

Universität Bielefeld  
Fakultät für Linguistik und Literaturwissenschaft  
Forschungsmethoden  
Dr. Wiltrud Mihatsch  
Sommersemester 2007

01.11.2007

---

# Die Verteilung anaphorischer Pronomen in der rhetorischen Struktur von Texten

---

Eine kleine Korpusstudie

von  
Daniel Jettka  
djettka@uni-bielefeld.de

---

# INHALT

1	Einleitung.....	1
2	Theoretische Grundlagen.....	2
2.1	Theorie der Pronomen.....	2
2.1.1	Anaphorische Pronomen vs. Deiktische Pronomen.....	3
2.1.2	Arten von Pronomen.....	4
2.2	Die rhetorische Struktur von Texten.....	5
3	Das Korpus.....	9
3.1	Die Texte.....	9
3.2	Annotationen.....	10
3.2.1	Annotation semantischer Relationen – CHS.....	10
3.2.2	Annotation der rhetorischen Struktur.....	12
4	Statistische Analyse.....	14
4.1	Vorgehen.....	14
4.2	Ergebnisse.....	14
5	Fazit.....	23
	Literaturverzeichnis.....	24

## 1 EINLEITUNG

Die Beschäftigung mit dem Thema der Verteilung anaphorischer Pronomen in der rhetorischen Struktur von Texten entstand aus meiner Arbeit in dem Projekt „Sekimo“ – „Sekundäre Informationsmodellierung und vergleichende Diskursanalyse“ – im Rahmen der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“. Hier bin ich damit beauftragt, Statistiken zu anaphorischen Beziehungen in einem Korpus von wissenschaftlichen Artikeln und Zeitungsartikeln zu erstellen. Das Korpus umfaßt u.a. vier Texte, für die Annotationen der anaphorischen Beziehungen und der rhetorischen Struktur<sup>1</sup> vorliegen. Mehr Informationen zu den Annotationen bietet der Abschnitt 3.2.

Zunächst ging es im Zusammenhang mit Pronomen nur darum, eine erste Übersicht zur Verteilung von Demonstrativpronomen in der rhetorischen Struktur aufzustellen. Hier interessierten mich aber die Verteilungen der anderen Pronomenarten, also habe ich die vorliegende Studie begonnen, um zu sehen, ob sich die eingehende Beschäftigung mit diesem Gebiet, auch im Hinblick auf meine Masterarbeit, lohnen könnte. Die Idee besteht darin, aus der Verbindung von semantischer Struktur (anaphorischen Beziehungen) und rhetorischer Struktur mögliche Hinweise für automatische Verfahren der pronominalen Anapherresolution zu gewinnen.

Im zweiten Kapitel stelle ich zunächst die theoretischen Grundlagen zu Pronomen (im Speziellen: anaphorischen Pronomen) und der rhetorischen Struktur von Texten vor. Das Kapitel 3 befaßt sich nachfolgend mit der Vorstellung des kleinen Korpus, das der Studie als Informationsgrundlage dient. Hier geht es um die Texte und deren Annotationen. Zur statistischen Analyse komme ich im vierten Kapitel. Nachdem das generelle Vorgehen der Analyse erläutert wurde, gebe ich einen Überblick über die Ergebnisse zu den einzelnen Pronomenarten. Im letzten Abschnitt werde ich ein kurzes Fazit zu der kleinen Korpusstudie liefern.

---

1 Die Annotationen der rhetorischen Struktur entstammen dem Gießener Projekt „Generische Dokumentstrukturen in linear organisierten Texten“ der DFG-Forschergruppe 437

## 2 THEORETISCHE GRUNDLAGEN

### 2.1 Theorie der Pronomen

Die Beschäftigung mit Pronomen hat eine lange Tradition. Die grundlegenden Erkenntnisse zu den Eigenschaften der Pronomen gehen zurück auf Grammatiker der griechischen Antike. Insbesondere Apollonios Dyskolos hat in seinem Werk „Über das Pronomen“ wichtige Grundlagen zum heutigen Verständnis der Pronomen geliefert. Apollonios Definition der Pronomen ist in zwei Fassungen überliefert. In der ersten Fassung sagt er zu den Eigenschaften von Pronomen:

„Das Pronomen ist also wie folgt zu definieren: Ein Wort, das an Stelle eines Nomens (steht und) identifizierte Referenten anzeigt; unregelmäßig in Kasus und Numerus, wenn es in der Wortform auch das Genus nicht deutlich bezeichnet“ (Übersetzung durch Brandenburg, 2005: 178).

Vor allem der erste Teil dieser Definition ist für die Studie von Bedeutung. Pronomen haben demnach eine Stellvertreterfunktion für Nomen, die wiederum identifizierte Referenten anzeigen.

Die enge Verbindung zwischen Nomen und Pronomen in Form einer Stellvertreterfunktion ließe sich nach Kim (2003) mit der morphologischen Entwicklung der Pronomen erklären. Demnach könnten „Pronomen als eine Verschmelzung aus einem Artikel und einem Nomen“ (Kim, 2003: 163 in Anlehnung an Hoang, 1982: 384, 1983: 6), als ein sog. Portmanteau-Wort im engeren Sinne, angesehen werden. „Beispielsweise ist das Pronomen im Deutschen 'er' eine Verschmelzung aus 'der Mann', 'sie' aus 'die Frau', 'jemand' aus 'ein Mensch' usw.“ (ebd.).

Nach Kim (2003: 162 in Anlehnung an Hoang, 1983: 6) muss eine Wortverschmelzung drei Kriterien erfüllen, um als Portmanteau-Wort im engeren Sinne bezeichnet werden zu können:

1. die Ausgangswörter kommen aus verschiedenen Wortklassen
2. das Ergebnis gehört keiner dieser Klassen an
3. als Ausgangswörter fungieren höchstens ein lexikalisches und mindestens ein grammatisches Wort

Weitere Beispiele für Portmanteau-Wörter sind 'an', 'am', 'zum', usw.<sup>2</sup> Ein Beispiel für ein Portmanteau-Wort im weiteren Sinne wäre 'fog'. Es handelt sich hierbei um eine Wortverschmelzung aus 'smoke' und 'fog'; diese erfüllt die oben genannten drei Bedingungen jedoch nicht.

In der späteren Fassung seiner Definition spezifiziert Apollonios die Art von Referenz zwischen dem Pronomen und dem Referenzierten:

„Demnach ist ein Pronomen das, was mit Deixis oder Anaphorik an die Stelle der Nomina tritt und bei dem kein Artikel steht“ (Übersetzung durch Brandenburg, 2005: 178).

Hier tritt eine wichtige Unterscheidung nach anaphorischer und deiktischer Verwendungsweise von Pronomen zu Tage. Da es in dieser Studie explizit um anaphorische Pronomen geht, werde ich diese im nächsten Abschnitt in Abgrenzung zu den deiktischen Pronomen vorstellen.

Die Definition bietet noch eine zweite wichtige Beobachtung: bei Pronomen stehen keine Artikel. Diese muss jedoch noch insofern erweitert werden, als daß Pronomen im eigentlichen Sinne nicht als Artikel fungieren können. In der herkömmlichen Grammatik wird das erste Wort in einer Nominalphrase wie „dieser Tisch“ häufig mit dem Begriff Demonstrativpronomen identifiziert. Genau genommen müßte es jedoch als Demonstrativartikel bezeichnet werden (vgl. Kim, 2003: 163). In der vorliegenden Studie werden daher auch nur Pronomen im oben genannten eigentlichen Sinne berücksichtigt. Als anaphorische Pronomen können folglich nur alleinstehende Pronomen, die für sich gesehen eine vollständige Nominalphrase darstellen, fungieren.

### 2.1.1 Anaphorische Pronomen vs. Deiktische Pronomen

Im Hinblick auf die Unterscheidung anaphorischer und deiktischer Pronomen ist es zunächst sinnvoll, sich den generellen Unterschied zwischen Anaphora und Deixis zu verdeutlichen.

---

2 Die Portmanteau-Wörter bestehend aus einer Präposition und einem Artikel (Bsp.: 'an') werden in der späteren Analyse übrigens tatsächlich als zwei getrennte Worte mit den Eigenschaften der Ausgangswörter behandelt.

Nach Consten (2004) ist Deixis „das unmittelbare (erste) Hinweisen auf das in der Aussenwelt wirklich gegenwärtige oder als gegenwärtig zu denkende Object“ (Consten, 2004: 5 in Anlehnung an Windisch, 1869b: 394). Mit Anaphora sei „dagegen die Rückbeziehung auf ein in der Rede schon genanntes Object“ (ebd.) gemeint. Ein solches in der Rede schon genanntes Objekt wird als Antezedens bezeichnet. Die anaphorischen Pronomen unterliegen einer syntaktischen Beschränkung:

„Pronomen können in allen drei Personen deiktisch verwendet werden, nur in der dritten Person kommt die Möglichkeit einer anaphorischen Verwendung hinzu“ (Brandenburg, 2005: 184)

Dieser Umstand erleichtert im Allgemeinen die Identifizierung anaphorischer Pronomen. Trotzdem muss natürlich bei Pronomen der 3. Person entschieden werden, ob sie in deiktischer oder anaphorischer Verwendung vorliegen. Auch die Art der anaphorischen Verwendung kann hierbei unterschiedlich sein. Weitere Informationen zu verschiedenen Arten von Anaphern bietet das Kapitel 3.2.1.

### 2.1.2 Arten von Pronomen

Im „Metzler Lexikon Sprache“ werden unter syntaktisch-semantischen Gesichtspunkten insgesamt sieben verschiedene Subklassen von Pronomen angenommen (vgl. Glück, 2000: 552f.). Diese decken sich mit den in dieser Studie untersuchten Pronomenarten, für die im Folgenden Beispiele angeführt werden. Dabei ist zu beachten, daß aus Sicht der Anaphernresolution zunächst einmal nur diejenigen Pronomen in der 3. Person interessant sind. Der Grund hierfür ist, wie oben bereits erwähnt, daß die 1. und 2. Person nur eine deiktische Verwendungsweise erlauben. Etwas komplizierter ist es bei den Pronomen der 3. Person. Hier muss im Einzelfall geklärt werden, ob deiktischer oder anaphorischer Gebrauch vorliegt. Dieses Problem tritt vornehmlich in der gesprochenen Sprache auf. Es wird in dieser Studie keine Rolle spielen, da hier nur wissenschaftliche Texte und Zeitungsartikel untersucht werden. Es kann an dieser Stelle davon ausgegangen werden, daß keine deiktische Verwendung vorhanden ist. Die nachstehende Übersicht beinhaltet einige Beispiele zu den verschiedenen Pronomenarten:

Personalpronomen:	ich, du, er, sie, es, wir, ihr, sie
Demonstrativpronomen:	der, dieser, jener, derjenige, derselbe
Relativpronomen:	der, die, das, welcher
Reflexivpronomen:	mich, dich, sich, uns, euch
Possessivpronomen:	mein, dein, sein, unser, euer, ihr
Indefinitpronomen:	jemand, alle, einer, keiner, mancher, man, wer, beide
Interrogativpronomen:	wer?, wessen?, wem?, wen?, was?, welcher?

## 2.2 Die rhetorische Struktur von Texten

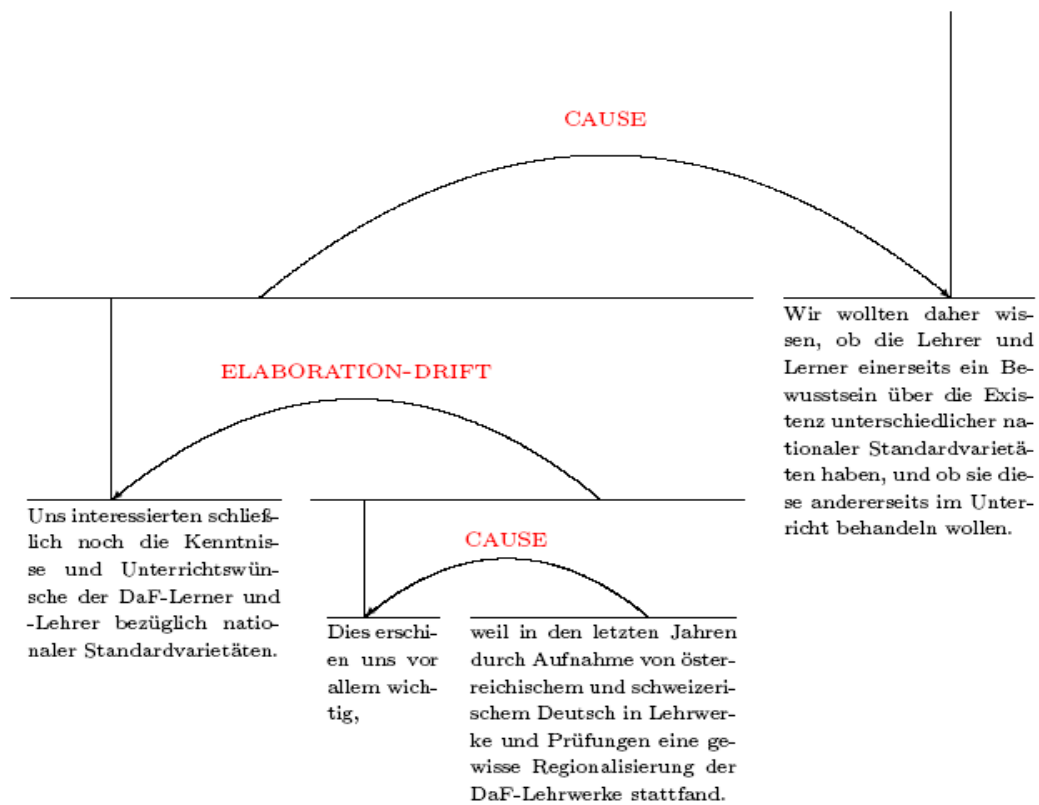
Im Folgenden werde ich kurz auf das Konzept der rhetorischen Struktur von Texten im Sinne der Rhetorical Structure Theory (RST) eingehen. Dabei kann ich in diesem Rahmen nur eine grundlegende Einführung geben<sup>3</sup>. Mann/Thompson (1988) fassen den Zweck der RST wie folgt zusammen:

„Rhetorical Structure Theory is a descriptive theory of a major aspect of the organization of natural text. It is a linguistically useful method for describing natural texts, characterizing their structure primarily in terms of relations that hold between parts of the text“ (Mann/Thompson, 1988: 243).

Die RST bietet die theoretischen und praktischen Hintergründe für die Erfassung der rhetorischen Struktur von Texten. Sicherlich läßt sich dieses am anschaulichsten an einem Beispiel verdeutlichen. Die Abbildung zeigt einen Ausschnitt eines RST-Baums für ein kurzes Textsegment:

---

<sup>3</sup> Eine umfassende Einführung in die RST bieten Mann/Thompson (1988), sowie die RST-Website <http://www.sfu.ca/rst/>



In dieser Abbildung sind die grundlegenden Konzepte der RST nachvollziehbar:

1. Segmentierung von Texten (i.A. in Sätze)
2. Abbildung der hierarchischen Struktur von Texten
3. Identifizierung rhetorischer Relationen zwischen Textteilen

Diese einzelnen Schritte hin zu einer Abbildung der rhetorischen Struktur eines Texts sind auf verschiedene Weise realisierbar. Generell sind ein manuelles Vorgehen und die automatische Verarbeitung denkbar. Für die manuelle Anwendung der RST auf einen beliebigen Text steht beispielsweise das Werkzeug RSTTool von Michael O'Donnell zur Verfügung<sup>4</sup>. Auf der anderen Seite gibt es aber durchaus vielversprechende automatische Anwendungen der RST. Zu beachten ist hier z.B. der „cue-phrase-based rhetorical parser“ von Marcu (2000), der grundlegende Möglichkeiten zur automatischen Erschließung der rhetorischen Struktur von Texten aufzeigt.

Die Abbildung der hierarchischen Struktur in der RST erfolgt durch zwei verschiedene Arten rhetorischer Relationen, die generell zwei nicht-überlappende Textsegmente (Nukleus und Sa-

<sup>4</sup> siehe <http://www.wagsoft.com/RSTTool/>



tellit) miteinander verbinden (vgl. Mann/Thompson, 1988: 245). Die Relationen sind aufgeteilt in sogenannte „Presentational Relations“, mit pragmatischem Hintergrund, und „Subject Matter Relations“, mit semantischem Hintergrund (vgl. Mann/Thompson, 1988: 257). In der obigen Abbildung sind drei Beispiele für „Subject Matter Relations“ zu sehen. Diese Nukleus-Satellit-Relationen können als subordinierende Relationen angesehen werden, da ein Segment, nämlich der Satellit, in seiner Bedeutung direkt vom Nukleus abhängig ist. Dieser Zusammenhang wird z.B. bei der *Cause*-Relation in der obigen Abbildung deutlich. Hier ist der Satz „Dies erschien uns vor allem wichtig“ als Nukleus mit dem von ihm abhängigen Satellit, der Begründung „weil in den letzten Jahren ...“, verbunden.

Demgegenüber gibt es aber auch multinukleare Relationen, die zwei gleichwertige Textsegmente miteinander verbinden. Sie wirken koordinierend. Als Beispiele hierfür wären Relationen wie *Conjunction*, *List* oder *Sequence* zu nennen. Eine Übersicht der in der RST verwendeten grundlegenden rhetorischen Relationen ist im Folgenden dargestellt:

### **Nukleus-Satellit-Relationen**

#### **Presentational Relations**

- Antithesis
- Background
- Concession
- Enablement
- Evidence
- Justify
- Motivation
- Preparation
- Restatement
- Summary

#### **Subject Matter Relations**

- Circumstance
- Condition
- Elaboration
- Evaluation
- Interpretation
- Means
- Non-volitional Cause
- Non-volitional Result
- Otherwise
- Purpose
- Solutionhood
- Unconditional
- Unless
- Volitional Cause
- Volitional Result

### **Multinukleare Relationen**

- Conjunction
- Contrast
- Disjunction
- Joint
- List
- Multinuclear Restatement
- Sequence

Die rhetorischen Relationen, die für die Annotation der vorliegenden vier Texte verwendet wurden, decken sich größtenteils mit denen der RST. Zum Teil wurden die grundlegenden Relationen aber auch erweitert, dieses gilt insbesondere für die Relation *Elaboration*. An den Stellen der statistischen Analyse, an denen diese Abweichungen eine Rolle spielen, werde ich hierauf allerdings explizit eingehen.

Im nächsten Abschnitt werde ich kurz das verwendete Korpus vorstellen. Danach werde ich zur statistischen Analyse der Verteilung anaphorischer Pronomen in der rhetorischen Struktur von Texten übergehen. Es geht hierbei darum, zu prüfen, inwieweit die verschiedenen Pronomenarten auf charakteristische Art und Weise in die rhetorische Struktur eingebettet sind. Die Ergebnisse könnten Hinweise zu Eigenschaften von Pronomen liefern, die speziell für die Anaphernresolution von Bedeutung sein könnten. So wäre es möglich Hypothesen in Bezug auf Suchstrategien zu generieren. Diese Hypothesen könnten folgendermaßen lauten: Für eine bestimmte Art anaphorischen Pronomens gilt, daß die Textsegmente, in denen es selbst steht, und diejenigen in denen sein Antezedens steht, besonders häufig durch eine bestimmte Art von rhetorischer Relation miteinander verbunden sind. Inwiefern solche Aussagen tatsächlich getroffen werden können, wird sich im Abschnitt 4.2 über die statistischen Analysen zu den einzelnen Pronomenarten zeigen.

## 3 DAS KORPUS

### 3.1 Die Texte

Für die Studie wird ein Korpus mit 4 Texten verwendet. Diese Texte liegen u.a. in den für den Untersuchungsgegenstand wichtigen Annotationen von CHS- und RST-Struktur vor (vgl. Abschnitte 3.2.1 und 3.2.2). Die beiden genannten Annotationen werden genutzt, um die semantischen Relationen (anaphorischen Beziehungen) mit der rhetorischen Struktur (nach der Rhetorical Structure Theory – RST) in Beziehung zu setzen. Die folgende Tabelle bietet einen Überblick über die Texte und ihren Umfang. Bei token und de handelt es sich um die Angabe der Instanzen von Token bzw. Diskurseinheiten in den einzelnen Texten. Die weiteren Angaben beziehen sich auf die Vorkommen der verschiedenen Arten von anaphorisch genutzten Pronomen.

Text	token	de	Pers	Dem	Rel	Refl	Poss	Indef	Interr
ling-deu-003	10048	2606	37	20	97	-	-	1	-
ling-deu-010	1895	501	4	-	6	-	1	-	1
Luckhardt-1995a	2542	698	18	4	28	6	-	4	-
zeit-15-2005	1226	369	27	4	5	-	-	-	-
alle Texte	15711	4174	86	28	136	6	1	5	1

Drei der vier Texte sind wissenschaftliche Artikel. Lediglich der Text zeit-15-2005 ist ein Zeitungsartikel. Der Texttyp scheint allerdings keine große Rolle für die Vorkommenshäufigkeit von Pronomen zu spielen. Der einzige auffällige Wert ist in diesem Zusammenhang die Anzahl der Personalpronomen im Zeitungsartikel. Während im Gesamtdurchschnitt der Texte Personalpronomen 0,6% aller Token ausmachen, sind es 2,2% im Zeitungsartikel. Diese Beobachtung könnte ein Hinweis auf verschiedenartige Verwendung von Personalpronomen in Zeitungs- und wissenschaftlichen Artikeln sein. Diesen Zusammenhang werde ich an dieser Stelle (mangels weiteren Materials) jedoch nicht weiterverfolgen.

Im nächsten Abschnitt gehe ich auf die Vorverarbeitung der Texte durch Annotationen ein. Mit deren Hilfe werden die Texte durch wichtige Informationen zu ihrer rhetorischen und semantischen Struktur angereichert. Auf Basis der Annotationen erfolgt im späteren Verlauf der Studie die statistische Analyse der Verteilung von Pronomen in der rhetorischen Struktur.

## 3.2 Annotationen

Die oben erwähnten Texte liegen in zwei verschiedenen Annotationen vor. Die Annotation der sogenannten CHS-Ebene (kohäsive Struktur) stammt aus dem Projekt „Sekimo“<sup>5</sup>. Hier sind die Annotationen der anaphorischen Beziehungen enthalten. Desweiteren stehen Annotationen der rhetorischen Struktur aus dem Projekt „SemDok“<sup>6</sup> zur Verfügung. Diese beiden Informationsquellen für die Analyse der Verteilung anaphorischer Pronomen in der rhetorischen Struktur werden im Folgenden vorgestellt.

### 3.2.1 Annotation semantischer Relationen – CHS

Der Annotation semantischer Relationen, im Speziellen der anaphorischen Beziehungen, gehen Verarbeitungsschritte voraus, die die Grundlage für die Identifizierung von Anapher-Antezedens-Paare darstellen. Es handelt sich hierbei insbesondere um Part-of-speech Tagging sowie die Annotation von Diskursentitäten, zwischen denen später anaphorische Relationen identifiziert werden. Das Ergebnis der ersten Annotationsschritte in Form von XML-Auszeichnungen sieht folgendermaßen aus:

---

5 Projekt A2 „Sekundäre Informationsstrukturierung und vergleichende Diskursanalyse“ der DFG-Forschergruppe 437 „Texttechnologische Informationsmodellierung“,

Website: <http://www.text-technology.de/Sekimo>

6 Projekt C1 „Generische Dokumentstrukturen in linear organisierten Texten“ der DFG-Forschergruppe 437, Website: <http://www.text-technology.de/SemDok>

```

1 <chs:para skip="no">
2   <chs:sentence id="w2077" auto="no">
3     <chs:de deID="de570" headRef="w2078" deType="nom" deAuto="1" deIsAna="yes"
4       deIsAnte="no">
5       <chs:token id="w2078" text="Dieser" dependHead="w2079" pos="PRON" syntax="
6         QNH" heur="no"
7         lemma="dieser"
8         dependValue="subj"
9         morpho="Dem_MSC_SG_NOM"
10        auto="yes"
11        cmx-output="correct">Dieser</chs:token>
12    </chs:de>
13    <chs:token id="w2079" text="wird" dependHead="w2091" pos="V" syntax="@AUX"
14      heur="no"
15      lemma="werden"
16      dependValue="v-ch"
17      morpho="IND_PRESENT_SG_P3"
18      auto="yes"
19      cmx-output="correct">wird</chs:token>
20    ...
21  </chs:sentence>
22 </chs:para>

```

Die einzelnen Wörter (Token) wurden automatisch als `chs:token` annotiert und dabei mit morpho-syntaktischen Eigenschaften versehen. Auf die Einzelheiten der Annotation werde ich in diesem Zusammenhang nicht eingehen. Desweiteren gibt es `chs:de`-Elemente, die die Diskursentitäten anzeigen. In diesem Fall wurden nur Diskursentitäten, die als Nominalphrasen angesehen werden können, beachtet. Es wäre durchaus denkbar, größere syntaktische Einheiten (z.B. Sätze) oder Elemente der logischen Dokumentstruktur (z.B. Paragraphen) als Diskursentitäten zu annotieren<sup>7</sup>. Hier ging es aber zunächst nur um die besagten Nominalphrasen.

Die Informationen zu anaphorischen Beziehungen wurden manuell annotiert. Die gewonnenen Daten hat man mittels Standoff-Annotationen hinzugefügt. Das Ergebnis sieht z.B. folgendermaßen aus:

```

1 <chs:standoff>
2   <chs:semRel>
3     <chs:cospecLink relType="ident" phorIDRef="de4" antecedentIDRefs="de3"/>
4     <chs:bridgingLink relType="hasMember" phorIDRef="de12" antecedentIDRefs="de8,
5       de10"/>
6     <chs:bridgingLink relType="poss" phorIDRef="de14" antecedentIDRefs="de12"/>
7     ...
8   </chs:semRel>
9 </chs:standoff>

```

<sup>7</sup> Anaphern können sich nicht nur auf Nominalphrasen, sondern z.B. auch auf ganze Sätze beziehen. Bsp.: 'Der Mann ist hingefallen. **Das** hatte einen guten Grund.'

Hier sind die ersten drei festgestellten anaphorischen Relationen innerhalb des Textes „zeit-15-2005“ dargestellt. Jede Relation ist mit vier grundlegenden Informationen versehen. Zunächst haben alle die Angabe der Anapher und des/der Antezedens/Antezedentien gemeinsam. Diese erfolgt über die Angabe einer ID, mit deren Hilfe Diskursentitäten eindeutig referenziert werden. Darüberhinaus wird zwischen Kospezifikations- und Bridging-Relationen unterschieden. Dieses schlägt sich im jeweiligen Elementnamen nieder: `cospecLink` vs. `bridgingLink`. Kospezifikation ist hierbei so definiert, „dass ein Ausdruck einen Vorangegangenen sprachlich wieder aufgreift“ (Goecke et al., 2007: 3). Bridging bzw. „Indirekte Anaphorik besteht dann, wenn das Antezedens zu einem anaphorischen Element nicht explizit im Text genannt ist, sondern aus dem Kontext erschlossen werden muss“ (Goecke et al., 2007: 14). Außerdem werden auch Possessivrelationen (erkennbar an verwendetem Possessivpronomen oder Genitiv-NP) als `bridgingLink` annotiert.

Bridging und Kospezifikation sind als die primären, anaphorischen Relationen anzusehen. Darüberhinaus werden diesen weitere sekundäre Relationstypen zugeordnet, die in der folgenden Übersicht aufgelistet sind:

<b>cospecLink</b>	ident	<b>bridgingLink</b>	poss
	propName		setMember
	namedEntity		hasMember
	synonym		meronym
	hyperonym		holonym
	hyponym		bridging
	paraphrase		
	addInfo		
	isA		

Ich werde an dieser Stelle nicht auf jede einzelne der sekundären Relationen eingehen. Vielmehr beschränke ich mich auf deren Nennung und komme ggfs. im Zusammenhang mit der statistischen Analyse auf einzelne von ihnen zurück.

### 3.2.2 Annotation der rhetorischen Struktur

Die zweite wichtige Informationsquelle für die spätere Analyse ist die Annotation der rhetorischen Struktur der vier Texte. Sie steht ebenfalls im XML-Format zur Verfügung<sup>8</sup>. Nachfolgend

<sup>8</sup> Die Annotationen der rhetorischen Struktur entstammen dem Gießener C1-Projekt der DFG Forschergruppe „Texttechnologische Informationsmodellierung“.

ist ein Beispielausschnitt der Annotation aufgeführt:

```

1 <hypo relname="consequence-s" id="i188">
2   <n>
3     <hypo relname="result" id="i190">
4       <s>
5         <para relname="list-coordination" id="i93">
6           <n>
7             <t id="i4">Maik hat kein eigenes Fahrrad,</t>
8           </n>
9           <n>
10            <t id="i6">und Marie* ährt nicht in den Urlaub.</t>
11          </n>
12        </para>
13      </s>
14      <n>
15        <hypo relname="preparation-other" id="i98">
16          <s>
17            <t id="i8">Zwei Kinder, eine Gemeinsamkeit:</t>
18          </s>
19          <n>
20            <t id="i9">Ihre Eltern haben nur halb so viel Geld im Monat wie
21              ein deutscher üDurchschnittsbrger.</t>
22          </n>
23        </hypo>
24      </n>
25    </hypo>
26  </n>
27  <s>
28    <hypo relname="background" id="i99">
29      <n>
30        <t id="i10">Damit gelten sie als "arm"</t>
31      </n>
32      <s>
33        <t id="i7">- zwei von rund 1,5 Millionen Kindern in Deutschland, die
34          gemeint sind, wenn von Kinderarmut gesprochen wird.</t>
35      </s>
36    </hypo>
37  </s>
38 </hypo>

```

Die Elemente `hypo` und `para` stehen hierbei für rhetorische Relationen, im Speziellen für hypotaktische (subordinierende) und parataktische (koordinierende) Relationen (vgl. Abschnitt 2.2). Diese Elemente enthalten je nach Relationstyp, der im Attribut `relname` angegeben ist, `n`- und `s`-Elemente. Diese stehen für Nuklei und Satelliten. Darin wiederum sind entweder weitere Relationen, die zwischen den enthaltenen Textsegmenten gelten, oder aber direkt einzelne Textsegmente beinhaltet. Auf diese Weise kann die Architektur der RST auf das XML-Format abgebildet werden. Mit Hilfe der nun zur Verfügung stehenden Informationen können die rhetorische Struktur und die anaphorischen Beziehungen miteinander in Beziehung gesetzt werden. Dieses geschieht im nächsten Schritt, der statistischen Analyse.

## 4 STATISTISCHE ANALYSE

### 4.1 Vorgehen

Aufgrund der XML-Architektur des Korpus bietet es sich für die statistische Analyse der vorhandenen Daten an, die XML-basierte Programmiersprache XSLT (Extensible Stylesheet Language for Transformations) zu verwenden. Eine umfassende Einführung in XSLT bietet Kay (2004).

Im Weiteren werde ich nicht auf die Einzelheiten der Programmierung eingehen. Stattdessen stelle ich das generelle Prinzip der statistischen Analyse dar. Die XML-Annotationen der vier Texte des Korpus können mit Hilfe von XSLT gleichzeitig analysiert werden. Hierzu werden die einzelnen Dateien in das Stylesheet eingebunden und sind so in jeder Phase der Analyse verfügbar. Auf diese Weise entstanden die einzelnen Statistiken (in Form von HTML-Tabellen) zu den verschiedenen Arten von Pronomen, die im nächsten Abschnitt dargestellt sind.

Zum Zweck der Analyse werden zunächst für jede Art von Pronomen alle anaphorischen Relationen gesucht, bei denen das Pronomen als Anapher fungiert. Dieser erste Analyseschritt erfolgt lediglich auf Basis der Annotation der CHS-Ebene. Nachfolgend kommen die Informationen der rhetorischen Struktur hinzu. Hier wird für die gesammelten semantischen Relationen mit pronominaler Anapher des jeweiligen Pronomentyps ermittelt, in welchen Textsegmenten die Anapher und das zugehörige Antezedens stehen. Der nächste Schritt besteht darin zu ermitteln, welche Art von rhetorischer Relation zwischen den beiden Textsegmenten herrscht. Sollten sich für die verschiedenen Pronomenarten unterschiedliche charakteristische Verteilungen von Anapher und Antezedens ergeben, könnten diese einen Beitrag zur pronominalen Anapherresolution leisten.

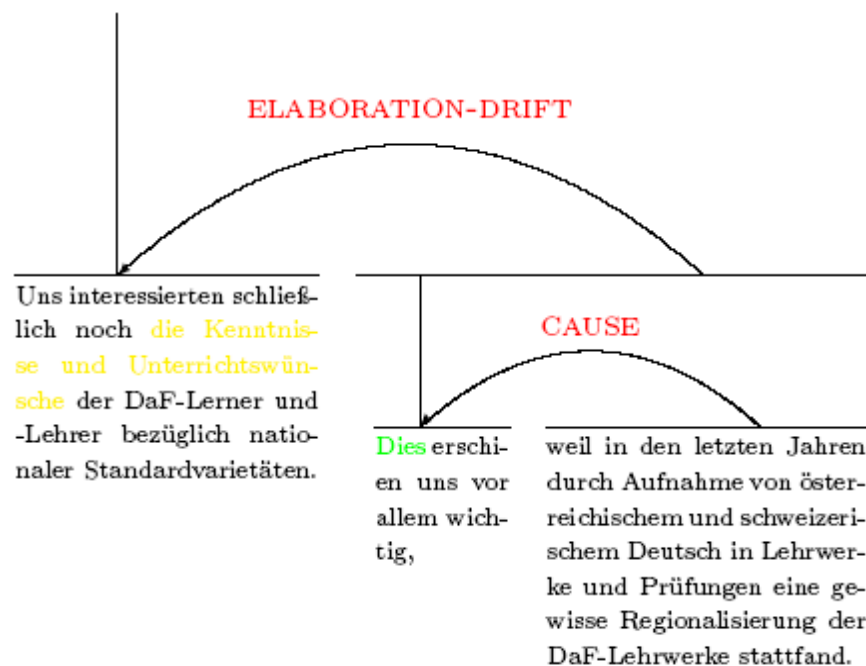
### 4.2 Ergebnisse

In diesem Abschnitt werden die Ergebnisse der statistischen Analyse zu jeder einzelnen untersuchten Pronomenart dargestellt. Die Ergebnisse beziehen sich auf alle vier zugrundeliegenden Texte gemeinsam. Da es sich bei den Texten lediglich um drei wissenschaftliche Artikel und einen Zeitungsartikel handelt, besteht an dieser Stelle nicht die Absicht etwas zu Unterschie-



den in verschiedenen Texttypen herauszufinden.

In den Tabellen sind spaltenweise die verschiedenen Arten von anaphorischen Beziehungen, in denen die Anapher-Antezedens-Paare mit pronominaler Anapher bestimmten Typs stehen, abgetragen. In den Zeilen stehen die unterschiedlichen Arten von rhetorischen Relationen, die zwischen den Textsegmenten gelten. Es handelt sich hierbei um die Textsegmente, die die Anapher und das Antezedens der jeweiligen anaphorischen Beziehung enthalten. In diesem Zusammenhang ist es wichtig zu beachten, daß die Textsegmente nicht immer direkt durch eine rhetorische Relation miteinander verbunden sind. Vielmehr kann es vorkommen, daß die Segmente wiederum in eine tiefere Struktur eingebettet sind. Dieser Sachverhalt schlägt sich in der Angabe der Merkmale `anaEmbed` und `anteEmbed` nieder. Diese geben die durchschnittliche Einbettungstiefe der Textsegmente von Anapher (`anaEmbed`) und Antezedens (`anteEmbed`) an. Anhand der folgenden Abbildung kann dieses nachvollzogen werden:



Die rhetorische Relation, die zwischen den Textsegmenten mit Anapher (grün) und Antezedens (gelb) gilt, kann als *Elaboration-drift* benannt werden. Allerdings ist die Anapher in eine tiefere rhetorische Struktur (*Cause*-Relation) eingebunden. Deshalb bekommt sie für `anaEmbed` den Wert 1. Je mehr rhetorische Relationen zwischen der ersten gemeinsamen Relation von Anapher- und Antezedens-Textsegment liegen, desto größer wird der Wert für

anaEmbed und anteEmbed. Insgesamt ist in diesem Kontext zu sehen, daß die Anapher im Allgemeinen in eine tiefere Struktur eingebettet ist, als das Antezedens. Dieses würde mit Beobachtungen von Marcu (2000) übereinstimmen. Er postuliert für sein System des automatischen rhetorischen Parsens von Texten, daß „the best discourse trees are often those that are skewed to the right“ (Marcu, 2000: 137). Das bedeutet, daß die Struktur rhetorischer Relationen in RST-Bäumen eher rechtslastig ist. Marcu erklärt sich diesen Zusammenhang wie folgt: „I believe that the explanation of this observation is that text processing is, essentially, a left-to-right process“ (ebd.).

Im Allgemeinen ist zu der Verteilung der anaphorischen Pronomen auf die verschiedene Arten von anaphorischen Relationen zu bemerken, daß sich der absolute Großteil von 98,5% in einer Kospezifikationsbeziehung mit dem sekundären Merkmal *ident* befindet. Dieses ist nicht verwunderlich, schaut man sich die Definition dieser anaphorischen Beziehung an.

Eine Relation der Kospezifikation „beruht auf einer phorischen Beziehung zwischen sprachlichen Ausdrücken, wobei der phorische Ausdruck ein Pronomen, eine NP oder ein Eigename sein kann“ (Goecke et al., 2007: 16). Desweiteren wird „der Wert *ident* [...] vergeben, wenn sich ein Pronomen auf eine NP bezieht“ (ebd.). Die vier übrigen anaphorischen Pronomen stehen in einer Bridging-Relation mit dem Relationstyp *setMember*. Hierbei gilt: „Der phorische Ausdruck [also das anaphorische Pronomen] ist Element einer durch die Bezugsgröße beschriebenen Menge“ (Goecke et al., 2007: 28). Ein Beispiel ist durch das folgende Anapher-Antezedens-Paar aus dem Korpus gegeben:

#### **Die Ergebnisse – die (der Schüler)**

In der statistischen Analyse des Korpus wurde das „die“ als anaphorisches Demonstrativpronomen erkannt. Es ist allerdings zu bemerken, daß die Annotation als Bridging-Relation *setMember* wahrscheinlich von der Semantik der komplexen Diskursentität „die der Schüler“ abhängt. In diesem Fall handelt es sich nicht um eine pronominale Nominalphrase nach der dieser Studie zugrundeliegenden Definition. Aufgrund dieser Beobachtung und der ohnehin sehr homogenen Verteilung der anaphorischen Pronomen auf semantische Relationen, werde ich mich im Folgenden auf die Betrachtung der Verteilung auf die rhetorischen Relationen konzentrieren. Den in den Annotationen vorhandenen rhetorischen Relationen habe ich zwei weitere hinzugefügt, die im Grunde gar keine Relationen sind. Der Typ *same* bezeichnet den

Fall, daß Anapher und zugehöriges Antezedens in dem selben Textsegment stehen. Es kann aber auch vorkommen, daß die Textsegmente der Anapher und des Antezedens überhaupt nicht durch eine rhetorische Relation miteinander verbunden sind. Für diesen Fall wurde der Typ *no* ergänzt.

Nachfolgend werden die Analyseergebnisse für jede einzelne Pronomenart aufgeführt und auf eventuell vorhandene charakteristische Verteilungen hin untersucht.

### Personalpronomen

Anaphoric Personal Pronouns		Semantic Relations		total
		cospecLink		
		ident		
Rhetoric Relations	background	2	anteEmbed: 1 anaEmbed: 1	2 anteEmbed: 1 anaEmbed: 1
	cause	2	anteEmbed: 0.5 anaEmbed: 0.5	2 anteEmbed: 0.5 anaEmbed: 0.5
	circumstance	3	anteEmbed: 0 anaEmbed: 0.67	3 anteEmbed: 0 anaEmbed: 0.67
	concession	6	anteEmbed: 0.67 anaEmbed: 1.17	6 anteEmbed: 0.67 anaEmbed: 1.17
	consequence-n	2	anteEmbed: 0.5 anaEmbed: 0	2 anteEmbed: 0.5 anaEmbed: 0
	consequence-s	1	anteEmbed: 2 anaEmbed: 1	1 anteEmbed: 2 anaEmbed: 1
	contrast-multi	8	anteEmbed: 0.13 anaEmbed: 0.5	8 anteEmbed: 0.13 anaEmbed: 0.5
	elaboration-continuation-other	15	anteEmbed: 0.47 anaEmbed: 0.8	15 anteEmbed: 0.47 anaEmbed: 0.8
	elaboration-definition	1	anteEmbed: 0 anaEmbed: 1	1 anteEmbed: 0 anaEmbed: 1
	elaboration-drift	4	anteEmbed: 0.5 anaEmbed: 1.5	4 anteEmbed: 0.5 anaEmbed: 1.5
	list-coordination	6	anteEmbed: 0.5 anaEmbed: 0.17	6 anteEmbed: 0.5 anaEmbed: 0.17
	means	2	anteEmbed: 0.5 anaEmbed: 0.5	2 anteEmbed: 0.5 anaEmbed: 0.5
	one-segment	2	anteEmbed: 0 anaEmbed: 0	2 anteEmbed: 0 anaEmbed: 0
	purpose-s	2	anteEmbed: 0 anaEmbed: 0	2 anteEmbed: 0 anaEmbed: 0
	no	1	anteEmbed: 0 anaEmbed: 0	1 anteEmbed: 0 anaEmbed: 0
	same	28	anteEmbed: 0 anaEmbed: 0	28 anteEmbed: 0 anaEmbed: 0
	sequence	1	anteEmbed: 1 anaEmbed: 0	1 anteEmbed: 1 anaEmbed: 0
	<b>total</b>		<b>86</b>	anteEmbed: 0.29 anaEmbed: 0.44

Eine Interpretation der Verteilung von anaphorischen Personalpronomen und ihrer Antezedentien zu treffen, erscheint aufgrund der nebenstehenden Statistik eher schwierig. Ungefähr ein Drittel der Anaphern befinden sich im gleichen Textsegment wie ihr Antezedens. Die restlichen Anapher-Antezedens-Paare verteilen sich recht gleichmäßig auf verschiedene Arten rhetorischer Relationen. Es erscheint mir mit Hilfe der vorliegenden Daten nicht möglich, Regularitäten zur Verteilung der anaphorischen Personalpronomen und ihrer Antezedentien in der rhetorischen Struktur zu erkennen. Das könnte natürlich dadurch bedingt sein, daß es keine derartigen regelhaften Vertei-

lungen gibt. Es scheint vielmehr so, als daß die rhetorische Struktur keine direkten Hinweise für die Zuordnung von anaphorischen Personalpronomen zum jeweiligen Antezedens liefern kann.

### Demonstrativpronomen

Anaphoric Demonstrative Pronouns		Semantic Relations		total
		cospecLink	bridgingLink	
		ident	setMember	
Rhetoric Relations	attribution	1 anteEmbed: 0 anaEmbed: 0	-	1 anteEmbed: 0 anaEmbed: 0
	elaboration-continuation-other	3 anteEmbed: 0.33 anaEmbed: 0.33	-	3 anteEmbed: 0.33 anaEmbed: 0.33
	elaboration-definition	1 anteEmbed: 0 anaEmbed: 1	-	1 anteEmbed: 0 anaEmbed: 1
	elaboration-derivation	1 anteEmbed: 0 anaEmbed: 3	2 anteEmbed: 0 anaEmbed: 1	3 anteEmbed: 0 anaEmbed: 1.67
	elaboration-drift	3 anteEmbed: 0 anaEmbed: 0.67	-	3 anteEmbed: 0 anaEmbed: 0.67
	elaboration-specification-other	2 anteEmbed: 0.5 anaEmbed: 0	-	2 anteEmbed: 0.5 anaEmbed: 0
	no	2 anteEmbed: 1 anaEmbed: 1	1 anteEmbed: 3 anaEmbed: 2	3 anteEmbed: 1.67 anaEmbed: 1.33
	same	12 anteEmbed: 0 anaEmbed: 0	-	12 anteEmbed: 0 anaEmbed: 0
<b>total</b>	<b>25</b> anteEmbed: 0.16 anaEmbed: 0.36	<b>3</b> anteEmbed: 1 anaEmbed: 1.33	<b>28</b> anteEmbed: 0.25 anaEmbed: 0.46	

Im Zusammenhang mit der Verteilung der anaphorischen Demonstrativpronomen und ihrer Antezedenten möchte ich zunächst auf die gefundenen Anaphern in Bridging-Relationen eingehen. Da hier nur drei Instanzen identifiziert wurden, erscheint es lohnenswert, diese genauer zu untersuchen<sup>9</sup>. Zwei dieser Anaphern sind im folgenden Satz gefunden

worden:

Die **Ergebnisse** sollen im Folgenden ausführlicher erläutert und kommentiert werden, zunächst **die** der Schüler, danach **die** der Lehrer.

Ich denke, daß die Bridging-Relation *setMember* für die Demonstrativpronomen an dieser Stelle falsch interpretiert wurde. Hierzu muß man wissen, daß die anaphorische Beziehung laut Annotation nur zwischen dem Demonstrativpronomen „die“ und „Die Ergebnisse“ gilt. Es sind nicht etwa die Konstruktionen „die der Schüler“ oder „die der Lehrer“, die als Anaphern fungieren. Würden diese komplexen Nominalphrasen als Anapher gelten, so wäre die Bridging-Relation *setMember* nachvollziehbar. Dann würde es sich aber auch nicht mehr um Demonstrativpronomen nach der hier geltenden Definition handeln. Somit ist davon auszuge-

<sup>9</sup> Die Instanz, bei der Anapher und Antezedens in der rhetorischen Struktur nicht miteinander verbunden sind (no) lasse ich hier außer Acht. Sie wurde ohnehin in der Annotation mit dem Kommentar „anapher?“ (d.h. sie ist nicht eindeutig) versehen

hen, daß bei den Demonstrativpronomen nur anaphorische Kospezifikationsbeziehungen vom Typ *ident* herrschen.

Zur Verteilung in der rhetorischen Struktur ist zu bemerken, daß ca. die Hälfte der Anaphern im gleichen Textsegment wie ihre Antezedentien anzutreffen sind. Der Großteil der restlichen Verteilungen beschränkt sich auf *elaboration*-Relationen. Dieser Umstand erschließt sich z.T., schaut man sich die Definition der Demonstrativpronomen an: „Subklasse der Pronomen mit der Funktion des ausdrücklichen Hinweisens“ (Glück, 2000: 141). Diese Definition kann sowohl für deiktischen als auch für anaphorischen Gebrauch gelten.

Im Allgemeinen kann man wohl davon ausgehen, daß die meisten anaphorischen Demonstrativpronomen entweder im gleichen Textsegment wie ihr Antezedens zu finden sind, oder daß Anapher und Antezedens in durch eine *elaboration*-Relation verbundenen Textsegmenten stehen.

### Relativpronomen

Anaphoric Relative Pronouns		Semantic Relations	total
		cospecLink	
		ident	
Rhetoric Relations	elaboration-continuation-other	1 anteEmbed: 1 anaEmbed: 1	1 anteEmbed: 1 anaEmbed: 1
	elaboration-specification-other	1 anteEmbed: 0 anaEmbed: 0	1 anteEmbed: 0 anaEmbed: 0
	same	132 anteEmbed: 0 anaEmbed: 0	132 anteEmbed: 0 anaEmbed: 0
	samesegment	2 anteEmbed: 0 anaEmbed: 0	2 anteEmbed: 0 anaEmbed: 0
total		136 anteEmbed: 0.01 anaEmbed: 0.01	136 anteEmbed: 0.01 anaEmbed: 0.01

Die Verteilung der Anapher- Antezedens-Paare mit Relativpronomen als Anapher ist recht eindeutig. 98,5% der Anaphern befinden sich in dem selben Textsegment wie ihr Antezedens. Diese Verteilung ist bedingt durch die Funktion von Relativpronomen, einen Relativsatz einzuleiten.

Dabei beziehen sich die Relativpronomen meist auf ein direkt vorausgehendes Substantiv oder eine Substantivgruppe (beide aus dem Hauptsatz) oder einen ganzen Satz (vgl. Glück, 2000: 575).

Lediglich zwei Anaphern befinden sich in Textsegmenten, die über eine *elaboration*-Relation mit dem Segment des Antezedens verbunden sind. Hierbei handelt es sich um die Folgenden:

Die Lehrer füllten einen im Vergleich zu **dem** für Lerner leicht modifizierten  
Fragebogen aus.

Wichtig erscheint uns ebenfalls die Einbindung **authentischer Gesprächssituationen** aus dem Alltag (Verkaufsdialoge in Geschäften, Gespräche mit Vermietern, Dienstleistungsdialoge mit Handwerkern, Institutionenkommunikation in dörflichen Gegenden, Gespräche in ländlichen Freizeitgruppen), in **denen** Laiensprecher ihre regionale Varietät in bedeutend authentischerem Maße sprechen als dies bisher in den Lehrwerken der Fall ist.

Im ersten Fall handelt es sich um eine falsche Annotation. Im zweiten um eine komplexe verschachtelte rhetorische *elaboration*-Relation, die Anapher und Antezedens. Für dieses Ergebnis ist allerdings eine Schwäche der statistischen Analyse verantwortlich, denn eigentlich stehen Anapher und Antezedens im gleichen Textsegment. Dieses wird jedoch durch ein eingebettetes Segment unterbrochen. Hierdurch kommt die Klassifizierung als *elaboration*-Relation zustande.

Daraus kann folgender Schluss gezogen werden: Handelt es sich bei der Anapher um ein Relativpronomen, ist ihr Antezedens im gleichen Textsegment zu finden.

### Reflexivpronomen

Anaphoric Reflexive Pronouns		Semantic Relations		total
		cospecLink		
		ident		
Rhetoric Relations	one-segment	1 anteEmbed: 0 anaEmbed: 0	1 anteEmbed: 0 anaEmbed: 0	
	purpose-s	1 anteEmbed: 0 anaEmbed: 1	1 anteEmbed: 0 anaEmbed: 1	
	same	4 anteEmbed: 0 anaEmbed: 0	4 anteEmbed: 0 anaEmbed: 0	
total		6 anteEmbed: 0 anaEmbed: 0.17	6 anteEmbed: 0 anaEmbed: 0.17	

Die Erwartung zur Verteilung der Reflexivpronomen wäre zunächst, daß sie sich zusammen mit ihrem Antezedens in einem Textsegment befinden. Diese Hypothese läßt sich allerdings mit Blick auf die Statistik nicht ganz bestätigen.

Zwei der sechs gefundenen Anapher-Antezedens-Paare sind auf verschiedene Textsegmente verteilt. Hierbei ist allerdings anzumerken, daß die Bezeichnung *one-segment* aussagt, daß zwei Textsegmente eng zueinander gehören, also zuerst zwar als getrennte Segmente erkannt wurden, aber in der späteren Analyse durch die rhetorische Relation *one-segment* zusammengefasst wurden. Dieser Umstand gilt so für das folgende Anapher-Antezedens-Paar:

**Ein Systembenutzer** ist desorientiert, wenn er **sich** im Gewirr möglicher

### Verknüpfungen verliert

Hier ist desweiteren zumindest diskutabel, ob die Anapher „sich“ nicht eher das Pronomen „er“ als Antezedens haben könnte. Dieses kann an dieser Stelle aber vernachlässigt werden, da die Analyse der rhetorischen Struktur die beiden ohnehin Teilsätze als *one-segment* klassifiziert.

Das zweite anaphorische Reflexivpronomen, für das gilt, daß sich das Antezedens nicht im selben Textsegment befindet, ist im folgenden Satz zu finden:

Einen Browser benötigt **man** auf seinem Rechner, um **sich** im WWW zu bewegen

Die beiden Textsegmente sind durch die Relation *purpose-s* verbunden. Hier wird deutlich, daß es sich nicht um eine außergewöhnliche Konstruktion handelt. Für den Fall, daß sich ein anaphorisches Reflexivpronomen in einem infiniten Nebensatz mit „zu“ befindet, kann es demnach als Normalfall angesehen werden, daß sich das Antezedens in einem anderen Textsegment befindet. Die beiden Textsegmente können z.B. durch rhetorische Relationen wie *purpose-s*, die für o.g. infinite Nebensätze charakteristisch sind, miteinander verbunden sein.

Darüberhinaus kann jedoch festgehalten werden, daß sich die Verteilung von anaphorischem Reflexivpronomen und seinem Antezedens oft auf ein und das selbe Textsegment beschränkt.

### Possessivpronomen

Die Analyse der possessiven Pronomen in Anaphernfunktion bietet leider nicht viele Interpretationsmöglichkeiten. Es ist durchaus nachvollziehbar, daß in dem Korpus von vier Texten nur sehr wenige Instanzen eines solchen Pronomens gefunden wurden. Es handelt sich genau genommen nur um eine Instanz, die obendrein als falsch annotiert angenommen werden muss. Das gefundene Anapher-Antezedens-Paar hat folgende Bestandteile:

**die Interaktion** – die **ihr** gewidmete Aufmerksamkeit

Das Pronomen „ihr“ wurde automatisch mit morpho-syntaktischen Eigenschaften versehen und hierbei fälschlicherweise als Possessiv- und nicht als Personalpronomen erkannt. Ein korrektes Beispiel für ein anaphorisches Possessivpronomen nach der hier gültigen Definition wäre:

Der Mann fand **eine Tasche**. Er erkannte, daß es **seine** war.

Eine solche Konstruktion ist im Prinzip vorstellbar, taucht in diesem Korpus jedoch nicht auf. Somit lassen sich an dieser Stelle über die Verteilung anaphorischer Possessivpronomen in der rhetorischen Struktur keine Aussagen treffen.

### Indefinitpronomen

Anaphoric Indefinite Pronouns		Semantic Relations		total
		cospecLink	bridgingLink	
		ident	setMember	
Rhetoric Relations	elaboration-continuation-other	2 anteEmbed: 0 anaEmbed: 3	-	2 anteEmbed: 0 anaEmbed: 3
	elaboration-drift	2 anteEmbed: 2 anaEmbed: 1	1 anteEmbed: 1 anaEmbed: 3	3 anteEmbed: 1.67 anaEmbed: 1.67
total		4 anteEmbed: 1 anaEmbed: 2	1 anteEmbed: 1 anaEmbed: 3	5 anteEmbed: 1 anaEmbed: 2.2

Beim Blick auf die Verteilung der anaphorischen Indefinitpronomen und ihrer Antezedentien fällt auf, daß es kein Paar gibt, das sich im gleichen Textsegment befindet.

Zwar ist die Datenlage auch hier nicht allzu umfangreich. Trotzdem halte ich diese Verteilung für nicht ganz zufällig.

Allerdings kann diese Vermutung an dieser Stelle nicht weiterverfolgt werden, da selbst unter den wenigen gefundenen Instanzen anaphorischer Indefinitpronomen noch Zweifelsfälle vorhanden sind. Ich halte es jedenfalls für fragwürdig, daß sich das Indefinitpronomen „man“ anaphorisch auf ein vorhergehendes Indefinitpronomen „man“ bezieht. In zwei der fünf gefundenen Anaphern ist diese Interpretation erfolgt. Deshalb belasse ich es bei der Vermutung, daß anaphorische Indefinitpronomen selten im gleichen Textsegment wie ihr Antezedens vorkommen.

### Interrogativpronomen

Für die Klasse der Interrogativpronomen wurde nur ein Beispiel für anaphorischen Pronomengebrauch gefunden. Bei näherer Betrachtung ist schon dieser eine Fund verwunderlich; denn man könnte sich im Vorhinein fragen, ob Interrogativpronomen überhaupt anaphorisch genutzt werden können. Die Datenlage spricht gegen diese Möglichkeit: Die gefundene Instanz eines anaphorischen Interrogativpronomens stellt sich als falsch analysiertes Relativpronomen „welches“ heraus.



## 5 FAZIT

Die Analyse der Verteilung von anaphorischen Pronomen in der rhetorischen Struktur von Texten gestaltet sich nicht einfach. Es hat sich herausgestellt, daß das Korpus von drei wissenschaftlichen Artikeln und einem Zeitungsartikel nicht ausreicht, um eindeutige Aussagen treffen zu können. Trotzdem sind die Daten aussagekräftig genug, Tendenzen aufzudecken.

Generell ist wohl davon auszugehen, daß anaphorische Pronomen in einer Kospezifikationsrelation (hier bezeichnet mit *cospecLink - ident*) mit ihrem Antezedens stehen, wobei sich beide auf den selben außersprachlichen Referenten beziehen.

Mit Blick auf die Verteilung in der rhetorischen Struktur läßt sich festhalten, daß die verschiedenen Pronomenarten durchaus unterschiedliche Ergebnisse liefern, die für die jeweilige Klasse als charakteristisch erscheinen. Interrogativpronomen und Possessivpronomen fallen aus der Analyse heraus, da sie in diesem Korpus, aber anscheinend auch im Allgemeinen, eher nicht in anaphorischer Verwendung vorkommen. Dabei darf natürlich nicht vergessen werden, daß nach der vorliegenden Definition nur pronominale Nominalphrasen bestehend aus einem einzelnen Pronomen betrachtet wurden.

Für Relativ- und Reflexivpronomen hat sich eine sehr homogene Verteilung herausgestellt. Die Antezedentien dieser beiden Pronomenarten sind, abgesehen von zweifelhaften Fällen, ausnahmslos im gleichen Textsegment wie die pronominale Anapher zu finden. Das mag trivial erscheinen, stellt jedoch im Bereich der pronominalen Anaphernresolution einen wichtigen Hinweis für Suchstrategien dar, der empirisch nachvollzogen werden kann.

Ganz so eindeutig wie bei den genannten Pronomenarten ist die Lage bei Personal-, Demonstrativ- und Indefinitpronomen nicht. Hier sind mit Hilfe der vorliegenden Daten sicherlich leichte Tendenzen festzustellen. Z.B. verteilen sich anaphorische Demonstrativpronomen und ihre Antezedentien entweder auf das gleiche Textsegment oder auf durch *elaboration*-Relationen verbundene Segmente. Trotzdem bietet das Korpus im Hinblick auf die Verteilung in der rhetorischen Struktur dieser drei Pronomenarten keine entscheidenden Hinweise.

**LITERATURVERZEICHNIS**

- Brandenburg, P. (2005). *Apollonios Dyskolos. Über das Pronomen. Einführung, Text, Übersetzung und Erläuterungen*. München/Leipzig: K. G. Saur Verlag.
- Consten, M. (2004). *Anaphorisch oder deiktisch? Zu einem integrativen Modell domänengebundener Referenz*. Tübingen: Max Niemeyer Verlag.
- Goecke, D., Stührenberg, M. & Holler, A. (unveröffentlicht). *Koreferenz, Kospezifikation und Bridging: Ein Schemaentwurf*. Interne Reports der DFG-Forscherguppe 437 „Texttechnologische Informationsmodellierung“, Projekt A2 „Sekimo“.
- Glück, H. (Hrsg.) (2000). *Metzler Lexikon Sprache*. 2. überarbeitete und erweiterte Auflage. Stuttgart/Weimar: Verlag J. B. Metzler.
- Kay, M. (2004). *XSLT 2.0 Programmers Reference*. 3. Auflage. Indianapolis: Wiley Publishing.
- Kim, K.-Y. (2003). *Funktion, Index und Pronomen*. Frankfurt a. M.: Peter Lang.
- Marcu, D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, Mass.: MIT Press.
- Mann, W. C./Thompson, S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. In: *Text*, Vol. 8, No. 3, S. 243-281.