Produzent des Textes etwas sagen möchte, bzw. das Wissen, dass der Produzent über das Thema und seine Aspekte besitzt, die größten Auswirkungen auf den Prozess der Selektion hat.

Was als inhaltliche Fortführung einer Äußerung zulässig ist, ist demnach unter anderem abhängig von dem Wissen des Produzenten über das Thema und seine Aspekte, d.h. dem Sachwissen, das in Form von Sachverhaltsstrukturen bzw. Situationsmodellen mental repräsentiert ist (siehe Kapitel 3.2.1). Ein Aspekt aus einem evozierten Frame bzw. ein Teil eines Situationsmodells ist jedoch nur unter bestimmten Voraussetzungen selektierbar, nämlich dann, wenn im aktuellen Diskurskontext (in den unmittelbar vorausgegangen Redebeiträgen bzw. dem angrenzenden Textausschnitt) bereits ein assoziierter Teilaspekt thematisiert wurde und sich das aktuelle Thema über eine Kohärenzrelation an diesen vorhergehenden Aspekt anschließen lässt, oder wenn der thematische Bezug zwischen einem der vorhergehenden Aspekte und dem aktuellen Aspekt durch sprachliche Mittel expliziert wird – im erstgenannten Fall wird die kognitive Sachverhaltsstruktur im Text abgebildet, im letztgenannten Fall wird sie durch den Text erweitert. Die einzige Passage des Textes, die nicht von dieser Beschränkung betroffen ist, ist der initiale Einleitungsteil, z.B. die Überschrift oder

---

<sup>234</sup> Anders sähe es aus, wenn man die thematische Relevanz unabhängig vom aktuellen Diskurskontext bewerten könnte – in diesem Fall könnte man auflisten, welche verschiedenen Themen und Subthemen in einem Text zu einem bestimmten Thema relevant sein können, unabhängig von ihrer Stellung in der Dialog- oder Textsequenz. Diese Art von thematischer Relevanz würde sich allein aus den Verknüpfungen von Inhalten in einem kognitiven Schema ergeben. Sieht man thematische Relevanz bzw. die Bewertung thematischer Relevanz allerdings als einen interaktiven kommunikativen Prozess an, ist eine solche – rein inhaltsbezogene – Interpretation von thematischer Relevanz nicht sinnvoll.

der erste Satz bzw. Absatz – an dieser Stelle muss das Thema erst noch gesetzt werden; der Textproduzent muss hier nur die Beschränkungen, die ihm das kommunikative Ziel und der kommunikative Rahmen auferlegt, berücksichtigen, ansonsten ist er beim Thematisieren vollkommen frei. Durch die initiale Setzung des Themas – die „Themenankündigung“ (Fritz 1994, 193) – wird schließlich der thematische Rahmen des Textes begründet, der dann auf Seiten des Produzenten das weitere Thematisieren als Selektieren beschränkt (er muss im weiteren Verlauf thematisch relevant bleiben) und auf Seiten des Rezipienten Erwartungen bezüglich des weiteren Verlaufs weckt und gleichzeitig einen „Verstehenshorizont“ (Fritz 1994, 193) bildet.

Neben der Abhängigkeit vom Sachwissen bzw. den Sachverhaltstrukturen muss beim Thematisieren als Selektieren auch die Abhängigkeit von der prototypischen generischen Texttypstruktur sowie der visuellen und logischen Textstruktur beachtet werden bzw. die Abhängigkeit vom Wissen über globale und lokale Textmuster (siehe Kapitel 3.2.2). Hierbei ist – wie im Übrigen auch bei dem Zusammenspiel mit dem Sachwissen – allerdings nicht mehr klar zwischen dem Thematisieren als Selektieren und dem Thematisieren als Strukturieren zu trennen. Die beiden Funktionen des Thematisierens gehen ineinander über. Wie in Kapitel 2.6.3.6 dargelegt, ist das, was an einer bestimmten Stelle im Text gesagt werden darf (an einer bestimmten Stelle der Textstruktur – Thematisieren als Strukturieren), nicht frei wählbar, sondern unterliegt texttypspezifischen kategorialen Beschränkungen. In manchen Texttypen sind diese Beschränkungen stärker wirksam, in anderen weniger.

Sowohl die Abhängigkeit vom Sachwissen als auch von den prototypischen globalen Textstrukturen speist sich aus dem Prinzip der Setzung und Erfüllung von Erwartungen – der Textproduzent setzt mit seinem initialen Textteil das Thema des Textes und muss im weiteren Verlauf des Textes die dadurch evozierten thematischen Erwartungen des Rezipienten erfüllen, um thematische Kohärenz zu gewährleisten. Damit ist das Prinzip der Setzung und Erfüllung von Erwartungen nicht nur zentral für die Textproduktion, sondern auch für die Rezeption (siehe Kapitel 3.4.2).

Im Folgenden soll eine zweite Funktion des Thematisierens vorgestellt werden: die des Strukturierens. Hierbei werden sowohl die Äußerungseinheiten des Textes als auch die Aspekte des Themas linearisiert und in einen Zusammenhang zueinander gesetzt, d.h. in eine Textstruktur überführt. Dabei können folgende, miteinander interagierende Teilhandlungen des Thematisierens als Strukturieren unterschieden werden: Thematisieren als *Orientieren*, als *In-Bezug-Setzen* und als *Fokussieren und Gewichten*. Mit dem Thematisieren als Orientieren ist gemeint, dass Texte als Ganzheit auf ein Thema bezogen sind. Ein Thema wird im Text zwar ausdifferenziert und in den einzelnen sprachlichen Teilhandlungen bzw. Äußerungseinheiten in Form verschiedenster Aspekte behandelt, dabei muss jedoch immer ein Bezug zum Thema erkennbar bleiben – die Aspekte müssen sich in den thematischen Rahmen des Textes einordnen lassen. Das Thema bzw. der thematische Rahmen dient demnach als durchgehende Orientierungs- und Einordnungsinstanz. Die Orientierung auf ein zentrales Thema kann vom Produzenten sprachlich signalisiert werden durch Wiederaufnahme

der zentralen Aspekte des Themas (in Form von lexikalischer und referenzieller Rekurrenz). Wie bereits in Kapitel 3.2.1 dargestellt, basiert der thematische Rahmen auf dem vorgängigen Wissen der Kommunikanten über einen Sachverhaltskomplex und die diesen Sachverhaltskomplex konstituierenden Entitäten und deren Relationen. Das Thematisieren als Orientieren ist demnach eng auf das Sachwissen bezogen. Darüber hinaus spielt in diesem Zusammenhang jedoch auch das Sprachwissen eine zentrale Rolle, insofern sowohl auf das abstrakte Wissen über die Möglichkeiten der Signalisierung des Themas Bezug genommen werden muss, als auch auf das konkrete Wissen darüber, mit welchem lexikalischem Material und welchen grammatischen Strukturen in der Regel im Kontext einer kommunikativen Situation über ein spezifisches Thema geredet wird. Dass hierbei natürlich keine Eins-zu-Eins-Übernahmen von Wissen möglich sind, sondern dass immer eine Adaption an die spezifische Kommunikationssituation (den Adressaten, das Publikationsmedium etc.) erfolgen muss, ist selbstverständlich.

Insofern die Aspekte des Themas die Bezugspunkte für die Teilhandlungen des Textes darstellen und über diese Teilhandlungen miteinander verknüpft werden, stellt sich das Thematisieren darüber hinaus als In-Bezug-Setzen dar – diese Bezugnahme erfolgt durch die Verknüpfung von Äußerungseinheiten bzw. sprachlichen Teilhandlungen durch Kohärenzrelationen. Kohärenzrelationen wiederum können entweder explizit signalisiert werden, z.B. durch kohäsive Mittel wie Konjunktionen, z.B. „weil“ (Diskursmarker für die Relation „Ursache“ oder „Grund“) oder „danach“ (Diskursmarker für die Relation „Sequenz“ oder „temporale Folge“), oder sie bleiben implizit – in diesem Fall rechnet der Produzent damit, dass es dem Rezipienten möglich ist, den Zusammenhang auf Basis seiner Wissensbestände zu inferieren. Beim Thematisieren als In-Bezug-Setzen findet demnach sowohl eine Interaktion mit dem Sachwissen (dem Wissen über Relationen zwischen Entitäten bzw. Aspekten des Themas) statt als auch eine Interaktion mit dem Sprachwissen (dem Wissen darüber, wie Bezüge zwischen den Aspekten des Themas sprachlich realisiert sein können) als auch mit dem Wissen über die Kommunikationssituation (insbesondere den Annahmen des Produzenten über das Sachwissen des Rezipienten).

Die dritte Teilhandlung des Thematisierens als Strukturieren stellt das Fokussieren und Gewichten dar. Dieses Fokussieren und Gewichten wird auf der Ebene des Satzes realisiert durch eine spezifische Informationsstruktur.<sup>235</sup> Vor allem in funktionalen Grammatiktheorien wird die Thema-Rhema-Gliederung als eines der zentralen Ausdrucksmittel für eine Strukturierung und Perspektivierung der Informationen im Satz betrachtet (siehe Halliday 1967; Daneš 1970, 1976; Halliday 1985; Fries 1995). Auch in den über die einzelne Äußerungseinheit hinaus-

---

<sup>235</sup> Kindt stellt in diesem Zusammenhang fest: „In der Linguistik werden seit einiger Zeit unter dem Begriff der Informationsstruktur verschiedene Aspekte einer funktionalen Satz- und Textgliederung subsumiert, die zu bekannten dichotomischen Kategorisierungen wie ‚Thema vs. Rhema‘, ‚alte Information vs. neue Information‘, ‚Hintergrund vs. Fokus‘ etc. gehören (vgl. Jacobs 1992). Man muß heute davon ausgehen, daß die Produktion und Rezeption von Äußerungen maßgeblich durch informationsstrukturelle Gegebenheiten gesteuert wird.“ (Kindt 1995, 261).

gehenden größeren Einheiten des Textes wird das Thematisieren als Fokussieren und Gewichten realisiert: Themen werden im Text so angeordnet, dass sie einen bestimmten globalen Informationsaufbau bewirken. Dieser globale Informationsaufbau ist je nach Textklasse und kommunikativem Zweck unterschiedlich. In einem narrativen Text über einen Unfall sind andere Aspekte des Themas von Bedeutung und erwartbar als in einem Ergebnisbericht. In Textsorten wie Zeitungsnachrichten gibt es einen anderen thematischen Aufbau als in wissenschaftlichen Artikeln: In erstgenanntem Texttyp wird mit dem wichtigsten Aspekt des Themas begonnen und es wird dann in absteigender Reihenfolge auf die unwichtigeren, d.h. weniger fokussierten Themen übergegangen. In wissenschaftlichen Artikeln werden die Aspekte des Themas dagegen zunächst den für diesen Texttyp spezifischen Texttypstrukturkategorien (z.B. Theorie, Methode) zugeordnet und dann in einer spezifischen Reihenfolge im Text präsentiert. Fokussierte Themen werden in dieser Textsorte darüber hinaus meist schon zu Beginn des Textes eingeführt (im vorangehenden Abstract oder der Einleitung) und am Schluss noch einmal aufgegriffen (in der Schlussbemerkung oder Zusammenfassung). Das Thematisieren als Strukturieren wird in diesem Zusammenhang demnach beschränkt durch das Wissen über Textmuster. Das Thematisieren als Fokussieren und Gewichten wirkt sich – wiederum in Interaktion mit dem Wissen über Textmuster – auch auf die Gestaltung von Einheiten des Textes aus, die zwischen den Makroeinheiten (die Einheiten der logischen und texttypspezifischen globalen Textstruktur) und den Mikroeinheiten (dem Satz bzw. der einzelnen Äußerungseinheit) liegen: Die Verteilung der Aspekte des Themas auf die Einheiten der logischen Textstruktur wie Überschriften, Absätze, Fußnoten, visuelle Hervorhebungen etc. erfolgt demnach nicht zufällig, sondern dient der Fokussierung und Gewichtung der Aspekte des Themas: Überschriften fungieren – zumindest in expositorischen Texten – häufig als Mittel zur Signalisierung des Themas. Fußnoten dagegen beschäftigen sich in der Regel mit einem eher nebensächlichen, nicht fokussierten Aspekt des Themas. Absätze können dazu verwendet werden, einen Themawechsel zu signalisieren. Visuelle Hervorhebungen können dazu dienen, einen bestimmten Aspekt des Themas hervorzuheben.

Die betrachteten Teilhandlungen des Thematisierens als Strukturieren sind nicht als aufeinanderfolgende Handlungen aufzufassen, sondern als ineinander verschränkte. Abschließend soll deshalb noch einmal der Blick auf den übergreifenden Prozess des Thematisierens als Strukturieren gerichtet werden. Beim thematisch gesteuerten Strukturieren werden die Äußerungseinheiten zunächst in eine sachlogische Ordnung gebracht. Das Thema hat hierbei eine strukturierende Funktion im Hinblick auf den Inhalt des Textes. Insbesondere bei explikativen Texten ist die inhaltliche Struktur stark von dem außersprachlichen „Sachverhaltszusammenhang“ geprägt (vgl. Jahr 2000, 394) (siehe hierzu Kapitel 3.2). Dieser Produktionsebene entspricht die Stufe Zwei der in Feilke (1996, 1186) genannten „Stufenfolgen in der Ausbildung von Kohärenzprinzipien“, die Stufe der „sachlogischen Ordnung“, bei der Schreiber lernen, Texte nach sachlogischen Gesichtspunkten zu strukturieren und dabei „Sach-Strukturen“ oder „thematische frames“ nutzen. Diese Strukturen haben dabei sowohl „heuristische“ als auch „textorganisierende“ Funktionen (vgl. Feilke 1996, 1187). Auf den



Zusammenhang von Thema (bzw. thematischer Kohärenz), Wissensstrukturen und Selektions- und Linearisierungsprozessen weisen auch Brinker/Hagemann hin:

Lokale wie globale thematische Kohärenz wird unter Bezugnahme auf repräsentationale Strukturen wie Schemata, ‚frames‘, ‚scripts‘ (...) in Gestalt verschiedener Selegierungs- und Linearisierungs- bzw. Sequenzierungsprinzipien rekonstruiert, die zwischen Aufbau und Modellierung von Wissensstrukturen und deren Verstehen vermitteln sollen. (Brinker und Hagemann 2001, 1254)

In Kapitel 2.7.3 wurde dargestellt, dass Frames auch als Listen von Handlungsanweisungen oder Fragen zu stereotypischen Situationen aufgefasst werden können. Insofern Fragen auch als unterspezifizierte Themen betrachtet werden können, wird die Textproduktion demnach geleitet durch die in einem Frame repräsentierte unterspezifizierte thematische Struktur zu einer stereotypischen Situation.

Neben den kognitiven Sachverhaltsstrukturen (als Repräsentation des Sachwissens) wird die Handlung des Thematisierens als Strukturieren jedoch auch beeinflusst durch das Handlungswissen, u.a. dem Wissen über Kommunikationsprinzipien. Über die Bedeutung der Maxime der Relevanz wurde bereits bei der Erörterung des „Thematisierens als Selektierens“ gesprochen. Die Maxime der Verständlichkeit wirkt im Gegensatz zur Maxime der Relevanz vor allem auf die thematische Strukturierung des Textes ein, insofern „eine der sachlichen Strukturierung des Sachverhalts entsprechende Anordnung der Informationseinheiten bei der Entfaltung des Themas“ zur inhaltlichen Verständlichkeit beiträgt und den Aufbau einer kohärenten Wissensstruktur erleichtert (Jahr 2000, 395).

Das Thematisieren als Strukturieren steht darüber hinaus in einem engen Zusammenhang mit dem Wissen über prototypische Textmuster (siehe Schank 1977a; van Dijk 1980; Rothkegel 1984b; van Dijk 1985; 1997; Stede 2007). Dieser Zusammenhang wurde in der Forschung vor allem unter dem Terminus des „Vertextungsmusters“ beschrieben (siehe Kapitel 2.3.3). Vertextungsmuster haben sich

im Lauf der historischen Entwicklung auf Grund erfolgreicher kommunikativer Erfahrungen vieler Generationen der Kommunizierenden als feste Grundmuster der Textkonstitution herausgebildet. (Heinemann und Heinemann 2002, 187)

Sie stellen demnach ein Grundgerüst für die Herstellung eines Textes und damit u.a. auch für die Entfaltung des Themas dar, an dem sich Textproduzenten orientieren können – vorausgesetzt sie haben die sprachliche Fähigkeit, diese kommunikativen Routinen in einer konkreten Kommunikationssituation adäquat einzusetzen (vgl. Heinemann und Viehweger 1991, 112). Diese Art der Strukturierung von Texten im Akt der Textproduktion entspricht der Stufe Drei der in Feilke (1996, 1186) genannten „Stufenfolgen in der Ausbildung von Kohärenzprinzipien“, der Stufe der „formalen Ordnung“. Diese Stufe setzt die „Fähigkeit zu einer Orientierung an rein formalen Textordnungskriterien voraus“, z.B. argumentativen Schemata wie „Pro-Contra-Conclusio“ (vgl. Feilke 1996, 1187).

Im Hinblick auf den Prozess der Textproduktion stellt sich allerdings die Frage, inwiefern es sich bei den Vertextungsmustern tatsächlich um Muster handelt,

die ein Textproduzent bei der Erstellung eines Textes nutzt, d.h. ob diese Muster Teil des Textstrukturwissens sind, das Produzenten beim Textproduktionsprozess nutzen (siehe Heinemann und Viehweger 1991).<sup>236</sup> Insbesondere bei komplexeren Texttypen, für die eine Mischung aus verschiedenen Vertextungsmustern charakteristisch ist, wie z.B. wissenschaftliche Artikel (sowohl explikativ als auch argumentativ, teilweise auch deskriptiv), stellt sich die Frage, inwiefern Textproduzenten sich an abstrakten Vertextungsmustern oder konkreten Textsortenmustern orientieren, oder möglicherweise auch an „natürlichen Ordnungsstrukturen“ (z.B. kausale oder temporale Verknüpfungen) (vgl. Wrobel 1995, 139).

In diesem Kapitel wurde dargelegt, dass das *Thematisieren als textorganisierende Handlung* betrachtet werden kann. Die Handlung des Thematisierens interagiert dabei mit verschiedenen Faktoren. Zu diesen Faktoren gehören sowohl die Konversationsprinzipien der Relevanz und Verständlichkeit, als auch das Wissen über Sachverhaltstrukturen, Texttypstrukturen, das Wissen über sprachliche Mittel zur Etablierung, Beibehaltung und zum Wechsel von Themen. Im Hinblick auf die Maxime der Relevanz wurde dargelegt, dass Textproduzenten beim Verfassen von Texten daran gebunden sind, thematisch relevante Äußerungen zu einem kohärenten Text zu verknüpfen. Mögliche Aspekte des Themas müssen zunächst selektiert und dann sowohl auf der lokalen als auch globalen Ebene kohärent miteinander in Bezug gesetzt werden. Eine Teilhandlung des Thematisierens als textorganisierender Handlung stellt demnach das *Thematisieren als Selektieren* dar, durch das in Interaktion mit der Maxime der Relevanz eine Auswahl aus der Menge der potentiellen Themen bzw. Aspekte des Themas getroffen wird. An das Selektieren schließt sich das *Thematisieren als Strukturieren* an (wobei diese Teilhandlungen des Thematisierens ineinander übergehen und nicht etwa als aufeinanderfolgende Handlungen aufzufassen sind). Das Thematisieren als Strukturieren wird in Form von drei Teilhandlungen realisiert: Dem Thematisieren als *Organisieren*, *In-Bezug-Setzen* und *Fokussieren und Gewichten*. Sowohl beim Thematisieren als Selektieren als auch beim Thematisieren als Strukturieren erfolgt eine Interaktion mit verschiedenen Wissensbeständen, insbesondere mit dem Sachwissen und dem Handlungswissen (insbesondere Wissen über Textmuster, aber auch dem Wissen über Kommunikationsprinzipien), dem Sprachwissen (insbesondere dem Wissen über die Möglichkeiten der Signalisierung des Themas sowie der Zusammenhänge zwischen Themen – und zwar als Wissen über informationsstrukturelle, lexikalische, referenzielle und grammatische Ausdrucksformen in einer spezifischen kommunikativen Situation) sowie dem Wissen über die Kommunikationssituation. Diese Wissensbestände können sich wiederum durch die Handlung des Thematisierens verändern und durch den Textproduktionsprozess erweitert, aktualisiert und neu strukturiert werden (epistemische Funktion des Schreibens).

---

<sup>236</sup> Wobei hier natürlich auch zu berücksichtigen ist, dass sich Schreibfähigkeiten entwickeln und Schreibende sich demnach in ihren Möglichkeiten, auf Schreibroutinen zuzugreifen und diese zu nutzen, unterscheiden (vgl. Feilke 1996; Molitor-Lübbert 1996).

### 3.4.2 Thema als „Verstehenshorizont“

Für Rezipienten dient das Thema als „Verstehenshorizont“<sup>237</sup> (Fritz 1994, 193) bzw. „Einordnungsinstanz für das Verstehen“ (Hoffmann 1997, 513), in die sie Äußerungen einordnen können. Der Rezipient baut mit Beginn des Textverstehensprozesses und auf Basis der ersten rezipierten Äußerungen des Textes eine Erwartung darüber auf, was das Thema des Textes ist. Ausgelöst werden Erwartungen durch „situative oder textuelle Präsignale“ (Heinemann und Viehweger 1991, 260). Eines dieser textuellen Präsignale, vielleicht das Wichtigste überhaupt, ist der Titel des Textes, der die Funktion einer Themenankündigung bzw. Themenangabe hat. Weitere Quellen von Erwartungen sind z.B. Strategien zur Informationsstrukturierung (Wissen über lokale sprachliche Muster, z.B. Wissen über die Thema-Rhema-Gliederung), die Konventionen der Textsorte (Wissen über globale sprachliche Muster), der unmittelbare Kontext des Textes (Wissen über die Kommunikationssituation) sowie das stereotypische Sachwissen (vgl. de Beaugrande und Dressler 1981, 153-157).

Erwartungen und Wissen sind demnach eng miteinander verknüpft: Das Zusammenspiel von Präsignalen (datengeleitetes Textverstehen, Bottom-up-Strategie) und Wissensbeständen (wissensgeleitetes Textverstehen, Top-Down-Strategie) produziert zunächst initiale Erwartungen, die dann sukzessive im weiteren Verlauf der Textverarbeitung durch neue Signale bestätigt oder revidiert werden. Für die Rolle des Themas beim Textverstehen heißt das, dass ein Thema zu Beginn eines Textes zunächst als Präsignal gesetzt wird, indem es durch den Titel signalisiert oder durch die ersten Äußerungseinheiten eingeführt wird, und auf diese Weise bestimmte initiale Erwartungen auslöst, die das Textverstehen leiten.

The initial or final position of topical expressions follows from their semantic functions. At the beginning they help the hearer to form a hypothesis about the topic of the discourse or the episode, so that the following sentences can be interpreted top down relative to that macroproposition, whereas in the final position, topical expressions serve to check, remind, or correct the already established macropropositions of the hearer, as well as to repeat what was already known. (van Dijk und Kintsch 1983, 202)

Im Verlauf des Textverstehensprozesses können diese thematischen Erwartungen bestätigt oder revidiert werden. Allerdings ist zu vermuten, dass diese Revisionen sich innerhalb eines bestimmten Rahmens bewegen. Initiale Themensetzungen eröffnen demnach einen Verstehenshorizont, der als Einordnungsinstanz oder thematischer Rahmen fungiert, in den die sprachlichen Teilhandlungen und Teilthemen des Textes, auf die diese Handlungen bezogen sind, eingeordnet

<sup>237</sup> Der Begriff des „Horizonts“ wird in der Literatur unterschiedlich definiert. Während Fritz davon ausgeht, dass das Verstehen des Textes bezogen ist auf einen spezifischen inhaltlichen Rahmen (den Horizont), der durch ein spezifisches Thema definiert ist und thematisch relevante Textinhalte umfasst, versteht Scherner (1984, 92) den Begriff des „Horizonts“ als (Wahrnehmung-)Hintergrund, von dem sich das Thema als Vordergrund abhebt. Der Horizont umfasst demnach keinen spezifischen thematischen Bereich, sondern stellt einen allgemeinen Verstehenshintergrund dar. Für die Textrezeption bedeutet das, dass vor dem allgemeinen Horizont des Lesers bestimmte „Konstellationen des Bewußtseinsfeldes“ (bestimmte „thematische Felder“) als relevant hervorgehoben werden.

werden können. Aus der Setzung des Themas als Verstehenshorizont resultieren gleichzeitig Erwartungen über das, was kommen wird: "Once given a certain theme about some event we may predict what will happen" (van Dijk und Kintsch 1983, 196). Diese Erwartungen beziehen sich auf die Teilthemen, die bei der Behandlung eines Textthemas auftreten können – das Thema wirkt als „top-down-Orientierung“ für das Textverstehen (s.o.). Dass ein Rezipient Erwartungen über Teilthemen entwickelt, basiert in kommunikativ-interaktionaler Perspektive auf dessen Wissen über die Konversationsmaxime der Relevanz (siehe Kapitel 3.4.1): Der Rezipient nimmt an, dass der Produzent sich relevant äußert. Er versucht deshalb, die sprachlichen Handlungen als thematisch relevant zu interpretieren. Die Suche nach Relevanz im Prozess der Textrezeption entspricht der Suche nach und Konstruktion von thematischer Kohärenz (vgl. Blakemore 2001, 106).

Zweck der Thematisierung ist die Herstellung einer gemeinsamen Orientierung auf ein Thema. Er ist erreicht, wenn ein Thema (...) als Einordnungsinstanz für das Verstehen eingeführt ist. Mit sprachlichen Mitteln wird der Adressat dazu gebracht, den betreffenden Gegenstand oder Sachverhalt als Thema dessen, was folgt, zu übernehmen. (Hoffmann 1997, 513f.)

Dass initiale Themensetzungen einen Verstehenshorizont für den Text schaffen und Erwartungen bezüglich des Inhalts des Texts evozieren, hängt aus einer kognitionsorientierten Sicht darüber hinaus damit zusammen, dass Entitäten (Gegenstände, Ereignisse, Handlungen, Zustände etc.) im Gedächtnis nicht einzeln abgespeichert sind, sondern in größeren Wissensseinheiten, d.h. zusammen mit verwandten Entitäten, und dass bei der Aktivierung einer Entität gleichzeitig auch die verknüpften Entitäten mit aktiviert werden:

Wenn eine Wissensseinheit aktiviert wird, so werden offenbar andere im Speicher eng verbundene Einheiten ebenfalls aktiv (...). Dieses Prinzip nennt man oft AKTIVIERUNGS-VERBREITUNG (engl. „spreading activation“). (de Beaugrande und Dressler 1981)

Wissenseinheiten bilden demnach eine „globale Wissenskonfiguration“ (de Beaugrande und Dressler 1981, 166). In einem Text wird diese Wissenskonfiguration aktiviert, wenn eine Wissensseinheit bzw. Entität, die Teil des Sachverhaltskomplexes ist, im Text sprachlich referenziert wird: "The key words and ideas of a discourse evoke substantial thematic or scenario structures, drawn from memory with rich default assumptions." (Minsky 1980, 16). Was im Text zu Beginn kommuniziert wird (z.B. der Titel oder der Einleitungssatz, insbesondere zu Beginn des Textes genannte Nominalphrasen oder nominalisierte Verbalphrasen), kann so „zum thematischen Ausgangspunkt framegestützter Wissensexplikationen“ gemacht werden (vgl. Konerding 1993, 211). D.h., wenn eine Entität A, die Teil des Frames X ist, im Text referenziert wird, ist deshalb nicht nur Entität A mental aktiviert, sondern auch die ebenfalls in Frame X enthaltenen Entitäten B-Z (diese allerdings nur semi-aktiv) (vgl. de Beaugrande und Dressler 1981, 93; Chafe 1998, 100f.).<sup>238</sup> Das Verstehen der thematischen Organisation

<sup>238</sup> Während de Beaugrande/Dressler davon ausgehen, dass ein im Text ausgedrücktes Thema einen bestimmten Rahmen aktiviert, definiert Chafe das Thema selbst als einen solchen Rahmen: "A topic is a unified thought complex that occupies, for a longer period of time, the speaker's semiactive consciousness" (Chafe 1998, 100f.).

eines Textes basiert nach dieser Auffassung ebenfalls auf dem Wissen über den Zusammenhang von Thema und Teilthema:

Given Topic – Sub-Topic relations likewise contribute to topical continuity, but in a less direct way, mediated through frames which allow for bridging assumptions in which missing links are reconstructed. A Given Topic – Sub-Topic sequence is interpreted as coherent if the interpreter, aided by his frame knowledge, can reconstruct the step(s) which allow S to assume that, once the Given Topic has been established, the Sub-Topic may be assumed to be available as well. (Dik 1997, 437)

Wenn in einem Text beispielsweise eine Wohnung beschrieben wird, die „Wohnung“ also als zentrales Textthema aufgefasst werden kann, können die Entitäten „Schlafzimmer“, „Wohnzimmer“, „Küche“ etc. als erwartbare – weil semiaktive – Teilthemen des Textes gelten, da sie Teil des gemeinsamen kognitiven Frames „Wohnung“ sind, der durch das Textthema gleich zu Beginn des Textes aktiviert wird. Ob alle Teile des Frames in gleichem Maße aktiviert sind und für das wissensgeleitete Verstehen des Textes zur Verfügung stehen, ist allerdings unklar. Während Johnson-Laird (1988) und de Beaugrande/Dressler (1981) keine weitere Differenzierung in der Gradienz der Aktivierung des Sachwissens vornehmen, gehen Sanford/Garrod (1981, zit. n. Rickheit et al. 2002) bzw. Sanford (1995) und Garrod (1995) davon aus, dass es einen impliziten und einen expliziten Fokus gibt. Im expliziten Fokus stehen jene Teile des mentalen Modells, die durch den aktuellen Textabschnitt aktiviert wurden – „the currently relevant discourse entities“ –, im impliziten Fokus dagegen die aktuellen kognitiven Szenarios (bzw. Schemata) – „the currently relevant scenarios“ –, die auf Grundlage des expliziten Fokus und des allgemeinen Weltwissens aktiviert wurden (Garrod 1995, 4). Der Aufbau von Kohärenz im Verstehensprozess wird umso schwieriger, je mehr Wechsel es bei den aktivierten Frames und Teilthemen gibt<sup>239</sup> bzw. desto mehr Entitäten aus verschiedenen Szenarios in den expliziten Fokus gesetzt werden – „die mentale Kohärenzbildung [wird] durch häufige Topic-Wechsel erschwert“ (Schnotz 1994, 191); ein kontinuierlicher thematischer Textaufbau führt dagegen zu „signifikant höheren Leistungen in einem Verstehenstest“ (Schnotz 1994, 191).

Das im Text repräsentierte Thema als das den kognitiven Frame aktivierende Konzept ist demnach zentral für den Textverstehensprozess, indem es das Textverstehen von Beginn an in eine spezifische Richtung lenkt (Erwartungen weckt über die Inhalte des Textes). Wie in den diversen, in Kapitel 2.8.2 dargestellten Textverstehensmodellen deutlich wurde, verläuft das Textverstehen zyklisch, d.h. Erwartungen werden durch die Interaktion von initialem Textmaterial und verschiedenen Wissenskomponenten geweckt und schließlich durch darauf folgendes neues Textmaterial revidiert oder modifiziert. Das Textverstehen wird so durch die Themasetzung angestoßen, indem Erwartungen über ein mögliches mentales Textmodell geweckt werden; die weitere Themenentwicklung im Textverlauf führt dann zu Revisionen dieses Modells.

<sup>239</sup> In einem Text wird in der Regel mehr als nur ein Frame sprachlich abgehandelt; Konearding (1993, 210) geht deshalb davon aus, dass es komplexe „Superframes“ gibt, die aus elementaren Frames aufgebaut sind.

### 3.4.3 Architektur und Komponenten eines Systems zur maschinellen Themenidentifikation

In der vorliegenden Arbeit wurde ein Ebenen-basiertes Thema-Modell vorgestellt, das als Grundlage für ein System zur maschinellen Themenidentifikation dienen kann. In diesem Thema-Modell wird das Thema auf drei Haupt- und sechs Unterebenen verortet: Auf der Ebene der Kognition (Thema als Sachverhaltskomplex), der Kommunikation (Thema als Orientierungsprinzip und Einordnungsrahmen) und der Ebene des Textes mit seinen sechs Beschreibungsebenen (Informationsstruktur, referenzielle, lexikalische, relationale, visuelle oder logische Struktur sowie die generische Texttypstruktur). Insofern das Thema als emergente Größe des Textes aufgefasst wird, ist das Thema auf der Textebene nur mittelbar, über die Reflexionen des Themas im Text enthalten. Eine besondere Rolle spielen hierbei die Diskursentitäten, über die im Text etwas gesagt wird; sie können als Reflexionen der Aspekte des Themas aufgefasst werden. Darüber hinaus werden im Text aber auch die Relationen zwischen den Aspekten des Themas reflektiert, sowohl durch Kohärenzrelationen als auch durch lexikalisch-semantische Relationen als auch durch Relationen auf Basis von Texttypstrukturschemata. Im Folgenden soll zunächst noch einmal auf das Zusammenspiel der verschiedenen Ebenen des Thema-Modells eingegangen werden. Die zu diesem Zweck erstellte, auf dem allgemeinen Modell zum Zusammenspiel von Textproduktion/-rezeption, Text und Wissen (siehe Abbildung 6, Kapitel 2.9) basierende Visualisierung (siehe Abbildung 25) dient dabei gleichzeitig als Vorlage für die Architektur eines Systems zur maschinellen Themenidentifikation (siehe Abbildung 26), welches im Anschluss näher beschrieben wird.<sup>240</sup>

Das in Abbildung 25 visualisierte Modell zum Zusammenspiel der Ebenen des Themas besteht aus vier zentralen Komponenten: A.1 Wissen, B.1 Thematisieren als Textproduktionshandlung, C.1 Text, D.1 Thema als Verstehenshorizont für die Textrezeption. Diese Komponenten und ihr Zusammenspiel sollen im Folgenden noch einmal kurz erläutert werden.

Die textorganisierende Handlung des Thematisierens mit ihren Teilhandlungen des Selektierens und Strukturierens (Abbildung 25, Komponente B.1) wird einerseits geprägt durch die individuellen Wissensbestände des Textproduzenten (z.B. wirkt sowohl das Sachwissen über das Thema als auch das Wissen über thematische Textmuster darauf ein, wie ein Text zu einem bestimmten Thema organisiert ist), andererseits können die Wissensbestände des Produzenten über die Handlung des Thematisierens verändert und aktualisiert, erweitert oder umstrukturiert werden (Abbildung 25, Komponente A.1). In ähnlicher Weise findet auch beim Textverstehen eine Interaktion mit den Wissensbeständen statt. Die ersten rezipierten Äußerungseinheiten des Textes enthalten potentielle indexikalische Zeichen für das Thema (Abbildung 25, Komponente C.1), die vom Rezi-

---

<sup>240</sup> Die Bezüge zwischen den drei Modellen spiegeln sich in der Benennung ihrer Komponenten wider: In Abbildung 6 werden die vier Komponenten Wissen, Textproduktion, Text und Textrezeption mit den Buchstaben A bis D, in Abbildung 25 als A.1 bis D.1 und in Abbildung 26 als A.2 bis D.2 bezeichnet.

pienten unter Rekurs auf sein Wissen über Sachverhalte – aber auch sein Sprach- und Handlungswissen (z.B. Wissen über die Funktion des Titels in expositorischen Texten im Hinblick auf die Signalisierung des Themas) – zur Bildung einer initialen Hypothese über das Thema des Textes genutzt werden.<sup>241</sup> Diese Hypothese fungiert als Verstehenshorizont für den weiteren Rezeptionsprozess; sie kann allerdings auf Basis des Text-Inputs im weiteren Verlauf des Rezeptionsprozesses zyklisch aktualisiert und revidiert werden. Ergebnis des Verstehensprozesses ist schließlich ein mentales Modell des Themas als Sachverhaltskomplex (Abbildung 25, Komponente D.1). Insofern sich die Wissensbestände verschiedener Rezipienten (und Produzenten) interindividuell unterscheiden können, konstruiert jeder Rezipient dabei ein individuelles mentales Modell des Textes.<sup>242</sup>

---

<sup>241</sup> Die Textrezeption wirkt dabei – wie die Textproduktion – auf die Wissensbestände des Rezipienten bzw. Produzenten zurück.

<sup>242</sup> In der Abbildung steht die Abkürzung „aE“ für „aktualisierte Erwartung“, „rE“ für „revidierte Erwartung“.

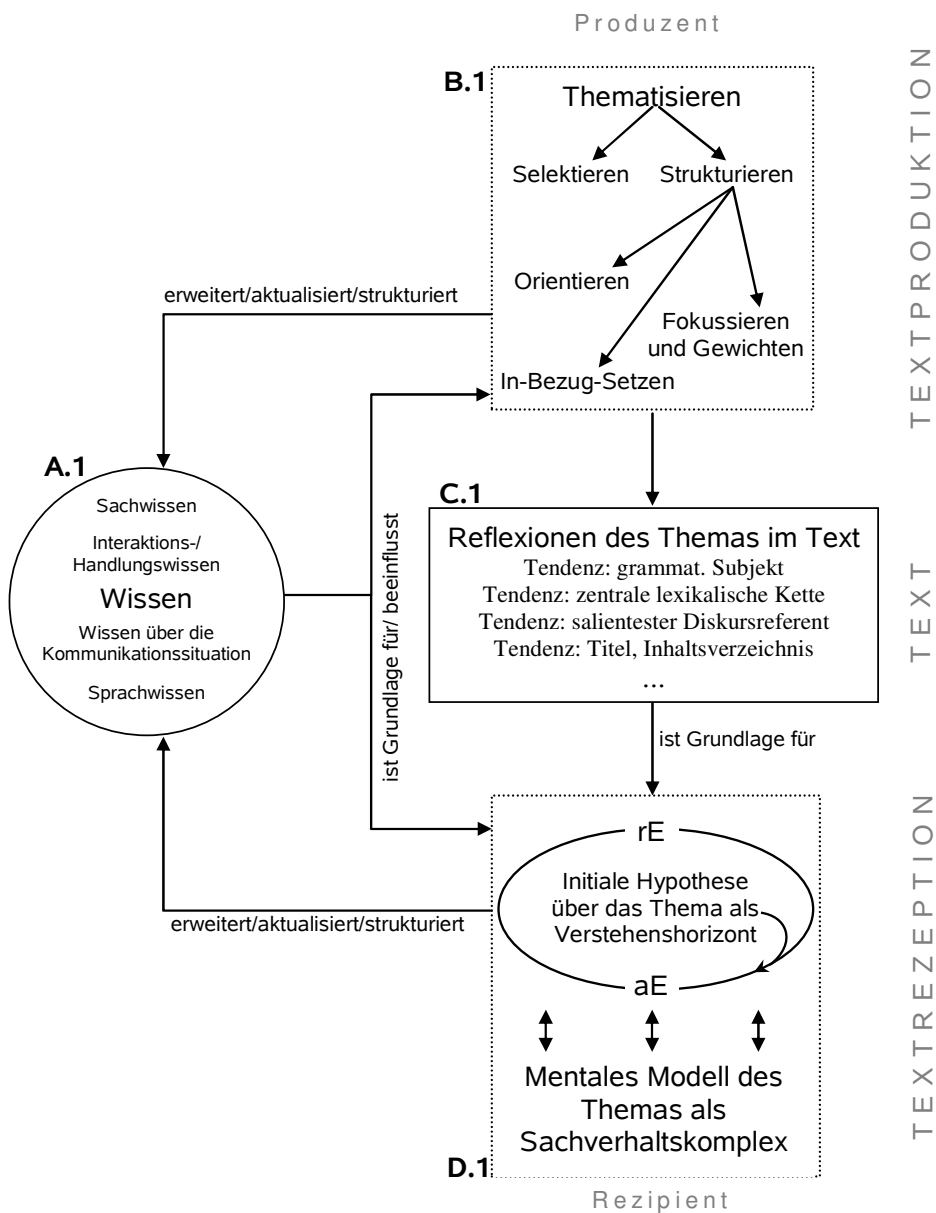


Abbildung 25: Prozessmodell zum Zusammenhang der Ebenen des Themas



Nachfolgende Tabellen dienen der weiteren Spezifizierung und Konkretisierung der Komponenten A.1 und D.1. Tabelle 12 gibt eine Übersicht über die Wissensbestände bzw. Erwartungen, die das Thematisieren sowie das Verstehen der thematischen Organisation beeinflussen; Tabelle 13 stellt noch einmal die Ebenen der Textstruktur dar, auf denen das Thema durch unterschiedliche sprachliche und visuelle Mittel reflektiert wird.

	<b>Sachwissen</b>	<b>Handlungswissen</b>	<b>Sprachwissen</b>	<b>Wissen über die Kommunikationssituation</b>
<b>Produzent</b>	Orientierung am Wissen über stereotypische Sachverhaltsstrukturen als Grundlage für die Selektion und Strukturierung der Aspekte des Themas im Text	Orientierung am Wissen über unter-spezifizierte texttyp-spezifische Thementypen sowie Orientierung an Konversationsprinzipien (z.B. Sei relevant! Sei verständlich!) als Grundlage für die Selektion und Strukturierung der Aspekte des Themas im Text	Umsetzung des Wissens über sprachliche Möglichkeiten der Signalisierung des Themas (Koreferenz, lexikalische Rekurrenz, Titel ...) als Grundlage für die Selektion und Strukturierung der Aspekte des Themas im Text	Wissen über die Faktoren, die die Kommunikationssituation prägen, als Voraussetzung für die verständliche inhaltlich-thematische Organisation des Textes
<b>Rezipient</b>	Erwartung, dass sich der Produzent an diesen Sachverhaltsstrukturen orientiert und diese so als Quelle für die Konstruktion eines mentalen Korrelats des Textinhalts/-themas genutzt werden können	Erwartung, dass sich der Produzent an diesen Kategorien, Mustern und Prinzipien orientiert und diese so als Quelle für die Konstruktion eines mentalen Korrelats des Textinhalts/-themas genutzt werden können	Erwartung, dass das sprachliche Material als Quelle für die Konstruktion eines mentalen Korrelats des Textinhalts/-themas genutzt werden kann	Erwartung, dass der Produzent die Kommunikationssituation beim Verfassen des Textes berücksichtigt hat und diese so als Quelle für die Konstruktion eines mentalen Korrelats des Textinhalts/-themas genutzt werden können

Tabelle 12: Wissensbestände und Erwartungen, die das Thematisieren sowie das Verstehen der thematischen Organisation beeinflussen

<b>Informationsstruktur</b>	Thema als Mitteilungsgegenstand des Satzes: Tendenz: grammat. Subjekt; Kasusrolle AGENS; Satzerstposition; Definitheit
<b>Lexikalische Struktur</b>	Thema als rekurrente Information; Thema als (eine) Quelle für Wortschatz: Tendenz: Gruppe zentraler Inhaltswörter; zentrale lexikalische Kette mit den meisten Querbezügen zu anderen Ketten
<b>Referenzielle Struktur</b>	Thema als zentraler Gegenstand des Textes: Tendenz: salientester Diskursreferent bzw. Diskursreferent mit den meisten Bezügen zu anderen Diskursreferenten
<b>Relationale Struktur</b>	Zusammenhänge zwischen Aspekten des Themas: Tendenz: semantische Relationen, anaphorische Relationen, lexikalische Ketten, Diskurs- bzw. Kohärenzrelationen, indirekte Relationen über Einheiten der Texttypstruktur
<b>Generische Texttypstruktur</b>	Aspekte des Themas als instanziierte Einheiten der Texttypstruktur bzw. der texttypspezifischen Thementypenstruktur
<b>Logische Textstruktur</b>	Tendenz: Titel, Inhaltsverzeichnis, Hervorhebung Tendenz: Themawechsel wird signalisiert durch Absatzbildung Tendenz: spezifische Position im Satz, Absatz, Text

Tabelle 13: Reflexionen des Themas auf den verschiedenen Ebenen der Textstruktur

Das Modell zum Zusammenspiel der Ebenen des Themas kann als Vorlage für die Entwicklung einer Architektur für ein System zur maschinellen Themenidentifikation genutzt werden (da es bei der automatischen Themenidentifikation allerdings ausschließlich um die Perspektive des Textverstehens und nicht des Textproduzierens geht, fällt Komponente B in einem System zur automatischen Themenidentifikation weg). Um das Modell in eine Software-Architektur zu transformieren; müssen zunächst einmal die einzelnen Komponenten in Input-, Output- und Prozess-Komponenten überführt werden. Zu den Input-Komponenten zählen Wissen und Text – beide müssen allerdings formal repräsentiert sein, um maschinell verarbeitet werden zu können. In Kapitel 3.2.5 wurde zu diesem Zweck ein Überblick über die zur Verfügung stehenden maschinell verarbeitbaren Wissensressourcen gegeben (Komponente A.2). In Kapitel 3.3.3 wurde dargestellt, in welcher Weise die einzelnen Reflexionen des Themas im Text bzw. die potentiellen indexikalischen Zeichen für das Thema in Form von Multi-Ebenen-Annotationen formal repräsentiert und mittels verschiedener Auswertungsverfahren und -werkzeuge miteinander verrechnet werden können; in den Kapiteln davor wurden darüber hinaus computerlinguistische Methoden vorgestellt, die zur maschinellen Analyse der unterschiedlichen Reflexionen des Themas im Text genutzt werden können (Komponente C.2) – wobei diese Methoden teilweise wissensbasiert sind, so dass bereits zur Erstellung der einzelnen Annotationsschichten Wissen benötigt wird. Komponente D.1 (Rezeption) wird im Architekturmodell in zwei Teilkomponenten untergliedert: in eine Prozess-

Komponente (D.2.a), d.h. dem eigentlichen Verarbeitungsprozess, der von der initialen Hypothese über Revisions- und Aktualisierungsprozesse zum Output führt, und in eine Output-Komponente (D.2.b), d.h. dem Ergebnis des Rezeptions- bzw. Verarbeitungsprozesses. Die Komponente D.2 muss einerseits im Hinblick auf das Ausgabeformat (die formale Repräsentation des Modells des Sachverhaltskomplexes) spezifiziert werden (Komponente D.2.b), andererseits im Hinblick auf die Prozesse (D.2.a), die nötig sind, um aus den Input-Komponenten A.2 und C.2 die Output-Komponente zu erzeugen. Prozesskomponente D.2.a. hat dabei die Aufgabe, die Top-Down-Constraints (A.2) und Bottom-Up-Cues (C.2) iterativ zu verarbeiten und miteinander zu verrechnen (zu den Verrechnungen von Multi-Ebenen-Annotationen siehe auch Kapitel 3.3.3). Aus diesen Verrechnungen und den (statistisch ermittelten) spezifischen Kombinationen von Cues und Constraints können einerseits Regeln bzw. Constraints zur Themaidentifizierung bzw. Themagewichtung abgeleitet werden. Andererseits resultiert aus dieser Verarbeitung der Input-Komponenten eine Output-Komponente, in der die zentralen Aspekte des Themas und ihre Zusammenhänge sowie die thematische Organisation des Textes festgehalten sind. Komponente D.2.a könnte in diesem Kontext zum einen ein graphisches Modell der im Text reflektierten Sachverhaltsstruktur ausgeben (D.2.b.2), in dem die im Text behandelten Aspekte des Themas als Knoten und ihre kognitiven Zusammenhänge als Kanten dargestellt sind.<sup>243</sup> Alternativ wäre auch denkbar, eine neue Annotationsschicht n+1 auszugeben (D.2.b.1), die z.B. eine validierte Variante der thematischen Organisation des Textes repräsentiert, die durch einen Vergleich der relationalen Diskursstruktur mit den auf lexikalischen Ketten basierenden Textsegmenten beruht.

In der in Abbildung 26 skizzierten Architektur eines Systems zur maschinellen Themenidentifikation sind die genannten Komponenten in Relation zueinander gesetzt und es wird dargestellt, welche Input- und Output-Komponenten ein solches Software-System aufweisen müsste. Offen bleibt in dem Modell die Umsetzung der Prozesskomponente, d.h. die „Verrechnung der Reflexionen des Themas“ (D.2.a) – diese ist momentan als Black Box dargestellt, d.h. die Prozesse, die im Inneren dieser Komponente ablaufen, bleiben – bis auf die in Kapitel 3.3.3 grob skizzierten Abläufe bei der „Verrechnung“ der Multi-Ebenen-Annotationen 1-n untereinander – unterspezifiziert. Welche konkreten Prozesse in dieser Verarbeitungskomponente neben den „Verrechnungen“ der Reflexionen auf den verschiedenen Textstrukturebene untereinander sowie mit den Wissensressourcen 1-n nötig und sinnvoll sind, um zu den dargestellten Output-Komponenten zu gelangen, und in welcher Weise diese Abläufe programmier-technisch umzusetzen sind, muss im Software-Entwicklungsprozess von Programmierexperten selbst geklärt werden.

---

<sup>243</sup> Insofern der im Text thematisierte Sachverhaltskomplex einen eindeutigen und unverwechselbaren thematischen Daumenabdruck darstellt, würde eine solche Graphenstruktur auch den Vergleich des analysierten Textes mit anderen Texten zum gleichen Thema ermöglichen, durch den Texte auf Basis ihrer thematischen Spezifika voneinander abgegrenzt werden können.

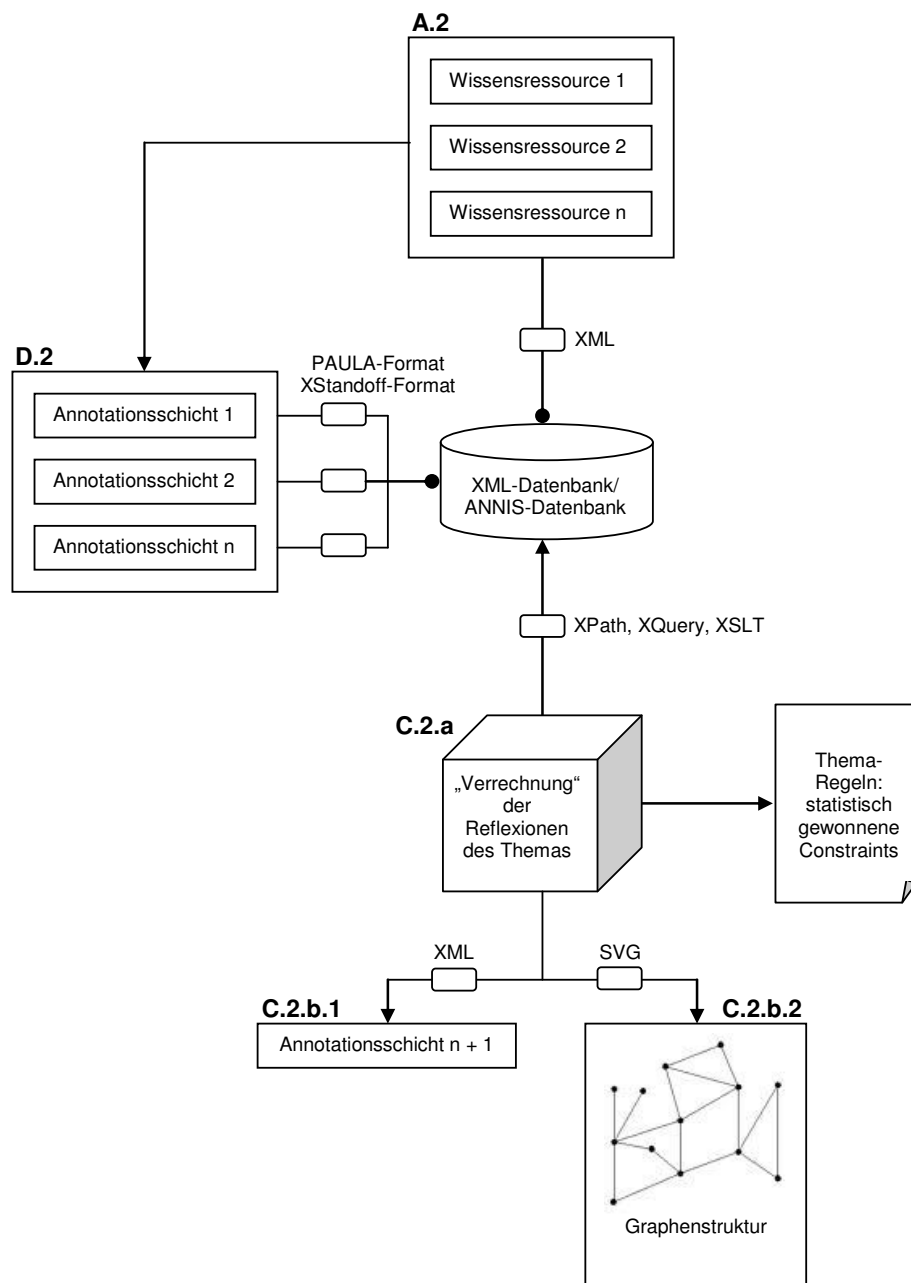


Abbildung 26: Architektur eines auf dem Ebenen-basierten Thema-Modell basierenden Software-Systems zur maschinellen Themenidentifikation

## 4 Zusammenfassung und Ausblick

Die vorliegende Arbeit stellt einen Beitrag zur textlinguistischen und computerlinguistischen Thema-Forschung dar, der zwei zentrale Desiderate erfüllt: Erstens die Entwicklung eines konsensfähigen, integrativen Thema-Konzepts und zweitens die Entwicklung eines theoretisch fundierten Modells zur Operationalisierung der Thema-Analyse. Im Hinblick auf Letzteres wurde aufgezeigt, welche Analyseschritte mittels welcher computerlinguistischen Verfahren automatisiert werden können, welche Ressourcen für eine „Verrechnung“ der Analysen der einzelnen Ebenen nötig und bereits verfügbar sind und wie die Architektur eines auf den theoretischen Überlegungen der vorliegenden Arbeit basierenden Softwaresystems zur maschinellen Themenidentifikation aussehen könnte.

Grundlage für die Entwicklung eines integrativen Thema-Konzepts war eine Neubewertung bisheriger Forschungsansätze, durch die zwei zentrale Erkenntnisse gewonnen wurden: Erstens konnte gezeigt werden, dass die meisten der linguistischen Thema-Konzepte keineswegs konkurrierende und einander ausschließende Thema-Auffassungen darstellen, sondern dass sie lediglich verschiedene Perspektiven auf das Thema widerspiegeln, d.h. als einander ergänzende Auffassungen interpretiert werden können. Zweitens wurde belegt, dass computerlinguistische Forschungsansätze zur maschinellen Themenidentifikation ein Theoriedefizit im Hinblick auf die diesen Arbeiten zugrunde liegende Thema-Konzeption aufweisen. Dieses Defizit ist mit dafür verantwortlich, dass zwischen den meisten Ansätzen zur Themenidentifikation und zur Keyword- bzw. Schlüsselwort-Identifikation kein großer Unterschied besteht – was in der computerlinguistischen Forschungsdebatte allerdings keineswegs reflektiert wird. Darüber hinaus wurden in der vorliegenden Arbeit zwei zu Beginn anvisierte Ziele erreicht: Zum einen wurde ein Thema-Konzept entwickelt, das die bisherigen, sehr vielfältigen linguistischen Thema-Konzepte in ein komplexes Modell zu integrieren erlaubt und das als Grundlage für computerlinguistische Arbeiten zur maschinellen Themenidentifikation geeignet ist, insofern auf Basis dieses Thema-Modells die Themenidentifikation eindeutig von der Keyword-Identifikation differenziert werden kann. Zum anderen wurde in der vorliegenden Arbeit ein Analysemodell für das Thema präsentiert, das sowohl eine Operationalisierung der Themenanalyse ermöglicht als auch auf einem theoretisch evaluierten Thema-Konzept basiert.

In der vorliegenden Arbeit werden in vielerlei Hinsicht neue theoretische und methodische Wege beschritten, so dass hier nicht nur ein Beitrag zur Thema-Forschung, sondern auch zur textlinguistischen Forschung insgesamt geleistet wurde. Im Folgenden werden die Ergebnisse der Arbeit deshalb detailliert zusammengefasst.

In Kapitel 2 wurde zunächst eine Übersicht über die geläufigsten linguistischen Thema-Konzepte gegeben. Neuartig an dieser Bestandsaufnahme ist zum einen die integrative Darstellung von Satz- und Textthema-Konzepten (diese werden in den bisher vorliegenden Überblicksarbeiten zum Thema üblicherweise als voneinander unabhängige Konzeptionen aufgefasst). Zum anderen werden die linguistischen Thema-Konzepte anhand der linguistischen Teilbereiche Grammatik, Semantik, Pragmatik und Kognition kategorisiert. Eine solch abstrakte Einordnung existierte bisher ebenfalls nicht. Stattdessen überwiegen in bisherigen Arbeiten Kategorisierungen auf Basis sehr spezifischer Kategorien (z.B. „Thema als Makroproposition“, „Thema als Fragestellung“, „Thema als Informationskern“). Die Vorstellungen von dem, was Linguisten als Thema auffassen, divergieren stark. Linguistische Thema-Konzepte unterscheiden sich einerseits im Hinblick auf ihre Haltung zur Frage, welche sprachliche Einheit (Satz oder Text) vom Thema beeinflusst wird (Satz- vs. Textthema-Konzepte), und andererseits, auf welcher sprachlichen Ebene das Thema verortet wird.<sup>244</sup> Auf Basis der zweiten Fragestellung wurden in der vorliegenden Arbeit grammatische, semantische, kommunikativ-pragmatische und kognitive Thema-Konzepte unterschieden (wobei deutlich gemacht wurde, dass Kreuzklassifikationen von Thema-Konzepten möglich sind). Auf der grammatischen Ebene kann das Thema als sprachliche Einheit des Satzes aufgefasst werden, die über Merkmale wie Satzposition, grammatische Rolle sowie thematische/semantische Rolle charakterisierbar ist. Insbesondere in den Arbeiten zur Informationsstruktur wird dabei deutlich, dass das Thema im Satz als pragmatische Einheit betrachtet wird, die in der Regel Teil einer bipolaren Einheit ist, z.B. Thema-Rhema, Topik-Kommentar oder Topik-Fokus. Semantische Thema-Konzeptionen unterscheiden sich untereinander, insofern das Thema teilweise als Gegenstand bzw. Diskursentität, über die im Satz oder Text etwas gesagt wird, aufgefasst wird, teilweise als Makro- oder Basisproposition oder als Prädikat-Aktanten-Struktur, die den semantischen Kern des Textes repräsentiert. Auch die kommunikativ-pragmatischen Auffassungen vom Thema divergieren. Das Thema wird teilweise als sprachliche Einheit des Satzes betrachtet, teilweise als Einheit, die auf den Text insgesamt bezogen ist. Besonders prominent im Hinblick auf die kommunikativ-pragmatischen Textthema-Auffassungen ist zum einen das „Thema-als-Fragliches“-Konzept, zum anderen das Thema-Konzept, das das Thema als Mittel zur Strukturierung von Diskursen auffasst. Auch die kognitiv orientierten Thema-Konzepte können weiter differenziert werden. Das Thema wird hier zum Teil über das Merkmal „Bekanntheit“ oder „mentale Aktiviertheit“ definiert, zum Teil wird es aber auch als Kern eines konzeptuellen Netzes aufgefasst, d.h. in einen Zusammenhang mit Wissensstrukturen gebracht. Die linguistischen Thema-Konzeptionen wurden in Kapitel 2 des Weiteren kontrastiert mit computer-

---

<sup>244</sup> In der vorliegenden Arbeit wurde allerdings – wie bereits angesprochen – auf eine generelle Differenzierung von Satz- und Textthema-Konzepten verzichtet, insofern insbesondere in den Arbeiten im Kontext der Prager Funktionalen Satzperspektive sowie der Systemisch-Funktionalen Linguistik die auf den Satz bezogenen Thema-Konzepte auf eine Folge von Sätzen angewandt bzw. übertragen wurden, um so die thematische Progression bzw. die Themenentwicklung im Text zu beschreiben. In diesen Ansätzen werden demnach keine grundsätzlichen konzeptuellen Unterschiede zwischen Satz- und Textthema gemacht.

linguistischen Auffassungen vom Thema. Dabei wurde deutlich, dass der Terminus „Thema“ in der Computerlinguistik häufig theoretisch unreflektiert und/oder zweckorientiert verwendet wird. Darüber hinaus zeigte sich, dass in der Computerlinguistik sehr selten auf linguistische Thema-Konzepte Bezug genommen wird – umgekehrt scheint die Textlinguistik maschinelle Verfahren zur Textverarbeitung allerdings ebenso wenig wahrzunehmen. Im Zuge einer stärkeren Wahrnehmung der Korpuslinguistik und ihrer Methoden ist hier allerdings seit einiger Zeit ein allmählicher Wandel zu konstatieren.

Auf Basis der Bestandsaufnahme und Kategorisierung linguistischer Thema-Konzepte wurde in den nachfolgenden Kapiteln der Frage nachgegangen, warum so viele, teils disparate linguistische Thema-Konzepte existieren und ob diese Konzepte – aus einer übergeordneten Perspektive gesehen – wirklich unvereinbar sind. Um diese Fragen beantworten zu können, wurde in Kapitel 4 ein Ebenen-basiertes Thema-Modell entwickelt, das das Thema als Konzept auffasst, das auf den drei Ebenen „Text“, „Kommunikation“ und „Kognition“ wirkt und auf diesen Ebenen jeweils unterschiedliche, aber dennoch aufeinander bezogene Funktionen hat (Kapitel 3.1.1): Auf der Ebene der Kognition kann das Thema als Sachverhaltskomplex aufgefasst werden, der die Funktion eines kognitiven Einordnungsrahmens hat. Auf der kommunikativen Ebene stellt das Thema bzw. die Orientierung am Thema (als Sachverhaltskomplex) ein Organisationsprinzip dar, das die Textproduktion und -rezeption steuert. Das Thema hat auf dieser Ebene die Funktion eines kommunikativen Einordnungsrahmens, an dem sich Produzenten und Rezipienten beim Produzieren bzw. Verstehen des Textes orientieren können. Die Orientierung an einem solchen thematischen Einordnungsrahmen führt wiederum dazu, dass das Thema auf der Ebene des Textes sprachlich sichtbar wird – indem Produzenten etwas thematisieren, wird dieses Thema auf den Ebenen der Textstruktur durch die Menge der zentralen Diskursentitäten, über die etwas gesagt wird, realisiert und durch sprachliche und visuelle Mittel reflektiert. Rezipienten wiederum dienen die Reflexionen des Themas im Text als indexikalische Zeichen, die die Konstruktion eines thematischen Rahmens bzw. Sachverhaltskomplexes unterstützen (siehe Abbildung 27).

Anhand der Einordnung der in Kapitel 2 dargestellten Thema-Konzepte in die Ebenen dieses integrativen Thema-Modells (Kapitel 3.1.2) konnte belegt werden, dass die beschriebenen Thema-Konzepte keineswegs unvereinbar sind, sondern dass sie das Phänomen „Thema“ lediglich aus sehr unterschiedlichen Perspektiven betrachten. Die scheinbar divergierenden Thema-Konzeptionen können deshalb als Facetten eines komplexen Konzepts verstanden werden, die jeweils eine unterschiedliche Ebene in den Fokus nehmen: die Rolle des Themas auf Textebene (Thema als Mitteilungsgegenstand/Redegegegenstand, über den im Satz/Text etwas gesagt wird bzw. der im Text in spezifischer Weise entwickelt wird; Thema als Element in Satzerstposition/als Ausgangspunkt der Mitteilung; Thema als Cluster von Schlüsselentitäten; Thema als zentraler, salienter Gegenstand des Textes), die Rolle des Themas auf der Ebene der Kommunikation (Thema als Fragliches, das im Text beantwortet/behandelt wird; Thema als „Verstehenshorizont“; Thema als Mittel zur Strukturierung von Diskursen) oder die Rolle des Themas auf der Ebene der Kognition (Thema als kognitiv am

stärksten aktivierte Diskursentität; Thema als Kern eines konzeptuellen Netzes/Schemas).

Die drei Ebenen des Thema-Modells (Kommunikation, Kognition und Text) spielen in der vorliegenden Arbeit nicht nur für die Einordnung der in Kapitel 2 beschriebenen Thema-Konzepte eine zentrale Rolle, sondern auch für das Analysemodell und – darauf aufbauend – das skizzierte Softwaresystem zur maschinellen Themenidentifikation. Für die Analyse des Themas werden die Ebenen der Kognition und des Textes als modulare, miteinander interagierende (Input-)Komponenten aufgefasst. Die kognitive Komponente fungiert für die Themenanalyse als Menge von Top-Down-Constraints, die Text-Komponente als Menge aufeinander bezogener Bottom-Up-Cues. Die Ebene der Kommunikation dient als Prozesskomponente bzw. Schnittstelle zwischen dem Thema auf Text- und dem Thema auf Wissensebene. Die Relevanz dieser drei Komponenten wurde sowohl in Kapitel 0 als auch Kapitel 3 ausführlich begründet und anhand exemplarischer Analysen illustriert.

Im Hinblick auf die erste (Input-)Komponente „Kognition“ wurden die für das Produzieren und Verstehen des Themas bzw. der thematischen Organisation des Textes zentralen Wissensbestände zunächst eingeführt (Kapitel 2.7.1) und schließlich im Detail entfaltet (Kapitel 3.2). Dabei wurde für jeden der vier Wissensbestände (Sachwissen, Handlungs-/Interaktionswissen, Sprachwissen und Wissen über die Kommunikationssituation) dargestellt, welches spezifische Wissen insbesondere beim Produzieren und Verstehen des Themas bzw. der thematischen Organisation eine Rolle spielt. Darüber hinaus wurde die Verfügbarkeit der jeweiligen Wissensbestände in Form ihrer digitalen Korrelate (bzw. Wissensressourcen oder Language Resources) beurteilt (Kapitel 3.2.5). Dabei wurde deutlich, dass zum jetzigen Zeitpunkt sämtliche Wissensbestände (außer dem Sprachwissen) nur in ausreichender bis mangelhafter Weise als digitale Ressourcen zur Verfügung stehen. Insbesondere im Hinblick auf das Interaktions- und Handlungswissen existieren zur Zeit nur sehr marginale Ansätze zur formalen Modellierung – diese sind bezogen auf die formale Beschreibung und Repräsentation des Wissens über bestimmte Typen von Textmustern (z.B. Wissen über die abstrakte Thementypen- bzw. Texttypstruktur wissenschaftlicher Artikel). Für andere Arten des Interaktionswissens ist dagegen bislang völlig offen, in welcher Form Ressourcen erstellt werden könnten, d.h. wie dieses Wissen erfasst, strukturiert und repräsentiert werden könnte bzw. sollte. Ähnliches gilt für das Sprachgebrauchswissen (als einem zentralen Bestandteil des Sprachwissens, der mit dem Sach- und Handlungswissen sowie dem Wissen über kommunikative Situationen interagiert). Das Sachwissen (als Wissen über den Zusammenhang von Entitäten) wird dagegen seit einigen Jahren vor allem in Form von Ontologien modelliert – sowohl der Umfang als auch die sprachliche Kodierung (in den meisten Ontologien Englisch) verhindert zum jetzigen Zeitpunkt allerdings eine befriedigende Verwendung im Kontext der (maschinellen) Themenanalyse deutschsprachiger Texte.

Im Hinblick auf die Frage nach der Repräsentation der Wissensbestände (sowie der Frage nach dem Zusammenwirken von Wissen und Text im Kommunikationsprozess) wurde in der vorliegenden Arbeit auf den Frame-Ansatz von Minsky



(1980) und Konearding (1993) rekurriert, in denen Frames als Listen von Fragen zu stereotypischen Situation aufgefasst werden, wobei diese Fragen den Status von Themen haben, an denen sich der Textproduzent bei der Textproduktion orientiert und die er für die Kommunizierung eines Sachverhaltskomplexes nutzt – während umgekehrt Rezipienten über das sprachliche Material einen Bezug zum thematisierten Frame herstellen können (Kapitel 2.7.2 und 2.7.3).

Grundlegend für die Modellierung der zweiten (Input-)Komponente „Text“ sind drei texttheoretische Überlegungen, die auf semiotischen, handlungstheoretischen und textstrukturalistischen Auffassungen basieren (Kapitel 2.6):

1. Texte sind komplexe indexikalische Zeichen (im Sinne von Antos 2009, 408f.), die zur Bezugnahme auf eine stereotypische kommunikative Situation (bzw. allgemeiner: ein Schema, eine Wissensstruktur) verwendet werden können bzw. als Indizien für diese Situation interpretiert werden können (Kapitel 2.6.2). Indexikalische Zeichen basieren dabei auf dem Prinzip des *pars pro toto*, d.h. durch eine sprachliche Einheit der Textoberfläche (z.B. ein Wort) oder ein Merkmal dieser Einheit (z.B. eine bestimmte Tempusform, aber auch ein bestimmtes Textstrukturmuster) wird das Schema bzw. die kommunikative Situation als Ganzes evoziert, wobei die Relation zwischen Ausdrucksmuster und Schema vom Produzenten/Rezipienten auf Basis seiner Erfahrungen, und speziell seines Sprachgebrauchswissens, hergestellt werden muss und nicht *per se* existiert. Voraussetzung für die Interpretation einfacher und komplexer indexikalischer Zeichen ist demnach die Regelmäßigkeit und Konventionalisiertheit der Verwendung eines Oberflächenmerkmals (bzw. einer Menge von Merkmalen) zur Bezugnahme auf einen bestimmten Typus einer kommunikativen Situation, d.h. die potentielle Verfügbarkeit dieser Beziehung im Sprachwissen (bzw. Sprachgebrauchswissens). In der vorliegenden Arbeit wird davon ausgegangen, dass Reflexionen des Themas im Text als indexikalische Zeichen für einen bestimmten Sachverhaltskomplex (als Thema) interpretiert werden können (Kapitel 2.7.2 und 3.2.1).
2. Texte sind komplex strukturierte Äußerungseinheiten, mit denen eine sprachliche Texthandlung (als komplexe sprachliche Handlung, die sich aus sprachlichen Teilhandlungen zusammensetzt) vollzogen wird (Kapitel 2.6.1). Zwischen sprachlicher Handlung und Äußerung sowie zwischen komplexer sprachlicher Handlung und Teilhandlung bestehen dabei „indem“-Zusammenhänge: Eine komplexe Handlung wird vollzogen, indem eine Folge von Teilhandlungen vollzogen wird, indem eine Folge von Äußerungseinheiten realisiert wird. Thema und Handlung stehen dabei insofern in einem Zusammenhang, als das Thema als Gegenstand aufgefasst wird, auf den mit einer Handlung Bezug genommen wird. Eine Analyse der Zusammenhänge zwischen Themen muss dementsprechend auch Zusammenhänge zwischen den sprachlichen Handlungen, die auf einen Gegenstand (ein Thema bzw. einen Aspekt des Themas) Bezug nehmen, berücksichtigen (Kapitel 3.3.2.1) und nicht nur Zusammenhänge zwischen den Diskursentitäten selbst.

3. Texte können aus zwei Hauptperspektiven beschrieben werden: der Produkt- und der Gebrauchsperspektive. Texte sind immer in einen bestimmten Gebrauchskontext eingebettet – als Teil einer solchen Gebrauchssituation können Texte auf einer kommunikativen und einer kognitiven Ebene betrachtet werden. Die Produktperspektive verengt den Blick auf den Text als Produkt. Aus einer solchen textinternen Sicht heraus können Texte als auf mehreren lokalen und globalen Ebenen strukturierte komplexe Äußerungseinheiten aufgefasst werden. Für den Gegenstand (expositorische Texte) und den Zweck der vorliegenden Arbeit (Operationalisierung der Analyse des Themas) wurde ein Textbeschreibungsmodell vorgeschlagen, in dem vier lokale und zwei globale, in vielfältiger Weise miteinander interagierende Textstrukturebenen differenziert wurden (Kapitel 2.6.3): Auf der lokalen Textstrukturebene lassen sich Texte in Hinsicht auf ihre Informationsstruktur, lexikalische, referenzielle und relationale Struktur beschreiben, auf der globalen Ebene in Hinsicht auf die generische Texttypstruktur sowie die visuelle und logische Textstruktur (siehe Abbildung 27). Die in Kapitel 2 diskutierten Thema-Konzepte lassen sich jeweils einer spezifischen Ebene zuordnen (Subklassifikation der Ebenen des Themas, siehe Kapitel 3.1.2, Tabelle 5). Das Thema interagiert demnach in spezifischer Weise mit jeder Textstrukturebene und leistet einen Beitrag zu der durch die jeweilige Strukturierungsebene konstituierten Kohärenz. Die sprachlichen und visuellen Mittel



Abbildung 27: Ebenen des Textes – Ebenen des Themas

auf den lokalen und globalen Textstrukturebenen haben dabei den Status kommunikativer Mittel oder – im Hinblick auf das Thema – indexikalischer Zeichen für das Thema (siehe Kapitel 3.3).

Diese drei texttheoretischen Überlegungen stellen grundlegende Eckpunkte für die Analyse des Themas auf Textebene dar. Texte – und speziell deren thematische Organisation – werden als sowohl intentionale als auch emergente Gebilde betrachtet, das Thema als intentionale und emergente Größe des Textes. Gegenstand der Themenanalyse müssen demnach sowohl die unter Umständen nicht-intentionalen als auch die intentionalen Reflexionen des Themas im Text sein. Zu letzteren zählen insbesondere die Reflexionen der intentionalen thematischen Textorganisation, d.h. die intentional gesteuerte Verknüpfung von Aspekten des Themas über die Verknüpfung der auf sie bezogenen sprachlichen Teilhandlungen. Insofern Texte als auf mehreren Ebenen strukturierte Äußerungseinheiten betrachtet werden, wurde in der vorliegenden Arbeit die Analyse des Themas durch Multi-Ebenen-Analysen realisiert, die durch XML-basierte Multi-Ebenen-Annotationen formal repräsentiert werden können. Ziel dieser Multi-Ebenen-Analysen ist, für jede der Textstrukturebenen die sowohl intentionalen als auch nicht-intentionalen Reflexionen des Themas im Text zu ermitteln. Die Merkmale des Textes, die in Kapitel 3.3 als Reflexionen des Themas beschrieben wurden, basieren dabei zum einen auf den Merkmalen, die dem Thema in den in Kapitel 2 dargestellten Thema-Konzeptionen zugeschrieben werden (und die deshalb als mehr oder weniger stark konventionalisierte Merkmale betrachtet werden können). Zum anderen wurde die Menge der Typen von Thema-Reflexionen dadurch erweitert, dass auf Basis sonstiger Forschungsarbeiten (z.B. zur Rolle des Titels im Text) sowie eigener Untersuchungen weitere Reflexionen des Themas im Text ermittelt werden konnten, die auf der visuellen bzw. logischen Textstruktur, der Texttypstruktur sowie der relationalen Struktur zu verorten sind. Die Texttypstruktur und die relationale Textstruktur im Sinne der RST wurden dabei als zentrale Textstrukturebenen im Hinblick auf die Analyse der Zusammenhänge zwischen den Aspekten des Themas aufgefasst (Kapitel 3.3.2).

Die Ebene der Kommunikation (Textproduktion und -rezeption) wurde in der vorliegenden Arbeit – wie bereits angesprochen – als Prozesskomponente bzw. Schnittstelle zwischen den Ebenen der Kognition und des Textes aufgefasst, da angenommen wird, dass im Prozess der Textproduktion und -rezeption Wissen und Text miteinander interagieren. Bei der Textproduktion<sup>245</sup> beeinflusst das Wissen darüber, was (im Hinblick auf die Sachverhaltsstruktur) zum Thema gehört, das, was im Text über das Thema gesagt wird. Umgekehrt kann die Textproduktion aber auch auf die Wissensbestände des Produzenten zurückwirken. Im Schreibprozess kann Wissen aktualisiert, erweitert oder in anderer Weise neu strukturiert werden (epistemische Funktion des Schreibens). Auch bei der Textrezeption interagieren Wissen und Text (Kapitel 3.4.2): Auf Basis des Textmate-

---

<sup>245</sup> Insbesondere bei der textorganisierenden Handlung des Thematisierens mit ihren Teilhandlungen des Selektierens, Strukturierens, Orientierens, In-Bezug-Setzens, Fokussierens/Gewichtens (Kapitel 3.4.1).

rials bzw. der Bottom-Up-Cues (insbesondere des Titels und der ersten Sätze als initialem Input) und ihres Wissens bzw. der Top-Down-Constraints bauen Rezipienten eine Erwartung im Hinblick auf das Textthema auf. Sie konstruieren einen Verstehenshorizont, der durch neuen Input korrigiert und modifiziert werden kann. Auf Grundlage dieses Erwartungs- und Verstehenshorizonts bezüglich des Themas wird schließlich sukzessive versucht, ein kohärentes mentales Modell des Textes (und damit auch des Themas als Sachverhaltskomplex) aufzubauen. Der Zusammenhang von Thema und Wissen ist dabei durch Erwartbarkeit und Adäquatheit geprägt. Was aufgrund gespeicherter kognitiver Schemata bzw. Frames thematisch erwartbar ist (z.B. der Zusammenhang von „Pizzabestellung“, „Pizzalieferung“ und „Bezahlung des Pizzaboten“), wird auch im Text als adäquater thematischer Zusammenhang bewertet. Das Wissen von Produzent und Rezipient sowie Annahmen über dieses Wissen stellen demnach eine Voraussetzung und gleichzeitig Beschränkung für das Produzieren und Rezipieren dar. Sowohl das Produzieren als auch das Verstehen von Texten ist demnach als komplexes Zusammenspiel von Textmaterial (Cues), Wissensstrukturen und sprachlichen und kommunikativen Prinzipien zu sehen (Constraints).

Das in der vorliegenden Arbeit präsentierte Analysemodell zielt darauf ab, die Themenanalyse zu operationalisieren und darüber hinaus mittels computerlinguistischer Verfahren und Ressourcen automatisierbar zu machen. Zu diesem Zweck wurden zunächst die drei Komponenten des Analysemodells theoretisch hergeleitet (s.o.). Auf Basis dieser theoretischen Herleitung wurde ein Prozessmodell zum Zusammenspiel von Kommunikation, Kognition und Text entwickelt (Kapitel 2.9, Abbildung 6), das im weiteren Verlauf der Arbeit im Hinblick auf das Produzieren und Verstehen des Themas bzw. der thematischen Organisation spezifiziert (Kapitel 3.4.3, Abbildung 25) und schließlich in ein prozessorientiertes Architekturmodell zur maschinellen Themenidentifikation überführt wurde (Kapitel 3.4.3, Abbildung 26). In diesem Architekturmodell müssen die Analyseergebnisse der einzelnen Ebenen miteinander sowie mit den Wissensressourcen „verrechnet“ werden. Ziel dieser „Verrechnung“ (im Sinne von Ortner und Sitta 2003, 23 – siehe auch Kapitel 4.3.3) ist zum einen die Identifizierung des Themas und seiner Aspekte, und in diesem Zusammenhang insbesondere die Differenzierung von wichtigen vs. weniger wichtigen sowie lokalen vs. globalen Aspekten des Themas, sowie zum anderen die Identifizierung der Relationen zwischen dem Thema und seinen Aspekten sowie zwischen den Aspekten untereinander. Hierzu müssen die einzelnen Reflexionen bzw. indexikalischen Zeichen in eine verständnisrelevante Ordnung gebracht werden und diejenigen Reflexionen bzw. indexikalischen Zeichen, die in einer vertikalen oder horizontalen Kontiguitätsbeziehung stehen, privilegiert behandelt werden, d.h. ein größeres Gewicht zugemessen bekommen. Diese Gewichtung geschieht unter Rekurs auf Wissen (bzw. dessen digital verfügbare Korrelate, d.h. Wissensressourcen bzw. Language Resources) sowie auf Basis statistischer Auswertungen. Dabei werden die indexikalischen Zeichen, für die signifikante Korrelationen bzw. Kontiguitätsbeziehungen zu indexikalischen Zeichen der gleichen oder einer anderen Textstrukturebene festgestellt werden können bzw. für die ein sinnvoller Bezug auf das Thema als Sachverhaltskomplex konstruiert werden kann, privilegiert behandelt. Für die Identifizierung der Relationen zwischen dem Thema und seinen

Aspekten sind dagegen vor allem die indexikalischen Zeichen relevant, zwischen denen horizontale Kontiguitätsbeziehungen konstruiert werden können. Im Hinblick auf die „Verrechnung“ der Reflexionen des Themas im Text untereinander sowie mit den Wissensbeständen ist anzunehmen, dass bestimmte Textstrukturebenen bzw. bestimmte Einheiten von Textstrukturebenen eine größere Bedeutung für die Identifizierung des Themas besitzen als andere. In der vorliegenden Arbeit wurde insbesondere auf die Bedeutung der ersten rezipierten Einheiten des Textes hingewiesen (insbesondere des Titels) sowie auf die Bedeutung der visuell hervorgehobenen Einheiten (z.B. Überschriften). Eine weitergehende Hierarchisierung der Reflexionen des Themas sollte allerdings auf Basis empirischer Korpusanalysen bzw. statistischer Korpusauswertungen erfolgen<sup>246</sup>, was in der vorliegenden Arbeit jedoch nicht geleistet werden konnte, da die Voraussetzungen zur Implementation des vorgestellten Analysemodells momentan noch nicht gegeben sind. Eine manuelle oder semi-automatische Analyse ist zwar – wie durch die exemplarischen Analysen zweier Beispieltex-te gezeigt werden konnte (Kapitel 3.3) – möglich, allerdings relativ aufwändig und für ein größeres Korpus, wie es für die Validierung und Weiterentwicklung des Analysemodells nötig wäre, nur in einem größeren Forschungsprojekt mit mehreren Mitarbeitern zu realisieren. In einem solchen Kontext müssten zunächst Richtlinien zur Analyse und Annotation der einzelnen Ebenen der Textstruktur entwickelt und die entstehenden Annotationen auf ihre Intra- und Interrater-Reliabilität geprüft werden.

Einer Implementation des Softwaresystems zur maschinellen Themenanalyse steht momentan insbesondere die mangelhafte Verfügbarkeit der Wissensressourcen sowie die teilweise unbefriedigende Präzision und Vollständigkeit der vorgestellten computerlinguistischen Verfahren zur maschinellen Analyse der Reflexionen des Themas auf den lokalen und globalen Textstrukturebenen im Weg. Da allerdings alle nötigen Analysekomponenten Gegenstand computerlinguistischer Forschungsarbeiten sind, sind diese Hindernisse nicht als grundsätzliche, sondern lediglich als temporäre Hindernisse aufzufassen, deren zumindest teilweise Überwindung mittel- oder langfristig zu erwarten ist.

Auf Basis des in der vorliegenden Arbeit entwickelten Thema- und Analysemodells ergibt sich eine Reihe von Anschlussmöglichkeiten für die weitere Forschung: Insbesondere im Hinblick auf das vorgestellte Analysemodell selbst sind verschiedene Weiterentwicklungen nötig. Zu nennen ist an dieser Stelle zunächst die Übertragung bzw. Anpassung des Modells an andere Texttypen. Die vorliegende Arbeit beschränkte sich auf die Entwicklung eines Analysemodells für expositorische Texte. Da in anderen Texttypen die konative/appelative, emotive/expressive oder poetische Funktion eine stärkere Rolle spielen können als in expositorischen Texten, sind bei der Analyse möglicherweise andere Ebenen und

---

<sup>246</sup> Hierzu könnten bereits existierende, auf mehreren Ebenen annotierte Korpora genutzt werden (z.B. das Potsdam Commentary Corpus oder das SemDok-Korpus wissenschaftlicher Zeitschriftenartikel). Allerdings decken diese Korpora nicht alle Ebenen und Einheiten des in der vorliegenden Arbeit vorgeschlagenen Analysemodells ab, so dass hier ergänzende Annotationen nötig wären.

Typen von Reflexionen des Themas im Text zu berücksichtigen. Neben der Frage nach der Möglichkeit der Übertragung des Analysemodells auf andere Texttypen blieb in der vorliegenden Arbeit auch offen, wie die verschiedenen indexikalischen Zeichen konkret miteinander „verrechnet“ werden können. Nötig wäre hier eine stärkere Formalisierung und Operationalisierung der als Black Box modellierten Prozesskomponente. In diesem Kontext müsste auch eine (empirisch validierte) Hierarchisierung der Reflexionen des Themas im Text vorgenommen werden. Dabei sollte geprüft werden, ob diese Hierarchien texttypspezifisch oder texttypunabhängig sind, d.h. ob die Bedeutung verschiedener Typen von Reflexionen des Themas sich je nach Texttyp ändert.

Daneben ergeben sich aus der vorliegenden Arbeit aber auch weitergehende Anknüpfungspunkte zu korpuslinguistischen und soziolinguistischen Forschungsarbeiten. Ein Beispiel hierfür ist die Forschung zu Sprachgebrauchsmustern. Insofern das Sprachgebrauchswissen (als Schnittstelle zwischen Ausdrucksebene, sprachlichem Handeln und kognitiver Ebene) eine zentrale Ressource für die Themenidentifikation darstellt, muss die formale Modellierung dieses Wissens (bzw. die Konstruktion eines digitalen Korrelats dieses Wissens) als wichtige Voraussetzung für die Implementation des dargestellten Softwaresystems betrachtet werden. Sprachgebrauchsmuster (als sichtbares Zeichen des Sprachgebrauchswissens) lassen sich über statistische Korpusauswertungen ermitteln. Dabei ist Ziel dieser Auswertungen, typische Ausdrucksmuster für bestimmte Typen kommunikativer Situationen (oder Themen als Sachverhaltskomplexen) zu identifizieren. Obwohl der Zusammenhang von Ausdrucksmuster und Schema dynamisch und in Abhängigkeit vom Kontext von den Kommunikanten selbst zu konstruieren ist, basiert er doch auf einer gewissen Regelmäßigkeit und Konventionalität. Die in jeweils bestimmten Kontexten bzw. Typen kommunikativer Situationen verwendeten typischen Ausdrucksformen könnten deshalb formal als Sprachgebrauchspräferenzen erfasst werden. Bubenhofer (2009, 319f.) schlägt in diesem Zusammenhang vor, auf Basis von Korpusanalysen eine „semantische Matrix“ zu erstellen, in der musterhafte, typische Sprechweisen Kategorien wie Thema, Textsorte etc. zugeordnet werden. Er geht dabei davon aus, dass die „musterhaften, typischen Sprechweisen semantische Felder strukturieren, also bestimmen, was auf welche Art und Weise wann und wo sagbar und verstehbar ist“ (Bubenhofer 2009, 320). Eine solche Matrix könnte darüber hinaus durch empirische Rezeptionsexperimente, die ähnlich wie die von Feilke (2003b) konstruiert sein könnten, validiert werden.

Weitere Forschungsfragen, die sich im Anschluss an die vorliegende Arbeit stellen, sind z.B. die nach der Abhängigkeit der im Text thematisierten Typen von Teilthemen von dem ontologischen Typus des Textthemas, und zwar unabhängig vom Texttyp. Auf Basis der Ansätze von Schank (1977b) und Adamzik (2004b) wurde zu dieser Frage eine eigene exemplarische Analysestudie durchgeführt, die die Annahme, dass Thementypkonstellationen abhängig sind vom Typus des Textthemas, aber unabhängig vom Texttyp – stützt. Aufgrund des geringen Umfangs der Studie bedarf diese Annahme allerdings der weiteren Validierung, z.B. durch größere Korpusstudien. Auch die Hypothese, dass sich die thematische Organisation von Texten mittels des eingeführten Inventars an RST-Relationen

(insbesondere der Subtypen von ELABORATION) beschreiben lässt, sollte auf Basis einer Evaluation der Intra- und Interrater-Reliabilität im Hinblick auf ihre Validität bzw. Praktikabilität geprüft werden. Voraussetzung hierfür ist die Entwicklung von Annotationsguidelines und die initiale Schulung der Annotatoren.

Die vorliegende Arbeit versteht sich auch als Beitrag zu einer besseren Verknüpfung von textlinguistischen und computerlinguistischen Ansätzen zur Textanalyse. Obwohl die manuelle/maschinelle Textanalyse sowohl für die Textlinguistik als auch die Computerlinguistik von großem Interesse ist, arbeiten beide Disziplinen mehr oder weniger autark voneinander. Ein Grund hierfür ist sicherlich darin zu sehen, dass textlinguistische Theorien und Modelle oft zu komplex sind, um mittels computerlinguistischer Verfahren und Ressourcen in konkrete Systeme umgesetzt werden zu können. Diese Arbeit versucht die Kluft zwischen beiden Bereichen zu überwinden, indem sie exemplarisch – am Beispiel des spezifischen Forschungsgegenstands Thema – aufzeigt, wie textlinguistische Theorien und Verfahren systematisiert und implementiert werden können.

Ausgangspunkt für eine solche Verknüpfung von Text- und Computerlinguistik muss dabei zunächst die gegenseitige Wahrnehmung sein. Nur wenn Textlinguisten sich damit vertraut machen, welche Methoden in der Computerlinguistik entwickelt werden und wie gut diese sind, und umgekehrt, nur wenn Computerlinguisten die Theorien und Modelle der Textlinguistik zur Kenntnis nehmen, sind die Voraussetzungen dafür geschaffen, dass beide Bereiche auch voneinander profitieren können. Die vorliegende Arbeit hat versucht, die Gemeinsamkeiten und Unterschiede sowie die Chancen einer Verknüpfung beider Disziplinen anhand des spezifischen Forschungsgegenstands Thema zu verdeutlichen. Da eine über die Grenzen des eigenen Fachgebiets hinausgehende Sichtweise dabei helfen kann, alte Phänomene neu zu entdecken oder aus einer neuen Perspektive zu betrachten, wird an dieser Stelle für eine stärkere gegenseitige Wahrnehmung plädiert. Für die Textlinguistik ergibt sich eine solche insbesondere im Rahmen der Korpuslinguistik, für die Computerlinguistik teilweise im Rahmen der Texttechnologie.

## 5 Literatur

### 5.1 Forschungsliteratur

- Adamzik, Kirsten (2004a): Sprache: Wege zum Verstehen. Tübingen, Basel: A. Francke.
- Adamzik, Kirsten (2004b): Textlinguistik. Eine einführende Darstellung. Tübingen: Max Niemeyer. (= Germanistische Arbeitshefte, 40).
- Agirre, Eneko und Philip Edmonds (2006): Introduction. In: Agirre, Eneko und Philip Edmonds (Hg.): Word Sense Disambiguation. Algorithms and Applications. Dordrecht: Springer, S. 1-28.
- Agricola, Erhard (1976): Vom Text zum Thema. In: Danes, Frantisek und Dieter Viehweger (Hg.): Probleme der Textgrammatik. Berlin: Akademie Verlag, S. 13-28. (= studia grammatica XI).
- Agricola, Erhard (1979): Textstruktur, Textanalyse, Informationskern. Leipzig: VEB Verlag Enzyklopädie.
- Aiello, Marco, Christof Monz, Leon Todoran und Marcel Worring (2002): Document understanding for a broad class of documents. In: International Journal on Document Analysis and Recognition 5 (1), S. 1-16.
- Allan, James, Jaime Carbonell, George Doddington, Jonathan Yamron und Yiming Yang (1998): Topic detection and tracking pilot study. Final report. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop. S. 194-218.
- Altmann, Hans (1993): Fokus-Hintergrund-Gliederung und Satzmodus. In: Reis, Marga (Hg.): Wortstellung und Informationsstruktur. Tübingen: Max Niemeyer, S. 1-38. (= Linguistische Arbeiten, 306).
- Andrews, Pierre und Martin Rajman (2004): Thematic Annotation: extracting concepts out of documents. Technical Report IC/2004/68. Lausanne: Swiss Federal Institute of Technology.
- Antos, Gerd (1989): Textproduktion: Ein einleitender Überblick. In: Antos, Gerd und Hans P. Krings (Hg.): Textproduktion: ein interdisziplinärer Forschungsüberblick. Tübingen: Max Niemeyer, S. 5-57. (= Konzepte der Sprach- und Literaturwissenschaft, 48).
- Antos, Gerd (1997): Texte als Konstitutionsformen von Wissen. In: Antos, Gerd und Heike Tietz (Hg.): Die Zukunft der Textlinguistik. Traditionen, Transformationen, Trends. Tübingen: Max Niemeyer, S. 43-63. (= Germanistische Linguistik, 188).
- Antos, Gerd (2000): Ansätze zur Erforschung der Textproduktion. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 105-112. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Antos, Gerd (2009): Semiotik der Text-Performanz. Symptome und Indizien als Mittel der Bedeutungskonstitution. In: Linke, Angelika und Helmuth Feilke (Hg.): Oberfläche und Performanz. Untersuchungen zur Sprache als dynamischer Gestalt. Tübingen: Max Niemeyer, S. 407-427.
- Aone, Ghinatsu und Scott William Bennett (1995): Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies. In: Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL). 26-30 June 1995, MIT, Cambridge, Massachusetts, USA, S. 122-129.
- Ariel, Mira (2001): Accessibility theory: an overview. In: Sanders, Ted, Joost Schilperoord und Wilbert Spooren (Hg.): Text Representation. Linguistic and psycholinguistic aspects. Amsterdam, Philadelphia: John Benjamins, S. 29-87.
- Asher, Nicholas (2004a): Discourse Topic. In: Theoretical Linguistics 30 (2-3), S. 163-201.



- Asher, Nicholas (2004b): Troubles with topics: Comments on Kehler, Oberlander, Stede and Zeevat. In: *Theoretical Linguistics* 30 (2-3), S. 255-262.
- Asher, Nicholas und Alex Lascarides (2003): *Logics of Conversation*. Cambridge, UK: Cambridge University Press.
- Auer, Peter (1986): Kontextualisierung. In: *Studium Linguistik* 19, S. 22-47.
- Bärenfänger, Maja (2008): RST Relationsdefinitionen in SemDok. Technischer Report. Gießen: Justus-Liebig-Universität Gießen, Institut für Germanistik, Fachgebiet Angewandte Sprachwissenschaft und Computerlinguistik. (= Interne Reports der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“). [Stand: 22. April 2008].
- Bärenfänger, Maja, Daniela Goecke, Mirco Hilbert, Harald Längen und Maik Stührenberg (2008): Anaphora as an Indicator of Elaboration: A Corpus Study. In: *Journal for Language Technology and Computational Linguistics (JLCL)* 23 (2), S. 49-72.
- Bärenfänger, Maja, Mirco Hilbert, Henning Lobin und Harald Längen (2008): OWL ontologies as a resource for discourse parsing. In: *LDV-Forum. GLDV-Journal for Computational Linguistics and Language Technology* 23 (1), S. 17-26.
- Bärenfänger, Maja, Mirco Hilbert, Henning Lobin, Harald Längen und Csilla Puskas (2006): Cues and constraints for the relational discourse analysis of complex text types - the role of logical and generic document structure. In: *Proceedings of the 2nd conference on Constraints in Discourse (CiD 2006)*. Maynooth, Ireland.
- Bärenfänger, Maja, Harald Längen, Mirco Hilbert und Henning Lobin (2010): The role of logical and generic document structure in relational discourse analysis. In: Kühnlein, Peter, Anton Benz und Candace L. Sidner (Hg.): *Constraints in discourse 2*. Amsterdam: John Benjamins, S. 81-104. (= *Pragmatics & Beyond New Series*, 194).
- Barzilay, Regina und Michael Elhadad (1999): Using lexical chains for text summarization. In: Mani, I. und M. T. Maybury (Hg.): *Advances in Automatic Text Summarization*. Cambridge, MA: MIT Press, S. 111-121.
- Bateman, John, Jörg Klein, Thomas Kamps und Klaus Reichenberger (2001): Towards Constructive Text, Diagram, and Layout Generation for Information Presentation. In: *Computational Linguistics* 27 (3), S. 409-449.
- Baumann, Stefan, Caren Brinckmann, Silvia Hansen-Schirra, Geert-Jan Kruijff, Ivana Kruijff-Korbayova, Stella Neumann, Erich Steiner, Elke Teich und Hans Uszkoreit (2004): The MULI Project: Annotation and Analysis of Information Structure in German and English. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*. Lissabon, Portugal.
- Bayer, Josef (1980): Diskursthemen. In: Weigand, Edda und Gerhard Tschauder (Hg.): *Perspektive: textintern. Akten des 14. Linguistischen Kolloquiums*. Bochum 1979. Bd. 1. Tübingen: Max Niemeyer, S. 213-223. (= *Linguistische Arbeiten*, 88).
- Bayerl, Petra Saskia, Daniela Goecke, Harald Längen und Andreas Witt (2003): Methods for the semantic analysis of document markup. In: *Proceedings of the ACM-Symposium on Document Engineering (DocEng 2003)*. Grenoble, France.
- Becker, Markus und Anette Frank (2002): A stochastic topological parser for German. In: *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002) - Vol. 1*. Taipei, Taiwan, S. 1-7.
- Beneš, Eduard und Josef Vachek (1971): Einleitung: Die Prager Schule - Erbe und Gegenwart. In: Beneš, Eduard und Josef Vachek (Hg.): *Stilistik und Soziolinguistik. Beiträge der Prager Schule zur strukturellen Sprachbetrachtung und Spracherziehung*. 2. Aufl. Berlin: Paul List Verlag, S. IV-XXVI.
- Berry, Margaret (1995): Thematic options and success in writing. In: Ghadessy, Mohsen (Hg.): *Thematic Development in English Texts*. London: Pinter, S. 55-84.
- Bieler, Heike und Stefanie Dipper (2008): Measures for Term and Sentence Relevance: an Evaluation for German. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC-2008)*. Marrakech, Morocco.

- Bieler, Heike, Stefanie Dipper und Manfred Stede (2007): Identifying formal and functional zones in film reviews. In: Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue. Antwerpen, Belgien, S. 75-78.
- Blakemore, Diana (2001): Discourse and Relevance Theory. In: Schiffrin, Deborah, Deborah Tannen und Heidi E. Hamilton (Hg.): The Handbook of Discourse Analysis. Malden, Massachusetts: Blackwell Publishers, S. 100-118.
- Boguraev, Branimir K. und Mary S. Neff (2000a): Discourse Segmentation in Aid of Document Summarization. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS-33). Maui, Hawaii.
- Boguraev, Branimir K. und Mary S. Neff (2000b): Lexical Cohesion, Discourse Segmentation and Document Summarization. In: Association for Computational Linguistics: Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000) - Vol. 1. Saarbrücken, Germany, S. 76-82.
- Brandt, Margareta (1996): Subordination und Parenthese als Mittel der Informationsstrukturierung in Texten. In: Motsch, Wolfgang (Hg.): Ebenen der Textstruktur: sprachliche und kommunikative Prinzipien. Tübingen: Max Niemeyer, S. 211-240. (= Germanistische Linguistik, 164).
- Brandt, Margareta und Inger Rosengren (1992): Zur Illokutionsstruktur von Texten. In: Zeitschrift für Literaturwissenschaft und Linguistik 22 (86), S. 9-51.
- Brinker, Klaus (1980): Textthematik als spezifisch textlinguistischer Forschungsbereich. In: Kühlwein, Wolfgang und Albert Raasch (Hg.): Sprache und Verstehen II. Kongressberichte der 10. Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL). Tübingen: Gunter Narr, S. 138-141.
- Brinker, Klaus (1997): Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden. Berlin: Erich Schmidt Verlag.
- Brinker, Klaus (2000): Textstrukturanalyse. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 164-175. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Brinker, Klaus und Jörg Hagemann (2001): Themenstruktur und Themenentfaltung in Gesprächen. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 2. Halbband. Berlin, New York: de Gruyter, S. 1252-1263. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.2).
- Brown, Gillian und George Yule (1983): Discourse Analysis. Cambridge: Cambridge University Press. (= Cambridge textbooks in linguistics).
- Bubenhofner, Noah (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin: Walter de Gruyter. (= Sprache und Wissen, 4).
- Bühler, Karl (1978 [1965]): Sprachtheorie. Die Darstellungsfunktion der Sprache. Frankfurt a. M., Berlin, Wien: Ullstein.
- Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó und Manfred Pinkal (2009): FrameNet for the semantic analysis of German: Annotation, representation and automation. In: Boas, Hans C. (Hg.): Multilingual FrameNets in Computational Lexicography. Methods and Applications. Berlin, New York: de Gruyter, S. 209-244.
- Büring, Daniel (1997): The Meaning of Topic and Focus. The 59th Street Bridge Accent. London, New York: Routledge.
- Bußmann, Hadumod (2002): Lexikon der Sprachwissenschaft. Stuttgart: Alfred Kröner Verlag.
- Carlson, Greg N. (1984): Thematic roles and their role in semantic interpretation. In: Linguistics 22 (3), S. 259-279.

- Carlson, Lynn und Daniel Marcu (2001): Discourse Tagging Reference Manual. ISI Tech Report ISI-TR-545. URL: <http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf> (zuletzt abgerufen am 12.09.2010).
- Carstensen, Kai-Uwe (2010): Nicht-sprachliches Wissen. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hg.): Computerlinguistik und Sprachtechnologie. 3., überarb. u. erw. Aufl. Heidelberg: Spektrum Akademischer Verlag, S. 532-543.
- Chafe, Wallace (1976): Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In: Li, Charles N. (Hg.): Subject and Topic. New York: Academic Press, S. 25-56.
- Chafe, Wallace (1987): Cognitive constraints on information flow. In: Tomlin, Russell S. (Hg.): Coherence and grounding in discourse. Amsterdam: John Benjamins, S. 21-51.
- Chafe, Wallace (1998): Language and the Flow of Thought. In: Tomasello, Michael (Hg.): The New Psychology of Language. Cognitive and Functional Approaches to Language Structure. Mahwah, New Jersey: Lawrence Erlbaum, S. 93-111.
- Chenoweth, N. Ann und John R. Hayes (2001): Fluency in Writing. Generating Text in L1 and L2. In: Written Communication 18 (1), S. 80-98.
- Cheung, Jackie Chi Kit und Gerald Penn (2009): Topological field parsing of German. In: Association for Computational Linguistics: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Conference on Natural Language Processing of the AFNLP - Vol. 1. Suntec, Singapore, S. 64-72.
- Chiarcos, Christian, Stefanie Dipper, Michael Götze, Ulf Leser, Anke Lüdeling, Julia Ritz und Manfred Stede (2008): A Flexible Framework for Integrating Annotations from Different Tools and Tagsets. In: Traitement Automatique des Langues 49 (2), S. 271-293.
- Chomsky, Noam (1965): Aspects of the Theory of Syntax. Cambridge, Massachusetts: The M.I.T. Press.
- Christmann, Ursula (2000): Aspekte der Textverarbeitungsforschung. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 113-122. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Cieri, Christopher (2000): Multiple Annotations of Reusable Data Resources: Corpora for Topic Detection and Tracking. In: Proceedings of the 5es Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 2000). Lyon, France.
- Clark, Herbert H. (1977): Bridging. In: Johnson-Laird, Philip N. und P. Wason (Hg.): Thinking: Readings in Cognitive Science. Cambridge: Cambridge University Press, S. 411-420.
- Clifton, Chris, Robert Cooley und Rennie Jason (2004): TopCat: Data Mining for Topic Identification in a Text Corpus. In: IEEE Transactions on Knowledge and Data Engineering 16 (8), S. 949-964.
- Coseriu, Eugenio (1980): Textlinguistik. Eine Einführung. Tübingen: Gunter Narr. (= Tübinger Beiträge zur Linguistik, 109).
- Coursey, Kino und Rada Mihalcea (2009): Topic Identification Using Wikipedia Graph Centrality. In: Proceedings of the Human Language Technology Conference and Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2009). Boulder, Colorado, S. 117-120.
- Coursey, Kino, Rada Mihalcea und William Moen (2009): Using encyclopedic knowledge for automatic topic identification. In: Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL 2009). Boulder, Colorado, S. 210-218.

- Cramer, Irene, Marc Finthammer, Alexander Kurek, Lukas Sowa, Melina Wachtling und Tobias Claas (2006): Experiments on Lexical Chaining for German Corpora: Annotation, Extraction, and Application. In: LDV-Forum 21 (1), S. 1-14.
- Cramer, Irene, Tonio Wandmacher und Ulli Waltinger (2011): Exploring Resources for Lexical Chaining: A Comparison of Automated Semantic Relatedness Measures and Human Judgements. In: Mehler, Alexander, Kai-Uwe Kühnberger, Henning Lobin, Harald Lüngen, Angelika Storrer und Andreas Witt (Hg.): Modeling, Learning and Processing of Text Technological Data Structures. Dordrecht: Springer.
- Cristea, Dan, Nancy Ide, Daniel Marcu und Valentin Tablan (1999): Discourse Structure and Co-Reference: An Empirical Study. In: Proceedings of the ACL'99 Workshop on the Relationship Between Discourse/Dialogue Structure and Reference. University of Maryland, USA.
- Cristea, Dan, Oana Postolache und Ionut Pistol (2005): Summarisation Through Discourse Structure. In: Gelbukh, Alexander (Hg.): Computational Linguistics and Intelligent Text Processing. 6th International Conference, CICLing 2005, Mexico City, Mexico, February 13-19, 2005, Proceedings. Berlin, Heidelberg: Springer, S. 632-644. (= Lecture Notes in Computer Science, 3406).
- Cruse, D. A. (1986): Lexical Semantics. Cambridge: Cambridge University Press. (= Cambridge textbooks in linguistics).
- Culotta, Aron, Michael Wick, Robert Hall und Andrew McCallum (2007): First-order probabilistic models for coreference resolution. In: Proceedings of the Human Language Technology Conference and Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2007). Rochester, NY, USA.
- Dahl, Trine (2000): Lexical Cohesion-Based Text Condensation. An Evaluation of Automatically Produced Summaries of Research Articles by Comparison with Author-Written Abstracts. PhD Thesis. Bergen: University of Bergen.
- Daneš, František (1970): Zur linguistischen Analyse der Textstruktur. In: Folia Linguistica 4, S. 72-78.
- Daneš, František (1976): Zur semantischen und thematischen Struktur des Kommunikats. In: Daneš, František und Dieter Viehweger (Hg.): Probleme der Textgrammatik. Berlin: Akademie Verlag, S. 29-40. (= studia grammatica XI).
- de Beaugrande, Robert und Wolfgang Ulrich Dressler (1981): Einführung in die Textlinguistik. Tübingen: Max Niemeyer.
- Deerwester, Scott C., Susan T. Dumais, Thomas K. Landauer, George W. Furnas und Richard A. Harshman (1990): Indexing by Latent Semantic Analysis. In: Journal of the American Society of Information Science 41 (6), S. 391-407.
- Dik, Simon C. (1997): The Theory of Functional Grammar. Edited by Kees Hengeveld. Part 2: Complex and Derived Constructions. Berlin, New York: de Gruyter. (= Functional Grammar Series, 21).
- Dipper, Stefanie (2005): XML-based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation. In: Proceedings of the Berliner XML Tage 2005 (BXML 2005). Berlin, Germany, S. 39-50.
- Dipper, Stefanie, Michael Götze, Uwe Küssner und Manfred Stede (2007): Representing and Querying Standoff XML. In: Rehm, Georg, Andreas Witt und Lothar Lemnitzer (Hg.): Data structures for linguistic resources and applications. Proceedings of the Biennial GLDV Conference 2007. Tübingen: Gunter Narr, S. 337-346.
- Dipper, Stefanie, Michael Götze und Stavros Skopeteas (2007): Information Structure in Cross-Linguistic Corpora. Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure. Technical report. Universität Potsdam. (= ISIS. Working Papers of the SFB 632).
- Dörre, Jochen, Peter Gerstl und Roland Seiffert (2001): Volltextsuche und Text Mining. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Endriss, Susanne Jekat, Ralf

- Klabunde und Hagen Langer (Hg.): Computerlinguistik und Sprachtechnologie. Eine Einführung. 1. Aufl. Heidelberg: Spektrum Akademischer Verlag, S. 425-441.
- Dressler, Wolfgang (1973): Einführung in die Textlinguistik. Tübingen: Max Niemeyer.
- Duden - Deutsches Universalwörterbuch. Elektronische Ausgabe. München, Berlin: Langenscheidt o.J. URL: <http://services.langenscheidt.de/hebis/> (zuletzt abgerufen am 31.03.2009).
- Dürr, Michael und Peter Schlobinski (2006): Deskriptive Linguistik. Grundlagen und Methoden. Göttingen: Vandenhoeck & Ruprecht. (= Studienbücher zur Linguistik, 11).
- Ebert, Christian, Michael Schiehlen, Ralf Klabunde und Stefan Evert (2010): Semantik. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hg.): Computerlinguistik und Sprachtechnologie. 3., überarb. u. erw. Aufl. Heidelberg: Spektrum Akademischer Verlag, S. 330-393.
- Edmundson, H. P. (1969): New methods in automatic extracting. In: Journal of the Association for Computing Machinery 16 (2), S. 264-285.
- Eggs, Ekkehard (2000): Vertexungsmuster Argumentation: Logische Grundlagen. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 397-414. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Eigler, Gunther, Thomas Jechle, Gabriele Merziger und Alexander Winter (1990): Wissen und Textproduzieren. Tübingen: Gunter Narr.
- Elhadad, Noemie und Kathleen R. McKeown (2001): Towards generating patient specific summaries of medical articles. In: Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2001). Pittsburgh, Pennsylvania, USA, S. 32-40.
- Ercan, Gonenc und Ilyas Cicekli (2008): Lexical Cohesion Based Topic Modeling for Summarization. In: Gelbukh, Alexander (Hg.): Computational Linguistics and Intelligent Text Processing, 9th International Conference, CICLing 2008, Haifa, Israel, February 17-23, 2008, Proceedings. Berlin, Heidelberg: Springer, S. 582-592. (= Lecture notes in computer science, 4919).
- Erk, Katrin und Sebastian Padó (2006): SHALMANESER – A Toolchain For Shallow Semantic Parsing. In: Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006). Genoa, Italy.
- Eroms, Hans-Werner (2000): Der Beitrag der Prager Schule zur Textlinguistik. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 36-43. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Farzindar, Atefeh und Guy Lapalme (2004): Legal Text Summarization by Exploration of the Thematic structures and Argumentative Roles. In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004). Barcelona, Spain, S. 27-34.
- Feilke, Helmuth (1993): Sprachlicher Common sense und Kommunikation: Über den „gesunden Menschenverstand“, die Prägung der Kompetenz und idiomatische Ordnung des Verstehens. In: Der Deutschunterricht 6, S. 6-21.
- Feilke, Helmuth (1996): Die Entwicklung der Schreibfähigkeiten. In: Günther, Hartmut und Otto Ludwig (Hg.): Schrift und Schriftlichkeit: ein interdisziplinäres Handbuch internationaler Forschung. 2. Halbband. Berlin, New York: de Gruyter, S. 1178-1191. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 10).
- Feilke, Helmuth (2003a): Kontext - Zeichen - Kompetenz. Wortverbindungen unter sprachtheoretischem Aspekt. In: Steyer, Kathrin (Hg.): Wortverbindungen - mehr

- oder weniger fest. Jahrbuch 2003 des Instituts für deutsche Sprache. Berlin, New York: de Gruyter, S. 41-64.
- Feilke, Helmuth (2003b): Textroutine, Textsemantik und sprachliches Wissen. In: Linke, Angelika, Hanspeter Ortner und Paul R. Portmann-Tselikas (Hg.): Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis. Tübingen: Max Niemeyer, S. 209-229.
- Feilke, Helmuth (2010): "Aller guten Dinge sind drei" – Überlegungen zu Textroutinen & literalen Prozeduren. In: Bons, Iris, Thomas Gloning und Dennis Kaltwasser (Hg.): Fest-Platte für Gerd Fritz. Gießen, 17.05.2010.
- Feilke, Helmuth und Gerhard Augst (1989): Zur Ontogenese der Schreibkompetenz. In: Antos, Gerd und Hans P. Krings (Hg.): Textproduktion: ein interdisziplinärer Forschungsüberblick. Tübingen: Max Niemeyer, S. 297-327. (= Konzepte der Sprach- und Literaturwissenschaft, 48).
- Feilke, Helmuth und Angelika Linke (2009): Oberfläche und Performanz - zur Einleitung. In: Linke, Angelika und Helmuth Feilke (Hg.): Oberfläche und Performanz. Untersuchungen zur Sprache als dynamischer Gestalt. Tübingen: Max Niemeyer, S. 3-17.
- Feilke, Helmuth und Siegfried J. Schmidt (1995): Denken und Sprechen. Anmerkungen zur strukturellen Kopplung von Kognition und Kommunikation. In: Trabant, Jürgen (Hg.): Sprache denken. Positionen aktueller Sprachphilosophie. Frankfurt a. M.: Fischer Taschenbuch Verlag, S. 269-297.
- Ferret, Olivier (2002): Using collocations for topic segmentation and link detection. In: Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002). Taipei, Taiwan.
- Ferret, Olivier und Brigitte Grau (2002a): A bootstrapping approach for robust topic analysis. In: Natural Language Engineering 8 (2-3), S. 209-233.
- Ferret, Olivier und Brigitte Grau (2002b): A Topic Segmentation of Texts based on Semantic Domains. In: Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002). Taipei, Taiwan, S. 1-7.
- Ferret, Olivier, Brigitte Grau und Nicolas Masson (1998): Thematic segmentation of texts: two methods for two kinds of texts. In: Proceedings of the 17th International Conference on Computational Linguistics (COLING 1998) - Vol. 1. Montreal, Quebec, Canada, S. 392 - 396.
- Fillmore, Charles J. (2006 [1982]): Frame Semantics. In: Geeraerts, Dirk (Hg.): Cognitive Linguistics: Basic Readings. Berlin, New York: de Gruyter, S. 373-400. (= Cognitive Linguistics Research, 34).
- Fillmore, Charles J. und Collin F. Baker (2001): Frame Semantics for Text Understanding. In: Proceedings of the NAACL Workshop on WordNet and Other Lexical Resources. Pittsburgh, USA.
- Firbas, Jan (1992): Functional sentence perspective in written and spoken communication. Cambridge: Cambridge University Press.
- Firbas, Jan (1995): A contribution on a panel discussion on Rheme. In: Ghadessy, Mohsen (Hg.): Thematic Development in English Texts. London: Pinter, S. 213-222.
- Fries, Peter H. (1995): A personal view of Theme. In: Ghadessy, Mohsen (Hg.): Thematic Development in English Texts. London: Pinter, S. 1-19.
- Fritz, Gerd (1982): Kohärenz. Grundfragen der linguistischen Kommunikationsanalyse. Tübingen: Gunter Narr. (= Tübinger Beiträge zur Linguistik, 164).
- Fritz, Gerd (1994): Grundlagen der Dialogorganisation. In: Fritz, Gerd und Franz Hundsnurscher (Hg.): Handbuch der Dialoganalyse. Tübingen: Max Niemeyer, S. 177-201.
- Gabrilovich, Evgeniy und Shaul Markovitch (2007): Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007). Hyderabad, India, S. 1606-1611.

- Gansel, Christina und Frank Jürgens (2002): Textlinguistik und Textgrammatik. Eine Einführung. Wiesbaden: Westdeutscher Verlag. (= Studienbücher zur Linguistik, 6).
- Garrod, Simon (1995): Distinguishing between explicit and implicit focus during text comprehension. In: Rickheit, Gert und Christopher Habel (Hg.): Focus and Coherence in Discourse Processing. Berlin, New York: de Gruyter, S. 3-17. (= Research in Text Theory, 22).
- Ge, Niye, John Hale und Eugene Charniak (1998): A statistical approach to anaphora resolution. In: Proceedings of the 6th ACL Workshop on Very Large Corpora (WVLC-6). Montreal, Quebec, Canada, S. 161-171.
- Ghadessy, Mohsen (1995): Thematic development and its relationship to registers and genres. In: Ghadessy, Mohsen (Hg.): Thematic Development in English Texts. London: Pinter, S. 129-146.
- Giora, Rachel (1983): Functional Paragraph Perspective. In: Petöfi, János S. und Emel Sözer (Hg.): Micro and Macro Connexity of Texts. Hamburg: Helmut Buske Verlag, S. 153-182. (= Papiere zur Textlinguistik, 45).
- Givón, Talmy (1983a): Topic Continuity in Discourse: An Introduction. In: Givón, T. (Hg.): Topic Continuity in Discourse: A Quantitative Cross-Language Study. Amsterdam, Philadelphia: John Benjamins, S. 5-41. (= Typological Studies in Language; 3).
- Givón, Talmy (1983b): Topic Continuity in Spoken English. In: Givón, T. (Hg.): Topic Continuity in Discourse: A Quantitative Cross-Language Study. Amsterdam, Philadelphia: John Benjamins, S. 347-363. (= Typological Studies in Language; 3).
- Givón, Talmy (1992): The grammar of referential coherence as mental processing instructions. In: Linguistics 30, S. 5-55.
- Givón, Talmy (1993): English Grammar. A Function-Based Introduction. Vol. II. Amsterdam, Philadelphia: John Benjamins.
- Givón, Talmy (1995): Coherence in text vs. coherence in mind. In: Gernsbacher, M.A. und T. Givón (Hg.): Coherence in Spontaneous Text. Amsterdam, Philadelphia: John Benjamins, S. 59-115. (= Typological Studies in Language; 31).
- Glanzberg, Michael (2002): Context and Discourse. In: Mind and Language 17 (4), S. 333-375.
- Gloning, Thomas (1994a): Dialoganalyse und Semantik. In: Fritz, Gerd und Franz Hundsnurscher (Hg.): Handbuch der Dialoganalyse. Tübingen: Max Niemeyer, S. 259-280.
- Gloning, Thomas (1994b): Praktische Semantik und Linguistische Kommunikationsanalyse. In: Fritz, Gerd und Franz Hundsnurscher (Hg.): Handbuch der Dialoganalyse. Tübingen: Max Niemeyer, S. 113-129.
- Goecke, Daniela, Harald Lüngen, Dieter Metzger, Maik Stührenberg und Andreas Witt (2010): Different Views on Markup. Distinguishing Levels and Layers. In: Witt, Andreas und Dieter Metzger (Hg.): Linguistic Modeling of Information and Markup Languages. Contributions to Language Technology. Dordrecht: Springer, S. 1-21. (= Text, Speech and Language Technology, 40).
- Goecke, Daniela, Anke Holler und Maik Stührenberg (2008): Coreference, Cospecification and Bridging: Annotation Scheme. Technical Report. Bielefeld: Universität Bielefeld. (= Internal Report of the DFG-Research Group 437 "Texttechnological Modelling of Information"). [Stand: 27. August 2008]. URL: <http://coli.lili.uni-bielefeld.de/cgi-bin/publications.pl> (zuletzt abgerufen am 23.09.2010).
- Goecke, Daniela, Maik Stührenberg und Tonio Wandmacher (2008): A hybrid approach to resolve nominal anaphora. In: LDV-Forum 32 (1), S. 43-58.
- Goecke, Daniela und Andreas Witt (2006): Exploiting logical document structure for anaphora resolution. In: Proceedings of the Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006). Genoa, Italy.

- Gómez-González, María Ángeles (2001): The theme-topic interface: Evidence from English. Amsterdam, Philadelphia: John Benjamins. (= Pragmatics and Beyond, 71).
- Götze, Michael und Stefanie Dipper (2006): ANNIS: Complex Multilevel Annotations in a Linguistic Database. In: Proceedings of the 5th EACL Workshop on NLP and XML: Multi-dimensional Markup in Natural Language Processing (NLPXML-2006). Trento, Italy, S. 61-64.
- Graefen, Gabriele (1997): Der wissenschaftliche Artikel - Textart und Textorganisation. Frankfurt a. M.: Peter Lang.
- Green, Georgia M. (2004): Some Interactions of Pragmatics and Grammar. In: Horn, Laurence R. und Gregory Ward (Hg.): Handbook of Pragmatics. Malden, MA: Blackwell Publishers, S. 407-426.
- Green, Stephen (1997): Building hypertext links in newspaper articles using semantic similarity. In: Proceedings of the 3rd Workshop on Applications of Natural Language to Information Systems (NLDB 1997). Vancouver, Canada, S. 178-190.
- Greimas, Algirdas Julien (1971): Strukturele Semantik. Methodologische Untersuchungen. Braunschweig: Vieweg.
- Grice, Paul H. (1979): Logik und Konversation. In: Meggle, Georg (Hg.): Handlung, Kommunikation, Bedeutung. Frankfurt a. M.: Suhrkamp, S. 243-265.
- Grosz, Barbara J., Aravind K. Joshi und Scott Weinstein (1995): Centering: A framework for modeling the local coherence of discourse. In: Computational Linguistics 21 (2), S. 203-225.
- Grosz, Barbara und Candace Sidner (1986): Attention, intention and the structure of discourse. In: Computational Linguistics 12 (3), S. 175-204.
- Gülich, Elisabeth, Heiko Hausendorf und Wolfgang Raible (1977): Linguistische Textmodelle. Grundlagen und Möglichkeiten. München: Wilhelm Fink.
- Gülich, Elisabeth und Heiko Hausendorf (2000): Vertextungsmuster Narration. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 369-385. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Gumperz, John (1982): Discourse strategies. Cambridge: Cambridge University Press. (= Studies in interactional sociolinguistics, 1).
- Gundel, Jeanette K. (1977): Role of Topic and Comment in Linguistic Theory. Bloomington: Indiana University Linguistics Club.
- Gundel, Jeanette K. und Thorstein Fretheim (2004): Topic and Focus. In: Horn, Laurence R. und Gregory Ward (Hg.): Handbook of Pragmatics. Malden, MA: Blackwell Publishers, S. 175-196.
- Hagemann, Jörg (2007): Typographie und logisches Textdesign. In: Roth, Kersten Sven und Jürgen Spitzmüller (Hg.): Textdesign und Textwirkung in der massenmedialen Kommunikation. Konstanz: UVK, S. 77-91.
- Hagenbruch, André (2010): Flache Satzverarbeitung. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hg.): Computerlinguistik und Sprachtechnologie. Eine Einführung. 3., überarb. u. erw. Aufl. Heidelberg: Spektrum Akademischer Verlag, S. 264-279.
- Halliday, Michael (1967): Notes on transitivity and theme in English. Part 2. In: Journal of Linguistics 3, S. 199-244.
- Halliday, Michael (1978): Language as social semiotic. The social interpretation of language and meaning. London: Edward Arnold.
- Halliday, Michael (1985): Introduction to Functional Grammar. London: Edward Arnold.
- Halliday, Michael und Ruqaiya Hasan (1976): Cohesion in English. London, New York: Longman.



- Harabagiu, Sanda (1999): From Lexical Cohesion to Textual Coherence: A Data Driven Perspective. In: *International Journal of Pattern Recognition and Artificial Intelligence* 13 (2), S. 247-265.
- Harras, Gisela, Edeltraud Winkler, Sabine Erb und Kristel Proost (2004): *Handbuch deutscher Kommunikationsverben. Teil 1: Wörterbuch*. Berlin, New York: de Gruyter. (= Schriften des Instituts für deutsche Sprache, 10.1).
- Hasan, Ruqaiya (1984): Coherence and Cohesive Harmony. In: Flood, James (Hg.): *Understanding Reading Comprehension: Cognition, Language and the Structure of Prose*. Newark, Delaware: International Reading Association, S. 181-219.
- Hayes, John R. und Linda S. Flower (1980): Identifying the Organization of Writing Processes. In: Gregg, Lee W. und Erwin R. Steinberg (Hg.): *Cognitive Processes in Writing*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, S. 3-30.
- Hearst, Marti A. (1994): Multi-paragraph Segmentation of Expository Text. In: *Proceedings of the 32nd annual meeting on Association for Computational Linguistics (ACL '94)*. Las Cruces, New Mexico, S. 9-16.
- Hearst, Marti A. (1997): TextTiling: segmenting text into multiparagraph subtopic passages. In: *Computational Linguistics* 23 (1), S. 33-64.
- Heinemann, Margot und Wolfgang Heinemann (2002): *Grundlagen der Textlinguistik. Interaktion - Text - Diskurs*. Tübingen: Max Niemeyer. (= Germanistische Linguistik, 230).
- Heinemann, Wolfgang (2000a): Das Isotopiekonzept. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. 1. Halbband. Berlin, New York: de Gruyter, S. 54-60. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Heinemann, Wolfgang (2000b): Textsorte - Textmuster - Texttyp. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. 1. Halbband. Berlin, New York: de Gruyter, S. 506-522. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Heinemann, Wolfgang (2000c): Vertextungsmuster Deskription. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. 1. Halbband. Berlin, New York: de Gruyter, S. 356-369. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Heinemann, Wolfgang und Dieter Viehweger (1991): *Textlinguistik. Eine Einführung*. Tübingen: Max Niemeyer.
- Helbig, Hermann (2006): *Knowledge Representation and the Semantics of Natural Language*. Berlin, Heidelberg: Springer.
- Hellwig, Peter (1984a): Grundzüge einer Theorie des Textzusammenhangs. In: Rothkegel, Anneli und Barbara Sandig (Hg.): *Text - Textsorten - Semantik: linguistische Modelle und maschinelle Verfahren*. Hamburg: Buske, S. 51-79. (= Papiere zur Textlinguistik, 52).
- Hellwig, Peter (1984b): Titulus oder Über den Zusammenhang von Titeln und Texten. Titel sind ein Schlüssel zur Textkonstitution. In: *Zeitschrift für Germanistische Linguistik* 12 (1), S. 1-20.
- Hernandez, Nicolas und Brigitte Grau (2003): What is this text about? Combining topic and meta-descriptors for text structure presentation. In: *Proceedings of the Annual International Conference of the Special Interest Group for Design of Communication of the Association for Computing Machinery (ACM-SIGDOC 2003)*. October 12-15, 2003, San Francisco, USA.
- Herrmann, Theo und Siegfried Hoppe-Graff (1989): Textproduktion. In: Antos, Gerd und Hans P. Krings (Hg.): *Textproduktion: ein interdisziplinärer Forschungsüberblick*.

- Tübingen: Max Niemeyer, S. 146-164. (= Konzepte der Sprach- und Literaturwissenschaft, 48).
- Herrmanns, Fritz (2002): ‚Bilder im Kopf. Zur Wiederauferstehung des Begriffes der ‚Idee‘ und der ‚Vorstellung‘ in den Begriffen des ‚Stereotyps‘, des ‚Schemas‘ des *frame* sowie ähnlicher Begriffe. In: Wiesinger, Peter (Hg.): Akten des X. Internationalen Germanistenkongresses Wien 2000. «Zeitenwende – Die Germanistik auf dem Weg vom 20. ins 21. Jahrhundert» Bd. 2. Bern: Peter Lang, S. 291-297.
- Heyer, Gerhard, Uwe Quasthoff und Thomas Wittig (2006): Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse. Herdecke, Bochum: W3L.
- Hilbert, Mirco, Oliver Schonefeld und Andreas Witt (2005): Making CONCUR Work. In: Proceedings of the Extreme Markup Languages. 1-5 August 2005, Montreal, Quebec, Canada.
- Hinrichs, Erhard W., Katja Filippova und Holger Wunsch (2005): A Data-driven Approach to Pronominal Anaphora Resolution for German. In: Proceedings of the Recent Advances in Natural Language Processing (RANLP). Borovets, Bulgaria.
- Hobbs, Jerry R. (1979): Coherence and Coreference. In: Cognitive Science 3 (1), S. 67-90.
- Hoey, Michael (1991): Patterns of Lexis in Text. Oxford: Oxford University Press.
- Hoey, Michael (2004): Lexical priming and the properties of text. In: Partington, Alan, John Morley und Louann Haarman (Hg.): Copora and Discourse. Bern: Peter Lang, S. 385-412.
- Hoffmann, Lothar (1988): Vom Fachwort zum Fachtext. Beiträge zur Angewandten Linguistik. Tübingen: Gunter Narr. (= Forum für Fachsprachen-Forschung, 5).
- Hoffmann, Ludger (1997): Thematische Organisation von Text und Diskurs. In: Hoffmann, Ludger, Gisela Zifonun und Bruno Strecker (Hg.): Grammatik der deutschen Sprache. Bd. 1. Berlin, New York: de Gruyter. (= Schriften des Instituts für deutsche Sprache, 7.1).
- Hoffmann, Ludger (2000): Thema, Themenentfaltung, Makrostruktur. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 344-356. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Hovy, Eduard und Chin-Yew Lin (1999): Automated Text Summarization in SUMMARIST. In: Mani, I. und M. Maybury (Hg.): Advances in Automatic Text Summarization. Cambridge (Mass.), London: MIT Press, S. 81-94.
- Hovy, Eduard und Elisabeth Maier (1995): Parsimonious or profligate: How many and which discourse structure relations? : Unveröffentlichtes Manuskript. [Stand: 15.09.2010]. URL: <http://www.isi.edu/natural-language/people/hovy/publications.html> (zuletzt abgerufen am 23.09.2010).
- Institut für deutsche Sprache (Hg.): (2003ff.): eLexiko. In: OWID – Online Wortschatz-Informationssystem Deutsch. Mannheim.
- International Organization for Standardization (ISO) (1986): ISO 8879:1986. Information Processing – Text and Office Systems – Standard Generalized Markup Language.
- Jackendoff, Ray (1988): Conceptual Semantics. In: Eco, Umberto, Marco Santambrogio und Patrizia Violi (Hg.): Meaning and Mental Representations. Bloomington and Indianapolis: Indiana University Press, S. 81-98. (= Advances in Semiotics).
- Jackendoff, Ray (1998): Why a Conceptualist View of Reference? A Reply to Abbott. In: Linguistics and Philosophy 21, S. 211-219.
- Jäger, T. Florian und D. Oshima (2002): Towards a Dynamic Model of Topic-Marking. In: Proceedings of the Workshop on Information Structure in Context. Stuttgart, Germany.
- Jahr, Silke (1996): Das Verstehen von Fachtexten. Rezeption - Kognition - Applikation. Tübingen: Gunter Narr. (= Forum für Fachsprachen-Forschung, 34).

- Jahr, Silke (2000): Vertextungsmuster Explikation. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 385-397. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Jakobs, Eva-Maria (1997): Textproduktionsprozesse in den Wissenschaften. In: Jakobs, Eva-Maria und Dagmar Knorr (Hg.): Schreiben in den Wissenschaften. Frankfurt a. M.: Peter Lang, S. 1-11. (= Textproduktion und Medium, 1).
- Jakobson, Roman (1971): Linguistik und Poetik (1960). In: Ihwe, Jens (Hg.): Literaturwissenschaft und Linguistik. Ergebnisse und Perspektiven. Bd. II/1. Zur linguistischen Basis der Literaturwissenschaft. Frankfurt a. M.: Athenäum, S. 142-178.
- Jasinskaja, Ekaterina, Jörg Mayer und David Schlangen (2004): Discourse Structure and Information Structure: Interfaces and Prosodic Realization. In: Ishihara, S., M. Schmitz und A. Schwarz (Hg.): Working Papers of the SFB632. Interdisciplinary Studies on Information Structure. 1. Potsdam: Universitätsverlag Potsdam, S. 151-206.
- Jobbins, Amanda C. und Lindsay J. Evett (1998): Text Segmentation Using Reiteration and Collocation. In: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL 1998). Montréal, Canada, S. 614-618.
- Johnson, Christopher und Charles J. Fillmore (2000): The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In: Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2000). Seattle, Washington, S. 56-62.
- Johnson-Laird, Philip N. (1988): How is Meaning Mentally Represented? In: Eco, Umberto, Marco Santambrogio und Patrizia Violi (Hg.): Meaning and Mental Representations. Bloomington and Indianapolis: Indiana University Press, S. 99-118. (= Advances in Semiotics).
- Jürgens, Frank (1995): Textthema und Textsorte. In: Pohl, Inge (Hg.): Semantik von Wort, Satz und Text. Beiträge des Kolloquiums „Semantik von Wort, Satz und Text“ in Rostock 1994. Frankfurt a. M.: Peter Lang, S. 247-260. (= Sprache - System und Tätigkeit, 14).
- Kando, Noriko (1997): Text-level Structure of Research Papers: Implications for Text-Based Information Processing Systems. In: Proceedings of the 19th Annual Colloquium of the British Computer Society on Information Retrieval Research (BCS-IRSG). Aberdeen, Scotland, S. 68-81.
- Kando, Noriko (1999): Text Structure Analysis as a Tool to Make Retrieved Documents Usable. In: Proceedings of the 4th International Workshop on Information Retrieval with Asian Languages. Taipei, Taiwan, S. 126-135.
- Karamuftuoglu, Murat und Olga Vechtomova (2008): Lexical Cohesion and Term Proximity in Document Ranking. In: Information Processing and Management 44 (5), S. 1485-1502.
- Karttunen, Lauri (2003): Discourse referents. In: Gutiérrez-Rexach, Javier (Hg.): Semantics: Critical Concepts in Linguistics. Vol. III. London: Routledge, S. 20-39.
- Keenan, Elinor O. und Bambi B. Schieffelin (1976): Topics as a Discourse Notion: A Study of Topic in the Conversations of Children and Adults. In: Li, Charles N. (Hg.): Subject and Topic. New York: Academic Press, S. 335-384.
- Kim, Su Nam und Min-Yen Kan (2009): Re-examining Automatic Keyphrase Extraction Approaches in Scientific Articles. In: Proceedings of the Proceedings of the ACL/IJCNLP-Workshop on Multiword Expressions (ACL-IJCNLP 2009). Singapore, S. 9-16.
- Kindt, Walther (1995): Informationsstruktur und Inferenzbildung. In: Pohl, Inge (Hg.): Semantik von Wort, Satz und Text. Beiträge des Kolloquiums „Semantik von Wort,

- Satz und Text“ in Rostock 1994. Frankfurt a. M.: Peter Lang, S. 261-284. (= Sprache - System und Tätigkeit, 14).
- Kintsch, Walter (1995): How readers construct situation models for stories: The role of syntactic cues and causal inferences. In: Gernsbacher, M. A. und Talmy Givón (Hg.): *Coherence in Spontaneous Text*. Amsterdam, Philadelphia: John Benjamins, S. 139-160. (= *Typological Studies in Language*, 31).
- Klein, Wolfgang und Christiane von Stutterheim (1992): Textstruktur und referentielle Bewegung. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 22 (86), S. 67-92.
- Klein-Berning, Jette (2007): Multilinguales Information Retrieval mit Latent Semantic Indexing. In: Rehm, Georg, Andreas Witt und Lothar Lemnitzer (Hg.): *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen*. Proceedings of the Biennial GLDV Conference 2007. Tübingen: Gunter Narr, S. 219-225.
- Klimanow, Gerda (1977): Überlegungen zu einigen definitorischen Kriterien des Textbegriffs. In: Danes, Frantisek und Dieter Viehweger (Hg.): *Probleme der Textgrammatik II*. Berlin: Akademie Verlag, S. 181-184.
- Klink, Stefan und Thomas Kieninger (2001): Rule-based Document Structure Understanding with a Fuzzy Combination of Layout and Textual Features. In: *International Journal of Document Analysis and Recognition* 4 (1), S. 18-26.
- Knott, Alistair, Jon Oberlander, Michael O'Donnell und Chris Mellish (2001): Beyond Elaboration: The interaction of relations and focus in coherent text. In: Sanders, Ted, Joost Schilperoord und Wilbert Spooren (Hg.): *Text representation: Linguistic and psycholinguistic aspects*. Amsterdam, Philadelphia: John Benjamins, S. 181-196.
- Koch, Peter und Wulf Oesterreicher (1994): *Schriftlichkeit und Sprache*. In: Günther, Hartmut und Otto Ludwig (Hg.): *Schrift und Schriftlichkeit*. Berlin, New York: de Gruyter, S. 587-604. (= *Handbücher zur Sprach und Kommunikationswissenschaft (HSK)*, 10.1).
- Kohlmann, Ute (1997): *Objektreferenzen in Beschreibungen und Instruktionen. Eine empirische Untersuchung zum Zusammenhang von Textstruktur, referentieller Bewegung und Formen von Objektreferenzen*. Frankfurt a. M.: Peter Lang. (= *Europäische Hochschulschriften: Reihe 21, Linguistik*, 187).
- Konerding, Klaus-Peter (1993): *Frames und lexikalisches Bedeutungswissen. Untersuchungen zur linguistischen Grundlegung einer Frametheorie und zu ihrer Anwendung in der Lexikographie*. Tübingen: Max Niemeyer. (= *Germanistische Linguistik*, 142).
- Konerding, Klaus-Peter (2001): *Sprache im Alltag und kognitive Linguistik: Stereotype und schematisiertes Wissen*. In: Lehr, Andrea, Matthias Kammerer, Klaus-Peter Konerding, Angelika Storrer, Caja Thimm und Werner Wolski (Hg.): *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*. Berlin: Walter de Gruyter, S. 151-172.
- Kosta, Peter (2009): *Prototypensemantik und Stereotypen*. In: Kempgen, Sebastian, Peter Kosta, Tilman Berger, Karl Gutschmidt und Sebastian Kempgen (Hg.): *Die slawischen Sprachen/The Slavic Languages. Ein internationales Handbuch zu ihrer Geschichte, ihrer Struktur und ihrer Erforschung. Band 1*. Berlin, New York: Mouton de Gruyter, S. 828-848.
- Krohn, Wolfgang und Günter Küppers (1992): *Glossar*. In: Krohn, Wolfgang und Günter Küppers (Hg.): *Emergenz: Die Entstehung von Ordnung, Organisation und Bedeutung*. 2. Aufl. Frankfurt a. M.: Suhrkamp, S. 388-395.
- Kruijff-Korbayova, Ivana und Mark Steedman (2003): *Discourse and information structure*. In: *Journal of Logic, Language and Information* 12 (3), S. 249-259.
- Kübler, Sandra (2008): *The PaGe 2008 Shared Task on Parsing German*. In: *Association for Computational Linguistics: Proceedings of the Proceedings of the Workshop on Parsing German*. Columbus, Ohio, S. 55-63.

- Kunze, Claudia (2010): Lexikalisch-semantische Ressourcen. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hg.): *Computerlinguistik und Sprachtechnologie*. 3., überarb. u. erw. Aufl. Heidelberg: Spektrum Akademischer Verlag, S. 504-514.
- Kunze, Claudia und Lothar Lemnitzer (2002): GermaNet - representation, visualization, application. In: *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*. S. 1485-1491.
- Kupiec, Julian, Jan Pedersen und Francine Chen (1995): A trainable document summarizer. In: *Proceedings of the Proceedings of the 18th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1995)*. S. 68-73.
- Lambrecht, Knud (1994): *Information Structure and Sentence Form. Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Landauer, Thomas, Peter Foltz und Darrell Laham (1998): An Introduction to Latent Semantic Analysis. In: *Discourse Processes* 25 (2-3), S. 259-284.
- Lang, Joel und Mirella Lapata (2010): Unsupervised Induction of Semantic Roles. In: *Association for Computational Linguistics: Proceedings of the Human Language Technology Conference and Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2010)*. Los Angeles, California, S. 939-947.
- Langer, Hagen (2001): Syntax und Parsing. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Endriss, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hg.): *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg, Berlin: Spektrum Akademischer Verlag, S. 203-245.
- Langer, Hagen, Harald Lüngen und Petra Saskia Bayerl (2004a): Text type structure and logical document structure. In: *Proceedings of the ACL Workshop on Discourse Annotation*. Barcelona, Spain, S. 49-56.
- Langer, Hagen, Harald Lüngen und Petra Saskia Bayerl (2004b): Towards automatic annotation of text type structure: experiments using an XML-annotated corpus and automatic text classification methods. In: *Proceedings of the LREC Workshop on XML-based richly annotated corpora (Xbrec)*. Lissabon, Portugal.
- Lappin, Shalom und Herbert J. Leass (1994): An Algorithm for Pronominal Anaphora Resolution. In: *Computational Linguistics* 20 (4), S. 535-561.
- Lavid, Julia (1997): Controlling thematic choices in discourse: towards a specification of contextual constraints. In: *Proceedings of the First International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 1997)*. Rio de Janeiro, Brazil, S. 107-118.
- Lavid, Julia (2000): Contextual Constraints on Thematization in Written Discourse: An Empirical Study. In: Bonzon, P., M. Cavalcanti und R. Nossun (Hg.): *Formal Aspects of Context*. Dordrecht: Kluwer Academic Publishers, S. 37-47.
- Lawrie, Dawn und W. Bruce Croft (2000): Discovering and Comparing Topic Hierarchies. In: *Proceedings of the 6th International RIAO Conference*. Paris, France, S. 314-330.
- Lawrie, Dawn, W. Bruce Croft und Arnold Rosenberg (2001): Finding Topic Words for Hierarchical Summarization. In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New Orleans, Louisiana, USA, S. 349-357.
- Leinfellner, Elisabeth (1992): *Semantische Netze und Textzusammenhang*. Frankfurt a. M.: Peter Lang. (= *Arbeiten zur Sprachanalyse*, 14).
- Lemnitzer, Lothar und Claudia Kunze (2007): *Computerlexikographie*. Tübingen: Gunter Narr.
- Lemnitzer, Lothar und Heike Zinsmeister (2006): *Korpuslinguistik. Eine Einführung*. Tübingen: Gunter Narr.

- Lenerz, Jürgen (1977): Zur Abfolge nominaler Satzglieder im Deutschen. Tübingen: Stauffenburg. (= Studien zur deutschen Grammatik, 5).
- Lenz, Eva Anna und Angelika Storrer (2003): Annotationsschicht: Thematische Strukturen/Themenentwicklung. Technischer Report. TU Dortmund. (= Interne Reports der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“). URL: <http://coli.lili.uni-bielefeld.de/Texttechnologie/Forschergruppe/publications/hytext-themenentwicklung.pdf>.
- Lerch, Hans-Jürgen (1991): Strukturen der Wissensrepräsentation. Empirische Analysen von Informationsverarbeitungsprozessen beim Verstehen von Texten. Regensburg: S. Roderer. (= Theorie und Forschung; 123; Psychologie; 49).
- Leskovec, Jure, Lars Backstrom und Jon Kleinberg (2009): Meme-tracking and the Dynamics of the News Cycle. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009). New York, NY, USA, S. 497-506.
- Levy, Elena (1982): Towards an objective definition of discourse topic. In: Proceedings of the 18th regional meeting of the Chicago Linguistic Society. Chicago, USA, S. 295-304.
- Li, Charles N. und Sandra A. Thompson (1976): Subject and Topic: A New Typology of Language. In: Li, Charles N. (Hg.): Subject and Topic. New York: Academic Press, S. 457-490.
- Lin, Chin-Yew (1995): Knowledge-based Automatic Topic Identification. In: Proceedings of the 33rd annual meeting of the Association for Computational Linguistics (ACL 1995). Cambridge, Massachusetts, S. 308-310.
- Lin, Chin-Yew (1997): Robust automated topic identification. PhD Thesis.: Faculty of the Graduate School, University of Southern California.
- Lin, Chin-Yew (1998): Assembly of topic extraction modules in SUMMARIST. In: Proceedings of the AAAI Spring Symposium on Intelligent Text Summarization. S. 44-50.
- Lin, Chin-Yew und Eduard Hovy (1997): Identifying Topics by Position. In: Association for Computational Linguistics: Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP 1997). Washington, D.C., USA, S. 283-290.
- Lin, Chin-Yew und Eduard Hovy (2000): The Automated Acquisition of Topic Signatures for Text Summarization. In: Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000) - Vol. 1. Saarbrücken, Germany, S. 495-501.
- Linke, Angelika, Markus Nussbaumer und Paul R. Portmann (2004): Studienbuch Linguistik. Tübingen: Max Niemeyer. (= Germanistische Linguistik, 121).
- Lobin, Henning (2001): Informationsmodellierung in XML und SGML. Berlin: Springer.
- Lobin, Henning (2003): Textauszeichnung und Dokumentgrammatiken. In: Lobin, Henning und Lothar Lemnitzer (Hg.): Texttechnologie. Perspektiven und Anwendungen. Tübingen: Stauffenburg, S. 51-82.
- Lobin, Henning (2004): Dokumentgrammatiken als Grundlage von XML-Tools. In: Mehler, Alexander und Henning Lobin (Hg.): Automatische Textanalyse. Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte. Wiesbaden: Westdeutscher Verlag, S. 23-38.
- Lobin, Henning (2009): Inszeniertes Reden auf der Medienbühne. Zur Linguistik und Rhetorik der wissenschaftlichen Präsentation. Frankfurt a. M./ New York: Campus. (= Interaktiva. Schriftenreihe des Zentrums für Medien und Interaktivität (ZMI), Gießen, Bd. 8).
- Lobin, Henning, Maja Bärenfänger, Mirco Hilbert, Harald Lungen und Csilla Puskas (2010): Discourse relations and document structure. In: Metzling, Dieter und Andreas Witt (Hg.): Linguistic modeling of information and Markup Languages. Con-

- contributions to Language Technology. Dordrecht: Springer. (= Text, Speech and Language Technology).
- Lobin, Henning, Harald Lüngen, Mirco Hilbert und Maja Bärenfänger (2011): Processing Text-technological Resources in Discourse Parsing. In: Mehler, Alexander , Kai-Uwe Kühnberger, Henning Lobin, Harald Lüngen, Angelika Storrer und Andreas Witt (Hg.): *Modelling, Learning and Processing of Text-Technological Data Structures*. Springer. (= Studies in Computational Intelligence).
- Longacre, Richard E. (1979): The Paragraph as a Grammatical Unit. In: Givón, Talmy (Hg.): *Syntax and Semantics*. San Diego, New York: Academic Press, S. 115-134. (= *Discourse and Syntax*, 12).
- Lönneker, Birte (2003): Konzeptframes und Relationen. Extraktion, Annotation und Analyse französischer Corpora aus dem World Wide Web. Berlin: Akademische Verlagsgesellschaft.
- Lötscher, Andreas (1987): *Text und Thema. Studien zur thematischen Konstituenz von Texten*. Tübingen: Max Niemeyer. (= *Germanistische Linguistik*, 81).
- Luhn, H. (1958): The automatic creation of literature abstracts. In: *IBM Journal of Research and Development* 2 (2), S. 159-165.
- Lundquist, Lita (1989): Coherence in Scientific Texts. In: Heydrich, Wolfgang, Fritz Neubauer, János Petöfi und Emel Sözer (Hg.): *Connexity and Coherence. Analysis of Text and Discourse*. Berlin, New York: de Gruyter, S. 122-149.
- Lüngen, Harald, Csilla Puskas, Maja Bärenfänger, Mirco Hilbert und Henning Lobin (2006): Discourse segmentation of German written text. In: Salakoski, T. , F. Ginter, S. Pyysalo und T. Phikkala (Hg.): *Advances in Natural Language Processing. 5th International Conference on Natural Language Processing, FinTAL 2006*. Turku, Finland, August 2006, Proceedings. Springer, S. 245-256. (= *Lecture Notes in Artificial Intelligence*, 4139).
- Lutzeier, Peter Rolf (1995): *Lexikologie: ein Arbeitsbuch*. Tübingen: Stauffenburg.
- Makkonen, Juha, Helena Ahonen-Myka und Marko Salmenkivi (2004): Simple Semantics in Topic Detection and Tracking. In: *Information Retrieval* 7 (3-4), S. 347-368.
- Mann, William C. und Sandra A. Thompson (1988): Rhetorical Structure Theory: Towards a functional theory of text organisation. In: *Text* 8 (3), S. 243-281.
- Marcu, Daniel (1997): From discourse structures to text summaries. In: *Proceedings of the ACL/EACL Workshop on Intelligent Scalable Text Summarization*. Madrid, Spain, S. 82-88.
- Marcu, Daniel (1999): Discourse Trees are good indicators of importance in text. In: Mani, I. und M. Maybury (Hg.): *Advances in Automatic Text Summarization*. Cambridge (Mass.), London: MIT Press, S. 123-136.
- Marcu, Daniel (2000): *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge (Mass.), London: MIT Press.
- Martin, Jim R. (1995): More than what the message is about: English Theme. In: Ghadessy, Mohsen (Hg.): *Thematic Development in English Texts*. London: Pinter, S. 223-258.
- Mathesius, Vilém (1971 [1929]): Die funktionale Linguistik. In: Benes, Eduard und Josef Vachek (Hg.): *Stilistik und Soziolinguistik. Beiträge der Prager Schule zur strukturellen Sprachbetrachtung und Spracherziehung*. 2. Aufl. Berlin: Paul List Verlag, S. 1-18.
- Matthiessen, Christian (1995): THEME as an enabling resource in ideational 'knowledge' construction. In: Ghadessy, Mohsen (Hg.): *Thematic Development in English Texts*. London: Pinter, S. 20-54.
- Mehler, Alexander (2004): Quantitative Methoden. In: Lobin, Henning und Lothar Lemnitzer (Hg.): *Texttechnologie. Perspektiven und Anwendungen*. Tübingen: Stauffenburg Verlag, S. 83-107.

- Metzelin, Michael und Harald Jaksche (1983): Textsemantik. Ein Modell zur Analyse von Texten. Tübingen: Gunter Narr. (= Tübinger Beiträge zur Linguistik, 224).
- Minsky, Marvin L. (1980): A framework for representing knowledge. In: Metzling, Dieter (Hg.): Frame conceptions and text understanding. Berlin: de Gruyter, S. 1-25.
- Mitkov, Ruslan, Branimir Boguraev und Shalom Lappin (2001): Introduction to the Special Issue on Computational Anaphora Resolution. In: Computational Linguistics 27 (4), S. 473-477.
- Moens, M. (2007): Summarizing court decisions. In: Information Processing and Management 43 (6), S. 1748-1764.
- Moens, Marie-Francine und Rik De Busser (2001): Generic topic segmentation of document texts. In: Proceedings of the 24th ACM SIGIR Annual International Conference on Research and Development in Information Retrieval. New York, USA, S. 418-419.
- Molitor-Lübbert, Sylvie (1996): Schreiben als mentaler und sprachlicher Prozeß. In: Günther, Hartmut und Otto Ludwig (Hg.): Schrift und Schriftlichkeit: ein interdisziplinäres Handbuch internationaler Forschung, 2. Halbband. Berlin, New York: de Gruyter, S. 1005-1027. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 10).
- Molnár, Valéria (1993): Zur Pragmatik und Grammatik des TOPIK-Begriffes. In: Reis, Marga (Hg.): Wortstellung und Informationsstruktur. Tübingen: Max Niemeyer, S. 155-202. (= Linguistische Arbeiten, 306).
- Moore, Johanna D. und Martha E. Pollack (1992): A problem for RST: The need for multi-level discourse analysis. In: Computational Linguistics 18 (4), S. 538-544.
- Morris, Jane und Graeme Hirst (1991): Lexical cohesion computed by thesaural relations as an indicator of the structure of text. In: Computational Linguistics 17 (1), S. 21-48.
- Morris, Jane und Graeme Hirst (2004): Non-classical lexical semantic relations. In: Proceedings of the Workshop on Computational Lexical Semantics at the Human Technologies Conference of the North American Chapter the Association for Computational Linguistics (HLT-NAACL 2004). Boston, Massachusetts, USA.
- Morris, Jane und Graeme Hirst (2005): The Subjectivity of Lexical Cohesion in Text. In: Shanahan, James G., Yan Qu und Janyce Wiebe (Hg.): Computing attitude and affect in text. Dordrecht: Springer.
- Motsch, Wolfgang (1996): Ebenen der Textstruktur. Begründung eines Forschungsprogramms. In: Motsch, Wolfgang (Hg.): Ebenen der Textstruktur: sprachliche und kommunikative Prinzipien. Tübingen: Max Niemeyer, S. 3-36. (= Germanistische Linguistik, 164).
- Neumann, Günter (2010): Text-basiertes Informationsmanagement. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hg.): Computerlinguistik und Sprachtechnologie. 3., überarb. u. erw. Aufl. Heidelberg: Spektrum Akademischer Verlag, S. 576-615.
- Nomoto, Tadashi und Yuji Matsumoto (1996): Exploiting Text Structure for Topic Identification. In: Proceedings of the 4th Workshop on Very Large Corpora (WVLC-4) at the Conference on Computational Linguistics (COLING 1996). Copenhagen, Denmark, S. 101-112.
- Nöth, Winfried (2000): Handbuch der Semiotik. Stuttgart/Weimar: J. B. Metzler.
- Nuopponen, Anita (1993): Begriffssysteme und Textstruktur. Am Beispiel deutscher und finnischer Enzyklopädieartikel. In: Schröder, Hartmut (Hg.): Fachtextpragmatik. Tübingen: Gunter Narr, S. 99-113.
- OASIS – Organization for the Advancement of Structured Information Standards (2001): RELAX NG Specification. Committee Specification. URL: <http://www.relaxng.org/spec-20011203.html> (zuletzt abgerufen am 12.10.2010).
- Oberlander, Jon (2004): On the reduction of discourse topic. In: Theoretical Linguistics 30 (2-3), S. 213-225.



- Ortner, Hanspeter und Horst Sitta (2003): Was ist der Gegenstand der Sprachwissenschaft? In: Linke, Angelika, Hanspeter Ortner und Paul R. Portmann-Tselikas (Hg.): Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis. Tübingen: Max Niemeyer, S. 3-64.
- Paul, Herrmann (1898): Prinzipien der Sprachgeschichte. Halle a. S.: Max Niemeyer.
- Poesio, Massimo, Rahul Mehta, Axel Maroudas und Janet Hitzeman (2004): Learning to resolve bridging references. In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004). Barcelona, Spain.
- Ponzetto, Simone Paolo und Michael Strube (2007a): Deriving a Large Scale Taxonomy from Wikipedia. In: Proceedings of the 22nd National Conference on Artificial Intelligence (IAAI-07) - Vol. 2. Vancouver, British Columbia, Canada, S. 1440-1445.
- Ponzetto, Simone Paolo und Michael Strube (2007b): Knowledge Derived From Wikipedia For Computing Semantic Relatedness. In: Journal of Artificial Intelligence Research 30, S. 181-212.
- Poon, Hoifung und Pedro Domingos (2008): Joint Unsupervised Coreference Resolution with Markov Logic. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008). Honolulu, Hawaii, S. 650-659.
- Postolache, Oana, Ivana Kruijff-Korbayová und Geert-Jan M. Kruijff (2005): Data-driven approaches for information structure identification. In: Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005). Vancouver, Canada, S. 9-16.
- Power, Richard, Donia Scott und Nadjet Bouayad-Agha (2003): Document Structure. In: Computational Linguistics 29 (2), S. 211-260.
- Prince, Ellen F. (1981): Toward a Taxonomy of Given-New Information. In: Cole, Peter (Hg.): Radical Pragmatics. New York: Academic Press, S. 223-255.
- Pufahl, Ingrid (1992): Topics in German and US television news reports. In: Linguistics 30 (1), S. 217-241.
- Putnam, Hilary (1990 [1975]): Die Bedeutung von ‚Bedeutung‘ [The Meaning of 'Meaning']. Frankfurt a. M.: Klostermann.
- Ramm, Wiebke und Claudia Villiger (1995): Global Text Organization and Sentence-Grammatical Realization: Discourse-Level Constraints on Theme Selection. Claus-Report Nr. 61. Saarbrücken: Universität des Saarlandes.
- Rauh, Gisa (1988): Tiefenkasus, thematische Relationen und Thetarollen. Die Entwicklung einer Theorie von semantischen Relationen. Tübingen: Gunter Narr. (= Tübinger Beiträge zur Linguistik, 309).
- Redeker, Gisela (2000): Coherence and structure in text and discourse. In: Black, William und Harry Bunt (Hg.): Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics. Amsterdam: Benjamins, S. 233-263.
- Rehm, Georg (2005): Hypertextsorten. Definition - Struktur - Klassifikation. PhD Thesis. Gießen: Justus-Liebig-Universität Gießen.
- Reimer, Ulrich (1991): Einführung in die Wissensrepräsentation: Netzartige und schemabasierte Repräsentationsformate. Stuttgart: Teubner. (= Leitfaden der angewandten Informatik).
- Reinhart, Tanya (1982): Pragmatics and Linguistics: An Analysis of Sentence Topics. Bloomington: Indiana University Linguistics Club.
- Reis, Marga (1977): Präsuppositionen und Syntax. Tübingen: Max Niemeyer. (= Linguistische Arbeiten, 51).
- Reis, Marga (1993): Satzfügung und kommunikative Gewichtung. Zur Grammatik und Pragmatik von Neben- vs. Unterordnung am Beispiel ‚implikativer‘ und-Konstruktionen im Deutschen. In: Reis, Marga (Hg.): Wortstellung und Informationsstruktur. Tübingen: Max Niemeyer, S. 203-250. (= Linguistische Arbeiten, 306).
- Rickheit, Gert, Lorenz Sichelschmidt und Hans Strohner (2002): Psycholinguistik. Tübingen: Stauffenburg.

- Rickheit, Gert und Hans Strohner (1993): Grundlagen der kognitiven Sprachverarbeitung. Modelle, Methoden, Ergebnisse. Tübingen, Basel: A. Francke.
- Rose, Stuart, Dave Engel, Nick Cramer und Wendy Cowley (2010): Automatic keyword extraction from individual documents. In: Berry, Michael W. und Jacob Kogan (Hg.): Text Mining: Applications and Theory. Chichester, UK: John Wiley & Sons, S. 3-20.
- Rothkegel, Annely (1984a): Frames und Textstruktur. In: Rothkegel, Annely und Barbara Sandig (Hg.): Text - Textsorten - Semantik: linguistische Modelle und maschinelle Verfahren. Hamburg: Buske, S. 238-273. (= Papiere zur Textlinguistik, 52).
- Rothkegel, Annely (1984b): Thematisieren als Texthandlung. In: Krenn, H., J. Niemeyer und U. Eberhardt (Hg.): Sprache und Gesellschaft. Tübingen: Max Niemeyer, S. 63-73.
- Rothkegel, Annely (1989): Textualisierung von Wissen. Einige Forschungsfragen zum Umgang mit Wissen im Rahmen computerorientierter Textproduktion. In: LDV-Forum 6 (1), S. 3-13.
- Rothkegel, Annely (1997): Strategien der thematischen Textentfaltung. In: Gather, A. und H. Werner (Hg.): Semiotische Prozesse und natürliche Sprache. Stuttgart: Steiner, S. 461-468.
- Saggion, Horacio und Guy Lapalme (2000): Concept identification and presentation in the context of technical text summarization. In: Proceedings of the ANLP-NAACL Workshop on Automatic Summarization. Seattle, USA.
- Saggion, Horacio und Guy Lapalme (2002): Generating indicative-informative summaries with SumUM. In: Computational Linguistics 28 (4), S. 497-526.
- Salton, G., A. Singhal, C. Buckley und M. Mitra (1996): Automatic text decomposition using text segments and text themes. In: Proceedings of the 7th ACM Conference on Hypertext (Hypertext '96). Washington, DC, USA, S. 53-65.
- Salton, G., A. Singhal, M. Mitra und C. Buckley (1999): Automatic Text Structuring and Summarization. In: Mani, I. und M. Maybury (Hg.): Advances in Automatic Text Summarization. Cambridge (Mass.), London: MIT Press, S. 341-356.
- Sandig, Barbara (1997): Formulieren und Textmuster. Am Beispiel von Wissenschaftstexten. In: Jakobs, Eva-Maria und Dagmar Knorr (Hg.): Schreiben in den Wissenschaften. Frankfurt a. M.: Lang, S. 25-44.
- Sanford, Anthony und Linda M. Moxey (1995): Aspects of coherence in written language: a psychological perspective. In: Gernsbacher, M.A. und T. Givón (Hg.): Coherence in Spontaneous Text. Amsterdam, Philadelphia: John Benjamins, S. 161-187. (= Typological Studies in Language, 31).
- Schank, Gerd (1977a): Über einige Regeln der Themenverwendung in natürlichen Gesprächen. In: Muttersprache 87 (4), S. 234-244.
- Schank, Roger C. (1977b): Rules and Topics in Conversation. In: Cognitive Science 1 (4), S. 421-441.
- Scherner, Maximilian (1984): Sprache als Text. Ansätze zu einer sprachwissenschaftlich begründeten Theorie des Textverstehens. Forschungsgeschichte - Problemstellung - Beschreibung. Tübingen: Max Niemeyer. (= Germanistische Linguistik; 48).
- Scherner, Maximilian (2000): Kognitionswissenschaftliche Methoden in der Textanalyse. In: Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband. Berlin, New York: de Gruyter, S. 186-195. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 16.1).
- Schlobinski, Peter und Stephan Schütze-Coburn (1992): On the topic of topic and topic continuity. In: Linguistics 30 (1), S. 89-122.
- Schmid, Helmut (2008): Tokenizing and part-of-speech tagging. In: Lüdeling, Anke und Merja Kytö (Hg.): Corpus Linguistics. An International Handbook, Vol. 1. Berlin,

- New York: de Gruyter, S. 527-551. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 29.1).
- Schnotz, Wolfgang (1994): Aufbau von Wissensstrukturen. Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten. Weinheim: Beltz. (= Fortschritte der psychologischen Forschung, 20).
- Schonefeld, Oliver (2007): XCONCUR and XCONCUR-CL: A Constraint-Based Approach for the Validation of Concurrent Markup. In: Rehm, Georg, Andreas Witt und Lothar Lemnitzer (Hg.): Data Structures for Linguistic Resources and Applications. Proceedings of the Biennial GLDV Conference 2007. Tübingen: Gunter Narr, S. 347-356.
- Schröder, Hartmut (1993): Semiotische Aspekte multimedialer Texte. In: Schröder, Hartmut (Hg.): Fachtextpragmatik. Tübingen: Gunter Narr, S. 189-214. (= Forum für Fachsprachen-Forschung, 19).
- Schröder, Thomas (2003): Die Handlungsstruktur von Texten. Ein integrativer Beitrag zur Texttheorie. Tübingen: Gunter Narr.
- Schwarz, Lara, Sabine Bartsch, Richard Eckart und Elke Teich (2008): Exploring automatic Theme identification: A rule-based approach. In: Storrer, Angelika, Alexander Geyken, Alexander Siebert und Kay-Michael Würzner (Hg.): Text Resources and Lexical Knowledge. Selected Papers from the 9th Conference on Natural Language Processing KONVENS 2008. Berlin, New York: Mouton de Gruyter, S. 15-26.
- Schwarz, Monika (1996): Einführung in die kognitive Linguistik. Tübingen, Basel: Francke Verlag.
- Schwarz, Monika (2001): Establishing coherence in text. Conceptual continuity and text-world models. In: Logos and Language 2 (1), S. 15-23.
- Schwarz, Monika und Jeannette Chur (2004): Semantik. Ein Arbeitsbuch. Tübingen: Gunter Narr.
- Searle, J. R. (1996 [1969]): Was ist ein Sprechakt? In: Hoffmann, Ludger (Hg.): Sprachwissenschaft. Ein Reader. Berlin, New York: de Gruyter, S. 143-162.
- Seuren, Pieter A. M. (1991): Präsuppositionen. In: von Stechow, Arnim und Dieter Wunderlich (Hg.): Semantik. Ein internationales Handbuch der zeitgenössischen Forschung. Berlin, New York: de Gruyter, S. 286-318. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 6).
- Sgall, Petr, Eva Hajicová und Eva Benesová (1973): Topic, Focus and Generative Semantics. Kronberg/ Taunus: Scriptor.
- Silber, H. Gregory und Kathleen F. McCoy (2002): Efficiently Computed Lexical Chains As an Intermediate Representation for Automatic Text Summarization. In: Computational Linguistics 28 (4), S. 487-496.
- Spitzmüller, Jürgen (2009): Typographisches Wissen. In: Linke, Angelika und Helmuth Feilke (Hg.): Oberfläche und Performanz. Untersuchungen zur Sprache als dynamischer Gestalt. Tübingen: Max Niemeyer, S. 459-486.
- Stede, Manfred (2007): Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik. Tübingen: Gunter Narr.
- Stede, Manfred und Florian Kuhn (2009): Identifying the Content Zones of German Court Decisions. In: Abramowicz, Witold und Dominik Flejter (Hg.): Business Information Systems Workshops - BIS 2009. International Workshops, Poznan/Poland, April 2009, Revised Papers. Berlin/Heidelberg: Springer, S. 310-315.
- Stede, Manfred und Arthit Suriyawongkul (2010): Identifying logical structure and content structure in loosely-structured documents. In: Witt, Andreas und Dieter Metzger (Hg.): Linguistic Modeling of Information and Markup Languages. Contributions to Language Technology. Dordrecht: Springer, S. 81-96.
- Stede, Manfred und Carla Umbach (1998): DiMLex: A lexicon of discourse markers for text generation and understanding. In: Proceedings of the 17th International Con-

- ference on Computational Linguistics (COLING 1998). Montreal, Canada, S. 1238-1242.
- Stein, Stefan (2003): Textgliederung. Einheitenbildung im geschriebenen und gesprochenen Deutsch: Theorie und Empirie. Berlin, New York: de Gruyter. (= *Studia Linguistica Germanica*, 69).
- Steinhoff, Torsten (2007): Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten. Tübingen: Max Niemeyer. (= *Germanistische Linguistik*, 280).
- Stokes, Nicola (2003): Spoken and Written News Story Segmentation using Lexical Chains. In: *Proceedings of the Human Language Technology Conference and Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2003)*. Edmonton, Canada, S. 49-54.
- Stokes, Nicola (2004): Applications of Lexical Cohesion Analysis in the Topic Detection and Tracking Domain. PhD Thesis. Dublin: National University of Ireland, Dublin.
- Strauß, Gerhard und Gisela Zifonun (1985): Die Semantik schwerer Wörter im Deutschen. Teil 1: Lexikologie schwerer Wörter. Tübingen: Gunter Narr. (= *Forschungsberichts des Instituts für deutsche Sprache Mannheim*, 58).
- Strawson, Peter Frederick (1950): On Referring. In: *Mind* 59, S. 320-344.
- Strube, Michael (2010): Anaphernresolution. In: Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hg.): *Computeringuistik und Sprachtechnologie*. 3., überarb. u. erw. Aufl. Heidelberg: Spektrum Akademischer Verlag, S. 399-409.
- Stuckardt, Roland (2001): Design and Enhanced Evaluation of a Robust Anaphor Resolution Algorithm. In: *Computational Linguistics* 27 (4), S. 479-506.
- Stührenberg, Maik und Daniela Goecke (2008): SGF - An integrated model for multiple annotations and its application in a linguistic domain. In: *Proceedings of the Balisage: The Markup Conference 2008*. Montréal, Quebec, Canada.
- Stührenberg, Maik und Daniel Jettka (2009): A toolkit for multi-dimensional markup: The development of SGF to XStandoff. In: *Proceedings of the Balisage: The Markup Conference 2009*. Montréal, Quebec, Canada.
- Taboada, Maite und Julia Lavid (2003): Rhetorical and thematic patterns in scheduling dialogues. In: *Functions of Language* 10 (2), S. 147-148.
- Teich, Elke und Peter Fankhauser (2003): WordNet for Lexical Cohesion Analysis. In: *Proceedings of the 2nd Global WordNet Conference*. January 20-23, 2004, Brno, Czech Republic.
- Teufel, Simone (1999): Argumentative zoning: Information extraction from scientific text. PhD Thesis. Edinburgh: University of Edinburgh.
- Teufel, Simone, Jean Carletta und Marc Moens (1999): An annotation scheme for discourse-level argumentation in research articles. In: *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL '99)*. June 8-12, 1999, Bergen, Norway.
- Teufel, Simone und Marc Moens (1999): Argumentative Classification of Extracted Sentences as a First Step Towards Flexible Abstracting. In: Mani, I. und M. Maybury (Hg.): *Advances in Automatic Text Summarization*. Cambridge (Mass.), London: MIT Press, S. 155-175.
- Teufel, Simone und Marc Moens (2002): Summarizing Scientific Articles. Experiments with Relevance and Rhetorical Status. In: *Computational Linguistics* 28 (4), S. 409-445.
- Thione, Gian Lorenzo, Martin van den Berg, Livia Polanyi und Chris Culy (2004): Hybrid Text Summarization: Combining external relevance measures with Structural Analysis. In: *Proceedings of the ACL Workshop Text Summarization Branches Out*. Barcelona, Spain.

- van Deemter, Kee und Roger Kibble (2001): On Coreferring: Coreference in MUC and related annotation schemes. In: *Journal of Computational Linguistics* 26 (4), S. 629-637.
- van Dijk, Teun (1980): *Textwissenschaft. Eine interdisziplinäre Einführung*. Tübingen: Max Niemeyer.
- van Dijk, Teun (1985): Structures of News in the Press. In: van Dijk, Teun (Hg.): *Discourse and Communication*. Berlin: de Gruyter, S. 69-93.
- van Dijk, Teun (2008): *Discourse and Context. A sociocognitive approach*. Cambridge: Cambridge University Press.
- van Dijk, Teun und Walter Kintsch (1983): *Strategies of Discourse Comprehension*. New York: Academic Press.
- van Kuppevelt, Jan (1993): Intentionality in a Topical Approach of Discourse Structure. In: *Association for Computational Linguistics: Proceedings of the ACL Workshop on Intentionality and Structure in Discourse Relations*. Ohio State University, Columbus, Ohio, USA, S. 136-139.
- van Kuppevelt, Jan (1995): Discourse Structure, Topicality and Questioning. In: *Journal of Linguistics* 31 (1), S. 109-147.
- van Mol, Maarten und Mick O'Donnell (2004): Automatic recognition of generic structure: medical discharge notices. In: Banks, David (Hg.): *Text and Texture, Systemic Functional viewpoints on the nature and structure of text*. Paris.
- Vater, Heinz (2001): *Einführung in die Textlinguistik. Struktur und Verstehen von Texten*. München: Wilhelm Fink.
- Vater, Heinz (2005): *Referenzlinguistik*. München: Wilhelm Fink.
- Ventola, Eija (1995): Thematic development and translation. In: Ghadessy, Mohsen (Hg.): *Thematic Development in English Texts*. London: Pinter, S. 85-104.
- Versley, Yannick (2006): A Constraint-based Approach to Noun Phrase Coreference Resolution in German Newspaper Text. In: *Proceedings of the KONVENS 2006*. Konstanz, Germany.
- Versley, Yannick, Simone Paolo Ponzetto, Massimo Poesio, Vladimir Eidelman, Alan Jern, Jason Smith, Xiaofeng Yang und Alessandro Moschitti (2008): BART: A Modular Toolkit for Coreference Resolution. In: *Association for Computational Linguistics: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08): Human Language Technologies (HLT) Demo Session (Companion Volume)*. Columbus, Ohio, USA, S. 9-12.
- Viehweger, Dieter (1989): Coherence – Interaction of Modules. In: Heydrich, Wolfgang, Fritz Neubauer, János Petöfi und Emel Sözer (Hg.): *Connexity and Coherence. Analysis of Text and Discourse*. Berlin, New York: de Gruyter, S. 256-274.
- Villiger, Claudia (1996): Theme, Discourse Topic, and Information Structuring in German Texts. ESPRIT Basic Research Action: Dandelion, EP6665; Deliverable R1.2.2c. Technical Report. Saarbrücken: Universität des Saarlandes.
- von der Gabelentz, Georg (1972 [1891]): *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Tübingen: Verlag TBL (Tübinger Beiträge zur Linguistik).
- von Polenz, Peter (1985): *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens*. Berlin, New York: de Gruyter. (= Sammlung Götschen, 2226).
- von Stechow, Arnim (1991): Syntax und Semantik. In: von Stechow, Arnim und Dieter Wunderlich (Hg.): *Semantik. Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin, New York: de Gruyter, S. 90-148. (= Handbücher zur Sprach- und Kommunikationswissenschaft (HSK), 6).
- von Stutterheim, Christiane (1994): Quaestio und Textaufbau. In: Kornadt, Hans-Joachim, Joachim Grabowski und Roland Mangold-Allwinn (Hg.): *Sprache und Kognition. Perspektiven moderner Sprachpsychologie*. Heidelberg: Spektrum Akademischer Verlag, S. 251-272.

- von Stutterheim, Christiane (1997): Einige Prinzipien des Textaufbaus. Empirische Untersuchungen zur Produktion mündlicher Texte. Tübingen: Max Niemeyer. (= Germanistische Linguistik 184).
- W3C – World Wide Web Consortium (2004a): Web Ontology Language (Version 1: 2004, Version 2: 2009). URL: <http://www.w3.org/TR/owl-features/> (zuletzt abgerufen am 12.10.2010).
- W3C – World Wide Web Consortium (2004b): XML Schema Part 1: Structures Second Edition, Part 2: Datatypes Second Edition. URL: <http://www.w3.org/TR/xmlschema-1/> und <http://www.w3.org/TR/xmlschema-2/> (zuletzt abgerufen am 12.10.2010).
- W3C – World Wide Web Consortium (2006): Extensible Markup Language (XML), Version 1.1 [Version 1.0: 1998]. URL: <http://www.w3.org/TR/xml11/> (zuletzt abgerufen am 12.10.2010).
- W3C – World Wide Web Consortium (2007a): XML Path Language 2.0 [Version 1.0: 1999]. URL: <http://www.w3.org/TR/xpath20/> (zuletzt abgerufen am 12.10.2010).
- W3C – World Wide Web Consortium (2007b): An XML Query Language 1.0. URL: <http://www.w3.org/TR/xquery/> (zuletzt abgerufen am 12.10.2010).
- W3C – World Wide Web Consortium (2007c): XSL Transformations (XSLT) 2.0 [Version 1.0: 1999]. URL: <http://www.w3.org/TR/xslt20/> (zuletzt abgerufen am 12.10.2010).
- Wegener, Philipp (1991 [1885]): Untersuchungen über die Grundfragen des Sprachlebens. Amsterdam/Philadelphia: John Benjamins Publishing Company. (= Series II - Classics in Psycholinguistics, 5).
- Weischenberg, Siegfried (1990): Nachrichtenschreiben. Journalistische Praxis zum Studium und Selbststudium. Opladen: Westdeutscher Verlag.
- Welke, Klaus (2002): Deutsche Syntax funktional. Perspektiviertheit syntaktischer Strukturen. Tübingen: Stauffenberg. (= Stauffenberg Linguistik, 22).
- Werlich, Egon (1975): Typologie der Texte. Entwurf eines textlinguistischen Modells zur Grundlegung einer Textgrammatik. Heidelberg: Quelle & Meyer.
- Wikberg, Kay (1990): Topic, Theme, and Hierarchical Structure in Procedural Discourse. In: Aarts, Jan und Willem Meijs (Hg.): Theory and Practice in Corpus Linguistics. Amsterdam, Atlanta: Rodopi, S. 231-254. (= Language and Computers: Studies in Practical Linguistics, 4).
- Wilson, Deirdre und Dan Sperber (2004): Relevance Theory. In: Horn, Laurence R. und Gregory Ward (Hg.): Handbook of Pragmatics. Malden, MA: Blackwell Publishers, S. 607-632.
- Winter, Alexander (1992): Metakognition beim Textproduzieren. Tübingen: Gunter Narr.
- Witt, Andreas (2004a): Linguistische Informationsmodellierung mit XML. In: Mehler, Alexander und Henning Lobin (Hg.): Automatische Textanalyse: Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte. Opladen: Westdeutscher Verlag, S. 39-54.
- Witt, Andreas (2004b): Multiple hierarchies: new aspects of an old solution. In: Proceedings of the Extreme Markup Languages 2004. Montréal, Québec, Canada.
- Witt, Andreas, Harald Lungen, Felix Sasaki und Daniela Goecke (2005): Unification of XML Documents with Concurrent Markup. In: Literary and Linguistic Computing 20 (1), S. 103-116.
- Wöllstein-Leisten, Angelika, Axel Heilmann, Peter Stepan und Sten Vikner (1997): Deutsche Satzstruktur: Grundlagen der syntaktischen Analyse. Tübingen: Stauffenberg.
- Wrobel, Arne (1995): Schreiben als Handlung. Überlegungen und Untersuchungen zur Theorie der Textproduktion. Tübingen: Max Niemeyer. (= Germanistische Linguistik, 158).
- Wulff, Hans J. (1979): Texte, Themen, Titel. Die Überschrift im Rahmen der Textsemantik. In: Papiere des Münsteraner Arbeitskreises für Semiotik (PAPMAKS): Zur Textsemiotik des Titels 12, S. 199-239.

- Yule, George und Terrie Mathis (1992): The role of staging and constructed dialogue in establishing speaker's topic. In: *Linguistics* 30 (1), S. 199-215.
- Zesch, Torsten (2010): Study of Semantic Relatedness of Words Using Collaboratively Constructed Semantic Resources. PhD Thesis. Darmstadt: TU Darmstadt.
- Zesch, Torsten und Iryna Gurevych (2009): Wisdom of crowds vs. wisdom of linguists - measuring the semantic relatedness of words. In: *Natural Language Engineering* 16 (1), S. 25-59.

## 5.2 Quellen zur Textsammlung „Waldsterben“

- Lexikonartikel I:  
Munzinger-Online/Brockhaus-Enzyklopädie in 30 Bänden: „Waldsterben“. 21., völlig neu bearbeitete Auflage. München, Ravensburg.  
[<http://www.munzinger.de/search/document?index=mol-12&id=12023054105&type=text/html&query.key=kRkmmi1u&template=/publikationen/brockhaus/document.jsp>] (zuletzt abgerufen am 14.11.2010).
- Lexikonartikel II:  
AEIOU-Lexikon: „Waldzustand/Waldsterben“. Verlag Ed. Hölzel, Wien.  
[<http://www.aeiou.at/aeiou.encyclp.w/w117732.htm>] (zuletzt abgerufen am 14.11.2010).
- Zeitungsbericht:  
Schmidt, Heike: „Weltweites Waldsterben made in Germany“. In: die tageszeitung (taz), 25.1.2006, S.3. Online:  
[<http://www.taz.de/1/archiv/archiv/?dig=2006/01/25/a0157>] (zuletzt abgerufen am 14.11.2010).
- Pressemeldung I:  
Greenpeace-Redaktion: „Waldsterben durch Klimawandel. Greenpeace fordert drastische Senkung der Luftschadstoffe.“ Pressemeldung vom 08.12.2004, Hamburg.  
[[http://www.greenpeace.de/themen/klima/presseerklaerungen/artikel/waldsterben\\_durch\\_klimawandel/](http://www.greenpeace.de/themen/klima/presseerklaerungen/artikel/waldsterben_durch_klimawandel/)] (zuletzt abgerufen am 14.11.2010).
- Pressemeldung II:  
Robin Wood: „Der Wald stirbt – die Politik schaut zu.“ Pressemeldung vom 21.03.2005. [<http://www.robinwood.de/german/presse/050318.htm>] (zuletzt abgerufen am 14.11.2010).
- Offener Brief:  
Feldpolitik-Blog: „Frau Künast und das Waldsterben.“ Artikel vom 07.11.2004.  
[<http://feldpolitik.de/feldblog/item.php?i=239>] (zuletzt abgerufen am 14.11.2010).

## 6 Anhang: Beispieltexte im Original

### 6.1 Beispiel 1: „Gewalt gegen Castor nimmt zu“<sup>247</sup>

# Gewalt gegen Castor nimmt zu

## Anschläge auf Bahnstrecken – Grüne machen bei Atomtransport-Blockade mit

**Hannover (dpa) – Mit einer Welle von Bahnanschlägen hat der gewaltsame Protest gegen den bevorstehenden Castor-Transport ins atomare Zwischenlager nach Gorleben am Dienstag bundesweit einen neuen Höhepunkt erreicht.**

In Brandenburg, Berlin, Hessen, Niedersachsen und Schleswig-Holstein zerstörten vermutlich Atomkraftgegner am frühen Morgen mit Wurfankern die Oberleitungen von Bahnstrecken. Durch die insgesamt neun Anschläge entstand nach Angaben der Deutsche Bahn AG erheblicher Sachschaden. Verletzt wurde niemand. Zum Teil verspäteten sich Züge des Nah- und Fernverkehrs bis zu drei Stunden.

Der Verfassungsschutz schreibt die Aktionen gewalttätigen Autonomen zu. Das Bundeskriminalamt geht von Einzel Tätern

aus. Die Behörde in Wiesbaden hat bisher keine Anhaltspunkte dafür, daß die bundesweiten Anschläge von einer Gruppe koordiniert waren.

Anfang März sollen sechs Castor-Behälter per Bahn ins Zwischenlager Gorleben rollen. Drei stammen aus dem Atomkraftwerk Neckarwestheim (Baden-Württemberg), einer aus dem Reaktor Gundremmingen (Bayern) und zwei Atommüllbehälter aus der französischen Wiederaufbereitungsanlage La Hague. Rund 19 000 Polizisten und Bundesgrenzschutzbeamte werden Ende dieser Woche den Transport begleiten. Nach bisherigen Informationen sollen die Behälter am 28. Februar in Walheim nahe Neckarwestheim per Bahn eintreffen. Von dort soll der rund 700 Tonnen schwere Zug auf dem kürzesten Weg durch Baden-Württemberg, Hessen und Niedersachsen nach Gorleben gehen. Der Zug soll

in Walheim am kommenden Sonntag abends abfahren und am Montag an der Verladestation Dannenberg eintreffen.

Die Parteispitze und Fraktionsmitglieder von Bündnis 90/Die Grünen in Bonn und Niedersachsen kündigten am Dienstag an, sich an den Protesten zu beteiligen. Mit einer Fuhrer Mist aus der Region Gorleben protestierte Greenpeace am Kernkraftwerk Gundremmingen gegen den Transport quer durch Deutschland.

Die Gewerkschaft der Polizei (GdP) bezeichnete die Transporte als unsinnig. „Die Atomlobby hat seit 20 Jahren ihre Hausaufgaben nicht gemacht, um für den enormen giftigen Dreck, den sie produziert, Vorsorge zu treffen“, sagte GdP-Vorstandsmitglied Eugen Roth. Die geschätzten Kosten von rund 100 Millionen Mark müsse nun der Steuerzahler tragen. (*Tagesthema „Rückfall“ und Region Seite 22*)

- (1) Gewalt gegen Castor nimmt zu [Titel]  
[Zeilenumbruch]
- (2) Anschläge auf Bahnstrecken - Grüne machen bei Atomtransport-Blockade mit  
[Unterüberschrift]  
[Zeilenumbruch]
- (3) Hannover (dpa)  
[Zeilenumbruch]
- (4) Mit einer Welle von Bahnanschlägen hat der gewaltsame Protest gegen den bevorstehenden Castor-Transport ins atomare Zwischenlager nach Gorleben am Dienstag bundesweit einen neuen Höhepunkt erreicht. [Lead]  
[Zeilenumbruch]
- (5) In Brandenburg, Berlin, Hessen, Niedersachsen und Schleswig-Holstein zerstörten vermutlich Atomkraftgegner am frühen Morgen mit Wurfankern die Oberleitungen von Bahnstrecken.
- (6) Durch die insgesamt neun Anschläge entstand nach Angaben der Deutsche Bahn AG erheblicher Sachschaden.
- (7) Verletzt wurde niemand.
- (8) Zum Teil verspäteten sich Züge des Nah- und Fernverkehrs bis zu drei Stunden.

---

<sup>247</sup> Deutsche Presseagentur (dpa): „Gewalt gegen Castor nimmt zu“ (dpa). In: Stuttgarter Nachrichten vom 26.02.1997. S. 1.



- (9) [Zeilenumbruch]
- (10) Der Verfassungsschutz schreibt die Aktionen gewalttätigen Autonomen zu.
- (11) Das Bundeskriminalamt geht von Einzeltätern aus.
- (12) Die Behörde in Wiesbaden hat bisher keine Anhaltspunkte dafür, daß die bundesweiten Anschläge von einer Gruppe koordiniert waren.  
[Zeilenumbruch]
- (13) (...)  
Anfang März sollen sechs Castor-Behälter per Bahn ins Zwischenlager Gorleben rollen. Drei stammen aus dem Atomkraftwerk Neckarwestheim (Baden-Württemberg), einer aus dem Reaktor Gundremmingen (Bayern) und zwei Atom-  
müllbehälter aus der französischen Wiederaufbereitungsanlage La Hague. Rund 19000 Polizisten und Bundesgrenzschutzbeamte werden Ende dieser Woche den Transport begleiten. Nach bisherigen Informationen sollen die Behälter am 28. Februar in Walheim nahe Neckarwestheim per Bahn eintreffen. Von dort soll der rund 700 Tonnen schwere Zug auf dem kürzesten Weg durch Baden-Württemberg, Hessen und Niedersachsen nach Gorleben gehen. Der Zug soll in Walheim am kommenden Sonntag abends abfahren und am Montag an der Verladestation Dannenberg eintreffen.
- Die Parteispitze und Fraktionsmitglieder von Bündnis 90/Die Grünen in Bonn und Niedersachsen kündigten am Dienstag an, sich an den Protesten zu beteiligen. Bundesumweltministerin Angela Merkel will sich heute in Lüchow mit Vertretern der Atomkraftgegner im Gorleben-Landkreis Lüchow-Dannenberg treffen.
- Die Gewerkschaft der Polizei (GdP) bezeichnete die Transporte als unsinnig. "Die Atomlobby hat seit 20 Jahren ihre Hausaufgaben nicht gemacht, um für den enormen giftigen Dreck, den sie produziert, Vorsorge zu treffen", sagte GdP-Vorstandsmitglied Eugen Roth. Die geschätzten Kosten von rund 100 Millionen Mark müsse nun der Steuerzahler tragen. (Tagesthema "Rückfall" und Region Seite 22)

## 6.2 Beispiel 2: „Heringe“<sup>248</sup>

- (1) Heringe, Clupeidae, Familie der Knochenfische (Ordnung der Heringsartigen) mit rund 180 Arten, die meist in tropischen Meeresgebieten leben.
- (2) Die bis 50 cm langen Heringe sind überwiegend Schwarmfische, die sich von Plankton ernähren.
- (3) Die wirtschaftlich wichtigste Gattung Clupea umfasst neben dem an Küsten des nördlichen Pazifik lebenden Pazifischen Hering (Clupea pallasii) den für Europa bedeutsamen Atlantischen Hering (Clupea harengus; mit graugrünem Rücken, Seiten und Bauch silbrig), der in planktonreichen Meeresgebieten riesige Schwärme bildet.
- (4) Er tritt in zahlreichen Gruppen auf, die sich in Individuengröße und Laichzeiten, -plätzen und -bedingungen unterscheiden.
- (5) So laicht z. B. der Herbsthering im Spätjahr vor der schottischen Nordseeküste als Blankhering, in der südlichen Nordsee als Down(s)hering und in schwedischen und norwegischen Gewässern als Bohuslänhering.

---

<sup>248</sup> „Heringe“. In: Brockhaus-Enzyklopädie in 30 Bänden. 21., völlig neu bearbeitete Auflage. Aktualisiert mit Artikeln aus der Brockhaus-Redaktion und ergänzt um Verweise auf Munzinger-  
Texte. o.J.

- (6) Der Fröhjahrshering laicht im Fröhjahr in den norwegischen Fjorden ab und führt anschließend Nahrungswanderungen bis zur Ostküste Islands durch, um im Spätjahr zur norwegischen Küste zurückzukehren.
- (7) Weiterhin unterscheidet man den bis 20 cm langen Strömling (Ostseehering), der in der Ostsee lebt.  
[Zeilenumbruch]
- (8) Ein Weibchen des Atlantischen Herings legt rund 20 000 bis 70 000 Eier ab, die auf den Meeresgrund sinken und dort an Steinen, Algen oder Ähnlichem festkleben.
- (9) Die schlüpfenden Larven steigen an die Oberfläche,
- (10) sie machen ihre Jugendentwicklung im Küstenbereich durch.
- (11) Erst die mit zwei bis drei Jahren etwa 20 cm langen Jungheringe wandern von der Küste ab.
- (12) Die Geschlechtsreife tritt im Alter von drei bis sieben Jahren ein;
- (13) die Lebensdauer beträgt etwa 20 Jahre.
- (14) Einige Heringsarten sind anadrom und ziehen zum Laichen in die Flüsse, z. B. die Alsen.  
[Zeilenumbruch]
- (15) (...)  
Wirtschaft: Heringe, Sardinen und die ihnen nahe verwandten Sardellenarten sind die wirtschaftlich wichtigsten Objekte der Meeresfischerei. Sie stehen als Planktonfresser an der Basis der Nahrungskette, was ihre hohe produktionsbiologische Leistung bedingt. Der Weltfangertrag des Atlantischen Herings betrug (2002) 1,87 Mio. t. Das schwarmbildende Verhalten der Heringsfische bedingt hohe Konzentrationen von Fischbiomasse im Meer auf engem Raum. Diese kann mithilfe moderner Echolottechnik geortet und mit Schwimmschleppnetzen oder Ringwaden bei relativ geringem Zeit- und Energieaufwand gezielt und nahezu vollständig abgefischt werden. Fänge bis zu 300 t pro Geräteinsatz sind möglich. Daraus ergibt sich die Gefahr einer Überfischung, gegen die sich die Bestände als anfällig erweisen. Hauptfanggebiete sind die Schelfmeere des Nordatlantiks.

Geschichtliches: Heringsfischerei wurde schon im 7. Jahrhundert in Europa betrieben. Um 900 war der gesalzene Hering bereits im Handel begehrt. Im 13. Jahrhundert entwickelte sich die Heringsfischerei besonders in der westlichen Ostsee bei der schwedischen Landschaft Schonen. Der Handel lag in den Händen der Hanse; das Salz für die Heringsalzerei kam aus den Bergwerken in Lüneburg.