

Qualitative Inhaltsanalyse durch Computer - ein uneinlösbarer Anspruch?

-

Untersuchungen zur algorithmischen Textinhaltserschließung
am Beispiel der referentiellen Interpretation

Inauguraldissertation
zur Erlangung des Grades eines Doktors der Philosophie
im Fachbereich Gesellschaftswissenschaften
der Johann Wolfgang Goethe-Universität
zu Frankfurt am Main

vorgelegt von
Roland Stuckardt
aus Frankfurt am Main

1999
(Einreichungsjahr)

2000
(Erscheinungsjahr)

1. Gutachter: Prof. Dr. Dr. Dieter Mans
2. Gutachter: Prof. Dr. Peter Ph. Mohler
3. Gutachter: PD Dr. Winfried D'Avis

Tag der mündlichen Prüfung: 7. Januar 2000

Erschienen im TENEA Verlag für Medien, Berlin

Vorwort

Inhaltsanalyse ist eine sozialwissenschaftliche Schlüsselmethode, weil sprachliche Texte das dominierende Ausgangsmaterial empirischer Untersuchungen sind. In bezug auf die Vielzahl von inhaltsanalytischen Verfahren gilt bislang eine einfache dichotome und folgenreiche Klassifikation: Neben den kategorien- und wörterbuchbasierten quantitativen Methoden, deren wesentlicher Analyseschritt durch Computer durchgeführt werden kann, gibt es die qualitativen - oft als "hermeneutisch" charakterisierten - Methoden, die aufwendige nicht computerisierte Interpretationsschritte erfordern. Die Idee, daß man die potentiellen Vorteile der qualitativen Analyse durch eine algorithmische Präzisierung ihrer Vorgehensweise nutzen könne, ohne die prinzipiellen Nachteile, auf die in der "Qualitativ-Quantitativ"-Kontroverse immer wieder hingewiesen wurde, in Kauf zu nehmen, ist nicht populär. Sie scheint unvereinbar mit festgefügtten wissenschaftlichen Weltbildern und (Vor-)Urteilen.

Die Arbeit von Roland Stuckardt führt diese alte sozialwissenschaftliche Kontroverse mit einer sehr anspruchsvollen Methodik fort. Ein quantitativer Analytiker könnte sie als einen wichtigen Beitrag zur Lösung des Problems pronominaler Referenzen lesen. Mit den von Stuckardt theoretisch vorgestellten und praktisch implementierten Verfahren läßt sich die Referenz von Wörtern wie "er", "sein", "sich" etc. bestimmen, und dies ermöglicht eine wesentliche Verbesserung der quantitativen Inhaltsanalyse.

Aber seine Arbeit enthält noch einen zweiten methodischen Beitrag - oder soll man vielleicht besser von einer listigen Botschaft sprechen? Die Theorien und Technologien zur Verbesserung der quantitativen Analyse weisen über die quantitativen inhaltsanalytischen Verfahren hinaus und zeigen, daß eine qualitative Inhaltsanalyse mit dem Computer möglich ist. Eine erfolgreiche referentielle Interpretation ist eingebettet in Verfahren der Künstlichen Intelligenz und Kognitionswissenschaften, die eine partielle Erschließung von Textinhalten ermöglichen. Die Bedeutung der Arbeit besteht gerade darin, in die sozialwissenschaftliche Textanalyse neue Techniken am Beispiel der referentiellen Interpretation einzuführen, deren Rezeption auch zu einer besseren qualitativen Texterschließung mit dem Computer führen wird. Indem Stuckardt beispielhaft wesentliche Voraussetzungen einer qualitativen, algorithmisch präzisen Textinterpretation beschreibt, bereichert er eine alte Kontroverse mit neuen Argumenten.

Aus den Ergebnissen der Dissertation von Herrn Stuckardt ist zu folgern, daß die Zeit für eine intensive Beschäftigung mit den Chancen computergestützter qualitativer Textanalyse reif ist. Es ist der Arbeit zu wünschen, daß sie in den Sozialwissenschaften mithilft, die Diskussionen über die Möglichkeiten qualitativer Analyse durch Computer zu intensivieren.

Frankfurt am Main, im März 2000

Prof. Dr. Dr. Dieter Mans

Danksagung

Mein erster und ganz besonderer Dank gilt meinem Betreuer, Herrn Prof. Dr. Dr. Dieter Mans. Seine Aufgeschlossenheit gegenüber interdisziplinär ausgerichteten Forschungsthemen und seine konstruktiv-kritischen Anmerkungen in vielen Stunden fruchtbarer Diskussion waren für das Gelingen der vorliegenden Dissertation ebenso ausschlaggebend wie seine "moralische" Unterstützung in einigen schwierigen Phasen der Arbeit.

Sodann möchte ich Herrn Prof. Dr. Peter Ph. Mohler für sein Interesse und seine Offenheit gegenüber meinen Forschungsarbeiten herzlich danken. Seine zahlreichen Kommentare und Verbesserungsvorschläge haben sich als sehr wertvoll erwiesen.

Auch Herrn Pasi Tapanainen, Department of General Linguistics, University of Helsinki, sei an dieser Stelle für seine freundliche Unterstützung gedankt. Durch seine wiederholte Bereitschaft, computergestützte Syntaxanalysen für die evaluationsrelevanten Textkorpora mit seinem *Dependency Parser for English* durchzuführen und die Resultate in dem benötigten Format zur Verfügung zu stellen, hat er in wesentlichem Maße dazu beigetragen, daß die Untersuchungen der vorliegenden Arbeit auf einer vergleichsweise konkreten Ebene angesiedelt werden konnten und die Ergebnisse die entsprechend große Aussagekraft aufweisen.

Für die Unterstützung bei der Schlußredaktion des Manuskripts danke ich Frau Irmhild Hehlein. Ihrer tatkräftigen Mithilfe verdanke ich einige unerwartete Einsichten betreffend die Subtilitäten von deutscher Rechtschreibung und Grammatik.

Schließlich verdient auch Herr Bernd Weidmann ein großes Dankeschön. Auf seinen wertvollen Hinweis zur rechten Zeit ist es zurückzuführen, daß das Thema meiner Dissertation im interdisziplinären Kontext von Sozialwissenschaften, Linguistik und Informatik/Computerlinguistik angesiedelt ist.

Frankfurt am Main, im Januar 1999

Roland Stuckardt

Inhaltsverzeichnis

Vorwort	v
Danksagung	vii
1 Einleitung	1
1.1 Qualitative Sozialforschung	2
1.2 Algorithmen als Explikationsmodell	3
1.3 Das Problem der referentiellen Interpretation	4
1.4 Zentrale Thesen und wissenschaftlicher Beitrag	5
1.5 Gliederung	7
I Inhaltsanalyse - Verfahren, Theorie und Algorithmen	9
2 Inhaltsanalyse	11
2.1 Die sozialwissenschaftliche Inhaltsanalyse	11
2.1.1 Definitionen und grundlegende Begriffe	12
2.1.2 Inhaltsanalyse als Kategorienanalyse	16
2.1.3 Qualitative vs. Quantitative Inhaltsanalyse	19
2.2 Qualitative Aspekte der Inhaltsanalyse	20
2.2.1 Festlegung des Kategoriensystems	20
2.2.2 Kategorisierung (Codierung)	21
2.2.3 Deskriptionserstellung und Auswertung	22
2.2.4 Interpretation und Validierung der Auswertungsergebnisse	23
3 Computereinsatz in der Inhaltsanalyse	25
3.1 Computerunterstützte versus Computergestützte Analyse	25
3.2 Computerunterstützte Analyse	26
3.2.1 Unterstützung der Verwaltung von Kategorisierungsentscheidungen	26
3.2.2 Unterstützung der Bildung von Code-Abstraktionen	27
3.3 Computergestützte Analyse	29
3.3.1 Algorithmisierung des Kodierschritts	29
3.3.2 Das Softwaresystem TEXTPACK	30
3.3.3 Das Softwaresystem INTEXT	33
3.3.4 Das Softwaresystem GENERAL INQUIRER	33
3.4 Diskussion: Computereinsatz für die Inhaltsanalyse	34
4 Theorie der wörterbuchbasierten Kategorisierung	37

4.1	Das Wort als Bedeutungseinheit der Kategorienanalyse	38
4.2	Isotopie	41
4.3	Konstitution von Themen - eine textlinguistische Fallstudie	42
4.3.1	Begriffsdefinitionen	43
4.3.2	Die Fallstudie	45
4.4	Determinanten von Textthemen	50
4.4.1	Valenz und Abhängigkeitsstruktur	52
4.4.2	Parataxe und Hypotaxe	54
4.4.3	Hintergrundinformation	55
4.4.4	Zusammenfassung	55
4.5	Grenzen der wörterbuchbasierten Themenanalyse	56
5	Verfeinerte algorithmische Inhaltsanalyse	61
5.1	Anforderungen der Computergestützten Inhaltsanalyse	62
5.2	Morphologische Analyse	63
5.3	Morphologische und lexikalische Disambiguierung	64
5.4	Verfeinerung der algorithmischen Kategorisierung	65
5.4.1	Syntaktische Analyse	66
5.4.2	<i>Information Extraction</i>	69
5.5	Zusammenfassung	87
II	Algorithmische Interpretation referentieller Ausdrücke	91
6	Referentialität	93
6.1	Sprachliche Ausdrücke und referentieller Wiederaufgriff	94
6.1.1	Elementare Fallbeispiele	95
6.1.2	Komplexere Instanzen referentiellen Wiederaufgriffs	96
6.2	Begriffsdefinitionen	99
6.2.1	Anaphern und anaphorische Bezugnahme	99
6.2.2	Referenz, Koreferenz und Kospezifikation	100
6.3	Anapherninterpretation als sprachverarbeitender Prozeß	103
6.4	Eingrenzung des Untersuchungsgegenstands	105
7	Bausteine der referentiellen Interpretation	107
7.1	Restriktionen	109
7.1.1	Morphologische Kongruenz	109
7.1.2	Syntaktisch-konfigurationale Restriktionen	112
7.1.3	Semantische Skopus-Restriktionen	122
7.1.4	Fokusbasierte Bedingungen	123
7.1.5	Einschränkung kataphorischer Bezugnahmen	132
7.1.6	Sonstige Restriktionen	133
7.2	Präferenzkriterien	134
7.2.1	Fokusbasierte Präferenzbedingungen	135
7.2.2	Elementare Präferenzbedingungen	142
7.3	Zusammenfassung	145
8	Anapherninterpretationsverfahren	149
8.1	Der "naive" Algorithmus von Hobbs	150

8.2	Die fokusbasierte Anapherninterpretation von Sidner	151
8.3	Das “ <i>Shallow Processing</i> ”-Verfahren von Carter	152
8.4	Der Centering-Ansatz von Brennan, Friedman und Pollard	154
8.5	Der Mehrstrategieansatz von Carbonell und Brown	155
8.6	Der dependenzbasierte Algorithmus von Lappin und Leass	156
8.7	Der parserlose Algorithmus von Kennedy und Boguraev	157
8.8	Teilnehmer der “ <i>Coreference Task</i> ”-Evaluation von MUC-6	159
8.9	Diskussion	160
9	Basisalgorithmus und referentielle Datenstrukturen	163
9.1	Auswahl relevanter und algorithmisierbarer Strategien	164
9.2	Datenstrukturen für referentielle Information	164
9.3	Algorithmisierung der gewählten Strategien	167
9.3.1	Strategie MORKONGRU	168
9.3.2	Strategie SYNKONFIG	169
9.4	Der Basisalgorithmus	178
9.4.1	Integration der Strategien in das dreiphasige Rahmenverfahren	178
9.4.2	Laufzeitkomplexität	180
9.4.3	Eine asymptotisch effizientere Implementierungsvariante	181
9.5	Diskussion	182
9.5.1	Unterspezifikation des Basisalgorithmus	182
9.5.2	Berücksichtigung bindungstheoretischer Feinheiten: leere Kategorien . . .	182
10	Robuste referentielle Interpretation	187
10.1	Robustheit	188
10.1.1	Robustheit von Sprachverarbeitungsprozessen	189
10.1.2	Robustheit und referentielle Interpretation	189
10.1.3	Zwei Modelle robuster referentieller Interpretation	191
10.2	Fragmentarische syntaktische Beschreibungen	191
10.2.1	Ursachen fragmentarischer Syntax	192
10.2.2	Fragmentarische Syntax und referentielle Interpretation	193
10.3	Bindungsbedingungen auf fragmentarischer Syntax	196
10.3.1	Grundlegende Beobachtungen	197
10.3.2	Nichtheuristische Entscheidungsregeln für den fragmentarischen Fall . . .	197
10.3.3	Eine robuste Version des Basisalgorithmus	202
10.4	Ein vollständiger Algorithmus	208
10.4.1	<i>Packed Shared Forests</i> zur Repräsentation struktureller Ambiguität . . .	208
10.4.2	Dominanzrelationen in PSFs	210
10.4.3	Verifikation der Bindungsbedingungen auf PSFs	213
10.4.4	Verfeinerung des Basisalgorithmus	215
10.4.5	Strukturelle Disambiguierung durch referentielle Evidenz	217
10.5	Diskussion	218
10.5.1	Der ultimative Algorithmus?	219
10.5.2	Ausblick	219
11	Implementierung und Evaluation	221
11.1	Eingrenzung des Evaluationsszenarios	223
11.1.1	Teildisziplinen der Evaluation	223

11.1.2	Wahl geeigneter Textkorpora	225
11.1.3	Präzisierung der inhaltlichen und formalen Vorgaben	227
11.1.4	Definition formaler Evaluationsmaße	234
11.2	Implementierung: das ROSANA-System	243
11.2.1	Übersicht der Systemarchitektur	243
11.2.2	Syntaktische und morphologische Analyse	245
11.2.3	Das Kernsystem ROSANA	247
11.2.4	Bewertung der systemgenerierten Ergebnisse	259
11.3	Evaluation	260
11.3.1	Überblick der Ergebnisse für das Korpus <i>“Presse-Meldungen”</i>	260
11.3.2	Diskussion der Ergebnisse für die OV-Aufgabe	262
11.3.3	Diskussion der Ergebnisse für die KV-Aufgabe	263
11.3.4	Diskussion der Ergebnisse für die PS-Aufgabe	266
11.3.5	Überblick der Ergebnisse für das Korpus <i>“Mozart-Opern”</i>	267
11.3.6	Stellenwert und Skopus der Prominenzfaktoren und Gewichte	269
11.4	Diskussion	272
11.4.1	Vergleich mit anderen Systemen zur Interpretation von Pronomen	272
11.4.2	Möglichkeiten zur Verfeinerung von ROSANA	274
12	Zusammenfassung und Ausblick	275
12.1	Neubewertung der Computergestützten Inhaltsanalyse	276
12.2	Algorithmische Lösbarkeit des Pronomen-Problems	279
12.3	Der nächste Schritt: anwendungsbezogene Studien	282
A	ROSANA-Interpretation einer Pressemeldung	285
A.1	Pressemeldung Nr. 57	285
A.2	Satzweise Interpretation (OV- und KV-Aufgabe)	286
A.2.1	Satz 1	287
A.2.2	Satz 2	287
A.2.3	Satz 3	288
A.2.4	Satz 4	289
A.2.5	Satz 5	293
A.2.6	Satz 6	294
A.2.7	Satz 7	296
A.2.8	Satz 8	297
A.2.9	Satz 9	297
A.3	Bestimmung nichtpronominaler Substitute (PS-Aufgabe)	299
A.4	Bewertung der Interpretationsleistung	299
A.4.1	Bestimmung objektspezifizierender Vorkommen (OV-Aufgabe)	299
A.4.2	Bestimmung der Kospezifikations-Äquivalenzklassen (KV-Aufgabe)	300
A.4.3	Bestimmung lexikalischer Substitute für Pronomen (PS-Aufgabe)	302
B	ROSANA-Interpretation eines <i>“Mozart-Opern”</i>-Texts	305
B.1	Text <i>“Don Giovanni”</i>	305
B.2	Antezedensentscheidungen	308
B.3	Nichtpronominale Substitute	312
B.4	Bewertung der Interpretationsleistung	315
	Literaturverzeichnis	317

Kapitel 1

Einleitung

Die klassische maschinelle Textinhaltsanalyse der Sozialwissenschaften basiert auf einem Wörterbuchmodell. Auf der Grundlage von Wortlisten, die unterschiedliche Kategorien definieren, wird eine automatische Codierung der zu analysierenden Texte bewerkstelligt, deren Ziel in einer Aufschlüsselung des Materials nach untersuchungsrelevanten *Themen* besteht. Theoretischer Hintergrund dieser Wörterbuchanalyse ist die traditionelle *quantitative inhaltsanalytische Methode*, deren Anspruch es ist, den Textinhalt möglichst *objektiv*, d.h. unter weitgehender Ausschaltung subjektiv-interpretativer Faktoren, zu erschließen, um eine *maximale intersubjektive Reproduzierbarkeit* der Codierentscheidungen zu erreichen. Die in den Aufgabenbereich des Forschers fallende Definition der Kategorien in Form von Wortlisten soll auf der Grundlage eines Verständnisses von *Durchschnittsbedeutungen* der verwendeten Wörter erfolgen, wobei das kommunikative Wissen über den spezifischen Entstehungskontext der vorliegenden Texte unberücksichtigt bleibt. Auf diesem Weg soll der sog. *manifeste Inhalt* der Textdokumente *beschrieben* werden, der als von einem an subjektive Interpretationen gebundenen sog. *latenten Inhalt* operational abgrenzbar angesehen wird. Die Auswertung des Ergebnisses der wortorientierten Kategorisierung geschieht typischerweise auf quantitativem Wege - z.B. per elementarer, vom Ausgangsmaterial numerisch abstrahierender Häufigkeitsauszählung der den unterschiedlichen Dokumenten zugeordneten Kategorien, gefolgt von einer mathematisch-statistischen Analyse.

Die Vercodung des Textmaterials erfolgt auf der Basis elementarer Zeichenketten-Vergleiche der Textwörter mit den Elementen der kategoriedefinierenden Wortlisten. Die algorithmische Umsetzbarkeit dieser Aufgabenbeschreibung in ein Computerprogramm ist unmittelbar und ohne Einschränkung gewährleistet. Da die Vercodung somit durch den Computer bewerkstelligt werden kann, wird der Anspruch der intersubjektiven Reproduzierbarkeit des Klassifikationsvorgangs in bestmöglicher Weise eingelöst.

Zwar hat es den Anschein, als werde der Objektivitätszielsetzung durch das beschriebene Verfahren tatsächlich in ultimativer Form Rechnung getragen. Jedoch offenbart eine nähere Analyse einige grundlegende Probleme des Modells der wörterbuchbasierten maschinellen Inhaltsanalyse. Ist es theoretisch überhaupt gerechtfertigt, von "*Durchschnittsbedeutungen*" sowie von "*manifestem*" vs. "*latentem*" Inhalt zu sprechen und eine scharfe *Trennlinie zwischen quantitativ-objektiver vs. interpretativ-subjektiver (qualitativer) Analyse* zu ziehen? Auf welcher Grundlage kann von einer Erschließung der "*Themen*" des ausschließlich unter Bezug auf die Einzelwortebene klassifizierten Textmaterials gesprochen werden? Wie steht es mit der *Gültigkeit*, d.h. der Aussagefähigkeit der erzielbaren Ergebnisse relativ zum Gegenstandsbereich der inhaltsanalyti-

schen Untersuchung?

Da es in der Tat fragwürdig erscheint, von einer intersubjektiv unkontroversen Definierbarkeit von “Durchschnittsbedeutungen” durch den Forscher auszugehen, kann jedenfalls von einer vollkommenen Objektivität der wörterbuchbasierten Inhaltsanalyse nicht die Rede sein (vgl. Lisch und Kriz: [LiKr78], S. 45 f.). Während der eigentliche Codiervorgang mit hundertprozentiger Genauigkeit maschinell reproduzierbar ist, ist dies jedoch keineswegs für die vorgelagerte Phase der Entwicklung des Kategoriensystems gewährleistet, die in jedem Fall eine interpretativ-subjektive Operation bleibt. Insofern erweist sich die Unterscheidung zwischen quantitativ-objektiver vs. interpretativ-subjektiver (qualitativer) Analyse in bezug auf das Gesamtverfahren als problematisch. Als “*quantitativ*” im eigentlichen Sinne des Wortes sind lediglich die typischerweise numerisch-reduktiven Operationen der Auswertephase zu bezeichnen, die sich in der wörterbuchbasierten Analyse an die Klassifikation anschließt. Auch die durch den Computer bewerkstelligten Kategorisierungen sind Elementarinstanzen *qualitativer* Verarbeitung.

Der mit der wörterbuchbasierten Inhaltsanalyse verbundene Objektivitätsanspruch wird somit nicht mit der ursprünglich postulierten Stringenz eingelöst, da qualitative Elemente niemals vollständig eliminiert werden können. An die Stelle der somit nicht einlösbaren Forderung einer Erschließung objektiver, “manifeste” Inhalte tritt nunmehr die Zielsetzung einer möglichst guten intersubjektiven Reproduzierbarkeit der inhärent qualitativen inhaltsanalytischen Schritte. Aufgrund der algorithmischen Explikation genügt der einzelwortbasierte Codiervorgang der maschinellen Inhaltsanalyse diesen Anforderungen in ultimativer Weise, jedoch ist er der Kritik einer unzureichenden texttheoretischen Fundierung ausgesetzt (vgl. die Arbeit Fühlaus, in der u.a. aus diesem Grund die “*Sprachlosigkeit der Inhaltsanalyse*” attestiert wird ([Fühl82])). Die “Themen”, die auf der Grundlage von Einzelwortvergleichen erschlossen werden, verkörpern nur die Oberfläche der durch den jeweiligen Text kommunizierten Inhalte. Insofern sind zwar die Codierentscheidungen reproduzierbar, jedoch genügen die erschlossenen inhaltlichen Entitäten nicht den Anforderungen einer aussagefähigen Inhaltsanalyse. *Es stellt sich unmittelbar die Frage, ob die maschinelle wörterbuchbasierte Inhaltsanalyse auf eine solidere linguistische Basis gestellt werden könnte, durch die die begrenzten Möglichkeiten einer wortorientierten Textinhaltserschließung unter Beibehaltung der algorithmischen Explikation des Klassifikationsvorgangs transzendiert werden.*

1.1 Qualitative Sozialforschung

Die beschriebene Problemstellung verkörpert den Ausgangspunkt der vorliegenden Arbeit, in der es folglich primär darum geht, eine inhärent qualitative Operation der maschinellen Inhaltsanalyse - den Codierschritt - auf der Grundlage von Algorithmen *mit verfeinertem Inhaltsbezug* zu explizieren. Die Untersuchungen betreffend die algorithmische Formalisierbarkeit qualitativer Prozesse sind jedoch vor dem breiteren Hintergrund der sog. *qualitativen Wende* zu sehen, die sich durch einen Trend hin zu betont subjektiv-interpretativ ausgerichteten Verfahren in der sozialwissenschaftlichen Forschung sowie in verwandten Disziplinen wie Erziehungswissenschaften und Psychologie seit Beginn der siebziger Jahre abzeichnet. Die auf strukturiert-standardisierten - und damit tendenziell präskriptiven - Instrumenten wie Multiple-Choice-Fragebögen, Bewertungsskalen etc. basierende, den Forschungsgegenstand regelrecht *vermessende* Methodik ist zunehmender Kritik ausgesetzt. In sozialen Prozessen, so wird argumentiert, stellen normativ fundierte, d.h. auf starren, fest vorgegebenen Entscheidungsgrundlagen basierende Handlungs-

muster nicht den Regelfall dar; vielmehr seien soziale Interaktionen an subjektive *interpretative Prozesse* geknüpft. Für die Erforschung sozialer Strukturen, so wird gefolgert, sollten *interpretative Methoden* zum Einsatz kommen, die den Forschungsgegenstand *entlang qualitativer Dimensionen* erfassen; ein im Kern empirisch-quantitatives Vorgehen sei inadäquat.¹

Eine Vielzahl qualitativer Verfahren wurde vorgeschlagen, deren Ziel in einer möglichst präzisen Explikation der *Rahmenbedingungen* des interpretativen Vorgehens besteht (vgl. etwa Mayring: [Mayr96]). Die Ansätze werden ausnahmslos durch vergleichsweise abstrakte Ablaufschemata beschrieben, in deren Einzelschritten auf die subjektive Sicht des Interpreten rekuriert wird, ohne den eigentlichen Interpretationsvorgang vollständig zu explizieren. Zur Erschließung der i.d.R. an den kulturellen, sozialen und historischen Kontext gebundenen *Regelmäßigkeiten*² sozialer Strukturen und Prozesse, die das Ziel der sozialwissenschaftlichen Forschung verkörpern, wird somit bewußt auf die kontextsensitive Interpretation des menschlichen Interpretieren zurückgegriffen, die zwar *regelmäßig* eingegrenzt, nicht jedoch *formal* beschrieben werden soll. Um die erzielten Ergebnisse zusätzlich abzusichern, kommen unterschiedliche Validierungs-Verfahren wie etwa gezielte *Variationen der Kontextbedingungen* oder *Diskussionsrunden des Plenums der Interpreten* zum Abgleich der individuellen Deutungen zum Einsatz. Die hierdurch erzielte Explikation ist nicht maximal (i.S.v. algorithmisch), jedoch wird dies angesichts der andersweitig nicht gewährleistbaren Qualität der Untersuchungsergebnisse als unvermeidlich angesehen; auf der Basis der Rahmenverfahren soll eine zumindest hinreichend hohe intersubjektive Reproduzierbarkeit erzielt werden.

1.2 Algorithmen als Explikationsmodell

Die Untersuchungen der vorliegenden Arbeit, die in bezug auf ein qualitatives Teilproblem der maschinellen Textinhaltsanalyse durchgeführt werden, sollen demnach vor dem Hintergrund der allgemeineren Problemstellung der formalen Explizierbarkeit *beliebiger* qualitativer Prozesse gesehen werden. Protagonisten der philosophisch-hermeneutischen Lehre vertreten die Position, daß eine vollständige Explikation des in eine subjektive Textinterpretation einfließenden *impliziten Vorverständnisses* unmöglich sei. Eine uneingeschränkt formale Beschreibung qualitativer Verstehensprozesse sei demnach ein aus theoretischen Gründen uneinlösbarer Anspruch; gemäß einer stringenten Auslegung der hermeneutischen Theorie sind somit jegliche Algorithmisierungsversuche zum Scheitern verurteilt. (Vgl. etwa die insbesondere von den Arbeiten Gadamer ([Gada60]) und Heideggers ([Heid79]) ausgehende Kritik, die Winograd und Flores in bezug auf "rationalistische" Verstehens-Modelle formulieren ([WiFl89])). Hieraus wird auf die prinzipielle Inadäquatheit jeglichen Computereinsatzes in der Bewerkstelligung der notwendig qualitativen Aufgaben der sozialwissenschaftlichen Inhaltsanalyse geschlossen.

In der vorliegenden Arbeit wird indes eine gegenteilige Position vertreten und anhand der Ergebnisse theoretischer und praktischer Untersuchungen zur maschinellen Textinhaltsanalyse untermauert. *Algorithmen werden als ein prinzipiell geeignetes Beschreibungsmittel für die vollständig*

¹Mayring bringt diese Kritik folgendermaßen auf den Punkt ([Mayr96], S. 1):

“Das rein quantitative Denken ist brüchig geworden; ein Denken, das sich den Menschen annähert, indem es testet und vermisst, mit ihnen experimentiert und ihre statistische Repräsentanz überprüft, ohne vorher den Gegenstand verstanden zu haben, seine Qualität erfaßt zu haben.”

²im Unterschied zu den universell gültigen *Gesetzmäßigkeiten*, die in den Naturwissenschaften untersucht werden

formale Explikation qualitativer Prozesse angesehen. Die Argumente der hermeneutischen Kritik werden jedoch insofern als aussagekräftig angesehen, als sie Aufschluß darüber geben, warum die formale Explikation des Hintergrundwissens der menschlichen Textinterpretation *schwierig* ist. Somit wird eine differenzierte Position eingenommen, derzufolge die Aufgabenstellung der formalen Explikation des Hintergrundwissens der intellektuellen (menschlichen) Interpretation als komplexes Problem einzustufen ist, jedoch keine prinzipielle Formalisierbarkeitsschranke theoretisch begründet werden kann. Ob eine konkrete qualitative Interpretationsaufgabe durch den Computer gelöst werden sollte, kann somit nicht unter Verweis auf die prinzipielle Nicht-Algorithmisierbarkeit entschieden werden. Vielmehr handelt es sich um eine Frage der Vertretbarkeit des technischen *Aufwands* sowie der auf dem gegenwärtigen Stand der Forschung erzielbaren *Analysegüte*; Vergleichsmaßstab sind Aufwand bzw. Analysegüte einer entsprechenden intellektuellen Inhaltsanalyse.

Da der zentrale Gegenstand der Untersuchungen in der Explikation qualitativer Aufgaben der Inhaltsanalyse mit dem Ziel einer maschinellen Ausführung besteht, wird die Formalisierung unter Rekurs auf Konzepte der *Informatik* bewerkstelligt. In erster Näherung lassen sich *Algorithmen* als *funktional eindeutige Charakterisierungen von Prozessen* definieren. Auf der Grundlage algorithmischer Programmbeschreibungen für die Turingmaschine, dem theoretischen Pendant des Computers, ergibt sich das mächtigste bekannte formale Modell von *Berechenbarkeit*; alle bislang definierten Berechenbarkeitsmodelle sind entweder äquivalent oder weniger mächtig. Gemäß der sog. *Church-Turing-These*, derzufolge *genau diejenigen Probleme "berechenbar" sind, die auf Turingmaschinen "berechenbar" sind*, existiert keine Alternative zur formalen Explikation von Prozessen, die die Grenzen des Modells der Turing-Berechenbarkeit transzendiert. Aus Sicht der Informatik, in der die Church-Turing-These weithin akzeptiert wird, verkörpern Turing-Algorithmen somit das Beschreibungsmittel der Wahl zur Explikation qualitativer Prozesse.

1.3 Das Problem der referentiellen Interpretation

Beginnend mit einer Analyse der theoretischen (insbesondere linguistischen) Hintergründe der sozialwissenschaftlichen Inhaltsanalyse werden in Teil I der Arbeit die auf der Basis des gegenwärtigen Stands der Forschung bestehenden Möglichkeiten einer algorithmischen Verfeinerung des Codierschritts der Themenanalyse von einem allgemeinen Standpunkt aus diskutiert; in Teil II wird ein *spezifisches Teilproblem* einer verfeinerten maschinellen Textinterpretation untersucht - *die Verarbeitung referenzierender (insbesondere pronominaler) sprachlicher Ausdrücke*. In bezug auf das Problem der referentiellen Interpretation werden Detailstudien durchgeführt, deren Ziel in der Entwicklung von Algorithmen besteht, durch die unterschiedliche Lösungswege für diese Aufgabe formal expliziert werden. Auf der Grundlage einer Implementierung eines der Algorithmen in einem konkreten Software-System soll der formale Nachweis der maschinellen Lösbarkeit dieses zentralen Teilproblems der Qualitativen Analyse von Texten erbracht werden. Da es sich um eine bekanntermaßen schwierige Aufgabe handelt, ist die algorithmische Explizierbarkeit entsprechender Prozesse als Beweis dafür zu werten, daß nichttriviale qualitative Teilschritte der Inhaltsanalyse tatsächlich mit gutem Erfolg durch den Computer bearbeitet werden können.

Daß das Problem der referentiellen Interpretation zentral für eine maschinelle Sprachinterpretation ist, die auf eine tiefere Erschließung der textuell kommunizierten Inhalte abzielt, soll anhand

eines ersten Beispiels veranschaulicht werden:³

“In der Bonner Koalition regte sich am Freitag erster offener Widerstand gegen die Pläne zur Neubewertung der Reserven. Der CDU-Verteidigungspolitiker Jürgen Augustinowitz kündigte an, er werde keiner Gesetzesänderung zustimmen, die die Stabilität der Mark oder das Vertrauen in die Bundesbank beeinträchtige. Spitzenpolitiker von Union und FDP bekräftigten hingegen, sie hielten an ihrem Vorhaben fest.”

Notwendige Voraussetzung für die Erschließung des propositionalen Gehalts der textkonstituierenden Äußerungsabfolge ist die korrekte *Interpretation der pronominalen Ausdrücke* “*sich*”, “*er*”, “*die*”, “*sie*” und “*ihrem*”. Ein menschlicher Interpret nimmt - ohne darüber zu reflektieren - etwa die sprachlichen Ausdrücke “*er*” und “*Jürgen Augustinowitz*” als inhaltlich äquivalent wahr, da sie (in einem noch näher zu explizierenden Sinne) dasselbe Weltobjekt - den Politiker Augustinowitz - referenzieren. Ohne diese Interpretationsleistung bliebe unklar, *wer* derjenige ist, der “*keiner Gesetzesänderung zustimmen*” würde. Das betrachtete Phänomen der *Koreferenz* ist keineswegs auf Instanzen von Pronominalisierung beschränkt, sondern betrifft in gleichem Maße nichtpronominaler Ausdrücke: In dem gezeigten Textausschnitt sind u.a. auch die Ausdrücke (Nominalphrasen) “*Neubewertung der Reserven*” (Satz 1) und “*ihrem Vorhaben*” (Satz 3) sprachliche Stellvertreter ein und derselben inhaltlichen Entität. Für eine auf der Einzelwortebene arbeitende und auf die Erschließung lexikalischer Information abzielende klassische Inhaltsanalyse jedoch verkörpern pronominaler Ausdrücke ein nicht ohne Weiteres lösbares Problem.

Auf den ersten Blick erscheint unklar, wie diese durch den Menschen anscheinend en passant erbrachte Interpretationsleistung durch ein Computerprogramm erbracht werden könnte.⁴ In der Tat handelt es sich um eine schwierige Aufgabe, für die bislang nur eine partielle (in Teilen heuristische) algorithmische Lösung existiert. Die Untersuchungen werden jedoch belegen, daß in bezug auf das Kernproblem - die *Resolution (Interpretation) pronominaler Ausdrücke* - bereits eine hohe maschinelle Analysequalität erzielt wird, die relativ nahe an die Leistungen menschlicher Textinterpreten heranreicht. Der Nachweis wird per formaler Evaluation erbracht, in der die Interpretationsergebnisse des Computers mit den Resultaten einer intellektuellen Interpretation derselben Texte abgeglichen werden.

1.4 Zentrale Thesen und wissenschaftlicher Beitrag

In Teil I der Arbeit werden die folgenden Thesen 1 und 2 untermauert, die sich auf allgemeine Aspekte des qualitativen Codierschritts der klassischen Inhaltsanalyse beziehen. Zunächst wird erarbeitet, warum die klassische maschinelle/Quantitative Inhaltsanalyse unter linguistischen Gesichtspunkten als unzureichend zu bewerten ist.

³aus: *Frankfurter Rundschau*, 31.5.97, S-Ausgabe, S. 1, Artikel “*Rückendeckung für die Bundesbank*”; Unterstreichungen durch den Verfasser der vorliegenden Arbeit

⁴Hiermit soll keineswegs der Anspruch einer “kognitiv adäquaten” Lösung erhoben werden, die die wesentlichen Aspekte der menschlichen Kognition algorithmisch abbildet. Die Untersuchungen sind primär dem Gebiet einer *ergebnisorientierten Künstlichen Intelligenz* als Teildisziplin der Informatik zuzuordnen; zumindest in bezug auf die Arbeitsweise der zu entwickelnden Algorithmen interessieren kognitionstheoretische Gesichtspunkte allenfalls am Rande.

These 1 *Die Beschränkungen der klassischen maschinellen/Quantitativen sozialwissenschaftlichen Inhaltsanalyse sind primär auf die unzureichende textlinguistische Fundierung des einzelwortorientierten Codierschritts zurückzuführen, durch den die "Themen" des zu analysierenden Texts nur auf einer elementaren lexikalischen Ebene erschlossen werden. Das Ziel einer thematischen Analyse wird somit nur in sehr eingeschränktem Umfang erreicht. Die in der Literatur vorgeschlagenen "Verfeinerungen" der maschinellen Inhaltsanalyse sind linguistisch unfundiert; die Grenzen der klassischen Verfahren werden nicht in entscheidendem Ausmaß transzendiert.*

Über eine Bestandsaufnahme und Analyse aktueller Verfahren der maschinellen Textinterpretation wird der Nachweis erbracht, daß die textlinguistischen Defizite der klassischen Inhaltsanalyse unter Beibehaltung der maximalen (algorithmischen) Explikation zumindest in Teilen eliminiert werden können.

These 2 *Bereits auf der Basis des gegenwärtigen Stands der Forschung sind die wesentlichen Voraussetzungen für eine algorithmische, insbesondere auf linguistischem Wissen basierende Verfeinerung des Codierschritts als qualitative Operation der maschinellen Inhaltsanalyse erfüllt. Weder theoretische noch technische Gründe sprechen für eine Begrenzung maschineller Themenanalysen auf die einzelwortorientierte, lexikalische Ebene. Die Grenzen der klassischen Verfahren sind somit in entscheidendem Ausmaß transzendierbar.*

Thesen 3 und 4 betreffen den qualitativen Prozeß der Koreferenzanalyse, dessen algorithmische Explikation das Ziel von Teil II der Arbeit bildet. Das zentrale Ergebnis besteht in der Erkenntnis, daß eine - wenn auch heuristische - Algorithmisierung bewerkstelligt werden kann, die primär auf linguistischem (morphologischem und syntaktischem) Wissen basiert. Für ein Computersystem, das diesen Algorithmus implementiert, wird der formale Nachweis erbracht, daß es das Teilproblem "Koreferenzanalyse" der referentiellen Interpretation insbesondere in bezug auf die Interpretation von Pronomen mit gutem Erfolg löst.

These 3 *Für den Teilprozeß "Koreferenzanalyse" des qualitativen Prozesses der referentiellen Interpretation ist eine algorithmische Explikation auf der Basis des gegenwärtigen Stands der Forschung möglich.*

Im Hinblick auf die Anwendung als Teilprozeß einer linguistisch verfeinerten maschinellen Inhaltsanalyse soll der algorithmische Ansatz der referentiellen Interpretation darüberhinaus den Anforderungen einer anwendungstauglichen, *robusten* Verarbeitung beliebiger Texte genügen. Nur ein kleiner Teil der in der Literatur beschriebenen Theorien bzw. Verfahren der referentiellen Interpretation erfüllt diese Bedingung. Analyseansätze, die allenfalls marginal relevante referentielle Phänomene abdecken oder von der uneingeschränkten orthographischen und grammatischen Korrektheit des Texts ausgehen, erweisen sich als ebenso ungeeignet wie Verfahren, die die Verfügbarkeit von Wissen voraussetzen, das auf der Basis aktueller Algorithmisierungstechniken nicht mit vertretbarem Aufwand oder nur partiell explizierbar ist. Aus der theoretischen Perspektive ist der Begriff der Robustheit weiter gefaßt und betrifft die analysetechnische Verschränkung unterschiedlicher Teilprozesse der Qualitativen Analyse. Betreffend das Problem der Koreferenzinterpretation kann der Zielsetzung einer robusten Textinterpretation auch in bezug auf diesen erweiterten Anspruch Rechnung getragen werden:

These 4 *Von zentraler Bedeutung für die algorithmische Explikation qualitativer Prozesse ist die Entwicklung robuster Analyseverfahren. Der Zielsetzung der robusten Verarbeitung wird durch eine möglichst weitreichende Verschränkung unterschiedlicher Teilprozesse der Analyse Rechnung getragen. Betreffend das Problem der Koreferenzinterpretation ist dieser Anspruch über ein Prozeßmodell einlösbar, in dem die Teilprozesse der syntaktischen und referentiellen Analyse interagieren.*

Die Ergebnisse der Teile I und II der Arbeit bilden die Grundlage für eine differenzierte Neubewertung der Möglichkeiten einer maschinellen qualitativen Textanalyse, die in folgendem Kernpostulat zum Ausdruck gebracht wird:

These 5 *Algorithmen stellen die geeignete Basis zur formalen Explikation qualitativer Prozesse der Textinhaltserschließung dar. Aus den Postulaten der hermeneutischen Philosophie ist keine generelle Inadäquatheit computergestützter qualitativer Analysen ableitbar, sondern lediglich die Aussage, daß die algorithmische Explikation qualitativer Prozesse im Einzelfall schwierig sein kann. Die Frage nach der Adäquatheit von Computeranalysen kann somit nicht a priori negativ beschieden werden; sie ist vielmehr vor dem Hintergrund konkreter inhaltsanalytischer Probleme sowie mit Blick auf den State-of-the-Art maschineller Textanalyseverfahren zu beantworten.*

Anhand der Ausführungen wird der *interdisziplinäre Charakter* der durchzuführenden Untersuchungen deutlich. Zentraler Bezugspunkt ist die Methode der *sozialwissenschaftlichen Inhaltsanalyse*, die unter *linguistischen* Gesichtspunkten zu untersuchen ist. Durch das Ziel der Algorithmisierung und Implementierung ergeben sich ferner Berührungspunkte mit den Disziplinen *Informatik* sowie insbesondere *Computerlinguistik*.

Somit ist der *wissenschaftliche Beitrag* der Dissertation in mehreren unterschiedlichen Gebieten lokalisiert. Er wird konstituiert durch: (1) die Untersuchung und Aufschlüsselung des textlinguistischen Hintergrunds der sozialwissenschaftlichen Inhaltsanalyse als Themenanalyse; (2) die Analyse aktueller computerlinguistischer Textanalyseverfahren als mögliche Grundlage einer adäquaten Verfeinerung der computergestützten Inhaltsanalyse; (3) den Entwurf und die Implementierung anwendungstauglicher Algorithmen zur Koreferenzinterpretation insbesondere pronominaler Ausdrücke, wobei Modelle robuster referentieller Textinterpretation entwickelt werden, die u.a. wichtige computerlinguistische Neuerungen darstellen.

1.5 Gliederung

In Teil I der Arbeit werden die klassische maschinelle/Quantitative sozialwissenschaftliche Inhaltsanalyse sowie im Speziellen die textlinguistische Grundlage des qualitativen Codierschritts untersucht. Kapitel 2 führt in die Thematik ein, indem einige unterschiedliche Definitionsvorschläge der inhaltsanalytischen Methode gegenübergestellt und ferner die qualitativen Teilprobleme eines zentralen inhaltsanalytischen Verfahrens, der sog. Kategorien- oder Themenanalyse, identifiziert werden. In Kapitel 3 wird eine Synopse der bislang bekannten Ansätze des Computereinsatzes in der Inhaltsanalyse für die Bearbeitung qualitativer Teilprobleme erarbeitet. Die Ergebnisse der Bestandsaufnahme legen eine Unterscheidung zwischen computerunterstützter und computergestützter Inhaltsanalyse nahe. Ausschließlich in Ansätzen der letzteren Art übernimmt die Maschine bestimmte Teilaufgaben der eigentlichen Textinterpretation. Eine Analyse

bekannter Systeme zur computergestützten Inhaltsanalyse zeigt auf, daß es sich vornehmlich um einzelwortorientierte Verfahren zur maschinellen Ausführung des Codierschritts handelt. Kapitel 4 untersucht die textlinguistischen Hintergründe der klassischen Themenanalyse, wodurch sowohl die elementare (lexikalische) Funktionsgrundlage als auch die inhärenten Beschränkungen des zugrundeliegenden Wörterbuch-Modells offengelegt werden. Es wird gezeigt, daß diese Grenze nur auf der Basis einer textlinguistisch fundierten Erschließung inhaltlicher Entitäten (“Themen”) transzendiert werden kann; im Speziellen wird nachgewiesen, daß die in bezug auf die wortorientierten Ansätze vorgeschlagenen Ad-Hoc-Verfeinerungen dieser Anforderung nicht gerecht werden (\Rightarrow These 1). Die Ergebnisse der Betrachtungen in Kapitel 5 legen den Schluß nahe, daß die identifizierten Grenzen der wörterbuchbasierten Kategorisierung auf der Grundlage aktueller algorithmischer, robuster Textanalyseverfahren der Computerlinguistik überwunden werden können (\Rightarrow These 2).

Teil II der Arbeit hat eine algorithmische Detailstudie zu einem speziellen qualitativen Teilproblem des Codierschritts einer verfeinerten Themenanalyse zum Gegenstand: die referentielle Interpretation sprachlicher Ausdrücke (insbesondere Pronomen). Kapitel 6 gibt eine Übersicht des Phänomens der Referentialität und definiert einige grundlegende Begriffe; der Untersuchungsgegenstand wird eingegrenzt auf das zentrale Teilproblem der *Koreferenzanalyse* von Ausdrücken mit *Objektbezug*, das insbesondere die Aufgabe der Interpretation der in Texten vergleichsweise häufig anzutreffenden Personalpronomen in dritter Person umfaßt. In Kapitel 7 wird eine Brücke von der theoretischen Diskussion hin zur operationalen Perspektive geschlagen. Unterschiedliche in der Literatur vorgeschlagene Strategien (Restriktionen und Präferenzen) für die computergestützte Koreferenzanalyse werden diskutiert und bezüglich einfließendem Wissen, Relevanz und Algorithmisierbarkeit aufgeschlüsselt. Hierdurch werden die Voraussetzungen für die Realisierung tatsächlich anwendungstauglicher Analyseverfahren geschaffen. Ein vergleichbares Ziel liegt der Synopse und Diskussion der wichtigsten bekannten Verfahren zur Koreferenzanalyse in Kapitel 8 zugrunde. Die Ergebnisse dieser Bestandsaufnahmen bilden den Ausgangspunkt für den Entwurf robuster Algorithmen, die den Anforderungen eines Einsatzes im Rahmen einer verfeinerten computergestützten Inhaltsanalyse genügen. Gegenstand von Kapitel 9 ist die Entwicklung des Basisalgorithmus. Die Schwerpunkte liegen auf der Auswahl und adäquaten Operationalisierung nutzbringender Einzelstrategien sowie ferner auf der Definition von Datenstrukturen, die den Anforderungen an die Repräsentation referentieller Information entsprechen. Kapitel 10 verfeinert den Basisalgorithmus und die Datenstrukturen im Hinblick auf die Robustheitszielsetzung. Zunächst wird das bisher nur intuitiv begründete Verständnis des Robustheitsbegriffs auf eine definitorische Basis gestellt. Zentrales Problem ist der Umgang mit unvollständigen syntaktischen Beschreibungen. Mehrere verfeinerte Versionen des Basisalgorithmus werden entwickelt, die der Robustheitsanforderung in graduell unterschiedlicher Form gerecht werden; eine theoretisch adäquate, maximal robuste Variante bewerkstelligt die Verschränkung der qualitativen Teilprozesse von syntaktischer und referentieller Analyse (\Rightarrow These 4). In Kapitel 11 werden Implementierung und Evaluation eines anwendungstauglichen Computersystems zur Koreferenzanalyse, des Systems ROSANA, beschrieben und diskutiert. Grundlage der Evaluation ist die formale Definition dreier unterschiedlicher Disziplinen der Erschließung referentiellen Textinhalts, in denen die maschinellen Interpretationsleistungen per Abgleich mit den Ergebnissen einer intellektuellen Interpretation numerisch bewertet werden. Anhand einer Evaluation auf zwei unterschiedlichen Textkorpora wird der Nachweis der hohen Interpretationsgüte von ROSANA insbesondere in der Disziplin der Interpretation von Pronomen erbracht (\Rightarrow These 3).

Kapitel 12 diskutiert die Ergebnisse der Untersuchungen von Teil I und Teil II in bezug auf das Kernpostulat der Dissertation (\Rightarrow These 5).

Teil I

Inhaltsanalyse - Verfahren, Theorie und Algorithmen

Kapitel 2

Inhaltsanalyse

Als Ausgangspunkt für die Diskussion der Kernfrage, welche Möglichkeiten bestehen, die Qualitative Analyse durch den Einsatz von Softwaresystemen zu unterstützen bzw. durch Algorithmen zu explizieren, soll die Methode der sozialwissenschaftlichen Inhaltsanalyse näher betrachtet werden. Ein Blick in die Literatur zeigt, daß unter diesem Begriff eine Vielzahl unterschiedlicher Analysetechniken subsumiert werden. Es erweist sich als notwendig und zweckmäßig, eine genauere Abgrenzung zu erarbeiten und eine Übersicht der grundlegenden Begriffe zu geben.

Als Ergebnis dieser Betrachtungen wird die *Themen-* oder auch *Kategorienanalyse* als das inhaltsanalytische Verfahren von zentraler Relevanz identifiziert. Gemäß traditionellem Verständnis ist die Kategorienanalyse *quantitativ* ausgerichtet: Per Konstruktion numerischer Deskriptionen wird vom Ausgangsmaterial abstrahiert; die Auswertung geht per statistischer Verfahren vonstatten. Gleichwohl stellt dieses spezifische Inhaltsanalyseverfahren den Anknüpfungspunkt für die Problemstellungen der vorliegenden Arbeit dar. Zum einen, weil - wie bereits in der Einleitung skizziert - der Großteil der Einzelschritte einer quantitativen Kategorienanalyse von inhärent qualitativer Gestalt ist; nicht zuletzt gilt dies für den zentralen Schritt der Kategorisierung, für den in der computergestützten Inhaltsanalyse bereits erste, grobkörnige Ansätze zur Algorithmisierung existieren, deren Verfeinerbarkeit im Rahmen der vorliegenden Arbeit diskutiert werden soll. Zum anderen, weil - wie sich zeigen wird - auch die neueren, qualitativ ausgerichteten Verfahren auf ein (hinreichend generalisiertes) sequentielles Ablaufschema der Kategorienanalyse bzw. auf Teilschritte davon abbildbar sind.

Im folgenden wird daher der Begriff der Kategorienanalyse in einem weitergefaßten Sinne als üblich, d.h. *nicht* mit implizitem Bezug auf quantitativ-reduktionistische Auswertestrategien, Verwendung finden. Entsprechend allgemein, d.h. *nicht* als Gegenpol zur Kategorienanalyse, soll von Qualitativer Analyse gesprochen werden.

2.1 Die sozialwissenschaftliche Inhaltsanalyse

Die Methode der Inhaltsanalyse blickt auf eine mehr als fünfzigjährige Geschichte zurück. Entsprechend groß ist die Vielfalt konkreter Verfahren, die sich z.T. recht stark im Hinblick auf Anwendungszielsetzungen und Analysetechniken unterscheiden. Es erscheint daher angebracht,

ein allgemeines Verständnis der inhaltsanalytischen Methode zu entwickeln. Zunächst sollen einige der bekanntesten Charakterisierungsversuche gegenübergestellt und diskutiert werden, um eine konsistente Begriffsgrundlage zu schaffen. In einem zweiten Schritt wird ein generisches Prozeßmodell erarbeitet, das den Ablauf der wichtigsten inhaltsanalytischen Verfahren schematisch beschreibt und sich als geeignet erweisen wird, die qualitativen Teilprobleme der Inhaltsanalyse präzise zu lokalisieren.

2.1.1 Definitionen und grundlegende Begriffe

Die klassische Charakterisierung stammt aus den inhaltsanalytischen Arbeiten Berelsons auf dem Gebiet der Kommunikationsforschung ([Bere52]):

Definition 2.1 (Content Analysis, nach Berelson) *Content analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication.*

Die Festschreibung der *quantifizierenden Vorgehensweise* und die Beschränkung des Analyseziels auf den sog. *manifesten Inhalt*, d.h. die Ausblendung der sog. *latenten Inhaltsebene*, worunter die pragmatische, an subjektive Verstehensprozesse gebundene Dimension verstanden wird, zielen in dieselbe Richtung: die Gewährleistung einer optimalen *Objektivität* und damit der idealen intersubjektiven Reproduzierbarkeit.

Die Definition Berelsons ist Kritik von mehreren Seiten ausgesetzt. Zunächst einmal wird das Ziel der Quantifizierung insbesondere deshalb als problematisch angesehen, da durch die hiermit einhergehende numerische Reduktion der Bezug zwischen Daten und originärem textuellen Kontext verloren gehe, womit die Validität der Analyseergebnisse potentiell tangiert werde. Bereits Kracauer entwickelte ein qualitatives Gegenmodell der Inhaltsanalyse ([Krac52]). Darüberhinaus ist die Abgrenzung zwischen Qualitativer und Quantitativer Inhaltsanalyse fragwürdig, denn auch in bezug auf das Berelsonsche Analysemodell erscheint es nicht als gerechtfertigt, von einer rein quantitativen Methode zu sprechen. Mittlerweile wird allgemein akzeptiert, daß zwischen Qualitativer und Quantitativer Analyse kein Verhältnis des Ausschlusses, sondern der *Komplementarität* besteht - m.a.W. sie *ergänzen* sich gegenseitig (Barton und Lazarsfeld: [BaLa55], George: [Geor59], Rust: [Rust80]). Am Beispiel der Kategorienanalyse als konkrete Klasse inhaltsanalytischer Verfahren wird dieser Sachverhalt vertiefend diskutiert werden (vgl. Abschnitt 2.1.3).

Ein zweiter zentraler Punkt betrifft das *unterliegende Kommunikationsmodell*, das sich auf ein behavioristisches *Reiz-Reaktions-Schema* zurückführen läßt, in dem der Kommunikationsinhalt übervereinfachend in einer elementaren Abfolge von Zeichen - aufgefaßt als *Stimuli* - lokalisiert wird, für die angenommen wird, daß deren Wirkung auf den Rezipienten vollständig vorhergesagt werden könne (vgl. Lasswell: [Lass27]).¹ Entgegen der hiermit verbundenen Erwartung erscheint es aus kommunikationstheoretischer Perspektive keineswegs allgemein gerechtfertigt, von einer vollständigen Objektivierbarkeit inhaltlicher Eigenschaften der Zeichen, d.h. von der Explizierbarkeit von "*Durchschnittsbedeutungen*" im eingangs beschriebenen Sinne, auszugehen. Demnach erschließt sich die Bedeutung i.a. nicht unter alleinigem Rekurs auf die kommunizierten Symbole, sondern in Relation zum jeweiligen Rezipienten in dessen spezifischer Disposition

¹Lasswell spricht in diesem Zusammenhang von "*quantitative semantics*" ([Lass65]).

zum Zeitpunkt der Interpretation; die pragmatische Handlungsrelevanz bzw. die *Wirkung* einer Abfolge von Zeichen ist prinzipiell vom jeweiligen Kommunikationskontext bzw. vom Vorwissen abhängig, das u.a. auch darüber entscheidet, ob die Kommunikation gelingt, d.h. der Rezipient den Kommunikator überhaupt *versteht*. Für die Berelsonsche Inhaltsanalyse heißt dies, daß tiefere semantisch-konnotative bzw. pragmatische Kommunikationsebenen nicht erschlossen werden. Da das zu analysierende Kommunikat entsprechend den Intentionen des Kommunikators sowie im Hinblick auf das Vorwissen des bzw. der spezifischen Adressaten in der ursprünglichen Kommunikationssituation konzipiert ist, entsteht für den Inhaltsanalytiker die theoretische Bedingung, sich in die Sichtweise der originären Kommunikationsteilnehmer hineinzuversetzen, um die Voraussetzungen für eine möglichst vollständige Erschließung der Kommunikationsinhalte zu erfüllen. Somit erscheint die klassische inhaltsanalytische Methode Berelsons nur für bestimmte Typen von Anwendungen wie z.B. die unmittelbar auf der Ebene der Zeichen angesiedelte Propagandatechnik-Analyse adäquat, da sich die analyserelevanten Inhaltsebenen aus kommunikationspragmatischen Gründen oftmals nur auf der Basis *interpretativer Inferenzen* erschließen.²

Aus Sicht der *sozialwissenschaftlichen* Inhaltsanalyse ist demnach die Einschränkung des Analyseziels auf die *Beschreibung* des sog. "*manifesten Inhalts*" insofern als problematisch anzusehen, als auf der Grundlage der Berelsonschen *Content Analysis* der zentrale Anspruch der sozialwissenschaftlichen Inhaltsanalyse, die Erschließung der als extratextueller Kontext der Kommunikation konstituierten *sozialen Wirklichkeit*, i.d.R. nicht eingelöst werden kann. In der späteren Literatur finden sich erweiterte Definitionen der Inhaltsanalyse, die dieser Zielsetzung Rechnung tragen. So definiert Krippendorff ([Krip80], S. 21)

Definition 2.2 (Content Analysis, nach Krippendorff) *Content analysis is a research technique for making replicable and valid inferences from data to their context.*

und, hierauf aufbauend, Merten ([Mert95], S. 59)

Definition 2.3 (Inhaltsanalyse, nach Merten) *Inhaltsanalyse ist eine Methode zur Erhebung sozialer Wirklichkeit, bei der von Merkmalen eines manifesten Textes auf Merkmale eines nichtmanifesten Kontextes geschlossen wird.*

Der extratextuelle Kontext wird hierbei grob untergliedert in Kommunikator, Rezipient und Situation der (in Textform vorliegenden) Kommunikation (vgl. Abbildung 2.1). Im Rahmen der

²Aus der kommunikationstheoretischen Kritik der Berelsonschen Inhaltsanalyse folgt jedoch keineswegs die prinzipielle Untauglichkeit maschineller Inhaltsanalysen. Es kommt dasselbe Argument wie gegenüber der eingangs umrissenen hermeneutischen Position zum Tragen: Die Untersuchungen in der vorliegenden Arbeit werden belegen, daß für bestimmte nichttriviale Teilprobleme der qualitativen Textinhaltserschließung bereits genügend *Vorwissen* algorithmisch expliziert werden kann; darüberhinaus wird keine prinzipielle Schranke betreffend die Explizierbarkeit des menschlichen Verstehensintergrunds gesehen. Obwohl noch bei weitem nicht alle intellektuellen Interpretationsleistungen auf dem gegenwärtigen Stand der Technik maschinell erbracht werden können, ist bereits heute erheblich mehr möglich als mit der klassischen Berelsonschen Inhaltsanalyse. Dies betrifft natürlich in erster Linie diejenigen Inhaltsebenen, für deren Erschließung weniger auf kontextuelles, sondern primär auf *innersprachliches* Wissen zurückzugreifen ist (Beispiel: die Interpretation von Pronomen). Eine formale, trennscharfe Abgrenzung zwischen manifesten (i.S.v. innersprachlich determinierten) und latenten (i.S.v. kontextuell determinierten) Inhalten erscheint jedoch problematisch, denn auch etwa für das Problem der Pronomeninterpretation lassen sich Beispiele angeben, in denen der Kommunikationskontext die Interpretationsoptionen determiniert. Entscheidend für den Erfolg einer algorithmischen Inhaltsanalyse ist, daß derartige Fälle (relativ zum gegenwärtig explizierbaren sprachlichen bzw. außersprachlichen Vorwissen) *selten* sind.

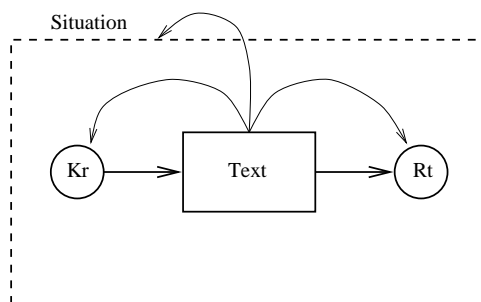


Abbildung 2.1: Inferenzziele der Inhaltsanalyse nach Merten ([Mert95], S. 56)

sozialwissenschaftlichen Inhaltsanalyse sind es die *gesellschaftlichen Situationsbedingungen*, die das Ziel der inferentiellen Theoriebildung bilden: normative Gegebenheiten, Wissens-, Glaubens- und Wertbestände, politische und ökonomische Rahmenbedingungen usw. Diese *soziale Wirklichkeit* liegt im Text i.d.R. nur *latent* vor und ist folglich auf der Grundlage einer interpretativen Inferenz zu erschließen.

In Abbildung 2.2 sind die unterschiedlichen Skopoi der Definition 2.1 einerseits sowie der Definitionen 2.2 und 2.3 andererseits gegenübergestellt. Während die Berelsonsche *Content Analysis*

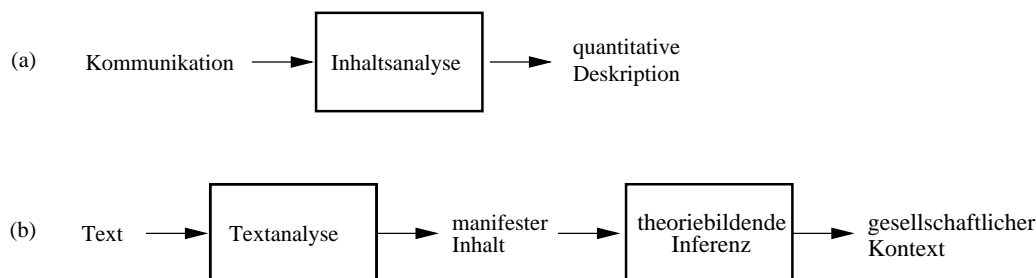


Abbildung 2.2: Skopus der Inhaltsanalyse gemäß (a) Berelson bzw. (b) Krippendorff/Merten

per definitionem alleine das Ziel verfolgt, eine (quantitative) *Beschreibung* des sog. manifesten Inhalts des Untersuchungsgegenstands anzufertigen, umfassen die Krippendorffsche sowie die Mertensche Inhaltsanalyse sowohl diese deskriptive Phase als auch den sich daran anschließenden Schritt der inferentiellen Theoriebildung über den Daten. Der eingeschränkte inhaltsanalytische Skopus in den Arbeiten Berelsons wird verständlich, wenn man berücksichtigt, daß dort neben der rein deskriptiven Inhaltsanalyse zwei weitere Methoden - *Intentions-* und *Wirkungsanalyse* - vorgestellt werden, die inferentiell-theoriebildend ausgerichtet sind und somit unter den weitergefaßten Begriff fallen; die Quantitative Inhaltsanalyse dient hier nur als Vorstufe oder aber sie wird eigenständig eingesetzt in den Spezialfällen, in denen sich die zu erschließenden gesellschaftlichen Situationsbedingungen ohne weitere Inferenz erschließen, d.h. *manifest im Text* vorliegen (vgl. auch Merten: [Mert95], S. 55).³ Folglich ist festzuhalten, daß sich der Inhaltsanalyseprozeß

³Warum Berelson der nichtinferentiellen Inhaltsanalyse eine derart große Eigenständigkeit einräumt, wird mit Blick auf die Entwicklungsgeschichte der *Content Analysis* offenbar: Die Propagandatechnik-Analyse, die ja der zentrale Ausgangspunkt der Entwicklung der inhaltsanalytischen Methode war, kann als Sonderfall angesehen werden, in der die relevanten Zielstrukturen identisch im Text vorzufinden sind.

in zwei Phasen unterteilen läßt: Textanalyse und theoriebildende interpretative Inferenz. Auch die theoriebildende Erschließung des extratextuellen Kontexts aus dem Text baut jedoch darauf auf, daß sich die relevanten gesellschaftlichen Situationsbedingungen zumindest *in wesentlichen Teilen* im Textinhalt *widerspiegeln*. Von der uneingeschränkten Gültigkeit dieses sog. *Repräsentationsmodells* auszugehen (*Reflexionshypothese*), ist nur in bestimmten Sonderfällen zulässig - beispielsweise, wenn die Inferenz auf den stilistischen Eigenschaften des Kommunikats aufbaut, oder wenn es sich um Texte mit Wahrheitsanspruch handelt. In vielen Fällen liegen die Verhältnisse komplizierter: Entsprechend dem *Instrumentalmodell* üben die Medien zumindest in bestimmtem Umfang eine Vorreiterfunktion für gesellschaftliche Prozesse aus (*Kontrollhypothese*). In diesem Fall verbietet sich natürlich ein unmittelbarer Analogieschluß auf den Kontext. Die Probleme, die hier bestehen, sind mit denen vergleichbar, die den Ausgangspunkt der kommunikationstheoretischen Kritik der klassischen inhaltsanalytischen Methode Berelsons bilden (vgl. o.); demnach sind auch die Intentionen des Kommunikators sowie darüberhinaus das vom Kommunikator supponierte Vorwissen der Rezipienten dafür ausschlaggebend, welche Inhalte im Kommunikat *manifest* werden. Aus diesem sowie aus anderen Gründen besteht für inhaltsanalytische Verfahren jedweder Art die Obligation des Nachweises der Validität der Inferenz auf den Kontext (*„externe Gültigkeit“*) per sog. *Außenvalidierung*.

Eine hiervon zu unterscheidende Facette der Validität inhaltsanalytischen Vorgehens betrifft die Textanalysephase, in deren Rahmen eine i.a. reduzierte Deskription⁴ des Textes erzeugt wird. Um die Objektivität des Verfahrens zu gewährleisten, hat das Erzeugen dieser Deskriptionen dem Anspruch der (insbesondere intersubjektiven, codiererunabhängigen) *Reproduzierbarkeit* zu genügen. Für diese sog. *interne Gültigkeit* wurde auch die Bezeichnung *Reliabilität* geprägt.

Bezüglich der Definition von Merten bleiben zwei weitere Punkte zu erörtern. Zum einen ist von der inhaltsanalytischen *Methode* die Rede, die als Sammelbegriff für eine große Anzahl konkreter *Verfahren* angesehen wird, die im Hinblick auf ihre spezifischen Ausrichtungen von höchst unterschiedlicher Gestalt sind. Da sich die Differenzierung zwischen Methode und Verfahren bewährt, soll sie in der vorliegenden Arbeit übernommen werden. Dieser Punkt ist natürlich gerade im Hinblick auf das Ziel von Teil I der Arbeit, der Gegenüberstellung inhaltsanalytischer Verfahren und linguistisch-symbolischer algorithmischer Ansätze, von Bedeutung, da die relevanten Punkte nur verfahrens-relativ erörtert werden können.

Zum anderen spricht Merten von der Inhaltsanalyse als eine Methode zur *Erhebung* gesellschaftlicher Rahmenbedingungen, die somit als *sekundäres Erhebungsinstrument* bezeichnet werden könne ([Mert95], S. 86 f.). Der hier unterliegende Erhebungsbegriff, der insbesondere die theoriebildende Kontextinferenz subsumiert, geht wesentlich über das hinaus, was Erhebungsverfahren im Sinne von Datenerfassungsverfahren leisten, wie sie beispielsweise auch im Rahmen der qualitativen Sozialforschung diskutiert werden ([Mayr96]). In bezug auf Abbildung 2.2 läßt sich diese letztere Art von Erhebung als Vorstufe auffassen, in deren Rahmen die Vertextung des Erhebungsgegenstands vonstatten geht.

Die Methode der sozialwissenschaftlichen Inhaltsanalyse kann somit zusammenfassend wie folgt charakterisiert werden:

1. *Eingabe* des Inhaltsanalyseprozesses ist ein *Text* allgemeinsten Zuschnitts, d.h. der Text kann entweder das originäre, vom Kommunikator selbst gewählte Medium oder aber das Ergebnis einer Verschriftung der Kommunikation per Anwendung eines adäquaten Aufbereitungsverfahrens sein.

⁴Die Qualifikation der erzeugten Beschreibung als *reduziert* gilt v.a. für die klassische, quantitativ orientierte Inhaltsanalyse. Im Falle einer qualitativ ausgerichteten Analyse fällt dieses Attribut i.a. weg.

2. Die eigentliche Inhaltsanalyse ist *nichtreaktiv*, da sie räumlich und zeitlich getrennt vom zu analysierenden Kommunikationsvorgang von einem a priori fixierten Text ausgeht.⁵
3. Ziel der Inhaltsanalyse ist die Ermittlung situational gegebener gesellschaftlicher Bedingungen per *theoriebildender Inferenz*.
4. In Abhängigkeit der jeweils relevanten gesellschaftlichen Situationsbedingung sowie der zur Anwendung kommenden Analysemittel lassen sich unterschiedliche *Verfahren* der inhaltsanalytischen Methode unterscheiden.
5. Der Inhaltsanalyseprozeß läßt sich grob untergliedern in eine *deskriptive Phase*, in der der Text auf Teilbereiche des sog. *manifesten Inhalts* reduziert⁶ wird, und eine *inferentielle, interpretative Phase* der Theoriebildung über die jeweilig relevanten Teilaspekte der sozialen Situationsbedingungen.⁷
6. Im Hinblick auf das Objektivitätsziel der Inhaltsanalyse sind bestimmte, i.a. qualitativ ausgerichtete *Gültigkeitsnachweise* zu erbringen, die einerseits die Reproduzierbarkeit der Vercodung (*Reliabilität*) und andererseits die Zulässigkeit der Kontextinferenz (*externe Validität*) betreffen.

und - im Vorgriff auf nachfolgende, präzisierende Ausführungen -:

7. Eine Dichotomisierung "*Qualitative vs. Quantitative Inhaltsanalyse*" im Sinne zweier sich gegenseitig ausschließenden Methoden ist inadäquat: Auch in Inhaltsanalyseverfahren, deren Textanalyse ausschließlich quantitative Deskriptionen erzeugt, sind *stets qualitative Verfahrenskomponenten* involviert.

2.1.2 Inhaltsanalyse als Kategorienanalyse

Da im folgenden näher expliziert werden soll, an welcher Stelle der Inhaltsanalyse qualitative Gesichtspunkte ins Spiel kommen, ist zunächst näher einzugrenzen, auf welche spezifischen inhaltsanalytischen *Verfahren* Bezug genommen wird. Die umfassende Typologie von Merten ([Mert95], S. 119 ff.) erweist sich als geeigneter Ausgangspunkt. Merten klassifiziert die inhaltsanalytischen Verfahren entlang zweier Dimensionen:

- *Ziel der Theoriebildung*: auf welche der drei Grundelemente des gesellschaftlichen Kontexts - Kommunikator, Rezipient oder Situation - geschlossen werden soll;
- *semiotische Ebene*, auf der die *Mittel der Analyse* einzuordnen sind; entsprechend einer in Anlehnung an die von Morris postulierten asymmetrischen relationalen Abhängigkeiten zwischen Syntax, Semantik und Pragmatik entstehen so insgesamt sechs unterschiedliche Klassen.

⁵Der Vorschlag von Merten, die Inhaltsanalyse mit Blick auf die pragmatischen Auswirkungen der vom Kommunikator intendierten Wirkung als "*interpretativ-reaktiv*" zu bezeichnen ([Mert95], S. 92 ff.), wird hier nicht aufgegriffen.

⁶Dies gilt wiederum v.a. für die Quantitative Inhaltsanalyse klassischen Zuschnitts, wie sie den obigen Definitionen zugrunde liegt.

⁷Es wird sich allerdings im folgenden sehr schnell zeigen, daß diese Trennung aus terminologischer Sicht *cum grano salis* zu nehmen ist, da auch die allgemein als deskriptiv bezeichneten inhaltsanalytischen Verfahren auf *Interpretationsleistungen* des Inhaltsanalytikers aufbauen.

Als das am häufigsten angewandte inhaltsanalytische Verfahren identifiziert Merten die sog. *Themen-* oder auch *Kategorienanalyse*. In dessen Basisversion wird es in die Klasse der *semantisch-semantischen Analyse des Kommunikators/der Situation* eingeordnet. Unter “*semantisch-semantisch*” ist zu verstehen: Die mittelbar zu erschließenden Zielgrößen - *Themen* - sind semantische Einheiten (\Rightarrow “*-semantisch*”); die hierfür heranzuziehenden Einheiten sind ebenfalls semantisch (\Rightarrow “*semantisch-*”).⁸ Die Themenanalyse wird außerdem als Basis einer Reihe weiterer, auf höheren semiotischen Ebenen angesiedelten Verfahren angesehen, in denen zusätzliche qualitative Gesichtspunkte eine Rolle spielen. Merten nennt hier die auf der semantisch-pragmatischen Ebene angesiedelten Verfahren der *Objektivitäts-* und *Symbolanalyse*, in denen neben einer rein kategorialen Aufschlüsselung unterschiedlicher Themen zusätzlich *Bewertungsaspekte*, d.h. die Polarität der vorgefundenen Kategorieninstanzen Berücksichtigung finden. Merten gelangt zu der Einschätzung, daß es sich bei der Themenanalyse und ihren unterschiedlichen anwendungsbezogenen Verfeinerungen insgesamt um das “*Standardverfahren für die Analyse von Massentexten - von Medien also*” handelt ([Mert95], S. 146f.).

Der Begriff “*Massentext*” ist hierbei weniger in linguistischen Strukturmerkmalen oder im Genre (d.h. in der Textsorte) der zu verarbeitenden Dokumente verankert, sondern steht vielmehr in unmittelbarem Zusammenhang mit der Komplexität der zu beantwortenden inhaltsanalytischen Fragestellungen. Eine i.d.R. mit einer erheblichen Reduktion der vorliegenden Information verbundene kategorisierende und evtl. numerisch abstrahierende (quantifizierende) Analyse bietet sich in erster Linie dann an, wenn eine Forschungsfrage vorliegt, die auf der Basis eines datenreduzierenden Erhebungsinstruments beantwortbar erscheint.⁹ Aus forschungspragmatischen Gründen wird dies freilich vor allem (jedoch nicht ausschließlich) dann der Fall sein, wenn eine große Datenmenge vorliegt und es somit auch formal gerechtfertigt ist, von Massendaten- oder Massentextverarbeitung zu sprechen.

Abbildung 2.3 schlüsselt das Grundschema der Analyse auf, das allen themenanalytischen Verfahren gemein ist. (1) Ausgangspunkt aller Verfahren dieser Klasse ist die theoriegeleitete, an den Zielen der Untersuchung orientierte Festlegung eines adäquaten *Kategoriensystems* K_1, \dots, K_n . Hierbei sind bestimmte Kriterien zu erfüllen, so z.B. die wechselseitige Exklusivität der Kategorien sowie deren eindeutige Definition. Schritt (2) der Themenanalyse bildet die Erhebungseinheiten, in die der Text untergliedert ist (Wörter, Absätze, Artikel o.ä.) auf einzelne Kategorien ab; dies entspricht dem ersten Teil der Textanalyse in Abbildung 2.2 (b). Als Zwischenergebnis liegt eine Menge von Instanzen I_1, \dots, I_m der Kategorien vor, die nun in Schritt (3) - dem zweiten Teil der Textanalyse - ggf. weiter reduziert wird, um zum Zwischenergebnis, einer Deskription des sog. manifesten Textinhalts, zu gelangen. In der klassischen Themenanalyse erfolgt auf dieser Ebene eine *quantitative Reduktion*, in deren Rahmen die Instanzen (Erhebungseinheiten) der unterschiedlichen Kategorien *ausgezählt* und u.U. unter Rekurs auf ein geeignetes Maß der *Größe* der instanzierenden Erhebungseinheiten zusätzlich *gewichtet* werden. Schritt (4) - die typadäquate Auswertung der vorliegenden Deskriptionen - stellt den ersten Schritt der Erschließung der situationalen Bedingungen dar. Im Falle der elementaren, d.h. nach Häufigkeiten quantifizierenden Themenanalyse sind es Prozeduren der statistischen Auswertung, die hier zum Einsatz kommen. Hiermit ist die theoriebildende Inferenz allerdings noch nicht vollzogen - diese bleibt in jedem Falle ein *interpretativer Prozeß* (Schritt (5)), der ausschließlich im Hinblick auf die Forschungsfrage und unter sensibler Berücksichtigung der Bedingungen der Außenvalidität

⁸Ein vielleicht besser nachvollziehbares Beispiel: eine *Verständlichkeitsanalyse*, die auf der Komplexität syntaktischer Strukturen aufbaut, ist auf der syntaktisch-pragmatischen Ebene einzuordnen, da die mittelbare Zielgröße - Verständlichkeit - ein pragmatisches Element ist.

⁹In besonderem Maße gilt dies für die klassische maschinelle Inhaltsanalyse, deren Codierschritt auf elementaren Zeichenkettenvergleichen basiert und daher einen entsprechend eingeschränkten Skopus aufweist (vgl. u.).

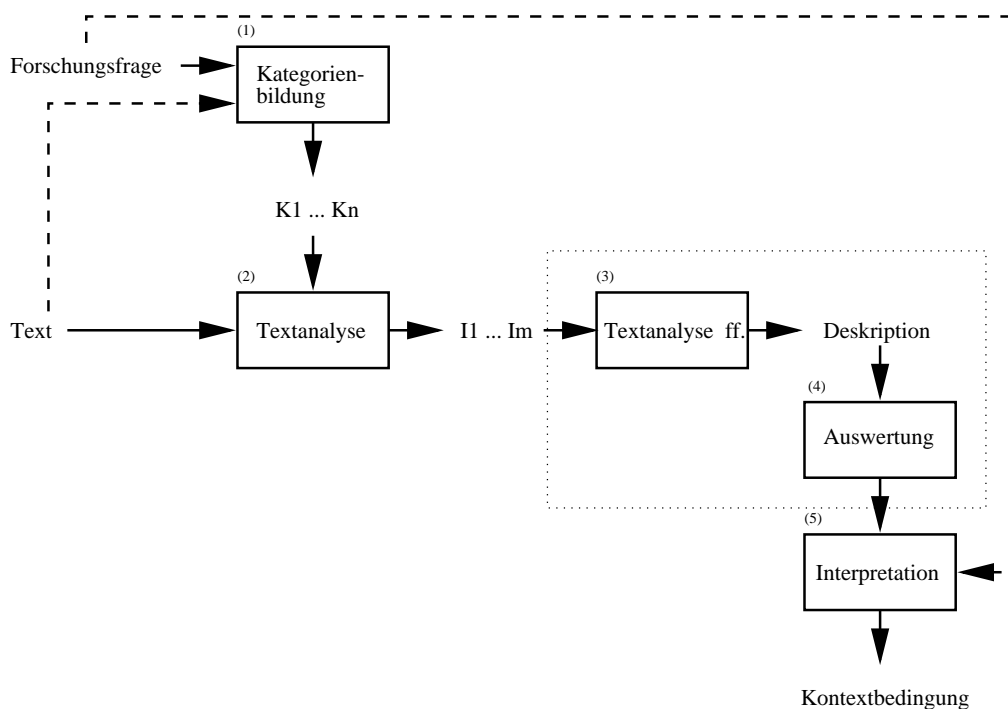


Abbildung 2.3: Die Kategorienanalyse als zentrales Verfahren der Inhaltsanalyse

vonstatten gehen kann.¹⁰

Die Weiterentwicklungen der Themenanalyse - Objektivitäts-, Symbol- und Bewertungsanalyse - unterscheiden sich insbesondere durch eine entsprechende Verfeinerung der deskriptionsbezogenen Schritte (3) und (4). Werden z.B. zusätzlich Bewertungsaspekte in die Analyse miteinbezogen, so kann dies seinen Niederschlag auf Seiten der statistisch-quantitativen Auswertung dergestalt finden, daß an Stelle der eindimensionalen, univariaten Analyse eine bi- oder multivariate Analyse tritt, die neben den Instanzenhäufigkeiten zusätzlich weitere Dimensionen der Bewertung einbezieht.

Im folgenden soll, wenn von der sozialwissenschaftlichen Inhaltsanalyse die Rede ist, die kategoriensystembasierte Analyse gemäß Abbildung 2.3 verstanden werden. Diese Annahme entspricht dem allgemeinen Verständnis, das dem Inhaltsanalysebegriff in der einschlägigen Literatur unterliegt. So betrachtet z.B. Früh ebenfalls die Themenanalyse sowie ihre wertungs- und argumentationsbezogenen Weiterentwicklungen ([Früh91], S. 76 ff.). Auch Lisch und Kriz betrachten die Methode der Inhaltsanalyse unter dem Gesichtspunkt der Informationsreduktion, die per Kategorisierung zu bewerkstelligen ist, ohne sich allerdings explizit darauf festzulegen, auf welche Analyseeinheiten die Kategoriedefinition abzielen hat ([LiKr78], S. 69 ff.).¹¹ Und schließlich lassen sich auch zumindest einige der übrigen Verfahren, die Merten in seiner Synopse vorstellt, auf dieses grundlegende Analyseschema abbilden - der Unterschied manifestiert sich in der abweichenden Gestalt des Kategorienschemas. So tritt beispielsweise im Verfahren der *syntaktischen*

¹⁰Im Falle der statistischen Auswertung führt man sich diesen Gesichtspunkt am einfachen Beispiel einer statistisch signifikanten Korrelation zweier Faktoren vor Augen, die ja ohne Weiteres noch keinen Kausalzusammenhang impliziert.

¹¹Die in [LiKr78], S. 72 ff. gegebenen Beispiele scheinen jedoch allesamt in die Klasse der themenanalytischen Kategoriensysteme zu fallen.

Komplexitätsanalyse ([Mert95], S. 139 ff.) eine grammatiktheoriebezogene Strukturklassifikation an die Stelle der themenorientierten Kategorisierung des Standardverfahrens; die weiteren, deskriptionsbezogenen Analyseschritte fallen hier wiederum unter das quantitative Paradigma, so daß die Entsprechung mit dem Standardschema der Inhaltsanalyse gewährleistet ist.

Gemäß diesem traditionellen Verständnis ist die Kategorien- oder Themenanalyse *quantitativ* ausgerichtet, d.h. sie basiert auf einer numerischen, reduktiven Deskription des Textinhalts. Es ist jedoch prinzipiell denkbar, daß an deren Stelle - insbesondere im Rahmen einer *qualitativ* orientierten Analyse - eine allgemeinere symbolische (nichtnumerische), u.U. sogar nichtreduktive, *vertiefende* Deskription des Textinhalts erzeugt wird, auf die ein nichtstatistisches, interpretatives Auswertungsverfahren zur Anwendung gelangt. Auch die neueren, beispielsweise von Mayring diskutierten Verfahren der Qualitativen Analyse ([Mayr96]) sind - wie sich zeigen wird - unter Bezugnahme auf das Prozeßschema 2.3 klassifizierbar. Unter Würdigung dieser Gesichtspunkte erscheint es somit gerechtfertigt, die Kategorienanalyse als grundlegenden Bezugspunkt für die weiteren Betrachtungen zu wählen.

2.1.3 Qualitative vs. Quantitative Inhaltsanalyse

Anhand von Abbildung 2.3 läßt sich die Inadäquatheit der Dichotomisierung "*Qualitativ vs. Quantitativ*" in bezug auf die Klasse der kategorienanalytischen Verfahren illustrieren. Zwar basiert die klassische Kategorienanalyse auf einem *quantitativen Kern* (Kasten umfassend Schritte (3) und (4)), jedoch existieren in jedem Fall Verarbeitungsschritte, die inhärent qualitativ sind (Schritte (1), (2) und (5)). Insbesondere ist der elementare Klassifikations- oder auch Kodierschritt (2) stets auf der qualitativen Ebene angesiedelt: Eine Datenreduktion per Abbildung auf eine Nominalskala - nichts anderes geschieht ja im Rahmen einer Kategorisierung - ist das elementare Beispiel einer qualitativen Operation. Speziell für dieses somit qualitative Teilproblem wird im weiteren Verlauf der Arbeit nach Möglichkeiten einer verfeinerten algorithmischen Explikation Ausschau gehalten.

Anders ausgedrückt: Wenn im Rahmen der klassischen Diskussion von *Quantitativer* Inhaltsanalyse gesprochen wird, so ist damit eine spezifische Ausgestaltung der deskriptionsbezogenen Kernschritte (3) und (4) gemeint. Wird an dieser Stelle ausgezählt bzw. in irgendeiner sonstigen Weise quantitativ gemessen und darauf aufbauend z.B. statistisch ausgewertet, hat man es - ungeachtet der qualitativen Orientierung der übrigen Schritte - mit *quantitativer* Analyse zu tun. Geschieht die Auswertung auf subjektiv-interpretative Weise, d.h. dient die Kategorisierung in Schritt (2) ausschließlich im Hinblick auf eine Vorabsystematisierung des immer noch vollständig zugänglichen Materials, so ist die Inhaltsanalyse *qualitativ* ausgerichtet.

Die klassische Quantitative Kategorienanalyse steht somit nicht im ausschließenden Gegensatz zur Qualitativen Analyse. Vielmehr besteht eine Beziehung gegenseitiger Ergänzung. Mit Blick auf den gemeinhin bestehenden Austauschverhältnis zwischen Validität und Reliabilität bringt Huber diese *Komplementarität* der beiden methodischen Ausrichtungen auf den Punkt, wenn er konstatiert ([Hube89], S. 41):

"Quantität für sich ist sinnlos, Qualität für sich genommen bleibt folgenlos."

2.2 Qualitative Aspekte der Inhaltsanalyse

Als Ausgangspunkt für die Diskussion des Themas der Computerunterstützung des Inhaltsanalyseprozesses im folgenden Kapitel soll nun eine detailliertere Diskussion der qualitativen Verfahrensschritte der Kategorienanalyse durchgeführt werden, um zu einem präziseren Verständnis der jeweiligen Explikationsproblematik zu gelangen, die sich im Rahmen von Versuchen der algorithmischen Operationalisierung stellt. Im Mittelpunkt der Betrachtungen stehen die theoretischen Bedingungen sowie die Explikationsansätze in der Gestalt interpretativer Verfahrensentwürfe, wie sie in Arbeiten zur Qualitativen Analyse vorgestellt werden (vgl. z.B. Mayring: [Mayr96]).

2.2.1 Festlegung des Kategoriensystems

Die Festlegung des Kategoriensystems ist ein qualitatives Teilproblem, an das bestimmte methodische Anforderungen gestellt werden. So identifiziert beispielsweise Holsti u.a. die folgenden Anforderungen ([Hols69]):

- Das Kategoriensystem soll mit den Zielen der Untersuchung korrespondieren, es soll also zumindest zu einem bestimmten Grad *theoretisch abgeleitet* sein;
- die Kategorien sollen *paarweise disjunkt* sein;
- das Kategoriensystem soll *vollständig sein*, d.h. es sollen alle möglichen Inhalte abgedeckt sein;
- die Kategorien sollen *eindeutig definiert* sein.

Die einzelnen Bedingungen sind insbesondere durch die *Gültigkeitszielsetzungen* des inhaltsanalytischen Vorgehens begründet. Während beispielsweise die *theoretische Orientierung* des Kategoriensystems im Hinblick auf die *Validität* der Analyseergebnisse relativ zur Ausgangsfragestellung gefordert wird, zielt der Bedingung einer *eindeutigen, trennscharfen Definition* der einzelnen Kategorien auf eine möglichst gute *Reliabilität* im Sinne von intersubjektiver Reproduzierbarkeit der Codierergebnisse. Unter diesen Gesichtspunkten ist somit Berelson uneingeschränkt beizupflichten, wenn er die zentrale Rolle des Kategorienschemas für das inhaltsanalytische Vorgehen betont ([Bere52], S. 147):

“Content analysis stands or falls by its categories.”

Die Aussage Berelsons ist vor einem meßtheoretischen Hintergrund zu sehen. Die inhaltsanalytischen Kategorien definieren eine *Nominalskala*, auf deren Grundlage *Prozesse qualitativen Messens* ablaufen. Die Qualität der Meßergebnisse hängt in entscheidendem Maße von der Qualität des zugrundegelegten *Meßinstruments*, d.h. des Kategoriensystems, ab. Durch die Bedingungen von Holsti soll die Güte des als eine *Serie qualitativer Meßoperationen* zu verstehenden Kategorisierungsvorgangs sichergestellt werden.

Eine nähere Betrachtung zeigt freilich, daß die Anforderungen Holstis teilweise in entgegengesetzte Richtungen weisen. So ist es im Hinblick auf eine möglichst gute Validität der Untersuchungsergebnisse sicherlich wünschenswert, ein möglichst fein differenziertes Kategorienschema einzusetzen. Andererseits steigt mit dem Grad der Differenzierung der Aufwand, der für eine trennscharfe, operationale Abgrenzung der einzelnen Kategorien zu leisten ist. Validität und Reliabilität stehen hier somit in einem Austauschverhältnis. Ein weiterer Gegensatz offenbart sich zwischen der Bedingung der *theoretischen Orientierung* sowie der Bedingung der *Vollständigkeit* des Kategoriensystems. Während der ersteren Anforderung mit einer *theoriegeleitet-deduktiven* Kategorienbildung entsprochen wird, legt letztere Anforderung die Entwicklung des Kategorienschemas per *induktiver, materialorientierter Verfeinerungszyklen* nahe, da i.a. erst im Rahmen des Codiervorgangs entscheidbar ist, ob das Kategoriensystem tatsächlich *vollständig* ist.

Die Konstruktion eines geeigneten Kategoriensystems läuft somit in einem Spannungsfeld zwischen Theorie und Empirie ab: Ist a priori wenig über den Gegenstandsbereich bekannt, bietet sich ein induktives Vorgehen an, das mit einem sehr allgemein gehaltenen, theoriebezogen-deduktiv gewonnenen Kategorienschema startet und auf eine iterative, materialorientierte Verfeinerung abzielt. Die Festlegung eines adäquaten Kategoriensystems erweist sich somit als qualitative Aufgabe, die i.a. erst unter interpretativer Bezugnahme auf das zu analysierende Material gelöst werden kann. In der Tat wird die Unterstützung der induktiven Entwicklung von Ordnungsschemata sogar als Grundcharakteristikum von Verfahren der Qualitativen Analyse angesehen (vgl. Tesch: [Te92]). Demnach läßt sich das Prozeßmodell der Kategorienanalyse (Abbildung 2.3) durch die Einführung eines Rückkopplungszyklus zwischen den Schritten (1) und (2) präzisieren.

In seiner Synopse qualitativer Verfahren unternimmt Mayring den Versuch der Explikation dieses qualitativen Analyseschritts der Konstruktion deskriptiver Systeme (vgl. [Mayr96], S. 78 ff.). Für die Offenlegung des empirischen Testschritts bedient er sich eines Verfahrens qualitativer inhaltlicher Zusammenfassung (ibid., S. 92 ff.). In beiden Fällen beschränken sich die Explikationsversuche allerdings auf die Formulierung relativ allgemein gehaltener Ablaufmodelle; auf diese Weise wird zwar die Systematik der Kategoriensystembildung transparent gestaltet, jedoch besteht nach wie vor eine breite Kluft zu einer formalen Operationalisierung¹².

2.2.2 Kategorisierung (Codierung)

Die Kategorisierung der Erhebungseinheiten baut unmittelbar auf den Ergebnissen der Kategoriensystemdefinition auf. Der Anspruch einer möglichst weitreichenden Explikation läßt sich auf die im vorhergehenden Abschnitt genannte Forderung einer eindeutigen und möglichst operationalen Definition der Kategorien abbilden. Insofern tragen die zuvor diskutierten qualitativen Verfahren zur Konstruktion eines adäquaten Kategorienschemas auch dazu bei, den Codierschritt selbst zu explizieren. Zudem bewerkstelligt das von Mayring vorgeschlagene Verfahren der qualitativen inhaltlichen Zusammenfassung nicht nur eine induktive Kategorienbildung, sondern gleichzeitig die Zuordnung von Textstellen zu den unterschiedlichen Kategorien und damit die Codierung selbst.

¹²Natürlich tritt Mayring auch nicht mit dieser Zielsetzung an. Vielmehr geht es zunächst um die Thematisierung dieses Bereichs, der in der Quantitativen Inhaltsanalyse "*meist übergangen*" ([Mayr96], S. 78) wird, und zwar mit der Zielsetzung, daß die "*Kategorienentwicklung nun systematischer angelegt*" (ibid., S. 92) wird. Was hierdurch tatsächlich gewonnen wird, bleibt allerdings unklar, da sich die angegebenen Ablaufmodelle doch relativ nahe am *Common Sense* bewegen.

Die Spezifikationen der Klassifikationskriterien werden üblicherweise in einem *Codierbuch* expliziert, das allen an der Inhaltsanalyse beteiligten Codierern zur Verfügung gestellt wird, um eine möglichst einheitliche, die Intercoderreliabilität fördernde Klassifikationsgrundlage zu schaffen. Die Entscheidungskriterien sind i.a. in Form von *natürlichsprachigen Erläuterungen und Ankerbeispielen* sowie ggf. zusätzlichen informellen *Kodierregeln* zur Behandlung von Zweifelsfällen gegeben. Der Codierungsschritt basiert auf der intellektuellen Interpretation der Kategoriebeschreibungen und stellt somit ein qualitatives Teilproblem dar; ob bzw. inwieweit der Grad der Explikation der Klassifikationskriterien ausreichend ist, läßt sich indirekt und partiell in Form der Intercoderreliabilität messen. Mayring schlägt vor, im *Prozeß der Bestimmung des Codierbuchs* dem Ablaufmodell einer sog. *strukturierenden qualitativen Inhaltsanalyse* zu folgen. Jedoch ist auch dieser Versuch der Explizierung eines qualitativen Prozesses wiederum der Kritik ausgesetzt, keine wirklich formale Operationalisierung zu verkörpern.

Im Vorgriff auf die späteren Ausführungen zur symbolischen, computergestützten Inhaltsanalyse sei darauf hingewiesen, daß das Spannungsfeld zwischen den offenbar im Gegensatz stehenden Anforderungen einer möglichst präzisen Kategorienbeschreibung einerseits und einer möglichst feinen, den Zielsetzungen der Analyse in idealer Weise entsprechenden Differenzierung der Kategorien auch an dieser Stelle in Erscheinung tritt. Eine Möglichkeit der operationalen Explikation besteht in der erschöpfenden Aufzählung kategorien-charakterisierender *Schlüsselwörter* als Zeichenketten im Codierbuch. Zwar wäre hiermit die Operationalität sichergestellt und somit die *Reliabilität* des Vercodungsprozesses optimal gewährleistet. Andererseits wird die *Validität* der Analyse potentiell beeinträchtigt, da die kontextuellen Bedingungen, die die konkreten syntaktischen, semantischen und pragmatischen Verwendungsweisen der Schlüsselwörter determinieren, keine Berücksichtigung finden. Dieses bereits aus der Diskussion der Kontroverse "*Qualität vs. Quantität*" bekannte Austauschverhältnis tritt somit auch speziell betreffend die symbolisch-formalen Ansätze zur Explikation des Codierschritts in Erscheinung.

2.2.3 Deskriptionserstellung und Auswertung

Während die zuvor diskutierten Inhaltsanalysekomponenten stets qualitativ ausgerichtet sind, beinhalten die Schritte der Deskriptionserstellung und der Auswertung in der klassischen Kategorienanalyse eine *Reduktion* auf *quantitative* Elemente (Häufigkeiten von Kategorieninstanzen, vom Text eingenommene Fläche etc.), an die sich eine statistische Auswertung anschließt. Es verbleibt zu erörtern, wie ein qualitativ ausgerichtetes Vorgehen aussieht und welche Explikationsansätze hierfür bekannt sind.

Anders als in der Quantitativen Analyse besteht das Ziel des Codiervorgangs i.a. nicht in einer Reduktion, sondern vielmehr in einer initialen *Systematisierung* des Materials. Mit anderen Worten: Durch die Kategorisierung der Erhebungseinheiten wird angestrebt, das Material gemäß bestimmter Kriterien vorzusortieren, um einen subjektiv-interpretativen, hermeneutischen Auswertungsprozeß zu unterstützen. Das Ausgangsmaterial bleibt weiterhin voll zugänglich. Kelle umschreibt dies folgendermaßen ([Kell96], S. 56)

"So coding and retrieval is nothing more than a mechanical preparation for interpretive analysis which is based on a careful inspection and analysis of raw data ..."

und leitet hieraus Anforderungen ab, die sich im Hinblick auf die Zweckmäßigkeit unterschiedlicher Ansätze zur Computerunterstützung einer qualitativ ausgerichteten Analyse als bedeutsam

erweisen werden (vgl. Abschnitt 3.2).

Es stellt sich die Frage nach geeigneten Modellen für die Explikation einer interpretativen Auswertung des Materials. Mayring nennt u.a. die folgenden *Ansätze zur qualitativen Auswertung* ([Mayr96]). Dem Verfahren der *sozialwissenschaftlich-hermeneutischen Paraphrase* liegt der Anspruch zugrunde, das zu evaluierende Material zunächst einer unabhängigen Interpretation mehrerer Forscher zu unterziehen (Heinze: [Hein87]); die Verfolgung der *hermeneutischen Spirale*¹³, d.h. die schrittweise, wechselseitige Erweiterung von Materialverständnis und Vorverständnis im Rahmen des Deutungsprozesses wird hier durch eine sog. *metakommunikative Diskussion* fragwürdiger Punkte im Plenum der Interpreten beschleunigt. Die *Objektive Hermeneutik* bedient sich der hermeneutischen Vorgehensweise zur Explikation von Handlungskontexten, wobei das Ziel zugrunde liegt, aufbauend auf einer gedankenexperimentellen Variation dieser Kontexte Strukturgeneralisierungen zu ermitteln, die als *objektive Bedeutungsstrukturen* des zu untersuchenden Materials zu verstehen sind (Oevermann, Allert, Konau, Krambeck: [OAKK79]). Auf die Erschließung *subjektiver, intentionaler Bedeutungsstrukturen* ausgerichtet ist das Verfahren der *Phänomenologischen Analyse*, die in der Tradition der Husserlschen Phänomenologie steht; die Erschließung von Bedeutungsstrukturen geschieht hier *an subjektiven Perspektiven ansetzend*, indem durch gedankenexperimentelle Kontextvariation zum Wesen des Phänomens vorgedrungen werden soll (*eidetische Reduktion*). Insofern das Ziel der Analyse in der Kategorisierung des Materials besteht, fällt auch das zuvor diskutierte Verfahren der inhaltlichen Zusammenfassung, das ja die induktive Kategoriebildung sowie die Codierung selbst unterstützt, in diese Klasse der Auswertungsverfahren. Zusammen mit dem ebenfalls bereits erwähnten Verfahren der strukturierenden Inhaltsanalyse sowie einer Technik zur systematischen Explikation des für Textstellen relevanten kontextuellen Explikationsmaterials subsumiert Mayring dieses Verfahren unter dem Oberbegriff der *Qualitativen Inhaltsanalyse*.

Wenn auch die typischen Anwendungsgebiete der einzelnen Verfahren voneinander abweichen - eines ist allen Ansätzen gemeinsam: *Die Systematisierung beschränkt sich auf die Angabe vage formulierter Ablaufschemata, die für die Verfahrensdurchführung einen vergleichsweise großen Spielraum lassen*. Folglich gilt auch hier: Der Nachweis einer hinreichenden Operationalisierung des jeweiligen Verfahrens läßt sich vor allem indirekt erbringen, indem eine nachgeschaltete quantitative Auswertung eine ausreichende Intercoderreliabilität aufzeigt.

2.2.4 Interpretation und Validierung der Auswertungsergebnisse

Ist die Auswertung vollzogen, so sind die gewonnenen Ergebnisse im Hinblick auf die Forschungsfrage zu interpretieren, um die theoriebildende Inferenz auf den gesellschaftlichen Kontext zu vollziehen. Im Unterschied zum Nachweis der internen Gültigkeit (insbes. der Reliabilität) sind die Möglichkeiten zur formalen Absicherung der Ergebnisse dieses inhaltsanalytischen Schritts begrenzt. Eine Überprüfung auf sog. *Kriteriumsvalidität*, d.h. die Validierung der Resultate per statistischem Abgleich (Bestimmung des Korrelationskoeffizienten) mit einem bekanntermaßen in engem inhaltlichem Zusammenhang mit den zu erschließenden Zielgrößen stehenden, möglichst *manifesten Außenkriterium* setzt dessen unmittelbare Verfügbarkeit bzw. verlässliche Operationalisierbarkeit voraus. Speziell in letzterem Fall entsteht jedoch rekursiv die Anforderung der Validierung des Verfahrens, auf dessen Grundlage das Vergleichskriterium operationalisiert ist. In verstärktem Maße betrifft dieses Problem den ebenfalls vorgeschlagenen Test auf sog.

¹³Es wird bewußt nicht von einem hermeneutischen Zirkel gesprochen, um herauszustellen, daß dem Verstehensprozeß nach dem Dafürhalten der vorliegenden Arbeit kein unaufösbarer *Circulus vitiosus* zugrunde liegt.

Konstruktvalidität, in dem vergleichbare Gegenproben auf der Grundlage einer Menge verwandter Hypothesen durchzuführen sind (vgl. Bortz und Döring: [BoDö95], S. 185 ff.). Somit bewegt sich die Güte einer derartigen indirekten Außenvalidierung innerhalb einer bestimmten Grenze, die durch die Validität und die Reliabilität der herangezogenen Instrumente determiniert wird. Auch Merten thematisiert diesen *unendlichen Rekurs des Validierungsproblems*, der die Inhaltsanalyse als grundlegendes Erhebungsinstrument in besonderem Maße beträfe ([Mert95], S. 313):

“Daß, vergleichbar etwa den Zuverlässigkeitsprüfungen, so wenig Prüfungsmöglichkeiten auf Gültigkeit vorhanden sind, liegt vor allem darin begründet, daß eine Validierung durch andere Instrumente jeweils die Validität dieser Instrumente voraussetzt. Da aber Inhaltsanalyse als basales Erhebungsinstrument gelten darf [...], ist gerade eine Validierung der Inhaltsanalyse durch andere Erhebungsinstrumente nicht unproblematisch.”

In jedem Fall stellt die Aufgabe der Validierung ein qualitatives Problem dar, in dessen Lösung quantitativ-statistische Tests allenfalls als Hilfskonstrukte einfließen.

In der qualitativen Sozialforschung wird das Validierungsproblem zwar zur Kenntnis genommen, jedoch werden ausschließlich allgemeine Anhaltspunkte dafür gegeben, wie die Gültigkeit der Interpretationsergebnisse sichergestellt werden könnte. Mayring nennt die Möglichkeiten, Interpretationen *argumentativ abzusichern* ([Mayr96], S. 119) oder einer *kommunikativen Validierung* zu unterziehen, in deren Rahmen die Ergebnisse mit den untersuchten Subjekten diskutiert und ggf. revidiert werden (ibid., S. 121). In den relativ allgemein gehaltenen Ausführungen manifestiert sich das Problem jedweder externer Validierung, wonach der Gültigkeitsnachweis letztendlich immer an subjektiven Bedeutungszuweisungen verankert werden muß, insofern kein Modell des zu inferierenden Kontextes zur Verfügung steht, auf dessen Basis ein formaler Beweis vonstatten gehen könnte.

Kapitel 3

Computereinsatz in der Inhaltsanalyse

Die Diskussion der qualitativen Prozeßkomponenten der Inhaltsanalyse hat aufgezeigt, daß bereits durchweg Ansätze zu deren Systematisierung in Gestalt von Verfahren der Qualitativen Analyse existieren. Mit Ausnahme der nur kurz angerissenen wörterbuchbasierten Operationalisierung des Kategorisierungsschritts ist der Grad der formalen Explikation allerdings als nicht hinreichend für eine algorithmische Automatisierung durch Softwarelösungen einzustufen. Um die Grundlagen für die Diskussion der Algorithmisierungsproblematik interpretativer Prozesse zu vervollständigen, wird nun die Perspektive gewechselt. Es soll eine Synopse bekannter Softwarelösungen für qualitative Prozeßkomponenten der Inhaltsanalyse erarbeitet werden. Auf der Basis dieses Überblicks wird es möglich sein, die Grenzen der aktuellen Ansätze zu erarbeiten. In einem weiteren, konstruktiven Schritt werden diese Betrachtungen dann als Ausgangspunkt der Identifikation inhaltsanalytischer Teilprobleme dienen, für die sich verfeinerte computerlinguistische Lösungsansätze anbieten.

3.1 Computerunterstützte versus Computergestützte Analyse

Mit Blick auf die unterschiedliche Ausrichtung der vorgeschlagenen Ansätze bietet sich folgende Unterscheidung an:

- *Computerunterstützte* Inhaltsanalyse,
- *Computergestützte* Inhaltsanalyse.

Softwarelösungen zur Computerunterstützung der Inhaltsanalyse stellen *Hilfsfunktionen* zur Verfügung, die den Forscher bei der Qualitativen Analyse *unterstützen*, ohne jedoch algorithmische Realisierungen des eigentlichen Interpretationsvorgangs anzubieten. Typische Aufgaben, die von solchen Systemen übernommen werden, sind Datenverwaltung und Retrievalfunktionen. In ihrer

Ausrichtung genügen derartige Programmsysteme den Anforderungen, der üblicherweise für die Unterstützung der Qualitativen Analyse gestellt werden. So konstatiert etwa Huber ([Hube92a], S. 9):

“Sich bei der Analyse qualitativer Daten eines Computers zu bedienen heißt nicht, das Interpretieren dem Computer zu überlassen.”

Während der Forscher also - beispielsweise unter Anwendung eines der oben genannten schematischen Verfahren zur Qualitativen Analyse - über die Entscheidungen sowie den Ablauf des Analysevorgangs bestimmt, leistet der Computer wichtige *Hilfsdienste* für Verwaltung, Revision und Verfeinerungen der Codierentscheidungen. Diese Funktionalität ist für die Qualitative Analyse um so mehr von Bedeutung, als ja hier i.a. das Kategoriensystem erst induktiv, d.h. materialorientiert und - der hermeneutischen Spirale folgend - in mehreren Verfeinerungsschritten zu bestimmen ist und der Arbeitsaufwand entsprechend umfangreich ausfällt.

Softwarelösungen zur computergestützten Inhaltsanalyse liegt der viel weitreichendere Anspruch zugrunde, *Interpretationsleistungen* zu erbringen, die andernfalls in den Aufgabenbereich des Forschers fielen. Es sollen also zumindest einige qualitative Prozeßkomponenten der Analyse *vollständig automatisiert* werden.¹ Bislang handelt es sich in erster Linie um Programmsysteme, die den Codierschritt auf der Basis von Kategoriensystemen algorithmisieren, die über *Wörterbücher* definiert sind. Derartige Lösungen werden v.a. in der klassischen Quantitativen Inhaltsanalyse eingesetzt; den spezifischen Anforderungen einer qualitativ ausgerichteten Inhaltsanalyse wird nicht Rechnung getragen, da u.a. (a) die interpretativ-induktive Entwicklung des Kategoriensystems am Material allenfalls partiell unterstützt wird und (b) als Ergebnis eine *Datenreduktion* auf Codehäufigkeiten vorliegt, womit vom Ausgangsmaterial abstrahiert wird und somit eine intellektuelle Interpretation nicht länger möglich ist.

3.2 Computerunterstützte Analyse

In einer typischen Qualitativen Analyse klassischen Zuschnitts werden die Vercodungen, d.h. die klassifizierten Textpassagen, mit Hilfe manueller Techniken verwaltet: Textpassagen werden im (papierförmigen) Ausgangsmaterial markiert, dann ggf. ausgeschnitten und in Codierbögen oder Karteikartenablagen systematisiert (*“Cut-and-Paste”*-Techniken). Es ist offensichtlich, daß diese Vorgehensweise in der Praxis recht schnell an ihre Grenze stößt: Die für die Qualitative Analyse charakteristische induktive Entwicklung des Kategoriensystems wird nur unzureichend unterstützt, da die damit einhergehende schrittweise Verfeinerung der Codierung zu einem erheblichen technischen Arbeitsaufwand führt.

3.2.1 Unterstützung der Verwaltung von Kategorisierungsentscheidungen

Auf der Ebene der Verwaltung von Codierungen lassen sich folgende Aufgaben identifizieren:

¹Für die statistischen Verfahren, die in der Auswertungsphase einer Quantitativen Analyse zum Einsatz kommen, ist die Algorithmisierung trivial. Sie fallen nicht in den Skopus der vorliegenden Arbeit, da keine hermeneutische Textinterpretationsleistung erbracht wird.

1. Textstellen markieren und kategorisieren,
2. zur Unterstützung der induktiven Verfeinerung:
 - Retrieval bestimmter Textstellen,
 - Sortieren der Textstellen nach Kategorien.

Erste Versuche, diese Aufgaben auf der Basis von Textverarbeitungs- oder Datenbanksoftware zu bewerkstelligen, bewiesen einerseits den potentiellen Nutzen eines elektronischen Cut-and-Paste, legten andererseits aber auch die Ungeeignetheit konventioneller Datenbanklösungen zur Verwaltung unstrukturierter Texte offen. Ausgehend von diesen Erfahrungen wurden seit Mitte der achtziger Jahre eine Anzahl von Spezialsystemen entwickelt, die den zuvor genannten Anforderungen gerecht werden - z.B. (ohne Anspruch auf Vollständigkeit) *The Ethnograph*, *HyperQual*, *MAX*, *QUALPRO*, *Textbase Alpha*.²

Es ist allgemein akzeptiert, daß diesen Systemen nicht nur eine reine Hilfsfunktion zukommt, sondern daß sie vielmehr eine wichtige methodologische Neuerung für die Qualitative Analyse darstellen. Die Unterstützung für die Verwaltung der aufgrund des Verzichts auf vorschnelle Datenreduktion oft umfangreichen Datenmengen schafft erst den notwendigen Freiraum für die propagierte induktive, schrittweise Verfeinerung des deskriptiven Systems. So bemerkt beispielsweise Tesch ([Te92], S. 68):

“Nicht nur kann mehr Zeit und intellektuelle Energie auf die eigentliche analytische und interpretative Aufgabe verwandt werden, sondern es ist heute selbstverständlich, daß Kodierungssysteme mehrfach verändert und verbessert werden [...].”

Auch Kelle kommt zu dem Ergebnis, daß sich die Akzeptanz dieser Ansätze gerade in deren idealer Kompatibilität mit der methodologischen Ausrichtung der Qualitativen Analyse begründet ([Kell96], S. 43):

“[...] second generation software represents a major methodological innovation for qualitative research since they allowed the researcher to analyse more data more systematically and carefully [...].”

3.2.2 Unterstützung der Bildung von Code-Abstraktionen

Auch auf dieser Ebene geht es um die Unterstützung des Kategorisierungsvorgangs, hier jedoch mit einer erweiterten Zielsetzung. Während zuvor die Textpassagen selbst Gegenstand der Codierung waren, sind es nunmehr die Codes (Kategorienzuweisungen), die analysiert werden. Es werden Regelmäßigkeiten in den Codierungen gesucht, die Anhaltspunkte für (im weitesten Sinne) konzeptuelle Zusammenhänge und damit den Ausschlag für die Einführung von Meta-Codes geben könnten. Tesch spricht in diesem Zusammenhang von einer *Theorie-konstruierenden Analyse*, deren Ziel, die theoretische *Erklärung*, vom Ziel des *Verstehens* einer rein beschreibenden

²Für Details und eine vollständigere Übersicht vgl. Fielding und Lee ([FiLe91]) oder Huber ([Hube92a]).

Analyse erster Ordnung zu unterscheiden sei ([Te92]). In bezug auf die Ausführungen in Kapitel 2 koinzidiert diese Aufgabenstellung mit dem Schritt der inferentiellen Theoriebildung der erweiterten Inhaltsanalysedefinition nach Krippendorff und Merten (vgl. Abbildung 2.2).

Nach Huber ([Hube92b], S. 139 ff.) lassen sich die möglichen Zusammenhänge unterteilen in

- einfache semantische Sequenzrelationen zwischen zwei Codes: hierarchisch, kausal, temporal, konzessiv, konditional etc.,
- komplexere Sequenzmuster: basierend z.B. auf motivations- oder handlungstheoretischen Erwartungsmodellen.

Entsprechend den technischen Anforderungen, die sich für eine solche Analyse ableiten lassen, wurden Softwarelösungen entwickelt, die eine automatische Identifikation von bestimmten Mustern in den Kategoriezuweisungen ermöglichen. Zu den unterstützten Suchfunktionen zählen:

- Ko-Okkurrenz,
- Nachbarschaft,
- Anordnung,
- Überlappung von Textpassagen,
- Einbettungsrelation zwischen Textpassagen,
- selektive Suche auf einer Teilmenge von Texten.

Mittlerweile wurde eine Reihe von Systemen realisiert, die eine Auswahl entsprechender Funktionen unterstützen: *Atlas/ti*, *HyperRESEARCH*, *Hypersoft*, *NUDIST*, *Qualog* und - als *Qualog*-Derivat - *AQUAD*, wobei sich das zuletzt genannte System durch besondere Flexibilität auszeichnet, indem es eine offene Programmierschnittstelle für die freie Spezifikation auch komplexer Suchmuster mit Sprachelementen der logischen Programmierung anbietet (Huber, *ibd.*). Zwar sehen die angebotenen Suchoptionen zum Teil beeindruckend aus, *der eigentliche Interpretationsvorgang wird jedoch nicht automatisiert*. Vielmehr handelt es sich um eine vergleichsweise komplexe Retrievalfunktionalität, die eine Suche nach oberflächlichen Regelmäßigkeiten in den Codeabfolgen unterstützt; *was* letztendlich gesucht wird, und *wie* die Suchergebnisse inhaltlich zu interpretieren sind, darüber entscheidet noch immer der Forscher.

Eine besondere Rolle kommt dem Verfahren des *fallorientierten qualitativen Vergleichens* von Ragin zu ([Ragi87, Ragi95], engl. "*qualitative comparative analysis*", Programm *QCA*, implementiert auch in *AQUAD*). Der Grundgedanke dieses Ansatzes besteht darin, unterschiedliche "Fälle" (entsprechend Folgen von Codezuweisungen zu einzelnen Texten) auf aussagenlogische Formeln abzubilden. Die vergleichende Suche nach Regelmäßigkeiten in den Codesequenzen geschieht durch Anwendung eines Verfahrens zur logischen Minimierung, dem sog. *Quine-McClusky-Algorithmus*. Ergebnis ist eine Liste reduzierter Codesequenzen - sog. *Prinimplikanten* -, die als Anhaltspunkte für Kausalzusammenhänge in der codierten Domäne interpretiert werden können. Das Verfahren ist insofern flexibel, als es sowohl in der Phase der induktiven Kategorienverfeinerung (zur schnellen Aufdeckung verdeckter Inkonsistenzen) als auch in der

Phase der Theoriekonstruktion (Ergebniszusammenfassung) oder gar im Rahmen eines Vergleichs mehrerer Untersuchungen eingesetzt werden kann (vgl. [Hube92b]). Alle Vorteile dieses algorithmischen Verfahrens dürfen jedoch nicht darüber hinwegtäuschen, daß auch hier letztlich gilt: *Die eigentliche Interpretationsleistung vollbringt der Forscher*; diese Tätigkeit beginnt bereits bei der Festlegung einer adäquaten Abbildung gegebener Fälle auf aussagenlogische Formeln³ und endet nicht vor der interpretativen Entscheidung, ob ein algorithmisch bestimmter Implikant tatsächlich einen konzeptuellen Zusammenhang kennzeichnet.

Im Unterschied zu den Ansätzen zur Computerunterstützung der Codierungsverwaltung haben sich die in diesem Abschnitt vorgestellten Verfahren zur Suche nach systematischen Relationen zwischen Code-Kookkurrenzen bisher nicht durchgesetzt. Während Tesch davon spricht, daß das Potential der Systeme zur Theorie-Konstruktion wenigstens *begonnen* hätte, die Methode der Qualitativen Analyse zu beeinflussen ([Te92], S. 68), äußert Kelle die Vermutung, die bisherige Inakzeptanz könne von einer potentiellen *Inkompatibilität* algorithmischer (d.h. nicht hermeneutisch-interpretativer) Verfahren mit der interpretativen Grundausrichtung der Qualitativen Analyse herrühren; diese Unverträglichkeit käme genau dann zum Vorschein, wenn die erweiterte Suchfunktionalität fälschlicherweise als Werkzeug zum *Testen* von Hypothesen angesehen wird. Als hiervon strikt abzugrenzende *Suchwerkzeuge*, die die (interpretierenden) Forscher beim kreativen Spiel mit ihren Daten im Rahmen der inferentiellen Theoriebildung *unterstützen*, hätten diese Systeme aber durchaus ihre Berechtigung ([Kell96], S. 51 ff.).

3.3 Computergestützte Analyse

Anders als im Falle der zuvor untersuchten Werkzeuge zur Unterstützung des inhaltsanalytischen Prozesses treten die nunmehr zu diskutierenden Verfahren mit dem ambitionierteren Anspruch an, bestimmte, gemäß dem Diktum der hermeneutischen Lehre durch den Forscher zu erbringende *Interpretationsleistungen* algorithmisch zu lösen. Bisher konzentrieren sich entsprechende Ansätze auf den Kategorisierungsprozeß.

3.3.1 Algorithmisierung des Kodierschritts

Gemeinsames Merkmal der in diese Klasse fallenden Ansätze sind Vercodungskriterien, die generell auf *oberflächlichen* Merkmalen der zu klassifizierenden Auswahlinheiten basieren: In der Regel sind es einzelne Wörter als Zeichenketten, die den Ausschlag für die Zuordnung zu einer bestimmten Kategorie geben; eine verfeinerte, semantische oder pragmatische Analyse des Textinhalts liegt den Entscheidungen nicht zugrunde. Das Kategorienschema wird durch ein *Wörterbuch* operationalisiert, in dem die unterschiedlichen Kategorien per Enumeration von Schlüsselwörtern beschrieben sind. Zur Lösung des Problems von Wortendungsvariationen⁴ werden unterschiedliche Ansätze verfolgt, i.a. entweder die Aufzählung aller möglicher Wortformen in einem *Vollformenlexikon* oder die heuristische Spezifikation von *Schlüsselwort-Präfixen*. Der Rahmenalgorithmus des eigentlichen Kategorisierungsvorgangs basiert auf einer Zerlegung

³Für große und daher potentiell "widersprüchliche" Fallmengen sind auch *quantitativ* ausgerichtete, auf Schwellenwerten basierende Zuordnungskriterien denkbar.

⁴In linguistischer Terminologie wird hier von *Flexion* oder auch *Beugung*, umfassend Deklination von Nomen, Konjugation von Verben und Komparation von Adjektiven gesprochen.

der zu analysierenden Auswahleinheiten (Textdokumente) in einzelne Wörter. Der zugrundeliegende Wortbegriff ist elementar: Als "Wörter" aufgefaßt werden maximale leerzeichen- und satzzeichenfreie ununterbrochene Abfolgen von Zeichen - Zeichenketten - im Text; eine morphologische Analyse zur Ermittlung der lexikalischen Stammformen der Wörter findet in der Computergestützten Inhaltsanalyse klassischen Zuschnitts nicht statt. Die Vercodung erfolgt wortweise: Für jedes nicht als irrelevant angesehene Wort⁵ findet ein Vergleich mit allen im Kategoriewörterbuch vorkommenden Schlüsselwörtern statt; besteht Übereinstimmung, so wird die Auswahleinheit derjenigen Kategorie zugeordnet, der das Schlüsselwort zugeordnet ist. Am Ende des Analysevorgangs liegt für jede Auswahleinheit eine Menge von Kategoriezuordnungen vor, die für die weitere Verarbeitung - beispielsweise in Form einer quantitativ-statistischen Auswertung - genutzt werden können.

Die Entscheidung, ob ein Wort auf eine Kategorie K abzubilden ist, geschieht auf der Basis eines buchstabenweisen Vergleichs mit den im Wörterbuch aufgezählten, K definierenden Schlüsselwörtern. Dieser Vergleich zweier Zeichenketten, der sich auch für Schlüsselwort-Präfixe nur unwesentlich schwieriger gestaltet, stellt ein Problem dar, das auf triviale Weise algorithmisch lösbar ist.⁶ Weitergehendes kontextuelles Wissen findet keine Berücksichtigung, womit - im Einklang mit der klassischen Definition Berelsons - der Anspruch der Erschließung ausschließlich des sog. manifesten Inhalts eingelöst werden soll. Die Vorteile der algorithmischen Operationalisierung des Codiervorgangs liegen auf der Hand. Die Reliabilität ist maximal, da der Prozeß jederzeit mit identischem Ergebnis wiederholt werden kann. Darüberhinaus ist die Laufzeitkomplexität des Klassifikationsalgorithmus niedrig, so daß eine Recodierung im Falle von Änderungen oder Erweiterungen des Codierwörterbuchs i.a. (in Abhängigkeit vom Gesamtumfang des zu analysierenden Materials) ohne großen Zeitaufwand durchgeführt werden kann. Allerdings werden diese Vorteile teuer erkauft. Im Vergleich zu einer interpretativen Kategorisierung ist die Entscheidungsgrundlage ungleich oberflächlicher, da die Codierung einzelwortbezogen und unter Außerachtlassung der kontextuell determinierten Semantik erfolgt. Es besteht somit die Gefahr, daß die hohe - hier sogar maximale - Reliabilität durch nicht mehr akzeptable Abstriche bei der *Validität* der Aussagen, die auf Basis einer solchermaßen erzeugten Codierung gewonnen werden können, entwertet wird. Bevor jedoch diese offenkundigen Einschränkungen ausführlicher untersucht werden, sollen zunächst einige Softwaresysteme besprochen werden, die eine automatische wörterbuchbasierte Codierung realisieren.

3.3.2 Das Softwaresystem TEXTPACK

Die Inhaltsanalysesoftware TEXTPACK, die auf frühe Arbeiten am Zentralarchiv für empirische Sozialforschung an der Universität zu Köln zurückgeht, wird am Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim weiterentwickelt (Mohler und Züll: [MoZü95]). Seine Grundfunktionalität deckt alle zuvor identifizierten Teilprobleme der wörterbuchbasierten

⁵I.a. besteht die Möglichkeit, bestimmte, z.B. als nicht-distinktiv angesehene Wörter wie beispielsweise Konjunktionen ("und", "oder", "aber", "denn", "weil", ...) und Präpositionen ("in", "auf", "für", "mit", "von", ...) a priori von der weiteren Analyse auszuschließen, indem man diese in sog. *Stopwort-Listen* spezifiziert.

⁶Informatisch gesprochen: Das Problem fällt in die Klasse der *Regulären Sprachen*, für die eine algorithmische Lösung sogar bereits auf der Basis eines stark restringierten Berechenbarkeitsmodells, den sog. *Endlichen Automaten*, möglich ist (vgl. z.B. Hopcroft und Ullman [HoU179]).

Textklassifikation ab.⁷ Ein Beispiel eines Klassifikationswörterbuchs für TEXTPACK ist⁸

```
002 +   TEXTPACK
001     bliss
001     hellwards
001     magic
001     mirror
002 -   program
001     thunder
```

Die erste Spalte enthält einen numerischen Kategorie-Identifikator. Er ordnet die in den folgenden Spalten gegebene Schlüsselwortspezifikation einer bestimmten Kategorie zu. Die Schlüsselwortspezifikation selbst besteht aus einer Zeichenkette sowie ggf. einer Marke “-” oder “+”, die festlegt, in welcher Weise die Zeichenkette mit den Wörtern des Texts verglichen wird. Marke “-” legt fest, daß die Zeichenkette als Schlüsselwort-*Präfix* zu interpretieren ist - beispielsweise als heuristische Lösung des Problems der Wortendungsvariation.⁹ Marke “+” spezifiziert, daß die Übereinstimmung *vollständig* sein muß; diese Option bietet sich im Falle von Eigennamen wie (im Beispiel) TEXTPACK an, falls davon ausgegangen werden kann, daß diese ausschließlich unflektiert vorkommen. Ist keine Marke spezifiziert, so zählen auch Übereinstimmungen als Treffer, in denen das Schlüsselwort als *beliebige* Teilzeichenkette eines Wortes der zu vercodenden Auswahlinheit vorkommt. Die Spezifikationen im Beispielwörterbuch würden somit u.a. folgende Treffer zulassen¹⁰

```
+ TEXTPACK  TEXTPACK
bliss       bliss, blissful, ...
hellwards   hellwards, ...
magic       magic, magical, ...
mirror      mirror, mirrored, ...
- program   program, programme, programmer, programmatic, ...
thunder     thunder, thundercloud, thunderstorm, ...
```

Speziell das Schlüsselwort-Präfix *program* erwies sich als sehr anfällig für Ambiguitäten. Im Beispielkorpus traten u.a. folgende Verwendungskontexte in Erscheinung:

```
software programs, nuclear bomb programme, atomic programme,
rural programs, soil conservation programs
```

⁷Darüberhinaus stellt TEXTPACK Werkzeuge zur Verfügung, die u.a. folgende Aufgaben unterstützen: Anlegung/Verwaltung von sowie Zugriff auf hierarchisch geordnete Textsegmentierungen als potentielle Auswahl- und Kodiereinheiten, Auszählen von Worthäufigkeiten, Vokabularvergleich unterschiedlicher Dokumente, Index- und Konkordanzherzeugung, Listen von Keywords-in-Context (KWIC).

⁸aus: Mohler und Züll, [MoZü95], S. 31

⁹Daß hiermit keine vollständige Lösung des Flexionsproblems erreicht werden kann, sieht man anhand des folgenden Beispiels: Die Spezifikation “- Mann” würde zwar die Singularformen “*Manns*” und “*Manne*” abdecken, jedoch schon nicht mehr die Pluralformen mit verändertem Stammvokal (“*Männer*” etc), für die somit ein eigener Eintrag notwendig wird. Aber selbst dann bleibt noch das Problem bestehen, daß auch Wörter wie “*Mannschaft*” oder gar “*Mannigfaltigkeit*” mitcharakterisiert werden.

¹⁰Die fiktiven Treffer dieses Beispiels wurden auf der Grundlage eines Wörterbuchs der englischen Sprache sowie eines Korpus von Tickermeldungen manuell generiert.

Zumindest in der älteren Version IV geht TEXTPACK über die Basisfunktionalität wörterbuch-basierter Vercodung hinaus.¹¹ Das Submodul TEXTLING schafft die stark eingeschränkte, da nicht wirklich inhaltsbezogene Möglichkeit, die Operationalisierung der Kategorien unter Bezugnahme auf den Kontext zu verfeinern. Das Wörterbuch darf *komplexe Ausdrücke* enthalten, die sich, ausgehend von Zeichenketten als elementare Schlüsselwörter, rekursiv unter Verwendung bestimmter Funktoren bilden lassen. Als Funktoren stehen u.a. zur Verfügung: UND (Nachbarschaft), ODER/LISTE (Alternativen), VOR/RFOLGE, NACH (Anordnung). In ihrer Semantik sind diese Operatoren mit den Konzepten zur Bildung von Code-Abstraktionen vergleichbar, die als Suchwerkzeuge für die Computerunterstützte Inhaltsanalyse vorgestellt wurden.

Im Ambiguitätsfall für die Präfixspezifikation – **program** könnte mit Hilfe dieses Funktoreninventars eine induktive Verfeinerung der Kategorienschema-Operationalisierung vorgenommen werden. Beispielsweise würde die Spezifikation

```
(software UND (program ODER programmer))
```

bewirken, daß **software** nur gemeinsam mit einem Vorkommen von entweder **program** oder **programmer** innerhalb derselben Auswahlinheit als Schlüsselwort wirkt. Eine weitere Verschärfung des Kriteriums könnte durch Festschreibung einer Reihenfolge erreicht werden:

```
(software VOR (program ODER programmer))
```

In diesem Fall werden nurmehr gemeinsame Vorkommen als Treffer gezählt, in denen **software** in der Auswahlinheit vor **program** oder **programmer** vorkommt.

TEXTPACK PC 5.0 bietet eine zwar nicht äquivalente, jedoch zumindest vergleichbare Möglichkeit der Spezifikation von *Wortfolgen*, für die eine vollständige Übereinstimmung (“+”) gegeben sein muß:

```
...
001 +   software program
001 +   software programmer
...
```

Auch mit der erweiterten Funktionalität zur Kontextbezugnahme in den Operationalisierungen der Kategorien werden jedoch bei weitem nicht alle Fehlcodierungen ausgeschlossen. TEXTPACK bietet ein Werkzeug KWIC an, das eine intellektuelle Validierung der automatisch getroffenen Codierentscheidungen ermöglicht. KWIC generiert Konkordanzbeschreibungen für die codierten Wörter in Form von Keyword-in-Context-Listen, die es ermöglichen, problematische Fälle zu erkennen und ggf. durch Verfeinerungen des Kategorienschemas zu beseitigen. Somit sind in TEXTPACK die Grenzen zwischen computergestützter und computerunterstützter Analyse fließend gestaltet, da ein semiinterpretativ-hermeneutischer Analysestil mit induktiven Verfeinerungsschritten ebenfalls unterstützt wird; jedoch geschehen auch hier die Klassifikationsentscheidungen letztendlich auf der Basis des Wörterbuchs.

¹¹Das schon seit 1981 verfügbare TEXTPACK IV (Höhe, Klingemann, Rademacher, Schicke: [HKRS81]) bietet die im folgenden beschriebene Funktionalität, die allerdings in der aktuellen PC-Version TEXTPACK PC, Release 5.0, nicht zur Verfügung steht.

3.3.3 Das Softwaresystem INTEXT

Die von Harald Klein, Institut für Soziologie der Universität Jena vertriebene Inhaltsanalysesoftware INTEXT für PCs ([Klei95, Klei97]) weist bezüglich der allgemeinen Merkmale eine große Ähnlichkeit mit TEXTPACK auf. Als echte Neuerung bietet INTEXT ein Werkzeug zur *interaktiven* Validierung des automatischen Codiervorgangs, das es beispielsweise ermöglicht, Mehrdeutigkeiten oder sonstige potentiell problematischen Fälle einer intellektuellen Entscheidung zuzuführen. Im Vergleich mit der PC-Version von TEXTPACK verfügt INTEXT über verfeinerte Konzepte zur Definition von Schlüsselwörtern: Neben Wortfolgen und Wortstammfolgen werden auch Wortpermutationen sowie die Verwendung der Metazeichen “?” und “*” als Jokersymbole unterstützt; darüberhinaus existiert eine - natürlich stark eingeschränkte - Möglichkeit zur automatischen Erkennung negierter Suchbegriffe. Auch wenn die Sprachkonzepte zur Operationalisierung der Kategorien an einigen Stellen über die von TEXTPACK bzw. dessen Ergänzungspaket TEXTLING gebotenen Beschreibungsmittel hinauszugehen scheinen, so bestehen doch prinzipiell die gleichen Einschränkungen, was die Möglichkeiten zur Bezugnahme auf den Kontext und damit die erzielbare Granularität der Analyse anbelangt. Auch in TEXTPACK basiert die Codierentscheidung letztendlich auf den in ihrer Ausdruckskraft stark restringierten *Regulären Ausdrücken*. Für die nachfolgende Diskussion der konzeptuellen Grenzen der wörterbuchbasierten automatischen Vercodung erweisen sich die Unterschiede als letztlich belanglos, so daß - ohne einer Wertung Ausdruck verleihen zu wollen - auf eine Diskussion der entsprechenden Konzepte von INTEXT verzichtet werden soll.

3.3.4 Das Softwaresystem GENERAL INQUIRER

Das Analyseprogrammpaket GENERAL INQUIRER ist der Urahn der wörterbuchbasierten Verfahren zur computergestützten Inhaltsanalyse. GENERAL INQUIRER ging aus Arbeiten hervor, die seit Beginn der sechziger Jahre von Philip Stone und Mitarbeitern durchgeführt wurden ([SBNO62, SDSO66]).

Im Rahmen der Verfeinerung der Basiskomponente wurde ein Schwerpunkt auf die Disambiguierung von Homonymen und Polysemen gelegt, da diese recht bald als grundlegendes Problem der Einwortanalyse identifiziert wurden. Auch die Technik, die im GENERAL INQUIRER zum Einsatz kommt, basiert lediglich auf einem stark eingeschränkten Modell eines (intratextuellen) Kontextes in Gestalt einer Operationalisierung der KWIC-Technik. Für Vorkommen mehrdeutiger Schlüsselwörter wird in deren unmittelbaren (satzlokalen) Nachbarschaft nach weiteren Schlüsselwortvorkommen gesucht, die - aufgrund von statistischen Analysen - dafür bekannt sind, tendenziell in *bestimmten Verwendungszusammenhängen* des lexikalisch mehrdeutigen Wortes aufzutreten. Diese (heuristische) Eingrenzung der Verwendungsart erlaubt nunmehr die Wahl einer spezifischen Lesart des zu disambiguierenden Schlüsselworts. Stone verweist auf Evaluationsergebnisse, denen zufolge mit Hilfe dieser Technik mehr als 90 Prozent aller “*word ambiguities*” resolvierbar sind, und gelangt insgesamt zu einer überaus positiven Bewertung der KWIC-Disambiguierungstechnik ([Ston69], S. 204):

“Not only is extra-sentence information not needed, but also it is usually unnecessary to have the complete sentence available, let alone an unraveling of its syntactic structure.”

3.4 Diskussion: Computereinsatz für die Inhaltsanalyse

Nachdem nunmehr eine Übersicht der unterschiedlichen Ansätze zur Computerunterstützten bzw. Computergestützten Analyse vorliegt, soll zunächst der Rückbezug zum allgemeinen Schema der Inhaltsanalyse (Abbildung 2.3) hergestellt werden. Welcher Analyseschritt wird durch welches Verfahren unter- bzw. gestützt? Was bleibt dem Forscher vorbehalten?

Die beschriebenen Ansätze zur Computerunterstützung der Inhaltsanalyse setzen praktisch ohne Ausnahme an den Schritten (1) Kategorienbildung und (2) Kategorisierung sowie speziell an der für die Qualitative Analyse charakteristischen induktiven Rückkopplungsschleife zwischen (1) und (2) an. Dies gilt sowohl für die Datenbankunterstützung zur Codierungsverwaltung als auch für die Suchfunktionen zur Bildung von Code-Abstraktionen, die in ihrer primären Funktion Werkzeuge zur Identifikation komplexerer Bedeutungseinheiten darstellen.

Als Instrument für die *inferentielle Theoriekonstruktion* kommt der letzteren Klasse von Verfahren jedoch noch eine weitere Aufgabe zu, in der sie Schritt (4) in Abbildung 2.3 zuzuordnen sind und somit ein qualitatives Gegenstück des im Rahmen der quantitativ ausgerichteten Inhaltsanalyse zum Einsatz gelangenden statistischen Instrumentariums darstellen. Sie verkörpern Werkzeuge, die den interpretierenden Forscher bei der Exploration der Daten unterstützen. Die speziell dem QCA-Verfahren zugeschriebene Rolle zum Vergleich der Ergebnisse mehrerer unterschiedlicher Untersuchungen ist sogar darüberhinaus mit Schritt (5) in Beziehung zu setzen, da sie sich als Technik zur relativen Außenvalidierung auffassen läßt. In Anbetracht dieser Möglichkeiten zur nicht-statistischen Generalisierung gelangt Huber zu dem Fazit ([Hube92b], S. 175)

“Meta-Analysen müssen nicht notwendig quantitativ sein”

Für die Computergestützte Inhaltsanalyse gestaltet sich die Zuordnung zu Schritt (2) eindeutig: Die in einer manuellen Inhaltsanalyse intellektuell vollzogene Codierung wird per Operationalisierung des Kategoriensystems durch ein Wörterbuch algorithmisiert. Die von TEXTPACK und INTEXT angebotenen Beschreibungsmittel zur Präzisierung des Kontextbezugs dienen ausschließlich zur Verfeinerung der Kategorienoperationalisierungen - eine Funktion als Werkzeug zur Theoriekonstruktion, wie sie für die vergleichbaren Operatoren zur qualitativen Exploration möglicher Code-Abstraktionen postuliert wird, ist nicht gegeben. Gemäß ihrer Ausrichtung sind diese Werkzeuge primär als datenreduzierende Voranalyssysteme einer quantitativ ausgerichteten Analyse konzipiert.

Zusammenfassend verbleibt festzuhalten, daß die Mehrzahl der Computerlösungen für die Inhaltsanalyse zu Schritt (2) der Kategorienanalyse beiträgt. Vergleicht man die Ansätze zur automatischen, computergestützten Kategorisierung mit den qualitativen Verfahren zur Exploration möglicher Code-Abstraktionen, so zeigen sich umfassende Überschneidungen. Das Inventar an Operatoren, das etwa durch das Ergänzungspaket TEXTLING von TEXTPACK zur Verfügung gestellt wird, läßt sich unmittelbar auf entsprechende Suchfunktionen der qualitativen Softwarelösungen abbilden (vgl. Abschnitt 3.2.2): Beispielsweise verkörpert der UND-Operator das Äquivalent der qualitativen Codekookkurrenz-Suche; auch zu den Operatoren VOR/RFOLGE und NACH existieren i.d.R. gleichwertige Entsprechungen. Der Unterschied liegt in der Art der Verankerung dieser Konzepte in der Quantitativen bzw. Qualitativen Analyse. Erstere bewerkstelligen eine Ausweitung der algorithmischen Operationalisierung des Codierschritts auf komplexere Bedingungen, die jedoch nach wie vor auf extensionalen, d.h. durch Wortlisten beschriebenen atomaren Kategorisierungskriterien aufbauen - ausschließlich der Computer trifft

hier die Entscheidung, welche Kategorie zugewiesen wird. Umgekehrt verhält es sich im letzteren Falle: In der Regel ist es der Forscher, der - jeweils mit Blick auf das Ausgangsmaterial - sowohl die erste, unstrukturierte Kategorisierung vornimmt als auch letztlich die Entscheidung trifft, ob und ggf. in welcher Weise die per Computerexploration bestimmten Regelmäßigkeiten (als Abstraktionen der elementaren Codezuweisungen) semantisch bzw. pragmatisch zu interpretieren sind; Ziel können hierbei sowohl einfache semantische Sequenzrelationen als auch komplexere, unmittelbar sozialwissenschaftlich relevante Sequenzmuster sein.

Vor diesem Hintergrund stellt sich die Frage, ob die Verfahren zur computergestützten Codierung notwendig auf das bisher Erreichte - die extensionale Operationalisierung der Klassifikationskriterien auf der Grundlage von Wörterbüchern - beschränkt sind. Auch die erste der in Abschnitt 3.2.2 genannten Zielsetzungen der qualitativen Codeabstraktion - die Identifikation einfacher semantischer Sequenzrelationen - zeigt einen unmittelbaren Anknüpfungspunkt für den Einsatz verfeinerter, auf linguistisch fundierten Analysetechniken basierender Verfahren. Die oben genannten Relationen könnten ausgehend von einer verfeinerten syntaktischen Analyse ermittelt werden, da die Schlüsselwörter, auf die sich etwa Huber in seiner Arbeit bezieht, nichts anderes als subordinierende bzw. koordinierende Konjunktionen sind ([Hube92b], S. 142 ff.). Eine alleine auf einem Zeichenkettenvergleich basierende Schlüsselwortanalyse greift zu kurz, da auf dieser Basis nicht die semantisch verknüpften Aussagen ermittelt werden können, sondern allenfalls die *Existenz* einer bestimmten semantischen Verknüpfung.

Aus diesen Erwägungen läßt sich folgendes Programm ableiten:

Es ist zu untersuchen, welche Möglichkeiten bestehen, eine algorithmische Operationalisierung des Codiervorgangs feinkörniger zu bewerkstelligen als auf der Basis von Wörterbüchern sowie oberflächlichen, ggf. durch zusätzliche Verknüpfungen und Kontextoperatoren ohne wirklichen Inhaltsbezug angereicherten Zeichenkettenvergleichen.

Eine Verfeinerung dieses Analyseschritts käme sowohl der computergestützt-quantitativen als auch der computerunterstützt-qualitativen Inhaltsanalyse zugute: im ersteren Fall, weil präzisere Ausdrucksmittel zur Operationalisierung von Kategorienschemata eine Verbesserung der Validität der erzielten Ergebnisse ermöglichen sollte; im letzteren Fall, weil damit verbesserte Werkzeuge zur Verfügung stünden, die auch zur weiterreichenden Unterstützung der Suche nach abstrakten Codes - die bislang ausschließlich auf Operatoren ohne Inhaltsbezug aufbaut - potentiell nutzbringend sind.

Bevor jedoch dieses Programm in Angriff genommen wird, sollen die Funktionsgrundlage sowie die prinzipiellen Grenzen der klassischen Computergestützten Inhaltsanalyse, d.h. der wörterbuchbasierten, extensionalen Algorithmisierung des Kategorisierungsvorgangs, unter Einnahme einer linguistischen Perspektive exploriert werden.

Kapitel 4

Theorie der wörterbuchbasierten Kategorisierung

Die aus der algorithmischen Explikation des Codierschritts folgende maximale Reliabilität ist nur eines der Kriterien zur Bewertung der wörterbuchbasierten Inhaltsanalyse. Die zentrale Anforderung besteht darin, daß das, *was* hierdurch beliebig reproduzierbar wird, überhaupt den *inhaltlichen* Zielsetzungen gerecht wird. Anders ausgedrückt: Der *Objektivitätsanspruch* der Inhaltsanalyse wird eben nicht alleine dadurch eingelöst, daß die durch den Computer bewerkstelligten Transformationen beliebig wiederholbar sind, sondern unterliegt darüberhinaus der Kernbedingung, daß der Input dieses Prozesses - die vom Forscher in Wörterbuchform vorgegebene Operationalisierung des Kategoriensystems - den Ansprüchen genügt, die im Hinblick auf die Forschungsfrage an die Untersuchung zu stellen sind. Welche Aussagekraft könnte den Ergebnissen einer computergestützten Inhaltsanalyse zugeschrieben werden, deren Ergebnisse zwar in technischem Sinne ideal reproduzierbar sind, die aber andererseits für zwei Texte, die im Hinblick auf die Forschungsfrage ähnlich zu interpretieren sind, zu völlig unterschiedlichen Resultaten führt? In den Worten von Lisch und Kriz ([LiKr78], S. 99):

“Ideale Reproduzierbarkeit von Kodierentscheidungen garantiert nämlich noch lange nicht ideale Reproduzierbarkeit von inhaltlichen Ergebnissen. ... Die Reproduzierbarkeit von Kodierentscheidungen spielt ... bestenfalls die Rolle einer notwendigen Voraussetzung.”

Die technische Reproduzierbarkeit gewährleistet keineswegs, daß die Ergebnisse tatsächlich *valide* sind. Was folglich zu leisten verbleibe, umschreibt Fühlau folgendermaßen ([Fühl82], S. 164):

“Für die Vertreter der maschinellen Inhaltsanalyse mag daher die Frage nach Objektivität und Systematik endgültig und zufriedenstellend gelöst sein - zu beweisen ist die Qualität von Computeranalysen nur durch eine Diskussion der Qualität von Computeranalysen und nicht durch den Verweis auf die Funktionstüchtigkeit des Instruments.”

In den folgenden Abschnitten soll diese Fragestellung in bezug auf die wörterbuchbasierte Computergestützte Inhaltsanalyse diskutiert werden.

4.1 Das Wort als Bedeutungseinheit der Kategorienanalyse

Betrachtet man den State-of-the-Art der maschinellen Sprachverarbeitung zu Beginn der sechziger Jahre, so wird verständlich, warum dem Urahn der Computergestützten Inhaltsanalyse, das in jenem Jahrzehnt entwickelte System GENERAL INQUIRER, ein Modell der *wortlistenbasierten* Operationalisierung des Kategorisierungsvorgangs zugrundeliegt. So bedienten sich etwa auch die ersten Ansätze zur maschinellen Übersetzung, dem Schwerpunktthema der frühen computerlinguistischen Forschung, im Wesentlichen einer elementaren und bekanntermaßen unzureichenden Wort-für-Wort-Übersetzungstechnik. Obwohl schon zum damaligen Zeitpunkt das Wissen um die Komplexität sprachlicher Phänomene verfügbar war: Die verlockenden Möglichkeiten des mit absoluter Präzision arbeitenden Werkzeugs Digitalrechner führten zu übertrieben optimistischen Einschätzungen betreffend die Algorithmisierbarkeit komplexer Probleme, für die zum Teil selbst heute - mehr als 30 Jahre später - noch keine annähernd befriedigenden Lösungen vorliegen. Eine treffende Darstellung der Ursache der überzogenen Euphorie, mit denen frühe computerlinguistische Forschungsprojekte initiiert wurden, findet sich bei Weizenbaum ([Weiz76], S. 247):

“Als das, was man damals für Riesencomputer hielt, mehr oder weniger zuverlässig zu arbeiten begann, wurden einige ansonsten ganz vernünftige Leute das Opfer eines Syndroms, das ich mit dem Satz kennzeichnen würde: “Jetzt, da wir (endlich) X haben, können wir auch Y tun”. In der damaligen Situation war X das, was man zu dieser Zeit für sehr umfangreiche Informationsspeicher (Gedächtnisse) und sehr hohe Rechengeschwindigkeiten hielt, und Y war die maschinelle Übersetzung von Sprachen.”

Daß die ersten Lösungsansätze auf dem Gebiet der Sprachverarbeitung die *Zeichenkette* als Bezugspunkt wählten, wird intuitiv verständlich, wenn man sich vergegenwärtigt, auf welchem Abstraktionsniveau ein Programmierer zur damaligen Zeit sein Werkzeug Computer wahrnahm. Die Daten liegen zeichenweise in linear angeordneten Speicherzellen ab, auf die per schrittweisem Inkrement numerischer Indices nacheinander zugegriffen werden kann; ein Vergleich zweier Zeichenketten läßt sich somit unmittelbar mit Hilfe zweier Indexvariablen durchführen. In der Tat sind einige der anspruchsvolleren Probleme, die auf der Ebene der Verarbeitung einzelner Wörter entstehen, auf der Grundlage von allenfalls moderaten Verfeinerungen dieses Basisalgorithmus zumindest approximativ lösbar. Hierzu zählt u.a. das bereits erwähnte *Flexionsproblem*, dem u.a. die folgenden Fälle zuzuordnen sind:

1. einfache Endungsvariationen, z.B. “*Frau*” → “*Frauen*”; “*leben*” → “*lebe*”, “*lebst*”, ...
2. Veränderungen des Stammvokals, z.B. “*Mann*” → “*Männer*”, ... ; “*sterben*” → “*starb*”, ...
3. Präfigierung, z.B. “*leben*” → “*gelebt*”, ...

4. abgespaltene Präfixe, z.B. “ablösen” → “Er löste den Bundeskanzler ab”, ...

Zumindest für die ersten drei Fallklassen ist eine Lösung durch Enumeration aller Formen (“*Vollformenlexikon*”) möglich. Als elegantere Alternative mit geringerer lexikalischer Redundanz kommt eine Technik in Frage, bei der zwischen Stamm- und Endungsanteil unterschieden wird.¹ Den Ausschlag für die Kategorisierungsentscheidung gibt jeweils der Stamm, wobei dem Endungsanteil eine disambiguierende Funktion zukommen kann. Dieses zuletztgenannte Verfahren kann auch durch ausschließliche Berücksichtigung des Stammanteils approximiert werden, wobei jedoch mit einem Verlust an Präzision zu rechnen ist: “*Haus*” und “*hausen*” - das am Satzanfang ja auch großgeschrieben wird - wären hier u.U. nicht mehr zu unterscheiden.

Problematisch hingegen erscheint Fallklasse 4, denn wie soll ohne eine i.a. viel aufwendigere syntaktische Analyse entschieden werden, daß es sich bei “*ab*” um ein abgespaltenes Präfix der Verbform “*löste*” handelt und nicht etwa um eine Präposition?

Auch in einem anderen wichtigen Fall, der *Wortform-Ambiguität*, ist eine Entscheidung zum Teil erst unter Berücksichtigung des syntaktischen Kontexts möglich. Zwar hilft in vielen Fällen die durch Groß- oder Kleinschreibung vorgegebene Differenzierung, jedoch existieren Belege, für die dieses Kriterium nicht greift: “*Würde*” kann sowohl als Substantiv oder als Hilfsverb (in Fragesätzen) satzinitial vorkommen und wäre dann ohne eine feinere Kontextanalyse nicht zu unterscheiden.

Ein wichtiges Problem stellen *lexikalische Ambiguitäten* dar: Im Falle von *Homonymie* mit Unterscheidung im grammatischen Geschlecht (“*der Heide*” vs. “*die Heide*”) kann die Disambiguierung - sofern die Genusdifferenz im konkreten Beleg abgebildet ist - unter Rückgriff auf den syntaktischen Kontext erfolgen; bei *Polysemie* hingegen (“*der Sturm*” in der Meteorologie bzw. im Fußball) ist ein solches oberflächlich-distinktives Merkmal nicht verfügbar. Auch hierfür existiert jedoch, wie zuvor ausgeführt, ein Verfahren - die von Stone entwickelte KWIC-Technik -, das es erlaubt, eine mit hoher Wahrscheinlichkeit korrekte Entscheidung zu fällen, ohne jedoch eine tiefgehende Analyse durchzuführen.

Darüberhinaus ergeben sich Probleme durch das sprachliche Phänomen der *Kompositabildung*. Insbesondere in deutschsprachigen Texten kommen i.a. eine Vielzahl zusammengesetzter Wörter vor, deren vollständige Enumeration aufgrund der prinzipiellen Unbeschränktheit impraktikabel ist. Eine nähere Analyse ist jedoch zumindest aus theoretischer Sicht unabdingbar, denn wie sonst soll entschieden werden, daß ein im Lexikon nicht enthaltenes Wort wie u.U. “*Hausbriefkasten*” nicht - entsprechend der Präfixübereinstimmung - als Gebäude zu interpretieren ist? Diese Aufgabe kann mit Hilfe von Spezialalgorithmen bearbeitet werden, die aber lediglich eine einfache Heuristik, umfassend eine Folge elementarer Zeichenkettenvergleiche, durchführen und daher nicht in allen Fällen die richtige Entscheidung treffen (vgl. etwa die “*Strategie der Wortzerlegung von links*” von Rohland: [Roh177]).

Die Ausführungen verdeutlichen, daß zumindest ein Teil der mit einer einzelwortbezogenen Analyse verbundenen Probleme ohne Tiefenanalyse, d.h. auf der Ebene oberflächlicher Zeichenkettenvergleiche durch heuristische Algorithmen mit hinreichender Präzision gelöst werden können. Jedoch wird ferner ersichtlich, daß es sich nur um *Ersatzverfahren* handelt, da keine Theorie zugrundeliegt, auf deren Basis *Textinhalte* erschlossen werden. Ein elementares Beispiel zeigt auf, wo die prinzipielle Grenze der Verfahren liegt, die ausschließlich mit zeichenweisem Vergleich von Wörtern arbeiten:

¹Für das Flexionsproblem existieren mittlerweile computerlinguistische Softwarelösungen, die auf dieser Technik basieren und einen extrem hohen Abdeckungsgrad bei einer sehr hohen Analysegeschwindigkeit erreichen. (Vgl. u., Abschnitt 5.2)

“Bundeskanzler Helmut Kohl trat den Urlaub an. Wie in jedem Jahr fuhr er an den Wolfgangsee.”

Wie soll das Pronomen “*er*” als weitere Erwähnung von “*Helmut Kohl*” erkannt werden? Für das Ergebnis der Inhaltsanalyse sollte es i.a. irrelevant sein, ob eine Referenz an der sprachlichen Oberfläche mit Hilfe eines nichtpronominalen Ausdrucks oder einer Pronominalanapher ausgedrückt wird.² Es ist unmittelbar ersichtlich, daß mit elementaren Zeichenkettenvergleichen alleine nichts auszurichten ist. Die Strategie, die für Synonymie angewandt wird - die möglichst umfassende Enumeration aller Synonyme in der entsprechenden Kategoriedefinition - erweist sich in diesem Fall als untauglich. Die Auflösung pronominaler Verweise bedarf einer tieferen, inhaltsbezogenen Analyse, in der auch satzübergreifende Bezüge erschlossen werden. Eine Verallgemeinerung der KWIC-Technik, deren Betrachtungsfenster sich ohne weiteres auch über Satzgrenzen hinaus ausdehnen ließe, greift zu kurz: Die Mehrdeutigkeit ist nicht lexikalisch, sondern textuell determiniert und damit potentiell *offen*; eine Lösung ähnlich der zuvor für Homonymie und Polysemie angegebenen kann es somit nicht geben. Unter Anbetracht der komplexen Determinanten dieses Problems der Pronomenresolution gelangt Fühlau zu folgender pessimistischer Bewertung ([Fühl82], S. 140, eingeschobenes Zitat nach Deichsel und Stone: [DeSt75], S. 114)

“Grundsätzlich werden Fragen wie Deixis und Pronominalisierung (“a nasty content analysis problem”) wohl kaum von Computern zu lösen sein, zumal dazu die ganze Spannbreite menschlichen Wissens und menschlicher Erfahrung notwendig ist.”

und folgert hieraus, pronominale Verweise müßten

“... als Teil der Textaufbereitung von Hand substituiert werden.”

(ibid, S. 171). Diese Einschätzung wird in der Literatur zur computergestützten Inhaltsanalyse generell vertreten.

Somit erweist sich das Pronomenproblem als in hohem Maße *relevant* für die automatische Kategorisierung, es läßt sich jedoch im Rahmen einer elementaren, zeichenkettenorientierten Oberflächenanalyse nicht zufriedenstellend lösen. Es liegt für Verfahren der Einwortanalyse jenseits der Grenze des Machbaren und stellt gleichzeitig eine anspruchsvolle Herausforderung für die Entwicklung inhaltsnäherer Analysestrategien dar. Ohne den detaillierten Ausführungen in Teil II der Arbeit vorweggreifen zu wollen - die pessimistische Einschätzung von Fühlau und Anderen betreffend die Machbarkeit algorithmischer Verfahren zur Pronomenauflösung ist mittlerweile als widerlegt anzusehen.

In den bisherigen Betrachtungen wurde die Perspektive des spezifischen Modells eingenommen, das der klassischen Computergestützten Inhaltsanalyse zugrunde liegt. Gegeben die Einzelwortanalyse, welche linguistischen Probleme bestehen? Inwieweit ist eine algorithmische Lösung dieser Probleme innerhalb des Analyseparadigmas möglich? Mindestens ebenso wichtig ist jedoch die Diskussion aus der Perspektive der inhaltsanalytischen *Zielsetzung*. Welchen Ansprüchen

² Anders hingegen könnte es sich verhalten, wenn “*Helmut Kohl*” mit Hilfe des Ausdrucks “*Birne*” wiederaufgegriffen wird. In diesem Falle könnte der neben der auch identischen Referenz transportierte (abweichende) *Sinn* des Ausdrucks durchaus relevant für die inhaltsanalytischen Zielsetzungen sein.

wird eine Einwortanalyse überhaupt gerecht? Anders ausgedrückt: Stellen Aufzählungen einzelner Wörter überhaupt ein geeignetes Instrument zur Spezifikation inhaltlich unterschiedlicher Kategorien dar? Wo liegen die Grenzen derartiger Beschreibungen?

4.2 Isotopie

Zur Erörterung dieser Frage bietet sich ein Exkurs in die Textlinguistik an. Textkonstituierender Zusammenhang wird auf vielfältige Weise etabliert. Referentieller Wiederaufgriff - zuvor am Spezialfall des Pronomenproblems diskutiert - stellt nur eine von vielen Ebenen dar, entlang derer sich Textualität entwickelt. Eine spezifische Variante ist die von A. J. Greimas mit dem Begriff *Isotopie* belegte (Bußmann: [Buß90], S. 357)³

“Wiederkehr von Wörtern desselben Bedeutungs- bzw. Erfahrungsbereichs in einem Text, z.B. Arzt, Fieber, Spritze, Honorar.”

Somit bezeichnet Isotopie die im allgemeinen nicht mit einer Koreferenzbeziehung einhergehende *Rekurrenz einer bestimmten Teilmenge semantischer Merkmale*.

Der Begriff ist recht vage gehalten und läßt offen, was unter Übereinstimmung des “Bedeutungs- bzw. Erfahrungsbereichs” genau zu verstehen ist. Auf jeden Fall aber ist erst auf Basis der Interpretation eines konkreten Textes zu entscheiden, welches die jeweils rekurrerten Merkmale sind. Während im Isotopiebeispiel der gerade betrachteten Definition ein semantisches Merkmal von “Arzt” wiederaufgegriffen wird, das sich vielleicht am besten als *Medizin: +* umschreiben läßt, könnte es in einem anderen Text, in dem von unterschiedlichen Berufen die Rede ist, ein Merkmal *Berufsbezeichnung: +* sein, welches eine bestimmte Isotopieebene umschreibt (“Arzt”, “Soziologe”, “Computerlinguist”, “Tierheilpraktikerin”, ...). In einem dritten Text ist es vielleicht die Kombination dieser beiden Merkmale, die eine der Isotopieebenen umschreibt (“Arzt”, “Krankenpfleger”, “Laborassistent”, “Rettungssanitäter” ...). In Spezialfällen kann die Gemeinsamkeit auch in einer Koreferenz auf ein durch den sprachlichen Ausdruck eindeutig bestimmtes Objekt bestehen (“Papst”, “Heiliger Vater”, “Johannes Paul II”, “Karol Wojtyla”, ...).

Spannt man den Bogen zurück zur Inhaltsanalyse, so läßt sich feststellen, daß die wörterbuchbasierte Operationalisierung der Kategorien in der computergestützten Inhaltsanalyse als Spezifikation einer Menge von Isotopieebenen auffaßbar ist. Jede Kategorie entspricht einer Isotopieebene; die Wortlisten sind vollständige Enumerationen der sprachlichen Ausdrücke, denen ein gemeinsames Merkmal - umschreibbar als *Kategorie_i: +* - zugeordnet wird. Folglich werden per Codierung und Häufigkeitsauszählung unterschiedliche Isotopieebenen eines Textes bestimmt: Kategorien, die häufig instanziiert werden, entsprechen Folgen von Bedeutungseinheiten

³Man könnte hier nahezu ebensogut auf den Begriff der *Kontiguität* zurückgreifen, der in Bußmann folgendermaßen definiert ist (ibid., S. 418):

“... (2) In der Semantik Relation zwischen Lexemen, die der gleichen semantischen, logischen, kulturellen oder situationellen Sphäre angehören. ... ”

Die Isotopie nach Greimas verankert Kontiguität als Gemeinsamkeit lexikalisch-semantischer Merkmale, die auf Basis einer sog. Komponentenanalyse gewonnen werden, im Lexikon (vgl. auch Hellwig: [Hell84], S. 57). Da die lexikalische Dimension in der computergestützten Inhaltsanalyse in Form des Wörterbuchs in Erscheinung tritt, wird im Text der Isotopiebegriff verwendet.

(Wörtern), die zum textuellen Zusammenhang über das den Vertretern dieser Isotopieklasse gemeinsame Bedeutungsmerkmal beitragen. Der computergestützten Inhaltsanalyse im klassischen Sinne liegt also implizit eine Isotopieanalyse entlang extensionaler Isotopieebenen-Definitionen im Wörterbuch zugrunde.

Aus textlinguistischer Perspektive kann Isotopie als ein auf der lexikalischen Ebene angesiedeltes Ausdrucksmittel zur *Etablierung textuellen Zusammenhangs* angesehen werden. Die unterschiedlichen Isotopieebenen eines Textes verkörpern - intuitiv gesprochen - elementare, in Einzelwörtern verankerte Text-“Themen”, die durch die einfachen Algorithmen einer wörterbuchbasierten Inhaltsanalyse bestimmbar sind. Die Entsprechung mit thematischen Einheiten wird auch andernorts wahrgenommen (Bußmann: [Buß90], S. 357):

“In der Anzahl der Isotopieebenen spiegelt sich die thematische Komplexität eines Textes.”

Eine vergleichbare Beobachtung findet sich bei Hellwig - allerdings nicht in bezug auf die wörterbuchbasierte Inhaltsanalyse, sondern vielmehr aus der Perspektive der Vergabe von Schlagwörtern in der Textarchivierung ([Hell84], S. 58):⁴

“Ich kann nur vermuten, daß sich Isotopie eines Tages als Rekurrenz von Schlagwörtern definieren läßt, die Redeteilen aufgrund der Zugehörigkeit zu bestimmten Wissensbereichen zugeteilt werden.”

Betrachtet man die Namen der unterschiedlichen Kategorien als Schlagwörter, so wird ersichtlich, daß sich die wörterbuchbasierte Inhaltsanalyse als *Algorithmisierung des Prozesses der Schlagwortvergabe* auffassen läßt.

Es gibt eine Vielzahl weiterer sprachlicher Konstrukte, die typische Stifter textuellen Zusammenhangs und somit potentiell “thematische” Elemente sind. Folglich liegt es auf der Hand, zu untersuchen, welchen Beitrag eine Einbeziehung dieser Ausdrucksmittel für eine Kategorienanalyse verfeinerten Zuschnitts leisten könnte. Die Ergebnisse der folgenden Fallstudie werden dazu beitragen, Aussagen zum zweiten Kernthema dieses Kapitels - der Frage nach den *Grenzen* der wortorientierten Kategorienanalyse - abzuleiten. Den geeigneten Ausgangspunkt bilden die Definitionen einiger grundlegender textlinguistischer Begriffe, umfassend zwei unterschiedliche Ebenen textuellen Zusammenhangs - *Kohäsion* und *Kohärenz* - sowie den bislang ausschließlich intuitiv verwendeten Begriff des *Themas*.

4.3 Konstitution von Themen - eine textlinguistische Fallstudie

In der vorausgegangenen Diskussion wurde die Übereinstimmung lexikalisch-semantischer Merkmale in Isotopieketten als sozusagen *atomares*, da einzelwortbasiertes sprachliches Mittel zur Herstellung textuellen Zusammenhangs identifiziert. Dem gegenüberzustellen sind nunmehr die

⁴Hellwig spricht allerdings ausdrücklich *nicht* von Themen, die für ihn auf einer pragmatischen Ebene der Textstruktur angesiedelt sind (vgl. u.).

in irgendeiner Weise *strukturierten* Konstrukte, die komplexere thematische Einheiten etablieren. Diese Aufgabe soll auf der Grundlage einer Fallstudie⁵ geschehen, deren erstes Ziel darin besteht, zu einem intuitiven Verständnis inhaltlicher Einheiten zu gelangen, die sich als Instanzen textueller Themen auffassen lassen. Zunächst aber sollen einige grundlegende textlinguistische Begriffe definiert werden.

4.3.1 Begriffsdefinitionen

Textueller Zusammenhang läßt sich zumindest zwei unterschiedlichen Ebenen zuordnen, die gemeinhin mit den Begriffen *Kohäsion* und *Kohärenz* belegt werden.⁶

Definition 4.1 (Kohäsion) *Ein Text heißt kohäsiv, wenn er einen durch lexikalisch-semantische oder formalgrammatische Mittel hergestellten Zusammenhang aufweist.*

Typische Spielarten von Kohäsion sind Isotopie/Kontiguität (lexikalische Kohäsion), Koreferenz (Pronomenverwendung usw.), Ellipse⁷ und Konjunktionen (vgl. u. bzw. Halliday und Hasan: [HaHa76]). Kohäsion ist jedoch keineswegs hinreichend dafür, daß eine Satzabfolge von einem kompetenten Sprecher der jeweiligen Sprache als *zusammenhängend* empfunden wird, wie an folgendem Beispiel ersichtlich wird:

“Helmut Kohl ist Bundeskanzler. Der Bundeskanzler wird vom Bundestag gewählt. Der Bundestag hat seinen Sitz in Bonn. Es liegt in der tiefsten Provinz.”

Zwar ist die Satzabfolge augenscheinlich kohäsiv - die Elemente, die die Kohäsion auf der referentiellen Ebene etablieren, sind unterstrichen -, jedoch fehlt offenbar der Zusammenhang auf der pragmatischen Ebene. Es ist schwerlich eine Verwendungssituation vorstellbar, in der diese Ansammlung von Äußerungen als *Text* interpretierbar wird. Es fehlt die *Kohärenz*.

Definition 4.2 (Kohärenz) *Ein Text heißt kohärent, wenn er relativ zu einem bestimmten Verwendungszusammenhang pragmatisch sinnvoll erscheint.*

Kohärenz ist demnach keine absolute Größe, da die Einschätzungen der Kommunikationspartner sowie der extratextuelle Kontext einfließen.

Auf der Grundlage dieser beiden Begriffe erarbeitet Hellwig ein zweidimensionales Textmodell, in dem folgende Formen der Textstrukturentwicklung unterschieden werden ([Hell84]):

- *referentielle Progression*: die *sequentielle* Entwicklung der Textstruktur über *kohäsive* Anknüpfungen,

⁵Die Studie wurde im Rahmen einer textlinguistischen Untersuchung zur Problematik der Pronomenresolution erarbeitet (Stuckardt und Haenelt: [StuH94]).

⁶Die im folgenden gegebenen Definitionen orientieren sich erneut an Bußmann ([Buß90]) und Hellwig ([Hell84]) sowie darüberhinaus an Halliday und Hasan ([HaHa76]).

⁷sprachliche Auslassungen, z.B. in “Beate ging und *sje* ärgerte sich.”

- *thematische Progression*: die *hierarchische* Entwicklung der Textstruktur durch *kohärente* Anknüpfungen.

Somit ist im Modell von Hellwig die thematische Textentwicklung auf der pragmatischen Ebene angesiedelt. Zwar soll der Begriff des Themas im Hinblick auf die Zielsetzung der Modellierung der sozialwissenschaftlichen Themenanalyse im folgenden abweichend definiert werden; jedoch erweist sich die Unterscheidung dieser beiden Progressionstypen als intuitiv ansprechend und aus textlinguistischer Sicht adäquat. Die Detailuntersuchung des Problems der Pronomenresolution in Teil II der Arbeit wird diese Aussage untermauern (vgl. Abschnitt 7.1.4).

Speziell die von Hellwig vorgeschlagene Explikation der pragmatischen Entwicklung eines Textes durch *zweifach kontextualisierte Fragen* stellt ein sowohl intuitiv als auch theoretisch ansprechendes Kriterium zur interpretativen Determination der Kohärenz eines Textes dar: Eine Äußerung U_{i+1} knüpft genau dann kohärent an den durch die Äußerungsabfolge U_1, \dots, U_i vermittelten intratextuellen Kontext sowie den extratextuellen Kontext an, wenn sich eine Frage F_{i+1} formulieren läßt, die die folgenden Bedingungen erfüllt:

- F_{i+1} ist nach der Lektüre von U_1, \dots, U_i bzw. im Verwendungskontext naheliegend;
- U_{i+1} trägt - zumindest ansatzweise - zur Beantwortung von F_{i+1} bei.

In den Worten von Hellwig (ibd., S. 62):

“Ich behaupte nun, daß die Kohärenz eines monologischen Textes tatsächlich im Fraglichen liegt, das vorab bestanden hat oder sich im Verlauf des Textes aus dem bisher Gesagten ergibt und so die Aussagen implizit verbindet.”

Hellwig geht noch einen Schritt weiter, indem er den textlinguistischen Begriff des Themas und das im obigen Sinne in bezug auf einen Text Fragliche (weitgehend) gleichsetzt (ibd., S. 65):

“Meine These ist, daß der Begriff des Themas gleichgesetzt werden kann mit dem Fraglichen, das ich bisher als konstitutiv für die zweite Dimension der Struktur monologischer Texte angesetzt habe.”

In der hierarchischen und somit baumförmigen Kohärenzstruktur entspricht somit jedem Teilbaum ein Fragesatz und damit ein (Teil-)“Thema”.

Für den Begriff des Themas werden in der Linguistik eine Vielzahl unterschiedlicher Definitionen vorgeschlagen. Der Ansatz von Hellwig führt eine Variante ein, die insbesondere strikt vom Themenbegriff zu unterscheiden ist, der der Thema-Rhema-Dichotomie in der Funktionalen Satzperspektive der Prager Schule zugrunde liegt; in letzterem Modell verkörpern “Themen” i.a. Entitäten des *referentiellen* Wiederaufgriffs, deren textbildendes Potential somit auf der Ebene der Kohäsion angesiedelt ist.⁸

⁸In voller Konsequenz spricht so z.B. Daneš von *thematischer Progression*, wenn Hellwig die Bezeichnung referentielle Progression gebrauchen würde (vgl. [Dane70]).

Weder der pragmatische Themenbegriff nach Hellwig noch das i.a. auf der referentiellen Ebene angesiedelte "Thema" der Funktionalen Satzperspektive erweisen sich im Hinblick auf den Untersuchungsgegenstand der vorliegenden Arbeit als geeignet. Soll Isotopie als Spezialfall thematischer Rekurrenz auffaßbar sein, so stellt eine Definition von "Thema" als konstitutive Entität der pragmatischen Textstruktur keinen geeigneten Ausgangspunkt dar. Eine ausschließlich referentielle Lesart erweist sich offensichtlich als zu eingeschränkt.

Es soll daher eine dritte Möglichkeit verfolgt werden, die in wesentlichen Teilen auf der Ebene der Kohäsion verankert ist. Ganz allgemein stellen Themen inhaltliche Einheiten dar, die im Text wiederholt aufgegriffen werden. Da die durch den Text kommunizierten Inhalte gleichzeitig den *textuellen Kontext* ausmachen, läßt sich der Begriff des Themas folgendermaßen definieren:

Definition 4.3 (Thema) *Ein Thema ist ein textuell etablierter (Teil-)Kontext, der im Text mehr als einmal instanziiert, d.h. wiederaufgegriffen wird.*

Dis Kohäsionsvarianten Isotopie und Koreferenz stellen sozusagen *atomare Spezialfälle* dar, in denen lexikalisch-semantische bzw. referentielle Entitäten reinstantiiert werden. I.a. beinhaltet der textuelle Kontext jedoch komplexere inhaltliche Einheiten, die ebenfalls für einen Wiederaufgriff in Frage kommen. Speziell um diese potentiellen Induktoren nichtatomarer Themenstränge geht es in der folgenden Fallstudie.

4.3.2 Die Fallstudie

Abbildung 4.1 zeigt den Text - einen Auszug einer Architektenbiographie -, auf dessen Basis die Fallstudie durchgeführt werden soll.⁹ Den sich anschließenden Betrachtungen wird kein spezifisches Modellierungsverfahren für den textuellen Kontext zugrunde gelegt;¹⁰ die Diskussion bewegt sich zunächst auf einer eher informellen Ebene. Im nachfolgenden Abschnitt sollen die Ergebnisse der Fallstudie systematisch aufbereitet werden, indem die für die Etablierung komplexerer Themen relevanten sprachlichen Ausdrucksmittel identifiziert und unterschiedlichen Beschreibungsebenen zugeordnet werden.

Zunächst sollen auch hier die wichtigsten Isotopieebenen des Beispieltexsts identifiziert werden, da diese in vielen Fällen die Grundlage der Konstitution komplexerer Themenstränge verkörpern. In der linken Spalte wird jeweils ein die Isotopieebene treffend umschreibendes "Schlagwort" angegeben; die rechte Spalte enthält (in unflektierter Form) die lexikalischen Entitäten, die die jeweilige Ebene im Text realisieren.¹¹

⁹Die fortlaufende Nummerierung der Sätze wurde als Bezugspunkt für die nachfolgende Diskussion hinzugefügt.

¹⁰In der Literatur zur Textlinguistik finden sich mannigfaltige Beschreibungsvorschläge, die auf den unterschiedlichsten Theorien basieren - vgl. etwa die Arbeit von Haenelt ([Haen95]), in der ein mehrschichtiges Textmodell entwickelt wird, das unterschiedliche Ansätze miteinander kombiniert.

¹¹Eine derartige Auflistung der Isotopieebenen illustriert die obige Aussage, wonach Schlagwortvergabe und Analyse der Isotopieebenen durch vergleichbare Prozesse beschreibbar sind; die postulierte Entsprechung zur wörterbuchbasierten Inhaltsanalyse ergibt sich unmittelbar, da sich die Wortlisten als Operationalisierung eines entsprechenden Kategoriensystems auffassen lassen.

(1) Behrens, Peter, *1868 in Hamburg, +1940 in Berlin. (2) Behrens entwickelte als einer der ersten Architekten des 20. Jahrhunderts eine architektonische Konzeption, die den Anforderungen der industrialisierten Zivilisation gerecht wurde - zu einer Zeit, in der die Gesellschaft noch in archaischen Vorstellungen dachte, gleichzeitig aber blind auf die überwältigenden Fortschritte der Technik vertraute. (3) Während die moralischen und sozialen Forderungen der expressionistischen Maler aus Dresden (Die Brücke) zu neuen Versuchen in der bildenden Kunst führten, stand Behrens am Beginn der modernen Architektur in Deutschland, auf die er zwischen 1900 und 1914 einen entscheidenden Einfluß ausübte. (4) Durch seine Tätigkeit als Formgestalter erschloß er außerdem um 1907 ein neues Spezialgebiet, das heute unter der Bezeichnung "industrial design" bekannt ist. (5) Auch hier hatte die Arbeit von Behrens nachhaltige Wirkung auf die Entwicklung von Stil und Technik; (6) die vom →Jugendstil inspirierten Formen des Kunsthandwerks erschwerten zu jener Zeit noch eine Verwirklichung der Gestaltungsprinzipien, die der neuen Lebensweise entsprachen.

Abbildung 4.1: Erster Absatz einer Architektenbiographie (aus: Lampugnani, [Lamp83])

<Architektur>	Architekt, architektonisch, Architektur
<bildende Kunst>	Architekt, architektonisch, Architektur, expressionistisch, Maler, Kunst, Formgestalter, "industrial design", Stil, Technik, Jugendstil, Kunsthandwerk
<wirken>	entwickeln, Einfluß, ausüben, Tätigkeit, erschließen, Arbeit, Wirkung, Entwicklung, Verwirklichung
<vorantreiben>	entwickeln, Einfluß, erschließen, Entwicklung, Verwirklichung
<künstlerischer Beruf>	Architekt, Maler, Formgestalter
<rückständig>	archaisch
<progressiv>	industriell, Zivilisation, Fortschritt, neu, modern

Welche der qua lexikalischem Gehalt möglichen Isotopieebenen letztendlich Themenstränge konstituieren, hängt - wie gesagt - von der Interpretation des Textes ab. In diesem Sinne werden Themen durch die pragmatische Dimension der Textentwicklung *determiniert*. (Anders als in der zitierten Arbeit von Hellwig wird jedoch nicht die Auffassung vertreten, daß Themen selbst explizierende Entitäten der pragmatischen Ebene sind.). Die angegebenen Ebenen erscheinen allesamt relevant im Hinblick auf eine kohärente Interpretation des Textes, einer Biographie des Architekten Peter Behrens:

- Da es in Architektenbiographien primär um *Architektur* geht, liegt es auf der Hand, daß verwandte semantische Merkmale auch als themenkonstituierend wahrgenommen werden;
- ebenso sollte eine Diskussion aus der allgemeineren Perspektive der *bildenden Kunst* themenrelevant sein;
- in Biographien steht allgemein das *Wirken* der beschriebenen Person im Vordergrund;
- in Architektenbiographien sind die unterschiedlichen Tätigkeiten dieser Person auf den unterschiedlichen Gebieten der bildenden Kunst von besonderem Interesse;
- gerade das Wirken eines Künstlers spielt sich oftmals im Spannungsfeld zwischen *Reaktion* und *Progression* ab.

Das zuletztgenannte Beispiel zweier in semantischem Kontrast stehenden Isotopieebenen gibt einen ersten Anhaltspunkt für die Beschränkung, der eine rein lexikalische, auf Isotopieketten basierende Themenkonstitution unterliegt. Zwar können in der vorliegenden Biographie jeweils mehrere lexikalische Instanzen der mit <rückständig> und <progressiv> beschriebenen Themenstränge identifiziert werden; das im Text offenbar im Vordergrund stehende, durch den Gegensatz zwischen Überkommenem und Neuem aufgeworfene *Spannungsverhältnis* wird jedoch nicht lexikalisch thematisiert, was durch Wörter wie “Wandel”, “Umbruch”, “Umwälzung” etc. auch allenfalls ansatzweise möglich wäre. Stattdessen kommen nichtatomare inhaltliche Elemente ins Spiel, die einer “*syntaktisch-strukturellen* Beschreibungsebene, einer kompositionalsemantischen Ebene oder gar der vom außertextuellen Kontext vermittelten *Hintergrundinformation* (dem Weltwissen der Interpreten) entspringen.

Da sich die Kontextkonstitution jedoch mit der linearen Entwicklung des Textes vollzieht, soll nicht mit diesem relativ komplexen Fall begonnen werden, der erst zu Ende des in Abbildung 4.1 gezeigten Textausschnitts in Erscheinung tritt. Die Äußerungen sollen der Reihe nach und unter dem Gesichtspunkt des thematischen und pragmatischen Anschlusses an den jeweiligen Kontext diskutiert werden. Die themenrelevanten Entitäten werden jeweils an der Stelle ihres erstmaligen Wiederaufgriffs, d.h. *rückschauend*, identifiziert und beschrieben.

Äußerung 1: “Behrens, Peter, *1868 in Hamburg, +1940 in Berlin.”

Auf der pragmatischen Ebene wird an den Verwendungskontext von Architekten-Kurzbiographien angeknüpft. Die kohärenzexplizierende Frage könnte etwa wie folgt lauten:

Welches sind die wichtigsten Fakten, die über Leben und Wirken von Peter Behrens bekannt sind?

Äußerung 1 trägt lediglich partiell zu deren Beantwortung bei, da nur die Geburts- und Todesdaten kommuniziert werden. Folglich bleiben Fragen offen, die vom verbleibenden Text beantwortet werden sollten.

Äußerung 2: “Behrens entwickelte als einer der ersten Architekten des 20. Jahrhunderts eine architektonische Konzeption, die den Anforderungen der industrialisierten Zivilisation gerecht wurde - zu einer Zeit, in der die Gesellschaft noch in archaischen Vorstellungen dachte, gleichzeitig aber blind auf die überwältigenden Fortschritte der Technik vertraute.”

Der thematische Zusammenhang mit Äußerung 1 manifestiert sich primär auf der referentiellen Ebene: In beiden Äußerungen ist von ein und derselben Person - dem Architekten Peter Behrens - die Rede. Die lexikalisch-semantisch vermittelte Rekurrenz des Ausdrucks “20. Jahrhundert” auf einen Ausschnitt des in Äußerung 1 auf komplexere Weise kommunizierten Lebenszeitraums stellt einen weiteren, wohl eher sekundären Faktor dar.

Aus der pragmatischen Perspektive sind Äußerungen 1 und 2 *nebengeordnet* - die Aussage von Äußerung 2 beantwortet nicht etwa eine Frage, die sich aus Äußerung 1 ergibt, sondern initiiert die Beantwortung des noch offenen Teils der Ausgangsfrage des Verwendungskontexts, da mit der Beschreibung des Lebenslaufs begonnen wird.¹²

¹²Statt von der schrittweisen Beantwortung *einer* generellen Frage auszugehen, könnte man alternativ postulieren, daß der Verwendungskontext für Biographien eine Abfolge *zweier* oder gar *mehrerer* spezifischer Fragen

Äußerung 3: “Während die moralischen und sozialen Forderungen der expressionistischen Maler aus Dresden (*Die Brücke*) zu neuen Versuchen in der bildenden Kunst führten, stand Behrens am Beginn der modernen Architektur in Deutschland, auf die er zwischen 1900 und 1914 einen entscheidenden Einfluß ausübte.”

Gemäß der atomaren Ebene des lexikalisch-semantischen Wiederaufgriffs erscheinen u.a. die Isotopieklassen <Architektur>, <künstlerische Berufe> und <wirken> als maßgeblich für den textuellen Zusammenhang. Diese Aussage kann auf der Grundlage tiefergehender Betrachtungen präzisiert werden. Wohl kaum ein menschlicher Interpret käme zu dem Ergebnis, daß die primäre Themenentwicklung etwa entlang der Aufzählung unterschiedlicher künstlerischer Berufe (*Architekt, . . . , Maler*) vonstatten geht. Vielmehr scheint der thematische Wiederaufgriff einer komplexen inhaltlichen Entität vorzuliegen, die aus einer Kombination von Instanzen der Isotopieebenen <wirken> (vermöge “*Einfluß*”/“*ausüben*”) und <Architektur> sowie dem referentiellen Wiederaufgriff von “*Behrens*” hervorgeht und sich wie folgt umschreiben läßt:

Peter Behrens wirkt als Architekt.

Speziell in Äußerung 3 wird dieses nichtatomare Thema nur sehr indirekt ausgedrückt. In den Prozeß der Determination, daß (erneut) etwas über Behrens’ Tätigkeit als Architekt ausgesagt wird, fließt Hintergrundinformation etwa der folgenden Art ein: *Jemand, der einen entscheidenden Einfluß auf die moderne Architektur ausübt (und bereits als Architekt in Erscheinung getreten ist), ist i.a. als Architekt tätig.* Bereits dieses inhaltlich vergleichsweise einfach gelagerte Beispiel zeigt auf, wie komplex sich der Prozeß der Themenentfaltung i.a. gestaltet. Äußerung 3 ist Äußerung 2 pragmatisch subordiniert, da Information beigetragen wird, die im Anschluß an die Interpretation von Äußerung 2 potentiell fraglich ist:

In welcher Relation zu innovativen Strömungen auf anderen Gebieten der bildenden Kunst steht Behrens’ Wirken als Architekt?

Äußerung 4: “Durch seine Tätigkeit als Formgestalter erschloß er außerdem um 1907 ein neues Spezialgebiet, das heute unter der Bezeichnung “*industrial design*” bekannt ist.”

Das zuvor beschriebene komplexe Themenmuster wird erneut aufgegriffen, allerdings nur partiell: Nicht mehr die Arbeit von Behrens als *Architekt*, sondern vielmehr dessen Tätigkeit als *Formgestalter* steht nun im Mittelpunkt der Ausführungen. Für diese Art der Themenkonstitution per Variation (von i.a. referenzierenden *Mitspielern* einer kommunizierten *Situation*, vgl. u.) soll im folgenden der Terminus *Modulation* verwendet werden. Im vorliegenden Fall geschieht die Modulation entlang der Isotopieebene <künstlerischer Beruf>, so daß sich das nunmehr abstraktere Thema folgendermaßen umschreiben läßt:

Peter Behrens übt einen künstlerischen Beruf aus.

nahelegt, die die Erwartungshaltung sequentialisieren, mit der man i.a. eine Kurzbiographie liest: Geburts- und Todesdaten folgt eine Beschreibung des Lebenslaufs sowie ggf. eine Liste wichtiger Veröffentlichungen.

Modulation stellt eine der grundlegenden Spielarten des Wiederaufgriffs komplexer Themenmuster dar. Daß ein erneuter Beitrag zu einem bereits installierten nichtatomaren Thema erfolgt, wird hier zusätzlich oberflächennah, und zwar über die Konjunktion “*außerdem*”, reflektiert.¹³ Die Relation des thematischen Wiederaufgriffs liefert den entscheidenden Anhaltspunkt für die pragmatische Interpretation. Äußerung 2 - der Initiator des Themas - und Äußerung 4 tragen beide zu der von Äußerung 1 induzierten Frage bei; sie sind daher auf der pragmatischen Ebene *nebengeordnet*.

Äußerung 5: “*Auch hier hatte die Arbeit von Behrens nachhaltige Wirkung auf die Entwicklung von Stil und Technik;*”

Der thematische Anschluß von Äußerung 5 ist erneut relativ komplex ausgestaltet. Auf der lexikalischen Ebene wird einmal mehr die Isotopieklasse <*wirken*> instantiiert. Eine weitergehende semantische Interpretation offenbart Parallelen zu einem strukturierten Thema, das von Äußerungen 2 und 3 gemeinsam beigetragen wird. Die Partikel “*auch*” gibt den entscheidenden Anhaltspunkt: In Verbindung mit dem anaphorischen Ausdruck “*hier*”, der den von Äußerung 4 kommunizierten Sachverhalt wiederaufgreift, legt diese nahe, nach einem komplexen thematischen Muster Ausschau zu halten, daß die zwischen den Äußerungen 4 und 5 bestehende Relation vorwegnimmt.¹⁴ Tatsächlich läßt sich ein entsprechender Präzedenzfall im Diskurs finden - auf einer relativ abstrakten Ebene entspricht die pragmatische Relation zwischen den Äußerungen 2 und 3 der Beziehung zwischen den Äußerungen 4 und 5: Die jeweils übergeordnete Äußerung thematisiert die Arbeit von Peter Behrens auf einem bestimmten Gebiet der bildenden Kunst; die jeweils untergeordnete Äußerung betont den entscheidenden Einfluß, den Peter Behrens für die Weiterentwicklung des jeweiligen Gebietes hatte. Dieses komplexe thematische Muster läßt sich folgendermaßen beschreiben:

Behrens als bildender Künstler bewirkt die Entwicklung des jeweiligen Gebiets der bildenden Kunst

In diesem Fall wird das Themenmuster gleich doppelt moduliert: entlang der Isotopieebenen <*künstlerischer Beruf*> und - synchron - <*bildende Kunst*>.

Auf der pragmatischen Ebene ist Äußerung 5 Äußerung 4 subordiniert; was sich durch folgende naheliegende Anknüpfungsfrage veranschaulichen läßt:

Welche Auswirkungen hatte Behrens' Arbeit als Formgestalter für die Entwicklung des Spezialgebiets “industrial design”?

¹³Liegt die Konjunktion “*außerdem*” vor, so läßt sich oftmals folgende Beziehung beobachten: Diejenigen Ausdrücke, die in einer Äußerung vorausgehen, realisieren den Teil des Themenmusters, der invariant bleibt; nachfolgende Ausdrücke tragen die modulierten Komponenten des Themas bei. Daß es sich hierbei jedoch nicht um ein scharfes Kriterium handelt, wird anhand des vorliegenden Falles deutlich: “*Formgestalter*” ist als Modulation zu “*Architekt*” zu interpretieren, geht aber dennoch der Konjunktion voraus.

¹⁴Die semantische Funktion der Partikel “*auch*” läßt sich folgendermaßen umschreiben: Es wird signalisiert, daß eine Prädikation über ein anderes Objekt erneut - moduliert - zur Anwendung gelangt. Die jeweiligen Realisierungen der Prädikation durch sprachliche Ausdrücke können hierbei recht stark divergieren; entscheidend ist, daß die Partikel “*auch*” als Schlüsselwort für die Suche nach derartigen Themenübereinstimmungen fungiert.

Äußerung 6: *“die vom →Jugendstil inspirierten Formen des Kunsthandwerks erschwerten zu jener Zeit noch eine Verwirklichung der Gestaltungsprinzipien, die der neuen Lebensweise entsprachen.”*

Unter isolierter Betrachtung kann Äußerung 6 zunächst der Wiederaufgriff des <rückständig>-<progressiv>-Gegensatzes zugeschrieben werden, der bereits in Äußerung 2 (hinter dem Gedankenstrich) thematisiert wird. In Äußerung 2 wird dieser Kontrast syntaktisch-strukturell sowie lexikalisch expliziert: Die koordinierende Konjunktion *“aber”* drückt eine Adversativrelation zwischen den beteiligten Satznuklei aus, wobei zusätzlich gilt, daß sich die verknüpften Aussagen nicht semantisch gegenseitig ausschließen. Die Partikel *“noch”* und das Zeitadverb *“gleichzeitig”*, die die erste bzw. zweite Aussage relativieren, zeigen an, daß der Gegensatz auf einen bestimmten Zeitraum eingegrenzt ist; speziell die erste Aussage ist - vermöge *“noch”* - in ihrer Gültigkeitsdauer beschränkt. Der eigentliche Gegensatz besteht zwischen Reaktion und Progression und wird insbesondere lexikalisch vermittelt: *“archaisch”* vs. *“Fortschritt”*. In Äußerung 6 hingegen wird der Alt-Neu-Kontrast indirekt ausgedrückt: Zwar ist das Neue lexikalisch thematisiert (Adjektiv *“neu”*), jedoch die Instanziierung des Überkommenen scheint allenfalls mit Hilfe komplexer semantischer Inferenzen sowie unter Rekurs auf Hintergrundwissen als solche erkennbar.¹⁵

Weitet man den Skopus der Betrachtungen aus und interpretiert Äußerung 6 in Verbindung mit Äußerung 5, so wird ersichtlich, daß sogar die von Äußerung 2 insgesamt kommunizierte inhaltliche Entität moduliert wiederaufgegriffen wird. Das entsprechend abstrahierte Themenmuster kann folgendermaßen umschrieben werden:

Behrens' Arbeit als bildender Künstler unter dem Gesichtspunkt bestehender Gegensätze zwischen Überkommenem und Neuem

Erneut ergibt sich eine pragmatische Subordination unter die unmittelbar vorausgehende Äußerung, die sowohl durch einen oberflächlichen Indikator - Semikolon statt Punkt - als auch durch die Art der Kohäsion - die Zeitanapher *“zu jener Zeit”* knüpft an Äußerung 5 an - angezeigt wird:

Auf welche Widerstände traf das Wirken von Behrens als Formgestalter?

4.4 Determinanten von Textthemen

Die Fallstudie hat offengelegt, daß eine Vielzahl von Determinanten unterschiedlicher Beschreibungsebenen zur Etablierung von Textthemen bzw. potentiell themenstiftenden inhaltlichen Einheiten beitragen. Isotopie deckt nur einen kleinen Ausschnitt der Möglichkeiten zur Bildung textueller Themen ab. In den meisten Fällen sind es nicht die atomaren lexikalischen Merkmale, sondern komplexe Relationen über atomaren inhaltlichen Entitäten mit lexikalisch-semantischem oder referentiellen Beitrag, wobei zusätzlich Freiräume für bestimmte Variationen bestehen.

¹⁵Es stellt sicherlich kein Problem dar, sich einen textuellen Kontext vorzustellen, in dem der sprachliche Ausdruck *“die vom Jugendstil inspirierten Formen des Kunsthandwerks”* für den künstlerischen Fortschritt steht und nicht als Hemmnis moderner Gestaltungsprinzipien zu interpretieren ist.

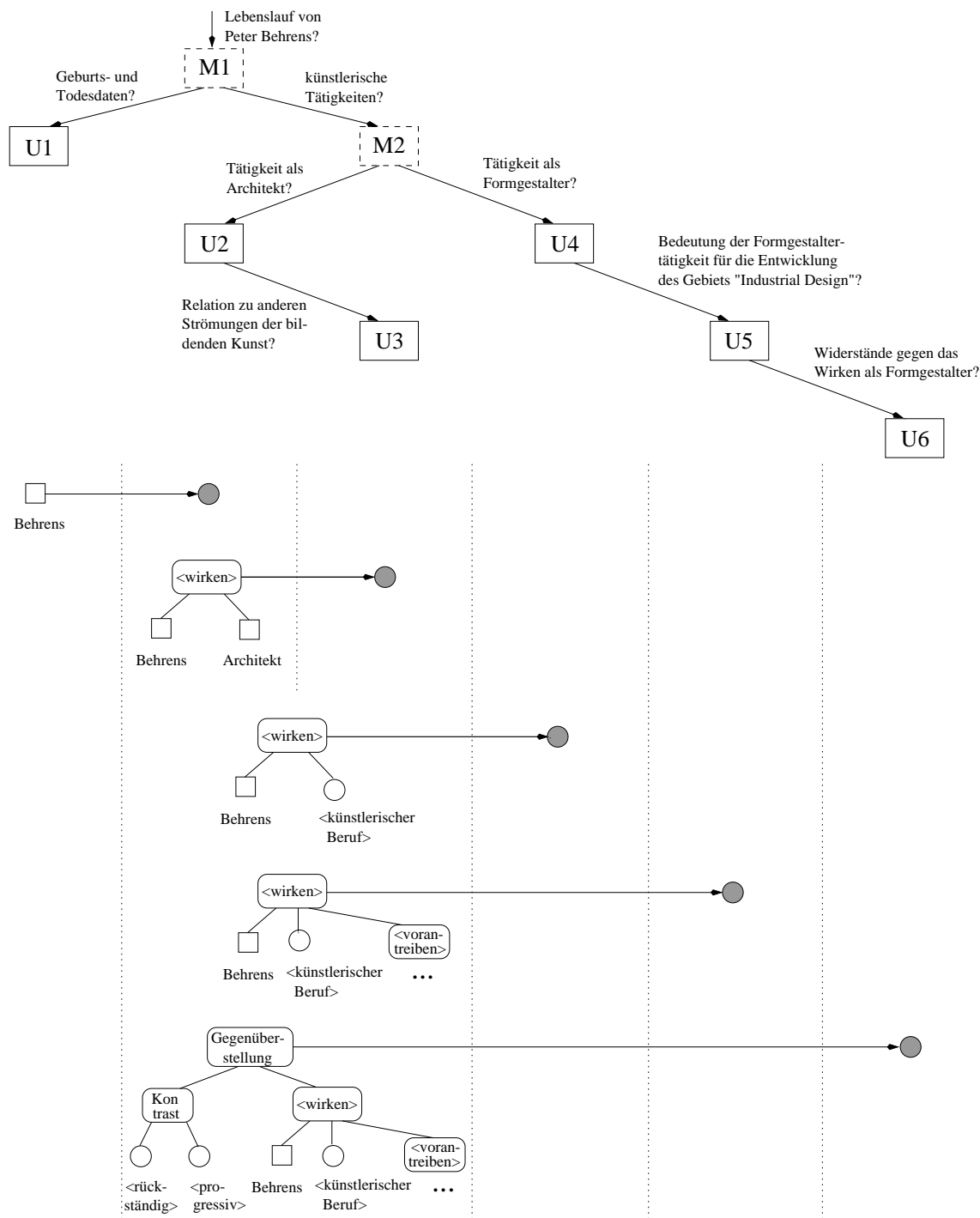


Abbildung 4.2: Pragmatische und thematische Textstruktur für die Fallstudie

Bislang wurden die thematischen Entitäten und ihre Bestimmungsfaktoren eher informell beschrieben. Die Darstellung soll nun präzisiert werden. Ein Blick auf die im Rahmen der Fallstudie identifizierten Belege für thematische Rekurrenz zeigt auf, daß die wiederaufgegriffenen kontextuellen Entitäten i.a. von *situationaler* Gestalt sind, d.h. sie beschreiben Relationen zwischen einer Anzahl von *Mitspielern* - den Partizipanten der Relation. Abbildung 4.2 faßt die wesent-

lichen Instanzen von Themenrekurrenz in der Fallstudie sowie ihre Zuordnung zu den Äußerungen U_1, \dots, U_6 des Texts zusammen und gibt darüberhinaus eine Skizze der pragmatischen Kohäsionsstruktur sowie der explizierenden Fragen. Die im unteren Teil der Grafik baumförmig dargestellten komplexen Themenmuster lassen sich auch textuell formulieren:¹⁶

U_2, U_3	<wirken>(Handelnder:Behrens,Beruf:Architekt,...)
U_2, U_4	<wirken>(Handelnder:Behrens,Beruf:<künstlerischer Beruf>,...)
U_{2-3}, U_5	<wirken>(Handelnder:Behrens,Beruf:<künstlerischer Beruf>, Ziel:<vorantreiben>(Handelnder:Behrens,Ziel:<bildende Kunst>,...))
U_{2-3}, U_6	Gegenüberstellung: <wirken>(Handelnder:Behrens,Beruf:<künstlerischer Beruf>,...) <Gegensatz>(<rückständig>, <progressiv>)

Es sind zwei Typen des Wiederaufgriffs atomarer Mitspieler zu unterscheiden: referentielle Rekurrenz und - dargestellt durch Angabe der entsprechenden Isotopieebene (in spitzen Klammern) - lexikalisch-semantiche Rekurrenz. Als zusätzliche Charakterisierung wird jedem Mitspieler - zunächst ohne theoretische Rechtfertigung (vgl. u.) - ein semantischer *Rollenbezeichner* (*Handelnder, Beruf, Ziel, ...*) zugeordnet.

Im Hinblick auf die Zielsetzung einer algorithmischen Verfeinerung der computergestützten Inhaltsanalyse sind diejenigen Fälle von besonderem Interesse, in denen die den Themen zugrundeliegenden relationalen Beziehungen auf der lexikalischen oder syntaktischen Ebene initiiert werden. Dieser Punkt soll nun ausführlicher diskutiert werden.

4.4.1 Valenz und Dependenzstruktur

Relationale Beziehungen sind oftmals in bestimmten Einzelwörtern verankert, denen für sich alleine genommen bereits das Potential zur Verknüpfung einer bestimmten Anzahl von Mitspielern zugeschrieben werden kann. Dies gilt insbesondere, jedoch nicht ausschließlich, für Verben. In der Linguistik wurde hierfür der Begriff der *Valenz* geprägt. Valenz läßt sich definieren als *linguistische Fügungspotenz*, die die Fähigkeit eine Lexems charakterisiert, andere sprachliche Konstituenten zu binden. Dabei werden in der Regel bestimmte grammatische Bedingungen an die unterschiedlichen Mitspieler, die hier auch als *Aktanten* bezeichnet werden, geknüpft - beispielsweise, welche Aktanten notwendig auftreten, in welchem grammatischen Kasus - Nominativ, Genitiv, Dativ, Akkusativ - diese stehen, welche semantische Rolle - Agens, Quelle, Ziel etc. - ihnen zuzuschreiben ist und welche semantischen Merkmale sie tragen. Folgende Beispiele explizieren einige Aspekte der Fügungspotenz des Verbs "entwickeln":

- (a) Peter Behrens entwickelte eine neuartige Konzeption.
- (b) * Peter Behrens entwickelte.
- (c) * Peter Behrens entwickelte einer neuartigen Konzeption.

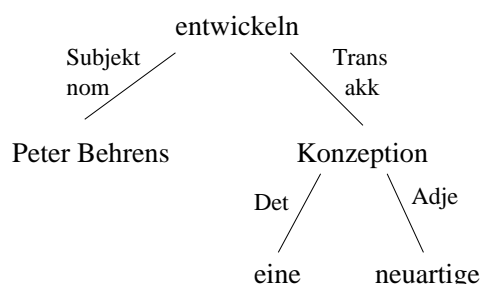
¹⁶Sowohl hier als auch in der Grafik sind nur die Wiederaufgriffe des jeweils spezifischsten, d.h. "maximalen" Themenmusters angegeben. Beispielsweise wird in Äußerung U_5 auch eine inhaltliche Entität reinstantiiert, die bereits in U_2, U_3 und U_4 in Erscheinung tritt, jedoch von dem letztlich zugeordneten Element graphentheoretisch subsumiert wird; auf eine Visualisierung der Rekurrenz derartiger Komponenten-Themen wird daher verzichtet.

Die Inakzeptabilität der Belege (b) und (c) ist ein Indikator dafür, daß offenbar ein Akkusativobjekt obligatorisch gebunden wird. Unter Explikation der semantischen Rollen ergibt sich folgende relationale Darstellung der Valenzbeziehung:

entwickeln(Agens:Peter Behrens,Ziel:Konzeption)

Die Semantische Rolle beschreibt die inhaltliche Funktion, die dem entsprechenden Mitspieler in der durch das Verb bezeichneten Relation zukommt. (In bestimmten Fällen werden auch mehrere semantische Rollen je Mitspieler - z.B. Agens und Quelle gleichzeitig - vergeben.) Wichtig ist an dieser Stelle primär die Ähnlichkeit dieser Repräsentation mit den zuvor als themenrelevant identifizierten relationalen inhaltlichen Entitäten.

Da die Valenz eine im Lexikon verankerte Eigenschaft darstellt, ist sie im Hinblick auf die Vergabe von semantischen Rollen der Ebene der *lexikalischen Semantik* zuzuordnen. Inwieweit sich die Valenz eines Lexems in der syntaktischen Struktur widerspiegelt, hängt von der Grammatik ab, die der syntaktischen Analyse zugrunde gelegt wird. In der von Lucien Tesnière entwickelten *Dependenzgrammatik* stellen die vermöge der Valenzen bestehenden Abhängigkeitsbedingungen die grundlegenden Determinanten der syntaktischen Struktur dar, die hier auch als *Dependenzstruktur* bezeichnet wird ([Tesn59]). So könnte etwa Satz (a) wie folgt modelliert sein:



(a) *Peter Behrens entwickelte eine neuartige Konzeption.*

Die Kanten des Dependenzbaums sind jeweils mit der syntaktischen Funktion sowie ggf. mit dem grammatischen Kasus markiert, in dem die entsprechenden sprachlichen Ausdrücke stehen. Modelliert die Grammatik die Vergabe der auch als *Tiefenkasus* bezeichneten semantischen Rollen, so spricht man von einer *Kasusgrammatik*¹⁷.

Da Themen inhaltliche Einheiten verkörpern, stellen relationale Repräsentationen, die nach Tiefenkasus geordnet sind, die bevorzugte Basis für die Determination von Themenwiederaufgriffen dar. Die grammatischen Kasus - auch *Oberflächenkasus* genannt - sagen alleine noch nichts über die semantischen Rollen aus, da die Beziehung zwischen Oberflächenkasus und Tiefenkasus von weiteren Bedingungen, insbesondere dem *Genus Verbi* - Aktiv oder Passiv - abhängig ist.¹⁸ Folgende Beispiele erläutern diese Gegebenheit:

¹⁷Diese Bezeichnung geht auf die Arbeiten von Fillmore zurück ([Fill68]).

¹⁸Aus computerlinguistischer Sicht besteht das Problem, daß eine Flexionsanalyse stets nur die grammatischen Kasus liefert. Zur Bestimmung der Tiefenkasus ist auf spezielle Theorien zurückzugreifen, die deren Ableitung aus den Flexionsmerkmalen und dem oberflächenstrukturellen Kontext erlauben. Die *Theorie der Semantischen Emphase* von Jürgen Kunze stellt einen geeigneten Ausgangspunkt dar, da in ihrem Rahmen Regeln über den Zusammenhang zwischen Tiefenkasus und Oberflächenkasus expliziert werden, die sogar dem Anspruch der Algorithmisierbarkeit gerecht werden ([Kunz91]). Für eine ausführlichere Diskussion dieser Problematik vgl. die Arbeit von Haenelt ([Haen95]).

- (d) Peter Behrens entwickelte eine neuartige Konzeption.
 (e) Die neuartige Konzeption wurde von Peter Behrens entwickelt.
 (f) Die Entwicklung der neuartigen Konzeption durch Peter Behrens ...

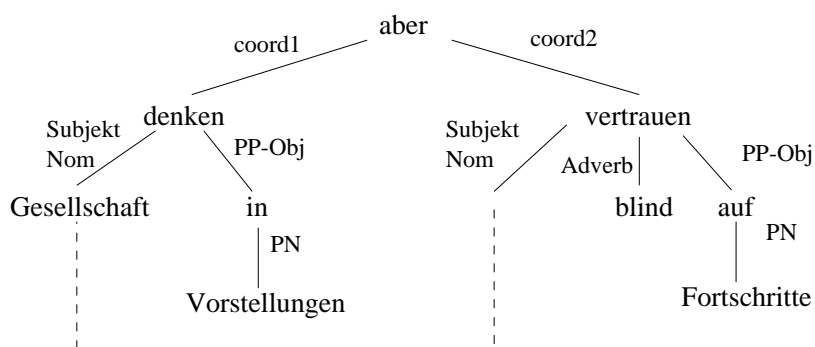
Beleg (f) zeigt darüberhinaus auf, daß nicht alleine Verben, sondern durchaus auch andere Wortarten - insbesondere Substantive - eine dependenzstrukturbildende Funktion zukommt. Für sog. *Deverbativa* wie “*Entwicklung*” ist dies unmittelbar einsichtig, da die Valenz des Ausgangsverbs zumindest in Teilen geerbt wird; die unterschiedlichen, nunmehr in Form von Nominalattributen wie Possessivpronomen, Genitivattribut und attributiven Präpositionalphrasen realisierten Mitspieler sind hier durchweg fakultativ.¹⁹ Ungeachtet dieser spezifischen Oberflächenrealisierung sollte auf dasselbe thematische Grundmuster, das sich entlang der Instanziierung der *semantischen* Rollen ergibt, erkannt werden. Wenn auch die Perspektive variiert, so ist die per Funktor und Mitspielern kommunizierte *Situation* in allen Fällen dieselbe.

In der obigen Fallstudie treten Deverbativa, die Themen wiederaufgreifen, gleich mehrfach in Erscheinung. Ein gutes Beispiel existiert in Äußerung 4: “*seine Tätigkeit als Formgestalter*” als sozusagen deverbativ kondensierte Variation von “*er ist als Formgestalter tätig*”.

4.4.2 Parataxe und Hypotaxe

Auch die Nebenordnung (Parataxe) oder Unterordnung (Hypotaxe) von Teilsätzen wird unmittelbar auf der Ebene der syntaktischen Struktur abgebildet. Geht man davon aus, daß die Teilsätze - u.a. vermöge der Valenz der jeweiligen Verben - Situationen kommunizieren, so können Hypotaxe und Parataxe als Ausdrucksmittel zur Bildung komplexerer *Situationsbündel* aufgefaßt werden.

In der Dependenzstruktur wird die Form des Bündels reflektiert. Beispielsweise ergibt sich für die durch die koordinierende Konjunktion “*aber*” vermittelte Adversativrelation in Äußerung 2 folgendes Bild einer parataktischen Satzstruktur:



- (2) [...] die Gesellschaft noch in archaischen Vorstellungen dachte, gleichzeitig aber blind auf die überwältigenden Fortschritte der Technik vertraute.

Die in dieser Darstellung aus Übersichtlichkeitsgründen nicht abgebildeten Partikel und Adverbien (“*noch*”, “*gleichzeitig*”) tragen zusätzliche Information bei, die in eine verfeinerte semantische

¹⁹Für detailliertere Ausführungen zur substantivischen Valenz vgl. Teuber ([Teub79]).

Interpretation der durch den Junktors “*aber*” vermittelten inhaltlichen Verknüpfung einfließt. In vielen Fällen kann die semantische Beziehung alleine auf der Basis von Konjunktion und ggf. Partikeln vollständig determiniert werden, weswegen diesen Wortarten oftmals (insbesondere in Arbeiten zur Diskursstrukturanalyse) die Rolle sog. *Schlüsselwörter* zugeschrieben wird. Aus ähnlichem Grund erweisen sie sich als relativ oberflächennahe Determinanten der Themenkonstitution von Bedeutung: Ein thematischer Aufgriff der obigen Adversativrelation könnte durch einen sprachlichen Ausdruck vollzogen werden, dessen dependenzgrammatische Modellierung mit der obigen Struktur zumindest weitgehend - bis auf Modulation - übereinstimmt; komplexe Inferenzen zur Verifikation, daß es sich tatsächlich um dieselbe inhaltliche Rahmenverknüpfung handelt, sind aufgrund der stereotypen Semantik i.a. nicht notwendig.

4.4.3 Hintergrundinformation

Im Text der Fallstudie geschieht der Wiederaufgriff der zuletzt diskutierten thematischen Entität auf indirekte Weise. Selbst unter abstrahierendem Rekurs auf semantische Rollen tritt die Übereinstimmung, die für Äußerung 6 identifiziert wurde, noch nicht zutage: Zwar kann Äußerung 6 als Wiederaufgriff der adversativen Relation zwischen Überkommenem und Neuem interpretiert werden; die Ähnlichkeit manifestiert sich jedoch nicht auf der dependenz-strukturellen Ebene, sondern erst auf der Ebene einer tiefensemantischen Repräsentation, denn erst unter Rückgriff auf Hintergrundinformation sollte sich erschließen, daß in etwa folgende abstrakte Entsprechung besteht:

<i>überkommen</i>	<i>“die vom Jugendstil inspirierten Formen des Kunsthandwerks”</i>
<i>neu</i>	<i>“Gestaltungsprinzipien, die der neuen Lebensweise entsprechen”</i>
<i>adversativ</i>	<i>“erschwerten zu jener Zeit noch eine Verwirklichung”</i>

Wie anhand dieses Beispiels deutlich wird, können thematische Rekurrenzen auch auf sehr komplexe Weise zustande kommen. Welche Entitäten als Themen wahrgenommen werden, ist letztlich immer auch eine Frage von kontextuellem Vorwissen und subjektiver Interpretation. Erneut zeigt sich die prinzipielle Inadäquatheit der Unterscheidung zwischen “manifeste” und “latente” Inhaltsebene.

4.4.4 Zusammenfassung

Im Rahmen der Fallstudie wurde ein Themenbegriff motiviert, der mit der Sichtweise der sozialwissenschaftlichen Themenanalyse in Einklang steht, indem er u.a. die mit der wörterbuchbasierten Operationalisierung des Kategorisierungsschritts implizit assoziierte Isotopieanalyse subsumiert. Themen sind definiert als inhaltliche Einheiten - textuelle (Teil-)Kontexte -, die im Text wiederholt instanziiert werden. Isotopie ist eine lexikalisierte Basisvariante. Ein verfeinertes Modell ergibt sich unter Einbeziehung von Situationen als komplexere Kontexteinheiten, die zumindest partiell durch die lexikalische Fügungspotenz (Valenz) bestimmter Wörter determiniert werden; insofern die Grammatik diese Relationen widerspiegelt, entstehen thematische Einheiten mit Entsprechung in der syntaktischen Struktur. Jedoch zeigen die Beispiele auch die prinzipiellen Grenzen der syntaktischen *Widerspiegelung* potentiell themenrelevanter Kontexteinheiten

auf, die ja per definitionem *semantische* Entitäten sind: Die thematischen Rekurrenzen werden nicht notwendig an der syntaktischen Oberfläche oder vermöge Tiefenkasus reflektiert, sondern sind in komplex gelagerten Fällen erst auf der Basis von tiefensemantischen Repräsentationen und Inferenzen ableitbar.

4.5 Grenzen der wörterbuchbasierten Themenanalyse

Die Betrachtungen in den vorangegangenen Abschnitten ermöglichen eine qualifizierte Beantwortung der eingangs postulierten Kernfrage nach Reichweite und prinzipiellen Grenzen einer wörterbuchbasierten Textkategorisierung.

Die auf Wortlisten basierende Textkategorisierung wurde als Themenanalyse identifiziert, der nur ein sehr eingeschränktes Konzept textueller Themenentwicklung zugrundeliegt: die in lexikalisch-semantischen Merkmalen verankerte *Isotopie*, definiert als “*Wiederkehr von Wörtern desselben Bedeutungs- bzw. Erfahrungsbereichs*”. Die Fallstudie hat jedoch aufgezeigt, daß Themen i.d.R. durch Fortführung wesentlich komplexerer Entitäten des textuellen Kontexts installiert werden. Aus textlinguistischer Perspektive stellt also Isotopie lediglich einen elementaren Spezialfall dar. Folglich unterliegt die klassische Computergestützte Inhaltsanalyse gravierenden Einschränkungen. Bereits die Partizipationsrelationen zwischen Mitspieler und Situation, die in dependenzstrukturellen Beschreibungen an der syntaktischen Oberfläche reflektiert werden, finden keine Berücksichtigung, obwohl sie als themendeterminierende Einheiten hochrelevant sind. Dies gilt erst recht für die noch komplexeren Themenmuster, die durch Bündelung von Situationen entstehen: Auch solche inhaltlichen Einheiten werden z.T. syntaktisch-strukturell reflektiert - etwa durch parataktische oder hypotaktische Strukturen.

In einer hypothetischen Anwendung der wörterbuchbasierten Inhaltsanalyse auf Künstlerbiographien würden sich bereits folgende Beispiele von Kategorisierungskriterien als problematisch erweisen:

- *Kategorie “Schaffens-Akte”*: Die zu klassifizierenden Auswahleinheiten beschreiben die Schaffens-Aktivitäten bestimmter Künstler und die daraus hervorgegangenen Werke,
- *Kategorie “Bildende Kunst im gesellschaftlichen Umfeld”*: Die zu klassifizierenden Auswahleinheiten stellen Innovationen in der bildenden Kunst und Widerstände des jeweiligen gesellschaftlichen Umfelds gegenüber,
- *Kategorie “Innovative Künstler”*: Die zu klassifizierenden Auswahleinheiten beschreiben das innovative Wirken von Künstlern unter der Perspektive der Überwindung von Widerständen im gesellschaftlichen Umfeld.

Da in den Kriterien Bezug auf Mitspieler *in bestimmten Situationen* genommen wird, ist eine Operationalisierung dieser Kategorien durch nicht weiter strukturierte Wortlisten offensichtlich inadäquat; mit Hilfe der wörterbuchbasierten Inhaltsanalyse könnte somit allenfalls eine heuristische Approximation erreicht werden.

Die Limitation der wortlistenbasierten Klassifikation spiegelt sich in der Entwicklung verfeinerter Beschreibungskonzepte für die Kategoriendefinition wider, wie sie u.a. vom TEXTPACK-Erweiterungspaket TEXTLING zur Verfügung gestellt werden (vgl. Abschnitt 3.3.2). Aus der

vorstehenden Untersuchung geht jedoch unmittelbar hervor, warum auch derartige Erweiterungen letztlich nur heuristische Ersatzverfahren mit i.a. ungenügender Reichweite verkörpern: Alleine auf der Basis von Kookkurrenztests lassen sich die themenrelevanten Partizipationsrelationen nicht zuverlässig erkennen. Eine Operationalisierung der zuvor beschriebenen Kategorie “*Schaffens-Akte*” vermöge

<ARCHITEKT> = Behrens ODER Utzon ODER Asplund ODER Gropius ...
 <SCHAFFEN> = entwickeln ODER erarbeiten ODER
 erschliessen ODER verwirklichen ...

001 <ARCHITEKT> VOR <SCHAFFEN>

deckt eben nur bestimmte Standardrealisierungen ab:²⁰

- (a) *Peter Behrens entwickelte eine architektonische Konzeption.*
- (b) *Behrens erschloß als Formgestalter ein neues Spezialgebiet.*

Wird die Kernaussage auf andere Weise ausgedrückt, so greift auch das verfeinerte Kriterium nicht:

- (c) *Als Formgestalter erschloß Behrens ein neues Spezialgebiet.*

Zudem kann es dazu kommen, daß Auswahleinheiten, die irrelevant sind, fälschlicherweise kategorisiert werden:

- (d) *Peter Behrens entwickelte sich zu einem Designer von Weltrang.*

In diesen problematischen Fällen setzt eine korrekte Klassifikationsentscheidung zumindest die Berücksichtigung des syntaktischen Kontexts voraus.

Natürlich kann die obige Kategoriedefinition weiter verfeinert werden. Jedoch wird mit solchen Reparaturversuchen letztendlich nur das Symptom kuriert: Eine adäquate Theorie, mit Hilfe derer ein Bezug zwischen sprachlichen Ausdrücken und kommunizierten Inhalten hergestellt wird, liegt einer derartigen Vorgehensweise nicht zugrunde. Zudem werden die Beschreibungen u.U. derart komplex, daß weitere Verfeinerungen nicht mehr praktikabel erscheinen. Als Beleg für diese Aussage sind die Erfahrungen aus einer Studie von Scheerer und Tarnai zu werten, die TEXTPACK zusammen mit dem Erweiterungspaket TEXTLING zur Auswertung von Verbalbeurteilungen in Grundschulzeugnissen einsetzten ([STa89]). Eine vergleichsweise breite Auswahl stereotyper Bewertungsformulierungen konnte mit einem relativ einfachen und übersichtlichen Kategoriensystem erfaßt werden; jedoch existierte eine kleine Zahl von Zeugnissen, die den Formulierungsvorgaben nicht folgten und deren Berücksichtigung im Wörterbuch zu einem nicht mehr vertretbaren Beschreibungsaufwand geführt hätte. Scheerer und Tarnai gelangen zu folgender Konklusion (ibid., S. 295, Anführungszeichen im Original):

²⁰Im Rahmen des Beispiels wird vereinfachend unterstellt, daß das Flexionsproblem gelöst ist.

“Es finden sich [...] noch immer Sätze, die nicht der richtigen Abstufung zugeordnet sind. [...] vor allem [...] beruht es auf der Vielfalt der deutschen Sprache, Einschränkungen, Wünsche und Entwicklungen im Kontext eines Satzes auszudrücken. Eine eindeutige Zuordnung wäre nur dann möglich, wenn die markanten Elemente der einzelnen fraglichen Sätze in das Wörterbuch aufgenommen würden.

Angesichts der Fülle des Materials wird dieser Versuch aufgegeben, da er wesentlich zeitaufwendiger ist als eine “manuelle” Vercodung.”

Die praktisch unbegrenzte Vielfalt sprachlicher Ausdrücke, mit der eine bestimmte inhaltliche Einheit kommuniziert werden kann, sprengt den Rahmen der Möglichkeiten einer einzelwortorientierten Analyse. Die Verfügbarkeit verfeinerter Beschreibungsmittel, wie sie von TEXTLING zur Verfügung gestellt werden, ändert hieran nichts, da die inhaltstheoretische Fundierung fehlt.²¹ Entsprechend dringlich erscheint die Exploration von *linguistisch fundierten* Erweiterungen der computergestützten Inhaltsanalyse, die eine inhaltliche Erschließung sprachlicher Ausdrücke per theorieadäquater, auf Strukturgeneralisierungen basierender syntaktischer und semantischer Analysealgorithmen bewerkstelligt.

Auch die Verfahren zur Computerunterstützung der Qualitativen Analyse könnten von einer verfeinerten Berücksichtigung inhaltlicher Einheiten profitieren. Ein Analyseverfahren, das Syntax und Semantik hypotaktischer und parataktischer Strukturen erschließt, wäre nicht nur im Rahmen einer verfeinerten computergestützten Themenanalyse hilfreich: Es würde ferner ermöglichen, die Erkennung semantischer Sequenzrelationen zur Unterstützung der inferentiellen Theoriekonstruktion - eine Aufgabe der computergestützten Qualitativen Analyse nach Huber - zumindest heuristisch zu algorithmisieren (vgl. o., Abschnitt 3.2.2).

Abschließend sollen die Ergebnisse der textlinguistischen Fallstudie zusammengefaßt werden:

1. Die wörterbuchbasierte Inhaltsanalyse funktioniert - wenn auch unter erheblichen Einschränkungen -, weil sie eine Isotopieebenenanalyse verkörpert, wobei Isotopie als lexikalisch-semantisch verankerter Spezialfall der textuellen Themenentwicklung definiert ist.
2. Die Limitationen der wörterbuchbasierten Inhaltsanalyse resultieren daraus, daß neben Isotopie noch eine Vielzahl anderer, i.a. wesentlich komplexerer Möglichkeiten zur Etablierung von Textthemen bestehen, die nicht berücksichtigt werden.
3. Verfeinerte Beschreibungsmittel, wie sie von Erweiterungspaketen wie TEXTLING zur Verfügung gestellt werden, werden den Zielsetzungen nicht gerecht, da sie keinen theoriebasierten Bezug zu den kommunizierten Inhalten vermitteln; auch hier basieren die Kategorisierungsentscheidungen letztlich auf oberflächlichen Kriterien ohne Inhaltsbezug.
4. Erfolgversprechend sind bereits Verfeinerungen, die auf (eventuell nur partiellen) Beschreibungen des syntaktischen Kontexts basieren; die Verfügbarkeit semantischer Rollen würde den Gebrauchswert derartiger Beschreibungen noch erhöhen.

Aus den Punkten 1 bis 3 folgt These 1 der Arbeit (vgl. S. 6).

In bezug auf die nunmehr offenliegende Problematik des Computereinsatzes für die sozialwissenschaftliche Inhaltsanalyse sollen im folgenden Kapitel algorithmische Lösungsansätze diskutiert

²¹Die Diskussion belegt auf drastische Weise, welche Kluft sich zwischen dem durch den Namen TEXTLING zum Ausdruck gebrachten Anspruch und dem tatsächlich Gebotenen auftut.

werden, die im Laufe der letzten Jahre in der computerlinguistischen Forschung entwickelt wurden. Besonderes Augenmerk soll hierbei der Zielsetzung einer verfeinerten, inhaltsorientierten Operationalisierung des Codierschritts der computergestützten Inhaltsanalyse zuteil werden. Speziell für letzteres Problem wird ein Vorschlag erarbeitet, die robusten Analysetechniken einer neuartigen Klasse computerlinguistischer Verfahren, die ursprünglich für ein anderes Anwendungsgebiet konzipiert wurden, als Ausgangspunkt zu wählen.

Kapitel 5

Verfeinerte algorithmische Inhaltsanalyse

Im Verlaufe der Ausführungen in den vorangegangenen Kapiteln wurden einige zentralen Probleme einer einzelwortorientierten Computergestützten Inhaltsanalyse identifiziert:

- *Flexion*: Je nach Wortart existieren für ein Lexem eine oder mehrere Wortformen, die im Rahmen des Kategorisierungsprozesses gleichzubehandeln sind.
- *Lexikalische Ambiguität*: Auf der Ebene einer einzelwortbezogenen Analyse können Mehrdeutigkeiten bestehen, die durch eine Analyse des syntaktischen oder semantischen Kontexts aufzulösen sind.
- *Pronomen*: Eine pronominale Erwähnung einer inhaltlichen Einheit sollte im Kategorisierungsvorgang nicht von einer vollständigen, "lexikalischen" Erwähnung dieser Einheit unterschieden werden.
- *Wörterbuchbasierte Operationalisierung der Kategorien*: Wortlisten verkörpern nur eine elementare, lexikalische Variante zur Charakterisierung textthematischer Entitäten; die Validität der Computergestützten Inhaltsanalyse sollte sich erheblich verbessern lassen, indem auch komplexere, relationale inhaltliche Einheiten als Kriterien der Kategorisierung einbezogen werden.

Dies sind ausnahmslos Fragestellungen, zu denen bereits extensive Vorarbeiten in Linguistik bzw. Computerlinguistik geleistet wurden. Es soll daher ein Blick auf die in diesen Disziplinen erarbeiteten Ergebnisse geworfen werden. Zu diskutieren ist, inwieweit neuartige Verfahren der Computerlinguistik zur Lösung der genannten Probleme beitragen könnten. Der Schwerpunkt der Betrachtungen wird auf die letztgenannte Problemstellung - der Frage nach einer inhaltlich fundierteren Basis für Kategorisierungsentscheidungen -, gelegt. Das Pronomenproblem bildet den Gegenstand der Ausführungen in Teil II der Arbeit.

5.1 Anforderungen der Computergestützten Inhaltsanalyse

Nicht jedes linguistische bzw. computerlinguistische Modell bzw. Verfahren ist jedoch in gleichem Maße dafür geeignet, den spezifischen Anforderungen der Computergestützten Inhaltsanalyse zu genügen. Da die Inhaltsanalysesoftware i.a. auf großen, wenig restringierten Textmengen arbeiten soll, ist die *Massendatentauglichkeit* ein zentrales Kriterium. Es läßt sich wie folgt aufschlüsseln:

- Das Verfahren darf ausschließlich auf *verfügbarem Wissen* aufbauen; ein Verfahren, dessen prozedurale Komponenten zwar algorithmisch umsetzbar sind, das jedoch auf unrealistischen Vorannahmen basiert, ist offensichtlich untauglich.
- Das Verfahren soll *robust*¹ gegenüber bestimmten Problemfällen sein, die bei der Verarbeitung der Eingabe - insbesondere im Falle großer Textmengen - auftreten können, u.a.
 - orthographische Fehler,
 - ungrammatische Ausdrücke,
 - nicht resolvierbare Ambiguitäten.
- Das Verfahren sollte von allgemeinem Nutzen für die Analyse eines hinsichtlich Genre und inhaltlicher Domäne *breiten Spektrums unterschiedlicher Texte* sein; es sollte es *nicht auf Spezialfälle beschränkt* sein, die sich in der praktischen Anwendung als marginal oder irrelevant erweisen.
- Die *Analysegeschwindigkeit* sollte den Anforderungen einer Verarbeitung großer Textmengen gerecht werden.
- Die *Analysegüte* hat bestimmten Anforderungen zu genügen; insbesondere darf nicht das Ergebnis der Inhaltsanalyse verzerrt werden.

Diesen Bedingungen kommt keineswegs die Rolle trivialer Selbstverständlichkeiten zu. Dies betrifft sowohl die im folgenden diskutierten computerlinguistischen Analyseverfahren als auch die Lösungsansätze für das Pronomenproblem, die Gegenstand von Teil II der Arbeit sind. Ein spezifischer Teilaspekt, der mit den ersten drei Punkten des Anforderungskatalogs verbunden ist, besteht in der Zielsetzung der eindeutigen operationalen Umsetzbarkeit (“Implementierbarkeit”) - informatisch gesprochen: Der jeweilige Ansatz muß ein *Algorithmus* sein. Es wird sich zeigen, daß bei weitem nicht alle computerlinguistischen Ansätze, die als “Algorithmen” etikettiert sind, die Bedingungen erfüllen, die sich aus der informatischen Lesart dieses Begriffs ergeben. Bauer und Goos geben folgende Definition ([BaGo82], S.47):

¹Der Begriff der Robustheit soll zunächst entsprechend folgendem intuitiven Verständnis gebraucht werden:

Eine algorithmische Umsetzung bzw. Implementierung eines Verfahrens heißt robust, wenn sie im Falle bestimmter Unregelmäßigkeiten in der Eingabe nicht ergebnislos abbricht, sondern die Verarbeitung fortführt und ein - bezogen auf den Grad des Defekts der Eingabe - relativ gutes Ergebnis liefert.

Eine verfeinerte, abgestufte Definition des Robustheitsbegriffs, die auf die spezifischen Anforderungen computerlinguistischer Problemstellungen zugeschnitten ist, wird im Rahmen der Diskussion von Lösungsansätzen für das Pronomenproblem gegeben.

Definition 5.1 (Algorithmus, nach Bauer und Goos) Ein Algorithmus ist eine präzise, d.h. in einer festgelegten Sprache abgefaßte, endliche Beschreibung eines allgemeinen Verfahrens unter Verwendung ausführbarer elementarer (Verarbeitungs-)Schritte.

Die Verfahrensbeschreibung kann durchaus semiformal oder gar natürlichsprachig sein, sofern nur sichergestellt ist, daß die eindeutige technische Umsetzbarkeit gewährleistet ist, die in der Definition mit den Eigenschaftswörtern *präzise* (i.S. von unzweideutig), *endlich* und *ausführbar* gefordert wird.² In bezug auf die Rahmenbedingungen einer Massentextanalyse geht dies insbesondere einher mit den oben identifizierten Bedingungen. Die Algorithmen-eigenschaft muß *relativ zu den zur Verfügung stehenden Ausgangsdaten* erfüllt sein - im vorliegenden Fall bedeutet dies, daß insbesondere auch solche Texte verarbeitet werden können, die im orthographischen oder grammatischen Sinne fehlerhaft sind.

5.2 Morphologische Analyse

Das Problem der morphologischen Analyse kann als weitestgehend gelöst angesehen werden. So beschreiben beispielsweise Karlsson und Karttunen den für die Computerlinguistik eher ungewöhnlichen Forschungsstand auf dem Gebiet der morphologischen Analyse folgendermaßen ([CMU+95], S. 111):

“In the last 10-15 years computational morphology has advanced further towards real-life applications than most other subfields of natural language processing. The quest for an efficient method for the analysis and generation of word-forms is no longer an academic research topic [...].”

Für viele Sprachen, darunter auch Deutsch, sind mittlerweile computerlinguistische Softwarelösungen mit extrem hohem Abdeckungsgrad und Analysegeschwindigkeiten verfügbar, die

²Letztlich kann es keinen formalen Beweis dafür geben, daß ein in einer nicht-oder semiformalen Sprache beschriebener vermeintlicher Algorithmus all diesen Ansprüchen genügt - selbst eine Implementierung erbrächte diesen Nachweis nicht, da dann immer noch der Beweis ausstünde, daß die Implementierung der Spezifikation genügt. Bei D’Avis findet sich eine treffende Beschreibung der auch hier letztlich gegebenen Verankerung einer formalen Sprache in einer als *semantisch selbstreferentiell* zu bezeichnenden Umgangssprache ([DAvi94], S. 97f, Hervorhebungen im Original):

“Die Abstraktionsleistung zum Formalismus gelingt [...] nur aufgrund der Präponderanz der semantischen Dimension der Umgangssprache, in der formale Zeichen überhaupt erst als Zeichen eingeführt werden können. Über diese semantische Erstbegründungsfunktion in der ersten Aneignung formaler Sprachen hinaus ist die Umgangssprache also auch letzte Metasprache, weil auch ein formales Zeichensystem ohne Rekurs auf natürliche Sprache weder entwickelbar noch vollständig explizierbar ist.”

Insofern ist also auch hier eine einhundertprozentige Intersubjektivität nicht gewährleistet. Es kann lediglich indirekt per Verweis auf die zeichentranszendierende, insbesondere handlungstheoretische Verankerung der Bedeutung des Algorithmusbegriffs argumentiert werden: Unter Informatikern existiert ein geschärftes Verständnis dafür, welche natürlichsprachigen Beschreibungen präzise formuliert und damit funktional eindeutig charakterisiert sind; die Disziplin Informatik gibt sich mit ausgezeichnetem Erfolg mit dieser Form der Explikation zufrieden.

den Anforderungen der Massendatenverarbeitung vollauf genügen.³ Auf nähere Ausführungen zu den theoretischen Grundlagen einer qualitativ hochwertigen morphologischen Analyse soll verzichtet werden; von primärer Bedeutung ist die Feststellung, daß somit die heuristischen Ad-Hoc-Lösungen für das Flexionsproblem, die vor allem in der älteren Literatur zur Computergestützten Inhaltsanalyse weit verbreitet sind, ihre Berechtigung verloren haben.

Daß das Flexionsproblem als gelöst angesehen werden kann, ist nicht nur für die jahrzehntealte wörterbuchbasierte Analyse von Bedeutung; als Komponententechnologie sind Systeme zur morphologischen Analyse natürlich auch für die im folgenden diskutierten inhaltsorientierten Ansätze zur Verfeinerung der Computergestützten Inhaltsanalyse relevant.

5.3 Morphologische und lexikalische Disambiguierung

Auf der Ebene einzelner Wörter lassen sich zwei Klassen von Ambiguität unterscheiden: Ambiguität der *Wortart* und *lexikalische Ambiguität* im engeren Sinne (Homonymie, Polysemie). Wie aus den Ausführungen in Abschnitt 4.1 hervorgeht, sind beide Fälle relevant für die Computergestützte Inhaltsanalyse.

Instanzen von Wortartenambiguität lassen sich in den meisten Fällen unter Berücksichtigung des syntaktischen Kontexts auflösen. Beispielsweise ist für

“Würde Ihnen der vorgeschlagene Termin passen?”

entscheidbar, daß *“Würde”* als Hilfsverb und nicht als Nomen gebraucht wird.

Die computerlinguistische Forschung hat algorithmische Verfahren zur sog. *morphologischen Disambiguierung* hervorgebracht, die die im Verwendungszusammenhang korrekte Wortart mit hoher Wahrscheinlichkeit ermitteln. Diese auch als *Part-of-Speech-Tagger* bezeichneten Systeme bewerkstelligen ihre Aufgabe, indem der syntaktische Kontext in sehr beschränktem Umfang heuristisch rekonstruiert wird. Es existieren bereits massendatenfähige Lösungen mit sehr hoher Ergebnisgüte. Einer der bekanntesten Ansätze, die *Constraint Grammar*, trifft für englische Texte in 99.7 % aller Fälle die richtige Entscheidung, wobei - je nach Text - für 2-6 % aller Mehrdeutigkeiten keine Entscheidung getroffen wird, da die Heuristiken nicht feinkörnig genug arbeiten (vgl. Karlsson, Voutilainen, Heikkilä und Antilla: [KVHA95]).⁴ Sowohl Ergebnisgüte als auch Analysegeschwindigkeit lassen das System für die Verarbeitung großer Dokumentmengen geeignet erscheinen

³Für den Sieger der 1994 von der Gesellschaft für linguistische Datenverarbeitung (GLDV) veranstalteten Evaluationskonferenz Morpholympics (Hausser: [Haus94]), das auf dem sog. *two-level model* von Koskenniemi ([Kosk83]) basierende System GERTWOL, werden folgende Zahlen genannt (Haapalainen und Majorin: [HaMa94]): ca. 85000 Wortformen, wobei die Abdeckung per Derivationsmorphologie und Kompositabildungsmechanismus noch erheblich erweitert wird; für orthographisch einwandfreie Texte beträgt der Abdeckungsgrad 99 %, für unbeschränkte Texte 98 % (gemessen über Korpora unterschiedlicher Anwendungsdomänen). ; auf einer Sun SPARCstation 2 wird eine Analysegeschwindigkeit von 200 Wortformen pro Sekunde erreicht. Das System wird von Lingsoft, Inc., Helsinki, kommerziell vertrieben.

⁴Diese Aufschlüsselung entspricht der in formalen Evaluationen von Informationserschließungssystemen üblichen Unterscheidung zwischen *Precision* und *Recall*. Zwischen diesen beiden Größen, die weiter unten noch näher definiert werden, besteht i.d.R. eine Austauschbeziehung; je besser die Precision, desto schlechter der Recall und umgekehrt.

Für die zweite Fallklasse wortbezogener Mehrdeutigkeit, der lexikalischen Ambiguität im engeren Sinne, liegen die Verhältnisse komplexer: Nur in einer kleinen Zahl von Homonymiefällen - etwa für Verben, deren Lesarten sich in der Valenz unterscheiden - kann sich die Mehrdeutigkeit auf der Ebene der syntaktischen Struktur widerspiegeln. Statistische Ansätze zur sog. *word sense disambiguation*, deren Grundidee darin besteht, eine Entscheidung auf Basis einer lexikalischen Kookkurrenzanalyse zu treffen, nehmen sich dieses Problems an. Die Entscheidungsgrundlage wird i.a. aus der automatischen Analyse großer Textkorpora gewonnen. Von ihrer Geschwindigkeit sind diese Verfahren für die Massendatenverarbeitung geeignet; jedoch reicht die erzielbare Ergebnisgüte nicht an die Leistungsfähigkeit etwa der Systeme zur morphologischen Disambiguierung heran.

5.4 Verfeinerung der algorithmischen Kategorisierung

Die Unzulänglichkeit der wörterbuchbasierten Inhaltsanalyse rührt jedoch vom restringierten Inhaltsbezug des auf Isotopie basierenden Kategorisierungskriteriums her. Was können also computerlinguistische Ansätze zur Lösung dieses Kernproblems beitragen?

In Abschnitt 4.4 wurde offengelegt, daß sich die Themenentwicklung oftmals per Wiederaufgriff von relationalen Bezügen - entsprechend Mitspielern in bestimmten Situationen - vollzieht. Zumindest in bestimmten Fällen werden diese inhaltlichen Einheiten auf der Ebene der syntaktischen Struktur reflektiert. Folglich sollte diskutiert werden, inwieweit syntaktische Beschreibungen durch computerlinguistische Verfahren zur Syntaxanalyse algorithmisch erstellt werden können, welches die Probleme und Limitationen dieser Ansätze sind und inwieweit sie den spezifischen Anforderungen der Computergestützten Inhaltsanalyse gerecht werden.

Anhand der im Rahmen der Fallstudie diskutierten Beispiele wurde jedoch ferner deutlich, daß eine unmittelbare Widerspiegelung inhaltlich-thematischer Elemente in der syntaktischen Struktur eher die Ausnahme denn die Regel ist. Aus der Perspektive der Anforderungen einer verbesserten Computergestützten Inhaltsanalyse wäre daher eine ausschließlich syntaktische Analyse der Kritik ausgesetzt, daß nur ein Bruchteil aller potentiell themenrelevanten inhaltlichen Einheiten Berücksichtigung fänden. Eine Einbeziehung semantischer Rollen entkräftet dieses Argument nicht, da auch kasusgrammatische Beschreibungen letztendlich nur einen recht eingeschränkten Ausschnitt thematischer Entitäten abdecken. Angesichts der Ergebnisse der Fallstudie liegt es auf der Hand, daß sich der inhaltliche Beitrag eines Textes zumindest in Teilen orthogonal, d.h. unabhängig oder nur mittelbar abhängig von diesen Beschreibungsebenen, entwickelt.

Wenn eine syntaktische und kasusgrammatische Analyse nur einen kleinen Ausschnitt des inhaltlichen Beitrags eines Texts erschließen kann, so stellt dies ihre Tauglichkeit für eine Verfeinerung der Computergestützten Inhaltsanalyse in Frage: Die Reproduzierbarkeit inhaltlicher Elemente sollte nicht ausschließlich auf der Gestalt der syntaktischen Oberflächenstruktur basieren. Diese Bedingung legt es nahe, die Erschließung des Textinhalts losgelöst von den spezifischen Ebenen der linguistisch-strukturellen Konstitution zu betrachten. Dies soll keineswegs heißen, daß die syntaktischen Beschreibungen der zu analysierenden sprachlichen Ausdrücke von den Betrachtungen ausgeschlossen werden. Vielmehr wird die Perspektive gewechselt. *Gesucht werden Inhalte. Ausgangspunkt für eine adäquate Analyse sollte daher die Struktur relevanter Inhalte sein*, und zwar a priori unabhängig davon, wie diese an der sprachlichen Oberfläche ausgedrückt werden.

Betreffend die Machbarkeit dieser ambitionierten Zielsetzung ist Skepsis angesagt. Wenn schon

- wie sich zeigen wird - eine rein syntaktische Analyse der Satzstruktur ein nichttriviales, ja komplexes Problem darstellt, das die Möglichkeiten einer massendatenfähigen Textanalyse übersteigt, wie kann dann dem auf den ersten Blick viel weiterreichenden Anspruch einer allgemeinen - nicht an der Syntax, sondern an der Zielstruktur der Information orientierten - Inhaltserschließung Genüge getan werden?

In der Computerlinguistik hat sich in den letzten Jahren ein Spezialgebiet namens *Information Extraction* etabliert, dessen Forschungsgegenstand die Untersuchung der praktischen Realisierbarkeit von Softwaresystemen ist, die dem zuvor genannten Anspruch gerecht werden. Zielsetzung ist die Entwicklung bedingt massendatenfähiger Softwaresysteme, die, orientiert an den Vorgaben konkreter *Anwendungsszenarien*, entwickelt werden. Für den internen Aufbau der Programme werden keine Vorgaben gemacht; wichtig ist alleine die Performanz, die diese Systeme im Rahmen einer *formalen Evaluation* erzielen, die in Form der bereits wiederholt durchgeführten sog. *Message Understanding Conference* ("MUC") institutionalisiert ist.

Im Unterschied zu vielen anderen Gebieten der Computerlinguistik befinden sich diese Arbeiten nicht mehr auf dem Stand der theoretischen Modellbildung. Mittlerweile werden diese Softwaresysteme vergleichsweise hohen Ansprüchen gerecht. Somit verkörpern die Ansätze des *Information Extraction* einen idealen Bezugspunkt für die Untersuchung der Möglichkeiten inhaltsorientierter Verfeinerungen der Computergestützten Inhaltsanalyse. Ihnen soll daher ein ausführlicher Abschnitt gewidmet werden.

5.4.1 Syntaktische Analyse

Zunächst jedoch sollen die zentralen Probleme der computergestützten Syntaxanalyse identifiziert werden - dies wirft u.a. auch ein Licht darauf, welches die Rahmenbedingungen der Ansätze des *Information Extraction* sind. Es wird sich zeigen, daß für Grammatiken, die die volle Komplexität natürlicher Sprachen abdecken, die Implementierung zufriedenstellender Parser⁵ alles andere als eine triviale Aufgabe ist.⁶

Jeglicher Parsingvorgang baut auf den Ergebnissen einer lexikalischen sowie morphologischen Analyse auf: Hierbei wird die als Buchstabenfolge vorliegende Eingabe in eine Sequenz von Wörtern (Stämme, lexikalischen Einheiten) zerlegt und deren Flexionsmerkmale (soweit anwendbar: morphologischer Kasus, Numerus, Genus, Person usw.) bestimmt. Wie in Abschnitt 5.3 deutlich wurde, kann es Fälle von Wortambiguität geben, die erst im Rahmen der syntaktischen Analyse oder aber durch Vorschaltung eines heuristischen Verfahrens zur morphologischen Disambiguierung eliminierbar sind; letzterer Ansatz ist zu bevorzugen, da auf diese Weise die Ambiguität während der eigentlichen (uneingeschränkten) Syntaxanalyse so gering wie möglich gehalten wird, was sich für die Laufzeiteffizienz als bedeutsam erweist. Es kann sogar vorkommen, daß die Mehrdeutigkeit der Wortform mit der syntaktisch-strukturellen Ambiguität einhergeht. Eines der klassischen Beispiele für die englische Sprache lautet

"Time flies like an arrow."

⁵Die deutschen Entsprechungen "*Syntaxanalysator*" bzw. "*Syntaxanalyse-system*" haben sich bislang nicht durchsetzen können; daher wird auch in der vorliegenden Arbeit stets die englische Bezeichnung "*Parser*" verwendet.

⁶Bestimmte Ausschnitte natürlichsprachiger Grammatiken lassen sich effizient implementieren, so z.B. die sehr oberflächliche Syntaxanalyse in Weizenbaums ELIZA-System ([Weiz65], zit. nach [Weiz76]).

Entsprechend der lexikalischen Lesarten für “*Time*” - Nomen, Verb oder Adjektiv - sind hier mindestens drei syntaktische Lesarten möglich.

Eine viel bedeutsamere Auswirkung auf die Performanz des Parsings hat jedoch eine Form von Ambiguität, die erst auf der Ebene der Rekombinierung syntaktischer Teilanalysen in Erscheinung tritt und als *strukturelle Mehrdeutigkeit* bezeichnet wird. Das Problem entsteht u.a. im Rahmen der Verarbeitung von Relativsätzen, Adverbialsätzen und Präpositionalphrasen, deren syntaktisch-strukturelle Unterordnung (“Anknüpfung”) oftmals erst auf Basis einer inhaltlichen Interpretation erfolgen kann. In folgendem Satz

Peter beobachtet den Mann im Park mit dem Teleskop.

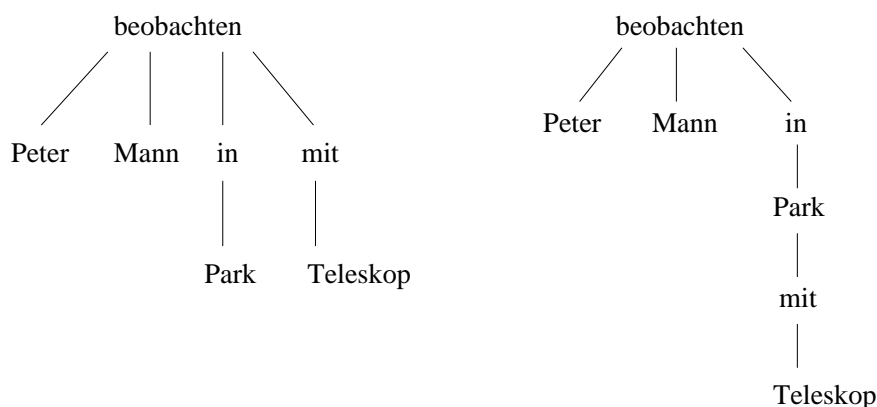
bestehen zwei Möglichkeiten, die Präpositionalphrase “*im Park*” zu interpretieren:

- *adverbial*: als Ort der Beobachtung,
- *attributiv*: als nähere Bestimmung des beobachteten Manns,

sowie drei Möglichkeiten, die Präpositionalphrase “*mit dem Teleskop*” zu interpretieren:

- *adverbial*: als Hilfsmittel der Beobachtung,
- *attributiv*: als nähere Bestimmung des beobachteten Manns,
- *attributiv*: als nähere Bestimmung des Parks.

Es ergeben sich fünf syntaktische Lesarten des Ausgangssatzes.⁷ Folgende Abbildung zeigt die Syntaxbäume (Dependenzstrukturen) zweier Interpretationsvarianten, die bei adverbialer Interpretation der Präpositionalphrase “*im Park*” möglich sind.



⁷Die Lesart, die der Kombination (“*adverbial*”, “*attributiv zu Mann*”) entspricht, ist aus strukturellen Gründen ausgeschlossen: Die entsprechende syntaktische Struktur ist - in diesem Fall unzulässigerweise - *nichtprojektiv*. Anhand dieses Falls wird deutlich, daß die Rahmenbedingungen auf der Ebene der syntaktischen Struktur bereits einen Teil der semantischen Interpretationsmöglichkeiten ausschließen. Das Thema der Interdependenz zwischen Syntax und Semantik wird sich erneut im Rahmen der Diskussion algorithmischer Lösungen des Pronomenproblems in Teil II der Arbeit stellen.

Das Beispiel zeigt, daß sich strukturelle Mehrdeutigkeiten, die darüberhinaus durch Anknüpfungsambiguitäten für Nebensätze sowie durch Mehrdeutigkeit der syntaktischen Funktion bedingt werden, multiplizieren. Dieses als *kombinatorische Explosion* bezeichnete Problem kann dazu führen, daß sich für komplexere Sätze einige tausend syntaktische Lesarten ergeben - ein Fall, der in der Praxis gar nicht selten auftritt und zu ernsthaften Effizienzproblemen des Parsingprozesses führen kann.⁸

Weitere Probleme, mit denen im Rahmen einer algorithmischen Syntaxanalyse umgegangen werden muß, ergeben sich durch koordinative Strukturen sowie Ellipsen. Auch hier kann es zur Ambiguität kommen. Aus Sicht der algorithmischen Syntaxanalyse besteht folgendes Dilemma: Wird die Entscheidung der Ambiguität offengelassen, so führt dies in vielen Fällen zur kombinatorischen Explosion und damit zu intolerabler Laufzeitperformanz; sollen hingegen die Ambiguitäten unmittelbar entschieden werden, so sind Interpretationsleistungen zu erbringen, die auf semantischen oder pragmatischen Kriterien basieren, deren Operationalisierungen robust gegenüber partiellen syntaktischen Beschreibungen sein müssen und i.a. nur als Heuristiken implementiert werden können, die nicht in allen Fällen die korrekte Entscheidung treffen.

Die mit den Zielsetzungen der Massendatentauglichkeit verbundene Robustheitsanforderung führt zu zusätzlichen Komplikationen. Die zugrundeliegenden Grammatiken schreiben vor, daß die zu analysierenden Sätze syntaktisch korrekt und vollständig sind - ein Blick in die Tageszeitung lehrt jedoch, daß diese Annahme selbst für seriöse Printmedien unrealistisch ist. Wird nun ein derartiger Satz syntaktisch analysiert, so liefert ein herkömmlicher Parser das Ergebnis "*Satz ungrammatisch*" - dies selbst dann, wenn die Abweichung von der zugrundeliegenden Grammatik marginal ist und die intendierte Interpretation für den menschlichen Leser auf der Hand liegt. Einen aus technischer Perspektive ähnlichen Fall stellt die sog. *Untergenerierung* dar: Die Grammatikregeln, auf der die Syntaxanalyse basiert, werden i.a. orientiert an einem Beispielkorpus entwickelt und verfeinert, das in sowohl lexikalischer als auch syntaktischer Hinsicht unvollständig ist. Eine Analyse anderer Texte wird oftmals Probleme bereiten, da i.d.R. Lexeme oder syntaktische Phänomene vorkommen, die durch die per Hand entwickelte Grammatik nicht abgedeckt werden. Briscoe nennt aufschlußreiche Zahlen ([CMU+95], S. 142):

"[...] in one project, a grammar for sentences from computer manuals containing words drawn from a restricted vocabulary of 3000 words which was developed over three years still failed to analyze 4% of unseen examples (Black, Lafferty, et al., 1992). This probably represents an upper bound using manual development of generative grammars; most more general grammars have far higher failure rates in this type of test."

Beide genannten Probleme - Unvollständigkeit von Daten bzw. Grammatik - sind im Rahmen einer robusten Parsingstrategie etwa dadurch zu lösen, daß die grammatischen Regeln nicht mehr als *strikte* Vorgaben einfließen. Diese Strategie wirft allerdings unmittelbar eine Folgefrage auf: Welche Formen eingeschränkter Regelverletzung sollten zugelassen werden? Zumindest für

⁸Hierbei ist zu bedenken, daß im Rahmen der syntaktischen Rekombinationsschritte eine Vielzahl unterschiedlicher Restriktionen zu verifizieren ist. Existieren mehrere Lesarten für einen bestimmten Ausschnitt der Eingabe, so sind auch die Rekombinationsschritte - und damit die Verifikation der Restriktionen - entsprechend häufig durchzuführen.

Die Entwicklung effizienter Verfahren zur vollständigen syntaktischen Analyse stellt ein offenes Forschungsproblem dar. Die Schwierigkeiten liegen in der Feinabstimmung von Parsingstrategie, Grammatik und Repräsentationssprache, in der die grammatischen Regeln und die linguistischen Beschreibungen codiert werden. Für eine ausführliche Diskussion dieses komplexen Themas sei auf die Dissertation von Böttcher verwiesen ([Bött96]).

unrestringierte Domänen liegen noch keine zufriedenstellenden Lösungen dieses Problems vor. Angesichts der Fülle von Schwierigkeiten, mit denen ein Parsing auf unrestringierter Anwendungsdomäne behaftet ist, liegt die Konklusion auf der Hand, daß eine umfassende syntaktische Analyse im Rahmen einer Verarbeitung großer Textkorpora *impraktikabel* ist.

Mit Blick auf diese Gegebenheit wurden im Laufe der letzten Jahre eine Reihe von Analysesysteme implementiert, die eine abgespeckte Zielsetzung - die sog. "seichte" *Syntaxanalyse* (engl. *shallow parsing*) - verfolgen und auf die Verarbeitung großer Textmengen zugeschnitten sind. Anstelle der Ableitung eines vollständigen Analysebaums verfolgen diese Ansätze das zwar weniger ambitionierte, doch dafür realistische Ziel einer partiellen syntaktischen Analyse, in deren Rahmen der Strukturbaum lediglich ausschnittsweise rekonstruiert wird.⁹ Somit wird die Massendatentauglichkeit mit einem Verlust an syntaktischer Analysetiefe erkauft.

Insgesamt darf hierbei jedoch nicht aus den Augen verloren werden, daß der Anspruch einer isolierten syntaktischen Analyse, wie er sich mit oberflächlichem Blick auf die entsprechende Unterscheidung der semiotischen Ebenen ableiten läßt, von vorneherein den theoretischen Anforderungen an ein adäquates Analysemodell nicht gerecht wird. *Zwischen Syntax und Semantik bestehen Interdependenzen, und diese Interdependenzen sind keineswegs unidirektional angelegt.* Da die syntaktische Struktur dazu beiträgt, Inhalte zu konstituieren, reflektiert sie notwendigerweise den semantischen Beitrag. In eine vollständige Analyse der syntaktischen Struktur fließen somit notwendig interpretative Elemente ein. Steht die entsprechende Information nicht zur Verfügung, so sind strukturelle Mehrdeutigkeiten die naheliegende Konsequenz. Folgerichtig identifizieren Karlsson und Karttunen die Verfeinerung des zugrundeliegenden Prozeßmodells als eines der zentralen Probleme zukünftiger Forschung auf dem Gebiet der massendatentauglichen Analyse ([CMU+95], S. 115):

"Much work remains to be done on interfacing morphological descriptions with lexicon, syntax, and semantics in a maximally informative way. This presupposes a global view of how the various processing components relate to one another."

Insofern hier keine Abhilfe geschaffen wird, verkörpert der Ansatz des seichten Parsings notwendig das Mittel der Wahl für die syntaktische Analyse großer Textmengen.

5.4.2 *Information Extraction*

Der Begriff "*Information Extraction*" findet gemeinhin Verwendung zur Bezeichnung eines Teilgebiets der computerlinguistischen Textanalyse, dessen Gegenstand die systematische, schablonenhafte Aufbereitung des Textinhalts nach bestimmten Zielkriterien ist. Typisches (generisches) Anwendungsszenario ist die automatische Anreicherung von Datenbanktabellen, deren Einträge aus Beschreibungen von Ereignissen sowie deren typischen Partizipanten - Handelnder, Begünstigter, Instrument der Handlung etc. - und sonstigen Umstandsbestimmungen - Ort, Zeit etc. - bestehen. Für Architektenbiographien könnten etwa die Schaffens-Aktivitäten der unterschiedlichen Architekten von Interesse sein; mögliche Analyseergebnisse ließen sich relational darstellen.¹⁰

⁹Für eine Übersicht der wichtigsten Systeme und die entsprechenden Literaturverweise vgl. Karlsson und Karttunen ([CMU+95], S. 113 f.).

¹⁰In der Terminologie der Künstlichen Intelligenz werden derartige Attribut-Wert-Strukturen auch als *Frames* ("Beschreibungsrahmen") bezeichnet.

<i>Aktivität</i>	<i>Schaffender</i>	<i>Ziel</i>	<i>Ort</i>	<i>Zeit</i>
<i>entwickeln</i>	<i>Behrens</i>	<i>Konzeption</i>		
<i>erschließen</i>	<i>Behrens</i>	<i>Spezialgebiet</i>		<i>um 1907</i>
<i>errichten</i>	<i>Utzon</i>	<i>Haus</i>	<i>Holte</i>	<i>1952/53</i>
<i>entwerfen</i>	<i>Utzon</i>	<i>Opernhaus Sydney</i>		<i>1956</i>
<i>bauen</i>	<i>Asplund</i>	<i>Waldkapelle</i>	<i>Südfriedhof Stockholm</i>	<i>1918-20</i>
...

Charakteristisches Merkmal von Softwaresystemen des *Information Extraction* ist die Anwendungsorientiertheit. Im Unterschied zu theoretisch ausgerichteten Textanalysemodellen steht nicht die Art der Ergebnisererschließung, sondern die formal gemessene Güte der ohne intellektuelle Intervention auf unrestringiertem Text gewonnenen Ergebnisse im Zentrum des Interesses. Es wird eine *ergebnisorientierte* Perspektive eingenommen; ob ein Analyseansatz kognitiv adäquat ist, spielt keine Rolle.

***Information Extraction* und Computergestützte Inhaltsanalyse**

Rahmenanforderung ist die robuste Verarbeitung größerer Textmengen, womit eine wesentliche Parallele zu den in Abschnitt 5.1 identifizierten Anforderungen der Computergestützten Inhaltsanalyse gegeben ist. Jedoch scheinen die unterschiedlichen Anwendungszielsetzungen auf den ersten, oberflächlichen Blick keine Gemeinsamkeiten aufzuweisen: In der Computergestützten Inhaltsanalyse klassischen Zuschnitts geht es nicht um die systematische, schablonenhafte, *qualitativ* ausgerichtete Aufbereitung des durch den Text kommunizierten Inhalts, sondern um die Klassifikation von Textpassagen, orientiert an den Vorgaben operationaler (algorithmisierter) Kategorisierungskriterien, gefolgt von einer i.a. *quantitativen* Auswertung.

Sieht man in den Algorithmen zur schematischen Informationserschließung hingegen einen Ausgangspunkt für die Operationalisierung feinkörnigerer Kategorisierungskriterien, so wird der Berührungspunkt offensichtlich. Insofern die Inhalte wissensorientiert, d.h. weitgehend unabhängig von der Realisierung an der sprachlichen Oberfläche, und mit akzeptabler Verlässlichkeit extrahiert werden können, wird der oben identifizierten Anforderung an verfeinerte Ansätze der algorithmischen Kategorisierung offenkundig entsprochen. Da die Softwarelösungen für Aufgaben des *Information Extraction* mit diesem Anspruch antreten, lohnt sich ein genauerer Blick auf Funktionsumfang, Leistungsfähigkeit und Aufbau einschlägiger Systeme. Auf der Basis dieser Betrachtungen wird es möglich sein, zu präziseren Aussagen über die Perspektiven einer verfeinerten Computergestützten Inhaltsanalyse zu gelangen. Der Kerngedanke besteht darin, die Zielstrukturen des *Information Extraction* als komplexe *thematische Schablonen* aufzufassen, die die Beschreibungsgrundlage verfeinerter Kategorienschemata bilden.

Message Understanding Conferences

Zunächst soll jedoch der Skopus der Betrachtung von Anwendungsszenarien und Softwaresystemen eingegrenzt werden. Die Entwicklung auf dem Gebiet des *Information Extraction* wurde seit Ende der achtziger Jahre in entscheidendem Maße durch die in relativ regelmäßigen Abständen veranstalteten *Message Understanding Conferences* (MUCs) vorangetrieben. Bei den *Message*

Understanding Conferences handelt es sich nicht um Tagungen herkömmlichen Zuschnitts, sondern vielmehr um *Evaluationskonferenzen*. Die mit Blick auf ein bestimmtes, im Vorfeld der Konferenz bekanntgegebenes Anwendungsszenario entwickelten Teilnehmerprogramme werden hier einem Blindtest auf neuen, unbekanntenen Texten der entsprechenden Domäne unterzogen. Als Ergebnis liegen untereinander vergleichbare Bewertungen der partizipierenden Systeme vor, die auf der Basis formal definierter Gütemaße errechnet werden. Obwohl sich nicht alle Arbeiten auf dem Gebiet des *Information Extraction* in MUC-Partizipationen niedergeschlagen haben, erscheint es dennoch gerechtfertigt, die nachfolgenden Betrachtungen auf die im Rahmen der *Message Understanding Conferences* dokumentierten Ergebnisse zu beschränken, da die MUCs mittlerweile allgemein als Motor der Weiterentwicklung auf diesem Gebiet angesehen werden. In den Worten der Organisatoren von MUC 6 (Grishman und Sundheim: [GrSu96a], S. 466):

“The MUCs are notable [...] in that they in large part have shaped the research program in information extraction and brought it to its current state.”

Die Aktivitäten im Rahmen der *Message Understanding Conferences* gehen auf Forschungsvorhaben militärischer Institutionen der USA zurück. Initiale Problemstellung war die Analyse militärischer, in Textform vorliegender Nachrichten. MUC-Organisator ist die Forschungsabteilung des *Naval Command, Control and Ocean Surveillance Center (NRAD)*; Geldgeber ist die *Defense Advanced Research Projects Agency (DARPA)*. Auch in den Anwendungsszenarien der früheren MUCs spiegelt sich das militärische bzw. nachrichtendienstliche Interesse der Auftraggeber: Die Domänen von MUC 1 und MUC 2 waren militärische Nachrichten; MUC 3 und MUC 4 hatten Nachrichten über Terroranschläge in Lateinamerika zum Gegenstand. Erst in den Konferenzen 5 und 6 fand ein Wechsel hin zu Problemstellungen auf Wirtschaftsmeldungen - die Extraktion von Fakten über Unternehmenskooperationen, personelle Veränderungen im Management etc. - statt.¹¹

MUC-Problemstellungen und die Portierbarkeitszielsetzung

Die im Rahmen der MUC-Konferenzen entwickelten Softwaresysteme haben somit einen *aufgabenspezifischen Zuschnitt*; die Adaptierung eines Programms an neue Anwendungsdomänen

¹¹Ein Blick auf die Teilnehmerschaft von MUC 6 zeigt auf, daß neben Instituten, die auf dem Gebiet der Rüstungstechnologie forschen bzw. direkt der Rüstungsindustrie zuzurechnen sind, private Softwareentwicklungsunternehmen und Forschungsgruppen an Universitäten mindestens ebenso zahlreich vertreten sind. Die Technologie des *Information Extraction* ist eben auch außerhalb der militärischen Anwendung - etwa als Such- und Indexierungsansatz in der Internet-Technologie - von großer Relevanz.

Wie in so vielen Bereichen der Informationstechnologie verkörpert die Begehrlichkeit von Militär und Nachrichtendiensten jedoch auch für das Gebiet des *Information Extraction* - und damit für die Entwicklung robuster Analysetechniken der Computerlinguistik - die initiale Triebfeder. Für den Forscher, der über eine diesem Thema gegenüber angemessene Sensibilität verfügt, bleibt der schale Beigeschmack, daß jeglicher Beitrag zur Verbesserung der Analysetechniken des *Information Extraction* ebenfalls einen Beitrag zur Perfektionierung staatlicher Überwachungsmaßnahmen verkörpert. Die zum Zeitpunkt des Entstehens dieser Zeilen entbrannte heftige Diskussion um den von einer breiten, große Teile des politischen Spektrums überspannenden parlamentarischen Mehrheit getragenen “Großen Lauschangriff” und die damit verbundene Zustimmung zu einer Grundgesetzänderung zeigt überaus deutlich, von welchem Ausmaß die Begehrlichkeit des Staates ist; steht die Technologie zur automatischen “Filterung” großer Textmengen einmal zur Verfügung, so ist davon auszugehen, daß die einschlägigen Überwachungsbehörden zu deren ersten Anwendern gehören werden.

kann durchaus zeitaufwendig oder gar mit konzeptuellen Schwierigkeiten verbunden sein. Obwohl sich diese Gegebenheit bei flüchtiger Betrachtung nicht mit den Anforderungen einer verfeinerten Computergestützten Inhaltsanalyse zu decken scheint, sind die von den MUC-Teilnehmern entwickelten Analysetechniken von entscheidender Relevanz: einerseits, weil sie einem durch formale Evaluation abgesicherten Szenario zur Entwicklung wirklich robuster Analyseverfahren für unrestringierten Text entspringen, das in Anspruch und Zielsetzung in der Computerlinguistik bislang einzigartig ist; andererseits, weil die Förderung der Entwicklung domänenunabhängiger Komponentensysteme zu einem Kernziel der aktuellen MUCs avanciert ist. Gerade der letztere Punkt ist von entscheidender Bedeutung für die Entwicklung praxistauglicher Anwendungen: Die Konfiguration eines Systems an neue Anwendungsszenarien sollte in möglichst breitem Rahmen durch den Anwender erfolgen können und nicht mehr in den Zuständigkeitsbereich des Entwicklers fallen.

Diese Zielvorgabe schlug sich in der Ausgestaltung der erweiterten Aufgabenmenge der *Message Understanding Conference 6* nieder, deren vier "Tasks" folgendermaßen definiert und motiviert sind (Grishman und Sundheim, [GrSu96b]):

- *Scenario Template Task*: die Basisaufgabe der Faktenextraktion - personelle Veränderungen im Management von Firmen -, modular gegliedert nach beteiligten Personen und Organisationen.
- *Template Element Task*: Zur Förderung einer domänenunabhängigen, anwendungsportablen Architektur wurde die Extraktion zweier als anwendungsübergreifend "elementar" angesehenen Objekttypen - Personen und Organisationen - als eigenständige Problemstellung aus der Basisaufgabe herausgelöst.
- *Named Entity Task*: Auch die Aufgabe der Identifikation von Namen wurde isoliert, um die Entwicklung weitestgehend domänenunabhängiger Komponententechnologie zu fördern; für dieses nicht zu tief gehende Problem schien das Ziel realistisch zu sein.
- *Coreference Task*: Um die Entwicklung auch tiefgründigerer Analysetechniken zu fördern, wurde das Teilproblem der referentiellen Interpretation von Nominalphrasen auf die Agenda gesetzt, wobei im Hinblick auf die Komplexität einer konsistenten Definition zunächst nur Koreferenzbeziehungen in die Aufgabe einbezogen wurden;¹² da die Annahme gerechtfertigt erscheint, daß Koreferenzresolution ein grundlegendes Teilproblem des *Information Extraction* darstellt, sollte auch diese Aufgabe die Entwicklung von Komponententechnologie fördern.¹³

Neben der Aufgabenmenge war auch die Gestaltung des Evaluationsszenarios darauf ausgerichtet, das Design domänenunabhängiger, portabler Submodule zu fördern. Ausgehend von den Erfahrungen aus einem Probedurchlauf, der die Gelegenheit zu einem ersten Test der Systemkomponenten gab, stand den Partizipanten nur ein Monat zur Verfügung, um die Adaptierung der Programme an die endgültige, evaluationsrelevante Version der "*Scenario Template*"-Aufgabe zu bewerkstelligen.

¹²Diese Förderung einer bestimmten Tiefenanalysetechnik geschah mit Blick auf die von den bisher entwickelten Systemen verfolgten "seichten", d.h. oberflächlichen Analysetechniken; entgegen der Grundausrichtung der MUCs - keine Vorgaben betreffend die Analysetechnik - wurde hierdurch erstmals ein Anreiz geschaffen, Systemkomponenten für als besonders relevant angesehenen Teilprobleme der Tiefenanalyse zu schaffen.

¹³Hierbei ist allerdings zu bedenken, daß - insbesondere für die Interpretation nichtpronominaler Ausdrücke - domänenabhängiges Wissen einfließen kann. Somit wird nicht alleine das Portabilitätsziel unterstützt.

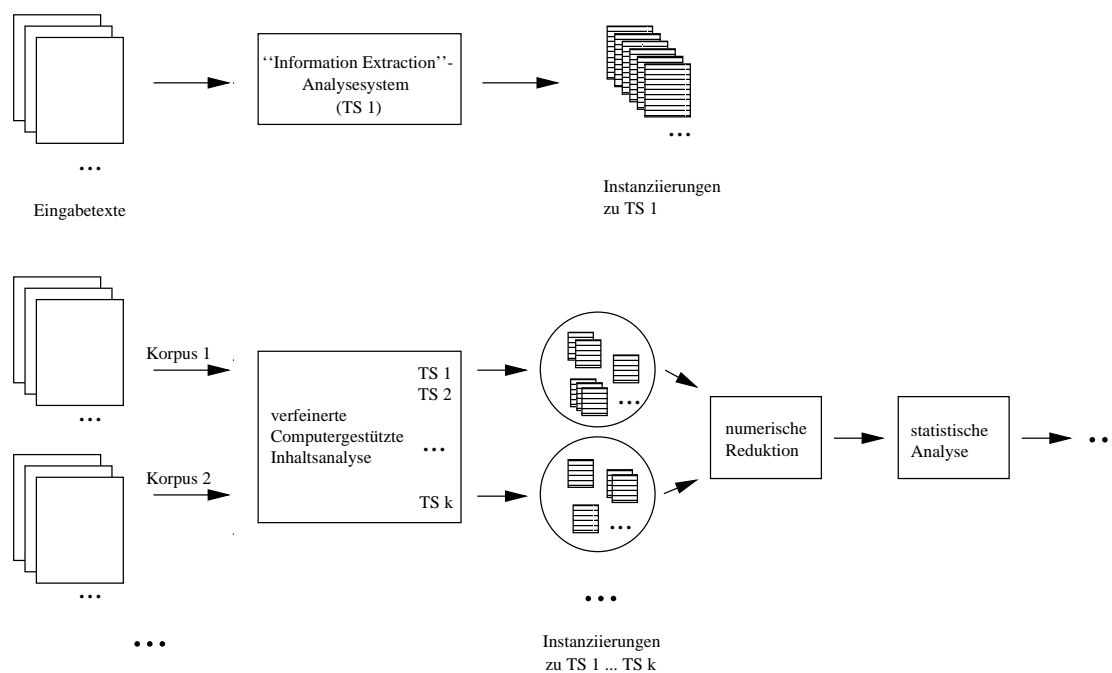


Abbildung 5.1: *Information Extraction* versus verfeinerte Computergestützte Inhaltsanalyse

“*Information Extraction*”-Technologie für die Computergestützte Inhaltsanalyse

Insofern die Ergebnisqualität den Anforderungen genügt und eine Adaptierung an unterschiedliche Anwendungsdomänen mit vertretbarem Aufwand möglich ist, sind Softwarelösungen für die “*Scenario Template*”-Aufgabe sowie die in diesem Zusammenhang entwickelten generischen Analysetechniken von potentieller Relevanz für die Verfeinerung der Algorithmisierung des Kategorisierungsvorgangs. Abbildung 5.1 stellt schematisch dar, in welcher Weise die “*Information Extraction*”-Technologie für die Computergestützte Inhaltsanalyse nutzbar gemacht werden könnte. Der obere Teil der Abbildung veranschaulicht die Arbeitsweise von Systemen des *Information Extraction*. Die Analyse ist *qualitativ* ausgerichtet; das Ergebnis besteht aus einer Menge symbolischer Faktenbeschreibungen, wobei das Analyseziel im Basisfall nur einen einzigen Szenario-Typus, entsprechend *einer* thematischen Schablone TS 1, umfaßt. Anders verhält es sich im Falle der verfeinerten Computergestützten Inhaltsanalyse, die im unteren Teil der Abbildung skizziert ist.¹⁴ Die Operationalisierungen der verschiedenen Kategorien bauen auf einer Menge von Szenario-Typen, d.h. auf i.d.R. *mehreren* thematischen Schablonen auf, wobei unterschiedliche Möglichkeiten der Zuordnung bestehen:

- Einerseits ist denkbar, daß jede thematische Schablone nur zur Definition einer bestimmten Kategorie herangezogen wird;
- alternativ könnte das Klassifikationskriterium an Variationen in spezifischen Attributinstanzen¹⁵ festmachen, wobei eine thematische Schablone die Operationalisierungsgrundlage mehrerer Kategorien sein kann.

¹⁴Abbildung 5.1 zeigt die quantitative Spielart einer verfeinerten CGI. Prinzipiell ist auch eine qualitativ ausgerichtete Weiterverarbeitung möglich.

¹⁵vgl. die Fallstudie in Abschnitt 4.3, Stichpunkt Modulation

In ersterem Fall unterscheiden sich die Kategorien in der jeweils relevanten Relation; im letzteren Fall ist die konkrete Instanziierung eines Mitspielers eines spezifischen Relationentypus die Klassifikationsgrundlage. Beiden Ansätzen ist gemeinsam, daß die Entscheidung unter Berücksichtigung relationaler Beziehungen erfolgt und damit *auf feinkörnigerer Basis* geschieht als in der wortlistenbasierten Inhaltsanalyse.

Es sollen nun die wesentlichen Punkte diskutiert werden, die sich bezüglich der Nutzbarmachung der “*Information Extraction*”-Technik für die Computergestützte Inhaltsanalyse ergeben. Ausgangspunkt ist die zentrale Frage

Inwieweit lassen sich die Analyseverfahren und Techniken, die im Rahmen der MUCs entwickelt wurden, für eine Verfeinerung der Computergestützten Inhaltsanalyse bewerten?

Diese Problemstellung läßt sich in mehrere Punkte untergliedern:

1. Sind die im Rahmen von MUC bearbeiteten *Aufgabenstellungen* (insbesondere des “*Scenario Template*”-Problems) hinreichend komplex, so daß entsprechende Lösungen auch den Anforderungen einer Verfeinerung der Computergestützten Inhaltsanalyse gerecht werden könnten?
2. Genügt die *Ergebnisgüte*, mit der diese Aufgaben algorithmisch gelöst werden können, den Anforderungen der Computergestützten Inhaltsanalyse?
3. Ist die *Portabilität* der realisierten Softwaresysteme bereits hoch genug, um den Ansprüchen unterschiedlicher Kategorisierungsaufgaben gerecht zu werden?¹⁶

Eine “*Scenario Template*”-Aufgabe von MUC-6

Zunächst soll veranschaulicht werden, welche vergleichsweise komplexen Probleme mit den im Rahmen der MUCs entwickelten Systemen bearbeitet werden können. Dies soll anhand einer konkreten “*Scenario Template*”-Aufgabenstellung geschehen. Abbildung 5.2 zeigt eine Zielbeschreibung, die im Rahmen des Testdurchlaufs von MUC 6 zu erzeugen war.¹⁷ Das zugrundeliegende “*Labour Negotiation*”-Szenario spezifiziert Verhandlungsereignisse zwischen Arbeitgebern und Gewerkschaften als relevante Inhalte.

Im oberen Teil der Abbildung ist eine Textpassage gezeigt, die im Hinblick auf diese Aufgabe relevant ist; die partizipierenden Systeme sollten eine Ereignisbeschreibung anlegen, die der Vorgabe entspricht, die im unteren Teil der Abbildung wiedergegeben ist.

Das Beispiel zeigt auf, daß die Zielbeschreibungen mit Hilfe von Verweisstrukturen (Zeigern)

¹⁶Das diesem Punkt zugrundeliegende Verständnis von Portabilität ist strikt zu unterscheiden von dem einschlägigen Portabilitätsbegriff der Informatik, der die Adaptierbarkeit an unterschiedliche Software- und Hardwareplattformen umfaßt. Als rein implementationstechnische Fragestellung ist letztere Portabilitätsform in den folgenden Ausführungen irrelevant.

¹⁷Das Beispiel ist einer offiziellen Beschreibung der Aufgaben von MUC 6 entnommen (im Internet verfügbar via `ftp cs.nyu.edu`, Unterverzeichnis `pub/nlp/muc6`). Die englischen Erläuterungen zu den Attributinstanziierungen entstammen dieser Quelle.

```
<s> No progress was reported in the strike against Nynex Corp. where
both sides are locked in a bitter dispute over health-care benefits. </s>
<s> The CWA represents about 40,000 of that company's workers, and IBEW
represents an additional 20,000 workers. </s>
```

```
<TEMPLATE-8908230100-1> :=
  DOC_NR: "8908230100"
  CONTENT: <NEGOTIATION-8908230100-5>
<NEGOTIATION-8908230100-5> :=
  PARTY: <PARTY-8908230100-8> [follow pointers down to Nynex object]
        <PARTY-8908230100-4> [follow pointers down to CWA object]
        <PARTY-8908230100-7> [follow pointers down to IBEW object]
  ISSUE: BENEFITS
  PROPOSAL_STATUS: PROPOSAL_REJECTED
  TALK_STATUS_1: IMPASSE
  TALK_STATUS_2: IMPASSE
  COMMENT: "I assume both unions are striking and that 'No progress was
           reported in the strike' means they're at an impasse"
           / "This NEGOTIATION object could be split into 2, one for each
           union, but the text makes it sound like there's just one
           dispute going on"
<PARTY-8908230100-8> :=
  NEGOTIATOR_ORG: <ORGANIZATION-8908230100-7>
  NEGOTIATOR_SIDE: EMPLOYER
<PARTY-8908230100-4> :=
  NEGOTIATOR_ORG: <ORGANIZATION-8908230100-2>
  NEGOTIATOR_SIDE: UNION
<PARTY-8908230100-7> :=
  NEGOTIATOR_ORG: <ORGANIZATION-8908230100-6>
  NEGOTIATOR_SIDE: UNION
<ORGANIZATION-8908230100-2> :=
  ORG_NAME: "Communications Workers of America"
  ORG_ALIAS: "CWA"
  ORG_DESCRIPTOR: "the union" [this fill comes from text that
                              appears earlier in the article]
  ORG_TYPE: OTHER
<ORGANIZATION-8908230100-6> :=
  ORG_NAME: "International Brotherhood of Electrical Workers"
           [this fill comes from text that appears earlier in the article]
  ORG_ALIAS: "IBEW"
  ORG_DESCRIPTOR: "the union" [this fill comes from text that
                              appears earlier in the article]
  ORG_TYPE: OTHER
<ORGANIZATION-8908230100-7> :=
  ORG_NAME: "Nynex Corp."
  ORG_DESCRIPTOR: "that company" [this fill comes from text that
                              appears earlier in the article]
  ORG_TYPE: COMPANY
```

Abbildung 5.2: Beispiel für "Scenario Template"-Aufgabe, Testdurchgang MUC-6

hierarchisch aufgebaut sind. Einer im Hinblick auf die “*Labour Negotiation*”-Aufgabe relevanten Textpassage entspricht eine <TEMPLATE-Beschreibung, die einen Identifikator des entsprechenden Dokuments sowie einen *Verweis* auf eine Beschreibung des Verhandlungsereignisses (<NEGOTIATION-...) enthält. Das Verhandlungsereignis wiederum wird beschrieben durch die Angabe der beteiligten Parteien, des Verhandlungsgegenstands sowie des Verhandlungsstatus. Die Verhandlungsparteien sind abermals per Verweis angegeben (<PARTY-...), über den die Information zur Verhandlungsseite - Arbeitgeber oder Gewerkschaft - sowie letztlich eine Charakterisierung der jeweiligen Organisation zur Verfügung gestellt wird. Letztere Beschreibungen werden durch Objekte vom Typ ORGANIZATION repräsentiert, die - Merkmal der Förderung eines höheren Portabilitätsgrads - eine eigenständige Zielgröße der “*Template Element*”-Aufgabe verkörpern.

Obwohl die Anzahl der Slots (zu füllende Argumentstellen der Zielrepräsentation) im Vergleich zu früheren MUCs reduziert ist¹⁸ - das wesentliche Ziel von MUC-6 bestand in der Förderung der Entwicklung portierbarer Systeme -, muß ein Analyseprogramm noch immer hohen Anforderungen gerecht werden, um die Zielbeschreibung auch nur partiell aus dem Quelltext zu rekonstruieren. Der Zusammenhang zwischen den streikinitiiierenden Gewerkschaften, dem bestreikten Unternehmen und dem Status der Verhandlungen wird nur sehr indirekt und auf Diskursebene, d.h. über Satzgrenzen hinweg, etabliert. Zur Ableitung der relationalen Ergebnisstruktur fallen somit Verarbeitungsschritte an, die die Ebene einer ohnehin bereits hinreichend problematischen syntaktischen Analyse transzendieren. Dieses schwierige Problem ist bereits aus der obigen Fallstudie zur textuellen Themenkonstitution bekannt.

Evaluationsergebnisse von MUC-6

Insofern diese Aufgabe des *Information Extraction* in einem zu definierenden Sinne “hinreichend gut” bewerkstelligt werden kann, ergibt sich die gewünschte Ausgangsbasis für eine Verfeinerung der Computergestützten Inhaltsanalyse. Die beiden grundlegenden Evaluationsmetriken zur Bewertung der Ergebnisgüte von “*Information Extraction*”-Systemen sind unter den Bezeichnungen *Precision* und *Recall* bekannt, die sich allgemein wie folgt definieren lassen:

Definition 5.2 (Precision, Recall) Sei E die Anzahl aller Einzelergebnisse, die das zu bewertende System erzeugt hat. Seien ferner K die Anzahl korrekter Einzelergebnisse des Systems und A die Anzahl korrekter gesuchter Einzelergebnisse, d.h. die Zielvorgabe. Die Größen P - *Precision* - und R - *Recall* - sind folgendermaßen definiert:¹⁹

$$P := \frac{K}{E}$$

$$R := \frac{K}{A}$$

¹⁸Das “*Joint Venture*”-Problem, das in MUC-5 zu bearbeiten war, ist über Faktenschablonen definiert, die 47 Slots umfassen und 11 Objekttypen unterscheiden.

¹⁹Auch hier wiederum wird auf eine Übersetzung speziell des Begriffs “*Recall*” verzichtet, da er ohne deutsches Pendant ist und allgemeine Verbreitung gefunden hat.

Bezugspunkt ist eine konkrete Teilaufgabe als Zielvorgabe für das zu evaluierende System, durch die u.a. determiniert wird, von welcher Gestalt die betrachteten "Einzelergebnisse" sind; die Aussagefähigkeit der gewonnenen Werte hängt von der Repräsentativität des der Evaluation zugrundeliegenden Testkorpus ab. Die Größe *Precision* beschreibt, inwieweit ein System dazu in der Lage ist, die gesuchte Information möglichst "präzise", d.h. flankiert mit möglichst wenig inkorrekt Information, zu extrahieren. Die Größe *Recall* charakterisiert die Fähigkeit eines Systems, die gesuchte Information möglichst *vollständig* wiederzufinden. Precision und Recall stehen in einem Austauschverhältnis: Je schärfer die verwendeten Kriterien, desto besser wird die Precision, was jedoch i.a. mit einer Verschlechterung des Recall erkauft wird.

In den Evaluationen der Systeme für die "Scenario Template"-Aufgabe von MUC stellen die Instanziierungen der einzelnen Schablonen-Slots die Bezugseinheiten der Bewertung dar.²⁰ Die Precision- und Recallwerte werden auf unterschiedlichen Granularitätsebenen ermittelt:

- separat für die einzelnen Argumentstellen - also etwa für die PARTY- oder ISSUE-Slots der Schablone aus Abbildung 5.2,
- kumuliert über die Argumentstellen bestimmter Komponenten der Schablone - z.B. alle Slots der Komponenten des Objekttyps ORGANIZATION,
- kumuliert über alle Argumentstellen der Schablone.

Die Gesamtbewertung der Systeme - in MUC-Terminologie die "All Objects"-Evaluation - basiert auf der zuletzt genannten Kumulation. Eine größere Anzahl von MUC-6-Teilnehmern erzielte Precision-Recall-Kombinationen mit Precision-Werten zwischen 0.6 und 0.7 und Recall-Werten zwischen 0.4 und 0.5 (vgl. [MUC696], S. 24 ff. und S. 295 ff.). Hierbei ist zu beachten, daß es "das" beste System nicht geben muß, da zwischen Precision und Recall das zuvor beschriebene Austauschverhältnis besteht. Das hieraus entstehende Vergleichsproblem wurde partiell durch die Einführung eines kombinierten Maßes, des sog. "F-Measure" begegnet, in das die Precision- und Recall-Werte mit wählbarer Gewichtung einfließen (ibid., S. 294). In der Evaluation wurden die F-Werte je System für drei unterschiedliche Gewichtungen ermittelt: 1:1, 1:2 und 2:1. Dies ergibt einen ersten Anhaltspunkt für die Güte sowie die spezifische Stärke der Systeme im Vergleich. Für die 1:1-Gewichtung lag das beste Ergebnis eines Teilnehmersystems bei 56.4.

Eine definitive Entscheidung, welches der Systeme für eine bestimmte Gewichtungskombination das beste ist, ist jedoch auch auf dieser Grundlage nicht möglich, weil die anhand des spezifischen Testkorpus gewonnenen Werte nicht notwendig einen *statistisch signifikanten* Qualitätsunterschied ausdrücken. Um diesem Problem zu begegnen, wurde eine computergestützte Analyse der statistischen Signifikanz der Ergebnisse durchgeführt (Chinchor, ibid., S. 39 ff.). Für das vermeintlich beste System mit F-Wert 56.40 etwa stellte sich heraus, daß 6 der 10 Konkurrenzsysteme, darunter eines mit F-Wert 48.14, als gleichwertig einzustufen sind. Als relativ vages Fazit der Evaluation für die 1:1-Gewichtung wurde festgehalten, daß 8 der 11 Teilnehmersysteme einen gleichermaßen hohen Leistungsstandard aufweisen, der sich in F-Werten um bzw. oberhalb von 50 niederschlägt. Für die Mehrheit der Partizipanten liegt der Precision-Wert über dem des

²⁰Für die MUC-Evaluation sind Precision und Recall etwas subtiler definiert als in den obigen elementaren Definitionen. U.a. ist berücksichtigt, daß das korrekte Ergebnis nicht immer eindeutig definierbar ist; für die partizipierenden Systeme bestehen mitunter Freiheitsgrade, die im Rahmen der Bewertung berücksichtigt werden. Formale Definitionen der MUC-Evaluationsmetriken finden sich in den Konferenzproceedings zur MUC-6 ([MUC696], S. 293 ff. und S. 36). Neben Precision und Recall werden dort noch vier weitere Metriken definiert, die es u.a. erlauben, die Fehler eines Systems näher aufzuschlüsseln.

Recalls.

Auf der Ebene der slotspezifischen Einzelbewertungen wird zwar das relative Verhältnis der Precision- und Recallwerte - P größer als R bzw. umgekehrt - i.a. reproduziert, jedoch gibt es große Unterschiede zwischen den absoluten Werten. Das System mit F-Wert 56.40 erzielt für die unterschiedlichen Slots (P, R)-Wertekombinationen von (0.97, 0.79) bis hin zu (0.35, 0.14). Die Divergenzen lassen sich z.T. dadurch erklären, daß einige der Argumentstellen inhärent schwierig zu füllen sind; in einigen Fällen sind sie jedoch Ausdruck unterschiedlicher Schwerpunktsetzungen im Systemdesign. Weist ein Teilnehmer einen im Vergleich zu Konkurrenzsystemen auffällig schlechten (P, R)-Wert auf, so kann dies ein Anhaltspunkt für mögliche Verbesserungen sein.

Neben der kumulierten und nichtkumulierten Berechnung von Precision und Recall für die Argumentstellen wurde im Rahmen von MUC-6 auch das sog. "Text Filtering"-Maß berechnet, das die Fähigkeit eines Systems charakterisiert, für Texte korrekt zu *entscheiden*, ob ein vorgegebener Szenario-Typus instanziiert wird. Diese Aufgabe ist i.d.R. leichter als die Bestimmung der relevanten Mitspieler - die besten Teilnehmer erzielten (P, R)-Kombinationen mit Eckwerten von (0.8+, 0.98) bis hin zu (0.96, 0.8+).

Portabilitätsaspekte

Wie zuvor beschrieben, stellte die Förderung der Entwicklung domänenunabhängiger Teilkomponenten die Motivation für die Ausgestaltung der Aufgaben sowie die Vorgabe einer engen zeitlichen Begrenzung für die Anpassung der Systeme an die Evaluationsdomäne im MUC-6 dar. Ultimatives Ziel sind *generische Systeme*, die sich vom Anwender über geeignete Schnittstellen mit wirtschaftlich vertretbarem Aufwand an unterschiedliche Anwendungsszenarien anpassen lassen. Da sich dies unmittelbar mit den Anforderungen einer verfeinerten Computergestützten Inhaltsanalyse deckt, soll nun mit Blick auf Systemarchitektur und Erfahrungsberichte einiger MUC-6-Partizipanten erarbeitet werden, inwieweit diese Zielsetzung als bereits erreicht angesehen werden kann. Neben der Diskussion von Problematik und Möglichkeiten der technischen Realisierung flexibel adaptierbarer Systeme steht auch die Frage im Vordergrund, wie groß der erwartete zeitliche Aufwand für eine Rekonfiguration ausfällt.

Allgemein als unter Portabilitäts Gesichtspunkten sinnvoll eingeschätzt wird eine möglichst weitgehende *Separation* von Prozeßkomponenten und Wissen in *domänenspezifischen* und *domänenunabhängigen Teil*. Insofern eine Systemarchitektur diese Trennung adäquat - z.B. über konzeptuell klare Schnittstellen - reflektiert, unterstützt sie die gezielte Reprogrammierung der domänenvariablen Anteile. Die zentrale Instanz des domänenunabhängigen und damit anwendungsinvarianten Anteils stellt die linguistische Verarbeitungskomponente dar, die ja auf grammatischem Wissen basiert, das für die Gesamtheit der zu analysierenden Dokumente identisch ist. Die generelle Gültigkeit ist somit das grundlegende Argument für den Einsatz linguistisch fundierter Analysetechniken; jedoch offenbart ein Blick auf die Komplexität eines generellen Parsings die technischen Probleme eines solchen Unterfangens.

Diese Gegebenheiten finden ihren Niederschlag in den Teilnehmersystemen von MUC-6, denen durchweg die Bestrebung zugrundeliegt, linguistische Generalisierungen möglichst weitgehend in die Analyse einzubeziehen. Da jedoch i.d.R. keine als wirklich vollständig zu bezeichnende Grammatikimplementierung verfügbar sein wird, besteht das Problem, daß mit einer neuen Anwendung oftmals syntaktische Konstrukte in Erscheinung treten, die noch nicht abgedeckt sind.²¹ Somit entsteht das Dilemma, daß es einerseits aus Portabilitätsgründen erstrebenswert

²¹Vgl. hierzu Abschnitt 5.4.1, Briscoe-Zitat.

ist, syntaktische Generalisierungen in möglichst großem Umfang in die Analyse einzubeziehen, daß es jedoch andererseits undenkbar erscheint, die in diesem Rahmen entstehende Obligation der Grammatikpflege auf den Systemanwender zu übertragen - die Implementierungsdetails unrestringierter Grammatiken sind i.a. zu komplex. Aus dieser Erwägung heraus gelangte etwa die Teilnehmergruppe von der New York University zu der Entscheidung, in ihrem "NYU"-System - im Unterschied zu den Versionen für die früheren MUCs - keine vollständige syntaktische Analyse mehr zu implementieren und anstelle dessen ein beschreibungstechnisch einfacheres, auf regulären Sprachen aufbauendes Verarbeitungsmodell zu verwenden;²² Ralph Grishman begründet diesen Schritt folgendermaßen ([MUC696], S. 168):

"Our Holy Grail, like that of many groups, is to eventually get the computational linguist out of the loop in adapting an information extraction system for a new scenario. This will be difficult, however, if the scenario requires the addition of some grammatical construct, albeit minor."

Das Ziel, eine Domänenadaption durch den Anwender zu ermöglichen, wäre ansonsten in weite Ferne gerückt.

Ein anderer Weg zur Lösung dieses Problems wird mit dem PIE-System der Entwicklergruppe der University of Manitoba, Winnipeg, beschritten ([MUC696], S. 113 ff.). Das Ergebnis wird auf der Basis einer möglichst vollständigen Abhängigkeitsstrukturanalyse konstruiert; die Erweiterungsobligation entfällt, da das System robust arbeitet, indem es für unbekannte Syntaxkonstrukte eine *fragmentarische Ergebnisstruktur* liefert. Der Vorteil dieses Ansatzes ist darin zu sehen, daß - zumindest im Falle eines nichtfragmentarischen Ergebnisses - die linguistischen Generalisierungen durch die der Analyse zugrundeliegende Abhängigkeitsgrammatik vollständig und konzeptuell klar getrennt von den domänenspezifischen Verarbeitungsschritten zur Verfügung gestellt werden. Kritik an diesem Verarbeitungsmodell kann jedoch zumindest von zwei Seiten geäußert werden. Zum einen findet sich keine Aussage darüber, in wieviel Prozent aller Fälle der fragmentarische Fall auftritt und inwieweit die grammatischen Generalisierungen dann überhaupt zur Verfügung stehen. Zum anderen scheint die Konstruktion der Ergebnisse der "Scenario Template"-Aufgabe, die größtenteils auf unmittelbarer Grundlage von Abhängigkeitsbäumen geschieht, nur im Falle von Zielbeschreibungen möglich zu sein, deren Argumente i.a. eine überschaubare Anzahl stereotyper abhängigkeitsstruktureller Relationen ausfüllen. Letzteres Problem betrifft zwar nicht mehr die Separation linguistischen Wissens, ist aber dennoch für die Diskussion der Portabilitätsfrage relevant, da es die Anwendbarkeit des Ansatzes für komplexere Szenarien fraglich erscheinen läßt.²³

Die meisten Teilnehmersysteme von MUC-6 - darunter das oben angesprochene "NYU"-System - setzen "Pattern Matching"-Techniken ein, mit Hilfe derer die syntaktische Struktur nur eingeschränkt rekonstruiert wird.²⁴ Die Implementierung grammatischer und domänenspezifischer Restriktionen basiert hierbei auf regulären Ausdrücken, für die zwar nicht der unter Anwendungsgesichtspunkten problematische Wartungsaufwand anfällt, die jedoch andererseits ohne

²²Dieses Beispiel belegt anschaulich, wie die spezifischen Vorgaben für MUC-6 die gezielte Weiterentwicklung der Systemarchitekturen bedingen.

²³Die Arbeit auf der Grundlage einer unrestringierten abhängigkeitsstrukturellen Syntaxanalyse stellt eine Neuerung der Manitoba-Gruppe dar, die für die früheren MUCs mit komplexeren "Scenario Template"-Aufgaben ebenfalls linguistisch seichtere Analyseverfahren einsetzte.

²⁴Neben dem zuvor diskutierten Ansatz der University of Manitoba existieren noch zwei weitere Systeme - die Beiträge der University of Durham sowie der University of Sheffield -, die entgegen diesem Trend unrestringiert parsen.

Weiteres nicht das Abstraktionsniveau erreichen, das auf der Basis vollständiger Grammatiken erzielbar wäre. Die Autoren des “NYU”-Systems identifizieren diesbezüglich zwei grundlegende Desiderata ([MUC696], S. 174):

1. den Erhalt der *Beschreibungsökonomie*, der sich auf lexikalischer Ebene dadurch niederschlägt, daß unterschiedlichen Lexemen mit übereinstimmender syntaktischer und evtl. semantischer Argumentstruktur eine strukturell identische Spezifikation zugeordnet werden kann;
2. die Erkennung der *logischen Prädikat-Argument-Struktur* unabhängig von der Oberflächenrealisierung - Aktiv- und Passivvarianten, äquivalente deverbativ konstituierte Relationen etc.

Diese in Umgebungen mit vollständiger syntaktischer Analyse implizit unterstützten Abstraktionsmechanismen sollten erhalten werden, weil durch sie der Aufwand des potentiell fehlerträchtigen Adaptierungsvorgangs minimiert wird. Insofern deren Re-Integration in einen “*Pattern Matching*”-Ansatz gelingt, wird der Portierbarkeitszielsetzung unmittelbar Rechnung getragen:

- Der Zeitaufwand für die Domänenadaption wird reduziert;
- die Fehleranfälligkeit wird minimiert;
- die automatische Erzeugung der (anwendungsspezifischen) Analysevarianten für die unterschiedlichen syntaktischen Oberflächenrealisierungen schafft die Voraussetzungen zur Domänenadaption durch den Anwender, der auf diese Weise von der tieferegehende linguistische Kenntnisse voraussetzenden Obligation befreit wird, alle Spielarten zu explizieren.

Der “NYU”-Ansatz ist in dieser Hinsicht vorbildlich, da er den obengenannten Desiderata mittels zweier spezieller Konzepte - eine sog. “*Pattern-by-Example*”-Technik sowie sog. *Metaregeln* - gerecht wird. Grishman umschreibt diese Vorgehensweise wie folgt ([MUC696], S. 175):

“The most recent explorations also indicate how syntax can ‘creep back’ into a system from which it was unceremoniously ejected. [...] Noun group syntax remains explicit, as one phase of pattern matching. Clause syntax is now utilized in the metarules for defining patterns and in the rules which analyze example sentences to produce patterns.”

Als vordringliches Problem zukünftiger Forschung wird die Entwicklung interaktiver Schnittstellen genannt, die eine für die anwenderseitige Domänenkonfiguration geeignete logische Sicht vermitteln. Hierbei seien allerdings schwierige Probleme betreffend die genaue Ausgestaltung der bisher allenfalls als prototypisch einzustufenden Abstraktionskonzepte zu lösen.

Ein weiterer “*Pattern Matching*”-Ansatz, in dem elaborierte Techniken zur Separation genereller und domänenspezifischer Komponenten zum Einsatz kommen, ist das FASTUS-System der SRI International, USA. Der größte Teil der domänenspezifischen Verarbeitung ist in der

sog. *“Domain”-Phase*, dem letzten Verarbeitungsschritt einer Sequenz hintereinandergeschalteter Transducer (regulärer Automaten mit Ausgabe), konzentriert. Innerhalb des *“Domain”-Schritts* selbst kommt auch linguistisches, d.h. generelles Wissen zum Einsatz, das jedoch ähnlich wie im *“NYU”-System* derart organisiert ist, daß es bei Domänenwechsel nicht recodiert zu werden braucht. Sog. *macro rules*, in etwa entsprechend den Metaregeln von *“NYU”*, codieren das domänenunabhängige grammatische Wissen; die Konfiguration an spezifische Domänen wird bewerkstelligt, indem die Makroregeln unter Angabe anwendungsabhängiger Parameter instanziiert werden. In Erfüllung des zweiten oben genannten Desideratums wird der Analysecode für syntaktische Ausdrucksvarianten automatisch generiert.

Als wesentliches Hindernis auf dem Weg hin zu einem anwenderkonfigurierbaren System identifizieren die FASTUS-Entwickler die schlecht lesbare, unintuitive Syntax der Transducer-Spezifikationssprache FASTSPEC, in der u.a. die Makroregeln sowie deren Instanzen codiert sind. Sie gelangen jedoch zu der Einschätzung, daß sich dieses Problem per Realisierung einer geeigneten Anwenderschnittstelle lösen lassen könnte: ([MUC696], S. 241 f.):

“[...] the information specified by the domain-dependent rules is relatively straightforward to provide, (although currently obscured by a rather opaque syntax) so that with the help of a suitable user interface, it is easy to imagine an analyst supplying the system with the information needed to customize it to a new extraction task. Developing such a tool is one of our next priorities.”

Eine weitere Schwierigkeit besteht darin, daß nicht alle domänenspezifischen Spezifikationen im Rahmen der *“Domain”-Phase* von FASTUS codiert sind; jedoch finden sich keine Angaben, auf deren Grundlage sich der zu erwartende Recodierungsaufwand sowie der Grad der Anwenderzugänglichkeit abschätzen ließe.

Zuletzt sollen die *Aussagen zum Zeitbedarf einer Domänen-Rekonfiguration* aus den drei zitierten Systembeschreibungen gegenübergestellt werden. Die obere Grenze ist durch die Einschränkung des Zeitraums für die Konfiguration an die MUC-6-Evaluationsdomäne auf einen Monat vorgegeben. Wenig aufschlußreich sind die Angaben für das PIE-System, da aus dessen Beschreibung nicht genau hervorgeht, inwieweit die Trennung zwischen domänenspezifischem und generellem Code vollzogen ist. Für die (nicht optimierte) Konfiguration der *“Scenario Template”-Komponente* zu MUC-6 wird ein Aufwand von fünf Personentagen genannt, jedoch scheint dies nicht die insgesamt notwendige Arbeit abzudecken, da anwendungsspezifische Modifikationen und Erweiterungen in weiteren Modulen - etwa der Syntax - notwendig waren.

Auch in der Beschreibung des *“NYU”-Systems* findet sich lediglich die undifferenzierte Aussage, daß für die Domänenanpassung weniger als drei Wochen zur Verfügung standen.

Am aussagekräftigsten sind die Angaben zum FASTUS-System, dessen Konzept zur getrennten Codierung genereller und domänenspezifischer Aspekte den vergleichsweise besten Eindruck hinterläßt: Zur Konfiguration der *“Domain”-Phase* wurde *weniger als ein Personentag* aufgewandt. Zwar bleibt die Frage offen, wie aufwendig die Recodierung der nicht in der *“Domain”-Phase* konzentrierten anwendungsspezifischen Anteile ausfallen kann. Zumindest für MUC-6 scheinen diese zusätzlichen Arbeiten jedoch einen eher marginalen Charakter aufzuweisen, so daß das Ergebnis als viel versprechend einzustufen ist.

Diskussion: “*Information Extraction*”-Technologie für die verfeinerte CGI

Die Beschreibung des “*Labour Negotiation*”-Szenarios hat gezeigt, daß die algorithmisch lösba- ren MUC-Aufgabenstellungen bereits ein relativ komplexes Format besitzen. Die Ergebnisgüte, die die meisten der an der “*Scenario Template*”-Evaluation teilnehmenden Systeme erzielen, erfüllt zumindest in Teilen hohe Ansprüche. Vergleichsmaßstab ist die Ergebnisqualität, die im Rahmen einer intellektuellen Analyse erreicht wird und die indirekt in Form der Intercoder-Reliabilität²⁵ meßbar ist. Wenn schon zwei menschliche Codierer für eine bestimmte Aufgabe der Inhaltserschließung keinen höheren Grad an Übereinstimmung erzielen, so kann von einem Computer kein besseres Resultat erwartet werden - zumindest, insofern die Ergebnisdiskrepanz auf Schwierigkeiten, die Informationserschließungsaufgabe präzise zu definieren, und nicht auf Flüchtigkeitsfehler der Codierer zurückzuführen ist. Unter diesem Gesichtspunkt liegen die Ergebnisse der Partizipanten für die “*Text Filtering*”-Aufgabe - dem impliziten Teilproblem der “*Scenario Template*”-Aufgabe, zu entscheiden, ob ein bestimmtes Szenario vorliegt - bereits nahe am möglichen Optimum.²⁶ Für die schwierigere “*All Objects*”-Aufgabe, bei der das Ziel in der Identifikation der relevanten Argumente eines vorgegebenen Szenarios besteht, liegt die Performanz mit F-Werten um 0.5 erwartungsgemäß weiter von der möglichen Bestmarke entfernt: Laut Sundheim ([MUC696], S. 27) beträgt die über alle Argumentstellen kumulierte Intercoder-Reliabilität 83 %. Bezüglich der Ergebnisse für die einzelnen Argumentstellen ist das Bild uneinheitlich. Zwar ist damit zu rechnen, daß die für die Computergestützte Informationsextraktion schwierigen Slots - i.a. solche, für die die Information typischerweise über komplexe Operationen, also z.B. aus dem Diskurs oder per Inferenz zu erschließen ist - auch für den menschlichen Codierer problematischer sind; andererseits sollte speziell hier der Vorteil der intellektuellen Analyse voll zum Tragen kommen. Für Argumentstellen hingegen, für die hohe (P, R)-Werte erzielbar sind, ist der Abstand zur menschlichen Performanz wohl nur noch gering.²⁷ Es bleibt der Rückbezug zu den spezifischen Problemstellungen einer verfeinerten Computergestützten Inhaltsanalyse herzustellen. Auf der Basis der durchgeführten Untersuchungen ergeben sich Antworten auf einer relativ generellen Ebene. Mit Blick auf die oben identifizierten zwei grundlegenden Möglichkeiten, eine Verfeinerung der Kategorisierungskriterien mit Hilfe der “*Information Extraction*”-Technologie zu bewerkstelligen, sind zunächst folgende Aussagen möglich:

- Sind die unterschiedlichen Kategorien über paarweise disjunkten Mengen von Szenario-Typen operationalisiert, so entspricht das Klassifikationsproblem der “*Text Filtering*”-Aufgabe von MUC. Da sich dieses Problem mit hohen (P, R)-Werten algorithmisieren läßt, wird den Anforderungen der verfeinerten CGI entsprochen.
- Geschieht die Klassifikation mit Bezug auf Variationen in Attributinstanzen eines Schematypus, der mehreren Kategorien zugeordnet ist, so ist die Möglichkeit einer algorithmischen Lösung davon abhängig, welche Performanz für das spezifische Attribut erzielbar

²⁵Die Intercoder-Reliabilität wird per definitionem in einem einzelnen Prozentwert gemessen, der den Grad der Übereinstimmung von zwei Ergebnisvorschlägen beschreibt. Eine verfeinerte Aufschlüsselung nach Precision und Recall ist nur möglich, wenn eine der beiden Vorgaben zum Referenzwert erklärt wird.

²⁶Zur “*Interannotator Variability*” für die Textfilterungsaufgabe finden sich in der Ergebniszusammenfassung in den Proceedings der MUC-6 (Sundheim, [MUC696], S. 13 ff.) keine Zahlen. Entsprechend den Erfahrungswerten aus den unterschiedlichen Forschungsgebieten der algorithmischen Informationserschließung stellen Werte oberhalb der 97-Prozent-Marke eher die Ausnahme dar.

²⁷In den Proceedings zu MUC-6 finden sich keine Angaben zur slotspezifischen Intercoder-Reliabilität.

ist. Liegen die (P, R) -Werte am oberen Ende des in den MUC-Evaluationen beobachteten Spektrums, so ist die Ergebnisqualität als hinreichend zu bewerten; ist das Kriterium hingegen an einem "schwierigen" Slot verankert, für den niedrige Werte erzielt werden, wird den Anforderungen an eine algorithmische Lösung nicht entsprochen.

Präzisere Aussagen zur Relation zwischen Anforderungen der verfeinerten Computergestützten Inhaltsanalyse und Möglichkeiten der "Information Extraction"-Technologie sind nur auf der Grundlage einer umfassenden Bestandsaufnahme der "typischen" tiefenanalytischen CGI-Probleminstanzen möglich. Ein solches Programm bedingte jedoch die Bezugnahme auf eine möglichst große Menge konkreter Anwendungen der Inhaltsanalyse und fällt somit nicht in den Skopus der vorliegenden Arbeit. Daher sollen lediglich die zwei zentralen Forschungsfragen identifiziert werden, die den Schwerpunkt einer derartigen Exploration bilden.

1. Es ist genauer einzugrenzen, welche *Anforderungen aus Sicht typischer CGI-Probleme* an eine robuste Tiefenanalyse bestehen und welche der im Rahmen der MUCs entwickelten Analysetechniken folglich für eine verfeinerte CGI potentiell relevant sind. Mit anderen Worten: Anstelle der beiden zuvor identifizierten Basismöglichkeiten zur Nutzbarmachung von MUC-Verfahren sollten Szenarien robuster Analyse betrachtet werden, die in möglichst präziser Entsprechung zu den Anforderungen einer verfeinerten CGI stehen.
2. Es ist genauer einzugrenzen, für welche typischen Anforderungen einer verfeinerten CGI auch im Rahmen der "Information Extraction"-Forschung noch keine adäquaten algorithmischen Verfahren entwickelt wurden.

Um nur ein Beispiel für die letztere Fragestellung zu geben, sei auf das für die Inhaltsanalyse so zentrale *Negationsproblem* verwiesen.²⁸ Hingegen scheint das Negationsproblem in den typischen Szenarien der Informationsextraktion zweitrangig zu sein: I.a. berichten Wirtschafts- und sonstige Nachrichtenmeldungen über Ereignisse, die tatsächlich stattgefunden haben.²⁹ Auch betreffend die Portabilitätsanforderung sind die durch die MUC-Evaluationen induzierten Entwicklungen als viel versprechend einzustufen. Speziell das FASTUS-System, in dem die Trennung domänenunabhängigen und domänenspezifischen Wissens besonders weitgehend und konzeptionell ansprechend vollzogen ist, ist als aussagekräftiger Anhaltspunkt dafür zu werten, daß die Anforderungen auch der verfeinerten Computergestützten Inhaltsanalyse erfüllt werden können: Im Hinblick auf die in einer Quantitativen Analyse üblicherweise große Textmenge fällt ein Konfigurationsaufwand zur Algorithmisierung der einzelnen Kategorien von weniger als einem Personentag unter wirtschaftlichen Gesichtspunkten nicht ins Gewicht.³⁰

²⁸Daß dieses Problem von großer Relevanz ist, sieht man am Beispiel der in Abschnitt 4.5 diskutierten Studie von Scheerer und Tarnai mit einer Anwendung der CGI zur Auswertung verbaler Beurteilungen in Zeugnissen ([STa89]): Die An- oder Abwesenheit eines Negationsausdrucks ist i.d.R. ausschlaggebend für die Kategorisierungsentscheidung. Entsprechend nicht ohne Grund bietet auch das Softwarepaket INTEXT eine (extrem eingeschränkte, da theoretisch inadäquate) Funktionalität zur automatischen Negationserkennung.

²⁹So wird etwa in der Beschreibung des PIE-Systems der University of Manitoba ein Satz aus dem Evaluationskorpus angegeben, für den die Nichtberücksichtigung einer Negation zur Generierung einer unerwünschten Ereignisbeschreibung führte ([MUC696], S. 124). Obwohl das System somit offenbar über keinen umfassenden Mechanismus zur Negationserkennung verfügt, gehörte es dennoch zur Gruppe der besten MUC-6-Partizipanten.

³⁰Als Vergleichsmaßstab kann hier eine Angabe von Merten herangezogen werden, derzufolge alleine für Auswahl, eventuelle Beschaffung sowie anwendungsspezifische Konfiguration eines Standardsoftwarepakets zur statistischen Analyse sechs Monate Vorlaufzeit einzuplanen sind ([Mert95], S. 324).

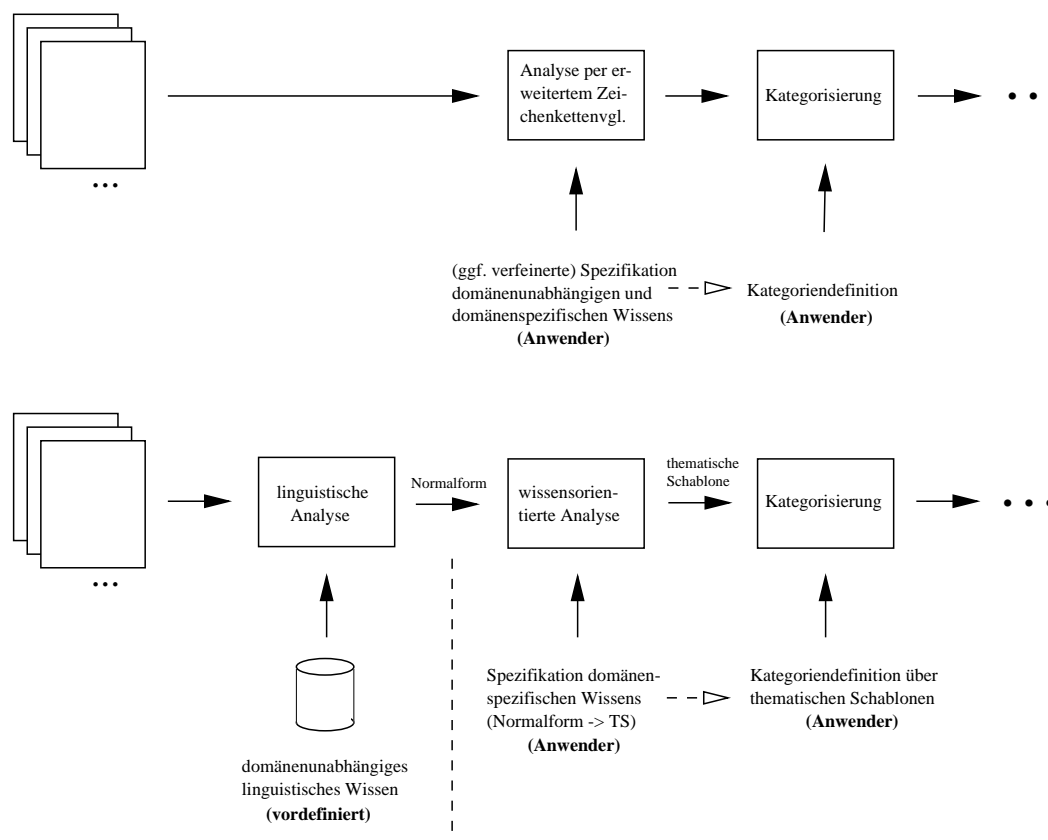


Abbildung 5.3: Herkömmliche versus verfeinerte, portablem Computergestützte Inhaltsanalyse

Zu berücksichtigen ist allerdings auch hier, daß sich die zeitliche Angabe auf die Konfiguration des FASTUS-Systems für die Domäne des Szenarios von MUC-6 bezieht. Eine definitive Aussage kann daher erst unter Bezugnahme auf typische Probleminstanzen einer verfeinerten CGI erfolgen, die möglicherweise komplexer ausfallen und für die ggf. ein größerer Zeitaufwand anzusetzen wäre. Eine weitere Forschungsfrage mit zentraler Relevanz, die im Rahmen einer umfassenden Exploration charakteristischer Aufgaben einer feinkörnigeren CGI zu beantworten wäre, ist daher:

3. Wie groß ist der zu erwartende Aufwand für eine Konfiguration des Analysesystems zur Algorithmisierung einer "typischen" Kategorie in der verfeinerten CGI?

Für die Praktikabilität der verfeinerten CGI unabdingbar ist die Verfügbarkeit von Spezifikationssprachen und Werkzeugen, die auf die Bedürfnisse der Anwender zugeschnitten sind. Die von den FASTUS- und "NYU"-Entwicklergruppen angestrebte Realisierung graphischer Anwenderschnittstellen und konzeptuell einfacher Spezifikationssprachen koinzidiert somit mit den Anforderungen der Computergestützten Inhaltsanalyse. Entsprechend den Erfahrungsberichten zur MUC-6-Evaluation ist zu erwarten, daß diese Zielsetzungen bereits mittelfristig erreichbar sind.

Die Techniken der Wissensseparation, die den wesentlichen Beitrag für die Konstituierung einer

anwendungsflexiblen, portablen Systemarchitektur leisten, gehen einher mit einem entscheidenden Vorteil gegenüber dem zuvor bereits als inadäquat identifizierten Beschreibungsmodell, das z.B. von der wortorientierten TEXTPACK-Erweiterung TEXTLING angeboten wird. Die im Rahmen der Diskussion in Abschnitt 4.5 identifizierte Kernproblematik, derzufolge die prinzipielle Unbeschränktheit der Variabilität sprachlicher Ausdrucksmittel zur Kommunikation ein und desselben Inhalts zu einer für den Anwender von TEXTPACK/TEXTLING prinzipiell unhandhabbaren Spezifikationsaufgabe führt, kann mit Hilfe der neuen Analysemodelle zwar nicht vollständig gelöst werden; jedoch ist es nunmehr möglich, eine Unterstützung in Form einer *automatischen Generierung von Analysevarianten für stereotype oberflächenstrukturelle (syntaktische) Ausdrucksmöglichkeiten einer relationalen inhaltlichen Entität* anzubieten.

Anhand der Skizzen in Abbildung 5.3 lassen sich die Vorteile einer solchen Lösung für die verfeinerte Computergestützte Inhaltsanalyse identifizieren. Durch die Trennung domänenunabhängigen (insbesondere linguistischen) und domänenspezifischen Wissens wird der Anwender von der zeitaufwendigen Aufgabe entbunden, für die unterschiedlichen Beschreibungsinstanzen stets erneut dieselben syntaktischen Ausdrucksvarianten zu explizieren; neben der Zeitersparnis bringt ein solches Vorgehen die bereits für den MUC-Anwendungskontext identifizierten Vorteile der Fehlervermeidung qua Algorithmisierung sowie der geringeren Anforderungen an das Vorwissen des Spezifikators, der nun nicht mehr über computerlinguistische Kompetenz zu verfügen braucht - letztere Vereinfachung ist Grundvoraussetzung für die allgemeine Verbreitung von Softwarelösungen der verfeinerten CGI. Anstelle sich unter Zuhilfenahme theoretisch inadäquater Beschreibungswerkzeuge immer wieder erneut mit den syntaktischen Ausdrucksvarianten zu befassen, überläßt der Spezifikator diesen Part dem Analysesystem, das über das entsprechende syntaktische Vorwissen verfügt. Bezugspunkt der Kategorienbeschreibung ist nunmehr eine - i.a. inhaltsorientierte - *Normalform*, die bereits von den oberflächenstrukturellen Varianten abstrahiert. Als klassisches Beispiel sind relationale Darstellungen zu nennen, deren Argumentstellen nach Tiefenkasus geordnet sind;³¹ auf dieser Grundlage könnte z.B. eine Operationalisierung der Beispielskategorie "Schaffens-Akte" aus Abschnitt 4.5 bewerkstelligt werden, vermöge derer eine Fehl kategorisierung der dort beschriebenen problematischen Fälle von vornherein ausgeschlossen wird, da die Passiv- und Deverbativvarianten automatisch Berücksichtigung finden und ferner der semantisch irrelevante Beleg auf Basis der einhergehenden verfeinerten Kontextanalyse als solcher erkannt und verworfen wird (vgl. S. 57).

Im Rahmen der Explikation des domänenspezifischen Teils der Analyse ist nunmehr zu spezifizieren, auf welche Weise die normalisierte, anwendungsunabhängige Repräsentation auf die letztlich gesuchten Wissensstrukturen - thematische Schablonen - abzubilden sind. In der vCGI schließt sich die Beschreibung der Kategorisierungskriterien an, in der Bezug auf Wissensschemata genommen wird, die an die Stelle inadäquater wortbasierter Spezifikationskonzepte treten.

Precision oder Recall?

Schließlich stellt sich die Frage nach der relativen Relevanz der Precision- und Recall-Werte, die ja - wie zuvor beschrieben - in einem Austauschverhältnis stehen: Welche Kombination ist aus Sicht der Anforderungen der Computergestützten Inhaltsanalyse zu bevorzugen?

Mit Blick auf deren Auswirkungen für die einzelnen - positiven oder negativen - Kategorisierungsentscheidungen erscheinen beide Größen für die Ergebnisse der CGI gleichermaßen wichtig:

³¹Hiermit soll jedoch keineswegs behauptet werden, daß speziell dieser Normalisierungsprozeß ohne Weiteres und uneingeschränkt algorithmisch bewerkstelligt werden kann.

- Eine geringere *Precision* der Algorithmisierung eines Klassifikationskriteriums schlägt sich in Fehlkategorisierungen nieder, die einem Verlust an negativer Evidenz entsprechen;
- ein geringerer *Recall* der Algorithmisierung eines Klassifikationskriteriums schlägt sich in irrtümlich unterlassenen Kategorisierungen nieder, die einem Verlust an positiver Evidenz entsprechen.

Eine genauere Betrachtung zeigt jedoch, daß speziell in diesem Punkt ein wichtiger Unterschied zu den Anforderungen der Problemstellungen der Informationsextraktion besteht. Aufgrund der qualitativen, einzelfallerhaltenden Ergebnisstruktur sind für Anwendungen des “*Information Extraction*” beide Werte als gleichermaßen bedeutsam einzustufen.³² Für die CGI, durch die in der Regel eine Abstraktion von den konkreten Einzelfällen bewerkstelligt werden soll, an die sich eine quantitative Auswertung anschließt, scheint der Precision-Wert die primär relevante Größe zu sein. Dies läßt sich mit Blick auf die unterschiedlichen Bezugsgrößen von Precision- und Recall-Faktor begründen. Ein nichtmaximaler Recall der Algorithmisierung eines Klassifikationskriteriums bewirkt, daß ein bestimmter Prozentsatz positiver Codierentscheidungen nicht getroffen wird; unter der idealisierenden Annahme, daß der Recall-Fehler in gleichem Maße für alle Dokumente bzw. Mengen von Dokumenten (Korpora) zum Tragen kommt, bleibt der Quotient der Codierungshäufigkeiten zweier Dokumente bezüglich dieser Kategorie gleich: Seien D_i und D_j zwei Dokumente, K eine Kategorie, $h(D_i, K)$ und $h(D_j, K)$ die Anzahl der Instanziierungen von K durch D_i bzw. D_j sowie $r(K)$ der Recall-Faktor für die Operationalisierung von K , dann ist

$$\frac{h(D_i, K)r(K)}{h(D_j, K)r(K)} = \frac{h(D_i, K)}{h(D_j, K)}$$

Hingegen schlägt sich eine nichtmaximale Präzision der Algorithmisierung eines Klassifikationskriteriums in *additiven* Komponenten nieder, deren Wert unter analoger Uniformitätsannahme abhängig von der Länge (der Anzahl der codierrelevanten Einheiten) eines Dokuments ist. Somit würde der Quotient der Codierungshäufigkeiten zweier Dokumente verändert: Seien zusätzlich $l(D_i)$ und $l(D_j)$ die Längen der Dokumente D_i bzw. D_j und $p(K)$ der Precision-Faktor für die Operationalisierung von K , so ergibt sich der Quotient nunmehr approximativ als

$$\frac{h(D_i, K) + (1 - p(K))l(D_i)}{h(D_j, K) + (1 - p(K))l(D_j)}$$

Dieser Unterschied erweist sich unter meßtheoretischem Gesichtspunkt als bedeutsam. Insofern die quantitative Auswertung unter Ausnutzung des *Rationalskalenniveaus* der Messung geschieht, ist nunmehr eine Verringerung des Recall, nicht jedoch der Precision hinnehmbar, da wie gezeigt - letzterer Fall mit einer Veränderung des *Verhältnisses* der Skalenwerte einhergeht. Auch dann, wenn die Auswertung auf Nominalskalenniveau erfolgt, kann Precision wichtiger sein als Recall. Mit der Länge eines bezüglich einer Kategorie irrelevanten Dokuments steigt die Wahrscheinlichkeit, daß eine Zuordnung irrtümlich geschieht. Ein niedriger Precision-Wert kann das Ergebnis aus statistischer Sicht viel nachhaltiger verwässern als ein niedriger Recall, dessen negative Auswirkungen sich zumindest im Falle einer großen Anzahl zu analysierender

³²Im Hinblick darauf, daß die weitere Auswertung der durch ein “*Information Extraction*”-System erzeugten standardisierten Faktenbeschreibungen i.a. durch einen menschlichen Interpreten geschieht, könnte - zumindest bei nicht allzu geringer Precision - der Recall-Wert als die zu optimierende Größe eingestuft werden: Inkorrekte Ergebnisse lassen sich a posteriori intellektuell filtern, wohingegen nicht extrahierte Ergebnisse nur unter Anschauung des *vollständigen* Eingabekorpus rekonstruierbar wären.

Dokumente in Grenzen halten. Eine definitive Aussage hängt von mehreren Parametern ab, so u.a. vom genauen Analyseszenario, von der durchschnittlichen Dokumentgröße und von den Eckwerten des Austauschverhältnis zwischen Precision und Recall.

Somit scheint die für die weitaus meisten MUC-Teilnehmer zu beobachtende Precision-Tendenz den Anforderungen einer verfeinerten Computergestützten Inhaltsanalyse entgegenzukommen.³³ Forschungsarbeiten, deren Gegenstand die Entwicklung von Analyseverfahren mit besonderer Ausrichtung des (P, R) -Austauschverhältnisses ist, sind daher auch für die Weiterentwicklung der CGI relevant.³⁴

5.5 Zusammenfassung

Die Ausführungen haben gezeigt, daß für einen großen Teil der in Abschnitt 4.1 identifizierten Probleme der wortorientierten Kategorienanalyse praxistaugliche computerlinguistische Lösungen verfügbar sind, die im Hinblick auf Abdeckungsgrad und Analysegeschwindigkeit hohen Ansprüchen genügen. Dies gilt sowohl für die morphologische Analyse als auch für die Wortart-Disambiguierung. Für die allgemeinere Problemstellung der lexikalischen Disambiguierung sind heuristische Lösungsansätze erforscht und verfeinert worden.

Der Schwerpunkt lag jedoch auf der Exploration der Machbarkeit einer verfeinerten Computergestützten Inhaltsanalyse, durch die die Einzelwortebene in inhaltsadäquater Weise transzendiert wird. Den Ausgangspunkt bildete die Fallstudie in Abschnitt 4.3, anhand derer die Zielsetzung der Berücksichtigung nichtlexikalischer, relationaler inhaltlicher Entitäten als Grundlage einer verfeinerten Themenanalyse abgeleitet wurde. Bereits eine rein syntaktische Analyse, durch die per Definitionem nur die sehr eingeschränkte oberflächenstrukturelle Widerspiegelung sprachlicher Inhalte erschlossen wird, erwies sich unter Operationalisierungsgesichtspunkten als problematisch. Dies führte zum Wechsel der Perspektive, wonach ein wissensorientierter Ansatz verfolgt werden sollte, in dem die relevanten Inhalte anforderungsorientiert und a priori unabhängig von spezifischen linguistischen Beschreibungsebenen betrachtet werden.

Der Analyseansatz der Wahl sollte den Rahmenanforderungen der Massendatentauglichkeit genügen. Ein wichtiges Unterkriterium ist die Robustheitszielsetzung, derzufolge zu gewährleisten ist, daß das Verfahren auch für Textdokumente arbeitet, die sich in orthographischem oder grammatischem Sinne als fehlerhaft erweisen - eine Eigenschaft, die für Verfahren der Computerlinguistik keineswegs selbstverständlich ist. In Abschnitt 5.4.2 wurden erste Anhaltspunkte dafür erarbeitet, daß die auf dem Gebiet des *Information Extraction* eingesetzten Textanalysetechniken den Anforderungen einer verfeinerten Kategorienanalyse in sowohl funktionaler als auch technischer Hinsicht entsprechen. Die Entwicklungen im Kontext der MUC-Evaluationen sind als besonders geeigneter Anknüpfungspunkt anzusehen, da aufgrund der gezielten Förderung anwendungsportabler Systemarchitekturen eine weitere Grundvoraussetzung für die Anwendung dieser Technologie auf dem Gebiet der CGI im Mittelpunkt der Forschung steht. Ob diese Zielsetzung letztlich erreicht werden kann, hängt entscheidend davon ab, inwieweit die Aufgabe der Domänenadaptation - in der CGI-Welt entsprechend der Spezifikation formaler Kategorisie-

³³Zwei unter diesem Gesichtspunkt besonders hervorsteckende MUC-6-Systeme erzielten für die "All Slots"-Aufgabe (P, R) -Kombinationen von (73,37) bzw. (70,47).

³⁴Die allgemeine Anerkennung der Relevanz dieser Problemstellung schlägt sich in aktuellen MUC-Beiträgen nieder: Einer der "Scenario Template"-Partizipanten - Systems Research and Applications, Fairfax - reichte drei Systemversionen zur Evaluation ein: Basisversion, Precision-Variante und Recall-Variante (vgl. MUC-6-Proceedings, [MUC696], S. 221 ff.).

rungskriterien - vermöge geeigneter Benutzerschnittstellen auf den Anwender übertragen werden kann.

Als Forschungsaufgabe wurde das Programm einer umfassenden Untersuchung formuliert, in deren Rahmen die genaue Relation zwischen typischen Probleminstanzen einer verfeinerten Computergestützten Inhaltsanalyse und den Anwendungsszenarios von “*Information Extraction*”-Systemen untersucht werden sollte. Auf diesem Weg sollten präzisere Aussagen zur Übertragbarkeit der Analysetechniken auf den neuen Anwendungskontext erarbeitet sowie eventuell bestehende Lücken identifiziert werden.

Somit kann den Defiziten der wörterbuchbasierten Inhaltsanalyse wirkungsvoller begegnet werden als durch Spezifikationskonzepte, wie sie etwa das Erweiterungspaket TEXTLING zu TEXTPACK zur Verfügung stellt. Beispielsweise könnte die in Abschnitt 4.5 exemplarisch diskutierte Operationalisierung der relationalen Kategorie “*Schaffens-Akte*” nun auf vereinfachte und adäquate Weise bewerkstelligt werden: In den Analyseansätzen des Information Extraction kommen Techniken zur Anwendung, die den Anwender von der - unzumutbaren - Obligation befreien, jede mögliche Oberflächenrealisierung in der Kategorienbeschreibung vorwegzunehmen. Auch unter Verzicht auf eine technisch problematische vollständige syntaktische Analyse können die auf der oberflächenstrukturellen Ebene bestehenden Variationsmöglichkeiten mit gutem Erfolg automatisch berücksichtigt werden. Insofern kann auch dem in Abschnitt 4.5 zitierten Problem von Scheerer und Tarnai abgeholfen werden, deren Versuch einer möglichst vollständigen Operationalisierung von Ausdrucksvarianten per Enumeration in TEXTPACK an der Vielfalt sprachlicher Ausdrucksmöglichkeiten scheiterte - die Trennung des generellen linguistischen Wissens von den anwendungsspezifischen Wissenskomponenten schafft hierfür die Voraussetzung und fördert überdies die Portierbarkeit.

Unter Rückbezug auf die eingangs zitierte Arbeit von Fühlau ([Fühl82]), derzufolge die zum damaligen Zeitpunkt verfügbaren algorithmischen Ansätze zur wortorientierten Inhaltserschließung als unter linguistischen Gesichtspunkten untauglich (“*sprachlos*”) - zu bewerten sind, ist nunmehr auf der Grundlage neuerer computerlinguistischer Forschungsergebnisse ein zumindest partielles Überdenken der dort geäußerten skeptischen Grundhaltung gegenüber einer Computergestützten Inhaltsanalyse angezeigt. Die Computerlinguistik hat mittlerweile robuste Analysetechniken hervorgebracht, denen vergleichsweise feine Konzepte textueller Inhaltserschließung zugrunde liegen. Zwar verbleibt im Rahmen von vertiefenden Studien der konkrete Nachweis zu erbringen, ob sich diese Ansätze für die spezifischen Anforderungen einer verfeinerten Computergestützten Inhaltsanalyse nutzbar machen lassen - jedenfalls stellen die Ergebnisse der obigen Betrachtungen bereits ein starkes Indiz dafür dar.

Die Ergebnisse der durchgeführten Untersuchungen sind als Evidenz für die Gültigkeit der eingangs postulierten These 2 zu werten. Insofern ein objektiver, am Ergebnis orientierter Bewertungsmaßstab angelegt und die Erfolge der neuartigen, auf elaborierten, robusten computerlinguistischen Analysetechniken basierenden Softwaresysteme anerkannt werden, bleibt nur die Möglichkeit, zu konzedieren, daß unter Anwendungsgesichtspunkten bereits ein entscheidender Schritt hin zur Algorithmisierung des jeder intellektuellen Analyse zugrundeliegenden Hintergrundwissens vollzogen wurde. Dieses Resultat stellt einen ersten Beleg für die Gültigkeit des Kernpostulats der Arbeit (These 5) dar. Eine zu stringente Auslegung der Aussagen der hermeneutischen Philosophie, derzufolge es aus theoretischen Gründen *unmöglich* ist, das einer intellektuellen Analyse zugrundeliegende “implizite Vorverständnis” im Rahmen der Realisierung algorithmischer Verfahren zur qualitativen Textanalyse in hinreichendem Umfang zu explizieren, wird diesen neueren Ergebnissen nicht gerecht. Somit läßt sich für eine *abgeschwächte, relativierende* Variante der hermeneutischen Betrachtungsweise argumentieren: Die Algorithmisierung bestimmter (prozeduraler und Wissens-)Komponenten des Verstehensprozesses ist u.U.

schwierig; jedoch kann hieraus nicht die theoretische Unmöglichkeit jedweder algorithmischer Inhaltsanalyse gefolgert werden. Obwohl die hermeneutische Theorie häufig dazu herangezogen wird, um gegen die Zielsetzung maschineller Intelligenz in toto zu argumentieren, sind aus ihr gleichwohl keine harten Machbarkeitsschranken ableitbar, wie auch von D’Avis herausgestellt wird ([DAvi94], S. 125 f., Hervorhebungen im Original):

“Zieht man sich nicht zurück auf die Position des prinzipiell Unaussprechlichen als Besonderheit des interpretationswirksamen ‘Hintergrundes’, sondern geht von seiner diskursiven Einlösbarkeit aus (auch wenn die Verbalisierung im Einzelfall mühsam sein und nicht immer gelingen mag), so liegt in der hermeneutischen Forderung nach einem Vorverständnis bisher keine prinzipielle Grenze der maschinellen Symbolverarbeitung, die mit den gegebenen theoretischen Mitteln gezeigt werden könnte.”

Auch eine Maschine kann über einen “impliziten Verstehenshintergrund” verfügen; es existiert keine theoretische Rechtfertigung dafür, das “implizite Vorverständnis” eines menschlichen Interpreten als prinzipiell unzugänglich für eine Algorithmisierung einzustufen. Eine einfache Unterscheidung “impliziten” und “expliziten” i.S.v. “explizierbaren” Wissens ist inadäquat. Vielmehr scheint es eine Frage der Vertretbarkeit des zu treibenden Aufwands zu sein, die mit Blick auf das jeweilige Interpretationsproblem sowie auf die Algorithmisierungstechniken, die auf dem aktuellen Stand der Forschung zur Verfügung stehen, zu beantworten ist.

Für die Problemstellungen der MUC-Evaluationsszenarien wurde bereits der Nachweis erbracht, daß die zu erbringenden - in weiten Teilen nichttrivialen - Interpretationsleistungen auf dem gegenwärtigen Stand computerlinguistischer Textinterpretationsverfahren mit vertretbarem Aufwand algorithmisch explizierbar sind. Zwar kann hieraus nicht gefolgert werden, daß bereits in nächster Zukunft beliebig subtile Interpretationsprobleme wie etwa die Konstruktion textueller Kohärenzstrukturen (vgl.o.) maschinell gelöst werden können. *Jedoch wurden zentrale Anhaltspunkte dafür gewonnen, daß den Anforderungen einer verfeinerten Computergestützten Inhaltsanalyse bereits heute Genüge getan werden kann.*

Bisher wurde ein Punkt ausgeklammert, dessen grundlegende Bedeutung für die Computergestützte Inhaltsanalyse bereits mehrfach herausgestellt wurde: die algorithmische Interpretation referentieller (insbesondere pronominaler) Ausdrücke. Im nun folgenden Teil II der Arbeit soll dieses Problem ausführlich behandelt und somit die Thesen 3 und 4 belegt sowie zusätzliche Evidenz für das Kernpostulat (These 5) gesammelt werden. Wie sich zeigen wird, fällt die als schwierig bekannte Aufgabe der Pronomenresolution in die Klasse derjenigen Probleme, die auf dem gegenwärtigen Stand der Technik robuster Textanalyse mit vertretbarem Aufwand lösbar sind. Die Ausführungen werden sich nicht auf eine allgemeine Darstellung beschränken: Für dieses wichtige Teilproblem der Textinterpretation sollen Detailstudien betrieben werden, in deren Verlauf konkrete Algorithmen entworfen werden, die den Rahmenanforderungen der vCGI - Massendatentauglichkeit, Robustheit und Portierbarkeit - gerecht werden. Auf der Basis einer Implementierung wird die Funktionsweise der Algorithmen anhand von praktischen Anwendungen demonstriert und deren Leistungsfähigkeit evaluiert. Auch hier wird sich somit die Einschätzung Fühlhaus, derzufolge das Problem der Interpretation von Pronomen *“kaum von Computern zu lösen sein”*³⁵ würde, widerlegen lassen.

³⁵vgl. Abschnitt 4.1 sowie [Fühl82], S. 140

Teil II

Algorithmische Interpretation referentieller Ausdrücke

Kapitel 6

Referentialität

Gegenstand von Teil I der Arbeit waren Untersuchungen zu Leistungsfähigkeit und Grenzen der Computergestützten Inhaltsanalyse, in deren Rahmen eine allgemeine Perspektive angenommen wurde. Als prinzipielles Defizit wurde der mangelnde Inhaltsbezug der vermögelisten von Zeichenketten sowie einfacher Mengen von Operatoren realisierten Themenanalyse identifiziert. Die hier implizit zugrunde gelegte lexikalische Ebene der Entwicklung textueller Themen, die Isotopie, verkörpert nur einen elementaren Spezialfall; für eine generellere Analyse der thematischen Entwicklung eines Texts erweist sich eine elaboriertere inhaltliche Analyse des Texts unabdingbar, wobei im allgemeinen ein Rekurs auf Wissen über den außersprachlichen Kontext notwendig ist. Auch wenn dieser Anspruch auf der Grundlage gegenwärtiger computerlinguistischer Analyseverfahren nur partiell eingelöst werden kann - die Diskussion der in den letzten Jahren auf dem Gebiet des *Information Extraction* erzielten Ergebnisse hat gezeigt, daß die Aussichten einer algorithmischen Textinhaltserschließung nicht so aussichtslos sind, wie es die einschlägigen Kritiker dieses Anspruchs stets betonen.

In Teil II der Arbeit wird ein zentrales Teilproblem der Interpretation natürlichsprachiger Texte, die Analyse der referentiellen Verweisstruktur, thematisiert. Die globale Betrachtungsperspektive soll zugunsten einer detaillierten, algorithmischen Ausrichtung der Untersuchung aufgegeben werden. Die Ausführungen in Teil I haben verdeutlicht, daß die referentielle Interpretation für eine verfeinerte Computergestützte Textinhaltsanalyse in zumindest zweierlei Hinsicht relevant ist: einerseits, da die referentielle Progression des Textes einen (nichttrivialen) Spezialfall der Themenentwicklung verkörpert; andererseits, da die Zuordnung *pronominaler* Referenzen zu Antezedenten mit konzeptuellem Gehalt (nichtpronominale Nomen, Namen) für sich betrachtet ein zentrales Problem der klassischen wörterbuchbasierten Inhaltsanalyse darstellt. Falls der Nachweis gelingt, daß dieser spezielle qualitative Teilprozeß der Textinterpretation algorithmisch expliziert und somit durch Computer ausgeführt werden kann, ist ein bedeutender Schritt hin zu einer inhaltsnäheren maschinellen Textanalyse vollzogen.

Hierbei ist eine Vorgehensweise zu wählen, durch die sichergestellt wird, daß die in Abschnitt 5.1 identifizierten Anforderungen - Massendatentauglichkeit, Robustheit, Algorithmeigenschaft, Analysegüte - erfüllt werden. Deshalb werden sich die Ausführungen nicht auf die alleinige *An-gabe* vermeintlicher Algorithmen beschränken; nur durch die tatsächliche Implementierung wird ein einwandfreier Nachweis der Operationalität erbracht. Hierdurch wird auch erst die Voraussetzung dafür geschaffen, die Tauglichkeit der Analyseverfahren im Rahmen einer formalen Eva-

lation der Performanz auf *realen* Evaluations-Korpora nachzuweisen.¹ Somit soll entsprechend der Methodik verfahren werden, die der im Rahmen der MUCs durchgeführten “*Coreference Task*”-Evaluation zugrunde liegt.

Ausgangspunkt der Entwicklung eines robusten, anwendungstauglichen Algorithmus wird eine Bestandsaufnahme bekannter theoretischer Modelle und Verfahren zur referentiellen Interpretation sein, in deren Rahmen eine Reihe wichtiger Gesichtspunkte zu diskutieren ist. Welche Teilprobleme werden abgedeckt? Welche Theorien liegen zugrunde? Inwieweit wird den zuvor postulierten Anforderungen entsprochen? Die Betrachtungen werden die zentrale Relevanz einer bestimmten Klasse konfiguratoraler Restriktionen herausstreichen, deren Anwendung die Verfügbarkeit syntaktischer Beschreibungen voraussetzt. Daß diese Annahme u.a. aufgrund von strukturellen Ambiguitäten unrealistisch ist, wurde in Abschnitt 5.4.1 hinlänglich diskutiert. An diesem Problempunkt setzt die vorliegende Arbeit an: Es werden Algorithmen entwickelt, die eine *robuste* Verifikation syntaktisch-konfiguratoraler Restriktionen auf *partiellen* syntaktischen Beschreibungen bewerkstelligen. Der bislang gemäß intuitivem Verständnis gebrauchte Begriff der *Robustheit* sprachinterpretierender Prozesse soll nunmehr ein solides definitorisches Fundament erhalten, dem eine *graduelle Differenzierung* unterliegt. Die ultimativ robuste Variante des Algorithmus zur Koreferenzinterpretation wird eine *Verzahnung von syntaktischer und referenzsemantischer Analyse* realisieren.

Zunächst sind jedoch einige allgemeine Vorbetrachtungen zum Themenkreis Referentialität angesagt. Es soll eine Grundlage geschaffen werden, die eine Klärung des theoretischen Status einiger zentraler Begriffe erlaubt und es ferner ermöglicht, Anspruch, Reichweite und Grenzen der zu entwickelnden algorithmischen Analysemodelle festzulegen. Hierzu sollen einige Fallbeispiele besprochen werden, die die Vielschichtigkeit des Phänomens der Referentialität aufzeigen und einen ersten Eindruck von der Komplexität der Problemstellung einer algorithmischen referentiellen Interpretation vermitteln.

6.1 Sprachliche Ausdrücke und referentieller Wiederaufgriff

Vorab sei bemerkt, daß mit den folgenden Ausführungen keineswegs der Anspruch einer vollständigen Taxonomie referenzierender sprachlicher Ausdrücke verbunden ist.² Das Ziel besteht vielmehr darin, eine Grundlage für die Diskussion des theoretischen Status folgender zentraler Begriffe zu schaffen, die bislang mehr oder weniger intuitiv gebraucht wurden:

- Anapher / anaphorischer Ausdruck,
- Antezedens,
- Vorkommen / Okkurrenz,
- Referenzobjekt,

¹Wie im folgenden anhand einiger klassischer computerlinguistischer Ansätze für das Problem der referentiellen Interpretation deutlich werden wird, genügt ein beachtlicher Teil der Analysemodelle diesen Anforderungen *nicht* bzw. nur teilweise - sei es, daß der Algorithmeigenschaft nicht Genüge getan ist, und/oder sei es, weil die zugrundeliegende Theorie anhand von Spielzeugbeispielen entwickelt wurde, die mit den für die relevanten Anwendungskorpora typischen sprachlichen Phänomenen wenig bis überhaupt nichts zu tun haben.

²In bezug auf die englische Sprache vergleichsweise umfassende Übersichten finden sich etwa in den Arbeiten von Hirst ([Hirs81]) sowie Halliday und Hasan ([HaHa76]).

- Referenz,
- Koreferenz / referentieller Wiederaufgriff.

Es wird sich zeigen, daß für den Kernbegriff der *Referenz* eine konzeptuelle Differenzierung indiziert ist, die sich in der Einführung des Begriffs der *Spezifikation* niederschlagen wird. Die im folgenden diskutierten Belege sind in deutscher Sprache abgefaßt. Die Übertragbarkeit der Beispiele in andere Sprachen ist zumindest soweit gewährleistet, daß der darauf aufbauenden theoretischen Begriffsklärung sprachübergreifende Gültigkeit zugeschrieben werden kann.

6.1.1 Elementare Fallbeispiele

Einige elementare Fälle *pronominalen* referentiellen Bezugs wurden bereits in Teil I der Arbeit diskutiert. Aus der Perspektive einer algorithmischen Inhaltserschließung ist eine (unmittelbare oder mittelbare) Zuordnung zu nichtpronominalen Antezedenten wünschenswert.

- (1a) Peter Behrens_i baut sich_i ein Haus_j.
 (1b) Peter Behrens_i baut seinem_i Kollegen_j eine Gartenlaube_k.
 (1c) Peter Behrens_i ist berühmt. Er_i baut Fabrikhallen_j.

Die referentiellen Wiederaufgriffsrelationen sind durch *Koindexierung*, d.h. per Kennzeichnung mit gemeinsamen *Referenzindexen* (tiefgestellten Kleinbuchstaben), hervorgehoben.³ Die Unterscheidung zwischen Reflexivpronomen (Beleg (1a)), Possessivpronomen (Beleg (1b)) und gewöhnlichen Personalpronomen in dritter Person (Beleg (1c)) wird sich im Rahmen der algorithmischen Behandlung als bedeutsam erweisen.

Weitere Formen anaphorischen Aufgriffs im engeren Sinne werden durch Relativpronomen, substantivische Demonstrativpronomen und Pronominaladverbien konstituiert:

- (2a) Behrens_j traf einen Architekten_i, mit dem_i er_j gemeinsam studierte.
 (2b) Behrens_j war mit Walter Gropius_i befreundet.
Dieser_i lehrte am Bauhaus_k zu Weimar_l.
 (2c) Behrens_j war der Vorgang_i peinlich.
 Er_j hüllte sich_j darüber_i in Schweigen_k.

Jedoch stellen Pronomen keineswegs die einzige Möglichkeit dar, einen referentiellen Wiederaufgriff zu bewerkstelligen. Insbesondere Nominalphrasen mit *bestimmtem* Artikel leisten oftmals Vergleichbares:

³Referenzindexe sind ein formalgrammatisches Beschreibungsmittel zur Postulierung identischer "Referenz", d.h. Wiederaufgriff desselben Redeobjekts oder zumindest des gleichen Konzepts (vgl. u., Abschnitt 6.1.2). Durch die Angabe von Referenzindexverteilungen wird somit bereits eine kontextuelle *Interpretation* des Satzes unter referentiellen Gesichtspunkten vorgenommen, die nicht unbedingt die einzig Mögliche sein muß.

In Chomskys *Government and Binding Theory* stellen Referenzindexe ein Beschreibungsmittel der satzgrammatischen Ebene dar ([Chom81]). Im Einklang mit der üblichen Praxis - und nicht zuletzt auch aus inhaltlichen Gründen (vgl. u.) - bietet es sich jedoch an, diese Notation auch zur Kennzeichnung *satzübergreifender* Koreferenz zu gebrauchen.

- (3) *Peter Behrens_i baut sich_i ein Haus_j.
Der große Architekt_i hat sich_i zwei Monate Urlaub_k genommen.*

In erster Annäherung läßt sich festhalten, daß der definite Artikel einen Indikator des Wiederaufgriffs eines im Diskurs *bereits erwähnten* Elements verkörpert, wohingegen mit indefinitem Artikel signalisiert wird, daß ein *neues* Element eingeführt wird.⁴

Auch Eigennamen kommt eine referentielle Funktion zu. So könnte der propositionale Inhalt der Satzabfolge (2b) alternativ wie folgt ausgedrückt werden:⁵

- (4) *Behrens_j war mit Walter Gropius_i befreundet.
Gropius_i lehrte am Bauhaus_k zu Weimar_l.*

Eine Funktion referentiellen Aufgriffs üben oftmals auch an der sprachlichen Oberfläche nicht-realisierte *elliptische Elemente* aus:

- (5) *Behrens_i begrüßte Gropius_j und ε_i bot ihm_j einen Tee_k an.*

Es ist ohne Weiteres ersichtlich, daß sich der propositionale Gehalt des zweiten Teilsatzes nur unter Erschließung des "leeren" Subjekts bestimmen läßt. Die Konstruktion einer adäquaten inhaltlichen Beschreibung setzt also dessen Identifikation sowie die Bestimmung des referentiellen Bezugs im Rahmen einer syntaktisch-semantischen Strukturanalyse voraus.

6.1.2 Komplexere Instanzen referentiellen Wiederaufgriffs

Die bisher angeführten Belege sind in gleich mehrfacher Hinsicht elementar. Sie geben insofern ein nur unvollständiges Abbild der Vielschichtigkeit des Phänomens der Referentialität und damit auch der Komplikationen, die im Rahmen einer Algorithmisierung zu lösen sind. Die erste und wesentlichste Einschränkung betrifft den *Typ* der Referenz: Die bisherigen Beispiele drehten sich ausnahmslos um Instanzen von *Objekt-Referenz*, in denen Elemente wiederaufgegriffen werden, die durch Nominalphrasen realisiert wurden. Durch referentielle Verweise können jedoch auch relationale inhaltliche Entitäten (Sachverhalte, Ereignisse, Situationen etc.) re-instantiiert

⁴Was als *bekannt* und was als *neu* angesehen wird, hängt nicht zuletzt auch vom angenommenen kontextuellen Wissen ab. Das mit definitivem Artikel markierte Nomen kann durchaus auf ein außertextuelles Element Bezug nehmen. Beispielsweise würde folgende Artikelüberschrift in einer deutschen Zeitung von den allermeisten Lesern korrekt interpretiert werden:

Der Kanzler kommt nach Frankfurt.

⁵In den Belegen (2b) und (4) würde die Verwendung des gewöhnlichen Personalpronomens "er" zu einer referentiellen Ambiguität führen bzw. die (falsche) Zuordnung des möglichen Antezedens "Behrens" nahelegen. Die Verwendung des Demonstrativpronomens oder der Namensform beseitigt diese Mehrdeutigkeit. Derartige Beziehungen, die den diskutierten Beispielen zugrundeliegen, sollen zunächst nicht näher herausgestellt werden. Sie werden sich jedoch weiter unten, im Rahmen der Festlegung pronomentyp-spezifischer algorithmischer Entscheidungsheuristiken, als bedeutsam erweisen. Es sei zunächst lediglich bemerkt, daß hierbei ähnliche Faktoren eine Rolle spielen, wie sie für das Problem der Erzeugung kontextadäquater referenzierender Ausdrücke im Rahmen der automatischen Sprach*generierung* relevant sind.

werden, die u.a. durch vollständige Sätze oder Verbphrasen kommuniziert werden. Einige Beispiele sind:⁶

- (6a) *Behrens verlor seine Entwürfe.*
Dies warf ihn um mehrere Wochen zurück.
- (6b) *Behrens überarbeitete seinen Vortrag.*
Dabei dachte er laut nach.
- (6c) *Behrens verlor seine Entwürfe.*
 Das Mißgeschick warf ihn um mehrere Wochen zurück.
- (6d) *Behrens überarbeitete die Entwürfe und sein Anwalt € den Finanzierungsplan.*

Zweitens besteht in den Beispielen zur Objektreferenz ausnahmslos die Relation der referentiellen *Identität*. Insbesondere mit nichtpronominalen definiten Nominalphrasen können jedoch durchaus auch *spezifische Wiederaufgriffsbeziehungen* verbunden sein, etwa Teil-Ganzes-Beziehungen und sonstige semantische Beziehungen, die durch den konzeptuellen Gehalt (vor dem Hintergrund eines angenommenen Bestands an Weltwissen, vgl. .u.) vermittelt werden.

- (7a) *Gropius plante den Erweiterungsbau der Fagus-Werke in Alfeld an der Leine.*
 Für die Fassade entwarf er eine neuartige Glas-Eisen-Konstruktion.
- (7b) *Peter Behrens besuchte seinen Kollegen in Weimar.*
Die Kinder gingen mittlerweile bereits in die Schule.

Zumindest im Lichte der Anforderungen einer verfeinerten Themenanalyse erscheint die Erschließung derartiger indirekter Relationen sinnvoll.

Die dritte Einschränkung besteht in der bislang stets textinternen, in der Terminologie von Halliday und Hasan als *endophorisch* bezeichneten Referenz ([HaHa76], S. 33 ff.). Sprachliche Ausdrücke können jedoch des Weiteren auch einen *exophorischen* Bezug auf den außertextuellen *situationalen Kontext* der Äußerung nehmen. Dies betrifft die Personalpronomen in erster und zweiter Person, aber auch bestimmte Verwendungsweisen der übrigen Typen von Pronomen, insbesondere Demonstrativpronomen:

- (8) *Dieses Bild gefällt mir sehr gut, jenes dort drüben aber nicht.*

Außerhalb eines weiteren textuellen Kontexts läßt sich diese Äußerung nur vor dem Hintergrund der situationalen Rahmenbedingungen interpretieren. Ein Spezialfall exophorischer Referenz, in dem die referenzierte Entität textintern (sozusagen *meta-textuell*) zur Verfügung steht, kann im Falle von direkter Rede gegeben sein:⁷

- (9) *“Ich denke, daß dieser Kanzler endlich zurücktreten sollte.”,
 sagte der Politiker der Grünen.*

⁶Im folgenden wird auf die Angabe von Referenzindexen bis auf Weiteres verzichtet, da die anaphorischen Bezüge i.d.R. eindeutig bestimmt sind. In Rahmen von Kapitel 7 werden Referenzindexe in einem theoretisch motiviertem Zusammenhang erneut ins Spiel kommen.

⁷Hierbei ist zu beachten, daß der Ausdruck “*dieser Kanzler*” (zumindest im hier angenommenen Kontext) *keine* situationsdeiktische Bezugnahme darstellt. Die Interpretation erfolgt vielmehr auf der Basis von Weltwissen.

Eine weitere Form referentieller Bezugnahme, die zum Teil pronominal, manchmal aber auch definit-nominal realisiert wird, beinhaltet eine *Summation* mehrerer Antezedenten:

- (10a) *Peter Behrens war mit Walter Gropius befreundet.
Sie sind berühmt für ihre modernen Entwürfe.*
 (10b) *Peter Behrens war mit Walter Gropius befreundet.
Die (beiden) Architekten sind berühmt für ihre modernen Entwürfe.*

Eine zusätzliche Komplikation ergibt sich durch die Existenz bestimmter *nichtreferentieller* Vorkommen des Pronomens “*es*”. Drei entsprechende Verwendungsweisen sind zu unterscheiden - Korrelat, thematisches bzw. expletives “*es*”:⁸

- (11a) *Es freute Behrens, daß seine Entwürfe wiedergefunden wurden.*
 (11b) *Es sprachen Peter Behrens und Walter Gropius.*
 (11c) *Es ist dreiviertel Acht.*

Als problematisch erweisen sich ferner die Heuristiken, die betreffend die Rolle des definiten bzw. indefiniten Artikels formuliert wurden:

- (12a) *Die Mannschaften betraten den Rasen. Ein Spieler trug einen Ball.*
 (12b) *Der Wal ist das größte Säugetier.*

Während Beleg (12a) die an der sprachlichen Oberfläche bestehende Ambiguität zwischen dem unbestimmten Artikel und dem Zahladjektiv “*ein*” thematisiert, liegt in (12b) (zumindest in den meisten Kontexten) ein Beispiel einer *generischen Lesart* vor, in der nicht ein bestimmtes, bereits bekanntes Exemplar der Gattung Wal referenziert wird, sondern die Gattung an sich. Beiden Fällen ist gemein, daß die oben beschriebenen Heuristiken falsche Vorhersagen liefern, daß jedoch deren Erkennung im allgemeinen den Einsatz tiefer gehenden Weltwissens erfordert. Schließlich ergeben sich komplexe Fragestellungen in bezug auf die Abgrenzung des Koreferenzbegriffs selbst:

- (13) *Während Behrens seinen Sohn zunächst eine Tischlerlehre absolvieren ließ, schickte Gropius ihn sogleich an die Universität.*

Sicherlich wird das Pronomen “*ihn*” dahingehend interpretiert werden, daß es sich nicht auf den Sohn von Behrens, sondern auf den Sohn von Gropius bezieht. Mit anderen Worten: Nicht die Referenz, sondern lediglich das mit dem Antezedens-Ausdruck einhergehende *Konzept* wird wiederaufgegriffen; die eigentliche referentielle Interpretation geschieht mit Bezug auf den lokalen Kontext der Pronominalanapher und unter Anwendung von Wissen über Analogiebildung.⁹ In

⁸In der englischen Sprache existiert neben der Entsprechung “*it*” das (auch als Ortsanapher fungierende) Demonstrativpronomen “*there*”, das in bestimmten Kontexten nicht referentiell, sondern *fokussierend* wirkt:

There is a bird singing in the tree.

⁹Für die englische Sprache gibt Hirst das in der einschlägigen Literatur gemeinhin als “*Paycheck-Sentence*” referenzierte Beispiel

The man who gave the paycheck to his wife was wiser than the man who gave it to his mistress.

Anlehnung an die von Frege ([Freg94]) getroffene Unterscheidung zwischen Sinn und Bedeutung (Denotation) prägen Grinder und Postal die unterscheidenden Termini “*Identity of Reference Anaphora*” vs. “*Identity of Sense Anaphora*” ([GrPo71]).

6.2 Begriffsdefinitionen

Aufbauend auf den Vorbetrachtungen zum Phänomen der Referentialität soll nun eine möglichst präzise Abgrenzung der bislang intuitiv gebrauchten Begriffe *Anapher*, *Antezedens*, *anaphorischer/referentieller Wiederaufgriff* und *Referenz/Koreferenz* erarbeitet werden, die sich im Hinblick auf die im folgenden zu erörternde Problemstellung der computergestützten referentiellen Interpretation als adäquat erweist. Die Definitionen, die in der Literatur gegeben werden, sind keineswegs einheitlich, sondern abhängig von den spezifischen Zielsetzungen, die im jeweiligen Kontext verfolgt werden. Es erscheint daher angebracht, zunächst den theoretischen Status der genannten Begriffe näher herauszuarbeiten.

6.2.1 Anaphern und anaphorische Bezugnahme

Die ersten, elementaren Belege in Abschnitt 6.1.1 legen es nahe, eine *Anapher* als sprachlichen Ausdruck zu definieren, der auf einen im Text vorangehenden sprachlichen Ausdruck, das *Antezedens*, Bezug nimmt und mit diesem *koreferent* ist. Dies ist die *textlinguistische Sicht* als Kohäsionsrelation, wie sie u.a. von Halliday und Hasan ([HaHa76]) vertreten und von Hirst übernommen wird (vgl. Abschnitt 4.3.1). Die Definition von Hirst eignet sich als Ausgangspunkt der weiteren Diskussion ([Hirs81], S. 4, Hervorhebungen im Original):

“Anaphora is the device of making in discourse an abbreviated reference to some entity (or entities) in the expectation that the perceiver of the discourse will be able to disabbreviate the reference and thereby determine the identity of the entity. The reference is called an anaphor, and the entity to which it refers is its referent or antecedent. A reference and its referent are said to be coreferential.

Hierbei wird unterschieden zwischen Phänomen (“*anaphora*”) und phänomeninduzierendem sprachlichem Ausdruck (“*anaphor*”).¹⁰

Diese Definition ist in mehrfacher Hinsicht eingeschränkt. Einerseits hat die Bezugnahme durch einen in informationstheoretischem Sinne *abkürzenden* Ausdruck zu erfolgen (vgl. ibd., S. 4f.). Wie Hirst selbst als diskussionswürdig herausstellt, werden hierdurch bei Weitem nicht alle Fälle kohäsionsinduzierender Ausdrücke subsumiert (vgl. z.B. Beleg (3), in dem die definite NP “*Der große Architekt*” zusätzliche Information über die als “*Peter Behrens*” eingeführte Entität beisteuert). Zweitens wird ausschließlich der Fall der Identitätsrelation abgedeckt; einige der

¹⁰Die Verwendung der Begriffe “*anaphora*” und “*anaphor*” wird jedoch in der Literatur keineswegs konsistent gehandhabt. So hat sich im Englischen der Terminus “*anaphora resolution*” eingebürgert, obwohl ja eigentlich nicht das Phänomen, sondern die phänomeninduzierende Instanz, d.h. die *anaphor*, resoliert wird.

in Abschnitt 6.1.2 besprochenen komplexeren Wiederaufgriffsmuster (Belege (7), Teil-Ganzes-Beziehungen; (10), Summation; (13), *Identity of Sense*) sind durch den obigen Abgrenzungsversuch also noch nicht erfaßt. Eine entsprechend allgemeiner gehaltene Beschreibung des Begriffs *Anapher* findet sich bei Bußmann ([Buß90], S. 82):

“[...] (1) *Sprachliche Einheit, die zu einer anderen sprachlichen Einheit (→ Antezedens) im vorangehenden Kontext in einer anaphorischen Beziehung steht, d.h. deren → Referenz nur durch Bezug auf ein Antezedens bestimmbar ist [...]. Hat die A. gleiche Referenz wie das Antezedens, spricht man von → Koreferenz. [...]*”

Der Begriff der anaphorischen Beziehung wird hier nicht unmittelbar, sondern nur indirekt per Verweis auf den Stichpunkt *Referenz* eingegrenzt. Wie im nächsten Abschnitt offengelegt wird, bleiben jedoch noch Fragen offen, denn die referenztheoretischen Ansätze der Sprachphilosophie haben keineswegs einen einheitlichen Referenzbegriff hervorgebracht. Darüberhinaus ist zu berücksichtigen, daß es neben Koreferenz den Typus der “*Identity of Sense*”-Bezugnahme gibt. Ein weiterer Unterschied besteht im wortwörtlichen Ausschluß exophorischer Referenzen - das Antezedens ist als *sprachliche* Entität eingegrenzt; in der Arbeit von Hirst hingegen können Anaphern außertextuellen Bezug haben ([Hirs81], S. 7).

Aufbauend auf diesen Betrachtungen läßt sich nunmehr eine Definition synthetisieren, die den Anforderungen der vorliegenden Arbeit gerecht wird. Da das Ziel in der algorithmischen Analyse von Texten besteht, die an keinen bestimmten “Lese-Kontext” gebunden sind, jedoch ein Vorwissen supponierten, sollen die Betrachtungen auf sprachliche Ausdrücke mit endophorischer Referenz eingegrenzt werden. Die Charakterisierung von Bußmann erweist sich somit als der geeignetere Ausgangspunkt für die folgende Festlegung.

Definition 6.1 (Anapher, Antezedens, anaphorischer Wiederaufgriff) *Eine Anapher ist ein sprachlicher Ausdruck (evtl. eine Ellipse), der als inhaltliche Entität zu interpretieren ist, die bereits durch (mindestens) ein weiteres sprachliches Element, das Antezedens, entweder direkt oder indirekt in den Diskurs eingeführt wurde.*

Auf der Ebene des sprachlichen Ausdrucks - zwischen Anapher und Antezedens - besteht eine Kohäsionsrelation, die als anaphorischer Wiederaufgriff bezeichnet wird.

Referenz-Identität bzw. Koreferenz stellt nur einen Spezialfall anaphorischen Wiederaufgriffs dar. Die Relation kann auch indirekt sein (konzeptueller Aufgriff des Sinns, Summation, Teil-Ganzes usw.) und sich ggf. erst unter Rekurs auf Weltwissen erschließen.

Zu klären bleibt der theoretische Status der Begriffe “*Referenz-Identität*” und “*Koreferenz*”.

6.2.2 Referenz, Koreferenz und Kospezifikation

Die Definition von Hirst basiert auf einem textlinguistischen Begriff der Koreferenz, der als Relation über der Menge aller sprachlichen Ausdrücke eines Textes definiert ist. Diese Relation besteht nicht unmittelbar; gemäß referenzsemantischer Sicht ist sie als *vermittelt* anzusehen. Sprachliche Zeichen sind Stellvertreter für die Objekte einer außersprachlichen Realität (“Welt”), der der Text entspringt bzw. die der Text beschreibt. Haben zwei Zeichen (sprachliche

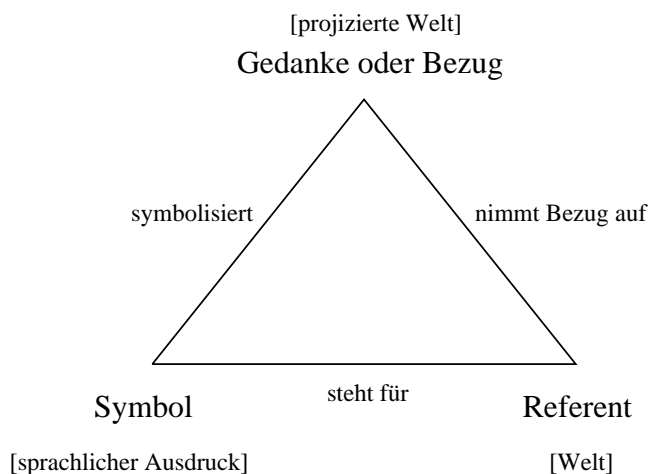


Abbildung 6.1: Semiotisches Dreieck

Ausdrücke) identische *Referenz*, so besteht zwischen ihnen die Relation der *Ko-Referenz*. Die theoretische Klärung des Begriff der Referenz, der den Bezug zur außersprachlichen Wirklichkeit bezeichnet, ist mit zahlreichen Problemen verbunden. Beispielsweise kann das durch ein sprachliches Element bezeichnete Objekt der Phantasie des Textautors entspringen und somit ohne Gegenstück in der Realität sein. Insofern das fiktive Objekt auch dem durchschnittlichen Leser bekannt ist, sollte dieser Entität dennoch eine Bedeutung zugewiesen werden.¹¹ Ohne auf die vielfältigen sprachphilosophischen Vorschläge zur Behebung der mit diesem naiven denotationstheoretischen Ansatz verbundenen Mängel genauer einzugehen, soll eine Lösungsweg gewählt werden, der sich im Hinblick auf den Skopus der vorliegenden Arbeit als hinreichend erweisen wird. Die Idee besteht in der Annahme einer *projizierten Welt*, die der realen Welt vorgeschaltet ist, und relativ zu der die Interpretation der sprachlichen Zeichen vonstatten geht (vgl. Habel, [Habe86], S. 26 ff.). Der Begriff der Referenz wird somit ersetzt durch eine Bezugnahme auf das Abbild der Welt im *kognitiven Substrat* von Autor/Leser bzw. sprachverarbeitendem Computersystem. Die hier thematisierten Beziehungen zwischen den drei Ebenen reale Welt, projizierte Welt und sprachliche Zeichen lassen sich im semiotischen Dreieck von Ogden und Richards lokalisieren ([OgRi74]; vgl. Abbildung 6.1).

Eine vergleichbare Sicht wird in den Arbeiten von Sidner angenommen; an Stelle der Begriffe Referenz und Koreferenz treten die Termini *Spezifikation* bzw. *Kospezifikation* ([Sidn83], S. 267 ff.).¹² Die Relation zwischen Anapher und Antezedens wird somit nicht mehr über die gemeinsame Referenzierung ein und desselben Objekts der realen Welt, sondern per *Spezifikation* ein und desselben *Spezifikats*¹³ im Welt-Modell des sprachverarbeitenden Systems (Mensch oder Maschine) konstituiert.

¹¹Beispiele für diesen Fall sind allgegenwärtig, z.B.:

Sherlock Holmes war ein berühmter Detektiv.

¹²Die englischen Originalbezeichnungen sind "*specification*" und "*co-specification*".

¹³Sidner verwendet den Begriff "*specification*" primär zur Bezeichnung der kognitiven Entitäten, auf die sprachlich verwiesen wird, jedoch andererseits (als Bestandteil des Begriffs "*co-specification*") anscheinend auch zur Bezeichnung des Verweises selbst. In der vorliegenden Arbeit soll feiner unterschieden werden zwischen *Spezifikaten* als *kognitiven Entitäten* und *Spezifikationen* als die durch sprachliche Ausdrücke bewerkstelligten *Verweise* auf kognitive Entitäten.

In diesem Rahmen sind nun auch die subtileren Fälle referentiellen Wiederaufgriffs rekonstruierbar. Beispielsweise lassen sich die Instanzen indirekter Bezugnahme in den Belegen (7) dahingehend interpretieren, daß in den Weltmodellen der beteiligten sprachverarbeitenden Systeme sowohl das Wissen über die durch den Text eingeführten konkreten Objektinstanzen als auch *konzeptuelles*, nicht an konkrete Entitäten geknüpftes Wissen zur Verfügung steht. In (7a) kann so z.B. das Spezifikat des Ausdrucks "Fassade" vermöge konzeptuellen Wissens mit dem Spezifikat des Ausdrucks "Erweiterungsbau" in Verbindung gebracht werden; Sidner prägt hierfür den Begriff der *abgeleiteten (impliziten) Spezifikation*¹⁴. Auf vergleichbare Weise lassen sich die übrigen Fälle rekonstruieren. Dies gilt insbesondere auch für den Fall der in Beleg (13) illustrierten *Identity of Sense Anaphora*; denn der Annahme, daß das mentale oder symbol-strukturelle Abbild der Welt neben der Spezifikation konkreter Objekte auch die Spezifikation des auf konzeptueller Ebene verankerten *Sinns* sprachlicher Ausdrücke (ggf. gefolgt von einer kontextuell determinierten Re-Instanzierung qua Analogieschluß) erlaubt, steht nichts im Wege. Auf der Grundlage dieser Betrachtungen sind nunmehr die Voraussetzungen gegeben, die Relation zwischen Anapher und Antezedens in adäquater Weise zu charakterisieren:

Anapher und Antezedens sind sprachliche Ausdrücke, die vermöge ihrer Spezifikate in der projizierten Welt in Beziehung stehen. Die Beziehung kann direkt sein (identische Objekte, d.h. Ko-Spezifikation) oder auch indirekt (abgeleitete Spezifikation), wobei sowohl Objekt-Wissen als auch konzeptuelles Wissen hineinspielt.

Unter einer dementsprechenden Ersetzung der Begriffe inhaltliche Entität, Referenz-Identität und Koreferenz gelangt man zu einer befriedigenden Version von Definition 6.1.

Abschließend sollen einige z.T. bereits intuitiv gebrauchte Begriffe abgegrenzt werden. Die erste Festlegung betrifft die Ebene der sprachlichen Symbole:

Definition 6.2 (Vorkommen (Okkurrenz)) *Ein Vorkommen (auch Okkurrenz genannt) ist ein sprachlicher Ausdruck, der "spezifiziert", d.h. der ein Spezifikat in der projizierten Welt symbolisiert.*

An die Stelle des in der realen Welt angesiedelten Referenzobjekts tritt der Terminus des in der projizierten Welt lokalisierten *Diskursreferenten*:

Definition 6.3 (Diskursreferent) *Der Begriff Diskursreferent ist ein Synonym für das Spezifikat, das ein sprachlicher Ausdruck eines bestimmten Textes (Diskurses) in der projizierten Welt spezifiziert (symbolisiert).*

Insgesamt ergibt sich folgendes Bild: Anapher und Antezedens sind Vorkommen, die denselben Diskursreferenten spezifizieren (Ko-Spezifikation) oder zwischen deren Diskursreferenten in der projizierten Welt eine inhaltliche Beziehung besteht (abgeleitete Spezifikation). Vorkommen, die einen bisher noch nicht ko- oder abgeleitet spezifizierten Diskursreferenten symbolisieren, führen diesen neu in den Diskurs ein.¹⁵

¹⁴engl. "implicit specification"

¹⁵Die Unterscheidung zwischen abgeleiteter Spezifikation und Einführung eines Diskursreferenten ist abhängig vom Wissensbestand der projizierten Welt; eine formale Fundierung ist daher ohne eine weitergehende Fixierung der Wissensstrukturen des Weltmodells nicht möglich.

Im soeben skizzierten Rahmen wird offen gelassen, welcher Art die inhaltlichen Entitäten sind, die durch Vorkommen spezifiziert werden. Spezifikate können sowohl Objekte als auch sonstige, relationale Entitäten repräsentieren. Die im folgenden entwickelten Verfahren zur Resolution anaphorischer Ausdrücke werden auf die Behandlung objektspezifizierender Okkurrenzen spezialisiert sein.

Jetzt, da der theoretische Hintergrund der grundlegenden Begriffe abgeklärt ist, soll allerdings wieder zu einer *informellen Sprechweise* zurückgekehrt werden. Nicht zuletzt erscheint dies dadurch gerechtfertigt, daß den zu untersuchenden linguistischen Theorien und computerlinguistischen Verfahren oftmals keine konsistenten bzw. sprachtheoretisch adäquaten Bezeichnungen zugrunde liegen. Wenn also im folgenden von referentiellm Bezug, Referenzobjekten, Koreferenz usw. die Rede ist, so ist dies *cum grano salis* zu lesen, d.h. eine Ersetzung durch die repräsentationalen Entsprechungen in der projizierten Welt vorzunehmen.

6.3 Anapherninterpretation als sprachverarbeitender Prozeß

In der bisherigen Diskussion wurden referentielle Phänomene aus einer rein statischen Perspektive heraus betrachtet. Die nachfolgenden Kapitel drehen sich jedoch um die dynamischen Gesichtspunkte einer algorithmischen referentiellen Interpretation von Texten. Es erscheinen daher abschließend einige Bemerkungen zum theoretischen Stellenwert dieses Teilgebiets sprachverarbeitender *Prozesse* angebracht.

Aus den obigen Betrachtungen wird deutlich, daß es sich bei sprachverarbeitenden Systemen (hier im Speziellen: Systeme zur referentiellen Interpretation von Texten) sowohl um *wissensverarbeitende* als auch um *wissensbasierte* Prozesse handelt: Sprachverarbeitende Systeme

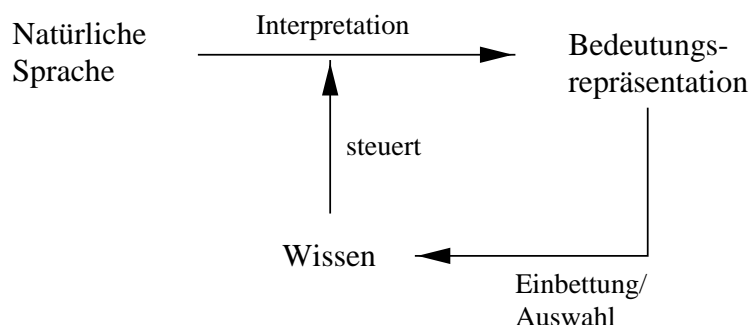
- *basieren* i.a. auf außersprachlichem Wissen über die Welt, und
- *produzieren* Wissen qua Erschließung der Bedeutung des zu analysierenden sprachlichen Inputs.

Die in Abschnitt 6.1 diskutierten Beispiele illustrieren diese beiden Facetten für die Aufgabenstellung der referentiellen Interpretation. Einerseits fließt Weltwissen ein, über das z.B. Fälle abgeleiteter Spezifikationen resolviert werden.¹⁶ Andererseits liefert eine schrittweise Analyse der referentiellen Bedeutung selbst Bedeutungsstrukturen (u.a. die durch die sprachlichen Ausdrücke symbolisierten Spezifikate), die mit dem ursprünglichen Weltmodell in Beziehung gesetzt werden, es eventuell modifizieren, und somit einen neuen Rahmen abstecken, in dem die Interpretation des noch nicht analysierten Teils der Eingabe vonstatten geht.¹⁷

Die Entsprechung dieses generischen Schemas sprachverarbeitender Prozesse mit der hermeneutischen Spirale - nicht jedoch mit einem hermeneutischen *Circulus vitiosus* - ist offensichtlich. Habel skizziert diese Interdependenz zwischen Wissen (Weltmodell) und Bedeutungserschließung in einem "*Grundschema sprachverarbeitender Systeme*" ([Habe86], S. 23):

¹⁶Weltwissen spielt natürlich auch eine wichtige Rolle im Rahmen der Interpretation von Pronominal-Anaphern.

¹⁷Die Annahme, daß die Analyse des sprachlichen Inputs *schrittweise* (i.d.R. satzweise) abläuft, läßt sich text- und diskurstheoretisch sowie insbesondere dadurch begründen, daß Versuche, einen Text von hinten nach vorne zu lesen, im allgemeinen erfolglos verlaufen werden.



Mit anderen Worten: Das Weltmodell, d.h. die projizierte Welt liegt im Schnittpunkt von Wissens- und Bedeutungsstrukturen; aus seiner kognitiven und sprachphilosophischen Perspektive plädiert Habel sogar dafür (ibid., S. 29)

“... daß Repräsentationen für Bedeutungen und Wissensrepräsentationen als zwei Seiten derselben Medaille angesehen werden.”

Dieser Punkt soll nicht weiter vertieft werden, da die damit verbundenen Fragestellungen außerhalb des Skopus der vorliegenden Arbeit lokalisiert sind. Prinzipiell fallen sprachverarbeitende Prozesse in die Schnittmenge der Fragestellungen folgender Disziplinen:

- *Kognitionswissenschaften*,¹⁸
- *Künstliche Intelligenz* (als Anwendungsdisziplin der Informatik).

Auch wenn an einem zentralen Punkt auf eine linguistische Theorie rekurriert werden wird, die auf dem Anspruch beruht, allgemeine und universelle Prinzipien der Sprachkompetenz zu formalisieren, so wird im folgenden nichtsdestotrotz die primäre Zielsetzung verfolgt, ein anwendungstaugliches, robustes Sprachverarbeitungssystem mit möglichst guter *Performanz* zu konzipieren; die Gestaltung des internen Aufbaus des Systems ordnet sich diesem Anspruch unter. Es wird eine *ergebnisorientierte Perspektive* eingenommen; die durchzuführenden Untersuchungen sind daher in erster Linie dem Gebiet der Künstlichen Intelligenz zuzuordnen. Fragen betreffend die *kognitionstheoretisch adäquate* Repräsentation von Wissens- und Bedeutungsstrukturen werden folglich nicht näher erörtert. Es sei diesbezüglich auf die Arbeit Habels

¹⁸Aus kognitionswissenschaftlicher Perspektive bieten sprachverarbeitende Prozesse die Möglichkeit einer empirischen Überprüfung von Theorien auf dem Computer: Führt die Implementierung einer Theorie zu einer Interpretationsleistung, die mit der menschlichen Performanz vergleichbar ist, so ist daraus zu folgern, daß die der Theorie zugrundeliegenden Prinzipien auch dem menschlichen sprachverarbeitenden Prozeß zugrunde liegt. Dieses Vorgehen basiert auf einem Prinzip von Miller ([Mill64], S. 108, zitiert in der Übersetzung Habels, [Habe86], S. 13):

“Wenn eine Maschine M eines I-Verhaltens (‘intelligenten’ Verhaltens) fähig ist, dann unterliegt M einem System P von Prinzipien, dem jede Maschine M’, die des I-Verhaltens fähig ist, ebenfalls unterliegt.”

Diese *Prinzipien* (und nicht das anwendungstaugliche, robuste Textanalyse-System) stellen das Erkenntnisinteresse der Kognitionswissenschaften dar.

verwiesen, in der diese Thematik aus der Sicht der Kognitionswissenschaften und mit speziellem Bezug auf die Repräsentation *referentiellen* Wissens diskutiert wird.¹⁹ In bestimmter Hinsicht ist die vorliegende Arbeit somit komplementär zu der von Habel - sie setzt dort an, wo dessen Untersuchungen enden: bei der *Implementierung referentieller Prozesse*, die bei Habel allenfalls fragmentarisch sowie ohne Anspruch auf Vollständigkeit und Anwendungstauglichkeit spezifiziert werden.

6.4 Eingrenzung des Untersuchungsgegenstands

Bereits in die Definition 6.1 des Anaphernbegriffs floß eine erste, einfach nachvollziehbare Restriktion ein: Ausschließlich anaphorische Ausdrücke mit *endophorischem* Bezug finden Berücksichtigung. Selbst in diesem intratextuellen Rahmen ist die Vielfalt immer noch enorm.

Aus der Sicht der Anforderungen einer verfeinerten Computergestützten Inhaltsanalyse kommt dem Problem der Interpretation von *Pronomen* zentrale Relevanz zu. Diese Aussage kann präzisiert werden: Die mit Abstand größte Anzahl von Instanzen pronominaler Anaphern ist dem Gebiet der *Objekt*-Spezifikation zuzuordnen. Fälle pronominalen Aufgriffs relationaler Entitäten sind vergleichsweise selten.²⁰

Für die elementare Form der Bezugnahme, der *Ko*-Spezifikation, läßt sich jedoch keineswegs behaupten, daß sie häufiger vorkommt als die komplizierteren Spielarten abgeleiteter Spezifikationen. Speziell die nichtpronominalen Nomen greifen in einer Vielzahl von Fällen nicht die durch das Antezedens spezifizierte Entität, sondern - vermittelt durch ihren konzeptuellen Gehalt - ein mit dieser in einem allgemeineren inhaltlichen Zusammenhang stehendes Objekt auf (vgl. Abschnitt 6.1.2, Belege (7a), (7b)).²¹ Die Pronomen hingegen *ko*-spezifizieren mit ihrem Antezedens in der Mehrzahl der Fälle. Instanzen pronominaler Summation (Beleg (10a), die ebenfalls mit einer abgeleiteten Spezifikation einhergehen, sind vergleichsweise selten.

Da eine formale Evaluation durchgeführt werden soll, sind nicht alleine den Desiderata der Anwendungs-Perspektive, sondern darüberhinaus methodischen Gesichtspunkten Rechnung zu tragen. Die Schlüssel, d.h. die Vorgaben, relativ zu denen die Interpretationsleistung der Algorithmen bewertet werden sollen, sind das Ergebnis intellektueller Annotation von Evaluationskorpora durch den Menschen. Um die Aussagefähigkeit der Evaluationsergebnisse zu gewährleisten, ist sicherzustellen, daß die Schlüssel mit guter Intercoderreliabilität reproduzierbar sind (vgl. Abschnitt 5.4.2). Folglich sollte die Aufgaben-Definition, auf deren Grundlage die Annotation mit referentieller Information erfolgt, unter Wahrung der Anwendungsrelevanz der zu bewertenden Interpretationsleistung so überschaubar wie möglich gestaltet werden. Unter diesen Gesichtspunkten erscheinen folgende Einschränkungen gerechtfertigt:

1. Es werden ausschließlich *Objekt*-Spezifikationen, induziert durch NP, betrachtet. (Dies beinhaltet die für die CGI primär wichtigen Instanzen von *Pronominalisierung*.)

¹⁹Ergebnis dieser Untersuchung ist die Repräsentationssprache "Referentielle Netze", die den aus referenztheoretischer Sicht zu stellenden Anforderungen genügt.

²⁰Diese Aussage beruht auf den Erfahrungen aus der formalen Evaluation (vgl. Kap. 11).

²¹Aussagefähige Zahlen zu gewinnen, ist insofern problematisch, als die Entscheidung, welche definite (nicht-pronominalen) NP *abgeleitet* spezifiziert, nur unter Annahmen über das zugrundeliegende Weltmodell getroffen werden kann.

2. Ausschließlich Instanzen von *Ko*-Spezifikation (unter Einschluß der in Beleg (13) beschriebenen “*Identity of Sense*”-Identität) werden betrachtet. (Eine Einbeziehung komplexerer Relationen würde den Annotationsvorgang erheblich verkomplizieren und die Intercoderreliabilität entsprechend beeinträchtigen. Zudem fallen die meisten Fälle von Pronominalisierung unter die Identitätsrelation.)
3. Elliptische, d.h. an der sprachlichen Oberfläche leere Vorkommen finden keine Berücksichtigung. (Andernfalls hätte die Annotation auf der Basis einer Repräsentation zu erfolgen, in der bereits vom Ausgangstext abstrahiert wird.)

Bis auf die abweichende Rahmenzielsetzung, die Verfeinerung der Computergestützten Inhaltsanalyse, decken sich die Eckdaten dieses Szenarios mit denen der MUC-“*Coreference*”-Aufgabe. Um die Vergleichbarkeit mit den MUC-Ergebnissen zu gewährleisten, soll die Annotation der Evaluationskorpora in weitgehender Entsprechung zu den inhaltlichen Vorgaben der “*Coreference Task Definition*” erfolgen ([CoTD97]). Eine präzisere Beschreibung der Aufgabenstellung wird in Abschnitt 11.1.3 gegeben.

Die Erfahrungen aus MUC-6 rechtfertigen die Einschränkung der Untersuchungen auf den Relationstyp *Ko*-Spezifikation. Bereits in diesem relativ restringierten Evaluationsszenario kommen subtile referentielle Phänomene zum Tragen, die den Prozeß der intellektuellen Annotation des Evaluationskorpus zu einer vergleichsweise schwierigen Aufgabe machen. Die formale Definition [CoTD97] umfaßt bereits 14 Seiten, ohne den Anspruch zu erheben, eine Entscheidungsgrundlage für auch nur annähernd alle Fälle darzustellen. Letztlich wird auf das intuitive Sprachverständnis der Annoteurin bzw. des Annoteurs rekuriert (ibid., S. 14):

“The coreference judgements should be based on the intelligent reader’s knowledge of the world resulting from his or her best understanding of the text. It should not be based on a theory of the structure of the text, or on a linguistic theory of how NPs are resolved, or on estimates of what the typical NLP system could do.”

Daß die abgespeckte Aufgabendefinition immer noch entsprechende Freiheitsgrade aufweist bzw. eine nicht zu unterschätzende Komplexität besitzt, läßt sich indirekt anhand der im Rahmen von MUC-6 ermittelten Zahlen zur Intercoderreliabilität der “*Coreference*”-Annotationen ablesen: 80 % Recall bei 82 % Precision (vgl. Sundheim: [Sund96], S. 20) - Werte, die erheblich unter der im Kriterienkatalog der Task-Definition genannten Zielvorgabe von 95 % liegen. Die Berücksichtigung weiterer Relationentypen würde diese Problematik noch verschärfen.

Letztlich tritt an dieser Stelle ein Dilemma zutage, das sich auf zwei konkurrierende Ansprüche der Evaluations-Methodik begründet. Greift man bei der Wahl der Annoteur auf *durchschnittliche Sprachbenutzer* zurück, so ist zur Vermittlung der Aufgabendefinition ein i.a. erheblicher Schulungsaufwand zu treiben, um eine hinreichende Intercoderreliabilität zu erzielen. Dieser Aufwand ließe sich unter Rückgriff auf linguistisch geschultes Personal reduzieren, das eine höhere Reliabilität auf der Grundlage eines verinnerlichten, evtl. jedoch auch in *unerwünschter* Weise normierenden theoretischen Vorwissens erzielt. Folglich gibt es gute Gründe für möglichst einfache und klar strukturierte Aufgabenbeschreibungen, die auf der Grundlage intuitiven Sprachverständnisses zugänglich sind.

Kapitel 7

Bausteine zur Operationalisierung der referentiellen Interpretation

Im vorangegangenen Kapitel wurde das Phänomen der Referentialität aus einer neutralen, d.h. linguistischen bzw. sprachphilosophischen Perspektive betrachtet; die Darstellung der Anapherninterpretation als sprachverarbeitender Prozeß bewegte sich auf einem sehr abstrakten Niveau. Diese Ebene grundsätzlicher Erwägungen soll nun zugunsten einer prozeßorientierten Sichtweise aufgegeben werden, in deren Rahmen die Möglichkeiten einer algorithmischen Lösung des Problems der Anapherninterpretation untersucht werden. Es geht somit darum, die *innere Struktur anapherninterpretierender Prozesse* näher einzugrenzen, die bisher lediglich durch das in Abschnitt 6.3 beschriebene *Grundschemata sprachverarbeitender Prozesse* charakterisiert sind. Unter der vorläufigen Annahme, daß die Teilmenge vorkommen-induzierender sprachlicher Ausdrücke eines referentiell zu interpretierenden Textes bereits bestimmt und ferner bekannt ist, *welche* der Vorkommen anaphorisch Bezug nehmen, stellt sich das Problem der Interpretation von Anaphern aus algorithmischer Sicht in erster Linie als *Entscheidungsproblem* dar:

Gegeben ist eine Menge V von Vorkommen und eine Teilmenge $A \subseteq V$ anaphorischer Vorkommen. Für alle Anaphern $a \in A$ ist unter Rekurs auf verfügbare Wissensquellen ein kospezifizierendes Antezedens $v_a \in V$ auszuwählen.

Ein erstes Beispiel soll die Grundzüge dieser Rahmenprozedur illustrieren:

- (1a) *Peter Behrens besuchte Walter Gropius in Weimar.
Er informierte ihn über eine Künstlergruppe in Berlin.*

Für das sprachverarbeitende System stellt sich die Aufgabe, den pronominalen Vorkommen “*Er*” und “*ihn*” je ein geeignetes Antezedens zuzuordnen. Unter der Annahme, daß auch die übrigen Nominalphrasen als Vorkommens-Induktoren erkannt wurden, hat die Auswahl aus folgender Menge von *Antezedens-Kandidaten* zu erfolgen: “*Peter Behrens*”, “*Walter Gropius*”, “*Weimar*”, “*Künstlergruppe*” und “*Berlin*”. Die Kandidaten “*Weimar*”, “*Künstlergruppe*” und

“*Berlin*” können jedoch aus grammatischen Gründen ausgeschlossen werden, da die Pronomen “*Er*” und “*ihn*” (beinahe) ausschließlich mit Antezedenten kospezifizieren, die die morphologischen Merkmale Numerus *Singular* und Genus *Maskulinum* aufweisen. Zur eigentlichen Auswahl verbleiben die Kandidaten “*Peter Behrens*” und “*Walter Gropius*”.

An dieser Stelle zeigt sich nun erneut die potentielle Komplexität des Problems der Anapherinterpretation. Wie soll *algorithmisch* entschieden werden, welches Pronomen welchen der verbleibenden Antezedens-Kandidaten wiederaufgreift? Klar (im Sinne von algorithmisch-formal erfaßbar) ist a priori nur, daß nicht beide Pronomen mit demselben Antezedens ko-spezifizieren. (Dieser Punkt wird weiter unten in größerer Ausführlichkeit diskutiert.) Für die verbleibenden Alternativen hingegen scheint es kein hartes Entscheidungskriterium zu geben: Was (zumindest) für den mit den grundlegenden Kenntnissen über die Vita von Peter Behrens bewehrten Leser auf der Hand liegt - Behrens wirkte für längere Zeit in Berlin -, mag für ein sprachverarbeitendes System, das über das entsprechende Vorwissen nicht verfügt, alles andere als klar sein. Ein möglicher Ansatz, die Reproduktion der Oberflächenanordnung qua Antezedenszuordnung (d.h. der relativ zur Anordnung der Ausdrücke an der sprachlichen Oberfläche “*erstere*” Kandidat wird dem entsprechend “*ersteren*” Pronomen zugeordnet usw.), stellt lediglich eine einfache Heuristik dar, die in vielen Fällen zu falschen Entscheidungen führt, wie die folgenden Variationen des Ausgangsbeispiels aufzeigen:

- (1b) ...
Er informierte *ihn* über eine Künstlergruppe am Bauhaus.

In diesem Fall könnte Weltwissen über die Beziehungen zwischen den mit den Ausdrücken “*Walter Gropius*”, “*Weimar*” und “*Bauhaus*” bezeichneten inhaltlichen Entitäten - Gropius lehrte am Bauhaus zu Weimar - eine umgekehrte Zuordnung nahelegen. Diese Präferenz kann sogar noch wesentlich stärker ausgeprägt sein:

- (2) *Behrens* übergab *Gropius* die halbfertigen Baupläne.
Er sollte *ihn* bei der weiteren Ausarbeitung unterstützen.

Anhand der ersten Beispiele wird bereits deutlich, daß die Aufgabenstellung der algorithmischen Interpretation anaphorischer Ausdrücke ein potentiell schwieriges Problem darstellt, zu dessen Lösung auf eine Vielzahl unterschiedlicher Faktoren bzw. Wissensquellen zurückzugreifen ist. Carbonell und Brown schlagen deshalb die Verfolgung eines “*multi-strategy approach*” vor - ein Paradigma, das sich in der Ausgestaltung gängiger Anapherinterpretations-Algorithmen durchgängig widerspiegelt ([CaBr88]). Die Kernhypothese besteht darin, daß (ibd., S. 96)

“... *anaphora resolution may be best accomplished through the combination of a set of strategies, rather than by a single monolithic method.*”

Mit Blick auf die obigen Beispiele läßt sich diese Aussage präzisieren. Offenbar unterscheiden sich die Faktoren im Grad der Stringenz, mit der sie wirken. Einige wirken obligatorisch, d.h. als *Restriktionen*; anderen wiederum scheint der Status eines *Präferenz*-Kriteriums zuzukommen, das nur eine Tendenzaussage über die *Plausibilität* von Kandidaten macht. Diese Differenzierung, die auf Carbonell und Brown zurückgeht, wird sich insbesondere im Rahmen der

Strategie-Integration in einen vollständigen Algorithmus zur Anapherninterpretation als relevant erweisen.

Im folgenden sollen die einzelnen Faktoren und deren theoretische Grundlagen näher betrachtet werden. Das Ziel besteht darin, eine Teilmenge von Strategien zu identifizieren, die algorithmisierbar sind, um zu einer Menge von *Bausteinen* zu gelangen, auf deren Grundlage die obige, abstrakte Spezifikation eines Algorithmus zur Anapherninterpretation konkretisiert werden kann.¹ In bezug auf die unterschiedlichen Faktoren, d.h. Restriktionen und Präferenzkriterien, sollen folgende Fragestellungen untersucht werden:

1. Auf welchen *Wissensquellen* baut die Strategie auf?
2. Ist die Strategie *algorithmisch* umsetzbar? Welche Voraussetzungen müssen erfüllt sein?
3. Wie *relevant* ist die Strategie für das Problem der Anapherninterpretation auf anwendungstypischem, unrestringiertem Text?

Wichtig ist somit neben der Algorithmisierbarkeit die in Punkt 3 formulierte *Relevanz-Zielsetzung*: Eine Theorie, die ausschließlich unter Rekurs auf Beispieldiskurse entwickelt wurde, ist u.U. nur von marginaler Bedeutung, falls die modellierten Phänomene Spezialfälle darstellen, die in Anwendungstexten selten oder sogar überhaupt nicht vorzufinden sind. Geht es um mehr als um ein theoretisches Verfahrensmodell, d.h. um konkrete Implementierungen, so sind solche Strategien natürlich irrelevant, selbst wenn sie sich einfach algorithmisieren lassen.

7.1 Restriktionen

Vorab sei bemerkt, daß die Dichotomisierung Restriktion vs. Präferenz in bezug auf Strategien zur Anapherninterpretation keineswegs unkontrovers ist (vgl. z.B. Mitkov, [Mitk97]). Für die meisten der in diesem Abschnitt beschriebenen "*stringenten*" Strategien lassen sich irgendwelche Ausnahmen finden, so daß es vielleicht adäquater wäre, von *Default-Regeln* anstelle von Restriktionen zu sprechen. Letzterer Begriff hat sich jedoch allgemein eingebürgert; von dieser Bezeichnungspraxis soll in der vorliegenden Arbeit nicht abgewichen werden.

7.1.1 Morphologische Kongruenz

Grundlegende Beispiele

Anhand von Beispiel (1a) wurde bereits deutlich, daß Pronomen offenbar nur mit solchen Antezedenten kospezifizieren können, mit denen sie in bezug auf die grammatischen Merkmale *Numerus*, *Genus* und *Person* übereinstimmen. Da sich diese Attribute zumindest partiell in der Flexion bzw. als lexikalische Merkmale der Stammform niederschlagen und damit durch eine

¹Diesen Bausteinen kommt ein Stellenwert zu, der vergleichbar ist mit dem der von Habel identifizierten *Referentiellen Prozesse* ([Habe86], S. 166 ff.). Ein Unterschied besteht in der größeren Aufgabennähe der hier diskutierten Strategien.

morphologische Analyse erschlossen werden können, spricht man diesbezüglich auch von der Bedingung der *morphologischen Kongruenz*.

Folgender Beleg illustriert das Wirken dieser Restriktion in größerer Ausführlichkeit:

- (3) *Die Ehefrau begrüßte den Architekten auf der Terrasse.
Er überreichte ihr einen Strauß Blumen.*

Zwar verkörpert eine Restriktion ihrem Anspruch gemäß kein Entscheidungskriterium zur Wahl eines bestimmten Kandidaten; jedoch zeigt dieses Beispiel, daß es Fälle geben kann, in denen alle Kandidaten bis auf einen eliminiert werden: Aus Kongruenzgründen verbleibt für das Pronomen “*Er*” nur die Antezedensoption “*den Architekten*”. Die Anapher “*ihr*” hingegen verbleibt mehrdeutig. Erst unter Anwendung weiterer Kriterien, die augenscheinlich auf außersprachlichem Wissen basieren, wird eine Entscheidung zugunsten von “*Die Ehefrau*” und zuungunsten von “*Terrasse*” möglich werden.

Nicht in allen Fällen sind die genannten morphologischen Merkmale durch eine rein morphologische Analyse bestimmbar: In dem Beleg

- (4) *Sie bot dem Architekten einen Drink an.*

ist es erst auf der Grundlage einer Analyse der syntaktischen Struktur möglich, zu entscheiden, daß für das Pronomen “*Sie*” nur die Singular-Lesart verbleibt.² Jedoch zeigt folgendes Beispiel, daß die Syntax durchaus nicht immer die Disambiguierung ermöglicht:

- (5) *Der Architekt traf sie auf der Veranda.*

Die Mehrdeutigkeit kann auch qua Antezedensauswahl, d.h. im Rahmen der Anapherninterpretation beseitigt werden:

- (6) *Peter Behrens begrüßte die Ehefrau.
Der Architekt kannte sie von der Universität.*

Nach der Wahl von “*die Ehefrau*” als Antezedens des Pronomens ist der Numerus der Anapher entsprechend fixiert. Eine anaphorische Anknüpfung an “*sie*” ist nurmehr durch solche Pronomen möglich, die eine Singular-Lesart besitzen. Insofern können sich vermöge dieses Restriktionstyps *Interdependenzen* zwischen Antezedens-Entscheidungen ergeben; dieser Punkt soll jedoch erst im Rahmen der Synthese des Gesamtalgorithmus näher erörtert werden.

Feinheiten

In Sprachen, in denen das grammatische Geschlecht (der *Genus*) eines Ausdrucks vom natürlichen Geschlecht (dem *Sexus*) des zugehörigen Spezifikats differieren kann, sind Abweichungen

²Dies wird ermöglicht durch eine Kongruenzbedingung, die zwischen Subjekt und Vollverb zum Tragen kommt: Da das Vollverb in 3. Person Singular steht, gilt dies notwendig auch für das zugehörige Subjekt “*Sie*”.

möglich. Für die deutsche Sprache verkörpert die Diminutivbildung von Nomen mit Personenbezug das Standardbeispiel:³

- (7) *Das Knäblein begrüßte die Pianistin.
Es/Er trug einen Matrosenanzug.*

Auch dieser Fall läßt sich unter Bezug auf die Entitäten der projizierten Welt erklären. Sprachliche Ausdrücke greifen einander nie direkt auf, sondern nur vermittelt durch deren Bezüge im Weltmodell. Die Bedingung der Genus-Kongruenz “emuliert” anhand der sprachlichen Realisierung die Bedingung der Sexus-Kongruenz der entsprechenden Spezifikate. Diese Sichtweise wird durch empirische Untersuchungen unterstützt: Das Pronomen steht nur dann im abweichenden (sprachlichen) Genus, wenn die Distanz zwischen Antezedens und Pronomen klein ist (vgl. Corbett, [Corb79]). Dies läßt sich dadurch erklären, daß die Information über die sprachliche Realisierung nicht dauerhaft in das Weltmodell integriert wird und daß somit das natürliche Geschlecht die allgemeingültige Entscheidungsgrundlage darstellt.

In Sprachen, in denen der Sexus des Spezifikats stets durch den Genus des sprachlichen Ausdrucks reproduziert wird, ist das in Beleg (7) beschriebene Phänomen nicht anzutreffen. Dies gilt u.a. für die englische Sprache. Da hier jedoch auch der definite Artikel nicht den Genus widerspiegelt (es gibt nur “the”), kommt es in verstärktem Maße zu Mehrdeutigkeiten, die nur unter Berücksichtigung des *semantischen Kontexts* resolviert werden können: So hängt etwa der Genus von “*speaker*” davon ab, ob dieser Ausdruck im Sinne von Lautsprecher oder Redner bzw. Rednerin zu interpretieren ist.

Diskussion

In Abschnitt 5.2 wurde bereits näher ausgeführt, daß das Problem der morphologischen Analyse als weitestgehend gelöst angesehen werden kann. Während es dort jedoch vor allem um die Stammform-Ermittlung ging, basiert die Anwendung der Strategie *Morphologische Kongruenz* auf der Verfügbarkeit der Merkmale Person, Numerus und Genus, die durch eine rein morphologische Analyse nicht immer eindeutig bestimmt werden können. Die Antwort auf die Frage der *Algorithmisierbarkeit* ist also zu relativieren. Die Operationalisierung des eigentlichen Kongruenztests ist trivial. Die Disambiguierung der morphologischen Attribute (besonders des Genus im Englischen) kann jedoch von semantischem Kontextwissen abhängen und daher beliebig schwierig werden. Je nach Sprache werden neben den Stammformen- und Genus-Lexika zumindest prinzipiell sowohl syntaktische als auch semantische Analyseressourcen benötigt. Auf dem gegenwärtigen Stand der Technik erscheint somit nur eine partielle Algorithmisierung mit vertretbarem Aufwand möglich.

Die *Relevanz* der Kongruenzstrategie hängt in wesentlichem Maße davon ab, wie groß der Grad der Ausdifferenzierung ist, der vermöge einer Genus-Klassifikation der vorkommeninduzierenden Ausdrücke erreicht wird. In englischen Texten, die sich in erster Linie um Sachen, d.h. nicht um

³Ein weiteres Beispiel mit literarischem Einschlag findet sich bei Dürrenmatt ([Dürr58], Kapitel 6):

“Dann hörte er plötzlich vom Hause her, hinter sich, einen Schrei wie von einem Tier. Er beschleunigte seinen Schritt und wußte nicht, ob es der Mann oder die Frau war, das so weinte.”

Personen drehen, leistet die Bedingung der Genus-Kongruenz keinen Beitrag, da hier allen Vorkommen das grammatische Geschlecht Neutrum zugeordnet ist. Für die deutsche Sprache liegen die Verhältnisse besser, da auch Nicht-Personen nach Genus differenziert werden. Der durch Beleg (7) beschriebene Sonderfall ist in der Praxis selten.

Wichtig für die Nützlichkeit des Kongruenzkriteriums ist darüberhinaus, inwieweit es gelingt, Mehrdeutigkeiten der morphologischen Attribute unter Rekurs auf den Kontext zu resolvieren. Entsprechend den Ausführungen im vorangegangenen Abschnitt dürfte dieser Punkt fürs Englische bedeutsamer sein als fürs Deutsche, da sich das für die Ausgestaltung von Pronominalisierungen relevante natürliche Geschlecht in nicht wenigen Fällen nicht eindeutig anhand eines Lexikoneintrags bestimmen läßt. Während die Numerus-Ambiguität etwa des Pronomens “*sie*” im Deutschen zumindest in einem Teil der Fälle syntaktisch entschieden werden kann, ist für die Resolvierung von Genus-Ambiguitäten im Englischen i.d.R. auf semantisches Kontextwissen zurückzugreifen.

Letztlich läßt sich die Frage der Relevanz also nur relativ zum Typ der zu analysierenden Texte (Vorkommens- sowie speziell Pronomendistribution) sowie unter Berücksichtigung der zur Verfügung stehenden Ressourcen (syntaktische und/oder semantische Kontextinformation) beantworten. Allerdings werden die Evaluationsergebnisse zeigen, daß das Kriterium in der Praxis von großem Nutzen ist, d.h. auch in algorithmischen Szenarien, in denen nur eine partielle Disambiguierung der morphologischen Merkmale erfolgt.

7.1.2 Syntaktisch-konfigurationale Restriktionen

Eine weitere Strategie, die sich als zentral für algorithmische Ansätze zur Anapherninterpretation erweisen wird, ist auf der Ebene der syntaktischen Oberflächenstruktur angesiedelt. Primärer Gegenstand dieses Restriktionstypus sind *satzinterne Kospezifikationsrelationen*; jedoch wird sich zeigen, daß auch Fälle der Fernwirkung auf satzübergreifende Wiederaufgriffsbeziehungen existieren.

Ausgehend von grundlegenden Beispielen soll daher ein zentraler Ansatz zur theoretischen Modellierung des Phänomens ausführlicher erörtert werden. Die entsprechenden Arbeiten entstammen dem Gebiet der Theorie der Generativen Grammatik, d.h. Modellen, denen der Anspruch zugrundeliegt, die *Kompetenz* eines Sprachbenutzers mithilfe möglichst *universeller Prinzipien* zu beschreiben. Da es hier jedoch ausschließlich um die Entwicklung eines *Performanzmodells* geht (vgl. Abschnitt 6.3), wird sich für die Algorithmisierung ein Spielraum für Vereinfachungen ergeben.

Grundlegende Beispiele

Ausgangspunkt sind einige elementare Beobachtungen, denen zufolge für unterschiedliche Typen anaphorischer Ausdrücke unterschiedliche Optionen satzinterner Kospezifikation bestehen. Für Reflexivpronomen erscheint ein (geeignetes) *lokales* Antezedens obligatorisch, hingegen scheinen nichtreflexive Personalpronomen (in erster Näherung) nur mit *nicht-lokalen* Okkurrenzen koduspezifizieren:⁴

⁴Entsprechend der allgemein üblichen Praxis kennzeichnet der Stern “*” diejenigen Belege, die (ggf. in der durch die Referenzindexe spezifizierten Lesart) gemäß intuitiver Einschätzung als *ungrammatisch* zu bewerten

- (8a) *Peter Behrens_i rasiert sich_i.*
 (8b) * *Peter Behrens_i rasiert ihn_i.*
 (8c) * *Peter Behrens_i verlangt, daß der Friseur_j sich_i rasiert.*
 (8d) *Peter Behrens_i verlangt, daß der Friseur_j ihn_i rasiert.*

Der Kontrast zwischen den Belegen (8a)/(8b) auf der einen und (8c)/(8d) auf der anderen Seite verdeutlicht, daß die Begriffe “*lokal*” und “*nicht-lokal*” nicht einfach in der Unterscheidung zwischen satzlokalem (intrasententiellem) und satzübergreifendem (intersententiellem) Bezug verankert werden können. Der Bezugsrahmen, anhand dessen sich der Lokalitätsbegriff manifestiert, scheint vielmehr die Ebene der einzelnen *Teilsätze* (Hauptsatz, Nebensätze) zu sein: Ein Reflexivum besitzt ein Antezedens innerhalb, ein nichtreflexives Pronomen hingegen (in erster Näherung) *nicht* innerhalb des Teilsatzes, in dem es vorkommt.

Für nichtpronominale anaphorische Vorkommen erweisen sich sowohl bestimmte teilsatz-lokale als auch bestimmte teilsatz-übergreifenden Kospezifikations-Relationen als unzulässig:

- (9a) * *Peter Behrens_i rasiert Peter Behrens_i.*
 (9b) * *Behrens_i informierte Gropius_j, bevor Behrens_i abreiste.*
 (9c) *Gropius_j informierte Behrens_i, bevor Behrens_i abreiste..*

Daß die Zulässigkeit oder Unzulässigkeit der Kospezifikation eines bestimmten Anaphern-Typs auch von der *strukturellen Position* des Vorkommens im Satz abhängt, d.h. eine noch näher zu charakterisierende strukturbezogene *Asymmetrie* besteht, wird anhand der unterschiedlichen Bewertung der Belege (9b) und (9c) deutlich.

Generell ist wichtig, daß Kospezifikation *keine* Beziehung asymmetrischen Wiederaufgriffs eines Antezedens durch eine Anapher, sondern vielmehr eine *symmetrische Relation* verkörpert. Insofern erschließt sich die Bedeutung der Unterstreichungen in den obigen Beispielsätzen durch eine entsprechend differenzierte Sichtweise: Es handelt sich hier nicht primär um die “zu resolvierende” Anapher, deren syntaktische Strukturbedingung einzig zählt, sondern um eines von mehreren Vorkommen, das *ob seiner spezifischen Position im Satz* für die Akzeptabilität bzw. Inakzeptabilität des jeweiligen Belegs ausschlaggebend ist; jedoch sind auch die übrigen Vorkommen für die konfigurationale Akzeptabilität bzw. Inakzeptabilität prinzipiell relevant.

Eine Theorie, die das beschriebene Phänomen unter Rekurs auf die Kospezifikations-Relation modelliert, hat somit zu gewährleisten, daß etwa die unterschiedliche Akzeptabilität der Belege

- (10a) *Behrens_i verlangt, daß der Friseur_j ihn_i rasiert.*
 (10b) * *Er_i verlangt, daß der Friseur_j Behrens_i rasiert.*

reproduziert wird: Relevant sind eben nicht alleine die Lokalitätsbeziehungen und Vorkommens-typen, sondern darüberhinaus die Positionierung der einzelnen Vorkommen im Satz.

Neben der Symmetrie legt die *Transitivitätseigenschaft* der Kospezifikationsrelation eine differenzierte Sicht anaphorischer Wiederaufgriffsbeziehungen nahe, die sich nicht auf die Ebene einzelner Antezedens-Zuordnungen beschränkt. In folgendem Beispiel

sind. In der Bewertung rekurriert der Autor ausschließlich auf seine eigene Sprachkompetenz. Die Gültigkeit scheint jedoch gewährleistet zu sein, da die hier diskutierten Beispiele in der einschlägigen Literatur i.d.R. als unkontrovers angesehen werden.

- (11) * *Behrens_i verlangt vom Friseur_j, daß er_i ihn_i rasiert.*

erscheint die Wahl des Antezedens “*Behrens*” für beide Pronomen *individuell zulässig*, jedoch erweisen sich die entsprechenden Entscheidungen als inkompatibel, da *transitiv* eine Kospezifikationsrelation zwischen den beiden Pronomen im Nebensatz entsteht, die - entsprechend Beleg (8b) - konfigural *unzulässig* ist.

Auf der bisherigen Betrachtungsebene wurden Begriffe wie “lokal”/“Teilsatz”, “strukturelle Position” und “Asymmetrie” verwendet, ohne sie unter Bezugnahme auf konkrete Strukturbeschreibungen zu formalisieren. Im Hinblick auf das Ziel der Algorithmisierung der Restriktionen ist dies jedoch unumgänglich. Im folgenden sollen theoretische Konzepte diskutiert werden, die diese Lücke schließen. Die empirischen Beobachtungen, die theoretisch-formal zu modellieren sind, lassen sich folgendermaßen zusammenfassen:

Die Kospezifikations-Relation unterliegt satzintern folgenden Bedingungen:

1. *Reflexivpronomen kospezifizieren mit einem “geeigneten” Vorkommen aus derselben “lokalen Domäne”.*
2. *Nichtreflexive Personalpronomen unterliegen Restriktionen, die bestimmte Kospezifikationen innerhalb derselben “lokalen Domäne” ausschließen.*
3. *Nichtpronominalen anaphorischen Ausdrücke unterliegen Restriktionen, die bestimmte Kospezifikationen sowohl innerhalb derselben “lokalen Domäne” als auch auf sonstig-satzlokaler Ebene ausschließen.*

Theoretische Modellierung

Eine für die beschriebene Zielsetzung geeignete theoretisch-formale Modellierung sollte einen konzeptuellen Rahmen zur Verfügung stellen, innerhalb dessen die zentralen Begriffe der *lokalen Domäne* sowie der *Geeignetheit* angemessen eingegrenzt werden können. Da das Ziel in der Entwicklung eines Performanzmodells besteht, geht es hierbei primär um die Reproduktion der intellektuellen Akzeptabilitätsbewertungen⁵. Geeignete Formalisierungen sind insbesondere in Theoriekomponenten der Generativen Grammatik erarbeitet worden, die mit dem weitergehenden Anspruch sowohl eines Kompetenzmodells als auch einer Theorie des Spracherwerbs antritt (Chomsky, [Chom81, Chom82, Chom86]; Grewendorf, [Grew88]; von Stechow und Sternefeld, [StSt88]). Die kompetenzlinguistische Ausrichtung der theoretischen Vorarbeiten bedingt, daß es primär um die Identifikation universeller, ggf. parametrisierter Prinzipien geht und die Aspekte einer prozeduralen Umsetzung allenfalls nachrangig diskutiert werden. Dennoch werden sich zumindest einige der formalen Konzepte - insofern ihre Algorithmisierbarkeit nachweisbar ist - als nützlich erweisen.⁶

Bezugspunkt der formalen Konzepte ist die Beschreibungsebene der sog. *Oberflächenstruktur*, d.h. die Ebene der *syntaktischen Struktur* der Generativen Grammatik.⁷ Als erstes Beispiel soll

⁵d.h. - in der Terminologie Chomskys - um die *Beobachtungs-Adäquatheit* der Theorie

⁶Eine Algorithmisierung der relevanten Prinzipien unter möglichstem Erhalt von deren Universalität hätte darüberhinaus den Vorteil der Sprachportabilität des Verfahrens.

⁷Für die hier verfolgten Zielsetzungen erscheint es verzichtbar, auf das Pendant, die zugrundeliegende *Tiefenstruktur*, sowie deren systematische Beziehung zur Oberflächenstruktur einzugehen. Es sei an dieser Stelle

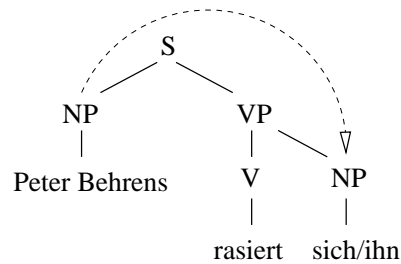


Abbildung 7.1: Oberflächenstruktur der Belege (8a)/(8b)

die typische phrasenstrukturelle Modellierung der Belege (8a)/(8b) betrachtet werden (vgl. Abbildung 7.1). Es geht hier in erster Linie darum, die theoretischen Konzepte einzuführen, die eine formale Definition der oben identifizierten Kospezifikations-Restriktionen ermöglichen. Die ebenfalls wichtige Frage, wie mit den typischen Problemen der Syntaxanalyse umzugehen ist, wird zunächst zurückgestellt.

Der strukturbezogene Begriff, dem in der hier ausschnittsweise darzustellenden Theorie eine zentrale Bedeutung zukommt, ist der der *K-Herrschaft*⁸ (vgl. Reinhart ([Rein76]) sowie die Weiterentwicklungen u.a. von Reinhart ([Rein83]) und Chomsky ([Chom81], [Chom86])). Ohne auf die teils subtilen Hintergründe einzugehen, die diesen Varianten zugrundeliegen und sich z.T. dadurch begründen, daß die K-Herrschafts-Definition den Anforderungen einer Vielzahl von Modulen der Generativen Grammatik zu genügen hat, soll auf eine Version zurückgegriffen werden, die sich für die verfolgten Zielsetzungen als hinreichend erweisen wird (vgl. von Stechow und Sternefeld: [StSt88], S. 36). Bezugspunkt der Definition sind die Knoten des Oberflächenstrukturbaums.

Definition 7.1 (Relation K-Herrschaft) *Knoten N_1 k-beherrscht N_2 genau dann, wenn folgende Bedingungen erfüllt sind:*

1. *Der nächste verzweigende Knoten, der N_1 dominiert, dominiert auch N_2 .*
2. *Weder wird N_2 von N_1 dominiert, noch wird N_1 von N_2 dominiert, noch ist $N_1 = N_2$.*

Der Pfeil in Abbildung 7.1 hebt eine K-Herrschafts-Beziehung zwischen den NPs “Peter Behrens” (N_1) und “sich”/“ihn” (N_2) hervor; dem S-Knoten kommt hierbei die Rolle des *nächsten verzweigenden Knotens* zu.

Als spezifisches Modul der Generativen Grammatik macht die sog. *Bindungstheorie* vom Begriff der K-Herrschaft Gebrauch, um ein formales Modell zu bilden, das die Akzeptabilitätsbewertungen der oben diskutierten Beispiele reproduziert. U.a. wird auf der Grundlage der K-Herrschafts-Relation ein Kriterium zur Charakterisierung derjenigen Vorkommen definiert, die als “geeignete” respektive “ungeeignete” lokale Antezedenten anzusehen sind. Notwendiger

lediglich darauf hingewiesen, daß die Theorie(n) zur Generativen Grammatik nicht ausschließlich die syntaktische Oberflächenstruktur als Bezugspunkt haben.

⁸im englischen Original “c(onstituent)-command”

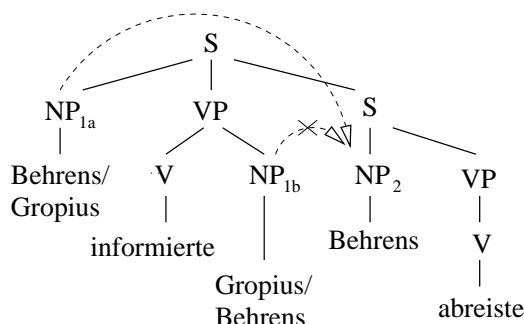


Abbildung 7.2: Oberflächenstruktur der Belege (9b)/(9c)

Zwischenschritt ist die Definition eines weiteren - zunächst rein strukturellen - Begriffs, der sog. *Bindungsrelation*:

Definition 7.2 (Bindungsrelation) *Knoten N_1 bindet Knoten N_2 genau dann, wenn N_1 und N_2 koindexiert sind und N_2 von N_1 k-beherrscht wird.*

Unter der Annahme, daß (die NP-Knoten von) “Peter Behrens” und “sich”/“ihm” koindexiert sind, besteht somit in dem in Abbildung 7.1 gezeigten Oberflächenstrukturbaum eine Konfiguration, in der erstere NP letztere NP bindet, nicht jedoch umgekehrt, da die K-Herrschafts-Relation nur in einer Richtung besteht.

Erst durch die bislang auf intuitiver Grundlage vollzogene Identifikation des formalen Konzepts der *Koindexierung* mit dem inhaltlichen Konzept der *Kospezifikation* wird das Konzept der Bindungsrelation mit der inhaltlichen Interpretation unterlegt, auf deren Grundlage sie im Rahmen der Bindungstheorie zum Einsatz gelangt. *Die Idee besteht darin, unter Rekurs auf die Definition der Bindungsrelation eine bestimmte Teilmenge aller möglichen Index-Distributionen und damit Kospezifikations-Relationen als zulässig zu charakterisieren.*

Der K-Herrschafts-Begriff erweist sich nun deshalb als relevant, weil er die im vorangegangenen Abschnitt identifizierte *strukturelle Asymmetrie*, die etwa bezüglich der für nichtpronominale Anaphern bestehenden Koindexierungsoptionen zum Tragen kommt, adäquat formalisiert: In den inakzeptablen Belegen (9a) und (9b) wird ein Nicht-Pronomen von einem Ko-Spezifikator k-beherrscht; in den akzeptablen Belegen (9c) und (8a) ist dies nicht der Fall. Abbildung 7.2 zeigt dies exemplarisch für die Belege (9b) und (9c): NP_{1a} , jedoch nicht NP_{1b} k-beherrscht NP_2 (in letzterem Fall ist VP der nächstverzweigende Knoten).⁹ Die Beispiele legen es also u.a. nahe, eine Restriktion zu formulieren, derzufolge Nichtpronomen nicht von einem Ko-Spezifikator k-beherrscht werden dürfen - unter Rekurs auf die Bindungsrelation: *Nichtpronomen sind niemals gebunden.*

Nun steht noch eine geeignete Formalisierung des *Lokalitätsbegriffs* aus, der in den Bedingungen für die übrigen Vorkommenstypen zum Tragen kommt. Das entsprechende Konzept, das hier eingeführt werden soll, firmiert unter der Bezeichnung *Bindende Kategorie*.¹⁰ Im Vorgriff

⁹Die phrasenstrukturelle Modellierung von Adverbialsätzen geschieht in Anlehnung an Grewendorf ([Grew88], S. 129f.).

¹⁰Dieser *bindungstheoretische Ansatz* entspricht dem neueren Modell, das etwa in von Stechow und Sternefeld ([StSt88]) zugrundegelegt wird. In Chomskys ursprünglichem Modell werden lokale Domänen unter Rekurs auf

auf eine geeignete Definition dieses Begriffs sollen zunächst die Restriktionen formuliert werden, denen die Koindexierungen der drei Anaphertypen unterliegen:

Definition 7.3 (Bindungsprinzipien A, B und C)

- (A) Ein Reflexiv- oder Reziprok-Pronomen ist in seiner Bindenden Kategorie gebunden.
- (B) Ein nichtreflexives Pronomen ist in seiner Bindenden Kategorie frei (d.h. nicht gebunden).
- (C) Eine nichtpronominale Anapher¹¹ ist in jeder Domäne frei (d.h. nicht gebunden).

Auf den ersten Blick bietet es sich an, lokale Domänen anhand der S-Knoten des Oberflächenstrukturbaums zu definieren. Dieser Ansatz erweist sich jedoch als unzureichend, da es Fälle (insbesondere infinitive Nebensätze) gibt, in denen diese Schranke durchbrochen werden darf. Als adäquater Bezugspunkt erweist sich die Existenz bzw. Nichtexistenz eines antezedenstauglichen lokalen "Subjekts"; hierbei liegt ein *verallgemeinerter Subjektbegriff* zugrunde, der auch bestimmte Formen logischer Subjekte der Nominalkategorie N berücksichtigt, die ebenfalls lokale Domäne induzieren. Relevant sind u.a. Possessivpronomen, sächsischer Genitiv und bestimmte Formen attributiver Präpositionalphrasen. Folgender Beleg-Set demonstriert die empirische Evidenz für den Fall des Possessiv-Pronomens.

- (12a) Der Friseur_i hört seine_j Geschichte von sich_j.
- (12b) * Der Friseur_i hört seine_j Geschichte von sich_i.
- (12c) * Der Friseur_i hört seine_j Geschichte von ihm_j.
- (12d) Der Friseur_i hört seine_j Geschichte von ihm_i.

Unter Berücksichtigung weiterer Feinheiten gelangt man zu folgender empirisch zufriedenstellenden Definition des Begriffs der Bindenden Kategorie:

Definition 7.4 (Bindende Kategorie) *Knoten N_1 ist Bindende Kategorie von Knoten N_2 genau dann, wenn N_1 der nächste Knoten ist, der N_2 dominiert und folgende Bedingungen erfüllt:*

1. N_1 dominiert ein $SUBJEKT^{12} N_3$, das N_2 *k*-beherrscht.

den Begriff der *minimalen regierenden Kategorie*, einem Element der *Rektionstheorie*, definiert. Die subtilen Unterschiede der beiden Ansätze erscheinen allenfalls unter theoretischen Gesichtspunkten als relevant - vgl. [StSt88], S. 242:

"Soweit wir sehen, werden damit die Begriffe der bindenden und der regierenden Kategorie empirisch äquivalent."

Da sich der bindungstheoretische Ansatz einfacher beschreiben läßt, wird er hier zugrunde gelegt.

¹¹In der Theorie findet sich hier der Begriff des *R-Ausdrucks*, d.h. des "selbsttätig referenzierenden" sprachlichen Ausdrucks. In der Literatur finden sich unterschiedliche Definitionen; im Rahmen dieser Arbeit wird die Sichtweise Haegemans zugrundegelegt, die hierunter neben *Namen* u.a. auch sonstige *nichtpronominale NP* subsumiert ([Haeg91], 214 f.).

¹²Für die Entwicklung eines prinzipiellen Verständnisses, das den Zielsetzungen der vorliegenden Arbeit gerecht wird, ist es hinreichend, diesen Begriff des *SUBJEKT* ("big subject") vereinfachend mit dem zuvor erläuterten Begriff des verallgemeinerten Subjekts gleichzusetzen. Im Rahmen einer präziseren Darstellung der Bindungs-

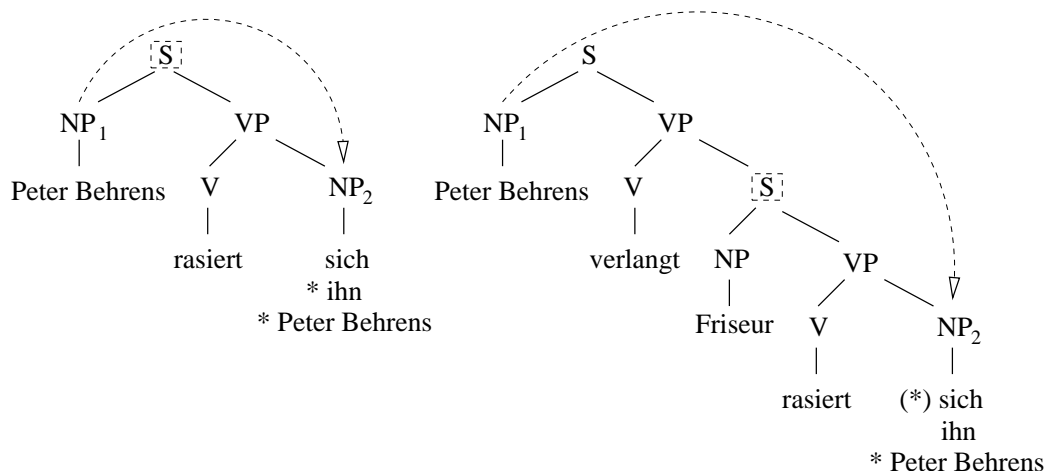


Abbildung 7.3: Lokale vs. nichtlokale Bindung in den Belegen (8a)/(8b) bzw. (8c)/(8d)

2. Falls N_2 ein Reflexiv- oder Reziprok-Pronomen ist, fällt eine Koindexierung von N_2 und N_3 nicht unter den *i-über-i-Filter*.

Der *i-über-i-Filter* ist folgendermaßen definiert:

Definition 7.5 (i-über-i-Filter) Oberflächenstrukturelle Koindexierungs-Konfigurationen der Form $[\alpha \dots [\beta \dots]_i]_i$ sind ausgeschlossen.

Für Reflexiva und Reziproka existiert eine bestimmte Phänomenklasse, die einen Rekurs auf den *i-über-i-Filter* bereits im Rahmen der Definition der Bindenden Kategorie indizieren. Darüberhinaus verkörpert der *i-über-i-Filter* jedoch eine *allgemeine Koindexierungs-Restriktion*, die unabhängig von den oben definierten Bindungsprinzipien zum Tragen kommt. U.a. sperrt erst diese Bedingung die Koindexierung einer Possessiv-Pronomen-NP mit deren Bezugs-NP

- (13) * Der Friseur_i rasiert seinen_j Kunden_j.

denn über die Bindungsprinzipien ist dieser Fall nicht abgedeckt, da zwischen den NP-Knoten eine Dominanzrelation und daher keine K-Herrschafts-Beziehung besteht.

Abbildung 7.3 zeigt exemplarisch für die oberflächenstrukturellen Konfigurationen der Belege (8a)/(8b) bzw. (8c)/(8d) auf, daß der Lokalitätsbegriff der Bindenden Kategorie sowie die darauf rekurrierenden Bindungsprinzipien (BPs) empirisch adäquat sind. In der linken Struktur ist

theorie wäre jedoch feiner zu unterscheiden zwischen finiten und infiniten Konstruktionen, für die sich empirische Unterschiede nachweisen lassen, die durch eine entsprechende Verfeinerung des SUBJEKT-Begriffs theoretisch erfaßt werden können: In finiten Konstruktionen fällt SUBJEKT mit einer künstlichen Kategorie AGR zusammen, die vom entsprechenden S-Knoten direkt dominiert wird und wie das gewöhnliche Subjekt den Inhalt von S k-beherrscht; in allen anderen Fällen koinzidiert es mit dem herkömmlichen Subjekt (vgl. von Stechow und Sternefeld: [StSt88], S. 215 ff).

der oberste S-Knoten die Bindende Kategorie von NP_2 . Somit wird NP_2 durch NP_1 *lokal* gebunden. Die BPs sagen korrekt voraus, daß ausschließlich im Falle des Reflexivpronomens eine Koindexierung mit NP_1 zulässig (und sogar erzwungen) ist. Für die rechte Struktur verkörpert der S-Knoten des Nebensatzes die Bindende Kategorie von NP_2 . Eine Koindexierung von NP_1 und NP_2 führt zu einer *nichtlokalen Bindung*; infolgedessen lizenziert BP B eine pronominale (nichtreflexive) NP_2 , wohingegen nichtpronominale NP_2 qua BP C noch immer ausgeschlossen sind. Am subtilsten gestalten sich die Verhältnisse im Falle einer reflexiven/reziproken NP_2 : Die Koindexierung mit NP_1 wird hier *nicht* durch BP A ausgeschlossen, sondern ergibt sich *transitiv*: BP A fordert die Koindexierung von NP_2 mit der Subjekt-NP des Nebensatzes; diese wiederum unterliegt BP C und darf daher nicht mit NP_1 koindexiert werden. In der Tat ist die nichtlokale Koindexierung von Vorkommen, die unter BP A fallen, an der Tagesordnung:

(14) *Behrens_i verspricht, daß er_i sich_i rasiert.*

Die Diskussion dieses Beispiels gibt somit erneut Aufschluß darüber, daß die in den Definitionen 7.3 und 7.5 formulierten bindungstheoretischen Restriktionen nicht auf der Ebene individueller Antezedensentscheidungen zum Tragen kommen, sondern *vollständige Indexverteilungen* betreffen. Eine adäquate algorithmische Umsetzung der Prinzipien und Filter hat dies zu respektieren. In der Theorie Chomskys kommt an diesem Punkt eine sog. *free indexing rule* zum Einsatz - ein theoretisches Konstrukt, daß alle möglichen Indexkombinationen generiert ([Chom81], S. 186 ff.); den Restriktionen der Bindungstheorie kommt die Aufgabe zu, genau diejenigen Kombinationen zu lizenzieren, die - interpretiert als Kospezifikationsmarker - intuitiv akzeptabel sind. Im folgenden wird sich herausstellen, daß es alternative Lösungen gibt, die unter Effizienzgesichtspunkten zu bevorzugen sind.

Diskussion

Unter der Voraussetzung, daß eindeutige Beschreibungen der syntaktischen Oberflächenstruktur verfügbar sind, ist die *Frage nach der Algorithmisierbarkeit* der bindungstheoretischen Restriktionen zu bejahen. Die unterschiedlichen Basisoperationen - etwa die Ermittlung der K-Herrschafts-Beziehungen oder die Bestimmung der Bindenden Kategorie - sind durch elementare Suchoperationen auf den oberflächenstrukturellen Beschreibungen operationalisierbar (vgl. von Stechow und Sternefeld: [StSt88], S. 229). Die Ausführungen in Abschnitt 5.4.1 haben allerdings belegt, daß diese Annahme aus unterschiedlichen Gründen unrealistisch ist. Durch ein elementares Beispiel läßt sich veranschaulichen, daß etwa je nach Entscheidung einer syntaktischen Anknüpfungs-Ambiguität unterschiedliche Koindexierungs-Optionen bestehen:

(15) *Behrens_i beobachtet den Eigentümer_j des Feldstechers_k mit ihm_k.*

Anhand von Abbildung 7.4 wird deutlich, daß die angegebene Koindexierung von "Feldstecher" und "ihm" nur unter der Voraussetzung zulässig ist, daß die PP "mit ihm" *adverbial* interpretiert wird; die Interpretation als *Attribut* zu "Feldstecher" führte zu einer Konfiguration, die durch den i-über-i-Filter (Definition 7.5) ausgeschlossen wird.

Strukturelle Mehrdeutigkeit ist jedoch nicht das einzige Hindernis einer Algorithmisierung der

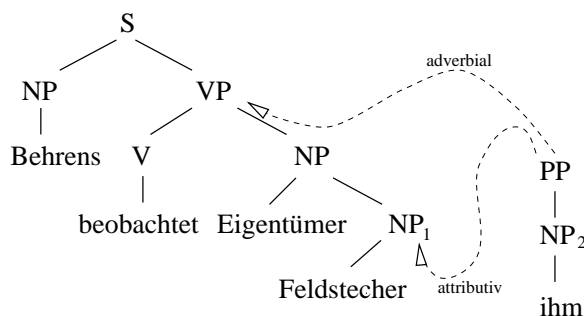


Abbildung 7.4: Anknüpfungs-Ambiguität in Beleg (15)

Koindexierungs-Restriktionen. Auch in Fällen, in denen keine Anknüpfungsambiguität besteht, kann es zu Problemen kommen - etwa bei der Bestimmung der Substrukturen, die eine lokale Bindungsdomäne beschreiben. In besonderem Maße betrifft dies die Instanziierung des lokalen Subjekts in NP-Bindungsdomänen (vgl. Stuckardt: [Stu96b]). Während sich die Verhältnisse im Falle expliziter Possessivmarker wie sächsischer Genitiv oder Possessivpronomen (vgl. Bsp. (12)) eindeutig gestalten, bestehen bei adjektivisch verwendeten Partizipien und deverbativen Nomen nichttriviale Entscheidungsprobleme. Für Deverbativa beispielsweise kann ein Genitivattribut entweder das Subjekt oder das Objekt instanzieren (*genitivus subiectivus* bzw. *genitivus obiectivus*, vgl. Teubert: [Teub79]). Die folgenden Beispiele zeigen, daß es im allgemeinen nicht ausreicht, die Existenz einer lokalen Domäne festzustellen, daß jedoch die Ermittlung der die Koindexierungsoptionen determinierender lokalen Mitspieler i.a. die Einbeziehung kontextabhängiger semantischer und pragmatischer Faktoren bedingt.¹³

- (16a) *Behrens_i akzeptiert die Entscheidung_k für ihn_i.*
 (16b) * *Behrens_i akzeptiert die Entscheidung_k für sich_i.*

Gemäß der intuitiven Akzeptabilitätsbewertung induziert das Deverbativnomen eine Bindungsdomäne, da der reflexive Wiederaufgriff des Subjekts des Hauptsatzes aus dem Präpositionalattribut ausgeschlossen zu sein scheint. Die Verhältnisse können sich jedoch ändern, wenn ein anderes Hauptsatzverb gewählt wird:

- (17a) *Behrens_i revidiert die Entscheidung_k für ihn_i.*
 (17b) *Behrens_i revidiert die Entscheidung_k für sich_i.*

Hier scheint auf den ersten Blick nicht klar zu sein, ob eine Bindungsdomäne zu eröffnen ist. Der Schlüssel liegt in der Beobachtung, daß in den Belegen (16) und (17) eine durch semantische/pragmatische Faktoren bedingte Interaktion von Hauptsatzverb und Deverbativum Auswirkungen auf die Instanzierungsoption des impliziten lokalen Subjekts hat: Jemand, der eine Entscheidung *akzeptiert*, kann nicht deren bewußter Agens sein. Für "*revidieren*" hingegen bleiben beide Möglichkeiten offen, was sich in der Akzeptabilität sowohl des reflexiven als auch

¹³In der folgenden Diskussion der Beispiele (16), (17) und (18) wird die Interpretation der PP "*für ihn*" als *Attribut* zu "*Entscheidung*" angenommen; ein entsprechender Kontext, in dem dies die einzig mögliche Lesart ist, könnte leicht angegeben werden.

des nichtreflexiv-pronominalen Wiederaufgriffs niederschlägt. Diese Beobachtung wird gestützt durch die folgenden Variationen mit expliziter Instanziierung des lokalen Subjekts:

- (18a) *Behrens_i revidiert Inas_j Entscheidung_k für ihn_i.*
 (18b) * *Behrens_i revidiert Inas_j Entscheidung_k für sich_i.*
 (18c) * *Behrens_i revidiert seine_i Entscheidung_k für ihn_i.*
 (18d) *Behrens_i revidiert seine_i Entscheidung für_k sich_i.*

Somit greifen Ansätze, die sich bei der Erkennung lokaler Bindungsdomänen auf das Vorhandensein sicherer Indikatoren wie eindeutige Possessivmarker beschränken, in Fällen wie (16) zu kurz: (16a) würde fälschlicherweise verworfen und (16b) fälschlicherweise akzeptiert.

Das ‐Ja‐ auf die Frage nach der Algorithmisierbarkeit bedarf somit einer Relativierung: Eine uneingeschränkte Operationalisierung der Koindexierungsrestriktionen ist nur unter der Voraussetzung möglich, daß die zuvor beschriebenen Fallklassen von Ambiguität resolviert werden können. Dies setzt jedoch voraus, daß geeignetes Hintergrundwissen bzw. Inferenzregeln zur Verfügung stehen - eine Anforderung, die in realistische Anwendungsszenarien unter Rekurs auf gegenwärtige Techniken i.d.R. nur partiell einlösbar ist. Es sind somit geeignete *Strategien robuster Verarbeitung* zu entwickeln, die dieses Problem lösen oder umgehen.

Des Weiteren werden in der Literatur *Problemfälle* diskutiert, für die die Bindungstheorie falsche Vorhersagen zu treffen scheint (nach Evans ([Evan80]) bzw. Sells):

- (19a) *Reagan_i voted for Reagan_i.*
 (19b) *Everyone likes John_i. Fred likes him_i. Mary likes him_i. Even John_i likes him_i.*

Die Koindexierung zu Beleg (19a) erscheint durch BP C ausgeschlossen; (19b) ginge mit einer transitiven Verletzung von BP B einher. Hierbei handelt es sich jedoch nur um Ausnahmefälle, die z.T. unter Rekurs auf die inhaltliche Ebene erklärbar sind: Für (19a) beispielsweise ließe sich argumentieren, daß die beiden Vorkommen von ‐Reagan‐ nicht kospezifizieren, sondern allenfalls eine Relation der abgeleiteten Spezifikation besteht, die nicht den bindungstheoretischen Restriktionen unterliegt.¹⁴ In realen, anwendungsrelevanten Texten kommt derartigen Fällen eine allenfalls marginale Rolle zu.

Daß es sich lohnt, den Aufwand einer robusten Implementierung der syntaktisch-konfiguralen Restriktionen zu betreiben, wird im folgenden Kapitel deutlich werden. Erfolgreiche Verfahren zur Anapherinterpretation basieren i.d.R. auf der Operationalisierung von Koindexierungsrestriktionen, die auf der Ebene der oberflächenstrukturellen Konfiguration lokalisiert sind. Allerdings wird i.d.R. von einem idealisierten Szenario ausgegangen, in dem eine eindeutige Beschreibung der Oberflächenstruktur vorgegeben ist. Im Rahmen der vorliegenden Arbeit soll diese Algorithmisierungslücke geschlossen und damit die Voraussetzung geschaffen werden, die bindungstheoretischen Restriktionen auch für realistische Anwendungsszenarien zu operationalisieren.

¹⁴Das zweite Vorkommen von ‐Reagan‐ könnte etwa dahingehend interpretiert werden, daß es eine relationale inhaltliche Entität der Art ‐Reagan becomes President‐ spezifiziert.

7.1.3 Semantische Skopus-Restriktionen

Eine weitere Klasse von Restriktionen ist auf der Ebene von Repräsentationen lokalisiert, die einer bestimmten Klasse diskursemantischer Theorien zugrundeliegen (u.a. Kamp, [Kamp81]; Kamp und Reyle, [KaRe93]; Heim, [Heim82]). Ähnlich wie in der syntaktischen Oberflächenstruktur lassen sich auch in diesen semantischen Strukturen konfigurationale Restriktionen verankern, die die Antezedensoptionen anaphorischer Vorkommen restringieren.

Elementare Beispiele

Folgender Minidiskurs stellt ein klassisches Beispiel dar:

- (20) *Jede Bäuerin_i, die_i einen Esel_j besitzt, schlägt ihn_j.*
 * *Er_j ist stets sehr störrisch.*

Die Inakzeptabilität der Kospezifikation des Pronomens “*Er*” und des Kandidaten “*Esel*” wird unter Bezugnahme auf die *Quantifikationsstruktur* erklärt. Die Objektrepräsentation des Vorkommens zu “*Esel*” ist nur innerhalb des Skopus des Allquantors, der durch die NP “*Jede Bäuerin*” induziert wird, *lokal gebunden*, und daher nicht von außen per anaphorischem Verweis zugreifbar:

$$\forall x.(B(x) \Rightarrow \exists y.(E(y) \wedge s(x, y)))$$

Eine solche Beschränkung kommt auch in Fällen impliziter Allquantifizierung zum Tragen:¹⁵

- (21) *Falls eine Bäuerin_i einen Esel_j besitzt, schlägt sie_i ihn_j.*
 * *Er_j ist stets sehr störrisch.*

In der *Diskursrepräsentationstheorie* von Kamp werden solche Restriktionen als *accessibility constraints* bezeichnet und unter Rekurs auf die hierarchisch-rekursiven Konfigurationen sog. *Diskursrepräsentations-Schemata* operationalisiert, die die Quantifikationsstruktur abbilden; die den Vorkommen zugeordneten Spezifikate werden als gebundene Variable - sog. *Diskursreferenten*¹⁶ - modelliert. Die Zugreifbarkeitsbedingungen kommen für inter- und intrasententielle Kospezifikation gleichermaßen zum Tragen.

¹⁵In diesem Fall besteht allerdings eine Mehrdeutigkeit, da der Existenzquantor sowohl engen als auch weiten Skopus haben kann. Die Inakzeptabilitätsbewertung bezieht sich auf den ersteren Fall.

¹⁶Im Rahmen der Diskursrepräsentationstheorie unterliegt diesem Begriff ein spezifisches Verständnis, das nur in den Grundzügen mit der Festlegung in Definition 6.3 übereinstimmt.

Diskussion

Die Tauglichkeit der in den genannten diskurssemantischen Theorien modellierten Restriktionen für anwendungsorientierte Ansätze zur Anapherninterpretation erscheint in gleich mehrfacher Hinsicht fragwürdig. Der erste und entscheidende Punkt betrifft die *Algorithmisierbarkeit* des Konstruktionsprozesses der zugrundeliegenden Repräsentationen - also etwa der Diskursrepräsentations-Schemata in der Theorie von Kamp. In einer Arbeit von Asher und Wada, in deren Rahmen der Versuch unternommen wird, ein formales Modell - einen "*computational account*" - der Anaphernresolution zu entwickeln, das u.a. die Zugreifbarkeitsrestriktionen der Diskursrepräsentationstheorie operationalisiert, werden Beispiele diskutiert, die illustrieren, daß es sich hierbei um ein potentiell schwieriges Problem handelt ([AsWa88]). Unklar ist u.a., in welchen Fällen ein an einen Allquantor mit globalem Skopus gebundener Diskursreferent intersententiell wiederaufgreifbar ist: In obigem Beispiel (20) scheint dies für den Diskursreferenten x (zu "Bäuerin") ausgeschlossen zu sein, jedoch gibt es Belege, in denen es sich augenscheinlich anders verhält ([AsWa88], S. 317):

- (22) *Every cadet_i receives a rigorous training.
First he_i goes to boot camp.
Then, he_i gets intensive flight instruction in a basic trainer.
Finally, he_i is given 200 hours of flight time in a supersonic aircraft.*

Asher und Wada gelangen zu dem Schluß, daß es schwierig sei, formale Regeln von allgemeiner Gültigkeit aufzustellen, mit Hilfe derer diese Fälle unterschieden werden können (ibd., S. 340, Randbemerkung 18):

"We have no idea how to capture such cases without appeal to detailed world knowledge; [...]"

Es erscheint somit ausgeschlossen, diesen Restriktionstypus für uneingeschränkten Diskurs mit hinreichendem Abdeckungsgrad zu algorithmisieren.

Ein zweiter, im Lichte dieses Sachverhalts nachrangiger Punkt betrifft die *Relevanz* der Skopus-Restriktionen für die angewandte Anaphernresolution. Eine Korpusstudie würde aufzeigen, daß die strukturellen Gegebenheiten der prototypischen Testfälle - etwa der berühmten "Esels-Sätze" - in realen Texten äußerst selten anzutreffen sind; aus performanzlinguistischer Sicht zeitigt das Algorithmisierungsproblem somit allenfalls marginale Auswirkungen.

7.1.4 Fokusbasierte Bedingungen

Neben den zuvor betrachteten Skopus-Restriktionen gibt es eine weitere Klasse von Bedingungen, die auf der Diskursebene angesiedelt sind. Empirische Untersuchungen zeigen, daß die zu einem bestimmten Zeitpunkt der Textanalyse verfügbare innertextuelle Kontextinformation einer *Prominenzordnung* unterliegt, die deren Relevanz für die Interpretation der aktuell zu verarbeitenden Äußerung determiniert. Mit anderen Worten: Durch die *Fokussierung* bestimmter

inhaltlicher Entitäten wird der Suchraum auf der Diskursrepräsentation einerseits *geordnet* und andererseits (für bestimmte Aufgaben) *begrenzt*. Es wird sich zeigen, daß das Problem der Anapherninterpretation zu denjenigen Aufgabengebieten gehört, für die sich Fokussierungsphänomene in beiderlei Hinsicht als relevant erweisen. Im Rahmen dieses Abschnitts sollen nun zunächst die *restringierenden* Aspekte sowie deren Algorithmisierbarkeit diskutiert werden.

Der Begriff der Fokussierung unterliegt in der Literatur keinem einheitlichen Verständnis. Es finden sich eine Vielzahl von Verwendungsweisen, deren Abgrenzung bisweilen auf subtilen Unterscheidungsmerkmalen basiert, die kaum algorithmisch reproduzierbar sein dürften. Die folgenden Ausführungen beziehen sich auf eine Theorie der Diskursstruktur, die sich als geeignet erweist, *zwei Ebenen der Fokussierung* zu lokalisieren, denen im Rahmen des Problems der Interpretation anaphorischer Ausdrücke eine unterschiedliche Rolle zufällt. Die Einschränkung der Betrachtung erscheint gerechtfertigt, da sich andere einschlägige Fokussierungstheorien i.a. auf eine der beiden Ebenen abbilden lassen.

Eine dreidimensionale Theorie der Diskursstruktur

Grosz und Sidner schlagen eine Theorie der Diskursstruktur vor, in deren Rahmen zwischen drei grundsätzlichen Beschreibungsebenen unterschieden wird ([GrSi86]):

- *Linguistische Struktur* (die Folge sprachlicher Ausdrücke, die die Äußerungen konstituiert),
- *Intentionale Struktur*,
- *Attentionaler Zustand* (Fokus-Strukturen).

Aus prozeduraler Sicht besteht eine *wechselseitige Abhängigkeit*: Intentionale Struktur und Attentionaler Zustand werden durch äußerungsweise Interpretation der Eingabe aufgebaut; umgekehrt verwendet der Interpretationsprozeß der einzelnen Äußerungen die Intentionale Struktur sowie den Attentionalen Zustand als Kontextinformation. Die Teilaufgaben der Äußerungsinterpretation, zu denen die Anapherninterpretation zu rechnen ist, profitieren in unterschiedlicher Weise von der Information der Fokus-Strukturen.

Gemäß dem Modell von Grosz und Sidner wird die Abfolge der Ausdrücke der Linguistischen Struktur qua Interpretation in sog. *Diskurs-Segmente* unterteilt, denen (entsprechend den öffentlichen Intentionen des Sprechers) elementare "*discourse segment purposes*" zugeordnet sind. Die Beziehungen zwischen diesen elementaren Intentionen konstituieren eine Hierarchie, die als *Intentionale Struktur* des Diskurses bezeichnet wird. Somit handelt es sich um eine pragmatische Dimension der Textualität, d.h. (in der Terminologie von Abschnitt 4.3.1) um eine Ebene, auf der sich die *Kohärenz* des Texts widerspiegelt.¹⁷

¹⁷Im Unterschied zu dem in Abschnitt 4.3.1 vorgestellten Modell der pragmatischen Relationen als *zweifach kontextualisierte Fragen* sehen Grosz und Sidner nicht zwischen *jedem* Paar benachbarter Äußerungen eine diskursstrukturell zu explizierende Beziehung. Äußerungen, die demselben "*discourse segment purpose*" zuzuordnen sind, gehören in der Intentionalen Struktur demselben Diskurssegment an.

Jedoch scheint dennoch eine weitgehende Übereinstimmung zwischen beiden Modellen zu bestehen: In Fällen, in denen Hellwig die pragmatische Anknüpfung vermöge der trivialen Frage "*Was weiter?*" umschreiben würde, scheinen Grosz und Sidner die beteiligten Äußerungen auf ein und dasselbe Diskurssegment abzubilden. Obwohl es sich lohnen sollte, diese Frage näher zu untersuchen, werden diese theoretischen Gesichtspunkte im folgenden keine Rolle spielen, da sich die beiden Ansätze unter Operationalisierungsgesichtspunkten als gleichermaßen problematisch erweisen werden.

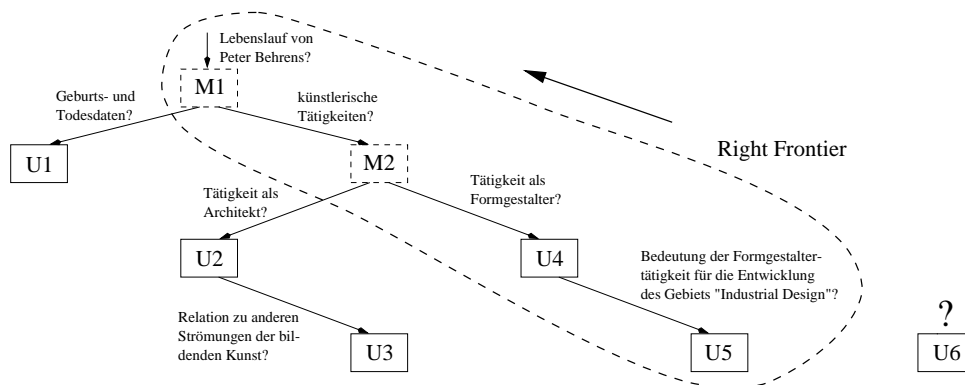


Abbildung 7.5: Globale Fokusstruktur der Kohärenzebene

Während in anderen Theorien der Diskursstruktur eine feinere Differenzierung nach sog. *rhetorischen Relationen* vorgenommen wird (vgl. z.B. Hobbs, [Hobb79]; Mann und Thompson, [MaTo87]), beschränkt sich die hier vorgestellte Theorie auf ein vergleichsweise abstraktes Basisinventar zweier rein struktureller Relationen, die zur Modellierung beliebiger intentionaler Beziehungen zwischen den Diskurssegmenten herangezogen werden. Grosz und Sidner sehen die Vorgabe einer endlichen Liste von rhetorischen Relationen als inadäquat an und begründen die Abstraktion ferner unter Verweis auf deren einfachere Algorithmisierbarkeit. Als Argument für die Geeignetheit des intentionenbezogenen Modells wird auf Studien zu sog. *aufgabenorientierten Texten bzw. Dialogen* verwiesen, in denen die Struktur der *Aufgabenbearbeitung* - als Spezialfall der Intentionalen Struktur - als fokussierungsrelevant identifiziert wird (Grosz, [Gros77]).

Attentionaler Zustand und globaler Fokus

Der *Attentionale Zustand* bezeichnet eine Fokus-Struktur, vermöge derer relevante inhaltliche Entitäten des Diskurses unter Rekurs auf einen bestimmten Ausschnitt der *Intentionalen Struktur* geordnet, d.h. fokussiert werden. Anhand von empirischen Studien wurde für eine Reihe von Teilproblemen der Äußerungsinterpretation nachgewiesen, daß der jeweils *“offene Teil”* der intentionalen Hierarchie einen eingeschränkten intratextuellen Kontext definiert, der die zur Verfügung stehenden Interpretationsmöglichkeiten abgrenzt. Der *“offene Teil”* wird definiert als der Pfad von dem der zuletzt analysierten, aktuellen Äußerung zugordneten Diskurssegment zum Wurzelsegment der Intentionalen Struktur. Abbildung 7.5 zeigt dies schematisch für den Zwischenzustand, der sich nach Interpretation von Äußerung 5 des Textes der Fallstudie in Abschnitt 4.3.2 ergibt (vgl. dort, Abbildung 4.2). Bestimmte Aspekte der Interpretation von Äußerung 6 laufen nun ausschließlich unter Bezug auf den Kontext ab, der durch diesen Pfad, den sog. *Right Frontier*, definiert wird. Dies gilt einerseits für die pragmatische Zuordnung des entsprechenden Diskurssegments selbst. Andererseits betrifft es Operationen wie die Resolution deiktischer Pronomen: Anhand von empirischen Studien weist Webber nach, daß sich die englischen Pronomen *“this”* und *“that”* ausschließlich auf Entitäten (hier: die Diskurssegmente selbst) beziehen, die in Äußerungen des *Right Frontier* erwähnt werden ([Webb88, Webb89]). Als weitere Anwendung der restriktiven Funktion des durch den Attentionalen Zustand definierten globalen Fokus böte es sich an, diese Restriktion auf beliebige *Pronominal-Anaphern* anzuwenden; Objekte, die aus-

schließlich in Diskurssegmenten außerhalb des *Right Frontier* vorkommen, können demzufolge nur nichtpronominal re-istanziiert werden.¹⁸

Der durch den offenen Teil der Intentionalen Struktur definierte Fokus enthält nicht nur die Diskurssegmente bzw. die damit verknüpften *discourse segment purposes* selbst, sondern *alle* inhaltlichen Entitäten, die in den entsprechenden Diskurssegmenten direkt spezifiziert bzw. per abgeleiteter Spezifikation zugänglich sind.¹⁹ Insofern bestehen Freiheitsgrade, da nicht operational abgegrenzt ist, in welchem Ausmaß abgeleitete Spezifikationen zu berücksichtigen sind.

Neben der Begrenzung des Suchraums wird durch die kellerspeicherartige Organisation des Attentionalen Zustands eine *Ordnung* über den fokussierten Entitäten definiert, die sich ebenfalls als relevant für die Erbringung bestimmter Interpretationsleistungen erweist. Wie der aufwärtsgerichtete Pfeil in Abbildung 7.5 andeuten soll, kann es (je nach vorliegender Aufgabe) naheliegen, die Diskurssegmente des *Right Frontier* von unten nach oben zu durchsuchen. Im Speziellen böte sich dies als *Präferenz*-Kriterium bei der Antezedensauswahl für anaphorische Elemente an.

Attentionaler Zustand und lokaler Fokus

In der Theorie von Grosz und Sidner wird eine zweite, lokale Ebene der Fokussierung identifiziert, die zwischen zwei benachbarten Äußerungen auch *innerhalb* ein und desselben Diskurssegments zum Tragen kommt. Im Unterschied zum globalen, durch die Kohärenzstruktur determinierten Fokus wird der lokale Fokus primär durch *kohäsionsstiftende* Faktoren determiniert. Der Mechanismus der lokalen Fokussierung wird als in besonderem Maße relevant für die Interpretation *pronominaler* Anaphern angesehen, da diese i.d.R. mit einem Antezedens aus den unmittelbar benachbarten Äußerungen kospezifizieren.

Ein in jüngster Zeit ausgiebig diskutiertes und auch von Grosz und Sidner vorgeschlagenes Modell der lokalen Fokussierung ist die *Centering-Theorie* (Grosz, Joshi, Weinstein: [GJW83, GJW95]). In deren Rahmen werden sowohl Restriktionen als auch Präferenzkriterien für die Anapherninterpretation formuliert. Ausgangspunkt ist eine Abfolge von Äußerungen U_1, \dots, U_n eines bestimmten Diskurssegments; jede Äußerung trägt eine Menge inhaltlicher Entitäten $\{e_{i,1}, \dots, e_{i,m_i}\}$ bei. Hierbei wird angenommen, daß die Elemente dieser Menge entsprechend ihrer relativen Prominenz (*“salience”*) partiell geordnet ist. Die intuitive Idee besteht darin, daß die Wahrscheinlichkeit einer erneuten Realisierung einer Entität in der Folgeäußerung umso größer ist, je höher die Position in der Rangordnung. Folgende Begriffe werden definiert:

Definition 7.6 (vorwärtsgerichteter Center²⁰)

Der *vorwärtsgerichtete Center* $Cf(U_i)$ einer Äußerung U_i ist die nach absteigender Prominenz topologisch sortierte Liste der inhaltlichen Entitäten:

$$Cf(U_i) = (e_{i,1}, \dots, e_{i,m_i})$$

¹⁸Es soll nochmals betont werden, daß auch die Entitäten außerhalb des Attentionalen Zustands weiterhin Bestandteil des interpretationsrelevanten Kontexts sind; die Elemente des *Right Frontier* werden jedoch als *kognitiv leichter zugänglich* aufgefaßt.

¹⁹Grosz und Sidner geben folgende Umschreibung ([GrSi86], S. 179):

“[...] ; this space contains those entities that are salient - either because they have been mentioned explicitly in the segment or because they became salient in the process of producing or comprehending the utterances in the segment ([...]).”

²⁰engl. “forward looking center”

Definition 7.7 (bevorzugter Center²¹)

Der bevorzugte Center $Cp(U_i)$ einer Äußerung U_i ist die inhaltliche Entität mit der höchsten Prominenz:

$$Cp(U_i) = first(Cf(U_i))$$

Während $Cf(U_i)$ und $Cp(U_i)$ statische Begriffe darstellen, die für einzelne Äußerungen definiert sind, ergibt sich der *lokale Fokus*, der sog. *rückwärtsgerichtete Center*, durch die Wiederaufgriffsrelationen, die zwischen zwei benachbarten Äußerungen bestehen:

Definition 7.8 (rückwärtsgerichteter Center²²)

Der rückwärtsgerichtete Center $Cb(U_{n+1})$ einer nicht diskurssegment-initialen Äußerung U_{n+1} ist definiert als eines derjenigen Elemente, die bereits in U_n realisiert sind und dort größtmögliche Prominenz haben, d.h.:

$$Cb(U_{n+1}) \in \{e \mid e \in Cf(U_{n+1}) \wedge e \in Cf(U_n) \wedge e \text{ hat max. Prominenz in } Cf(U_n)\}$$

Primärer Bezugspunkt der Theorie ist die Interpretation anaphorischer Elemente; in diesem Rahmen verkörpern die inhaltlichen Objekte die Spezifikate der sprachlichen Ausdrücke. In den Originalarbeiten zur Centering-Theorie wird nun u.a. folgende Regel definiert, durch die eine *Beschränkung der sprachlichen Realisierungsoptionen für die fokussierte Entität* zum Ausdruck gebracht wird:

Regel 7.1 (R1 der Centering-Theorie)

Ist irgendein Element e mit $e \in Cf(U_i) \wedge e \in Cf(U_{i+1})$ in U_{i+1} durch ein Pronomen realisiert, dann ist auch $Cb(U_{i+1})$ durch ein Pronomen realisiert.

D.h. die Pronominalisierung der lokal fokussierten Entität ist obligatorisch, falls irgendeine weniger prominente Entität pronominalanaphorisch realisiert wird. Umgekehrt ausgedrückt: *Bestimmte nichtpronominale Realisierungen des Fokus werden ausgeschlossen*. Unter der Annahme, daß jeweils das Subjekt der bevorzugte Center ist, ergibt sich für folgendes Beispiel²³

(23) Kurt_i hat ein Saxophon. Er_i spielt Anja_j darauf vor. Sie_j bewundert ihn_i/Kurt_i.

die Vorhersage, daß der nichtanaphorische Ausdruck "Kurt" im dritten Satz nicht den bevorzugten Center des zweiten Satzes spezifizieren könne.

Diskussion

Bezüglich der Frage nach der Algorithmisierbarkeit ist ebenfalls zwischen lokalem und globalem Fokus zu unterscheiden. Die inkrementelle Konstruktion der Intentionalen Struktur - Grundvoraussetzung für die Ermittlung des Attentionalen Zustands - betrifft die Kohärenzebene und ist

²¹ engl. "preferred center"

²² engl. "backward looking center"

²³ Die Unterstreichungen kennzeichnen die rückwärtsgerichteten Center der unterschiedlichen Äußerungen.

damit erwartet schwieriger als die Ermittlung des lokalen Fokus, der (primär) der Kohäsionsebene zuzuordnen ist.

Zwar vertreten Grosz und Sidner die These, daß die abstrahierende Intentionale Diskursstruktur einfacher zu erkennen ist als die feiner differenzierenden rhetorischen Relationen. In ihrer Arbeit gelangen sie jedoch nicht über eine allgemeine Charakterisierung des entsprechenden Prozesses hinaus. Neben der auf der Hand liegenden Aussage, derzufolge wechselseitige Abhängigkeiten zwischen den unterschiedlichen Beschreibungsebenen des Modells bestehen, werden drei Grundtypen von Information identifiziert, die für die intentionale Interpretation der Diskurssegmente sowie für deren strukturelle Einordnung in den partiellen Baum der Intentionalen Struktur relevant sind:

- Schlüssel-Wörter/-Ausdrücke²⁴
- Intentionen *einzelner* Äußerungen
- objektbezogenes und relationales Weltwissen

Paradebeispiele für Schlüsselwörter sind bestimmte Konjunktionen, Partikel und Adverbien, denen eine stereotype semantische Interpretation zukommt und die somit einen guten Anhaltspunkt für die Identifikation bestimmter rhetorischer bzw. pragmatischer Beziehungen abgeben. Während die Suche nach einer Menge vorgegebener Schlüsselwörter sowie die Identifikation von deren Bezugsentitäten auf der Ebene der syntaktischen Oberflächenstruktur geschehen kann, ist bereits die Algorithmisierung der Bestimmung der "atomaren" Intentionen von Einzeläußerungen ein offenes Problem. Entsprechend kommen Grosz und Sidner zu dem Schluß ([GrSi86], S. 190):

"It remains an open question what inferences are needed and how complex it will be [...] if the dominance relationship is not directly indicated in the discourse."

Auch in späteren Arbeiten, die auf eine formalgrammatische Beschreibung der Diskursstruktur abzielen, bleiben entscheidende Fragen offen; das Operationalisierungsproblem wird somit nicht wirklich gelöst (z.B. Scha und Polanyi, [SP88])

Es stellt sich somit die Frage, wieviel alleine auf der Basis von Schlüsselwörtern zu erreichen ist. Entsprechende Ansätze, die auf beliebigem Text lauffähig sind, wurden erst in jüngster Zeit vorgestellt. Der Algorithmus von Marcu bewerkstelligt ein unrestringiertes *rhetorisches Parsing* alleine auf der Basis von Schlüsselwörtern sowie von Information über deren typische Verwendungsweise, die a priori auf der Basis automatischer Korpusanalysen gewonnen wurde ([Marc97]). Auch wenn die ersten Analyseergebnisse (baumförmige Darstellungen der rhetorischen Diskursstruktur) beeindruckend aussehen, so stellt sich doch die Frage nach einer formalen Bewertung der Güte der Ergebnisse. Eine direkte Evaluation erweist sich insofern als problematisch, als die Aufgabe schwer zu definieren ist und die Übereinstimmung menschlicher Diskursstruktur-Codierer nur bei ca. $\frac{2}{3}$ liegt. Marcu unterbreitet den interessanten Vorschlag einer *indirekten Evaluation*, in deren Rahmen die Nützlichkeit der determinierten Strukturen für weitere Interpretationsprobleme wie u.a. der Anaphernresolution zu messen wäre. Die Untersuchungen befinden sich erst in einem Anfangsstadium; ein Einsatz der rhetorischen Diskursstruktur für die Anapherninterpretation bedingte u.a. die Untersuchung der Frage, inwieweit

²⁴engl. "*c(l)ue words/phrases*"

sich die oben beschriebenen globalen Fokussierungsphänomene auch in dieser Repräsentation lokalisieren lassen.

Auch die Implementierung der Strategien des lokalen Fokus, die sich aus den Regeln der Centering-Theorien ergeben, stellt ein nichtriviales Problem dar. Zunächst einmal fällt ins Gewicht, daß die Theorie in der oben präsentierten Fassung *unterspezifiziert* ist; eine algorithmische Umsetzung setzte u.a. voraus, daß folgende Lücken geschlossen werden:

1. Von welcher Gestalt sind die inhaltlichen Entitäten e , die die Center-Elemente konstituieren?
2. Welche Kriterien liegen der *Prominenzrelation* zugrunde, nach denen der vorwärtsgerichtete Center Cf geordnet ist?

Bezüglich Frage 1 wird in den Originalarbeiten (Grosz, Joshi, Weinsten: [GJW83, GJW95]) keine endgültige Festlegung getroffen: Die Lücke sei über die Annahme einer adäquaten semantischen Theorie zu schließen, die der Interpretation einzelner Äußerungen zugrundeliegt; bestimmte Rahmenanforderungen werden identifiziert, die sich aus der Zielsetzung einer Modellierung von Phänomenen der *lokalen Kohärenz*²⁵ ergeben. In den diskutierten Beispielen wird allerdings durchgängig von der Voraussetzung ausgegangen, daß es sich um Spezifikate von Nominalphrasen handelt. Auch unter dieser Annahme bleiben Fragen offen - etwa, inwieweit *abgeleitete Spezifikate* ebenfalls als Center-Elemente zu berücksichtigen sind.

Auch betreffend Punkt 2 sind Fragen offen. Zwar belegen psycholinguistische Studien, daß der *syntaktischen Funktion* eine wichtige Rolle zukommt: Eine durch ein syntaktisches Subjekt realisierte inhaltliche Entität wird i.d.R. als "prominenter" klassifiziert als ein durch ein Objekt oder gar ein durch ein Adverbial realisiertes Element (Gordon, Grosz, Gilliom: [GGG93]). Jedoch erscheint es evident, daß noch weitere Faktoren relevant sind.

Selbst wenn sich diese ersten Annahmen betreffend den Typ der inhaltlichen Entitäten und die zugrundegelegte Prominenzordnung als adäquat zur theoretischen Beschreibung von Phänomenen der anaphorischen Bezugnahme erweisen sollten, so ist das Problem der Algorithmisierung von Centering-Regel R1 noch immer nichttrivial. U.a. wird die Verfügbarkeit der syntaktischen Rolle vorausgesetzt - eine Annahme, die auf einer geeigneten Form von syntaktischer Analyse aufbaut. Die damit verbundenen Probleme sind mittlerweile hinlänglich bekannt. Ein weiterer Freiheitsgrad, der an dieser Stelle relevant wird, ergibt sich betreffend die Frage, in welcher Granularität sprachliche Ausdrücke auf Äußerungen abgebildet werden. Grosz, Joshi und Weinsten legen sich in diesem Punkt nicht fest ([GJW95], S. 209); geht man andererseits von der den Beispielanalysen implizit zugrundeliegenden Annahme aus, daß ausschließlich (ggf. komplexe) *Satzgefüge* relevant sind, so stellt sich unmittelbar die Frage, welche zusätzlichen Faktoren heranzuziehen sind, um etwa die Prominenzordnung zwischen Subjekt des Hauptsatzes und den Subjekten der einzelnen Nebensätzen festzulegen. In Abhängigkeit von den getroffenen Festlegungen ergeben sich also eine bzw. mehrere weitere Fragen:

3. Welche linguistischen Entitäten konstituieren Äußerungen?
4. Wie wird mit partiellen (z.B. mehrdeutigen) syntaktischen Analyseergebnissen verfahren?

²⁵Grosz, Joshi und Weinsten sprechen in bezug auf den Centering-Ansatz von einer "theory of local coherence". Letztendlich drehen sich die Betrachtungen jedoch primär um *Kohäsionsphänomene*, durch die die Kohärenz benachbarter Äußerungen eines Diskurssegments *vermittelt* wird.

5. Wie werden mehrfach vergebene syntaktische Rollen prominenztheoretisch bewertet?

Somit wird deutlich, daß die algorithmische Umsetzung der Centering-Regeln alles andere als trivial ist.

Über die *Relevanz* einer globalen Fokusstruktur für die Anapherninterpretation läßt sich bislang nur spekulieren, da die notwendigen Ressourcen erst in der Entwicklung begriffen sind (vgl. den rhetorischen Parser von Marcu). Sicherlich erscheint es von potentielltem Nutzen, zumindest in bestimmten Fällen den Suchraum auf die Entitäten des *Right Frontier* einzugrenzen. Allerdings stellt sich die Frage, ob eine stringente Auslegung dieses Kriteriums als *Restriktion* tatsächlich empirisch adäquat ist, oder ob nicht - gerade auch im Hinblick auf die zu erwartende Fehlerquote einer algorithmischen Diskursstruktur-Analyse - eine Verwendung als *Präferenzregel* mehr Erfolg verspräche. Vergleichbares scheint für die aus Regel R1 der Centering-Theorie ableitbare "Restriktion" zu gelten: Sollte in Variante 2 von Beispiel (23) tatsächlich das zweite Vorkommen "Kurt" anders interpretiert werden als das erste? Für Sprachgenerierungssysteme liegt der Nutzen von R1 auf der Hand; im Rahmen der Diskursinterpretation sollte R1 hingegen - falls überhaupt - als nichtstringente Präferenzstrategie Berücksichtigung finden.

Ein Prozeßmodell für die Interaktion der Fokusstrukturen

Die Diskussion hat gezeigt, daß bereits in bezug auf die individuellen Fokus- bzw. Diskursstrukturen nichttriviale Algorithmisierungs-Probleme bestehen. Hierbei könnte der Eindruck entstehen, daß es sich um Aufgaben handle, die aus prozeßtechnischer Sicht sequentiell zu bearbeiten sind: Zwar restringiert der globale Diskursfokus die Interpretation der anaphorischen Ausdrücke; jedoch nimmt umgekehrt der Konstruktionsprozeß der Kohärenzstruktur zumindest explizit keinen Bezug auf die Ergebnisse der Anaphernresolution. Grosz und Sidner beschränken sich darauf, eine wechselseitige Abhängigkeit zwischen Kohärenzstruktur und Äußerungs-Interpretation zu identifizieren ([GrSi86], S. 177):

"There is a two-way interaction between the discourse segment structure and the utterances constituting the discourse: linguistic expressions can be used to convey information about the discourse structure; conversely, the discourse structure constrains the interpretation of expressions ([...])."

Solange die algorithmische Diskursstruktur-Konstruktion alleine unter Bezug auf bestimmte Schlüsselwörter geschieht, sind die Verhältnisse klar. Für den allgemeinen Fall hingegen ergibt sich folgendes Problem: Da die Anapherninterpretation wesentlicher Bestandteil der Interpretation von Äußerungen ist, sind die beiden Interpretationsprozesse in adäquater Form *prozedural zu verzahnen*.

Ein erneuter Blick auf die Fallstudie in Abschnitt 4.3 erlaubt es, die bestehenden Abhängigkeiten näher einzugrenzen und ein Modell der Sequentialisierung relevanter Teilprobleme der Textinterpretation zu entwickeln. Ausgangspunkt ist die in 4.3.1 getroffene Unterscheidung zwischen referentieller, thematischer und pragmatischer Ebene der Textualität. Neben Isotopie wurde die *referentielle Ebene* als Spezialfall der auf der Kohäsionsebene angesiedelten thematischen Textentwicklung identifiziert. Dem gegenüber steht die pragmatische Dimension; für sie sei im

folgenden angenommen, daß das oben zugrundegelegte Hellwigsche Modell der Kohärenstruktur eine adäquate Umsetzung der Diskurssegmentstruktur der Grosz/Sidner-Theorie darstellt. Die Ergebnisse der Fallstudie in Abschnitt 4.3.2 ermöglichen eine feinkörnigere Charakterisierung der bestehenden Abhängigkeiten (vgl. Stuckardt und Haenelt: [StuH94]). Zunächst einmal wurde anhand der Studie deutlich, daß nicht nur die referentielle Entwicklung als Spezialfall, sondern allgemeiner die thematische (kohäsive) Entwicklung des Textes Indikatoren für den Aufbau der Kohärenzstruktur²⁶ liefert. Folglich liegt es nahe, neben der referentiellen und der Kohärenz-Interpretation eine Zwischenebene der *thematischen Interpretation* anzunehmen. Aus der Sicht eines sprachverarbeitenden Systems, als dessen Bestandteile die entsprechenden Prozesse zu realisieren sind, besteht folgendes Problem: Gegeben ist ein Text $\Theta = (U_1 \dots U_n)$ umfassend n Äußerungen; gesucht sind Repräsentationen $R_{1,n}$, $T_{1,n}$ und $P_{1,n}$ entsprechend referentieller, thematischer und pragmatischer Entwicklung des *gesamten* Textes. Die Ergebnisse sind inkrementell per äußerungsweiser Interpretation zu ermitteln. Zum Zeitpunkt der Interpretation von U_{i+1} , $0 \leq i \leq n - 1$ sind $R_{1,i}$, $T_{1,i}$ und $P_{1,i}$ sowie die auf der Basis äußerungslokaler Interpretation bekannten Charakteristika $R_{i+1,i+1}$, $T_{i+1,i+1}$ und $P_{i+1,i+1}$ bekannt;²⁷ gesucht sind $R_{1,i+1}$, $T_{1,i+1}$ und $P_{1,i+1}$. Die adäquate Verschränkung der entsprechenden Teilprozesse soll nun qua Identifikation *funktionaler Abhängigkeiten* zwischen diesen drei Ebenen ermittelt werden. Für den Prozeß der referentiellen Anknüpfung - d.h. für die Anapherninterpretation - ist zweierlei bekannt: Einerseits wird die Antezedenswahl zumindest partiell durch den globalen Fokus restringiert - maßgebend ist der Zustand der Intentionalen Diskursstruktur (genauer: von deren offenem Teil) *vor* Interpretation der aktuellen Äußerung; die Ergebnisse der referentiellen Interpretation wiederum fließen in die Bestimmung der thematischen Entwicklung ein, da die referentielle Progression einen elementaren Spezialfall der thematischen Progression verkörpert. Folglich sollte die referentielle Interpretation zuerst stattfinden. Die bestehenden Abhängigkeiten werden somit durch folgende dreistellige Funktion f_1 charakterisiert:

$$R_{1,i+1} = f_1(R_{1,i}, R_{i+1,i+1}, P_{1,i})$$

Erst auf der Grundlage der Ergebnisse der referentiellen Interpretation können Fälle komplexerer Themenprogressionen identifiziert werden, die u.a. durch Wiederholung oder auch Modulation auf der referentiellen Ebene partiell bestimmt werden. Andererseits scheint der Zustand der Intentionalen Struktur nur indirekt - in Gestalt der durch f_1 zum Ausdruck gebrachten Abhängigkeit - einzufließen, da die thematische Entwicklung orthogonal zur pragmatischen Entwicklung vonstatten geht.²⁸ Somit gilt folgende Abhängigkeitsbeziehung:

$$T_{1,i+1} = f_2(R_{1,i+1}, T_{1,i}, T_{i+1,i+1})$$

In einem dritten Schritt kann die pragmatische Einordnung der aktuellen Äußerung in die Intentionale Diskursstruktur stattfinden. In der Fallstudie wurde belegt, daß sich thematische

²⁶vgl. o., Reinstanzierung oder Modulation komplexer kontextueller Entitäten

²⁷Hiermit soll keineswegs ausgesagt werden, daß etwa die Zuordnung anaphorischer Verweise auf der satzinternen Ebene bereits vollzogen ist. Welche Information vermöge $R_{i+1,i+1}$, $T_{i+1,i+1}$ und $P_{i+1,i+1}$ zur Verfügung gestellt wird, bedarf einer gesonderten Abwägung; beispielsweise wird $R_{i+1,i+1}$ sicherlich die Information über die Anaphorizität der sprachlichen Ausdrücke, die U_{i+1} realisieren, umfassen.

²⁸Entscheidend ist, daß $T_{1,i+1}$ nicht von $P_{1,i+1}$ abhängt. Sollten textlinguistische Untersuchungen den Nachweis erbringen, daß der globale Fokus in $P_{1,i}$ auch für die thematische Interpretation in toto relevant ist, so kann f_2 problemlos entsprechend erweitert werden.

Wiederaufgriffe unmittelbar in der pragmatischen Struktur widerspiegeln können - vgl. die Beziehung zwischen U_2 und U_4 . Daß darüberhinaus die referentielle Ebene unmittelbar relevant ist, geht bereits aus der Theorie von Grosz und Sidner hervor: Die Eröffnung eines neuen Diskurssegments kann sich in der Wahl bestimmter referenzierender sprachlicher Ausdrücke widerspiegeln. Folglich besteht folgende Abhängigkeit:

$$P_{1,i+1} = f_3(R_{1,i+1}, T_{1,i+1}, P_{1,i}, P_{i+1,i+1})$$

Die Betrachtungen belegen erneut, daß der Weg von einer theoretischen Beschreibung von Fokussierungsmodellen hin zu einer algorithmischen Umsetzung noch weit ist. Sie geben aber auch Aufschluß darüber, welche konzeptuellen Verfeinerungen erfolgversprechend sind, um zu einer prozeßnäheren Beschreibung zu gelangen: Wenn es gelänge, nichttriviale Themenmuster auf der Basis von Schlüsselwörtern und oberflächenstruktureller Analyse heuristisch zu bestimmen, läge etwa eine entsprechende Verfeinerung des operationalen Ansatzes von Marcu auf der Hand.²⁹ Somit ergeben sich wichtige Anhaltspunkte betreffend die prozeß-strukturelle Relation von Interpretationsebenen, deren genaues Verhältnis in Modellen mit theoretischer Ausrichtung bislang allenfalls allgemein beschrieben wird.

7.1.5 Einschränkung kataphorischer Bezugnahmen

Eine weitere Restriktion erlaubt die Elimination bestimmter satzinterner Kandidaten, denen das Pronomen im Text *vorausgeht*. Anstelle von Antezedenten spricht man hier auch von *Postzedenten*; die Relation zwischen Pronomen und Postzedenten wird auch als *kataphorische* Bezugnahme bezeichnet. Kuno identifiziert ein sog. *Diskurs-Prinzip*, demzufolge das durch das Pronomen bezeichnete Spezifikat nicht wirklich kataphorisch erschlossen wird, da die Entität bereits aus dem Kontext bekannt sein muß ([Kuno87]), S. 87, Hervorhebung im Original):

“A pronoun can precede a coreferential NP only if the speaker can legitimately assume that the hearer can determine its referent from the preceding context (Kuno 1972).”

Die Regel wird anhand von Belegen illustriert, die indizieren, daß ausschließlich *bereits bekannte* Diskursreferenten spezifiziert werden können - indefinite Postzedenten erscheinen unzulässig:

- (24a) *Bevor er_i ging, verabschiedete sich_i Behrens_i von der Familie.*
 (24b) * *Bevor er_i ging, verabschiedete sich_i ein Architekt_i von der Familie.*

Aus den Beispielen geht zugleich hervor, daß die Restriktion offenbar nicht für Reflexivpronomen (allgemeiner: Pronomen des bindungstheoretischen Typs A) gilt, die in jedem Falle Bindungsprinzip A unterliegen.

Gemäß dem Diskursprinzip können somit nichtdefinite Postzedenten aus den Kandidatenmengen

²⁹Bereits im Rahmen der Fallstudie wurden Schlüsselwörter als relevant für die Ermittlung thematischer Entitäten identifiziert.

von Typ-B-Pronomen eliminiert werden. Es legt ferner nahe, *sämtliche* kataphorischen Wieder-
aufgriffe zu sperren, da das Spezifikat im Falle definiter Postzedenten ohnehin aus dem Kontext
bekannt sein sollte. Jedoch ergeben sich folgende Probleme:

1. Es kann i.a. schwer sein, definite und nichtdefinite Vorkommen algorithmisch zu unter-
scheiden - vgl. Abschnitt 6.1.2, Belege (12a) und (12b);
2. das Spezifikat ist nicht notwendig *unmittelbar* im textuellen Kontext realisiert, d.h. es kann
auch *abgeleitet* sein.

Die beiden Punkte werden anhand von folgender akzeptablen Variante zu Beleg (12a), Abschnitt
6.1.2, deutlich:

(25) *Die Mannschaften betraten den Rasen.
Bevor er_i den Anstoß ausführte, holte ein Spieler_i den Ball.*

Das Vorkommen “*ein Spieler*” spezifiziert einen Diskursreferenten, der durch den Ausdruck “*Die
Mannschaften*” bereits *indirekt* eingeführt wurde.

Die bestehenden Probleme legen es nahe, eine defensive Variante der Regel zu implementieren,
die Postzedenten zwar als weniger plausibel klassifiziert, jedoch nicht völlig ausschließt. Betref-
fend die Relevanz liegen noch keine Zahlen vor; es wird sich also noch zeigen müssen, ob sich
diese Restriktion - bzw. Präferenz - im Rahmen der angewandten Anaphernresolution bewährt.

7.1.6 Sonstige Restriktionen

Die bisher identifizierte Menge stringenter Kriterien deckt keineswegs alle in der Literatur unter-
breiteten Vorschläge ab. Sie umfaßt jedoch diejenigen Regeln, die für operationale Verfahren der
Anaphernresolution von zentraler Bedeutung sind bzw. denen die meiste Bedeutung zugemessen
wird. In der Arbeit von Carbonell und Brown werden zwei weitere Strategien angegeben, den-
nen die Eigenschaft zugeschrieben wird, die Antezedenskandidatenmenge a priori zu reduzieren
([CaBr88], S. 97).

(26) *Behrens_i gab Gropius_j einen Apfel_k. Er_j aß ihn_k.*

Die erste dieser Regeln - eine Restriktion, die sich vermöge *semantischer Rollenbedingungen*
(*selektionaler Restriktionen*) ergibt - wird für das anaphorische Vorkommen “*ihn*” wirksam:
Aufgrund der ausgefüllten Rolle kommen nur solche Antezedenten in Frage, die als *eßbar* anzu-
sehen sind. Im Hinblick auf den wahrscheinlichen kulturellen Kontext der Äußerung - Behrens
und Gropius gehören einer (i.d.R.) nicht anthropophagen Ethnie an - dürfte ausschließlich der
Kandidat “*Apfel*” übrig bleiben.

Für das Pronomen “*Er*” greift ein zweites Kriterium, das von Carbonell und Brown der Menge
der sog. “*Precondition/Postcondition Constraints*” zugeordnet wird. Aus der Interpretation der
ersten Äußerung ergibt sich als Ausgangspunkt der Interpretation der zweiten Äußerung, daß

nunmehr Gropius im Besitz des Apfels ist; folglich kann nur er derjenige sein, der den Apfel verspeist.³⁰

Die in der ersten Strategie zum Tragen kommenden semantischen Konzepteigenschaften lassen sich zwar grundsätzlich auf der lexikalischen Ebene *spezifizieren*, jedoch setzt deren Anwendung voraus, daß auch die Zuordnung der semantischen Rollenbedingungen an pronominale Situationspartizipanten algorithmisch vollzogen werden kann. Dies bedingt i.a. eine Analyse der syntaktischen Oberflächenstruktur, da zwischen Aktiv- und Passiv-Formen zu unterscheiden ist. Das Beispiel illustriert darüberhinaus, daß eine eindeutige Entscheidung z.T. erst unter Berücksichtigung des Kontexts getroffen werden kann.³¹ Ein Einwand ergibt sich im Hinblick auf die Flexibilität natürlicher Sprache: Auch Äußerungen wie die Folgende erscheinen verständlich:

(27) *Der Großmeister_i verspeist mit dem Turm_j den Freibauern_k.*

Was als *eßbar* anzusehen ist und was nicht, ist i.a. nur unter Rekurs auf kontextuelle Information zu entscheiden. Die semantische Rolle des Pronomens stellt nur einen Spezialfall dar. Die Strategie ist dennoch von potentielltem Nutzen, insofern sie in abgeschwächter Form als *präferenzsemantische Regel* (etwa im Sinne von Wilks, [Wilk75a, Wilk75b, Wilk75c]) Verwendung findet. Das entscheidende Problem besteht darin, eine Algorithmisierung dieser *selektionalen Präferenzstrategie* zu bewerkstelligen, die keine extensive intellektuelle Lexikonarbeit bedingt und mit dem Szenario einer partiellen oberflächenstrukturellen Analyse verträglich ist.

Die zweite Regel thematisiert einen einfach skizzierbaren Spezialfall situationspragmatischer Beziehungen zwischen zwei aufeinanderfolgenden Äußerungen, der sich auf der Basis von Weltwissen erschließt und nicht mehr lexikalisch-semantisch erfaßbar scheint. Es stellt sich die Frage, warum diese vergleichsweise schwierig zu operationalisierende spezielle Bedingung im Rahmen der Anapherninterpretation eine prominentere Rolle spielen soll als andere auf der pragmatischen Ebene angesiedelte Restriktionen. Die Voraussetzungen für eine algorithmische Umsetzung mit einer für eine Anwendung auf unrestringiertem Text hinreichenden Allgemeingültigkeit scheinen ohnehin nicht gegeben zu sein, wie auch bereits von Carbonell und Brown eingeräumt wird.

7.2 Präferenzkriterien

Mit den bisher identifizierten stringenten Strategien ist der Anspruch verbunden, die Menge der Antezedenskandidaten eines anaphorischen Elements zu *reduzieren*. Bereits im Rahmen der Diskussion wurde jedoch deutlich, daß zumindest einige der Restriktionen defensiver als *Präferenzen* umgesetzt werden sollten. Darüberhinaus existiert eine Menge von Kriterien, die entsprechend dem mit ihnen verbundenen theoretischen Anspruch von vornherein als *Präferenzregeln* zur *Auswahl* aus der nach Anwendung der stringenten Kriterien verbleibenden Kandidatenmenge

³⁰Durch eine Variation des Beispiels kann ferner gezeigt werden, daß Interdependenzen zwischen Antezedensentscheidungen auch auf der inhaltlichen Ebene bestehen können:

Der Löwe_i sah Gropius_j mit dem Apfel_k. Er_{i/j} aß ihn_{j/k}.

Auf der Basis von Weltwissen können zwei (jedoch nicht vier) zulässige Entscheidungskombinationen identifiziert werden. Eventuell erlaubt der Kontext eine definitive Entscheidung.

³¹Dies ist an sich noch kein Einwand, da es mit dem Wesen des Mehrstrategie-Modells in Einklang steht und auch für die übrigen zuvor diskutierten Restriktionen gilt.

klassifiziert werden. Diese Strategien sollen nun untersucht werden.

In Abschnitt 7.1.4 wurde bereits angedeutet, daß sich bestimmte Präferenzbedingungen auf den Ebenen der unterschiedlichen Fokusstrukturen lokalisieren lassen. Diese fokusbasierten Präferenzbedingungen sollen nun näher diskutiert werden. Da sich auch hier Probleme betreffend die operationale Umsetzungen ergeben werden, soll darüberhinaus ein zweites Modell besprochen werden, das auf elementaren, vergleichsweise einfach algorithmisierbaren Faktoren basiert, die sich für die praktische Anapherninterpretation als nutzbringend erweisen werden.

7.2.1 Fokusbasierte Präferenzbedingungen

Präferenzbedingungen des globalen Fokus

Auf der Basis des globalen Fokus, der sich durch den Attentionalen Zustand (den *Right Frontier*) der Intentionalen Diskursstruktur definiert, lassen sich neben Restriktionen, die für bestimmte Typen von Pronomen wirksam sind, *Präferenzkriterien* ableiten, die die inhaltlichen Entitäten eines Diskurses vermöge der zwischen den Diskurssegmenten bestehenden Einbettungsrelationen *ordnen*. Zumindest zwei Strategien bieten sich an:

1. Für Kandidaten, die in unterschiedlichen Diskurssegmenten des *Right Frontier* realisiert sind: Bevorzuge denjenigen Kandidaten, dessen Segment in der Diskursstruktur tiefer eingebettet ist (vgl. Abbildung 7.5).
2. Für anaphorische Elemente, für die auch Antezedenten außerhalb des *Right Frontier* zulässig sind: Bevorzuge Kandidaten des *Right Frontier* gegenüber Kandidaten, die im übrigen Teil der Diskursstruktur lokalisiert sind.

Die Diskussion in Abschnitt 7.1.4 hat jedoch belegt, daß eine algorithmische Umsetzung dieser Regeln für unrestringierte Anwendungsszenarien auf der Grundlage gegenwärtig verfügbarer Techniken noch große Probleme aufwirft und allenfalls ansatzweise möglich ist.

Präferenzbedingungen des lokalen Fokus

Geht man für den Moment davon aus, daß die bezüglich der Centering-Theorie bestehenden Unterspezifikationen in sowohl theoretisch als auch algorithmisch adäquater Form resolviert sind, so lohnt sich eine nähere Betrachtung der Präferenzkriterien, die aus diesem Modell des lokalen Fokus ableitbar sind. Bereits zuvor wurde dafür plädiert, Regel R1 - wenn überhaupt - als *präferentielle* Strategie einzusetzen, da deren Nutzen als *Restriktion* im Rahmen von Systemen zur Sprach*interpretation* als fragwürdig anzusehen ist.

In den Arbeiten zur Centering-Theorie wird eine zweite Regel diskutiert, die sich als Auswahlheuristik für die Anapherninterpretation anbietet. Ausgangspunkt ist die Beobachtung, daß bestimmte *Muster* anaphorischen Wiederaufgriffs zu einem nach intuitivem Dafürhalten *kohärenteren* Text führen. Wird nun davon ausgegangen, daß ein Autor bzw. Sprecher die sprachlichen Ausdrücke zum Erreichen ihres Kommunikationsziels derart wählt, daß der Text für den Leser

	$Cb(U_{i+1}) = Cb(U_i)$	$Cb(U_{i+1}) \neq Cb(U_i)$
$Cb(U_{i+1}) = Cp(U_{i+1})$	CONTINUE	SHIFT-1
$Cb(U_{i+1}) \neq Cp(U_{i+1})$	RETAIN	SHIFT

Abbildung 7.6: Wiederaufgriffsrelationen der Centering-Theorie

bzw. Kommunikationspartner maximal verständlich ist, liegt der Nutzen für die Anapherninterpretation auf der Hand: Insofern die Muster algorithmisch erkennbar und unterscheidbar sind, können diejenigen Antezedens-Zuordnungen bevorzugt werden, die zu der relativ besten (lokalen) Kohärenz führen; diese Strategie läßt sich unter Rekurs auf die *Konversationspostulate* von Grice rechtfertigen, derzufolge Kommunikationspartner u.a. einer Maxime der Modalität folgen, die sich in einer entsprechenden sprachlichen Ausgestaltung des Texts niederschlägt.³²

Die Muster unterschiedlicher Kohärenz werden anhand der Wiederaufgriffsbeziehungen definiert, die innerhalb der Center-Strukturen zweier aufeinanderfolgender, nicht diskurs-initialer Äußerungen U_i und U_{i+1} eines Diskurssegments bestehen. Abbildung 7.6 zeigt das zweidimensionale Klassifikationsschema. Primäres Kriterium ist die Beibehaltung bzw. der Wechsel des rückwärtsgerichteten Centers; das sekundäre Kriterium ergibt sich durch die Differenzierung, ob der rückwärtsgerichtete Center $Cb(U_{i+1})$ durch den bevorzugten Center $Cf(U_{i+1})$ realisiert wird oder nicht.³³

Als in besonderem Maße kohärent werden nun solche Wiederaufgriffsbeziehungen angesehen, die den rückwärtsgerichteten Center, d.h. den lokalen Fokus beibehalten und diesen in der Position des bevorzugten Centers realisieren. Unter Zugrundelegung der verfeinerten Taxonomie von Brennan, Friedman und Pollard ist Regel R2 folgendermaßen definiert (vgl. Randbemerkung 33 sowie [BFP87]):

Regel 7.2 (R2 der Centering-Theorie)

Bevorzuge CONTINUE vor RETAIN vor SHIFT-1 vor SHIFT.

Der potentielle Nutzen von Regel R2 für die Anaphernresolution läßt sich anhand von folgender Variante des Beispiels (23) erläutern:

(28) *Kurt_i hat ein Saxophon. Er_i spielt Peter_j darauf vor. Er_i hat ihn_j eingeladen.*

Unter der oben skizzierten vorläufigen Annahme bezüglich des Ordnungskriteriums der vorwärtsgerichteten Center verkörpert das Vorkommen "Er" in U_2 sowohl $Cf(U_2)$ als auch $Cb(U_2)$; ent-

³²Diese Aussage ist *cum grano salis* zu lesen, da sie - zumindest in bezug auf das Kommunikationsziel der Informationsübermittlung - nicht für alle Textkorpora gilt. Ein Blick in einschlägige wissenschaftliche Abhandlungen sollte ausreichen, um sich hiervon zu überzeugen.

³³In der Originalarbeit [GJW83] werden nur die drei Klassen CONTINUE, RETAIN und SHIFT unterschieden; die feinere Differenzierung zwischen SHIFT-1 und SHIFT geht auf Brennan, Friedman und Pollard zurück ([BFP87]).

sprechend verkörpert die Okkurrenz “*Er*” in U_3 den $Cf(U_3)$. Gemäß R2 sind diese beiden Vorkommen zu koindexieren, denn nur auf diese Weise entstünde die zu bevorzugende CONTINUE-Beziehung; das umgekehrte Zuordnungsmuster

(29) *Kurt_i hat ein Saxophon. Er_i spielt Peter_j darauf vor. Er_j hat ihn_i eingeladen.*

ergäbe indes nur ein RETAIN, da der $Cb(U_3)$ nicht mit maximaler Prominenz realisiert wäre. Daß Regel R2 kein striktes Kriterium verkörpert, sondern nur eine Präferenzregel darstellt, die modulo Erfüllung aller anwendbarer Restriktionen anzuwenden ist, wird anhand von folgender Variante deutlich:

(30) *Kurt_i hat ein Saxophon. Er_i spielt Peter_j darauf vor. Peter_j bewundert ihn_i.*

In diesem Fall wird durch die explizite Verwendung des nichtpronominalen Ausdrucks “*Peter*” das durch R2 vorgeschlagene Wiederaufgriffsmuster CONTINUE ausgeschlossen; vermöge der transitiven Fernwirkung des bindungstheoretischen Prinzips B verbleibt für die Pronominalanapher “*ihn*” nurmehr der Kandidat “*Kurt_i*” (bzw. “*Er_i*”).

Es liegt also am Sprecher, durch eine angemessene Wahl der linguistischen Realisierungsformen sicherzustellen, daß die Kommunikationspartner die intendierten Fokusverschiebungen erkennen können. In den Griceschen Konversationsmaximen wird postuliert, daß dies tatsächlich geschieht. In Beleg (30) erscheint die RETAIN-Relation auch unter alleinigem Rekurs auf Weltwissen erkennbar; jedoch illustriert folgendes Beispiel, in dem beide Partizipanten von U_3 pronominal realisiert sind, daß der *kognitive Aufwand* für die Interpretation der beiden anaphorischen Ausdrücke ungleich höher ausfällt:³⁴

(31) *Kurt_i hat ein Saxophon. Er_i spielt Peter_j darauf vor. Er_j bewundert ihn_i.*

Die Vorhersagekraft von R2 begründet sich somit darin, daß Diskurs (30) die gegenüber Diskurs (31) aus konversationspragmatischen Gründen bevorzugte sprachliche Realisierungsform der intendierten Aussagefolge darstellt.

Diskussion: Präferenzen des lokalen Fokus für die Anapherninterpretation

Grosz, Joshi und Weinstein sehen in Regel R2 eine “*constraint on speakers*”, die folglich in erster Linie für Sprachgenerierungs-Systeme relevant sei. Für Anwendungen der Sprachinterpretation

³⁴Grosz, Joshi und Weinstein sprechen diesbezüglich von der “*inference load placed upon the hearer*” ([GJW95], S. 208). Weiter heißt es dort:

“*We conjecture that the form of expression in a discourse substantially affects the resource demands made upon a hearer in discourse processing and through this influences the perceived coherence of the discourse.*”

Es wird nunmehr klar, warum Centering als eine Theorie der lokalen *Kohärenz* bezeichnet wird, wenngleich die betrachteten Phänomene prinzipiell der Kohäsionsebene zuzurechnen sind.

Grosz, Joshi und Weinstein weisen allerdings auch darauf hin, daß es zum Zeitpunkt der Veröffentlichung ihrer Arbeit noch keine psycholinguistische Evidenz für diese Behauptung in bezug auf Regel R2 gab (ibid., S. 215).

- besonders der Anaphernresolution - sei Regel R1 als *Restriktion* zentral und R2 allenfalls von sekundärer Relevanz ([GJW95], S. 215):

“This rule does not have the same direct implementation for interpretation systems; rather it predicts that certain sequences produce a higher inference load than others.”

Wie bereits in Abschnitt 7.1.4 ausgeführt, gibt es gute Gründe dafür, auch Regel R1 als *Präferenz*-Kriterium aufzufassen (vgl. die zweite Variante von Beleg (23)). Darüberhinaus erscheint die Einordnung von R1 als Interpretations- und R2 als Generierungsregel ungerechtfertigt: Entsprechend dem zuvor Gesagten ist im Hinblick auf das Konversationspostulat der Modalität für *beide* Regeln davon auszugehen, daß sie von Textproduzenten - zumindest von solchen mit lauterer Absichten - befolgt werden. Aufgrund dieses Sachverhalts sind *beide* Kriterien als potentiell nützlich für die Interpretation anaphorischer Ausdrücke anzusehen; diese Einschätzung spiegelt sich in ersten Ansätzen einer algorithmischen Umsetzung der Centering-Theorie wider (Brennan, Friedman und Pollard, [BFP87]).

Darüberhinaus stellt sich die Frage nach dem relativen Nutzwert der beiden Regeln. Anhand der Beispiele (28) bis (31) wurde deutlich, daß R2 in der Tat eine Heuristik verkörpert, die insbesondere *im Falle multipler pronominaler Okkurrenzen* ein Interpretationsmuster nahelegt, dem kognitionstheoretische Plausibilität zukommt. Der Nutzen von Regel R1 hingegen erscheint auf den ersten Blick fragwürdig. Bezüglich Beispiel (23) ergäbe sich lediglich eine Voraussage für die *Sprachgenerierung*: Demnach wäre ein Sprecher an die pronominale Realisierungsform für den rückwärtsgerichteten Center gebunden, um einen Text von maximaler Kohärenz zu erzeugen. Interessant für die Anapherninterpretation wird es erst dann, wenn für das andere Pronomen *alternative Interpretationsmöglichkeiten außerhalb des vorwärtsgerichteten Centers der vorhergehenden Äußerung* bestehen:

(32) *Kurt_i hat eine Oboe_k. Er_i spielt Anja_j darauf vor. Sie_{j/k} gefällt Kurt_i.*

Regel R1 besagt, daß für das Pronomen “*Sie*” das Antezedens “*Oboe_k*” zu bevorzugen ist. Es erscheint indes fraglich, ob diese Vorhersage tatsächlich zur Wahl der intuitiv zu bevorzugenden Interpretation führt. Andererseits stellt sich die Frage, ob Konfigurationen wie die in (32) skizzierte überhaupt in realen Texten vorkommen, da sie auch als inhärent ambig aufgefaßt werden können - hier sei einmal mehr auf das Konversationspostulat der Modalität verwiesen.³⁵ Zudem kann die ohnehin fragwürdige Vorhersage von R1 durch disambiguierende selektionale Bedingungen des Verbs offenbar leicht überstimmt werden:

(33) *Kurt_i hat eine Oboe_k. Er_i spielt Anja_j darauf vor. Sie_j bewundert Kurt_i.*

Darüberhinaus ist selbst die graduelle, nichtstringente Lesart von R1, derzufolge die Realisierungsmuster mit einer pronominalen Realisierung des rückwärtsgerichteten Centers stets mit einem relativ kohärenteren Diskurs einhergehen als diejenigen Muster, die R1 verletzen, nicht

³⁵Tatsächlich zeigen Korpusstudien, daß Personalpronomen in dritter Person praktisch ausschließlich Entitäten spezifizieren, die satzintern oder in der unmittelbar vorhergehenden Äußerung realisiert sind (vgl. die Evaluation in Kapitel 11 sowie u.a. Lappin und Leass: [LaLe94]).

unkontrovers. Mit einem psycholinguistischen Experiment erbringen Smith und Chambers den Nachweis, daß elementare Beispiele existieren, für die diese Grundaussage der Centering-Theorie nicht haltbar ist ([SmCh96], S. 277 f.).

Schließlich lassen sich Belege angeben, in denen sich die Vorhersagen der Regeln R1 und R2 widersprechen:

- (34) *Kurt_i hat ein Saxophon. Er_i spielt Peter_j darauf vor.
Kurt_{i/jk} kennt ihn_{j/i} seit der Schulzeit.*

Während R2 einen CONTINUE-Übergang favorisiert, plädiert R1 (indirekt) für ein RETAIN. Aus den besprochenen Beispielen läßt sich somit folgern, daß in erster Linie Regel R2 als Auswahlheuristik für die Anapherninterpretation relevant ist. Es stellt sich daher die Frage nach der Algorithmisierbarkeit dieser Präferenzstrategie. Auch unter der Annahme, daß die übrigen Voraussetzungen für die Centering-Implementierung erfüllt sind - insbesondere die Verfügbarkeit der syntaktischen Funktion als Ordnungskriterium der *Cf*-Prominenzordnung -, ergeben sich Probleme: Die Identifikation der durch ein bestimmtes Interpretationsmuster induzierten Wiederaufgriffsrelationen bedingt u.a. die Bestimmung des dadurch *resultierenden* rückwärtsgerichteten Centers $Cb(U_{i+1})$ (vgl. Abbildung 7.6) - im Unterschied zu den übrigen Determinanten ($Cf(U_{i+1})$, $Cb(U_i)$ und $Cf(U_i)$) kann dieser Faktor nur unter der Voraussetzung ermittelt werden, daß bereits das komplette Wiederaufgriffsmuster bekannt ist; die Präferenzregel macht also eine Vorhersage über die Plausibilität der vollständigen *Menge* von Antezedens-Entscheidungen. Ohne Weiteres erscheint somit eine Übertragung dieser Regel auf die heuristische Bewertung von *Einzelentscheidungen* nicht möglich. Zwar erbringen Brennan, Friedman und Pollard den (trivialen) Nachweis der *prinzipiellen* Implementierbarkeit der Centering-Regeln, jedoch bedient sich deren Algorithmus einer ineffizienten Enumerationsstrategie, die zunächst alle erdenklichen Entscheidungskombinationen generiert und erst dann eine Auswahl entsprechend der Centering-Prädikationen (insbesondere R2) trifft ([BFP87]). Der Ansatz scheint daher ausschließlich von theoretischem Interesse zu sein; eine Implementierung sollte sich effizienterer Auswahlregeln bedienen, die sich auf Einzelentscheidungen beziehen und somit mit einem Rahmenalgorithmus verträglich sind, der ein sequentielles Auswahlverfahren betreibt.

Folglich stellt sich die Frage, inwieweit die lokale Fokussierungsregel R2 überhaupt in anwendungsorientierten Ansätzen zur Anapherninterpretation in sinnvoller Weise operationalisierbar ist. Es läßt sich nun allerdings nachweisen, daß die Vorhersagen von R2 vermöge einer Verfahrensweise heuristisch emuliert werden können, in der ausschließlich Bezug auf die statischen Determinanten $Cf(U_{i+1})$ und $Cf(U_i)$ sowie den bereits induktiv bekannten dynamischen Faktor $Cb(U_i)$ genommen wird.³⁶ Die Verfahrensweise verkörpert eine *sequentielle Variante R2'* zu *R2*, durch die Plausibilitätsvorhersagen für Einzelentscheidungen gemacht werden. Gegenstand der heuristischen Bewertung sind daher Paare der Gestalt (a_j^{i+1}, e_k^i) , wobei a_j^{i+1} ein zu interpretierendes anaphorisches Vorkommen (d.h. ein sprachlicher Ausdruck) der aktuellen Äußerung U_{i+1} darstellt und e_k^i das Spezifikat eines (unter Berücksichtigung der Restriktionen) als Antezedens-kandidat in Frage kommenden sprachlichen Ausdrucks a_k^i der Vorgängeräußerung U_i ist. Die Reproduktion der Vorhersagen von R2 durch R2' ist zu gewährleisten

³⁶Eine ergänzende Festlegung ("Induktions-Verankerung") ist für die Interpretation der Äußerung U_2 zu treffen, da für deren Vorgängeräußerung U_1 noch kein rückwärtsgerichteter Center definiert ist. Naheliegender wäre etwa die Gleichsetzung $Cb(U_1) := Cf(U_1)$; Walker schlägt als Alternative vor, die Centering-Präferenzen nicht diskurs(segment)initial einzusetzen und stattdessen auf ein anderes Verfahren zurückzugreifen ([Walk89]).

- durch eine geeignete *lokale Plausibilitätsordnung* der Kandidatenmenge eines bestimmten anaphorischen Ausdrucks a_j^{i+1} ;
- durch eine geeignete *globale Plausibilitätsordnung* der möglichen individuellen Entscheidungen (a_j^{i+1}, e_k^i) aller anaphorischen Ausdrücke a_j^{i+1} - dies ist wichtig, da es zu Interdependenzen zwischen den Einzelentscheidungen kommen kann (vgl. Abschnitt 7.1.2).

Gegeben sei eine Menge von Paaren (a_j^{i+1}, e_k^i) , für die eine Plausibilitätsbewertung durchzuführen ist, die den zuvor beschriebenen Zielsetzungen gerecht wird. R2' kann nun in Gestalt folgender einzelentscheidungsbezogener Bewertungskriterien formuliert werden:

1. *Primäres Kriterium*: Höchste Plausibilität erhalten Paare mit $e_k^i = Cp(U_i)$, d.h. der Aufgriff des *bevorzugten Centers* von U_i .
Unterkriterium: Feinabstufung der Plausibilität gemäß der Einordnung von a_j^{i+1} - präziser: der künftigen Interpretation³⁷ von a_j^{i+1} - in $Cf(U_{i+1})$.
2. *Sekundäres Kriterium*: Nächsthöchste Plausibilität erhalten Paare mit $e_k^i = Cb(U_i)$, d.h. der Aufgriff des *rückwärtsgerichteten Centers* von U_i .
Unterkriterium: Feinabstufung der Plausibilität gemäß der Einordnung von a_j^{i+1} - präziser: der künftigen Interpretation von a_j^{i+1} - in $Cf(U_{i+1})$.
3. *Tertiäres Kriterium*: a_j^{i+1} realisiert den *bevorzugten Center* von U_{i+1} .
Unterkriterium: Feinabstufung der Plausibilität gemäß der Einordnung von e_k^i in $Cf(U_i)$.

Es kann nun gezeigt werden, daß dieser Kriterienkatalog in allen relevanten Fällen das Gewünschte leistet (vgl. Stuckardt, [Stu96c]). Für jedes Paar von Übergangsrelationen $\{t_1, t_2\}$ mit $t_1 \neq t_2$ ist der Nachweis zu erbringen, daß die "sequentielle" Befolgung von R2' zu derselben Vorhersage führt wie deren "paralleles" Pendant R2. Der Nachweis läuft über eine Fallstudie, in deren Rahmen Äußerungen mit multiplen pronominalanaphorischen Vorkommen betrachtet werden, für die Interpretationsalternativen bestehen, die entweder zu Muster t_1 oder zu Muster t_2 führen. Anhand von Beleg (28) kann etwa veranschaulicht werden, daß das Unterkriterium des primären Kriteriums die Bevorzugung des CONTINUE-Übergangs gegenüber der RETAIN-Alternative bewirkt.

Jedoch ist eine weitere Differenzierung in den betrachteten Fallbeispielen notwendig: Die Vorhersagen von R2' hängen *auch* davon ab, ob der rückwärtsgerichtete Center $Cb(U_i)$ mit dem bevorzugten Center $Cf(U_i)$ zusammenfällt oder nicht - etwa im Falle eines vorhergegangenen RETAIN. Gilt letzteres, so sind tatsächlich Abweichungen gegenüber R2 zu beobachten: (Die Unterstreichungen kennzeichnen die rückwärtsgerichteten Center der unterschiedlichen Äußerungen.)

(35) *Kurt_i ist musikalisch. Er_i spielt Saxophon. Peter_j bewundert ihn_i.
Er_{j/i} wohnt im selben Stadtteil.*

³⁷Da die Ordnung des Cf gemäß den Festlegungen der Centering-Theorie primär in der syntaktischen Funktion verankert ist, sollte die Einordnung ohne Ansehen der künftigen Interpretation möglich sein. Ruft man sich indes die aus theoretischer Sicht bestehende Interdependenz zwischen Oberflächenstrukturanalyse und referentieller Interpretation in Erinnerung, so offenbaren sich weitere Schwierigkeiten des Centering-Modells.

R2' erzeugt nun qua primärem Kriterium einen Übergang vom Typ SHIFT-1, obwohl R2 die ebenfalls mögliche CONTINUE-Interpretation vorschreibt.

Es stellt sich allerdings die Frage, ob die abweichende Vorhersage von Regel R2' tatsächlich als Mangel anzusehen ist, der deren Geeignetheit für die sequentielle Emulation der Originalregel R2 in Frage stellt. Gemäß intuitivem Dafürhalten könnte es durchaus naheliegen, für Beleg (35) die Vorhersage von R2' als die zu bevorzugende Interpretationsmöglichkeit anzusehen.³⁸ Eine vergleichbare Beobachtung wird bereits von Brennan, Friedman und Pollard gemacht ([BFP87], S. 160):

“Another question is whether the preference ordering of transitions [...] should always be the same. For some examples, particularly where U_n contains a single pronoun and U_{n-1} is a retention, some informants seem to have a preference for shifting, whereas the centering algorithm chooses a continuation ([...]).”

Auch Grosz, Joshi und Weinstein diskutieren den (weiterführenden) Vorschlag, *kontextsensitive Verfeinerungen von R2* zu untersuchen, in denen *Sequenzen* von Wiederaufgriffsrelationen - d.h. insbesondere der *unmittelbar vorhergehende Übergangstyp* - die Grundlage der Plausibilitätsaussage darstellen. Beispiel (35) scheint daher weniger ein Problem der sequenzialisierten Variante R2' aufzuzeigen, sondern verkörpert wohl eher einen Indikator für die *prinzipiellen Grenzen der originären Centering-Theorie*.

Regel R2' stimmt also mit ihrem Original in den meisten Fällen überein; Abweichungen ergeben sich ausschließlich dann, wenn ein RETAIN vorausgeht. Zumindest in zwei dieser drei Fälle gibt es ohnehin Anhaltspunkte dafür, daß die abweichende Vorhersage von R2' mit der intuitiv zu bevorzugenden Interpretation zusammenfällt. Der einzig verbleibende Problemfall entsteht für die Alternative RETAIN vs. SHIFT, die jedoch von ihrer Konstellation her eher selten zum Tragen kommen sollte.

Bei einer nochmaligen Betrachtung der konzeptuellen Ausrichtung der Centering-Regeln offenbaren sich vielmehr Probleme grundsätzlicherer Art. Gemäß ihrem Anspruch stellt R2 eine *Heuristik* dar, die - wie etwa anhand des Belegs (31) deutlich wurde - leicht durch andere semantische bzw. pragmatische Faktoren überstimmt werden kann. Es scheint daher wenig sinnvoll, nach einer weiteren Verfeinerung der sequentiellen Emulation von R2 Ausschau zu halten, da die Schwierigkeiten ohnehin tiefer liegen und die Centering-Theorie *in toto* betreffen.

Wenn allerdings die profan anmutende Sequentialisierung R2' bereits für die wesentlichen Fälle das operational umsetzt, was in R2 deklarativ beschrieben wird, stellt sich die grundsätzliche Frage, *ob* die in der Centering-Theorie formulierten Präferenzen überhaupt den Kern von algorithmischen Verfahren zur Anapherninterpretation bilden sollten. In den weitaus meisten Fällen zeichnet das primäre Kriterium von R2' bzw. dessen Unterkriterium verantwortlich für die Übereinstimmung der Vorhersage mit R2. Unter der bislang unterstellten Annahme, daß der bevorzugte Center *Cb* durch das Subjekt der Äußerung realisiert wird, kann das primäre Kriterium als simple Heuristik der Subjekt-Präferenz umschrieben werden; das Unterkriterium sorgt für die bevorzugte Kospezifikation mit anaphorischen Ausdrücken, die ebenfalls die syntaktische Rolle des Subjekts innehaben. Somit basieren die drei Kriterien von R2' einzig auf Heuristiken, die in einer (noch nicht theoretisch abgesicherten) Hierarchie syntaktischer Funktionen verankert sind. Die tiefergehende Frage, *wie* die hier angenommene Prominenzordnung auf möglichst

³⁸Diese Aussage wäre durch psycholinguistische Untersuchungen bzw. Korpusstudien zu untermauern: Wählt ein Sprecher in Beispielen, in denen die selektionalen Restriktionen des Verbs ausschließlich eine von der R2'-Vorhersage abweichende Interpretation erlauben, eine nicht-pronominale Realisierung für den Nicht-*Cf*-Mitspieler?

adäquate (und bevorzugt algorithmisierbare) Weise definiert werden sollte, bleibt größtenteils unbeantwortet.

Es liegt somit nahe, die Perspektive zu wechseln und die Frage in den Vordergrund zu stellen, welche *Faktoren* die *Prominenz* eines Antezedenskandidaten (ggf. relativ zur betrachteten Anapher) determinieren. Ein solches Vorgehen erscheint angemessen, da es das Problem der Antezedensauswahl von der Wurzel her angeht und auf diese Weise erst die Grundlage für eine *algorithmische* Umsetzung fokusbasierter Strategien schafft. Ein lokaler Fokus - etwa in der Gestalt des *Cb* - fällt eben nicht vom Himmel, sondern wird erst *determiniert* durch die Vorgaben der Präferenzordnung, und stellt damit eine abhängige Größe dar. Folglich bietet es sich an, von einem theoretischen Ansatz auszugehen, in dem die *unterschiedlichen Determinanten der Prominenzordnung* den primären Bezugspunkt bilden. Ein adäquates Verarbeitungsmodell wird von Alshawi vorgeschlagen ([Alsh87]). Dieses Modell multipler Präferenzfaktoren liegt einem Großteil der algorithmischen Verfahren zur Anapherninterpretation zugrunde; Lappin und Leass etwa weisen explizit auf diese Grundlage hin ([LaLe94], S. 558, Hervorhebungen im Original):

“In general, we agree with Alshawi (1987, p. 62) that an algorithm/model relying on the relative salience of all entities evoked by a text, with a mechanism for removing or filtering entities whose salience falls below a threshold, is preferable to models that ‘make assumptions about a single (if shifting) focus of attention.’ ”

Unter der ergänzenden Anmerkung, daß letztlich auch der Fokus als operational zu bestimmendes Element erst durch “*Salience*”-Faktoren *determiniert* wird (vgl. R2¹), stimmt diese Sichtweise mit der hier vertretenen überein.

Es sollen daher abschließend einige elementare Faktoren identifiziert werden, die den Grad der Prominenz der inhaltlichen Entitäten kodeterminieren. Einzelnen betrachtet können diese Faktoren als individuelle Präferenzstrategien aufgefaßt werden, die aber erst in ihrer Gesamtheit - qua kumulierter Prominenz-Bewertung - den Ausschlag für die Auswahl eines bestimmten Antezedens geben. Es wird deutlich werden, daß sich zumindest einige der in der Literatur vorgeschlagenen Faktoren auf die ein oder andere Weise in den Prädikationen der Centering-Regel R2 widerspiegeln.

7.2.2 Elementare Präferenzbedingungen

Prominenzfaktoren im engeren Sinne

Unter den Prominenzdeterminanten im engeren Sinne sollen diejenigen Faktoren verstanden werden, die die Prominenz eines Diskursreferenten auf eine Weise bestimmen, die unabhängig von dessen Plausibilität relativ zu einer *bestimmten* zu resolvierenden Anapher ist. Sie ergeben sich vermöge der Menge *aller* linguistischen Realisierungen.³⁹ U.a. erweisen sich die folgenden

³⁹Hier ist erneut die theoretisch adäquate Sichtweise ausschlaggebend, derzufolge nicht bestimmte Vorkommen (linguistische Ausdrücke) aufgegriffen werden, sondern inhaltliche Entitäten, für die im allgemeinen mehrere Realisierungen an der sprachlichen Oberfläche existieren, die in ihrer *Gesamtheit* die Prominenz der Entität determinieren. Somit wird ein weiterer Unterschied zur *Cf*-Prominenzordnung der Centering-Theorie ersichtlich, die sich lediglich auf die linguistischen Eigenschaften einer *einzigsten* sprachlichen Realisierung - nämlich derjenigen in der aktuellen Äußerung - begründet. Insofern erscheint das Alshawi-Modell feinkörniger und intuitiv plausibler.

Faktoren als relevant:

1. *Distanz*: Je weniger weit eine bestimmte sprachliche Realisierung des Diskursreferenten zurückliegt, desto mehr trägt sie zu dessen Prominenz bei;
2. Präferenz von qua Position in der syntaktischen Oberflächenstruktur *hervorgehoben*, d.h. *topikalisierten* Vorkommen;
3. Präferenz von Vorkommen in der syntaktischen Funktion des *Subjekts*;
4. (allgemeiner) Abstufung entsprechend einer *Hierarchie syntaktischer Funktionen*, z.B. Subjekt > direktes Objekt > indirektes Objekt > PP-Objekt.

Der Prominenzfaktor der *Distanz* kann u.a. als elementares Substitut für eine inhaltsorientierte, in der Präferenzordnung des globalen Diskursfokus verankerten Strategie angesehen werden. Die *Topikalisierungspräferenz* trägt Belegen wie folgendem Rechnung, in denen sich die Fokussierung eines Diskursreferenten in der syntaktischen Struktur niederschlägt.

(36) *Es war Gropius, der am Bauhaus zu Weimar lehrte.*

Die beiden übrigen Determinanten finden sich bereits in der Centering-Theorie bzw. speziell in der Operationalisierung R2' von R2 wieder: Das primäre Kriterium bringt eine *Subjektpräferenz* (jedoch nur des Subjekts der unmittelbar vorhergehenden Äußerung) zum Ausdruck; in den Unterkriterien spielt jeweils eine *Hierarchie syntaktischer Funktionen* hinein.

Unter der Annahme, daß syntaktische Rollen zur Verfügung stehen, ist die Implementierung der zuvor beschriebenen elementaren Faktoren vergleichsweise trivial. Da es im Unterschied zur Centering-Theorie jedoch nicht mehr nur *eine* Präferenzregel gibt, bestehen Freiheitsgrade in bezug auf die *relative Wertigkeit* der einzelnen Kriterien. Somit entsteht die Aufgabe, entsprechende *numerische Gewichte* festzulegen. Dies ist keineswegs als Nachteil anzusehen, da das Faktorenmodell insgesamt wesentlich feinkörniger ist und somit Differenzierungen ermöglicht, die im Centering-Modell nicht zum Ausdruck gebracht werden können.

Relative Prominenzfaktoren

Eine weitere Gruppe von Prominenzdeterminanten charakterisiert nicht die Prominenz eines Diskursreferenten im engeren Sinne, sondern ist *auf der Ebene der linguistischen Realisierungen* angesiedelt. Die Faktoren sind als *relativ* zu bezeichnen, weil sie den Grad der Plausibilität eines *bestimmten* Vorkommens des Diskursreferenten als Antezedens einer *bestimmten* Anapher determinieren. Es sind also jeweils nur die Eigenschaften der sprachlichen Realisierung eines konkreten Vorkommens einer inhaltlichen Entität relevant.

Das wichtigste Beispiel eines relativen Prominenzkriteriums stellt die *Rollenrätigkeits-Heuristik* dar, wodurch solchen Antezedenskandidaten eine höhere Plausibilität zugewiesen wird, die dieselbe *Rolle* ausfüllen wie die zu resolvierende Anapher. Je nachdem, welcher Rollenbegriff zugrundegelegt wird, ergeben sich unterschiedliche Spielarten, die allesamt als Anaphernresolutionsstrategien mit heuristischem Wert anzusehen sind.

(37a) *Behrens_i besucht den Architekten Gropius_j. Er_i zeigt ihm_j die Entwürfe.*

(37b) *Behrens_i zahlt seiner_i Assistentin_j eine Gratifikation_k.*

Er_i bedenkt sie_j jedes Jahr an Weihnachten.

Die einfachste Variante ist der *syntaktische Parallelismus*, durch den diejenigen Kandidaten in derselben *syntaktischen Funktion* bevorzugt werden. Für Beleg (37a) ergibt sich die Vorhersage, daß "Behrens" als Subjekt der initialen Äußerung ein besonders plausibler Kandidat der Pronominalanapher "Er" in der Subjektrolle der zweiten Äußerung ist - im Unterschied zur Vorhersage des Faktors der Subjektpräferenz besteht diese Präferenz ausschließlich relativ zur Anapher "Er", nicht jedoch relativ zum anaphorischen Vorkommen "ihm", das die Rolle des indirekten Objekts ausfüllt. Für letzteres Vorkommen ergibt sich auf der Grundlage der syntaktischen Funktion keine Präferenzaussage, da das andere in Frage kommende Antezedens "Architekt Gropius" ein direktes Objekt darstellt.

Die Güte der Vorhersage wird unter anderem durch den Grad der Übereinstimmung der Oberflächenstruktur der zugehörigen Äußerungen sowie durch die Existenz bestimmter Satzkonnektive determiniert, die deren Parallelität betonen (Smith, [Smit94]). Es ist jedoch klar, daß auf der Basis der syntaktischen Rolle nicht immer die richtige Präferenzentscheidung getroffen wird. Die Rollenträgheitsheuristik kann unter Rekurs auf *thematische Rollen* zumindest verfeinert werden. Für Beleg (37b) würde auf dieser inhaltsorientierten Grundlage die korrekte Entscheidung getroffen: Als *Begünstigter* einer "zahlen"-Situation ist der Kandidat "Assistentin" die zu bevorzugende Wahl für die Anapher "sie"; das syntaktische Kriterium führte hier zur Fehlentscheidung "Gratifikation".⁴⁰

Die Rollenträgheitsregeln erscheinen *relevant*, da in vielen Fällen multipler pronominaler Vorkommen auf alleiniger Grundlage der Subjektpräferenz keine eindeutige Präferenzaussage gegeben ist. Betreffend die Frage der *algorithmischen Unsetzbarkeit* gilt das zuvor Gesagte: Syntaktische Funktionen stehen modulo der Verfügbarkeit einer robusten Syntaxanalyse zur Verfügung; auch die thematischen Rollen erschließen sich auf dem gegenwärtigen Stand der Technik mit der notwendigen Reichweite und Granularität nur partiell.

Darüberhinaus stellt sich die Frage nach dem Verhältnis zwischen den Präferenzaussagen der Centering-Theorie und denen der Rollenträgheits-Heuristiken. Ein Blick auf das primäre Kriterium der Centering-Regel R2' zeigt, daß die Vorhersage der *syntaktischen* Rollenträgheitsheuristik zumindest für die Subjektrolle reproduziert wird;⁴¹ hingegen können sich Abweichungen für Belege ergeben, in denen ein Kandidat in Subjekt-Rolle nur durch ein Nicht-Subjekt-Pronomen wiederaufgreifbar ist: Gemäß dem priären Kriterium von R2' wird diese Zuordnung gegenüber einer eventuell ebenfalls möglichen rollenträgen Interpretation der Pronominalanapher bevorzugt. Die in zumindest einigen Arbeiten vertretene These, wonach die Vorhersagen der (syntaktischen) Parallelismusregel durch die Vorhersagen der Centering-Theorie subsumiert werden - vgl. etwa Brennan, Friedman und Pollard ([BFP87], S. 157)

"[...], structural parallelism is a consequence of our ordering [of] the Cf list by grammatical function and the preference for continuing over retaining."

⁴⁰Carbonell und Brown schlagen eine weitere, pragmatische Spielart der Rollenträgheitskriteriums vor, durch das solche Interpretationen anaphorischer Elemente bevorzugt werden, die zu einer "inertia of the underlying action" führen ([CaBr88], S. 97). Auf eine nähere Besprechung dieses Kriteriums soll verzichtet werden, da dessen Algorithmisierung die Verfügbarkeit von Weltwissen voraussetzt und sich somit auf dem gegenwärtigen Stand der Technik problematisch gestaltet.

⁴¹Es liegt erneut die Annahme zugrunde, daß die Strukturen Cf des vorwärtsgerichteten Center entsprechend einer syntaktischen Rollenhierarchie geordnet sind, an deren Spitze die Subjektfunktion steht.

erscheint somit unhaltbar. Tatsächlich gibt es psycholinguistische Evidenz für die kognitive Bedeutsamkeit oberflächenstruktureller Parallelitätseffekte, die nicht alleine in bezug auf die Subjektrolle, sondern insbesondere auch für die Objektrolle nachgewiesen werden können (vgl. die Experimente von Smith und Chambers, [SmCh96]). Somit ist zu folgern, daß das Centering-Modell alleine zu kurz greift, da die lokale Kohärenz nicht nur durch Faktoren der diskurssemantischen Ebene, sondern zusätzlich durch Determinanten bestimmt wird, die auf der oberflächenstrukturellen Ebene angesiedelt sind (ibid., S. 279):

“Given these results, it is reasonable to reject the account of attentional focusing advanced by centering as overly simplistic and unsatisfactory.”

Letztendlich ist also schwer zu sehen, was monolithische Fokusmodelle wie die Centering-Theorie von ihrem algorithmischen, prozeduralen Kern her mehr leisten als die simple Heuristik der Subjektpräferenz. Somit ergibt sich ein weiteres gewichtiges Argument für ein Modell der Kandidatenauswahl, das auf einzelnen Prominenzfaktoren basiert, zu denen sowohl die Subjektpräferenz als auch Parallelitätskriterien zählen.

Eine weitere Prominenzheuristik ergibt sich in Gestalt der defensiven Variante der in Abschnitt 7.1.5 motivierten Restriktion kataphorischen Wiederaufgriffs nichtdefiniter Postzedenten. Dieses auch als *Kataphernmalus* bezeichnbare Kriterium ist *relativ*, weil auf die Oberflächenanordnung von Anapher und Antezedenskandidaten Bezug genommen wird: Steht der Kandidat rechts von der Anapher, so ist dessen Plausibilität zu *reduzieren*. Der Hintergrund dieses Kriteriums ergibt sich aus der in Abschnitt 7.1.5 identifizierte Definitheitsanforderung. Aus den dort genannten Gründen sollte eine nichtstringente Implementierung als Präferenzkriterium gewählt werden, da eine allgemeine algorithmische Verifikation der Definitheitsanforderung problematisch ist und darüberhinaus der Fall einer abgeleiteten Spezifikation gegeben sein kann, in dem kein weiteres *kospezifizierendes* Antezedens im Diskurs existiert. Die Implementierung dieser Heuristik ist trivial; aufgrund ihrer Nützlichkeit zur Reduktion des Suchraums findet sie in vielen operationalen, anwendungsorientierten Ansätzen zur Anapherninterpretation Verwendung.

7.3 Zusammenfassung

Abbildung 7.7 faßt die Ergebnisse der Untersuchungen zusammen. Die Abfolge der Tabellenzeilen entspricht der obigen Diskussionsreihenfolge der Restriktionen und Präferenzheuristiken. Für jedes Kriterium werden die drei wesentlichen Punkte aufgeschlüsselt: einfließende *Ressourcen*, *Relevanz* und *Algorithmisierbarkeit* auf dem gegenwärtigen Stand der Technik. Folgende Ressourcen werden unterschieden: OBERF (Zeichenketten an der sprachlichen Oberfläche sowie deren Anordnung), MOLEX (morphologische Analyse und Lexikon, umfassend u.a. Stammformen, Wortarten, ggf. Genus), SELEX (lexikalisch-semantische Information), SYNPAR (Analysemodul zur Bestimmung der syntaktischen Oberflächenstruktur und die zugehörige Grammatik), SEMWI und PRAWI (semantische bzw. pragmatisches Kontextwissen sowie entsprechende Inferenzmechanismen).

Ressourcen, die im allgemeinen einfließen, sind mit “+” gekennzeichnet; solche, die nur in wenigen Fällen einfließen bzw. auf deren Verfügbarkeit ohne entscheidende Einbußen verzichtet

STRATEGIE	RESSOURCEN						BEWERTUNG	
	OBERF	MOLEX	SELEX	SYNPA	SEMWI	PRAWI	RELEV	ALGOR
MORKONGRU		+		•	•		⊕ S	⊕/⊖ S
SYNKONFIG				+			⊕	⊖
SEMSKOPUS				+	+	+	⊖	⊖
INT-FOKUS				+	+	+	⊖ (?) T	⊖
LOK-FOKUS				+			⊖	⊖
DEFKATAPH	+				•		⊖ (?)	⊕/⊖
SEMROLLEN			+	+	•	•	⊖ (?)	⊖/⊖
PRAE-POST					+	+	⊖	⊖
INT-FOKUS				+	+	+	⊖ (?) T	⊖
LOK-FOKUS				+			⊕	⊖
DISTANZ	+						⊕	⊕
TOPIKALIS				+			⊖	⊖
SUBJEKT				+			⊕	⊖
SYNROLLE				+			⊖ (?)	⊖
SYNPARALL				+			⊕	⊖
SEMPARALL			+	+	•		⊕	⊖
KATAMALUS	+						⊕	⊕

Abbildung 7.7: Ressourcen, Algorithmisierbarkeit und Relevanz der diskutierten Strategien

werden kann, tragen die Markierung “•”.⁴² Um die Darstellung überschaubar zu halten, werden offensichtliche Subsumptionsverhältnisse zwischen Ressourcen in der Tabelle nicht expliziert.⁴³ Die Einstufung von Algorithmisierbarkeit (ALGOR) bzw. Relevanz (RELEV) geschieht graduell: “⊕” = “*problemlos*” bzw. “*hoch*”, “⊖” = “*unter bestimmten Voraussetzungen*” bzw. “*mittel*”, “⊖” = “*auf dem gegenwärtigen Stand der Technik problematisch*” bzw. “*gering*”. Eine Einstufung ist mit “(?)” gekennzeichnet, falls die Angabe auf einer vorläufigen Schätzung basiert, die (ggf.) per Evaluation zu verifizieren ist; “S” bzw. “T” steht für Sprach- bzw. Anapherentyp-Abhängigkeit. Mehrfachangaben beziehen sich auf Fälle unterschiedlicher Schwierigkeit, die i.d.R. mit einem entsprechenden fakultativen Ressourcenbedarf einhergehen.⁴⁴

Ist die Algorithmisierbarkeit einer Strategie als voraussetzungsabhängig (“⊖”) klassifiziert, so bedeutet dies, daß es aufgrund der einfließenden Ressourcen zu Fällen unvollständiger Information kommen kann. Voraussetzung für die Implementierung derartiger Strategien ist die Entwicklung geeigneter Techniken robuster Verarbeitung, die beispielsweise durch *Heuristiken* oder vermöge einer *Generalisierung der Kriterien auf partielle Ergebnisse* bewerkstelligt werden kann. Ein Blick auf Abbildung 7.7 zeigt, daß dieser Fall nahezu ausschließlich dann gegeben ist, wenn eine Strategie auf der Verfügbarkeit der syntaktischen Oberflächenstruktur, d.h. auf der Parsing-Ressource SYNPA aufbaut.

Die Tabelle erlaubt es, diejenigen Kriterien zu identifizieren, die einerseits von hoher Relevanz für die praktische Anaphernresolution und andererseits auf dem Stand der Technik zumindest bedingt operationalisierbar sind. Die Voraussetzungen zur Festlegung eines algorithmischen

⁴²Beispielsweise basiert die Restriktion der morphologischen Kongruenz MORKONGRU primär auf den Ergebnissen einer morphologischen Analyse; jedoch *kann* - wie in Abschnitt 7.1.1 aufgezeigt - der syntaktische bzw. semantische Kontext zusätzliche Information zur Disambiguierung beisteuern, deren Fehlen jedoch nicht zur Nichtanwendbarkeit der Restriktion führt.

⁴³Z.B. basiert die Syntaxanalyse (SYNPA) in aller Regel auf einer morphologischen Voranalyse (MOLEX). Die Eigenschaften an der sprachlichen Oberfläche (OBERF) bilden den trivialen Ausgangspunkt jeglicher tiefergehenden Analyse.

⁴⁴Als Beispiel sei erneut auf die Strategie MORKONGRU verwiesen, deren Algorithmisierbarkeit erst im Falle syntaktisch- oder semantisch-kontextueller Disambiguierung potentiell schwierig wird.

Strategie-Mixes im Sinne von Carbonell und Brown sind also nunmehr gegeben. Es zeigt sich, daß eventuelle Einschränkungen in der Algorithmisierbarkeit von hochrelevanten Strategien beinahe ausnahmslos auf die Abhängigkeit von der Ressource SYNPA (oberflächenstrukturelles Parsing) zurückzuführen sind. Es steht zu erwarten, daß das Ziel einer robusten Verarbeitung für die in Abschnitt 7.1.2 beschriebenen syntaktisch-konfiguralen Restriktionen am schwierigsten zu erreichen sein wird, da auf die Struktur *in toto* und nicht nur auf bestimmte Merkmale wie etwa die syntaktische Rolle Bezug genommen wird. Hieraus ergibt sich der Schwerpunkt der folgenden Untersuchungen.

Bevor jedoch mit der Komposition eines anwendungsadäquaten Algorithmus zur Anapherninterpretation fortgefahren wird, sollen zunächst die wichtigsten in der Literatur beschriebenen Algorithmen und Verfahren analysiert werden. Dies schafft die Voraussetzungen für eine bessere Einordnung des verfolgten Ansatzes sowie für die Einschätzung der Relevanz der zu entwickelnden Strategien zur robusten referentiellen Interpretation.

Kapitel 8

Anapherninterpretationsverfahren

Im Rahmen der Betrachtungen zu Komponentenstrategien der Anapherninterpretation wurden grundlegende Beobachtungen betreffend Relevanz und Algorithmisierbarkeit gemacht. Dies schafft nicht nur die Voraussetzung zur eigenständigen Entwicklung anwendungstauglicher Verfahren, sondern ermöglicht darüberhinaus eine nähere Einordnung sowie die Bewertung der in der Literatur beschriebenen *klassischen Ansätze*. Das primäre Ziel besteht in der Offenlegung von *Unterspezifikationen* sowie *impliziten Annahmen* bezüglich der zur Verfügung stehenden Ressourcen (vgl. auch Walker, [Walk89]). Folgende Fragestellungen erweisen sich als relevant:

1. Welche *Strategien* - und damit: welche *Ressourcen* - finden Verwendung?
2. Wie steht es um den *Abdeckungsgrad* des Verfahrens, d.h.:
 - (a) Welche *Anapherntypen* und *Spezifikationsrelationen* werden berücksichtigt?
 - (b) Wie *relevant* sind die eingesetzten Strategien?
 - (c) Gibt es Beschränkungen der *inhaltlichen Domäne* oder des *Textgenres*?
3. Inwieweit ist die *Algorithmisierbarkeit* des Ansatzes gewährleistet?
4. Falls anwendbar: Wie ist es um die *Interpretationsleistung* von Implementierungen des Verfahrens bestellt, d.h. liegen Ergebnisse einer formalen Evaluation vor?

Diese Punkte sollen unter Rekurs auf die in Abbildung 7.7 vorliegende tabellarische Zusammenfassung der Strategieeigenschaften für die bekanntesten Ansätze zur Anapherninterpretation diskutiert werden. Die Ergebnisse der Untersuchung werden es ermöglichen, die im Hinblick auf die Anwendungszielsetzung bestehenden Defizite der klassischen Verfahren zu identifizieren. Die Analyse der Anaphernresolutionsverfahren geschieht in chronologischer Reihenfolge.¹

¹Der Aufstellung liegt nicht der Anspruch der Vollständigkeit zugrunde; vielmehr wurde eine Auswahl getroffen, die die *Meilensteine* in der Erforschung algorithmischer Verfahren der Anapherninterpretation umfaßt.

8.1 Der “naive” Algorithmus von Hobbs

Mitte der siebziger Jahre entwickelte Hobbs ein Verfahren zur Interpretation von Pronominalanaphern, das an frühe Arbeiten von Winograd² anknüpft und auf syntaktisch-konfiguralen Koindexierungsbeschränkungen aufbaut ([Hobb78]). Grundvoraussetzung ist die Verfügbarkeit *eindeutiger* und vollständiger Beschreibungen der syntaktischen Oberflächenstruktur für alle Sätze des zu interpretierenden Texts. Des Weiteren wird davon ausgegangen, daß geeignete Algorithmen für die *Resolution von Ellipsen* sowie für die *Identifikation verteilter Antezedenten* zur Verfügung stehen (vgl. Abschnitt 6.1, Beispiele (5) bzw. (10)).

Der Algorithmus ist auf die Behandlung von Personalpronomen in dritter Person - anscheinend inklusive Possessiva - mit *Objektbezug*³ spezialisiert. Für jedes Pronomen wird, ausgehend von dessen Position im Syntaxbaum, zunächst nach einem *intrasententiellen* Antezedens gesucht, indem der Pfad zur Wurzel schrittweise zur Wurzel zurückverfolgt wird und in benachbarten Zweigen nach morphologisch (bezüglich Numerus und Genus) sowie syntaktisch-konfigural zulässigen Kandidaten Ausschau gehalten wird. Der erste Kandidat, der diese Kriterien erfüllt, wird gewählt. Sollte die intrasententielle Suche erfolglos verlaufen, so wird auch in den vorausgehenden Sätzen gesucht, wobei die Suchreihenfolge wiederum durch den Oberflächenstrukturbaum determiniert wird und der erste morphologisch passende Kandidat ausgewählt wird (ibd., S. 315 f.).

Wichtig sind weniger die Details als vielmehr die grundlegenden Ideen des Verfahrens. Die Kernarbeit der Kandidatenelimination leisten die morphologische Kongruenzrestriktion sowie zwei syntaktisch-konfigurale Kospezifikationsbedingungen, denen vergleichbare Ideen wie den Filtern der in Abschnitt 7.1.2 diskutierten Bindungstheorie Chomskys zugrundeliegen. Hobbs selbst bezeichnet die Basisversion seines Algorithmus als (engl.) “*naive*”, da bestimmte syntaktisch-konfigurale Sonderfälle nicht adäquat behandelt werden und darüberhinaus der Mangel besteht, daß keine semantischen Kriterien zum Einsatz kommen, die es ermöglichen, problematische Entscheidungen zu vermeiden - der Algorithmus liefert in jedem Fall eine Lösung (ibd., S. 324).

Trotz dieser bereits von Hobbs hervorgehobenen Inadäquatheiten verkörpert der *naive Algorithmus* einen Meilenstein in der Erforschung algorithmischer Verfahren zur Anapherninterpretation. Durch eine Anwendung seines Ansatzes auf per Hand disambiguierte Syntaxbäume erbringt Hobbs den Nachweis, daß im allgemeinen (unter den getroffenen Annahmen) *gute Ergebnisse* erzielt werden: Das Basisverfahren erzielt eine Erfolgsquote von 88.3 % für Dritte-Person-Pronomen, die Objekte spezifizieren.⁴ Die nicht abgedeckten problematischen Fälle sind in der Praxis äußerst selten; auf semantische Kriterien kann in der Mehrzahl der Fälle ebenfalls verzichtet werden, da im näheren Kontext nur ein Kandidat existiert, der die syntaktischen und morphologischen Restriktionen erfüllt. Hobbs kommt zu der folgenden, bis heute nicht zu revidierenden Einschätzung (ibd., S. 324):

²Vgl. die Arbeit [Wino72], in deren Rahmen der erste Ansatz einer anwendungsorientierten, algorithmischen Anapherninterpretation entwickelt wird.

³Okkurrenzen von “*it*”, die relationale Entitäten spezifizieren, werden nicht unterschieden und somit nicht adäquat behandelt.

⁴Darüberhinaus beschreibt Hobbs eine Verfeinerung des Verfahrens, in deren Rahmen selektionale Restriktionen eingesetzt werden. Hierdurch erhöhe sich Korrektheit auf 91.7 %, jedoch werden wesentliche Details betreffend den Bezugsrahmen der Anwendung dieser lexikalischen Restriktionen nicht offengelegt.

“These results show that the naive approach is quite good. Computationally speaking, it will be a long time before a semantically based algorithm is sophisticated enough to perform as well, and these results set a very high standard for any other approach to aim for.”

Die Untersuchung von Hobbs lieferte somit die ersten greifbaren Anhaltspunkte für die Relevanz der morphologischen und syntaktisch-konfiguralen Restriktionen.

Der *naive Algorithmus* erweist sich aus mehreren Gründen als problematisch. Ein zentraler Kritikpunkt betrifft die *Bevorzugung intrasententieller Kandidaten*, die einer näheren Rechtfertigung bedarf, da sie beispielsweise der durch das Centering-Modell verfolgten Strategie der Präferenz intersententieller Antezedenten diametral gegenübersteht (vgl. Walker, [Walk89]). In ähnlichem Maße diskutabel ist die vermöge der Suchreihenfolge implizit zugrundegelegte *Prominenzordnung* der Antezedens Kandidaten: Lappin und Leass äußern die Vermutung, daß der Hobbssche Algorithmus nur für die vergleichbar kleine Teilmenge von Sprachen gelte, in denen die (in den Syntaxbäumen reflektierte) Oberflächenanordnung der Vorkommen mit der Hierarchie syntaktischer Rollen, die u.a. im Centering-Modell zugrundegelegt wird, koinzidiert - dies sei beispielsweise fürs Englische, nicht jedoch u.a. für die deutsche Sprache mit ihren expliziten morphologischen Kasusmarkern der Fall (vgl. etwa Primus, [Prim89]). Des Weiteren wird von Hobbs kein algorithmischer Mechanismus spezifiziert, mit dem transitive Interdependenzen zwischen Antezedensentscheidungen vermieden werden (vgl. Abschnitt 7.1.2, Beleg (11)).

Der *entscheidende Mangel* besteht jedoch in den unrealistischen Annahmen betreffend die Verfügbarkeit einer *eindeutigen oberflächenstrukturellen Beschreibung*, die in realen Anwendungskontexten nicht gegeben sein wird. Um zu einer tatsächlich anwendungstauglichen Variante von Hobbs' "Algorithmus" zu gelangen, sind Strategien zu entwickeln, die eine möglichst gute Umsetzung der als nur partiell algorithmisierbar klassifizierten Strategie SYNKONFIG leisten. Die Ergebnisse von Hobbs indizieren, daß der Aufwand sich lohnen sollte.

8.2 Die fokusbasierte Anapherninterpretation von Sidner

Das Verfahren von Sidner basiert auf Fokusstrukturen, die den in Abschnitt 7.1.4 vorgestellten ähneln ([Sidn79, Sidn81, Sidn83]). Es kann als einer der Ausgangspunkte der Arbeit von Grosz und Sidner ([GrSi86]) sowie der Centering-Arbeiten von Grosz, Joshi und Weinstein ([GJW83, GJW95]) aufgefaßt werden.

Der Kern der Arbeit Sidners besteht in der Beschreibung von Datenstrukturen für Fokusstrukturen sowie von Rahmenalgorithmen zur Initialisierung und Fortschreibung dieser Strukturen. Die elementaren Strukturen sind mit denen des Centering-Modells vergleichbar: Neben den fokussierten Entitäten - im Sidner-Modell wird zwischen *Diskursfokus* und *Aktorfokus* unterschieden - existieren entsprechende *Listen potentieller Foki*, die nach syntaktischen und semantischen Kriterien (insbesondere thematischen Rollen) geordnet sind. Durch die Verwendung einer Kellerspeicherstruktur sog. "*stacked foci*" kommen auch Kriterien einer nichtlokalen Fokussierung zum Einsatz, aus der sich Restriktionen für anaphorische Wiederaufgriffe ähnlich der oben beschriebenen "*right frontier*"-Bedingung ergeben.⁵ Die Anapherninterpretation verkörpert das

⁵Sidner nimmt hier noch keinen Bezug auf die in [GrSi86] beschriebenen global strukturierenden "Intentionen", sondern verweist lediglich auf die in der frühen Arbeit [Gros77] von Grosz identifizierten "*Task Structure*" aufgabenorientierter Dialoge als einer der möglichen Determinanten.

Bindeglied zwischen den Fokusstrukturen zweier Äußerungen: Sie geht von den Fokusstrukturen der vorhergehenden Äußerungen aus; gleichzeitig werden ihre Ergebnisse im Rahmen des zentralen Fokussierungsalgorithmus dazu herangezogen, die Fokusstrukturen der aktuellen Äußerung fortzuschreiben ([Sidn83], S. 292 ff.). Eine Vielzahl von Anaphertypen finden Berücksichtigung: Personalpronomen in dritter Person, Possessiva, Anaphern mit Situationsbezug, plurale Diskursanaphern mit verteilten Antezedenten sowie nichtpronominale definite NP einschließlich abgeleiteter Spezifikationen.

Trotz bzw. gerade aufgrund der prinzipiellen Bandbreite dieses Modells ist es als ungeeignet für die praktische Anwendung einzustufen. Lediglich die Verfahren zur Initialisierung und Fortschreibung der Fokusstrukturen werden informell prozedural beschrieben. Die konkrete algorithmische Ausgestaltung wesentlicher Teilprobleme dieser Prozesse - insbesondere der eigentlichen Anapherninterpretation - verbleibt *un- bzw. unterspezifiziert*. Eine nähere Aufschlüsselung der expliziten und impliziten Annahmen zeigt auf, daß auf Ressourcen zurückgegriffen wird, deren Algorithmisierung unter Verwendung gegenwärtiger Methoden nicht oder allenfalls partiell bewerkstelligbar ist:

- *syntaktische Oberflächenstruktur*: allgemein, besonders zur Ordnung der Listen des Potentiellen Fokus sowie als Ressource für syntaktisch-konfigurationale Restriktionen der Anapherninterpretation;
- *thematische Rollen*: zur Ordnung der Listen des Potentiellen Fokus sowie zur Unterscheidung zwischen Elementen des Diskursfokus und solchen des Aktorfokus;
- *semantische Rollenbedingungen*: insbesondere während der Anapherninterpretation;
- *allgemeines semantisches/pragmatisches kontextuelles Wissen und Inferenzen*: insbesondere für schwierige Probleme der Anapherninterpretation, z.B. als Entscheidungsgrundlage der nichtlokalen Fokusverschiebungen.

Der Ansatz von Sidner besitzt daher primär *theoretischen Stellenwert*. Der wesentliche Beitrag ist in der vergleichsweise elaborierten Skizze eines auf Fokusstrukturen aufbauenden *Prozeßmodells* zu sehen ([Sidn83], S. 292 ff. sowie S. 304). Die bereits als prinzipiell schwer algorithmisierbar identifizierte Fokustheorie von Grosz und Sidner stellt eine Weiterentwicklung dieses Verfahrens dar, in dem feiner und auf adäquate Weise zwischen lokalen und globalen Fokusphänomenen unterschieden wird (vgl. Abschnitt 7.1.4).

8.3 Das “*Shallow Processing*”-Verfahren von Carter

Carter schlägt einen Ansatz vor, der sich primär auf *linguistisches Wissen* stützt und damit insofern den Anforderungen einer anwendungstauglichen Anapherninterpretation entgegenkommt, als auf algorithmisch schwierig zu behandelndes nichtlinguistisches kontextuelles Wissen nur nachrangig zurückgegriffen wird ([Cart87]).⁶ Diese Vorgehensweise wird unter Rekurs auf zwei

⁶Carter unternimmt keinen Versuch der formalen Abgrenzung dieser beiden Wissenstypen, sondern bezieht sich auf eine Unterscheidung nach intuitivem Verständnis ([Cart87], S. 19):

“... *strictly speaking linguistic knowledge (at least about word meanings) is a subset of world knowledge. However, I will use the term ‘world knowledge’ to refer only to knowledge that is not deemed to be linguistic.*”

grundlegende Eigenschaften natürlicher Sprache gerechtfertigt, die sich aus dem Griceschen Konversationspostulat der Modalität ableiten: (1) Texte sind *redundant* in dem Sinne, daß Information i.d.R. wiederholt und ggf. auf unterschiedlichen Ebenen kommuniziert wird; (2) die unterschiedlichen Beschreibungsebenen *widersprechen sich i.a. nicht*. Mit anderen Worten: Wird ein bestimmter Inhalt auf der Ebene linguistischen Wissens kommuniziert, so kann i.a. auf die Heranziehung nichtlinguistischen Weltwissens verzichtet werden. Carter komprimiert diese Aussage zur sog. “*shallow processing hypothesis*”, einer der Kernthesen seiner Arbeit (ibd., S. 18 f.):

“A story processing system which exploits linguistic knowledge, particularly knowledge about focusing, as heavily as possible, and has access only to limited quantities of world knowledge, which it invokes only when absolutely necessary, can usually choose an appropriate antecedent for an anaphor even in cases where the inference mechanism by itself cannot do so.”

Den Beweis dieser These erbringt Carter in Gestalt der Implementierung und Evaluation eines “*Shallow Processing Anaphor Resolver (SPAR)*”, in der u.a. eine verfeinerte Version der zuvor skizzierten Sidnerschen Fokusalgorithmen sowie Inferenzprozeduren der Wilksschen Präferenzsemantik ([Wilk75a, Wilk75b, Wilk75c]) einfließen.

Anhand der Formulierung der “*shallow processing hypothesis*” werden bestimmte Einschränkungen betreffend die Reichweite ersichtlich. Zunächst einmal wird die Aussage dahingehend relativiert, daß nur auf Texte eines bestimmten Typs - sog. (chronologisch geordnete) “*stories*” - abgezielt wird. Obwohl Carter diese Einschränkung primär per Verweis auf den restringierten Abdeckungsgrad der durch *Boguraev’s analyser* bewerkstelligten linguistischen (syntaktischen und semantischen) Voranalyse rechtfertigt, so stellt sich zumindest die Frage, ob nicht die Regelbasis der Inferenzprozeduren im Falle einer Generalisierung der Anwendungsdomäne auf unrestringierten Text zu modifizieren bzw. erheblich zu erweitern wäre. Carter räumt dies ein, gelangt aber zu einer insgesamt positiven Einschätzung (ibd., S. 236):

“Although SPAR was designed to process stories describing sequences of events, it also proved able to handle, with only trivial extensions, some texts of other type [...] For non-story texts, the inference component might be generalized to look for coherence relations rather than causal chains, perhaps making use of the insights of Hobbs [1979] [...].”

Des Weiteren ist Skepsis betreffend die Implementierung des Fokus-Modells von Sidner angesagt. Die Diskussion in Abschnitt 8.2 hat gezeigt, daß zumindest in unrestringierten Szenarien Probleme entstehen, deren algorithmische Lösung mit gegenwärtigen Methoden nicht machbar erscheint. Carter belegt dies auch bereits für seine eingeschränkte Anwendungsdomäne, indem er Texte angibt, für die aufgrund von Phänomenen nichtlokaler Fokussierung falsche Entscheidungen getroffen werden. Geht man davon aus, daß derartige Fälle durch die Strukturen der “*stacked foci*” erfaßt werden, werden die Grenzen der Operationalisierung des Sidner-Verfahrens im Rahmen von SPAR offensichtlich.

Carter verweist selbst auf die noch nicht umfassende Abdeckung von Anapherntypen, die zwar die üblichen Spielarten (Personalpronomen, definite NP), nicht hingegen u.a. Ellipsen und “*do so*”-Anaphern umfaßt; die entscheidende Einschränkung sieht er jedoch bezüglich der zu erwartenden Komplexität typischer anwendungsrelevanter Texte (ibd., S. 245):

“Rather, what is needed in the long term is extensions of the system along many different dimensions together to enable it to deal with more complex texts. Work along these lines would encounter many of the unsolved problems of natural language processing; however, it is to be hoped that other unsolved problems, notably those of complex inference, would not be so severe in a shallow processing context.”

Es steht also zu erwarten, daß die Evaluationsergebnisse Carters - über 90 % Korrektheit für die einfach strukturierten Texte der “*Story*”-Domäne - nicht auf unrestringierte Anwendungsszenarien generalisierbar sind.⁷ Jedoch geben die Ergebnisse in anderer Hinsicht Anlaß zu Hoffnung: Wenn der Inferenzmechanismus - wie Carter aufschlüsselt - nur bei 12 % aller Pronomen ausschlaggebend für die Antezedensauswahl ist, wird offenbar in allen übrigen Fällen ausschließlich mit linguistischem Wissen ausgekommen. Offensichtlich ergibt sich ein Anhaltspunkt dafür, das “*Shallow Processing*”-Paradigma in sozusagen puristischer Form, d.h. unter völligem Verzicht auf potentiell komplexe Inferenzen, aber dennoch erfolversprechend auf ein uneingeschränktes Anwendungsszenario zu übertragen.⁸

8.4 Der Centering-Ansatz von Brennan, Friedman und Pollard

Die bereits in Abschnitt 7.2.1 partiell beschriebene Implementierung der Centering-Regeln von Brennan, Friedman und Pollard tritt ebenfalls mit dem Anspruch eines (wenn auch partiellen) Ansatzes zur Anapherninterpretation ([BFP87]). Der Schwerpunkt des Verfahrens liegt auf der Umsetzung der Centering-Regeln R1 und R2, wobei - entsprechend den Vorgaben der Originaltheorie - R1 als *Restriktion* und R2 als *Präferenz* realisiert wird. Als Option wird genannt, R1 ebenfalls nur präferentiell zu realisieren. Als weitere stringente Kriterien werden morphologische Kongruenz und syntaktisch-konfigurative Ausschlußbedingungen genannt. Somit setzt der Ansatz die *Verfügbarkeit oberflächenstruktureller Beschreibungen* voraus; er ist insofern als *partiell* einzustufen, als auf den Einsatz weitergehender semantischer und pragmatischer Ressourcen ausdrücklich verzichtet wird (ibid., S. 160). Eine weitere Einschränkung besteht im Hinblick auf die Anapherntypen: Anscheinend werden ausschließlich Pronominalanaphern in dritter Person sowie definite nichtpronominale NP berücksichtigt; Possessiva und Reflexiva/Reziproka bleiben außen vor.

Mit dem Rekurs auf die Centering-Theorie erbt das Verfahren von Brennan, Friedman und Pollard die oben aufgeschlüsselte Unterspezifiziertheit dieses Modells. Selbst unter der (diskussionswürdigen) Annahme, daß die Prominenzordnung der *Cf*-Listen per Hierarchie syntaktischer Rollen adäquat definiert und operationalisiert werden kann, bleiben entscheidende Fragen offen. Wie werden komplexe Satzgefüge behandelt? Wie läßt sich das Phänomen abgeleiteter Spezifikationen ins Centering-Modell integrieren? Wie sieht eine adäquate kontextsensitive Verfeinerung von R2 aus? Wie kann überhaupt der Bezugsrahmen der Centering-Fokussierungsregeln - das “lokale” Diskurssegment - algorithmisch bestimmt werden, ohne auf Ressourcen zurückzugreifen,

⁷Erschwerend kommt hinzu, daß die zu analysierenden Sätze z.T. ediert wurden, um überhaupt erst deren Verarbeitung durch das nur über eine beschränkte syntaktische Kompetenz verfügende Voranalyse-System zu ermöglichen.

⁸Zu lösen verbliebe freilich noch das Problem der Abhängigkeit von den semantischen Strukturen, die das Voranalyse-System nur in bezug auf die oberflächenstrukturell wenig komplexen “*Stories*” in der Lage ist, zu liefern.

deren Algorithmisierbarkeit auf dem gegenwärtigen Stand formaler Modellbildung nicht gewährleistet ist (vgl. Walker: [Walk89], S. 253)?

Ein weiterer Kritikpunkt ergibt sich durch die theoretisch nicht abgesicherte Bevorzugung von Antezedenten aus der unmittelbar vorhergehenden Äußerung gegenüber satzinternen Kandidaten. Zwar schließt die Formulierung des Algorithmus die Betrachtung intrasententieller Vorgänger nicht aus, jedoch scheinen intersententielle Zuordnungen durch die Centering-Präferenzen einseitig bevorzugt zu werden. Gerade im Hinblick auf die (per Handsimulation) erwiesene Performanz der auf der intrasententiellen Ebene beginnenden Suchstrategie des *naiven Algorithmus* von Hobbs erscheint eine solche Strategie hochrelevant. Eine vergleichbare Frage läßt sich bezüglich der nachrangigen Berücksichtigung von Interpretationsalternativen aus dem nichtlokalen Fokus formulieren.

Um zu einem praktikablen Verfahren zur Anapherninterpretation zu gelangen, sind somit nicht nur die bekannten Unterspezifikationen der Centering-Theorie zu eliminieren. Der Ansatz ist dahingehend zu erweitern, daß *intrasententielle Suchstrategien* integriert werden. Erst eine derartige Verfeinerung schafft die Voraussetzung, die Behandlung der bisher unberücksichtigten Typen von Pronominalanaphern - Reflexiva/Reziproka sowie Possessiva - in das Modell zu integrieren. Zur Vermeidung der ineffizienten Enumerationsstrategie sollte eine *sequentielle Variante von R2* implementiert werden. Wie bereits beschrieben, stellt sich dann jedoch die grundsätzliche Frage, warum nicht gleich auf ein erwartet leistungsfähigeres Modell elementarer Prominenzfaktoren übergegangen wird (vgl. Abschnitt 7.2.1).

8.5 Der Mehrstrategieansatz von Carbonell und Brown

Neben seiner grundsätzlichen Bedeutung als Paradigma (vgl. Kapitel 7) ist der Ansatz von Carbonell und Brown auch als praktisches Verfahren zur Anapherninterpretation relevant. In ihrer Arbeit [CaBr88] beschreiben Carbonell und Brown eine Implementierung, die auf einer Teilmenge der als prinzipiell relevant identifizierten Strategien basiert: Berücksichtigt werden Prä-Post-Restriktionen, semantische Rollenbedingungen sowie die Präferenzheuristiken syntaktische Topikalisierung, semantischer Parallelismus und Distanz, wobei die stringenten Kriterien vor den präferentiellen Strategien zur Anwendung gelangen. Sowohl pronominale als auch nicht-pronominale anaphorische Ausdrücke finden Berücksichtigung. Als Evaluationsergebnis für einen kleinen Beispieltext mit 31 anaphorischen Ausdrücken wird eine Korrektheit von 87 % genannt. Obwohl dieses Resultat prinzipiell gut ist, ergeben sich keine unmittelbaren Anhaltspunkte für eine Anwendung auf unrestringierten Text. Das Verfahren geht von der Verfügbarkeit vollständiger syntaktischer und semantischer Analyseergebnisse - insbesondere thematischer Rollen - für jeden zu interpretierenden Satz aus. Daß diese Voraussetzungen im Rahmen einer Analyse großer Textmengen unbekanntem Zuschnitts auf der Grundlage bisheriger Methoden nicht erfüllt werden können, sollte mittlerweile klar sein. Der Stellenwert des Mehrstrategieansatzes von Carbonell und Brown liegt daher primär auf der konzeptuellen Ebene: Anaphernresolutionsverfahren sollten sich eines modular aufgebauten *Komponentenmodells* bedienen; im Rahmen dieses Modells sollten aus Effizienzgründen *Restriktionen vor Präferenzen* zum Einsatz gelangen. Tieferliegende Probleme der Interaktion unterschiedlicher Beschreibungsebenen, wie sie beispielsweise im Modell von Sidner adressiert werden, bleiben zugunsten einer einfachen Operationalisierbarkeit der Architektur außen vor.

8.6 Der dependenzbasierte Algorithmus von Lappin und Leass

Der von Lappin und Leass beschriebene Algorithmus RAP (*Resolution of Anaphora Procedure*) verkörpert einen “*Shallow Processing*”-Ansatz im engeren Sinne, da gezielt auf Komponenten verzichtet wird, die für uneingeschränkte Anwendungsszenarien als gegenwärtig impraktikabel einzustufen sind ([LaLe94]). Zum Einsatz gelangen die morphologische Kongruenzrestriktion, syntaktisch-konfigurationale Bedingungen sowie syntaktisch verankerte Prominenzfaktoren und Parallelismuspräferenzen nach dem Modell von Alshawi, das als geeigneter angesehen wird als insbesondere das Centering-Modell der lokalen Fokussierung. Aufgrund dieser anwendungsorientierten Grundausrichtung stellt der Ansatz einen Meilenstein auf dem Weg hin zu einer praktischen, algorithmischen Anapherninterpretation dar.

Von besonderem Interesse sind die Implementierungstechniken, die für die Umsetzung der syntaktisch-konfiguralen Bedingungen der Bindungstheorie gewählt wurden. Das syntaktische Voranalyse-System basiert auf der “*Slot-Grammar*” von McCord und produziert daher *Dependenzbeschreibungen*⁹, durch die die konfiguralen Gegebenheiten, auf denen die originäre Formulierung der bindungstheoretischen Bedingungen aufbauen, nur partiell reproduziert werden. Lappin und McCord zeigen jedoch, daß die Vorhersagen der Bindungsprinzipien A, B und C durch eine kleine Menge dependenzkonfigural verankerter Bedingungen zumindest weitgehend reproduziert werden können - vgl. [LaCo90a] (Filter für Typ-B-Pronomen) bzw. [LaCo90a] (konstruktive Antezedenssuche für Typ-A-Pronomen). Die Implementierung der Strategie SYNKONFIG ist daher nicht an die Verfügbarkeit einer *phrasenstrukturellen* Beschreibung der syntaktischen Oberflächenstruktur gebunden; die Verwendung der dependentiellen “*Slot-Grammar*”-Strukturen scheint sogar vorteilhaft zu sein, denn Lappin und McCord argumentieren, daß bestimmte komplexe bindungs- und kontrolltheoretische Phänomene in diesem modifizierten strukturellen Rahmen einfacher zu algorithmisieren seien (ibd.).

Die Version von RAP, die in der Arbeit [LaLe94] beschrieben wird, beschränkt sich auf die Interpretation von Personalpronomen in dritter Person (inklusive Possessiva und Reflexiva/Reziproka); es wird ausschließlich nach Beziehungen der *Ko*-Spezifikation Ausschau gehalten. Eine umfassendere RAP-Version, in der u.a. definite (nichtpronominal) NP und nichtelementare Spezifikationsrelationen - verteilte Antezedenten für plurale Anaphern - Berücksichtigung finden, wird von Leass und Schwall beschrieben ([LS91]); jedoch ist deren Kontribution von eher theoretischem Wert, da die operationale Lösung dieser schwierigeren Probleme auf Teilprozessen aufbaut, deren algorithmische Explikation für uneingeschränkte Domänen auf dem gegenwärtigen Stand der Technik allenfalls partiell möglich erscheint.

Hingegen ist der in [LaLe94] beschriebene Ansatz prinzipiell domänenunabhängig, da auf problematische Ressourcen gezielt verzichtet wird; allerdings besteht eine Abhängigkeit vom Abdeckungsgrad der Dependenzanalyse des “*Slot Grammar*”-Parsers. In ihrer Evaluation von RAP greifen Lappin und Leass zu dem Mittel, einzelne Sätze zu modifizieren, um deren Verarbeitung zu ermöglichen (ibd., S. 549):

“These sentences [...] were edited slightly to overcome parse inaccuracies. Rather than revise the lexicon, we made lexical substitutions to improve parses. In some cases constructions had to be simplified. However, such changes did not alter the syntactic relations among the pronoun and its possible antecedents.”

⁹vgl. o., Abschnitt 4.4.1

Eine weitere, hier nicht erwähnte Einschränkung betrifft die unrealistische Annahme *eindeutiger* Parsing-Ergebnisse. Die vergleichsweise hohe Präzision des Algorithmus auf Texten aus einem Korpus von Computer-Manualen - ca. 86 % aller Pronomen (89 % intrasententiell, 74 % satzübergreifend) werden korrekt resolviert - ist daher nicht ohne Weiteres auf realistische Anwendungsszenarien generalisierbar. Die Ergebnisse haben in etwa denselben Stellenwert wie die Resultate der Handsimulation des *naïven Algorithmus* von Hobbs.¹⁰ Jedoch besteht ein wesentlicher Fortschritt in der Architektur des Interpretationsprozesses, durch die zwischen Restriktionen und Präferenzstrategien adäquat unterschieden wird. Hierdurch wird die Übertragbarkeit des Algorithmus auf andere Sprachen gewährleistet, für die die impliziten Annahmen, die der Hobbs'schen Präferenzordnung zugrundeliegen, nicht erfüllt sind (vgl. Abschnitt 8.1).

Für den RAP-Algorithmus werden einige inhaltsorientierte Verfeinerungen vorgeschlagen (Leass, [Leas91]); zumindest einer dieser Vorschläge erscheint auch unter Anwendungsgesichtspunkten als relevant. Einem Vorschlag von Dagan und Itai zufolge könnten *selektionale Präferenzen* anstelle durch aufwendige, für unrestringierte Texte impraktikable Lexikonarbeit alternativ *per automatischer Korpusextraktion* gewonnen werden und als Basis einer *statistischen* Auswahlheuristik Verwendung finden ([DaIt90]). Für eine entsprechende Erweiterung RAPSTAT von RAP konnte eine Performanzverbesserung um weitere 2.5 Prozentpunkte nachgewiesen werden. Aufgrund der *algorithmischen Ausrichtung* als "Shallow Processing"-Verfahren, der durch die Evaluation nachgewiesenen *Relevanz* der eingesetzten Strategien sowie des *flexiblen Verarbeitungsmodells* ist der Ansatz von Lappin und Leass als geeigneter Ausgangspunkt für die anwendungsorientierte Anapherninterpretation anzusehen. Jedoch verbleibt das Problem der Abhängigkeit von vollständigen und unambigen Beschreibungen der syntaktischen Oberflächenstruktur weiterhin ungelöst.

8.7 Der parserlose Algorithmus von Kennedy und Boguraev

Die Defizite des Ansatzes von Lappin und Leass in bezug auf die Zielsetzung einer domänenunabhängigen Anwendbarkeit stellen den Ausgangspunkt der Arbeiten von Kennedy und Boguraev dar ([KeBo96], S. 113):

"While one of the strong points of this algorithm is that it operates primarily on syntactic information alone, this also turns out to be a limiting factor for its wide use: current state-of-the-art parsing technology still falls short of robust and reliable delivery of syntactic analysis of real texts to the level of detail and precision that the filters and constraints described by Lappin and Leass assume."

Kennedy and Boguraev schlagen einen ersten Lösungsweg vor, der zu einem Anaphernresolutionsalgorithmus führt, der den Anforderungen einer unrestringierten Verwendbarkeit gerecht wird: An Stelle einer vollständigen syntaktischen Analyse treten die partiellen Ergebnisse des in

¹⁰Tatsächlich führen Lappin und Leass eine vergleichende Evaluation von Hobbs' *naïvem Algorithmus* auf demselben Textkorpus durch und ermitteln eine Korrektheit von 82 % (87 % intrasententiell, 81 % satzübergreifend). Zwar sieht die durch RAP erzielte Verbesserung um 4 Prozentpunkte auf den ersten Blick marginal aus; jedoch ist zu berücksichtigen, daß hierdurch immerhin (mindestens) 23 % aller verbleibenden (tendenziell *schwierigen*) Fälle gelöst werden.

Abschnitt 5.3 beschriebenen “*Constraint-Grammar*”-Taggers von LINGSOFT, der keine vollständige Analyse der Konstituentenstruktur bewerkstelligt, sondern sich auf die (in Teilen heuristische) Erzeugung von Beschreibungen umfassend morphologische Merkmale, Wortart und syntaktische Funktion beschränkt. Die Grundidee besteht somit darin, auf einer noch seichteren linguistischen Voranalyse aufzubauen.

Da nunmehr keine vollständigen Beschreibungen von Konstituenten- bzw. Dependenzstruktur zur Verfügung stehen, werden weitere Heuristiken benötigt, auf deren Grundlage die syntaktischen Koindexierungsbedingungen der unterschiedlichen Pronomentypen auf dieses wissensärmere Szenario übertragen werden können. Der erste und entscheidende Schritt von Kennedy and Boguraev besteht darin, ausgehend von den Ergebnissen des LINGSOFT-Taggers eine partielle, heuristische Rekonstruktion bestimmter konstituentenstruktureller Relationen zwischen den unterschiedlichen NPs - d.h. Vorkommenskontributoren - eines Satzes vorzunehmen. Dies geschieht durch Anwendung von *Analysemustern*, durch die heuristisch bestimmt wird, ob eine NP innerhalb eines Adverbs bzw. innerhalb einer weiteren NP vorkommt; da die Analysemustern durch *reguläre Ausdrücke* beschrieben werden, ist die Rekonstruktion des syntaktischen Kontexts notwendig heuristisch bzw. partiell.¹¹ In einem zweiten Schritt werden diese Merkmalsbeschreibungen des syntaktischen Kontexts gemeinsam mit der originären Information betreffend syntaktische Funktion und Oberflächenanordnung dazu verwendet, die syntaktischen Koindexierungsbedingungen heuristisch zu emulieren. Den Kern bilden drei Bedingungen, die *unter Rekurs auf die Kontextinformation* die Bindungsrestriktionen für nichtreflexive Personalpronomen nachbilden (ibid., S. 116);¹² Reflexiv- und Reziprokpronomen werden unter alleiniger Bezugnahme auf die Tagger-Ergebnisse resolviert.

Zusätzliche Filter bewerkstelligen eine (bezüglich der Reichweite *eingeschränkte* heuristische Elimination bestimmter nichtreferentieller Verwendungsweisen des Pronomens “*it*”. Die gegenwärtige Version des Algorithmus zielt primär auf die Interpretation von *Pronomen mit Objektbezug* ab; Anaphern mit relationalem Bezug werden nicht adäquat resolviert. Aus den gegebenen Beispielen geht ferner hervor, daß auch bereits einige (triviale) Instanzen anaphorischer definitiver (nichtpronominaler) NP resolviert werden; komplexere Fälle - z.B. abgeleitete Spezifikationen - bleiben unberücksichtigt.

In einer *Evaluation* des Algorithmus auf Texten eines breiten Spektrums von Genres wurde eine Präzision von 75 % für *anaphorische* Dritte-Person-Pronomen erzielt.¹³ Das Ergebnis ist jedoch insofern zu relativieren, als inkorrekt interpretierte Instanzen expletiver “*it*”-Verwendungen, die durch den erwähnten Filter nicht eliminiert wurden, sowie “*it*”-Vorkommen, die keine Objektentität spezifizieren, nicht berücksichtigt werden - werden diese Fehler mitgezählt, so sinkt die Korrektheit auf 68 % (ibid., S. 117, Randbemerkung 3).

Aufgrund des breiteren Spektrums an Textgenres ist kein unmittelbarer Vergleich mit den Ergebnissen von Lappin und Leass möglich. Kennedy und Boguraev stellen die größere Komplexität der in ihrer Evaluation zugrundegelegten Texte heraus und verweisen auf den (gemessen an den reduzierten Anforderungen an die Voranalyse) vergleichsweise geringen Verlust an Präzision gegenüber dem Algorithmus von Lappin und Leass. Dies kann jedoch nicht ohne Weiteres nachvollzogen werden, da sich in jener Arbeit kein Hinweis darauf findet, ob die Fehler des dort verwendeten Expletivumfilters in den Gesamtergebnissen mitgezählt wurden. Da die nichtreferentiellen “*it*”-Verwendungsweisen zumindest in Teilen durch den syntaktischen Kontext deter-

¹¹Kennedy und Boguraev machen keine genaueren Angaben betreffend dieses “*set of patterns*” (ibid., S. 114).

¹²Auch die Beschreibung dieser drei Regeln erfolgt nur auf vergleichsweise oberflächliche und allenfalls ansatzweise nachvollziehbare Weise.

¹³Aus den Ausführungen geht nicht hervor, ob dieses Resultat auch die (i.a. einfacher zu resolvierenden) Reflexiv- und Reziprok-Pronomen umfaßt.

miniert werden, kann nicht ausgeschlossen werden, daß im wissensärmeren Szenario bestimmte Grenzen der Filterbarkeit existieren, die sich im Gesamtergebnis niederschlagen.

Durch eine qualitative Evaluation der problematischen Fälle kommen Kennedy und Boguraev zu dem Ergebnis, daß die heuristische Rekonstruktion der syntaktisch-konfiguralen Restriktionen nur in wenigen Fällen fehlgeht (genaue Zahlen werden nicht genannt). Als wichtigste Fehlerquellen werden inkorrekte Genusinformation des LINGSOFT-Taggers sowie Fälle eingebetteter zitierter Rede genannt; die Ergebnisse ließen sich somit durch die Verwendung eines zusätzlichen morphologischen Lexikons bzw. zusätzlicher Kontextanalysemuster verbessern.

Obwohl aufgrund von in Teilen unvollständiger bzw. fehlender Information eine abschließende Bewertung des Ansatzes von Kennedy und Boguraev nicht möglich erscheint, so kann dessen Beitrag für die Entwicklung anwendungstauglicher Anapherninterpretationsverfahren nicht hoch genug eingeschätzt werden. Selbst unter Anlegung strenger Maßstäbe wird immerhin eine Güte von 68 % erzielt - ein Resultat, das für das als inhärent schwierig bekannte Problem der nicht domänenspezifischen Kospezifikationsinterpretation bis vor einigen Jahren wohl niemand für möglich gehalten hätte. Mit Blick auf die im Rahmen der Evaluation identifizierten wichtigen Problemklassen ergibt sich zumindest ein erster Anhaltspunkt für eine weitere Verbesserung: Läßt sich die Verifikation der syntaktischen Koindexierungsbedingungen im Speziellen bzw. die Anapherninterpretation im allgemeinen nicht auch auf den i. a. *partiellen* (ambigen bzw. fragmentarischen) Ergebnissen eines *herkömmlichen Parsing* bewerkstelligen? Damit entfielen zumindest ein Teil der heuristischen Regelanwendung und es bestünde Raum für Ergebnisverbesserungen, da zu erwarten ist, daß die zur Verfügung stehenden Beschreibungen des syntaktischen Kontexts eine höhere Güte aufweisen. Diese Idee wird den Anknüpfungspunkt für den im Rahmen der Arbeit zu entwickelnden Algorithmus verkörpern.

8.8 Teilnehmer der “*Coreference Task*”-Evaluation von MUC-6

Im Rahmen der *Message Understanding Conference 6* fand erstmals eine *formale* Evaluation implementierter Anapherninterpretationsalgorithmen statt, an der sieben Systeme partizipierten. Aufgrund der zugrundegelegten Methodik sind Operationalität und weitgehende Domänenunabhängigkeit der Ansätze gewährleistet. Im Unterschied zu den Systemen von Lappin und Leass bzw. Kennedy und Boguraev werden auch nichtpronominale Anaphern (definite NP bzw. Namen) betrachtet; da deren Erkennung und Resolvierung z.T. problematischer ist, wird in der Evaluation eine Differenzierung zwischen *Recall* und *Precision* vorgenommen (vgl. Abschnitt 5.4.2).

Die Systeme weisen grundlegende Unterschiede bezüglich der verwendeten Strategiekombinationen auf, was zumindest in einigen Fällen durch die im Gesamtsystem verfügbaren Analyse-Ressourcen bedingt ist (vgl. [MUC696]). Für vier der Ansätze stehen die Ergebnisse einer (evtl. fragmentarischen) syntaktischen Analyse zur Verfügung; ein Verfahren baut auf den Resultaten eines “*Part-of-Speech*”-Taggers auf; die übrigen zwei Systeme basieren auf einer partiellen syntaktischen Voranalyse, die sich regulärer Ausdrücke bedient. Ein unmittelbarer Zusammenhang zwischen Voranalyseszenario und Ergebnisgüte wurde nicht nachgewiesen: Das System mit dem besten Recall - PIE, University of Manitoba, $(P, R) = (0.63, 0.63)$ - basiert auf konventionellem Parsing; das System mit der höchsten Precision - FASTUS, SRI International, $(P, R) = (0.72, 0.59)$ - verwendet Analysetechniken, die auf regulären Sprachen aufbauen.

Einige Gemeinsamkeiten, die zum Teil auf das Layout von MUC-6 zurückführen sind, lassen sich

dennoch identifizieren. Allgemein als wichtig angesehen werden elaborierte Techniken zur Identifizierung kospezifizierender Namen auf der Basis oberflächlicher, zeichenkettenbasierter Heuristiken. Ein weiterer Trend besteht im Rückgriff auf bereits verfügbare, als massendatentauglich bekannte Systeme der morphologischen Analyse sowie des “*Part-of-Speech*”-Tagging. An die Stelle einer aufwendigen manuellen Erstellung von Lexika treten Techniken zur automatischen bzw. halbautomatischen Wissensakquisition, die auf verfügbare annotierte Korpora zurückgreifen; in einem der Teilnehmersysteme - das der University of Massachusetts - gelangte diese Technik sogar zur Bestimmung anapherntypspezifischer Strategiekombinationen (Entscheidungsbäume) zum Einsatz.

Das Algorithmienlayout von Lappin und Leass spiegelt sich am nachhaltigsten im erfolgreichen PIE-System wider. Zu den zumindest potentiell erfolgversprechendsten Ansätzen ist das System der University of Pennsylvania zu zählen ($(P, R) = (0.63, 0.55)$), das ebenfalls auf einer Vielzahl von Ressourcen und Strategien basiert, zu denen morphologische und syntaktisch-konfigurationale Kriterien ebenso gehören wie statistisch-semantische Heuristiken.¹⁴ Verfahren, die in erster Linie auf unscharfem konzeptuellen Wissen (vorgegebenen bzw. heuristisch akquirierten semantischen Netzen) basieren und auf oberflächennahe Kriterien verzichten, erscheinen weniger erfolgreich (LOLITA, University of Durham, $(P, R) = (0.44, 0.36)$).

Die Ergebnisse der MUC-6-Evaluation bestätigen somit die Grundaussagen der Arbeiten von Lappin und Leass bzw. Kennedy und Boguraev. Eine algorithmische Interpretation anaphorischer Elemente erscheint auf der Grundlage von Oberflächen- und syntaktischer Information mit Präzisionswerten in der Größenordnung von 70 % möglich; automatisch akquirierte semantische Information trägt allenfalls nachrangig zur Qualität der Ergebnisse bei. Die Auswertung im Rahmen von MUC-6 bestätigt ferner die Erwartung, daß die Resolution definiter (nichtpronominaler) NPs den eigentlichen Engpaß darstellt, der sich darin niederschlägt, daß die Recall-Resultate i.a. deutlich unter den erzielten Precision-Werten liegen.

8.9 Diskussion

Die Analyse der wichtigsten Anapherninterpretationsverfahren erbringt die Evidenz für die in Kapitel 7 vorgriffweise geäußerten Relevanzeinschätzungen zu den einzelnen Strategien. In der Tat erweisen sich die auf der Ebene der syntaktischen Oberflächenstruktur angesiedelten konfiguralen Bedingungen als zentral für die Resolution objektspezifizierender pronominaler Anaphern; zusammen mit weiteren auf der linguistischen Ebene angesiedelten Restriktionen wie etwa der morphologischen Kongruenz sowie zusätzlichen syntaxbasierten Auswahlheuristiken können Präzisionswerte von ca. 70 % erzielt werden. Die meisten der diskutierten Ansätze tragen zu diesem State-of-the-Art bei: der *naive Algorithmus* von Hobbs als erste Evidenz für den potentiellen Erfolg syntaxbasierter Ansätze; das Modell von Carbonell und Brown durch das Mehrstrategieparadigma sowie durch die implementierungstechnisch adäquate Differenzierung zwischen Restriktionen und Präferenzen; der Algorithmus von Lappin und Leass im Hinblick auf eine dementsprechende Formulierung eines auf dependenzstrukturellen Konfigurationen sowie Prominenzfaktoren basierten Algorithmus; der wichtige Schritt hin zu einem wirklich *seichten*, uneingeschränkt operationalen Voranalyseszenario durch den auf der Ausgabe eines “*Part-of-Speech*”-Taggers operierenden Algorithmus von Kennedy und Boguraev; und schließlich die im

¹⁴Die Autoren verweisen auf eine deutliche Erhöhung der Performanz ihres Systems ($(P, R) = (0.72, 0.63)$) nach der Beseitigung eines trivialen technischen Fehlers.

Rahmen von MUC-6 realisierten Verfahren, deren Stärken und Schwächen per formaler Evaluation offengelegt wurden.

Ein genauerer Blick auf die analysierten Ansätze zeigt jedoch auf, daß Spielraum für weitere Verfeinerungen besteht. Einige der wichtigsten Probleme des Algorithmus von Kennedy und Boguraev scheinen unmittelbar auf die vergleichsweise oberflächliche Analyse des syntaktischen Kontexts zurückzuführen zu sein. Folglich sollte sich die Präzision unter Rekurs auf - ggf. partielle bzw. heuristisch erstellte - syntaktische Strukturen herkömmlichen Zuschnitts erhöhen lassen. Beeinträchtigungen sind insbesondere für den Prozeß der Verifikation syntaktisch-konfiguratoraler Bedingungen zu erwarten, die allerdings für ein Szenario partieller - potentiell *fragmentarischer* - syntaktischer Beschreibungen in geeigneter Weise zu reformulieren wären. Durch die MUC-Teilnehmersysteme, die auf einer ausführlichen Syntaxanalyse basieren, wird diese Fragestellung nicht adressiert: Für das LaSIE-System der University of Sheffield stellt sich das Problem nicht, da die Oberflächenstrukturbäume a priori heuristisch disambiguiert bzw. vervollständigt werden;¹⁵ in der Beschreibung des Systems der University of Pennsylvania findet sich keine Angabe über die Behandlung partieller syntaktischer Information; das PIE-System der University of Manitoba arbeitet auf fragmentarischer Oberflächenstruktur, jedoch fehlen Aussagen betreffend die Generalisierung der Verifikation konfiguratoraler Bedingungen - die im Vergleich geringere Präzision von 63 % indiziert, daß Verbesserungen für die Pronomenresolution möglich sein sollten.

Im folgenden soll daher ein Algorithmus zur Anapherninterpretation entwickelt werden, der auf *fragmentarischen oberflächenstrukturellen Beschreibungen* arbeitet, die sich infolge von Ambiguitäten bzw. Konflikten während der syntaktischen Analyse ergeben. Der Schwerpunkt der Ausführungen liegt auf der Entwicklung von Methoden, die eine adäquate und möglichst vollständige Übertragung der syntaktisch-konfiguratoralen Bindungsbedingungen auf dieses realistische Voranalysezenario bewerkstelligen. Der Ansatz ist damit in gleichem Maße operational wie das Verfahren von Kennedy und Boguraev, denn Ressourcen zur fragmentarischen Syntaxanalyse auf unrestringiertem Text sind verfügbar. Im Unterschied zu den MUC-Partizipanten wird auf die Verwendung elaborierter Heuristiken sowie spezieller Lexika zur Behandlung von Namen und sonstiger nichtpronominaler NP verzichtet. Der Schwerpunkt der Untersuchungen liegt somit auf der Interpretation *pronominaler* Anaphern sowie auf der Entwicklung von Rahmenalgorithmen, die das Problem der robusten Verarbeitung partieller syntaktischer Strukturen unter Rekurs auf adäquate Strategien lösen.

¹⁵Aus der Beschreibung des Systems geht ohnehin nicht hervor, ob überhaupt syntaktisch-konfigurationale Restriktionen zum Einsatz gelangen.

Kapitel 9

Basisalgorithmus und referentielle Datenstrukturen

Auf der Grundlage der Untersuchungen in den vorangegangenen Kapiteln soll nun ein *Basisalgorithmus* zur Interpretation anaphorischer Ausdrücke entwickelt werden, der den identifizierten Anforderungen der Anwendungstauglichkeit entspricht (vgl. Stuckardt: [Stu96b, Stu97a]). Im ersten Schritt geht es hierbei um die Auswahl geeigneter Einzelstrategien sowie um deren adäquate Algorithmisierung und Integration in einen monolithischen Analyseprozeß. Entsprechend den Ergebnissen der obigen Analyse stellen die syntaktisch-konfigurationsalen Restriktionen der Bindungstheorie eine der Kernstrategien dar. Aus der Beschreibung in Abschnitt 7.1.2 geht hervor, daß eine Operationalisierung dieser Kriterienklasse aufgrund einiger zu berücksichtigender Subtilitäten nichttrivial ist. Es soll daher zunächst eine adäquate Algorithmisierung dieser Strategie entwickelt werden, wobei von der idealisierenden Annahme ausgegangen wird, daß eine eindeutige und vollständige Beschreibung der syntaktischen Oberflächenstruktur vorliegt. Im folgenden Kapitel wird das zentrale Problem fragmentarischer Syntaxanalyseergebnisse adressiert und eine robuste Version des Algorithmus erarbeitet, die auf partieller syntaktisch-konfigurationsaler Information arbeitet und somit uneingeschränkt operational ist.

Jede algorithmische Lösung des Problems der Anapherninterpretation basiert jedoch auf *Datenstrukturen*, die den unter Performanzgesichtspunkten zu stellenden Anforderungen an eine Repräsentation referentieller Information gerecht werden.¹ Bevor zu einer Diskussion der algorithmischen Aspekte geschritten werden kann, ist deshalb zunächst eine geeignete repräsentationale Basis zu erarbeiten. Somit ergibt sich für das vorliegende Kapitel die folgende Organisation:

1. Auswahl relevanter und operationa(b)ler *Strategien*;
2. Entwicklung geeigneter *Datenstrukturen* für referentielle Information;
3. *Algorithmisierung* der ausgewählten Strategien;
4. *Integration der Strategien* in einen Gesamtalgorithmus.

¹Es sei nochmals ausdrücklich darauf hingewiesen, daß im Rahmen der vorliegenden Arbeit keine Zielsetzung der *kognitiven Adäquatheit* verfolgt wird. Eine kognitionstheoretisch ausgerichtete Diskussion von Strukturen sowie - ansatzweise - Prozessen der Referentialität findet sich in der Arbeit von Habel ([Habe86]).

9.1 Auswahl relevanter und algorithmisierbarer Strategien

Mit Blick auf Abbildung 7.7 und die Ergebnisse von Kapitel 8 gestaltet sich die Auswahl der Kriterien, die sowohl anwendungsrelevant als auch auf dem gegenwärtigen Stand der Technik operational bzw. operationabel, d.h. partiell/heuristisch (robust) implementierbar sind, einfach. Von den *Restriktionen* sind dies die Strategien MORKONGRU (morphologische Kongruenz), SYNKONFIG (syntaktisch-konfigurationale Bedingungen), DEFKATAPH (Definitheitsbedingung kataphorischer Antezedenten) sowie - eingeschränkt - SEMROLLEN (selektionale Restriktion/Präferenz von Mitspielern eines bestimmten semantischen Typs). Zusätzlich soll eine weitere, bislang noch nicht näher betrachtete Klasse von Kriterien einfließen, die in den in Abschnitt 8.8 diskutierten Teilnehmersystemen der MUC-6-Evaluation eine zentrale Rolle spielt: Zumindest für bestimmte Anwendungsdomänen erweisen sich (vergleichsweise einfach zu implementierende) oberflächliche Bedingungen vollständiger oder partieller *Zeichenkettenübereinstimmung* von Namen als hochrelevant; auch für die allgemeine Anwendung sollte sich ein Grundinventar generell nützlicher Regeln dieser *Strategie ZEICHKETT* identifizieren lassen.

Betreffend die *Präferenzheuristiken* ist die Auswahl größer: Unter der oben diskutierten und begründeten Annahme, daß auf die Verwendung spezifischer Fokusregeln der Klasse LOK-FOKUS (z.B. die Centering-Regeln) zugunsten eines feinkörnigeren Inventars elementarer Prominenzkriterien verzichtet werden soll, kommen die Faktoren im engeren Sinne DISTANZ, TOPIKALIS, SUBJEKT und SYNROLLE sowie die relativen Faktoren KATAMALUS, SYNPARALL und - eingeschränkt - SEMPARALL in Frage.

Alle genannten Kriterien sollen im Basisalgorithmus umgesetzt werden. Die einzige Ausnahme betrifft die semantische Selektions-Restriktion/-Präferenz SEMROLLEN, deren Implementierung auf der aufwendigen Konstruktion eines umfassenden semantischen Lexikons basiert - eine Aufgabe, die über die grundsätzlichen Ziele der vorliegenden Arbeit hinausgeht. Es stellt jedoch kein prinzipielles Problem dar, den Kernalgorithmus entsprechend zu erweitern, wobei z.B., dem oben diskutierten Vorschlag von Dagan und Itai folgend, auf ein per Korpusextraktion automatisch konstruiertes Lexikon zurückgegriffen werden könnte ([DaIt90]). In Abhängigkeit der Qualität der verfügbaren Daten verbliebe zu entscheiden, ob das Kriterium als Restriktion oder als Präferenz sowie ggf. mit welchem relativen Gewicht realisiert wird.

Nachdem nun die zu berücksichtigenden Strategien festgelegt sind, besteht die nächste Aufgabe in der Definition geeigneter Datenstrukturen, mit denen die einfließende Information in einer Weise repräsentiert werden kann, die den Anforderungen einer möglichst effizienten Algorithmisierung der Anapherninterpretation entsprechen.

9.2 Datenstrukturen für referentielle Information

Ein System zur Interpretation anaphorischer Ausdrücke steht vor zwei klar voneinander abgrenzbaren Aufgaben: (1) die Identifikation spezifizierender Vorkommen; (2) die Ermittlung von Antezedenten für diejenigen Okkurrenzen, die einen anaphorischen Wiederaufgriff bewerkstelligen. Ergebnis von Schritt (1) ist eine Liste von *Vorkommensbeschreibungen*, in denen die für die eigentliche Anapherninterpretation in Schritt (2) relevante Information akkumuliert ist. Schritt (2) führt zur Ermittlung von Klassen kospezifizierender, d.h. ein und demselben *Diskursreferenten* zuzuordnender Vorkommen. Somit sind zwei entsprechende *Datentypen* zu vereinbaren, durch die Vorkommen und Diskursreferenten in adäquater Weise beschrieben werden können.

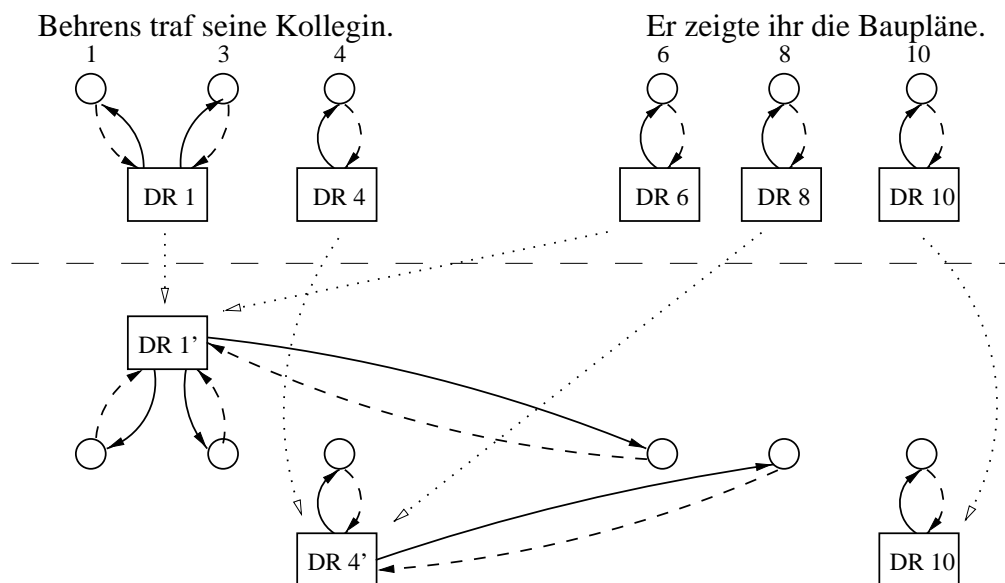


Abbildung 9.1: Verweisstrukturen zwischen Vorkommen und Diskursreferenten

Die Instanzen dieser Strukturen verweisen wechselseitig aufeinander: Jeder Okkurrenz ist ein eindeutig bestimmter Diskursreferent zugeordnet; umgekehrt umfaßt jeder Diskursreferent eine Menge von Vorkommen, die ihn an der sprachlichen Oberfläche realisieren. Daß die entsprechenden Verweisstrukturen repräsentational abgebildet werden, ist für Operationalisierung der Anapherninterpretation von zentraler Bedeutung: Die als Äquivalenzklassen kospezifizierender Vorkommen auffassbaren Diskursreferenten konstituieren das gesuchte *Ergebnis* des Anaphern-resolutionsprozesses; umgekehrt wird der Rückverweis von einem spezifischen Vorkommen zum zugehörigen Diskursreferenten benötigt, um die im Rahmen der schrittweisen Antezedenssuche anfallenden Identifikations- und Verschmelzungsoperationen auf Diskursreferenten durchführen zu können.

Abbildung 9.1 zeigt ein elementares Beispiel der referentiellen Interpretation eines Textes, der zwei Sätze umfaßt. Im oberen Teil der Darstellung ist der Zustand nach der Verarbeitung von Satz 1 skizziert. Diskursreferent DR₁ verkörpert das Spezifikat der beiden Vorkommen "Behrens" und "seine" (dargestellt als Kreise), die bereits als koreferent erkannt wurden; die Verweisstruktur ist in Form von Pfeilen zwischen Vorkommen und Diskursreferenten abgebildet. Diskursreferent DR₄ steht für den referentiellen Ausdruck *Kollegin*, für den noch keine Kospezifikation etabliert wurde. Der noch nicht interpretierte Teil des Diskurses - Satz 2 - wird durch eine Menge von Vorkommensbeschreibungen repräsentiert, deren zugehörige Diskursreferenten bislang sämtlich einelementig sind, da ja die referentielle Interpretation noch aussteht.

Der untere Teil der Abbildung zeigt die Diskursrepräsentation, die entsteht, nachdem Satz 2 (korrekt) interpretiert wurde. Die Entscheidungen, daß die Vorkommen "Behrens" und "Er" bzw. "Kollegin" und "sie" kospezifizieren, werden u.a. auf der Basis der entsprechenden Vorkommensbeschreibungen getroffen, die die morphologischen Beschreibungen des sprachlichen *Ausdrucks* enthalten. Ausgehend von den Vorkommen-Beschreibungen werden sodann die zugehörigen Diskursreferenten identifiziert, die der Interpretationsentscheidung zufolge zu verschmelzen (zu *uni-*

```

(defstruct (dr-okk)
  lexem      ;;; Stammform
  okknum     ;;; diskursglobale Nummer (eindeutiger Identifikator)
  satznum    ;;; Nummer des Satzes
  typ        ;;; Typ des Vorkommens
  bt-klasse  ;;; bindungstheoretische Klasse (A,B,C)
  disref     ;;; Zeiger auf den zugehoerigen Diskursreferenten
  ngp-liste  ;;; Liste morphologischer Merkmalsbuen-del (Numerus, Genus, Person)
  synfun     ;;; syntaktische Funktion
  semrolle   ;;; Tiefenkasus (thematische Rolle)
  quantor    ;;; zugehoeriges Artikelwort (fuer nichtpronominale NP)
  ostknoten) ;;; Zeiger auf den entsprechenden Knoten des Syntaxbaums

```

Abbildung 9.2: Datenstruktur für die Beschreibung von Vorkommen

fizieren) sind.² Im Ergebnis liegen zwei Diskursreferenten DR'_1 und DR'_4 vor, die die Vorkommen der verschmolzenen Diskursreferenten subsumieren. Um diese Operation implementierungstechnisch zu bewerkstelligen, sind Zugriffe auf alle Vorkommensbeschreibungen der “untergehenden” Diskursreferenten DR_6 und DR_8 notwendig, um die entsprechenden Zeiger neu zu setzen.³ Es verbleibt zu beschreiben, welche Information in den Beschreibungen von Vorkommen und Diskursreferenten im Hinblick auf die zu operationalisierenden Strategien zu akkumulieren ist.⁴ In Abbildung 9.2 ist eine erste Näherung der Datentypdefinition für *Vorkommensbeschreibungen* in der für die Implementierung des Algorithmus verwendeten Implementierungssprache COMMON LISP⁵ dargestellt. Es liegt der Anspruch zugrunde, diejenige Information zu bündeln, die einer *bestimmten Realisierung* eines Diskursreferenten zuzuordnen ist, und die für deren referentielle Interpretation relevant ist. Die für die Anwendung der morphologischen Kongruenzbedingung benötigte Beschreibung von Numerus, Genus und Person wird unter dem Attribut `ngp-liste` zur Verfügung gestellt; da es Fälle gibt, in denen diese Beschreibung a priori mehrdeutig ist (z.B. für das Personalpronomen “*sie*” im Deutschen), wird eine *Liste* solcher Beschreibungen verwaltet. Das Merkmal `lexem`, unter dem die morphologische Stammform abgelegt wird, ist primär für nichtpronominale Okkurrenzen relevant, da auf dessen Grundlage zeichenkettenbasierte Heuristi-

²Je nach Typ der Anapher können auf der Ebene der Diskursreferenten weitere Strategien zur Anwendung gelangen, die auf inhaltsbezogener Information basieren, die von den Charakteristika der jeweiligen sprachlichen Realisierung abstrahiert. Dies betrifft z.B. das natürliche Geschlecht des Referenzobjekts, das - wie oben ausgeführt - vom grammatischen Geschlecht spezifischer Realisierungen durchaus abweichen kann. In seinem Repräsentationsformalismus der Referentiellen Netze trifft Habel daher eine Unterscheidung zwischen Vorkommens- und RefO-Attributen, die auch im hier entwickelten Repräsentationsschema prinzipiell erfaßt ist ([Habe86]).

³Letztlich spielt es keine inhaltliche Rolle, *welcher* der beiden in die Unifikation einfließenden Diskursreferenten als Basis für die Beschreibung des umfassenderen Spezifikats gewählt wird. Unter Effizienzgesichtspunkten bietet es sich jedoch an, den “älteren” Diskursreferenten beizubehalten, da diesem i.d.R. mehr Vorkommen zugeordnet sind und dadurch die Anzahl durchzuführender Aktualisierungen vorkommenseitiger Verweise reduziert wird.

⁴Die Frage, *wie* diese Information im Rahmen des oben identifizierten Verarbeitungsschritts (1) gewonnen wird, wird hier zunächst übergangen, da sie erst in bezug die Ergebnisstrukturen konkreter, noch zu wählendes Voranalyse-systeme (morphologische und syntaktische Analyse) beantwortbar ist (s.u.).

⁵Für das Verständnis der im folgenden gegebenen Beispiele sind keine tieferegehenden Vorkenntnisse der Implementierungssprache COMMON LISP notwendig; ein intuitives Verständnis der entsprechenden programmiersprachlichen Konstrukte ist hinreichend.

Die Wahl von COMMON LISP als Implementierungssprache geschah primär aufgrund der zur Verfügung stehenden programmiersprachlichen Konstrukte zur Konstruktion und Manipulation komplexer Symbolstrukturen, wie sie für die Aufgabenstellungen der Verarbeitung natürlicher Sprache charakteristisch sind. Die Codierungen könnten prinzipiell in jeder anderen berechnungstheoretisch vollständigen (“turing-mächtigen”) Programmiersprache geschehen.


```
(defstruct (disref)
  disrefnum    ;;; eindeutiger Identifikator (Nummer einer der Okkurrenzen)
  okks        ;;; Liste kospezifizierender Okkurrenzen)
```

Abbildung 9.3: Datenstruktur für die Beschreibung von Diskursreferenten

ken für Namen und definite NP zur Anwendung gelangen. Um überhaupt entscheiden zu können, welcher Strategiemix relevant ist, wird unter `typ` festgehalten, um welche Form eines anaphorischen Ausdrucks (Reflexivum, Pronomen, Name usw.) es sich handelt. Die Anwendung der syntaktisch-konfiguralen Restriktionen (insbesondere der Bindungsprinzipien A, B und C) geschieht in Abhängigkeit der bindungstheoretischen Klasse, verwaltet unter der Strukturkomponente `bt-klasse`; der notwendige Zugriff auf die oberflächenstrukturelle Beschreibungsebene wird durch einen entsprechenden Zeiger `ostknoten` ermöglicht. Als Basis für die Entscheidung, ob eine bestimmte nichtpronominale NP *definit* ist und somit anaphorisch wirkt, wird ein Attribut `quantor` vereinbart, unter dem ggf. die Beschreibung eines Artikelworts der NP abliegt. Die Merkmale `synfun` und `semrolle` dienen zum Merken von Information betreffend die syntaktische oder thematische Rolle, die das Vorkommen in seinem spezifischen Kontext ausfüllt; dies schafft die Voraussetzungen für die Anwendung der Heuristiken der syntaktischen Prominenz sowie des syntaktischen bzw. semantischen Parallelismus. Die Implementierung der Distanzheuristik geschieht unter Rekurs auf die Strukturkomponente `satznum`. Das Attribut `okknum` dient einerseits als eindeutiger Identifikator (Schlüsselwert) eines Vorkommens; durch seine Gleichsetzung mit der (bzw. einer) Wortnummer des induzierenden sprachlichen Ausdrucks ermöglicht es ferner die Operationalisierung der Kataphernmalus-Strategie. Unter dem Merkmal `disref` wird der oben diskutierte Zeiger auf den zugehörigen Diskursreferenten verwaltet.

Die (ebenfalls vorläufige) Beschreibung der Datenstruktur für *Diskursreferenten* gestaltet sich elementar (vgl. Abbildung 9.3). Attribut `okks` enthält eine Liste aller Vorkommen, die den jeweiligen Diskursreferenten kospezifizieren. Unter `disrefnum` wird die Nummer des Diskursreferenten festgehalten - ein eindeutiger Identifikator, der i.d.R. mit der Nummer desjenigen Vorkommens zusammenfällt, das das Spezifikat in den Diskurs eingeführt hat (vgl. o.). Weitere Merkmale - etwa semantische Attribute, die dem Spezifikat in toto zuzuordnen sind - könnten problemlos ergänzt werden. Dies gilt u.a. für bestimmte Attribute, die dem Diskursreferenten vermöge selektionaler Restriktionen, die für dessen Realisierungen gelten, zuzuschreiben sind. Da die Strategie SEMROLLEN in dem zu beschreibenden Algorithmus vorerst keine Berücksichtigung findet, verbleibt alleine festzuhalten, daß der Datentyp für Diskursreferenten problemlos in adäquater Weise erweiterbar ist.

9.3 Algorithmisierung der gewählten Strategien

Unter der oben getroffenen vorläufigen Annahme, daß das als Voranalyse ablaufende syntaktische Parsing stets vollständige und unambige Ergebnisse liefert, ist die Implementierung der gewählten syntaktischen und oberflächlichen (anordnungsbezogenen) Präferenzkriterien trivial. Anders verhält es sich mit der Operationalisierung einiger der stringenten Kriterien, deren Diskussion in Kapitel 7 einige Feinheiten offengelegt hat, die im Rahmen einer adäquaten Implementierung zu berücksichtigen sind.

9.3.1 Strategie MORKONGRU

In Abschnitt 7.1.1 wurden Belege angegeben, die aufzeigen, daß sich die grammatischen Merkmale *Numerus*, *Genus* und *Person* nicht immer alleine per morphologischer Analyse des sprachlichen Ausdrucks eindeutig erschließen lassen. Ohne nun feiner danach zu unterscheiden, durch den Einsatz welchen Wissens die jeweilige Ambiguität zu resolvieren ist, läßt sich hieraus für Implementierungen der Strategie MORKONGRU die Bedingung ableiten, auf *Mengen* von Merkmalskombinationen zu arbeiten: Morphologische Kongruenz ist genau dann gegeben, wenn die *Schnittmenge* der Merkmalskombinationsmengen von Antezedenskandidat und pronominaler Anapher nichtleer ist. Dies ist der Grund dafür, warum in der in Abbildung 9.2 spezifizierten Datenstruktur für Okkurrenzbeschreibungen eine *Liste* (als spezifische Implementierungsvariante von Mengen) morphologischer Merkmalsbündel verwaltet wird.

Die zu berücksichtigenden Subtilitäten einer MORKONGRU-Algorithmisierung beschränken sich jedoch nicht auf diese in elementarer Weise codierbare Mengenoperation. Wie in Abschnitt 7.1.1 darüberhinaus aufgezeigt wird, können Antezedensentscheidungen morphologisch disambiguierend wirken in dem Sinne, daß eine vorgefundene Übereinstimmung in einer *Teilmenge* von Merkmalsbündeln auch für weitere pronominale Aufgriffe dieser Kospezifikationsklasse verbindlich werden - in bestimmten Fällen kann es hierdurch zu Interdependenzen zwischen individuellen Antezedensentscheidungen kommen:

- (1) *Die Vorstandsvorsitzende begrüßte die Praktikanten.
Der Abteilungsleiter sprach sie, bevor sie zum Mittagessen gingen.*

Das erste Vorkommen des Pronomens “*sie*” ist u.a. numerus-ambig - gemäß der Restriktion MORKONGRU kommen die Vorkommen “*Die Vorstandsvorsitzende*” (femininum,singular) und “*die Praktikanten*” (maskulinum,plural) gleichermaßen als Antezedenten in Frage. Hingegen ist die zweite “*sie*”-Okkurrenz qua syntaktischem Kontext - Kongruenzbedingung zwischen Subjekt und Vollverb - in seiner morphologischen Lesart auf die Merkmale (femininum,singular) festgelegt; zulässige Antezedenten sind daher “*die Praktikanten*” sowie das erstere Vorkommen von “*sie*”. Auf der Ebene der individuellen Entscheidungen wäre es somit möglich, die Antezedenszuordnungen “*sie*” (1. Vorkommen) → “*Die Vorstandsvorsitzende*” und “*sie*” (2. Vorkommen) → “*sie*” (1. Vorkommen) vorzunehmen - kombiniert sind diese Festlegungen jedoch inkompatibel, da sich eine Inkongruenz für das Numerus-Merkmal (singular bzw. plural) ergibt.

Präzise gesprochen handelt es sich im untersuchten Fall nicht um eine “morphologische” Mehrdeutigkeit, sondern um eine auf der inhaltlichen Ebene angesiedelte Ambiguität der *Kardinalität*, die im betrachteten Beispiel nicht unmittelbar an der sprachlichen Oberfläche - anhand der morphologischen Merkmale - abgelesen werden kann (daher die Anführungszeichen). Die beschreibungstheoretisch adäquate Lösung bestünde in der Einführung eines Kardinalitäts-Attributs auf der Ebene des *Diskursreferenten* (vgl. Habel, [Habe86]) sowie in der Operationalisierung einer entsprechend spezialisierten Interpretationsfunktion, die die Inferenz der ontologischen Kardinalität aus dem grammatischen Attribut des Numerus bewerkstelligt. Da die Abbildung jedoch i.d.R. elementar ist, sprechen Gründe der Beschreibungs- und Implementierungsökonomie dafür, diese eigentlich inhaltliche Kongruenzprüfung alleine auf der Basis der den Okkurrenzbeschreibungen zugeordneten morphologischen Beschreibungen durchzuführen.

Der entscheidende Schritt besteht darin, inkompatible Entscheidungskombinationen per inkrementeller Reverifikation der Kongruenzbedingungen während des schrittweisen Treffens einzelner Entscheidungen zu erkennen und zu unterbinden (vgl. Stuckardt: [Stu96a]). Die Strategie

```
(defstruct (ostknoten)
  synkateg   ;;; syntaktische Kategorie des Teilbaums - S, NP, VP, PP ... -
              ;;; bzw. Information, dass es sich um einen dr-okk-Knoten handelt
  inhalt     ;;; Liste von Zeigern auf die Kinder-ostknoten
              ;;; bzw. Zeiger auf die entsprechende Okkurrenz (dr-okk)
  dominator) ;;; Zeiger auf den Elter-ostknoten.
```

Abbildung 9.4: Datenstruktur für die Beschreibung von Knoten des Oberflächenstrukturbaums

MORKONGRU kommt somit in *zwei Phasen* des zu entwickelnden Algorithmus zum Einsatz: in einer ersten Phase, in der die lokale Verträglichkeit *individueller* Anapher-Antezedens-Paare getestet wird; in einer zweiten Phase der inkrementellen Auswahl, in der die Kongruenzbedingung erneut und unter Bezug auf Merkmalsmengen, die u.U. auf der Grundlage bereits getroffener Entscheidungen reduziert wurden, verifiziert wird. Da plausiblere Entscheidungsalternativen nicht zugunsten von weniger plausiblen Alternativen ausgeschlossen werden sollen, geschieht die schrittweise Auswahl in einer Ordnung absteigender Plausibilität. Vor der zweiten Phase der MORKONGRU-Anwendung ist daher (per Anwendung der Präferenzheuristiken) die Plausibilitätsbewertung vorzunehmen.

Anhand der Diskussion in Abschnitt 7.1.1 wurde ferner deutlich, daß die Operationalisierung der Strategie MORKONGRU auf entsprechenden Lexika basiert, in denen Stammformen sowie grammatisches Geschlecht spezifiziert werden. Falls der Stamm einer zu verarbeitenden Wortform im Lexikon der verwendeten morphologischen Analysekomponente *nicht* enthalten ist, sollten adäquate Techniken robuster Verarbeitung zum Einsatz gelangen. Hat das Lexikon einen hohen Abdeckungsgrad, so bietet sich die Verwendung einer Defaultregel an, derzufolge unbekannte Wortformen als *Namen* interpretiert werden, für deren morphologisches Merkmalsbündel heuristisch die Genus-Alternativen *maskulinum* und *femininum* (im Rahmen der weiteren Attributierungen *dritte Person* und *Singular*) angelegt werden. Als partielles Substitut könnte auf ein (domänenspezifisches) Lexikon mit eindeutigen morphologischen Spezifikationen (insbesondere des Genus) für die gängigen Namen der jeweiligen Anwendung zurückgegriffen werden.⁶ Die Operationalisierung eines solchen Analysemodells ist vergleichsweise elementar; sie wird im Rahmen der Beschreibung einer Implementierung des hier skizzierten Algorithmus näher erörtert werden.

9.3.2 Strategie SYNKONFIG

Anforderungen an die Datenstrukturen

Die Anwendung der syntaktisch-konfiguralen Koindexierungs-Restriktionen basiert auf den oberflächenstrukturellen Beschreibungen der zu interpretierenden Äußerungen. Für die systeminterne Modellierung dieser Syntaxbäume sind geeignete Datenstrukturen zu definieren, auf deren Grundlage die notwendigen Navigationsprozesse ablaufen können. Eine nähere Betrachtung der

⁶Für bestimmte Anwendungen, in denen die Menge der "typischen" Protagonisten einer relativ langsamen zeitlichen Änderung unterworfen sind - für Nachrichtenagentur-Meldungen etwa die Namen von Weltkonzernen bzw. von prominenten Politikern -, liegt die Verwendung und Pflege entsprechender Lexika durchaus innerhalb des Horizonts einer massendatentauglichen Textinhaltsanalyse.

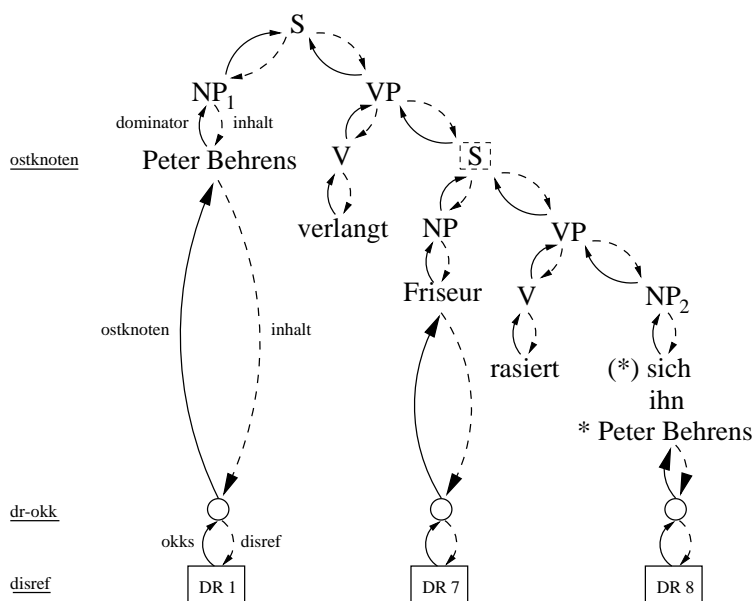


Abbildung 9.5: Verknüpfung der Beschreibungen von Okkurrenzen und Oberflächenstruktur

in Abschnitt 7.1.2 beschriebenen Definition des für die Bindungsbedingungen essentiellen K-Herrschafts-Begriffs ergibt, daß die Oberflächenstrukturbaume prinzipiell in zwei Richtungen zu traversieren sind - einerseits nach oben (zur Ermittlung des *nächsten verzweigenden Knotens*) und andererseits nach unten (zur Bestimmung von Dominanzrelationen zwischen zwei gegebenen Knoten). Darüberhinaus ist die syntaktische Kategorie eines Knotens relevant, denn diese Information determiniert insbesondere die Grenzknoten der lokalen Bindungsdomänen. Schließlich kann es notwendig sein, auf bestimmte, in den Vorkommensbeschreibungen verwaltete Information (etwa der syntaktischen Funktion) zurückzugreifen.⁷ Zusammen mit dem in Abschnitt 9.2 motivierten zeigerförmigen Merkmal *ostknoten* des Datentyps für Okkurrenzen ergibt sich somit eine *bidirektionale* datenstrukturelle Verknüpfung zwischen Vorkommens- und Oberflächenstrukturknoten-Deskriptionen. Abbildung 9.4 zeigt die Definition eines Datentyps *ostknoten*, die den identifizierten Anforderungen gerecht wird.

Operationalisierung der Koindexierungsbedingungen

Es soll nun näher erläutert werden, wie die Navigation auf den definierten Datenstrukturen im Rahmen der Verifikation syntaktisch-konfiguratoraler Bindungsbedingungen vonstatten geht. Abbildung 9.5 expliziert die internen Verweisstrukturen für eines der Beispiele aus Abbildung 7.3 (vgl. S. 118). Um etwa herauszufinden, ob die durch den Ausdruck "Peter Behrens" bezeichnete Okkurrenz einen in der vorliegenden syntaktischen Konfiguration zulässigen Antezedenten

⁷Die syntaktische Funktion geht natürlich aus der Beschreibung der Oberflächenstruktur hervor. Es handelt sich um eine im Rahmen der Implementierung getroffene Festlegung, diese nurmehr für die Verarbeitung der Vorkommen benötigte nichtstrukturelle Explikation der Rollen-Information als Bestandteil der Okkurrenzbeschreibungen zu modellieren. Unabhängig hiervon spiegelt der Syntaxbaum die Konfigurationsmuster wider, die mit bestimmten syntaktischen Rollen einhergehen - etwa die oben identifizierte Asymmetrie, derzufolge das Subjekt das Objekt k-beherrscht und nicht umgekehrt.

für das Pronomen “*ihn*” (zweite Variante des Beispiels) darstellt, sind zunächst die entsprechenden NP-Knoten (maximale Projektionen) NP_1 bzw. NP_2 zu ermitteln, was durch eine Navigation entlang der in den Attributen **ostknoten** (Vorkommensbeschreibung) bzw. **dominator** (Syntaxbaumbeschreibung) verwalteten Zeiger erfolgt. Ausgehend von NP_1 und NP_2 ist nun zu verifizieren, ob eine Koindexierung der beiden Okkurrenzen den Bindungsprinzipien genügt. Für das nichtreflexive Pronomen ist Bindungsprinzip B maßgebend, demzufolge NP_2 nicht innerhalb seiner bindenden Kategorie durch NP_1 k-beherrscht werden darf. Um dies zu überprüfen, wird ausgehend von NP_1 per Navigation entlang der **dominator**-Zeiger zunächst der *nächste verzweigende Knoten* i.S.d. K-Herrschafts-Definition ermittelt (Ergebnis: der S-Wurzelknoten des Syntaxbaums); in einem zweiten Schritt wird per rekursivem Abstieg entlang der in der Liste **inhalt** verwalteten Zeiger nach einer Dominanzbeziehung zwischen diesem Knoten und NP_2 gesucht - da gemäß Bindungsprinzip B nur *lokale* Koindexierungen auszuschließen sind, bricht die Suche beim Erreichen des inneren S-Knotens ab, da dieser den Beginn einer neuen lokalen Bindungsdomäne (dargestellt durch einen gestrichelten Kasten um den Knoten) kennzeichnet. Eine vergleichbare Suche ist von NP_2 aus zu starten, um die Bindungsbedingung des Antezedenskandidaten zu überprüfen, wobei in diesem Fall - es gilt BP C - unbegrenzt nach unten zu suchen ist; der Ablauf der Navigation ist ansonsten identisch.

Abbildung 9.6 gibt eine (bezüglich einiger Details leicht vereinfachte) Übersicht der Operationalisierung der zentralen Funktionen zur Verifikation der Bindungsprinzipien A, B und C. Die Funktionen **bp-konstruktiv-erfüllt?** und **bp-nicht-verletzt?** explizieren die Überprüfung der Bindungsbedingungen für die Anapher bzw. das Antezedens. Der Ablauf entspricht dem zuvor geschilderten Verfahren. In Schritt (1) werden die entsprechenden NP-Knoten der Oberflächenstruktur ermittelt; Schritt (2) initiiert eine Fallunterscheidung anhand der bindungstheoretischen Klasse des jeweilig relevanten Vorkommens: Ausgehend von dem per Aufruf der Funktion **naechst-verzweigender-knoten** ermittelten *nächstverzweigenden Knoten* wird nach Vorhandensein oder Fehlen entweder einer lokalen oder einer unbeschränkten Dominanzrelation Ausschau gehalten (Funktion **lokal-enthalten-in?** (Schritte (2A), (2B)) bzw. Funktion **enthalten-in?** (Schritt (2C))). Anhand der Beschreibung der Funktion **lokal-enthalten-in?** wird ersichtlich, daß die syntaktische Kategorie des jeweils zu traversierenden Knotens als Kriterium zur lokalen Begrenzung der Rekursion herangezogen wird (Randmarkierung (3), Aufruf der Prädikatfunktion **potentielle-bindende-kategorie?**). Die Suche bricht insbesondere dann ab, wenn ein S-Knoten eines finiten Satzes bzw. eine NP mit lokalem Subjekt (bes. Possessiva) vorliegt.⁸ Die Funktion **enthalten-in?** unterscheidet sich nur insofern, als auf dieses Abbruchkriterium verzichtet wird und die Traversierung des Syntaxbaums erst in den Blättern endet. Beiden Funktionen ist gemein, daß - entsprechend den Vorgaben der K-Herrschafts-Definition - die Traversierung des Teilbaums der potentiell k-beherrschenden NP, von der ausgegangen wurde, unterbleibt (Randmarkierung (4)).

⁸Zu berücksichtigen wäre ferner die in der Definition 7.4 der Bindenden Kategorie postulierte Zusatzbedingung für Typ-A-Pronomen, derzufolge eine (hypothetische) Koindexierung mit dem jeweiligen SUBJEKT nicht durch den i-über-i-Filter ausgeschlossen sein darf. Da ihr unter Anwendungsgesichtspunkten nur eine marginale Rolle zukommt, ist diese prinzipiell leicht zu operationalisierende Feinheit z.Zt. noch nicht implementiert (wohl aber der i-über-i-Filter als unabhängige bindungstheoretische Restriktion, vgl. u.). Die genaue Formalisierung des Begriffs der Bindenden Kategorie ist selbst in der Theorie nicht abschließend geklärt. Ein bekanntes Problem besteht in der Nicht-Komplementarität der Antezedensoptionen für Typ-A- und Typ-B-Pronomen in folgenden Belegen (zitiert nach Correa: [Corr88], S. 124):

- (a) *The children_i like each other's_j pictures.*
- (b) *The children_i like their_i pictures.*

Auch hier wird sich jedoch zeigen, daß die noch ausstehende adäquate Lösung keine Auswirkungen auf die Anwendungsperformanz des Algorithmus haben wird.

```

(defun bp-konstruktiv-erfuellt? (NP2-okk NP1-okk)
  (let ((NP1-ostkn (ostknoten-dominator (dr-okk-ostknoten NP1-okk)))      ;;; (1)
        (NP2-ostkn (ostknoten-dominator (dr-okk-ostknoten NP2-okk))))
    (case (dr-okk-bt-klasse NP2-okk)                                       ;;; (2)
      (A (let ((nv-knoten
                (naechst-verzweigender-knoten
                 (ueberspringe-synkateg-falls-pp-agens NP1-okk)))          ;;; (5)
              (lokal-enthalten-in? NP2-ostkn nv-knoten NP1-ostkn)))        ;;; (2A)
        (B (let ((nv-knoten
                (naechst-verzweigender-knoten
                 (ueberspringe-synkateg-falls-pp-agens NP1-okk)))          ;;; (5)
              (not (lokal-enthalten-in? NP2-ostkn nv-knoten NP1-ostkn)))) ;;; (2B)
          (C (let ((nv-knoten (naechst-verzweigender-knoten NP1-ostkn))
                  (not (enthalten-in? NP2-ostkn nv-knoten NP1-ostkn)))))) ;;; (2C)
            );;bp-konstruktiv-erfuellt?

(defun bp-nicht-verletzt? (NP1-okk NP2-okk)
  (let ((NP1-ostkn (ostknoten-dominator (dr-okk-ostknoten NP1-okk)))      ;;; (1)
        (NP2-ostkn (ostknoten-dominator (dr-okk-ostknoten NP2-okk))))
    (case (dr-okk-bt-klasse NP1-okk)                                       ;;; (2)
      (A t)                                                                 ;;; (2A)
      (B (let ((nv-knoten (naechst-verzweigender-knoten NP2-ostkn))
              (not
               (lokal-enthalten-in? NP1-ostkn nv-knoten NP2-ostkn))))      ;;; (2B)
        (C (let ((nv-knoten (naechst-verzweigender-knoten NP2-ostkn))
              (not
               (enthalten-in? NP1-ostkn nv-knoten NP2-ostkn))))))          ;;; (2C)
            );;bp-nicht-verletzt?

(defun lokal-enthalten-in? (ziel domaene domaeneninduktor)
  (unless (eq domaene domaeneninduktor)                                    ;;; (4)
    (do* ((ostknlste (ostknoten-inhalt domaene) (cdr ostknlste))
          (ostkn (car ostknlste) (car ostknlste)))
      ((or (eq ziel ostkn)
           (null ostknlste)
           (and (let ((sk (ostknoten-synkateg ostkn))
                     (not (or (eq sk *ost-dr-okk*)
                               (potentielle-bindende-kategorie? sk))))
                  (lokal-enthalten-in? ziel ostkn domaeneninduktor)))
            (not (null ostknlste))))))
    );;lokal-enthalten-in?

(defun enthalten-in? (ziel domaene domaeneninduktor)
  (unless (eq domaene domaeneninduktor)                                    ;;; (4)
    (do* ((ostknlste (ostknoten-inhalt domaene) (cdr ostknlste))
          (ostkn (car ostknlste) (car ostknlste)))
      ((or (eq ziel ostkn)
           (null ostknlste)
           (and (not (eq (ostknoten-synkateg ostkn) *ost-dr-okk*)
                       (enthalten-in? ziel ostkn domaeneninduktor)))
            (not (null ostknlste))))))
    );;enthalten-in?

```

Abbildung 9.6: Funktionale Operationalisierung der Bindungsprinzip-Verifikation

```
(defun i-ueber-i-bedingung? (NP2-okk NP1-okk)
  (let ((np2-ostkn (ostknoten-dominator (dr-okk-ostknoten NP2-okk)))
        (np1-ostkn (ostknoten-dominator (dr-okk-ostknoten NP1-okk))))
    (not
     (or (enthalten-in? np2-ostkn np1-ostkn)
         (enthalten-in? np1-ostkn np2-ostkn))))
  );; i-ueber-i-bedingung?
```

Abbildung 9.7: Funktionale Operationalisierung des i-über-i-Filters

Die Unterscheidung zweier Funktionen **bp-konstruktiv-erfüllt?** und **bp-nicht-verletzt?** erklärt sich in erster Linie durch die Berücksichtigung einer bereits in Abschnitt 7.1.2 identifizierten Feinheit, derzufolge das für Reflexiv- und Reziprokpronomen anzuwendende BP A zwar eine *lokale* Koindexierung vorschreibt, jedoch weitere *nichtlokale* Koindexierungen nicht ausschließt:

(2) *Behrens_i verspricht, daß er_i sich_i rasiert.*

Es wird daher zwischen einer *konstruktiven* und einer *nicht-konstruktiven* Verifikation der Bindungsprinzipien unterschieden: Erstere Funktion behandelt Typ-A-Pronomen als *zu resolvierende* Anapher - der aktuell vorgeschlagene Kandidaten wird daher im Hinblick auf seine Tauglichkeit als lokaler Binder untersucht; letztere Funktion realisiert die passive Sicht eines Reflexivums oder Reziprokums etwa als Antezedenskandidat - Bindungsprinzip A ist daher trivialerweise stets erfüllt.⁹

Ein weiterer Unterschied läßt sich anhand der Bewertung von Belegen der folgenden Art illustrieren:

(3a) *Der Brief wurde von Behrens_i an sich_i selbst adressiert.*

(3b) * *Der Brief wurde von Behrens_i an ihn_i adressiert.*

Da keine K-Herrschafts-Beziehung zwischen den koindexierten Vorkommen besteht, kommt die Bindungstheorie zu einer genau umgekehrten - intuitiv falschen - Vorhersage. Primus sowie Pollard und Sag unterbreiten Modellierungsvorschläge, denzufolge Vorkommen in der thematischen Rolle des *Agens* - unabhängig von der *syntaktischen* Funktion, in der sie realisiert sind - eine in der Bindungstheorie nicht erfaßte *nichtstrukturelle Prominenz* zukommt, die unabhängig von der oberflächenstrukturellen Konfiguration eine (ausschließliche) Zulässigkeit *reflexiven* lokalen Wiederaufgriffs bedingt (vgl. [Prim89] bzw. [PoSa92]). Insofern thematische Rollen zur Verfügung stehen, bewerkstelligt Funktion **uebersprunge-synkateg-falls-pp-agens** die gewünschte Differenzierung (Randmarkierung (5)).

Die Operationalisierung des *i-über-i-Filters* aus Definition 7.5 verläuft analog (vgl. Abbildung

⁹Dies kommt natürlich nur dann zum Tragen, falls Typ-A-Pronomen überhaupt als taugliche Antezedenskandidaten angesehen werden, denn es gibt durchaus Gründe dafür, den Wiederaufgriff von Reflexiva und Reziproka algorithmisch zu blockieren: Der entsprechende Diskursreferent ist (außer in wenigen exotischen Fällen) ohnehin im selben Satz und i.d.R. mit größerer Prominenz nichtreflexiv realisiert.

Im weiteren Verlauf der Diskussion wird sich jedoch ein weiteres Anwendungsgebiet der Verifikationsfunktion **bp-nicht-verletzt?** auftun, für das sich die beschriebene Eigenschaft unabhängig von der Frage der Antezedens-tauglichkeit von Typ-A-Pronomen als essentiell erweisen wird.

9.7). In diesem Fall ist auf unzulässige Dominanzbeziehungen in beiderlei Richtungen zu untersuchen. Die Tiefentraversierung geht unmittelbar von den NP-Knoten der beteiligten Okkurrenzen aus.

Entscheidungsinterdependenzen

Die bisherige Beschreibung der Operationalisierung syntaktisch-konfiguratoraler Restriktionen erfolgte ausschließlich aus der Perspektive individueller Entscheidungen: Betrachtet werden Paare bestehend aus anaphorischem Vorkommen und Antezedenskandidaten, deren Koindexierung auf konfigurationale Zulässigkeit hin zu überprüfen sind. Daß eine Beschränkung auf diese Betrachtungsebene aufgrund der Transitivität der Koindexierungsrelation zu unzulässigen Mustern von Referenzindexen führen kann, wurde bereits in Abschnitt 7.1.2 mit Blick auf folgendes Beispiel identifiziert:

(4) * *Behrens_i verlangt vom Friseur_j, daß er_i ihn_i rasiert.*

Obwohl individuell zulässig, dürfen nicht beide Personalpronomen des Nebensatzes zugleich mit dem Antezedenskandidaten “*Behrens*” koindexiert werden, da sonst - transitiv - Bindungsprinzip B der Anapher “*ihn*” verletzt wird, da eine lokale Bindung durch das Pronomen “*er*” die Folge wäre. Ähnlich wie zuvor bereits für die Strategie MORKONGRU ermittelt, bestehen also auch im Rahmen der Anwendung der bindungstheoretischen Bedingungen *Interdependenzen*, die zu unzulässigen Entscheidungskombinationen führen können, die durch ein geeignetes Design des Rahmenalgorithmus auszuschließen sind.

Das Problem der Entscheidungsinterdependenzen stellt kein Defizit der zugrundeliegenden Theorie dar, denn die deklarative Spezifikation der oberflächenstrukturellen Restriktionen schließt intuitiv unzulässige Indexdistributionen wie die in Beleg (4) beschriebene anforderungsgemäß aus. In einer Originalarbeit zur Bindungstheorie schlägt Chomsky darüberhinaus einen ersten Ansatz zur Operationalisierung der Bindungsbedingungen vor, der ausschließlich *zulässige* Indexverteilungen erzeugt ([Chom81], S. 186 ff.). Er basiert auf dem Konstrukt der sog. “*free indexing rule*”, durch die zunächst - ohne Berücksichtigung etwaiger Koindexierungsbedingungen - alle kombinatorisch möglichen Indexmuster generiert werden; erst in einem zweiten Schritt kommen die Kriterien der Bindungstheorie zur Anwendung, um die Teilmenge der theoretisch zulässigen Indexdistributionen zu ermitteln:

(5) *Peter Behrens berät seinen Kunden.*

	<i>i</i>	<i>j</i>	<i>k</i>
*	<i>i</i>	<i>j</i>	<i>i</i>
	<i>i</i>	<i>i</i>	<i>k</i>
	...		

Eine unmittelbare Implementierung dieses Ansatzes ergibt einen “*Generate-and-Test*”-Algorithmus, der zwar die Bindungsbedingungen adäquat abbildet, jedoch eine (in Abhängigkeit von der

Vorkommensanzahl) exponentielle Laufzeitkomplexität aufweist. Als Ad-Hoc-Nachweis der Operationalisierbarkeit der Bindungstheorie ist dieses Verfahren somit allenfalls von theoretischem Interesse.

Verfahren, die die bindungstheoretischen Bedingungen alleine unter Bezug auf Einzelentscheidungen, d.h. auf Paare bestehend aus anaphorischer Okkurrenz und Antezedenskandidat verifizieren, vermeiden zwar die exponentielle Laufzeitkomplexität des Originalansatzes; der Vorteil wird jedoch teuer erkauft, da diese Operationalisierungsstrategie, wie durch Beispiel (4) expliziert wird, die Bedingungen der Bindungstheorie nicht in adäquater Weise umsetzt. Da es mit der anwendungsorientierten Perspektive der Anaphernresolution in Einklang steht und darüberhinaus Implementierungstechnisch einfacher ist, die referentielle Interpretation unter Bezug auf elementare Teilprobleme in Form *einzelner* zu resolvierender anaphorischer Vorkommen anzugehen, liegt die zu kurz greifende einzelentscheidungsbezogene Anwendung der konfiguralen Kriterien förmlich auf der Hand. Dabei ist die algorithmische Berücksichtigung der Entscheidungsinterdependenzen anwendungsrelevant, denn Belege von der Gestalt des in (4) angegebenen - zwei Typ-B-Pronomen innerhalb derselben lokalen Bindungsdomäne, wovon eines das andere k-beherrscht - können (je nach Textsorte) durchaus häufig vorkommen. Insbesondere betrifft das geschilderte Problem *Relativpronomen*, die auch in Fällen, in denen deren Koindexierung syntaktisch determiniert¹⁰ ist, eine vergleichbare transitive Fernwirkung induzieren können:

(6) * Behrens_i kennt den Friseur_j, der_j ihn_j rasiert.

Unter den in Kapitel 8 beschriebenen Ansätzen, die die syntaktischen Koindexierungsrestriktionen operationalisieren, ignoriert u.a. der *naive Algorithmus* von Hobbs die Subtilität potentieller Entscheidungsinterdependenzen.

In der Literatur werden darüberhinaus einige Algorithmen beschrieben, die keine vollständigen Anaphernresolutionsverfahren darstellen, sondern die begrenzte Zielsetzung verfolgen, die bindungstheoretischen Bedingungen unter Vermeidung der inakzeptablen Laufzeitkomplexität des "Generate-and-Test"-Verfahrens von Chomsky zu operationalisieren. Eine nähere Untersuchung dieser Ansätze führt jedoch zu dem überraschenden Ergebnis, daß der Effizienzgewinn auch hier durch eine (teils stillschweigende) Ignorierung der Interdependenzproblematik bzw. - je nach Sichtweise - eine nur partielle Implementierung der Bindungsrestriktionen erkauft wird. Der *Algorithmus von Correa* basiert auf einer einmaligen Traversierung des (als vollständig und eindeutig vorausgesetzten) Oberflächenstrukturbaums, in deren Rahmen für alle Typ-A- und Typ-B-Pronomen je eine Menge bindungsprinzipiell zulässiger Antezedenskandidaten vorberechnet wird ([Corr88]); ausgehend von dieser Information bewerkstelligt eine nur sehr abstrakt spezifizierte Auswahlregel die Kandidatenauswahl, ohne das Interdependenzproblem zu berücksichtigen. In vergleichbarer Weise arbeitet das *Verfahren von Ingria und Stallard*, das jedoch einerseits genereller ist, da auch kataphorische Bezugnahmen Berücksichtigung finden ([InSt89]), und andererseits partiell ist, da keine Kandidatenauswahl stattfindet - das Interdependenzproblem wird also nicht aktuell und somit auch nicht gelöst.¹¹ Eine nähere Analyse

¹⁰Wie im Rahmen des folgenden Kapitels näher ausgeführt wird, fließt in die Konstruktion der oberflächenstrukturellen Beschreibung prinzipiell semantisches (auch: referentielles) Wissen ein; im Speziellen gilt dies für die strukturelle Interpretation von Relativsätzen, durch die die Koindexierung des Relativpronomens determiniert wird. Insofern ist die hier geschilderte Abhängigkeit nicht unidirektional.

¹¹Unabhängig von dieser Einschränkung stellt der Algorithmus von Ingria und Stallard eine äußerst elegante *partielle* Implementierung der Bindungsrestriktionen dar, durch die sich eine potentiell effizientere Alternative für den in den Abbildungen 9.6 und 9.7 beschriebenen Operationalisierungsansatz ergibt (vgl. u.).

zeigt auf, daß beide Ansätze eine (wiederum bezogen auf die Anzahl der Okkurrenzen bzw. NPs) quadratische Laufzeitkomplexität aufweisen.¹² Auch *Giorgi, Pianesi und Satta* stellen *zwei Algorithmen zur Operationalisierung der Bindungstheorie* mit quadratischer Laufzeitkomplexität vor, die die für die unter die Bindungsprinzipien A bzw. B fallenden Pronomen bestehenden Bedingungen aus der Perspektive *individueller* Entscheidungen verifizieren ([GPS90]).¹³ Zwar wird die Behandlung möglicher Entscheidungsinterdependenzen als wesentliche Problemstellung der Antezedensauswahlphase identifiziert (ibid., S. 124)

“[...] it is necessary to put together the constraints that have been separately computed for each item according to Principles A and B (and C); [...].”

ein konkreter Vorschlag zur algorithmischen Lösung dieses weitergehenden (ibid.)

“[...] problem of BT verification, i.e. whether a given index assignment for the NPs of a sentences complies with the restrictions of BT”

wird jedoch nicht unterbreitet.

Um die Kombination inkompatibler Einzelentscheidungen zu vermeiden, bedürfen die in den Abbildungen 9.6 und 9.7 spezifizierten Verifikationsfunktionen somit einer Verfeinerung durch einen anwendungspraktikablen (d.h. *laufzeit-effizienten*), in den einschlägigen Operationalisierungsansätzen noch nicht beschriebenen Mechanismus. Es bietet sich an, erneut auf das Verarbeitungsmodell zurückzugreifen, das in Abschnitt 9.3.1 zur Lösung des Interdependenzproblems der Strategie MORKONGRU skizziert wurde. Die adäquate Operationalisierung sowohl der bindungstheoretischen Bedingungen als auch der morphologischen Kongruenzrestriktion wird demnach durch einen *dreiphasigen Rahmenalgorithmus* geleistet (vgl. Stuckardt: [Stu96a]):

1. *Restriktionsanwendung* zur Ermittlung der *individuell* zulässigen Antezedenskandidaten;
2. *Präferenzbewertung* der Antezedenskandidaten;
3. *Schrittweise, plausibilitätsgeordnete Auswahl der Antezedenten*: Restriktionen, die qua Entscheidungsinterdependenz transitiv verletzt werden können, sind in geeigneter Weise zu *reverifizieren*.

Durch die Antezedensauswahl in der Reihenfolge absteigender Plausibilität wird verhindert, daß Einzelentscheidungen, die als weniger plausibel bewertet werden, qua Interdependenz zum Ausschluß von Alternativen mit höherer Plausibilität führen. In bezug auf die bindungstheoretischen Restriktionen kommt dem als psycholinguistisch evident identifizierten Kriterium des syntaktischen Parallelismus eine zentrale Rolle zu.

¹²Correa gibt für seinen Algorithmus eine lineare Komplexität an, läßt jedoch unberücksichtigt, daß die elementaren Operationen, auf denen sein Verfahren aufbaut, z.T. nichtkonstante Kosten verursachen.

¹³In Abgrenzung zur Arbeit von Ingria und Stallard liegt den Algorithmen von Giorgi, Pianesi und Satta der Anspruch zugrunde, eine adäquatere Operationalisierung von Domänen lokaler Bindung - insbesondere solcher, die durch NPs induziert werden - zu bewerkstelligen (ibid., S. 124).

```
(defstruct (disref)
  disrefnum      ;;; eindeutiger Identifikator (Nummer einer der Okkurrenzen)
  okks          ;;; Liste kospezifizierender Okkurrenzen
  lokale-okks)  ;;; temporaere Liste kospezifizierender Okkurrenzen
                ;;; (fuer Interdependenztest)
```

Abbildung 9.8: Datenstruktur für die Beschreibung von Diskursreferenten (verfeinert)

- (7a) *Behrens_i notiert, daß er_j ihn_k besuchen wird.*
 (7b) *Behrens_i läßt Gropius_j ein, weil er_k ihn_l bereits lange kennt.*

Einerseits etabliert es in Äußerungen der Gestalt (7a) eine Präferenzordnung, durch die eine eindeutige Auswahlreihenfolge geschaffen wird, die alleine auf der Grundlage der Subjektpräferenz nicht gegeben wäre - die Entscheidung "er" → "Behrens" würde demnach gegenüber der nicht die syntaktische Funktion erhaltenden interdependierenden Zuordnung "ihn" → "Behrens" bevorzugt. Andererseits bewirkt das syntaktische Rollenträgheitskriterium, daß Fälle von Entscheidungsinterdependenz oftmals a priori vermieden werden, da es in Belegen der Form (7b) die nichtinterdependierenden Entscheidungen "er" → "Behrens" und "ihn" → "Gropius" als erste Wahl klassifiziert. Es besteht somit die glückliche Situation, daß die als psycholinguistisch evident bekannte Parallelismus-Präferenz zumindest einen Teil der (zu zusätzlichem Berechnungsaufwand führenden) Auswahlkonflikte von vorneherein vermeidet. Dennoch werden - siehe Beispiel (7a) - nicht alle Interdependenzen vermieden, jedoch in solchen Fällen zumindest eine heuristisch plausible Entscheidungskombination bevorzugt.

Die Vorgehensweise, die plausibleren gegenüber den weniger plausiblen Einzelentscheidungen zu bevorzugen, ist insofern heuristisch, als nicht garantiert ist, daß die Auswahlkombination mit der größtmöglichen *kumulierten* Plausibilität generiert wird. Es ist jedoch nicht zu erwarten, daß dieses Defizit des Verarbeitungsmodells in der praktischen Anwendung zum Tragen kommt. Die Bewertung der Einzelentscheidungen geschieht ohnehin bereits unter Anwendung heuristischer Kriterien, durch die von tieferen inhaltlichen oder gar pragmatischen Gesichtspunkten weitgehend abstrahiert wird; der theoretische Status der nächsthöheren Ebene einer *kumulierten Plausibilität* ist somit ohnehin fragwürdig.

Um eine effiziente Algorithmisierung der erneuten Verifikation der bindungstheoretischen Bedingungen für die transitiv entstehenden Koindexierungen in Schritt 3 zu ermöglichen, bietet sich eine Verfeinerung der in Abbildung 9.3 beschriebenen Datenstruktur für Diskursreferenten an. Ausgangspunkt ist die Beobachtung, daß nur solche Vorkommen von Anapher bzw. Antezedenskandidat potentielle Auswahlkonflikt-Induktoren verkörpern, die zur aktuell interpretierten Äußerung gehören.¹⁴ Dies gilt unter der (unproblematischen) Voraussetzung, daß je anaphorischer Okkurrenz nur maximal ein Antezedens ausgewählt wird. Es bietet sich daher an, die satzlokalen Vorkommen der Diskursreferenten zunächst unter einem eigenständigen mengenwertigen Attribut `lokale-okks` zu verwalten (vgl. Abbildung 9.8). Im Rahmen des Interdependenztests wird das Kreuzprodukt - die Menge aller Okkurrenzpaare - über den `lokale-okks`-Mengen von Antezedenskandidat und Anapher gebildet; für jedes dieser - zur transitiven Koindexierung ausstehenden - Vorkommenspaare findet eine Verifikation der Bindungsbedingungen statt.¹⁵ Die Vorkommensbeschreibungen, die unter dem Attribut `okks` abliegen, sind hierbei irrelevant.

¹⁴Die Interdependenz kann jedoch vermöge einer identischen *satzübergreifenden* Antezedensentscheidung für diese beiden Okkurrenzen *vermittelt* werden.

¹⁵Für das hier ebenfalls enthaltene Paar (Kandidat, Anapher) kann die erneute Verifikation natürlich entfallen.

Mit Blick auf die Zulässigkeit der Index-Distribution von Beleg (2) wird ferner deutlich, daß sich der Interdependenztest ausschließlich der *schwachen* Version der Bindungsprinzipverifikation - Funktion `bp-nicht-verletzt?` - bedienen darf, da - wie oben beschrieben - transitiv entstehende Koindexierungen von Reflexiv- und Reziprokpronomen durch BP A nicht untersagt werden.

9.4 Der Basisalgorithmus

9.4.1 Integration der Strategien in das dreiphasige Rahmenverfahren

Ausgehend von den Ergebnissen der Untersuchungen zur Operationalisierung der Einzelstrategien soll nun eine erste vollständige Version eines adäquaten Verfahrens für die referentielle Interpretation anaphorischer Ausdrücke erarbeitet werden. Abbildung 9.9 zeigt den Basisalgorithmus, der die in Abschnitt 9.1 ausgewählten Strategien in das dreiphasige Rahmenverfahren integriert, das in Abschnitt 9.3 mit dem zentralen Ziel der Vermeidung inkompatibler Entscheidungskombinationen erarbeitet wurde.

Entsprechend dem grundlegenden Vorschlag, den Carbonell und Brown für Mehrstrategieverfahren zur Anapherninterpretation unterbreiten, werden die stringenten Kriterien (Phase 1 des Algorithmus) vor den Präferenzfaktoren (Phase 2 des Algorithmus) angewendet, um den Bewertungs- und Sortieraufwand in Phase 2 des Algorithmus durch eine möglichst frühe Elimination unzulässiger Antezedenskandidaten weitestmöglich zu reduzieren. Eine vergleichbare Zielsetzung liegt der Designentscheidung zugrunde, innerhalb der Restriktionsverifikation (Phase 1) zunächst die mit weniger Berechnungsaufwand einhergehende hochrelevante morphologische Kongruenzbedingung zu verifizieren, bevor die vergleichsweise aufwendigere Überprüfung der ebenfalls hochrelevanten bindungstheoretischen Restriktionen vonstatten geht. Zentrales Merkmal des Algorithmus ist die im vorigen Abschnitt motivierte Entkopplung von Präferenzbewertung/Sortierung in Phase 2 von der Phase 3 der Auswahl, in deren Rahmen eine erneute Verifikation derjenigen Kriterien stattfindet, die potentiell zu Entscheidungskonflikten führen. Im Rahmen der einzelentscheidungsbezogenen Phase 1 des Algorithmus werden insbesondere die als hochrelevant identifizierten bindungstheoretischen Kriterien verifiziert. Entsprechend den Vorbemerkungen von Abschnitt 9.3 zerfällt dieser Prozeß in drei Teilschritte: konstruktive Verifikation des Bindungsprinzips der Anapher (Schritt 1(b)i), nichtkonstruktive Verifikation des Bindungsprinzips des Antezedenskandidaten (Schritt 1(b)ii) und Filterung unzulässiger i-über-i-Konfigurationen (Schritt 1(b)iii). Die in den Abbildungen 9.6 und 9.7 beschriebenen funktionalen Operationalisierungen dieser Teilkriterien gelangen hier erstmals zur Anwendung.

Auf der Grundlage der in Abschnitt 9.1 ausgewählten Prominenzfaktoren im engeren bzw. weiteren Sinne erfolgt in Phase 2 zunächst die Plausibilitätsbewertung der für die einzelnen anaphorischen Elemente verbliebenen Antezedensoptionen (Schritt 2a). Dieser Schritt ist bislang insofern unterdeterminiert, als eine Festlegung der numerischen Plausibilitätsgewichte für die einzelnen Kriterien noch aussteht. Diese Lücke verbleibt im Rahmen der Beschreibung einer konkreten Implementierung und in Abhängigkeit der tatsächlich operationalisierten Faktoren zu schließen. Hierbei ist außerdem der Zielsetzung der *globalen* Bevorzugung erwartet plausiblerer *Typen* von Einzelentscheidungen Rechnung zu tragen - beispielsweise könnten zeichenkettenidentische Antezedenskandidaten von *Namensanaphern* generell als plausibler eingestuft werden als die Antezedenskandidaten für die i.a. schwieriger zu resolvierenden *Pronominalanaphern*. Die

1. Für jede anaphorische Okkurrenz Y , ermittle die Menge der möglichen Antezedenten X :
 - (a) Anapherentypabhängige Anwendung von Kongruenzrestriktionen:
 - (MORKONGRU) Falls Y pronominal ist, verifiziere die morphologische Kongruenz mit X .
 - (ZEICHKETT) Falls Y nichtpronominal, d.h. Name oder definite NP ist, verifiziere Namens- bzw. sonstige Übereinstimmung der lexikalischen Stammform mit X .
 - (b) (SYNKONFIG) Falls der Antezedenskandidat X intrasententiell ist, verifiziere, daß die Bindungsbedingungen von Y und X erfüllt sind: Für die vorgeschlagene Koindexierung
 - i. verifiziere, daß das Bindungsprinzip von Y *konstruktiv erfüllt* ist,
 - ii. verifiziere, daß das Bindungsprinzip von X *nicht verletzt* ist,
 - iii. verifiziere, daß keine i-über-i-Konfiguration vorliegt.
 - (c) (SEMROLLEN) Verifiziere die selektionale Verträglichkeit des Kandidaten X mit der Rolle, die das anaphorische Vorkommen Y ausfüllt.
 - (d) (DEFKATAPH) Falls Y ein Typ-B-Pronomen ist, der Antezedenskandidat X intrasententiell ist und X bezüglich der Oberflächenanordnung auf Y *folgt* (d.h. es liegt *kataphorischer* Wiederaufgriff vor), dann verifiziere, daß X *definit* ist.
2. Bewertung und Sortierung nach Plausibilität:
 - (a) Für jedes verbleibende Paar (Y_i, X_j) von Anapher und Antezedenskandidat: Ermittle - in Abhängigkeit des Typs von Y_i - einen numerischen Plausibilitätsgrad $v(Y_i, X_j)$, der X_j relativ zu Y_i bewertet, unter Rückgriff auf die Faktoren DISTANZ, TOPIKALIS, SUBJEKT, SYNROLLE, KATAMALUS, SYNPARALL und SEMPARALL.
 - (b) (*lokales Sortieren*) Für jede Anapher Y : Sortiere ihre individuellen Antezedenskandidaten X_j gemäß absteigender Plausibilität $v(Y, X_j)$.
 - (c) (*globales Sortieren*) Sortiere die Anaphern Y gemäß absteigender Plausibilität ihrer individuell besten Antezedenskandidaten.
3. Antezedensauswahl: Betrachte die Anaphern Y in der in Schritt 2c ermittelten Reihenfolge. Schlage die Antezedenskandidaten $X(Y)$ in der in Schritt 2b ermittelten Reihenfolge vor. Wähle $X(Y)$, falls keine Interdependenz besteht, d.h. genau dann wenn
 - (a) (MORKONGRU) die morphosyntaktischen Eigenschaften von Y und $X(Y)$ noch kompatibel sind,
 - (b) (SYNKONFIG) für alle Okkurrenzen Z_1 und Z_2 , die dem Diskursreferenten zu $X(Y)$ bzw. Y im *aktuellen* Lauf der Analyse zugeordnet wurden, gilt: Die Koindexierung von Z_1 und Z_2 , die transitiv aus der Wahl von $X(Y)$ als Antezedens für Y resultiert, *verletzt nicht* die Bindungsprinzipien und fällt nicht unter den i-über-i-Filter (Der Spezialfall $Z_1 = X(Y) \wedge Z_2 = Y$ braucht nicht erneut betrachtet zu werden.),
 - (c) Y und $X(Y)$ noch unterschiedlichen Diskursreferenten zugeordnet sind (Vermeidung zyklischer Wiederaufgriffsrelationen).

Abbildung 9.9: Der Basisalgorithmus

entsprechende Unterscheidung spiegelt sich in den sich anschließenden Sortierschritten wider: Zunächst werden die Kandidaten jeder Anapher *lokal* nach Plausibilität geordnet (Schritt 2b); um jedoch *global* die plausibelsten Einzelentscheidungen bevorzugen zu können, ist eine zweite Sortierphase notwendig, in deren Rahmen die zur Resolution anstehenden Anaphern entsprechend der Plausibilität ihres jeweils besten Kandidaten geordnet werden (Schritt 2c).

Die Auswahlphase 3 bedient sich der auf diese Weise gewonnenen anaphernlokalen bzw. globalen

Plausibilitätsordnungen. Auf der Grundlage von Information über bereits getroffene Entscheidungen wird nach Fällen von Entscheidungsinterdependenz Ausschau gehalten: Ist der individuell plausibelste Kandidat der betrachteten Anapher aufgrund eines transitiven Koindexierungskonflikts nicht mehr zulässig, so wird - basierend auf der lokalen Sortierung - der jeweils nächste Kandidat dieser Anapher als Substitut vorgeschlagen. Transitiv zu verifizieren sind die potentiell interdependenzbetroffenen Strategien MORKONGRU und SYNKONFIG. Entsprechend der obigen Diskussion kommt im Rahmen der transitiven Reverifikation der bindungstheoretischen Kriterien ausschließlich die nichtkonstruktive Verifikationsfunktion zur Anwendung. Der zusätzliche Schritt 3c zielt darauf ab, direkte und indirekte (transitiv entstehende) Relationen wechselseitigen Wiederaufgriffs zu verhindern. In folgendem Beleg

(8) ... \underline{Er}_i erwartet, daß Gropius \underline{ihn}_i anrufen wird.

würde eine wechselseitige Antezedenswahl $Er_i \rightarrow ihn_i$ und $ihn_i \rightarrow Er_i$ prinzipiell durch die morphologischen und syntaktisch-konfigurationsalen Restriktionen lizenziert; sie genügt jedoch nicht den unter Anwendungsgesichtspunkten zu stellenden Anforderungen, da für eines der beiden anaphorischen Elemente keine neue Information im Sinne der Zugehörigkeit zu einer bereits in den Diskurs eingeführten Kospezifikationsklasse beigetragen wird. Im betrachteten Beispiel könnte dies u.a. dazu führen, daß für die beiden pronominalen Erwähnungen des Diskursreferenten mit Index i keine konzeptuelles (nichtpronominales) Substitut gefunden wird. Die durch die Präferenzstrategie KATAMALUS induzierte Bevorzugung nichtkataphorischer Wiederaufgriffe bewirkt allerdings, daß die Interdependenzklausel 3c nur in wenigen Sonderfällen zum Tragen kommt, da in aller Regel ein links von der Anapher stehender Kandidat zur Verfügung steht und somit bevorzugt wird.

9.4.2 Laufzeitkomplexität

Eine auf die Schritte zur Verifikation der bindungstheoretischen Bedingungen bezogene Komplexitätsanalyse zeigt, daß der Algorithmus unter Gewährleistung der Korrektheit der generierten Indexdistributen die exponentielle Zeitkomplexität der in Abschnitt 9.3.2 diskutierten Regel der freien Indexierung aus Chomskys Originalarbeit vermeidet. Als geeignete Vergleichsbasis sei die Betrachtung beschränkt auf die Behandlung intrasententieller Antezedenskandidaten, deren Anzahl durch die Anzahl n der Knoten des Oberflächenstrukturbaums beschränkt ist.¹⁶ Es gibt somit nur $O(n^2)$ Paare von Anaphern und Antezedenskandidaten, für die die Bedingungen der Bindungstheorie sowie die übrigen Restriktionen zu verifizieren und ggf. Plausibilitätskriterien anzuwenden sind.¹⁷ Unter der vorläufigen Annahme, daß die Verifikation der bindungstheoretischen Bedingungen entsprechend der in den Abbildungen 9.6 und 9.7 spezifizierten Funktionen

¹⁶Gemäß einem Argument von Giorgi, Pianesi und Satta besteht für die Größe der oberflächenstrukturellen Beschreibung - und damit für deren Knotenanzahl n - wiederum eine obere Schranke, die linear in der Länge der eigentlichen Eingabe, d.h. der Wortanzahl des zu interpretierenden Satzes ist ([GPS90], S. 123). Die im folgenden durchgeführte Worst-Case-Abschätzung der Laufzeitkomplexität erscheint demnach unmittelbar auf den formal korrekten Bezugspunkt - die Größe des originären Inputs - übertragbar.

¹⁷Um die Gültigkeit dieses Arguments zu gewährleisten, ist es theoretisch notwendig, den Algorithmus um einen Mechanismus zu erweitern, ein bestimmtes Paar im Rahmen des Interdependenztests nur höchstens einmal bindungstheoretisch zu analysieren. Für das praktische Laufzeitverhalten erweist sich diese Feinheit jedoch als bedeutungslos.

bewerkstelligt wird, entstehen jeweils Kosten von $O(n)^{18}$, die durch die rekursive Traversierung des Oberflächenbaums bedingt sind; alle weiteren Operationen verursachen konstante Kosten. Da das zweistufige Sortieren in Schritt 2 nicht teurer ist - $O(n^2 \log n)$ entsprechend $O(n)$ Sortiervorgängen mit Komplexität $O(n \log n)$ -, ist die Laufzeitkomplexität des Algorithmus bezogen auf die intrasententielle Anapherninterpretation $O(n^3)$.

In bezug auf die Verifikation der bindungstheoretischen Bedingungen wird das *praktische* Laufzeitverhalten des Algorithmus durch zwei zentrale Faktoren determiniert: (I) die Anzahl der Paare, für die die syntaktisch-konfiguralen Bedingungen im Rahmen der Kandidatenfilterungsphase überprüft werden, und (II) die Anzahl der Paare, für die ein bindungstheoretischer Interdependenztest in der Antezedensauswahlphase durchgeführt wird. Der Beitrag von Faktor (I) wird reduziert, indem die preiswertere Kongruenzrestriktion zuerst angewendet wird. Der Beitrag von Faktor (II) wird durch das Rollenträgheitskriterium in Zusammenarbeit mit der Strategie, die Entscheidung mit der lokal höchsten Plausibilität zuerst zu betrachten, reduziert: Wie bereits anhand der Beispiele (7a) und (7b) diskutiert wurde, sind in dem für Interdependenz typischen Fall zweier in einer lokalen Domäne kopartizipierender Typ-B-Pronomen die individuell plausibelsten Antezedenskandidaten in der Regel nicht identisch.

Durch die Rahmenstrategie, die individuell plausibelsten Antezedenskandidaten zuerst in Erwägung zu ziehen, werden nicht alle Interdependenzkollisionen vermieden. Aus den genannten Gründen steht jedoch zumindest zu erwarten, daß die praktische Performanz des Algorithmus gut ausfällt und den Anwendungsanforderungen genügt.

9.4.3 Eine asymptotisch effizientere Implementierungsvariante

Unter Anwendung der gemeinsamen Grundidee, die den in Abschnitt 9.3.2 zitierten Ansätzen zur Verifikation der bindungstheoretischen Bedingungen zugrundeliegt, kann eine Verbesserung der Worst-Case-Laufzeitkomplexität des Basisalgorithmus erreicht werden, ohne auf die wichtige Eigenschaft der adäquaten Behandlung von Entscheidungsinterdependenzen zu verzichten. Demnach ist eine A-Priori-Berechnung *aller* aus den Bindungsprinzipien ableitbaren Bedingungen mit einem einmaligen quadratischen Berechnungsaufwand von $O(n^2)$ möglich (Correa, [Corr88]; Ingria und Stallard, [InSt89]; Giorgi, Pianesi und Satta, [GPS90]); diese Technik läßt sich unmittelbar auf die Anwendung des i-über-i-Filters übertragen. Für die $O(n^2)$ zu behandelnden Paare von Anapher und Antezedenskandidat verringert sich der Aufwand für die Verifikation der Bindungsbedingungen auf konstante Kosten, da nurmehr in einer vorberechneten Ergebnistabelle nachzuschlagen ist. Die theoretische Laufzeitkomplexität reduziert sich auf den Beitrag $O(n^2 \log n)$ des nunmehr teuersten Schritts, der Sortierphase in Phase 2.

Die Untersuchung der praktischen Eigenschaften der im folgenden vorgestellten konkreten Implementierung wird jedoch ergeben, daß eine entsprechend verfeinerte Ausgestaltung der Verifikation der bindungstheoretischen Bedingungen nicht notwendig ist. Unter Anwendungsbedingungen wird sich die in den Abbildungen 9.6 und 9.7 spezifizierte funktionale Implementierung als effizient erweisen. Ob es sich lohnt, alle zulässigen bzw. unzulässigen Koindexierungen a priori auszurechnen, hängt nicht zuletzt im jeweiligen Einzelfall davon ab, wie groß die Anzahl der auf der individuellen Entscheidungsebene morphologisch kongruenten Anapher-Antezedenskandidat-Paare ist und wie umfangreich demnach die Arbeit im Rahmen der Veri-

¹⁸Die O -Notation $g(n) = O(f(n))$ bezeichnet eine asymptotische obere Schranke. Demnach wächst die jeweilige Zielgröße $g(n)$ (hier: die Laufzeit des Basisalgorithmus) höchstens proportional zu $f(n)$, falls die Größe n des jeweiligen Inputs (hier: die Anzahl intrasententieller Okkurrenzen) gegen unendlich geht.

fikation der syntaktisch-konfiguralen Bedingungen ausfällt. Ist die zu erwartende Anzahl dieser Paare klein, so kann die oben beschriebene Operationalisierung der Kriterien durchaus die vorteilhaftere Implementierungstechnik sein.

9.5 Diskussion

9.5.1 Unterspezifikation des Basisalgorithmus

Die Spezifikation des Basisalgorithmus ist in mehrfacher Hinsicht partiell. Einer der wesentlichen Punkte, die im Rahmen einer Implementierung näher auszugestalten sind, betrifft - wie oben näher ausgeführt - die Festlegung numerischer Gewichte für die einzelnen Plausibilitätsfaktoren sowie ferner die anaphertyp-abhängige Auswahl adäquater Teilmengen anzuwendender Faktoren; z.B. ist unmittelbar ersichtlich, daß das Abstandskriterium DISTANZ für Typ-A-Anaphern irrelevant ist, da diese stets satzlokal zu binden sind. Vergleichbare Entscheidungen verbleiben bezüglich der jeweils anwendbaren Restriktionen in Phase 1 des Algorithmus zu treffen: Für die referentielle Interpretation von Namensanaphern sowie sonstiger nichtpronominaler definiter NP sind Subalgorithmen zu entwickeln, die geeignete Tests auf der Grundlage von Zeichenkettenvergleichen und evtl. Abkürzungsverzeichnissen spezifizieren. Da die Operationalisierung dieser Aufgaben zumindest in Teilen unter Berücksichtigung anwendungsspezifischer Gegebenheiten zu formulieren ist und deren uneingeschränkte Algorithmisierbarkeit nicht in Frage steht, sollen diese Teilprobleme nicht im Rahmen der Beschreibung des Basisalgorithmus diskutiert werden. Darüberhinaus verbleibt festzulegen, auf welche Weise überhaupt der *Input* des Basisalgorithmus erzeugt wird. Welcher sprachliche Ausdruck konstituiert ein spezifizierendes Vorkommen? Welches Vorkommen verkörpert eine Anapher, welches ein potentielles Antezedens? Auch diese Fragen sind erst unter Rekurs auf die durch ein konkretes Voranalyse-Szenario zur Verfügung gestellte morphologische und syntaktische Information zu beantworten; sie werden daher erst im Rahmen der Beschreibung der Implementierung in Kapitel 11 erörtert.

9.5.2 Berücksichtigung bindungstheoretischer Feinheiten: leere Kategorien

In der GB-Theorie wird eine große Bandbreite bindungstheoretischer Phänomene unter Rekurs auf sog. *leere Kategorien* modelliert, die sich in erster Näherung als oberflächenstrukturelle Elemente beschreiben lassen, für die kein Pendant auf der Ausdrucksebene existiert. Eine möglichst weitreichende Korrektheit der bindungstheoretischen Vorhersagen des Basisalgorithmus ist somit nur unter der Voraussetzung gewährleistet, daß auch dieser auf der oberflächenstrukturellen Ebene angesiedelte Typus *impliziter Mitspieler* im Prozeß der Identifikation zulässiger Indexdistributionen adäquat berücksichtigt werden kann. Dieser Punkt ist von primär theoretischem Interesse, da gegenwärtige anwendungstaugliche Syntaxanalytoren i.d.R. keine Oberflächenstruktur liefern, in denen leere Kategorien entsprechend den vollständigen Anforderungen der GB-Theorie angelegt sind. Ein Gewinn für die praktische Anaphernresolution wäre jedoch auch bereits dann gegeben, wenn die entsprechenden Feinheiten - in Abhängigkeit der verfügbaren oberflächenstrukturellen Beschreibungen - zumindest partiell operationalisiert werden könnten. Es soll daher abschließend untersucht werden, inwieweit der Basisalgorithmus auf die Verar-

beitung leerer Kategorien generalisierbar ist. Dies ermöglicht u.a. einen verfeinerten Vergleich mit den oben zitierten Ansätzen zur Algorithmisierung der bindungstheoretischen Bedingungen von Correa, Ingria und Stallard sowie Giorgi, Pianesi und Satta, in denen z.T. der Anspruch vertreten wird, bestimmte der im folgenden beschriebenen Phänomene adäquat zu operationalisieren. Es wird sich zeigen, daß diese Fälle durch den oben entwickelten Basisalgorithmus ohne grundlegende Modifikationen abgedeckt werden.

Sog. *Spuren*, die infolge von Bewegungstransformationen während des theoretischen Abbildungsprozesses der Tiefenstruktur auf die Oberflächenstruktur entstehen, konstituieren einen der Grundtypen leerer Kategorien.¹⁹ Als Platzhalter ist eine Spur t^{20} mit dem verschobenen Element koindexiert. Daß diese Vorgabe im Rahmen der Verifikation der bindungstheoretischen Restriktionen zu berücksichtigen ist, wird anhand der folgenden Belege deutlich:

- (9a) *[Who_i] does his_i mother love t_i?*
 (9b) ** [Who_i] does he_i love t_i?*
 (9c) *[Which picture of himself_j]_i does John_j like t_i?*

Unter der in der GB-Theorie getroffenen Annahme, daß die als “*Wh*”-Spur bezeichnete leere Kategorie t_i dem Bindungsprinzip B unterliegt, werden die intuitiven Akzeptabilitätsbewertungen der Sätze (9a) und (9b) theoretisch reproduziert: Das Pronomen “*he*”, nicht jedoch das Possessivum “*his*” k-beherrscht die Spur t_i in dessen lokaler Domäne und führt demnach zu einer Verletzung von BP B. Der oben entwickelte Basisalgorithmus deckt diesen Fall unmittelbar ab. “*Wh*”-Spuren werden als spezielle Klasse von Typ-B-Vorkommen behandelt, deren Koindexierung mit dem Originator als spezielle Form einer lokalen, *potentiell interdependierenden* und bereits vollzogenen Einzelentscheidung *durch die oberflächenstrukturelle Beschreibung vorgegeben* ist. Für den Beleg (9b) wird daher (spätestens) im Interdependenztest festgestellt, daß eine Koindexierung von “*Who*” und “*he*” die Bindungsbedingungen verletzen würde. Der Basisalgorithmus löst somit eines der offenen Probleme, das bezüglich der Bindungstheorie-Implementierung von Ingria und Stallard besteht ([InSt89], S. 269).²¹

Komplizierter gestalten sich die Verhältnisse in Fällen, in denen sich bestimmte strukturell

¹⁹Für die näheren Hintergründe dieser theoretischen Konzepte sei auf die einschlägige Literatur verwiesen (z.B. von Stechow und Sternefeld, [StSt88]).

Die vergleichsweise hohe Komplexität der sich anschließenden Diskussion verdeutlicht die Probleme und Grenzen, mit denen die verfolgte Vorgehensweise behaftet ist, mit einer Art von *Steinbruch-Mentalität* eine bestimmte Theoriekomponente (hier: die Bindungstheorie) eines Modells der Sprachkompetenz zu operationalisieren. Unter hinreichend feinkörniger Betrachtung offenbaren sich die Interaktionen mit anderen Modulen der Grammatik, die ja gleichzeitig weiteren Anforderungen gerecht werden müssen; eine unmittelbare Abbildung dieser GB-Vorgaben mag zu einer unter Implementierungsgesichtspunkten unerwünscht komplizierten Beschreibung des im Zentrum des Interesses stehenden Gegenstands - zulässige Index-Distributionen - führen. Die an den theoretischen Vorgaben unmittelbar orientierte Operationalisierung der bindungstheoretischen Restriktionen braucht daher prinzipiell nicht mit der endgültigen informatischen Umsetzung (als reinem Performanzmodell) übereinzustimmen; es wird sich jedoch zeigen, daß der verfolgte Ansatz auch den Anwendungsanforderungen gerecht wird.

²⁰ t wie engl. “*trace*”

²¹Lappin und McCord schlagen ein weiteres Lösungsmodell vor, in dem auf die Einführung leerer Kategorien als explizite Vorkommen verzichtet wird und die Koindexierung durch eine in allen entsprechenden oberflächenstrukturellen Positionen vorkommende gemeinsame, als Diskursreferent zu bezeichnende PROLOG-Variable zum Ausdruck gebracht wird ([LaCo90a, LaCo90b]). Zwar ist dieser Ansatz elegant, es erscheint jedoch mehr als fraglich, ob auf die Explikation von Vorkommensbeschreibungen allgemein verzichtet werden kann, da durch die Restriktionen und Präferenzkriterien in weiten Teilen auf die lokalen Eigenschaften der jeweiligen Konkurrenz Bezug genommen wird. Diesem Sachverhalt wird durch die im Rahmen der vorliegenden Arbeit beschriebene Operationalisierung Rechnung getragen.

zulässige Antezedenskandidaten erst unter Berücksichtigung von Spuren erschließen: Beleg (9c) verdeutlicht dies für ein Reflexivum “*himsel*”, für das der geforderte lokale Binder nur in der durch die Spur t_i bezeichneten originären Position zur Verfügung steht (“*John*”). Einem Vorschlag von Correa zufolge sollte daher die Antezedenssuche auf alle (durch Spuren bezeichnete) leeren Positionen der das Typ-A-Vorkommen enthaltenden Entität ausgedehnt werden ([Corr88], S. 127). Eine entsprechende Erweiterung des Basisalgorithmus ist durch Einführung zusätzlicher leerer Vorkommen der Anapher²² problemlos möglich, wobei lediglich darauf zu achten ist, daß insgesamt nur eine Antezedensentscheidung getroffen wird. Eine Suche ausgehend von den Trace-Positionen hat jedoch zu unterbleiben, wenn die Bindende Kategorie der Anapher bereits *innerhalb* des verschobenen Elements konstituiert wird; diese Feinheit wird im Modell von Correa nicht berücksichtigt.²³

Die GB-Theorie unterscheidet eine weitere Klasse leerer Kategorien - sog. *pro*-Elemente -, die in den oberflächenstrukturellen Beschreibungen bestimmter Sprachen angenommen werden, in denen das Subjekt eines finiten Satzes optional entfallen kann. *pro* dient als formaler Platzhalter des Subjekts und ist als Expletivum mit diesem koindexiert, wenn es an anderer Stelle im Satz nachgetragen wird. Eine dieser sog. *Pro-Drop-Sprachen*, in denen *pro* eine Rolle spielt, ist das Italienische. Die folgenden Belege (10a) und (10b) geben Beispiele der beiden genannten Verwendungsweisen von *pro* in den oberflächenstrukturellen Beschreibungen für diese Sprache (Beispiele zit. aus Giorgi, Pianesi und Satta, [GPS90]):

- (10a) pro_i telefona. (“*Er/Sie/Es* [kontextabhängig zu erschließen] telefoniert.”)
 (10b) pro_i telefona lui $_i$. (“*Er* telefoniert.”)
 (10c) Gianni $_i$ ha detto che pro_j arrivè a [la propria $_i$ madre] $_j$.
 “*Gianni* hat gesagt, daß seine Mutter ankommen wird.”

Wie bereits für die “*Wh*”-Spuren wird auch für das *pro*-Element aus theoretischen Erwägungen die Zuständigkeit von Bindungsprinzip B angenommen (vgl. Chomsky: [Chom86], S. 164), dem in allen obigen Beispielen Genüge getan ist. Umgekehrt sind jedoch Konfigurationen der Gestalt (10b) und (10c) *nicht* als Verletzungen der Bindungsprinzipien B bzw. C der postponierten Subjekte anzusehen - wie bereits von Giorgi, Pianesi und Satta herausgestellt wird, ist diese Unschädlichkeit der lokalen Bindung der Subjekte durch das zugehörige *pro* in der Operationalisierung der Bindungsprinzipverifikation zu berücksichtigen ([GPS90], S. 121). Der Basisalgorithmus ist somit um eine entsprechende Sonderbehandlung des Typ-B-Elements *pro* zu erweitern, was problemlos möglich ist. Bereits durch den Interdependenztest abgedeckt ist hingegen die Blockierung der Koindexierung des Possessivums “*propria*” mit der *pro*-Okkurrenz in Satz (10c), da diese zu einer transitiv vermittelten Verletzung des *i*-über-*i*-Filters führen würde, die in Schritt 3b des Algorithmus erkannt wird. Der Basisalgorithmus wird somit auch den von Giorgi, Pianesi und Satta postulierten Anforderungen an eine adäquate Verarbeitung von *pro*-Elementen gerecht, wobei der gewählte Lösungsweg bei einem insgesamt größeren Leistungsumfang - Bestimmung zulässiger Entscheidungskombinationen - als vorteilhafter, da vergleichsweise einfacher und modularer, erscheint.

²²Die bereits existierenden Spuren leisten in diesem Fall nicht das Gewünschte, denn sie sind Platzhalter der das anaphorische Element enthaltenden Makrostruktur und als solche mit dieser koindexiert.

²³Dieser Fall wäre beispielsweise in folgender Abwandlung des Belegs (9c) gegeben:

[Which of Peter's $_k$ pictures of himself $_k$] $_i$ does John $_j$ like t_i ?

Da die NP “*Peter's pictures of himself*” über ein (zugängliches) lokales Subjekt verfügt, konstituiert sie die Bindende Kategorie des Reflexivums “*himsel*”. Eine Antezedenssuche außerhalb der NP erübrigt sich somit.

Während eine korrekte algorithmische Handhabung von *pro*-Elementen von primär theoretischer Bedeutung zu sein scheint, ist die adäquate Berücksichtigung eines weiteren Typus leerer Kategorien von praktischem Nutzen für das Problem der Anapherninterpretation. In der GB-Theorie werden für bestimmte Formen von Infinitiv-Komplementen sog. *PRO*-Elemente als oberflächenstrukturelle Platzhalter für nichtrealisierte Subjekte angenommen, die vermöge ihrer Indexierung die Interpretationsoptionen etwaiger in der Infinitivstruktur vorkommender Anaphern (mit-)determinieren. In einigen Fällen wird die Indexierung des *PRO*-Elements in Abhängigkeit des Matrixsatz-Verbs determiniert, das das Infinitiv-Komplement bindet (sog. *Kontrollverben*): Für bestimmte Verben (z.B. “*versprechen*”) besteht Koindexierung mit dem Hauptsatz-Subjekt, wohingegen für andere Verben (z.B. “*überreden*”) die Koindexierung mit einem Objekt des Hauptsatzes vorbestimmt ist (*Subjekt-* bzw. *Objektkontrolle*). Die folgenden Belege zeigen auf, daß der jeweilige Index von *PRO* die Interpretationsoptionen eines (infinitivkomplement-lokal zu bindenden) Typ-A-Pronomens determiniert:

- (11a) *Der Friseur_i verspricht dem Kunden_j, PRO_i sich_i zu rasieren.*
- (11b) * *Der Friseur_i verspricht dem Kunden_j, PRO_i sich_j zu rasieren.*
- (11c) * *Der Friseur_i überredet den Kunden_j, PRO_j sich_i zu rasieren.*
- (11d) *Der Friseur_i überredet den Kunden_j, PRO_j sich_j zu rasieren.*

Für Typ-B- und Typ-C-Okkurrenzen ergeben sich in analoger Weise entsprechende Koindexierungs-*Ausschlüsse*.

Entsprechend der GB-Theorie unterliegt die leere Kategorie *PRO* in Abhängigkeit des jeweiligen Verwendungskontexts entweder BP A oder BP B.²⁴ U.a. aufgrund dieser auf den ersten Blick bindungstheoretisch widersprüchlichen Hybridrolle von *PRO* wurden Modellierungsansätze entwickelt, die dessen Bindungsoptionen in einer eigenständigen Theoriekomponente, der sog. *Kontrolltheorie*, beschreiben. Zumindest unter den hier verfolgten anwendungspraktischen Zielsetzungen erscheint ein solcher Schritt jedoch nicht notwendig, da in jeder konkreten *PRO*-Verwendung genau einer der beiden Fälle vorliegt. Entscheidend ist an dieser Stelle, daß der Basisalgorithmus leere Kategorien vom Typ *PRO* ohne weitere Verfeinerungen adäquat verarbeiten kann: Dies wird erreicht, indem *PRO* einerseits als geeigneter Antezedens-Kandidat für Reflexiva aufgefaßt wird (Belege (11)); die entsprechenden Koindexierungs-*Ausschlüsse* für Typ-B- und Typ-C-Vorkommen werden unter Berücksichtigung der vorgegebenen Koindexierungen im Rahmen des Interdependenztestschritts operationalisiert.

Die Diskussion der Operationalisierung bindungstheoretischer Bedingungen am Beispiel des theoretisch wichtigen Beschreibungskonstrukts der leeren Kategorie hat aufgezeigt, daß der Basisalgorithmus über die wünschenswerte Tragweite verfügt: Das spezifizierte Grundverfahren operationalisiert auch die durch leere Kategorien induzierten bindungstheoretischen Bedingungen, wobei allenfalls marginale Verfeinerungen ohne Modifikation des Rahmenalgorithmus vorzunehmen sind. Der Interdependenztestschritt spielt hierbei eine zentrale Rolle, da transitive Entscheidungskollisionen auch bezüglich der für die leeren Kategorien vorgegebenen Koindexierungen berücksichtigt werden. Letztlich gelingt die Operationalisierung der vermöge leerer Kategorien vermittelten bindungstheoretischen Bedingungen natürlich nur in dem Maße, in dem die durch einen Parser zur Verfügung gestellten oberflächenstrukturellen Beschreibungen den

²⁴Letztere Annahme wird bei sog. *arbiträrer Kontrolle* getroffen, in dem für das *PRO*-Element keine satzinterne (lokale) Koindexierung determiniert ist. Dieser Fall erscheint jedoch für die Aufgabenstellung der Anapherninterpretation weniger relevant und soll daher nicht näher diskutiert werden (vgl. von Stechow und Sternefeld, [StSt88]).

entsprechenden Grad der Ausdifferenzierung aufweisen. In Anwendungsszenarien kann jedoch nicht davon ausgegangen werden, daß leere Kategorien uneingeschränkt zur Verfügung stehen - insbesondere die algorithmische Analyse von Spuren ist mit vergleichsweise komplexen Problemen verbunden. Da es weiterer Typus impliziter, die Koindexierungsoptionen determinierender Mitspieler auf einer tieferen, durch semantische bzw. pragmatische Faktoren determinierten Ebene lokalisiert ist (vgl. Abschnitt 7.1.2), der unter der Zielsetzung einer Anwendung im Rahmen der Anaphernresolution prinzipiell gleichzubehandeln wäre, ist zu schließen, daß die operationale Abdeckung auf absehbare Zeit hin partiell bleiben wird. Die Diskussion hat gezeigt, daß der Basisalgorithmus eine adäquate Rahmenstrategie operationalisiert, die diese Fälle prinzipiell abdeckt.

Während sich der Abdeckungsgrad der Verarbeitung leerer Kategorien (nur) in der Qualität der Interpretationsleistung des Basisalgorithmus niederschlägt, stellen die noch ungelösten Probleme der Ambiguität bzw. Unvollständigkeit oberflächenstruktureller Beschreibungen die Anwendbarkeit des Verfahrens a priori in Frage. Bisher setzen die oben beschriebenen Prozesse zur Verifikation der Bindungsbedingungen die Verfügbarkeit eines eindeutigen und vollständigen Syntaxbaums voraus - eine Anforderung, die, wie bereits ausführlich erörtert, durch anwendungstaugliche Parser nicht eingelöst wird. Der Basisalgorithmus bedarf somit einer entsprechenden Verfeinerung. Dies ist der Kerngegenstand des folgenden Kapitels 10.

Kapitel 10

Robuste referentielle Interpretation

Die Entwicklung des Basisalgorithmus erfolgte unter der Prämisse, daß die Verfügbarkeit vollständiger und eindeutiger oberflächenstruktureller Beschreibungen für alle Sätze des zu interpretierenden Texts gewährleistet ist. Daß diese Bedingung im allgemeinen nicht erfüllt sein wird, wurde bereits im Rahmen der Diskussion der Algorithmisierbarkeit des Syntaxanalyseprozesses herausgestellt (vgl. Abschnitt 5.4.1). Folgende Beispiele

- (1a) *Peter beobachtet einen Elefanten mit einem langen Rüssel.*
- (1b) *Peter beobachtet einen Mann mit einem Teleskop.*
- (1c) ** Peter einen Mann mit einem Teleskop.*

illustrieren die grundlegenden Problemklassen:

- Die Disambiguierungskomponente der Syntaxanalyse verfügt nicht über das notwendige Wissen zur Interpretation eines Syntagmas mit aus menschlicher Sicht¹ eindeutigem Bezug (Beleg (1a), korrekte attributive vs. inkorrekte adverbiale Interpretation der Präpositionalphrase (PP));
- der zu analysierende Satz kann aus menschlicher Sicht *ambig* sein (Beleg (1b), attributive vs. adverbiale Interpretation der PP), wobei eine eindeutige Abgrenzung gegenüber dem zuvor beschriebenen Fall nicht in allen Fällen möglich sein wird;
- der zu analysierende Satz ist *ungrammatisch* (Beleg (1c), fehlendes Verb);
- die *linguistischen Ressourcen* des Syntaxanalyseprozesses - implementierte Grammatik, Parsingstrategien usw. - sind *unvollständig* und decken nicht alle Phänomene ab, die in einem zu analysierenden Text vorkommen (vgl. Abschnitt 5.4.1, Zitat S. 68).

Für die Belange einer algorithmischen referentiellen Interpretation, die auf der Basis von Beschreibungen der Oberflächenstruktur bewerkstelligt werden soll, sind die beschriebenen *Ur-*

¹i.S.v. "relativ zum Vorwissen eines durchschnittlichen Lesers"

sachen der erwarteten Unvollständigkeit von Syntaxanalyseergebnissen von sekundärer Bedeutung. Ausschlaggebend sind vielmehr deren *Auswirkungen*, wobei zwei grundlegende Fälle zu unterscheiden sind:

- syntaktisch-strukturelle Ambiguität, d.h. *mehr als eine* oberflächenstrukturelle Beschreibung,
- *keine* den ganzen Satz überspannende oberflächenstrukturelle Beschreibung.

Unter der Sichtweise des bislang stillschweigend unterstellten sequentiellen Prozeßmodells, demzufolge die referentielle Interpretation auf den Ergebnissen der bereits vollzogenen oberflächenstrukturellen Analyse aufbaut, werden sich die beiden Fälle unter einem gemeinsamen Problemtypus subsumieren lassen: Die Anaphernresolution ist auf syntaktischen Beschreibungen zu bewerkstelligen, die in einem noch näher zu definierenden Sinne *fragmentarisch* sind. Im Rahmen dieses Kapitels soll daher zunächst eine Verfeinerung des Basisalgorithmus entwickelt werden, die auf fragmentarischer Oberflächenstruktur arbeitet und damit den Anforderungen einer “*robusten*” und anwendungstauglichen Anapherninterpretation entspricht.

Das Modell der sequentiellen Verarbeitung, das den bislang veröffentlichten Anaphernresolutionsansätzen ausnahmslos zugrunde liegt, erweist sich jedoch unter genauerer Betrachtung als theoretisch inadäquat. Eine Herauslösung der bindungstheoretischen Bedingungen zwecks Anwendung auf eine bereits fertiggestellte Oberflächenstruktur ignoriert die in der GB-Theorie implizit modellierte *wechselseitige Abhängigkeit*, derzufolge die Bindungsbedingungen nicht “*auf*” der Syntax arbeiten, sondern - wie sich zeigen wird - in bestimmten Fällen die oberflächenstrukturellen Konfigurationen *ko-determinieren*. Das sequentielle Prozeßmodell als implizites Resultat der Herauslösung der Bindungsbedingungen aus dem umfassenden System *interagierender* Prinzipien und Parameter der GB-Theorie greift aus theoretischer Sicht zu kurz: Erneut ist es die bereits in anderem Zusammenhang als problematisch identifizierte, der isolierten Bindungstheorie-Implementierung zugrundeliegende *Steinbruch-Mentalität* (vgl. Abschnitt 9.5.2), die sich in konzeptuellen Schwierigkeiten des Operationalisierungsansatzes niederschlägt. Es soll daher ein weiterer algorithmischer Ansatz der “*robusten*” referentiellen Interpretation entwickelt werden, durch den eine adäquate *Verzahnung* des Anaphernresolutionsprozesses und der oberflächenstrukturellen Analyse bewerkstelligt wird.

Der Begriff “*robust*” wurde bisher in Anführungszeichen gesetzt, da er weiterer Präzisierungen bedarf. Gegen welche *Arten von Störungen* ist ein solchermaßen attribuiertes Verfahren resistent? Wie *weit* (in noch zu definierendem Sinne) reicht diese Resistenz? (gradueller Abstufung) Es wird daher mit einer näheren Eingrenzung dieses bislang intuitiv gebrauchten Eigenschaftsworts begonnen. Auf der Grundlage dieser Festlegungen wird sich ein gradueller Robustheits-Unterschied der beiden zu entwickelnden Anaphernresolutions-Algorithmen identifizieren lassen.²

10.1 Robustheit

Unter allgemeinem Bezug auf Informationsverarbeitungsprozesse kennzeichnet der Begriff der *Robustheit* die Eigenschaft, die jeweilige Interpretationsaufgabe auch auf solchen Eingaben zu

²Die Kerninhalte dieses Kapitels basieren auf der Vorveröffentlichung Stuckardt: [Stu97b].

lösen, die in bestimmter Hinsicht *defekt* sind.

10.1.1 Robustheit von Sprachverarbeitungsprozessen

Menzel schlägt eine verfeinerte Definition des Robustheitsbegriffs für Systeme der Interpretation natürlicher Sprache vor, die sich im Hinblick auf die weiteren Ausführungen dieses Kapitels als adäquat erweisen wird. Der Charakterisierung Menzels zufolge heißt ein Sprachverarbeitungsprozeß *robust*, wenn er in der Lage ist, auf (die Eingabe konstituierenden) linguistischen Beschreibungen zu arbeiten, die potentiell defekt sind ([Menz95]). Auf welche Ursache die Defekte im Einzelfall zurückzuführen sind - Mängel der Verarbeitungsressourcen oder der zu interpretierenden Äußerungen - spielt hierbei keine Rolle.

Ausgehend von dieser Basischarakterisierung identifiziert Menzel *drei Merkmale robuster Sprachinterpretation*, die im Einzelfall mehr oder weniger stark ausgeprägt sein können und insofern eine graduelle Abstufung robuster Verarbeitung induzieren. Die erste, obligatorische und daher i.d.R. nicht explizit genannte Eigenschaft robuster Sprachverarbeitung ist die der *Monotonie*: Je “*geringer*” der Defekt, den die Eingabe aufweist, desto “*besser*” fällt die Qualität der erbrachten Interpretationsleistung aus. Systeme mit feinabgestufter Monotonieeigenschaft sind diesbezüglich als *robuster* zu bezeichnen als Systeme, die ausschließlich eine triviale Form von Monotonie aufweisen.³ Ebenfalls mehr oder weniger implizit mit der Grunddefinition der Robustheit verwoben ist die Eigenschaft der *Autonomie*, derzufolge ein vollständiges Scheitern auf einer Ebene der Analyse möglichst nicht ein vollständiges Scheitern einer anderen, die entsprechende Beschreibung als Eingabe verarbeitende Analyseebene, oder gar den Abbruch des gesamten Analyseprozesses zur Folge haben sollte - mit anderen Worten: Der Prozeß sollte weder ergebnislos enden noch (unkontrolliert) zusammenbrechen. Eine dritte, fakultative Facette robuster Sprachverarbeitung betrifft das Prozeßmodell, das der Zusammenarbeit der unterschiedlichen Teilprozesse der Sprachinterpretation zugrundeliegt: Im Idealfall bestehen multiple Beziehungen der prozessualen *Interaktion*, die es ermöglichen, Defizite auf einer Ebene der Analyse unter Rekurs auf Information zu kompensieren, die auf anderen Verarbeitungsebenen gewonnen wird. Die Ausprägung der Interaktionsbeziehungen zwischen den Teilprozessen verkörpert ein weiteres Kriterium, mithilfe dessen eine graduelle Abstufung der Robustheit von Sprachverarbeitungsprozessen vorgenommen werden kann.

10.1.2 Robustheit und referentielle Interpretation

Die Abgrenzung des Robustheitsbegriffs von Menzel läßt sich unmittelbar auf das Sprachverarbeitungsproblem der referentiellen Interpretation anwenden. Die einfließenden Teilprozesse sind (u.a.) Wortsegmentierung, morphologische Analyse, syntaktische Analyse und - als übergeordneter, im Zentrum des Interesses stehender Prozeß - Anapherninterpretation, der den wichtigen Teilprozess der Bindungsbedingungs-Verifikation umfaßt. Bezüglich des beschriebenen Analyseszenarios besteht somit das Kernproblem der robusten Verarbeitung in der Abhängigkeit potentiell *fragmentarischer* oberflächenstruktureller Beschreibungen, auf deren Basis der Kernprozeß

³Mit “*trivial monoton*” sollen Systeme bezeichnet werden, die die durch den defekten Input gegebene partielle Information - unbeschadet der jeweiligen Schwere des Defekts - vollständig ignorieren oder stets dasselbe Default-Ergebnis - etwa eine “*leere*” Interpretation - liefern.

der referentiellen Interpretation abläuft. Ein erster Ansatz zur robusten Anapherninterpretation ergibt sich durch eine triviale Modifikation des Basisalgorithmus:⁴

Falls die syntaktische Analyse eine vollständige und unambige Beschreibung der Oberflächenstruktur der zu interpretierenden Äußerung liefert, so verfare, wie im Basisalgorithmus spezifiziert. Andernfalls - bei fragmentarischer Syntax - verzichte auf die Anwendung von Strategien, die auf der Verfügbarkeit syntaktischer Beschreibungen aufbauen.

Dieses Analysemodell ist zwar *robust* im Sinne der obigen Definition - es weist die Grundeigenschaften der Autonomie und der (trivialen) Monotonie auf -, jedoch keineswegs zufriedenstellend, da die Voraussetzungen, unter denen das Verfahren die für die Qualität der Interpretationsergebnisse so bedeutsame Evidenz der auf syntaktischen Beschreibungen arbeitenden Bindungsbedingungen heranzieht, in der praktischen Anwendung nur relativ selten erfüllt sein dürften. Das Ergebnis einer Syntaxanalyse in nichtrestringierten Anwendungsszenarien wird für Sätze mit nichttrivialer Struktur *in aller Regel* ambig und damit fragmentarisch ausfallen. Mit anderen Worten: Da nur eine triviale Form der Monotonie realisiert ist, erweist sich der *Grad* der Robustheit unter den beschriebenen Anwendungsbedingungen als unzureichend.

Die beschriebene Triviallösung bedarf also einer Verfeinerung, auf deren Grundlage auch für fragmentarischen oberflächenstrukturellen Input ein - in bezug auf den Grad des Defekts - *relativ gutes* Ergebnis erzielt wird. Die in Form von fragmentarischen syntaktischen Beschreibungen vorliegende partielle konfigurationale Information sollte *möglichst weitgehend* - eventuell auch unter Anwendung heuristischer Regeln - für die Verifikation der Bindungsbedingungen nutzbar gemacht werden. Diese Grundidee liegt der ersteren der nachfolgend beschriebenen robusten Versionen des Basisalgorithmus zugrunde: Im Vergleich zur Triviallösung ist diese als graduell robuster zu bezeichnen, da auch auf fragmentarischer Oberflächenstruktur eine akzeptable Ergebnisqualität erzielt wird.

Diesem ersten akzeptablen, bezüglich fragmentarischem syntaktischen Input hinreichend robusten Verfahren der referentiellen Interpretation liegt ein sequentielles Verarbeitungsmodell zugrunde, in dem die Ergebnisse der syntaktischen Analyse in den Prozeß der Anaphernresolution einfließen, jedoch nicht umgekehrt. Dieser bereits als theoretisch unbefriedigend identifizierte Sachverhalt spiegelt sich im Rahmen der Menzelschen Abgrenzung der Robustheitseigenschaften in Form der *fehlenden Interaktion* der beiden Verarbeitungsschritte wieder: Obwohl in der GB-Theorie implizit dargelegt und auch prinzipiell möglich - vgl. u. -, wird eine etwaige Evidenz der referentiellen Interpretation für die oberflächenstrukturelle Analyse nicht nutzbar gemacht. An diesem Punkt setzt das zweite im folgenden entwickelte Verfahren an, das eine bidirektionale, unter theoretischen Gesichtspunkten als *vollständig* zu bezeichnende Interaktion von referentieller Analyse und Parsing realisiert. Da auf diese Weise allen drei Robustheitseigenschaften - Autonomie, Monotonie *und* Interaktion - in optimaler Weise Rechnung getragen wird, kann dieser Ansatz auch als *ultimativ robust* bezeichnet werden.

⁴In der anschließenden Diskussion wird - marginal vereinfachend - außer Betracht gelassen, daß auch die Identifikation von Vorkommen durch die syntaktische Information ko-determiniert wird und somit ein eigenständiges Problem der robusten Verarbeitung konstituiert. Dies erscheint jedoch insofern gerechtfertigt, als bereits auf der Grundlage eines trivial robusten Verfahrens eine unter praktischen Gesichtspunkten völlig zufriedenstellende Lösung erzielbar ist.

10.1.3 Zwei Modelle robuster referentieller Interpretation

Als Basis für die Einordnung und Abgrenzung der im folgenden entwickelten Algorithmen sollen abschließend zwei grundlegende Verfahrensklassen einer anwendungstauglichen, robusten referentiellen Interpretation definiert werden. Das erste Modell basiert auf der Idee, auf eine uneingeschränkte syntaktische Analyse a priori zu verzichten:

Definition 10.1 (Modell der seichten Voranalyse) *An Stelle einer uneingeschränkten, potentiell fragmentarischen syntaktischen Analyse tritt eine seichte Voranalyse, die niemals ein defektes Ergebnis liefert. Aufbauend auf diesem Ergebnis wird die oberflächenstrukturelle Beschreibung in Teilen heuristisch rekonstruiert. Die auf syntaktisch-konfiguratoraler Evidenz basierenden Strategien der Anaphernresolution werden dahingehend modifiziert, daß sie auf den Ergebnissen dieser heuristischen Rekonstruktion arbeiten.*⁵

Ein Beispiel eines Ansatzes, dem das Modell der seichten Voranalyse zugrundeliegt, ist das in Abschnitt 8.7 beschriebene Verfahren der parserfreien Anaphernresolution von Kennedy und Boguraev, das auf den Ergebnissen eines Part-of-Speech-Taggers basiert.

Das Modell der seichten Voranalyse läßt insofern Spielraum für Verbesserungen, als auf das zwar in der Regel fragmentarische, jedoch im Vergleich mit der durch eine *heuristische* Rekonstruktion von Teilen der Oberflächenstruktur gewonnenen Information *potentiell verlässlichere* Analyseergebnis einer syntaktischen Analyse herkömmlichen Zuschnitts ohne zwingenden Grund a priori verzichtet wird. Die im Rahmen dieser Arbeit entwickelten Algorithmen setzen an diesem Punkt an. Es liegt ein Analysemodell zugrunde, durch das die verfügbare syntaktisch-konfigurational Evidenz so weit wie möglich genutzt wird:

Definition 10.2 (Modell der potentiell fragmentarischen Voranalyse) *Die auf syntaktisch-konfiguratoraler Information basierenden Strategien der Anaphernresolution werden dahingehend modifiziert, daß sie auf den potentiell fragmentarischen Voranalyseergebnissen eines uneingeschränkten und nichtheuristischen oberflächenstrukturellen Parsings arbeiten.*

10.2 Fragmentarische syntaktische Beschreibungen

Anhand der eingangs gegebenen Beispiele wurde der Begriff der fragmentarischen oberflächenstrukturellen Beschreibungen geprägt. Die Verwendung dieses Terminus geschah bislang auf intuitiver Basis. Es soll daher zunächst präziser ausgeführt werden, welche Spielarten fragmentarischer syntaktischer Beschreibungen existieren. Diese Vorbetrachtungen schaffen die geeigneten Voraussetzungen dafür, näher zu untersuchen, in welcher Form die Strategien einer algorithmischen referentiellen Interpretation durch die Defekte des oberflächenstrukturellen Inputs tangiert sein können.

⁵Strenggenommen handelt es sich hierbei um keine im ursprünglichen Sinne *robuste* referentielle Interpretation, da von vornherein keine syntaktischen Beschreibungen herangezogen werden, die potentiell defekt sind. Wichtig ist jedoch alleine die Anwendungstauglichkeit.

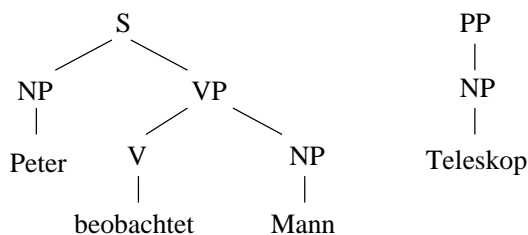


Abbildung 10.1: Fragmente der oberflächenstrukturellen Beschreibung für Beleg (2a)

10.2.1 Ursachen fragmentarischer Syntax

Die wesentlichen Fälle *struktureller Ambiguität*, die im Rahmen der algorithmischen syntaktischen Analyse zum Tragen kommen, entstehen bezüglich der Interpretation von Präpositionalphrasen, Relativsätzen und Adverbialsätzen sowie ferner bezüglich der Unsicherheit, die in einigen Fällen betreffend die Zuweisung der syntaktischen Funktionen *Subjekt* und *transitives Objekt* besteht.⁶

- (2a) *Peter beobachtet einen Mann mit einem Teleskop.*
 (2b) *Der Mann hat den Präsidenten besucht, der ihn überzeugte.*
 (2c) *Der Mann ist gegangen, bevor die Sitzung vertagt wurde, weil es schon spät war.*
 (2d) *Peter Behrens besuchte Walter Gropius zuerst.*

Die ersten drei genannten Fälle schlagen sich in einer Mehrdeutigkeit der strukturellen Anknüpfung des jeweiligen Syntagmas nieder - beispielsweise illustriert Beleg (2c) die mögliche Ambiguität der Interpretation eines Adverbialsatzes, für den ohne Weiteres nicht klar ist, ob er dem Hauptsatz oder dem ersten Nebensatz unterzuordnen ist. Beispiel (2d) zeigt einen Fall, in dem aufgrund fehlender morphologischer Merkmale offen ist, welche der unterstrichenen NP die Rolle des Subjekts füllt und welche das transitive Objekt darstellt.⁷

Abbildung 10.1 illustriert, wie sich die strukturelle Ambiguität von Beleg (2a) in einer fragmentarischen oberflächenstrukturellen Beschreibung niederschlägt, die zwei Teilbäume - informatisch gesprochen: Zusammenhangskomponenten - umfaßt: Ersteres Fragment enthält den Wurzelknoten des gesuchten vollständigen Syntaxbaums, letzteres Fragment entspricht der PP, deren Interpretation nicht entschieden werden konnte. Je nach verwendetem Parser steht i.a. auch die Information über die für das PP-Fragment bestehenden Interpretationsoptionen - Attribut der NP "Mann" oder adverbial - zur Verfügung (vgl. auch Abbildung 7.4, S. 120). Für die übrigen Beispiele ergeben sich die oberflächenstrukturellen Fragmente auf vergleichbare Weise. In den Fällen (2b) und (2c) sind die nichtangeknüpften Fragmente satzförmig (syntaktische Kategorie S). Für (2d) ergeben sich Fragmente der Kategorie NP, für die bekannt ist, daß sich deren Interpretationsoptionen wechselseitig bedingen.⁸ Die oberflächenstrukturellen Fragmente, die durch

⁶Ob die jeweilige Mehrdeutigkeit nur auf einen Mangel einer Komponente der algorithmischen Analyse zurückzuführen ist oder auch im Rahmen einer intellektuellen Analyse in Erscheinung tritt, spielt im folgenden keine Rolle.

⁷Im hier nicht betrachteten Fall der gesprochenen Sprache liefert i.a. die Intonation die benötigte Information.

⁸D.h.: Wird eine der beiden NP die Subjektkontrolle zugewiesen, so wird die andere NP das transitive Objekt und umgekehrt.

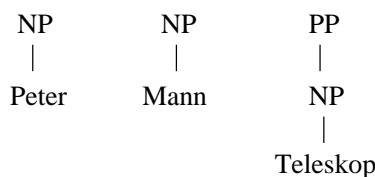


Abbildung 10.2: Fragmente der oberflächenstrukturellen Beschreibung für Beleg (3)

die einschlägigen Fälle struktureller Ambiguität induziert werden, sind somit Syntagmen der Kategorien PP, S und NP.

Weitere Typen von Fragmenten, die auch Syntagmen anderer Kategorien entsprechen können, entstehen in Fällen, in denen der Syntaxanalyseprozeß - gleich welchen Grunds - abbricht, *ohne* eine oder mehrere Lesarten gefunden zu haben. Auch in diesen Fällen steht in der Regel partielle Information zur Verfügung, die einen Zwischenzustand des Parsings repräsentiert, deren weitere Rekombination jedoch gescheitert ist. Das Problem, die entsprechenden syntaktischen Fragmente aus den Datenstrukturen des Parsers (i.a. der sog. *Parse-Tabelle*) in möglichst umfassender Weise zu extrahieren, ist in nicht allen Fällen trivial. Jedoch bewerkstelligen robuste, massendatentaugliche Parser diese Aufgabe i.d.R. zumindest auf heuristische Weise; daher erscheint es gerechtfertigt, davon auszugehen, daß entsprechende fragmentarische Beschreibungen der Oberflächenstruktur auch in diesem Fall zur Verfügung stehen. Für den eingangs besprochenen ungrammatischen Beleg

(3) * *Peter einen Mann mit einem Teleskop.*

könnte die fragmentarische Ausgabe der Syntaxanalyse aussehen wie in Abbildung 10.2 gezeigt: Die Syntagmen der Mitspieler werden gefunden, doch deren Rekombination scheitert, weil das verbindende Vollverb fehlt.⁹

10.2.2 Fragmentarische Syntax und referentielle Interpretation

Es soll nun näher untersucht werden, welchen Einschränkungen ein Algorithmus zur Anaphernresolution unterliegt, der auf fragmentarischen oberflächenstrukturellen Beschreibungen arbeitet. Anhand der Übersicht der in die Strategien des Basisalgorithmus einfließenden Wissensquellen (Abbildung 7.7, S. 146) wird deutlich, daß die Verfügbarkeit syntaktischen Wissens in den meisten Fällen vorausgesetzt wird. Betroffen sind die Restriktionen SYNKONFIG, SEMROLLEN und (partiell) MORKONGRU sowie die Präferenzheuristiken TOPIKALIS, SUBJEKT, SYNROLLE, SYNPARALL und - eingeschränkt - SEMPARALL.

Eine genauere Betrachtung zeigt jedoch auf, daß wesentliche Unterschiede bezüglich des Grads

Vergleichbare Interdependenzen existieren für die übrigen Fälle, in besonderem Maße aber für multiple PP-Ambiguitäten: Projektivitätsbedingungen des Parsing-Prozesses schließen bestimmte Kombinationen von Interpretationsentscheidungen aus. Auch diese Information wird von Syntaxanalytoren i.d.R. akkumuliert (vgl. Böttcher, [Bött96]) und könnte daher prinzipiell Bestandteil des Outputs des Parsingprozesses sein.

⁹U.U. wählt der Parser hier jedoch bereits die attributive Interpretation der PP, da kein Anhaltspunkt für eine mögliche Interpretationsalternative gegeben ist. Entsprechend würde sich die Zahl der Fragmente um eins reduzieren.

der Beeinträchtigung bestehen. Die Präferenzstrategien nehmen ohne Ausnahme Bezug auf die syntaktische Funktion bzw. auf die damit in mittelbarem Zusammenhang stehende semantische Rolle. Mit Blick auf die in Abschnitt 10.2.1 identifizierten Typen syntaktischer Fragmentierung wird ersichtlich, daß die Verfügbarkeit dieser Information durch die in ambigen Parsing-Ergebnissen am häufigsten anzutreffenden Fälle struktureller Ambiguität - PP-, Relativsatz- und Adverbialsatz-Anknüpfung - nicht bzw. allenfalls (im Falle nichtzugeordneter PP-Ergänzungen) marginal beeinträchtigt wird, da sie im Rahmen der nicht tangierten Satzfragmente lokal zur Verfügung steht. Für Instanzen der Rollen-Ambiguität (Beleg (2d)) sowie der auf unspezifische Weise induzierten Fragmentierung (Beleg (3)) ist dies nicht gewährleistet, da auch Fragmente entstehen können, die Partizipanten entsprechen, denen somit keine syntaktische Rolle zugeordnet wird. Jedoch ist es möglich, in solchen Fällen auf an der sprachlichen Oberfläche verfügbare Anordnungsinformation zurückzugreifen, um die syntaktische Rolleninformation auf heuristischer Basis zu rekonstruieren. Die in Abschnitt 7.2.2 beschriebene hierarchische Ordnung der syntaktischen Funktionen findet in unmarkierten Belegen ihre Entsprechung in der Oberflächenanordnung - das Subjekt steht i.d.R. am weitesten links (vgl. Primus, [Prim89]). Da es sich ohnehin bereits um *Heuristiken* der referentiellen Interpretation handelt, in die diese Information einfließt, ist die zu erwartende Verschlechterung der Interpretationsleistung gering.

Für die stringenten Strategien SEMROLLEN und MORKONGRU gilt Entsprechendes, da auch hier nur eine Abhängigkeit von der i.d.R. fragment-lokal verfügbaren syntaktischen Funktion besteht. Anders verhält es sich hingegen für die Restriktionsklasse SYNKONFIG, deren Anwendung nicht alleine auf lokaler Information wie der syntaktischen Rolle, sondern auf der syntaktischen Konfiguration insgesamt beruht. Ob eine Fragmentierung zu einer Beeinträchtigung der bindungstheoretischen Evidenz führt, hängt vom jeweiligen Fall ab: Gehören die Vorkommen, deren Koindexierung auf konfigurationale Verträglichkeit getestet werden soll, unterschiedlichen Fragmenten an, so ist diese Frage auf dem gegenwärtigen Stand der Ausführungen uneingeschränkt zu bejahen. Da die im Rahmen der Verifikation der Bindungsbedingungen betrachteten Vorkommenspaare in der oberflächenstrukturellen Beschreibung beliebig weit auseinanderliegen können, stellen Fälle, in denen eine syntaktische Fragmentierung die Anwendung der Strategie SYNKONFIG tangiert, eher die Regel denn die Ausnahme dar.

Es sollen nun einige Belege diskutiert werden, die die Beeinträchtigung der Verifikation der bindungstheoretischen Bedingungen durch die einschlägigen Typen struktureller Ambiguität illustrieren. In Abschnitt 7.1.2 wurde dies bereits für ein Fallbeispiel der *PP-Interpretation* geleistet:

(4) *Behrens beobachtet den Eigentümer des Feldstechers mit ihm.*

Je nachdem, ob die PP “*mit ihm*” adverbial oder attributiv (zu “*Feldstecher*”) interpretiert wird, ist eine Koindexierung der Vorkommen “*Feldstecher*” und “*ihm*” entweder möglich bzw. - aufgrund einer Verletzung des i-über-i-Filters - unmöglich (vgl. Abbildung 7.4, S. 120).

Vergleichbare Beispiele lassen sich für die übrigen zwei Typen von Anknüpfungs-Mehrdeutigkeit angeben. Auch die Wahl der Bezugs-NP eines *Relativsatzes* wirkt sich auf die potentiellen, qua Koindexierung herstellbaren Bindungsbeziehungen aus.

(5a) *Der Mann hat den Präsidenten besucht, der ihn überzeugte.*

(5b) *Der Mann hat den Präsidenten besucht, der das Staatsoberhaupt überzeugte.*

Wird die Lesart als Attribut zu “*Präsidenten*” gewählt, so ist eine Koindexierung von “*Präsidenten*” und “*ihn*” (Beleg (5a)) bzw. “*Staatsoberhaupt*” (Beleg (5b)) konfigural ausgeschlossen:

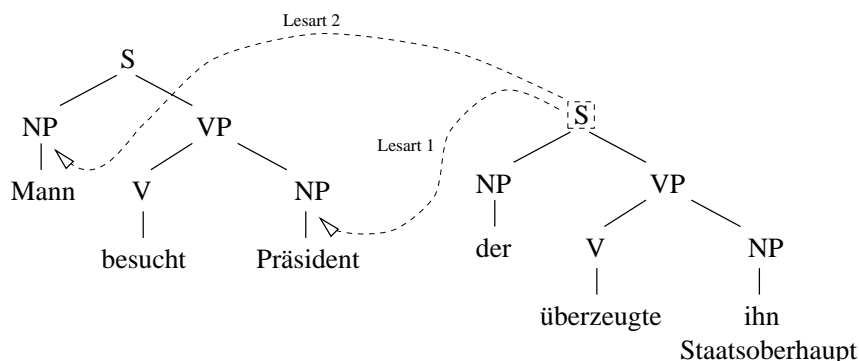


Abbildung 10.3: Syntaktische Fragmente für die Belege (5a) bzw. (5b)

unter der Annahme, daß das Relativpronomen “*der*” und dessen Bezugs-NP koindexiert sind¹⁰, ergibt sich eine transitive Verletzung der Bindungsprinzipien B bzw. C (vgl. Abbildung 10.3, Lesart 1).¹¹ Hingegen in der zweiten Lesart - Relativsatz als Attribut zu “*Mann*” - sind die entsprechenden Antezedens-Entscheidungen konfigurationsal zulässig, da das Relativpronomen mit einer anderen Bezugs-NP koindexiert ist und somit eine unzulässige lokale (transitiv vermittelte) Bindungs-Beziehung nicht entsteht (vgl. Abbildung 10.3, Lesart 2).

Zu einer Beeinträchtigung der bindungstheoretischen Evidenz kann es auch in Fällen struktureller Ambiguität kommen, die im Rahmen der Interpretation von *Adverbialsätzen* entstehen. In folgendem Beispiel hängen die Koindexierungs-Optionen des Vorkommens “*Staatsoberhaupt*” davon ab, ob der “*weil*”-Satz als Adverbial des Hauptsatzes oder des “*bevor*”-Satzes interpretiert wird.

(6) *Kohl ging, bevor der Präsident kam, weil das Staatsoberhaupt sich verspätete.*

Abbildung 10.4 veranschaulicht die konfigurationsalen Gegebenheiten der bestehenden Interpretations-Optionen. Unter der in der Theorie getroffenen Annahme, daß ein Adverbialsatz in der Oberflächenstruktur unmittelbar dem S-Knoten des Bezugssatzes subordiniert ist¹², besteht in Lesart 1 (“*weil*”-Satz als Adverbial des eingebetteten Nebensatzes) eine K-Herrschafts-Beziehung zwischen den NPs zu “*Präsident*” (NP₁) und “*Staatsoberhaupt*” (NP₂); eine Koindexierung der entsprechenden Vorkommen verstieße daher gegen Bindungsprinzip C. In Lesart 2 hingegen (“*weil*”-Satz als Adverbial des Hauptsatzes) ist diese Koindexierung möglich, da die maßgebliche K-Herrschafts-Relation nicht gegeben ist. Auch die bindungstheoretischen Vorhersagen in bezug auf die Interpretation von *Pronomen* werden durch strukturelle Ambiguitäten von Adverbialsätzen potentiell beeinträchtigt. In folgender Variante des soeben diskutierten Beispiels

(7) *Kohl ging, bevor er kam, weil das Staatsoberhaupt sich verspätete.*

¹⁰Aus der GB-Theorie scheint nicht eindeutig hervorzugehen, ob diese das Relativpronomen betreffende Koindexierungs-Annahme mit den theoretischen Vorgaben übereinstimmt - immerhin scheint sie im Widerspruch zur Vorhersage des *i-über-i-Filters* zu stehen. Für die im Rahmen dieser Arbeit verfolgten Anwendungszielsetzungen erweist sich diese Festlegung jedoch als adäquat.

¹¹Alternativ könnte dieser Koindexierungsausschluß über den *i-über-i-Filter* vorhergesagt werden.

¹²vgl. Grewendorf: [Grew88], S. 129 f.

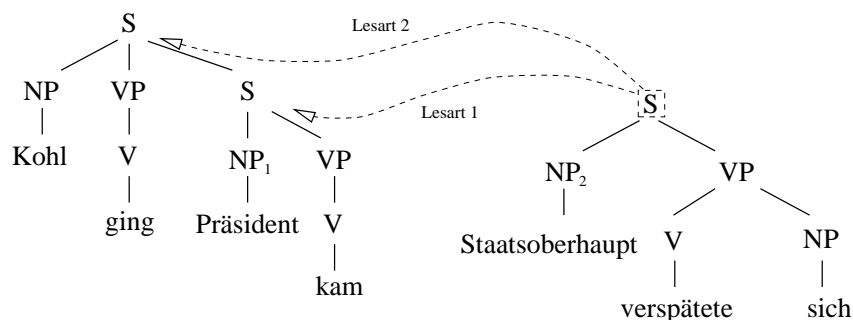


Abbildung 10.4: Syntaktische Fragmente für Beleg (6)

hängen die Antezedensalternativen des pronominalen Vorkommens “*er*” in analoger Weise von der für den “*weil*”-Satz getroffenen Interpretationsentscheidung ab.¹³

Für die übrigen Fallklassen fragmentarischer Syntax - Rollenunsicherheit bzw. nichtstereotype Formen der Fragmentbildung - ließen sich analoge Belege angeben. Die syntaktisch-konfigurative Information, auf der die Strategie SYNKONFIG aufbaut, steht auch hier nicht immer in dem benötigten Umfang zur Verfügung, um bestimmte bindungstheoretische Vorhersagen definitiv treffen zu können. Im Unterschied zu den übrigen auf oberflächenstruktureller Information aufbauenden Strategien, in denen lediglich auf die - ggf. heuristisch erschließbare - syntaktische Funktion Bezug genommen wird, scheint eine einfache Lösung des hier vorliegenden Problems nicht möglich, da auch auf Relationen der syntaktischen Konfiguration Bezug genommen wird, die nicht lokal eingrenzbar sind. Folglich sind weitergehende Überlegungen anzustellen, um die bereits als zentral für den Prozeß der algorithmischen referentiellen Interpretation identifizierte Strategie SYNKONFIG in adäquater Weise auf fragmentarische syntaktische Beschreibungen zu verallgemeinern.

10.3 Verifikation der bindungstheoretischen Bedingungen auf fragmentarischer Syntax

Es soll nun eine erste Verfeinerung des Basisverfahrens entwickelt werden, durch die auch im fragmentarischen Fall eine gute, möglichst nichtheuristische Operationalisierung der bindungstheoretischen Bedingungen erzielt wird. Ergebnis dieser Ausführungen wird ein Algorithmus sein, dem das Modell der potentiell fragmentarischen Voranalyse (Definition 10.2) zugrundeliegt und der im Vergleich zu der in Abschnitt 10.1.2 skizzierten Trivalllösung robuster referentieller Interpretation ein erhöhtes Maß an Robustheit aufweist.

¹³Hierbei braucht keineswegs eine Instanz kataphorischer Bezugnahme vorzuliegen, denn die Koindexierung von “*er*” und “*Präsident*” kann transitiv - etwa vermöge einer identischen Antezedenswahl für diese beiden Vorkommen - zustande kommen.

10.3.1 Grundlegende Beobachtungen

Ausgangspunkt der Entwicklung einer gegenüber fragmentarischer Syntax robusten Operationalisierung der Strategie SYNKONFIG ist die folgende elementare Beobachtung:

Kommen Anapher und Antezedenskandidat in derselben Zusammenhangskomponente der fragmentarischen oberflächenstrukturellen Beschreibung vor, so geht keine (unmittelbare) bindungstheoretische Evidenz verloren.

In diesem Fall ist die Verifikation der bindungstheoretischen Bedingungen auf nichtheuristische Weise möglich, da die syntaktisch-konfigurative Information, auf der in den Bindungsprinzipien A, B und C sowie im i-über-i-Filter Bezug genommen wird, vollständig zur Verfügung steht.¹⁴ Gehören die beiden Vorkommen hingegen unterschiedlichen Fragmenten an, so kann entscheidungsrelevante oberflächenstrukturelle Information fehlen.

Aufbauend auf diesen Überlegungen ist eine erste nichttriviale Lösung des Problems der Verifikation der Bindungsbedingungen auf potentiell fragmentarischen syntaktischen Beschreibungen möglich. Um für ein gegebenes Paar von Okkurrenzen entscheiden zu können, welcher der beiden Fälle vorliegt, werden die Vorkommensbeschreibungen mit einem Attribut ausgestattet, unter dem ein Identifikator des syntaktischen Fragments verwaltet wird, dem die jeweilige Okkurrenz zuzuordnen ist; dies kann ein Zeiger auf den eindeutig bestimmten Wurzelknoten des baumförmigen Fragments oder eine natürliche Zahl sein. Unter Bezugnahme auf diese Information kann die Strategie SYNKONFIG folgendermaßen auf fragmentarische Syntax adaptiert werden:

Kommen Anapher α und Antezedenskandidat γ in demselben oberflächenstrukturellen Fragment vor, so verifiziere die bindungstheoretischen Bedingungen wie im Falle von nichtfragmentarischer Syntax. Kommen Anapher α und Antezedenskandidat γ in unterschiedlichen Fragmenten vor, so betrachte die Koindexierung von α und γ als bindungstheoretisch zulässig, jedoch reduziere die Plausibilitätsbewertung der möglichen Entscheidung $\alpha \rightarrow \gamma$.

Falls Anapher und Kandidat unterschiedlichen Fragmenten angehören, wird somit auf die Anwendung der Strategie SYNKONFIG verzichtet; in diesem - erkennbaren - Fall erfolgt die Zusage der bindungstheoretischen Zulässigkeit der Koindexierung auf *heuristischer* Grundlage. Da der Ansatz auch im Falle fragmentarischer Syntax immer dann die Bindungsbedingungen verifiziert, wenn die involvierten Vorkommen demselben Fragment angehören, leistet er bereits mehr und ist graduell robuster als die in Abschnitt 10.1.2 beschriebene Triviallösung.

10.3.2 Nichtheuristische Entscheidungsregeln für den fragmentarischen Fall

Eine genauere Untersuchung zeigt jedoch auf, daß auch im nachteiligen Fall, in dem Anapher und Antezedenskandidat in unterschiedlichen Fragmenten lokalisiert sind, zumindest teilweise mehr

¹⁴Diese Aussage gilt ausschließlich für den *unmittelbaren* konfiguralen Vergleich der beteiligten Okkurrenzen. Im Falle von Entscheidungsinterdependenz, die durch Vorkommen außerhalb des lokalen Fragments von Anapher und Antezedens vermittelt wird, kann konfigurale Evidenz verlorengehen.

möglich ist, als die zuvor beschriebenen Verfahren leisten. Dies läßt sich anhand des folgenden Belegs erläutern, für den angenommen werden soll, daß die beiden “*er*”-Okkurrenzen aufgrund einer strukturellen Ambiguität des “*weil*”-Satzes in unterschiedlichen syntaktischen Fragmenten lokalisiert sind (für die konfiguralen Gegebenheiten vgl. o., Abbildung 10.4).

(8) *Der Präsident ging unmittelbar, nachdem er sprach, weil er sehr müde war.*

Gesetzt den Fall, die bindungstheoretische Zulässigkeit einer Koindexierung dieser beiden Pronomen sei zu testen. Der zuvor skizzierte erste Ansatz verzichtet auf eine weitergehende Überprüfung; die Koindexierung wird auf *heuristischer* Grundlage akzeptiert. Dabei ist für das beschriebene Beispiel eine *definitive* Entscheidung möglich. Da die syntaktischen Fragmente, in denen die Pronomen lokalisiert sind, deren Bindende Kategorien - die S-Knoten des jeweiligen Adverbialsatzes - enthalten, ist ungeachtet der Auflösung der bestehenden strukturellen Mehrdeutigkeit gewährleistet, daß keines der Pronomen das jeweils andere in dessen lokaler Bindungsdomäne k-beherrschen wird. Egal, welche Zuordnung des Adverbialsatzes letztlich die Korrekte ist - eine Verletzung des für beide Vorkommen anwendbaren Bindungsprinzips B wird in keinem Fall entstehen.

Die Vorhersagen des heuristischen Verfahrens stimmen jedoch nicht in allen Fällen mit den realen Gegebenheiten überein. Im Falle einer Typ-A-Anapher, deren lokale Bindungsdomäne vollständig bekannt ist, können fragment-externe Antezedenskandidaten definitiv ausgeschlossen werden. Unter denselben Ambiguitäts-Annahmen wie zuvor trifft also das heuristische Entscheidungsmodell für den folgenden Beleg die falsche Entscheidung - Zulässigkeit der Koindexierung von “*Kohl*” und “*sich*” im Rahmen der Antezedenssuche für das Reflexivum:¹⁵

(9) *Kohl ging, bevor der Präsident sprach, weil er sich verspätete.*

In den beiden beschriebenen Konfigurationen kann somit trotz syntaktischer Fragmentierung eine nichtheuristische, definitive Verifikation der Bindungsbedingungen erfolgen.

Es liegt nun auf der Hand, die in Abschnitt 10.3.1 beschriebene erste Lösung zu verfeinern, indem ein geeignetes *Inventar nichtheuristischer Kriterien für den fragmentarischen Fall* operationalisiert wird. Abbildung 10.5 zeigt eine Menge von acht entsprechenden Regelmustern. (Folgende notationelle Konventionen liegen zugrunde: Syntaktische Konstituenten werden durch runde Klammern begrenzt; eckige Klammern heben die Fragmentgrenzen hervor; $bk(X)$ kennzeichnet die Bindende Kategorie eines Knotens X ; $nvk(X)$ kennzeichnet den nächstverzweigenden Knoten eines Knotens X im Sinne der Definition des K-Herrschafts-Begriffs; der Index von $X_{typ Y}$ spezifiziert die bindungstheoretische Klasse des Vorkommens-Stifters X als Y mit $Y \in \{A, B, C\}$, z.B. $\gamma_{typ B}$ ist ein nichtreflexives Pronomen; $\sqrt{/*}$ zeigt an, ob das jeweilige Regelmuster die Koindexierung von γ und α zuläßt oder ausschließt.)

Die Anwendbarkeit einer Regel ist genau dann gegeben, wenn sich eine Entsprechung zwischen deren konfiguralen Fragment-Mustern und den konkreten syntaktischen Fragmenten eines Belegs, in denen Anapher α bzw. Antezedenskandidat γ vorkommen, herstellen läßt. Hierbei müssen alle in der Regel formalisierten Bedingungen erfüllt sein. Als Beispiele sollen die Regeln [F1] und [F2] diskutiert werden, die die anhand der Beispiele (8) bzw. (9) identifizierten

¹⁵Per Entscheidungskombination zustandekommende nichtlokale Koindexierungen, wie etwa im Falle einer für das Beispiel naheliegenden *transitiven* Identifikation der Vorkommen “*Präsident*” und “*sich*”, sind natürlich - wie oben beschrieben - statthaft.

[F1]	✓	{ ... $F_i = [\dots bk(\gamma)(\dots \gamma_{typ B} \dots)] \dots$, ... , $F_j = [\dots bk(\alpha)(\dots \alpha_{typ B} \dots)] \dots$... }
[F2]	*	{ ... $F_i = [\dots nvk(\gamma)(\dots \gamma_{typ A/B/C} \dots)] \dots$, ... , $F_j = [\dots bk(\alpha)(\dots \alpha_{typ A} \dots)] \dots$... }
[E1a]	✓	{ ... $F_d = [\dots \gamma_{typ A/B/C} \dots]$, ... , $F_e = [\dots bk(\alpha)(\dots \alpha_{typ B} \dots)] \dots$... }
[E1b]	✓	{ ... $F_d = [\dots \alpha_{typ B/C} \dots]$, ... , $F_e = [\dots bk(\gamma)(\dots \gamma_{typ B} \dots)] \dots$... }
[E2]	*	{ ... $F_d = [\dots \gamma_{typ A/B/C} \dots]$, ... , $F_e = [\dots bk(\alpha)(\dots \alpha_{typ A} \dots)] \dots$... }
[E3a]	*	{ ... $F_d = [\dots \gamma_{typ A/B/C} \dots]$, ... , $F_e = [\dots \alpha_{typ C} \dots]$... }, falls gilt: γ k-beherrscht α unabhängig von der Entscheidung der strukturellen Ambiguität
[E3b]	*	{ ... $F_d = [\dots \alpha_{typ A/B/C} \dots]$, ... , $F_e = [\dots \gamma_{typ C} \dots]$... }, falls gilt: α k-beherrscht γ unabhängig von der Entscheidung der strukturellen Ambiguität
[E4]	*	{ ... $F_d = [\dots \alpha_{typ A} \dots]$, ... , $F_e = [\dots nvk(\gamma)(\dots \gamma_{typ A/B/C} \dots)] \dots$... }

Abbildung 10.5: Regeln für die Bindungsprinzip-Verifikation auf fragmentarischer Syntax

Entscheidungsmuster schematisieren. Für Beleg (8) etwa besteht folgende strukturell-formale Entsprechung zu den in Regel [F1] angegebenen Fragment-Mustern:

F_i	=	Fragment des Kandidaten γ (Hauptsatz + “nachdem”-Satz)
γ	=	NP des im “nachdem”-Satz realisierten Typ-B-Kandidaten γ (“er”)
$bk(\gamma)$	=	S-Knoten des “nachdem”-Satzes - Bindende Kategorie von γ
F_j	=	Fragment der Anapher α (strukturell ambiger “weil”-Satz)
α	=	NP der im “weil”-Satz realisierten Typ-B-Anapher α (“er”)
$bk(\alpha)$	=	S-Knoten des “weil”-Satzes - Bindende Kategorie von α

Als Typ-B-Pronomen erfüllen sowohl α als auch γ die in der Regel spezifizierten jeweiligen bindungstheoretischen Typ-Bedingungen. Die Anwendungsbedingungen der Regel [F1] sind somit vollständig erfüllt und deren Vorhersage - Akzeptabilität (✓) der Koindexierung von α und γ - wird damit wirksam.

In analoger Weise ergibt sich die Entsprechung zwischen Beleg (9) und Regel [F2]:

F_i	=	Fragment des Kandidaten γ (Hauptsatz + “bevor”-Satz)
γ	=	NP des im Hauptsatz realisierten Typ-C-Kandidaten γ (“Kohl”)
$nvk(\gamma)$	=	S-Knoten des Hauptsatzes - nächstverzweigender Knoten zu γ
F_j	=	Fragment der Anapher α (strukturell ambiger “weil”-Satz)
α	=	NP der im “weil”-Satz realisierten Typ-A-Anapher α (“sich”)
$bk(\alpha)$	=	S-Knoten des “weil”-Satzes - Bindende Kategorie von α

Im Unterschied zu Regel [F1] wird bezüglich des Fragments F_i des Antezedenskandidaten γ die weniger stringente Anforderung gestellt, daß der nächstverzweigende Knoten $nvk(\gamma)$ enthalten ist. Dies ist in diesem Fall hinreichend, da hierdurch bereits sichergestellt ist, daß in keiner der bestehenden Möglichkeiten zur strukturellen Rekombination der beiden Fragmente F_i und F_j eine K-Herrschafts-Relation (als Vorbedingung einer lokalen Bindung) zwischen Typ-A/B/C-Kandidat und Typ-A-Anapher besteht. Hierfür ist nicht notwendig, daß F_i auch die Bindende Kategorie des Kandidaten γ enthält. (In Beleg (9) fallen $nvk(\gamma)$ und $bk(\gamma)$ akzidentuell zusammen.) Gemäß Regel [F2] ist die konstruktive Antezedensentscheidung $\alpha \rightarrow \gamma$ somit definitiv zu

blockieren (*).

Die bisher beschriebenen Regeln [F1] und [F2] nehmen zwar Bezug auf die Fragmentierung, sie sind jedoch nicht sensitiv für etwaige spezifische Relationen zwischen den Fragmenten von Anapher und Kandidat. Bereits zuvor wurde jedoch darauf hingewiesen, daß die Syntaxanalyse in bestimmten, durchaus nicht seltenen Fällen nähere Information betreffend die Arten zulässiger struktureller Rekombination zweier Fragmente zur Verfügung stellt - beispielsweise ist für die Adverbialsatz-Ambiguität der Belege (8) und (9) i.d.R. bekannt, daß zwischen Hauptsatz-Fragment und Nebensatz-Fragment eine *gerichtete* Einbettungsbeziehung besteht: Ungeachtet der Wahl einer Interpretationsalternative ist das Fragment des Adverbialsatzes stets dem Fragment des Hauptsatzes subordiniert. Liefert der Parser diese Information, so können die Anwendungsbedingungen der oben beschriebenen Entscheidungsmuster abgeschwächt werden. Im Falle der symmetrisch aufgebauten Regel [F1] ist für das jeweils dominierende Fragment nicht mehr länger zu fordern, daß die Bindende Kategorie des entsprechenden Vorkommens enthalten ist. Das asymmetrisch strukturierte Regelschema [F2] ist in zweierlei Hinsicht zu spezialisieren. Dominiert das Fragment des Kandidaten das der Anapher, so wird die Enthaltenseins-Bedingung des nächstverzweigenden Knotens des Kandidaten obsolet - die Verletzung von Bindungsprinzip A ergibt sich alleine aufgrund der verletzten Lokalitäts-Anforderung der Bindungs-Relation. Ist umgekehrt das Fragment der Anapher der Dominator, so wird alleine auf die Nichterfüllung der zweiten Bedingung von Bindungsprinzip A getestet - die geforderte K-Herrschafts-Beziehung zwischen Kandidat und Anapher ist bereits dann verletzt, wenn das Fragment des Kandidaten dessen nächstverzweigenden Knoten enthält.

In Abbildung 10.5 sind die entsprechenden Entscheidungsmuster unter den Bezeichnern [E1a] und [E1b] (Spezialisierungen zu [F1]) bzw. [E2] und [E4] (Spezialisierungen zu [F2]) beschrieben.¹⁶ (F_d steht hierbei für das dominierende, F_e für das subordinierte Fragment.) Liefert die Syntaxanalyse etwa für die Beispiele (8) bzw. (9) die zusätzliche Information der Einbettung des Fragments der Anapher in das Fragment des Kandidaten, so wird die jeweilige (positive bzw. negative) Entscheidung über die Regeln [E1a]/[E1b] (je nachdem, welches der beiden Typ-B-Pronomen als Anapher α betrachtet wird) bzw. [E2] bewerkstelligt; die Entsprechungen zwischen den Regelschemata und der vorliegenden Fragmentierungen ergeben sich in derselben Weise wie zuvor. Die Schemata [E1a] und [E1b] sind nicht völlig symmetrisch, denn für den Fall, daß die als Anapher α betrachtete Okkurrenz im dominierenden Fragment enthalten ist (Regel [E1b]), darf diese kein Reflexiv- oder Reziprok-Pronomen sein, da das entsprechende Bindungsprinzip A nicht - wie anaphernseitig zu fordern - konstruktiv erfüllt wäre.

Die bisherige Erläuterung der Regeln erfolgte auf der Basis von Beispielen, in denen eine Ambiguität der *Adverbialsatz-Interpretation* angenommen wurde. Die beschriebenen Entscheidungsmuster sind natürlich auch für die übrigen Fallklassen struktureller Mehrdeutigkeit relevant, wie folgendes Beispiel illustriert. Unter der Annahme, daß die *Anknüpfung des Relativsatzes* ambig ist und eine entsprechende syntaktische Fragmentierung besteht

(10) *Der Kanzler hat den Präsidenten unterstützt, der sich zur Wiederwahl stellte.*

kann auch hier unter Rekurs auf [F2] bzw. [E2] definitiv ausgeschlossen werden, daß die Okkurrenz "Kanzler" ein geeignetes Antezedens für das Reflexivum "sich" verkörpert. Anhand eines Beispiels der Fragmentbildung qua *Ambiguität einer PP* soll Regel [E3a] diskutiert werden.

¹⁶ "E" wie *Einbettung*

[F1]	α -/ γ -seitige BP-B-Erfüllung	Ausschluß der <i>lokalen</i> Bindung von α durch γ und umgekehrt
[F2]	α -seitige BP-A-Verletzung	Ausschluß der <i>lokalen</i> Bindung von α durch γ , bzw. Ausschluß, daß α durch γ k-beherrscht wird
[E1a]	α -seitige BP-B-Erfüllung	Ausschluß der <i>lokalen</i> Bindung von α durch γ
[E1b]	γ -seitige BP-B-Erfüllung	Ausschluß der <i>lokalen</i> Bindung von γ durch α
[E2]	α -seitige BP-A-Verletzung	Ausschluß der <i>lokalen</i> Bindung von α durch γ
[E3a]	α -seitige BP-C-Verletzung	Gewährleistung, daß α durch γ k-beherrscht wird
[E3b]	γ -seitige BP-C-Verletzung	Gewährleistung, daß γ durch α k-beherrscht wird
[E4]	α -seitige BP-A-Verletzung	Ausschluß, daß α durch γ k-beherrscht wird

Abbildung 10.6: Bindungstheoretischer Hintergrund der Regelschemata

(11) *Der Präsident sah seine Frau neben dem Amtskollegen.*

Unabhängig davon, ob die als strukturell mehrdeutig präsupponierte PP attributiv oder adverbial interpretiert wird, k-beherrscht die angenommene Kandidaten-NP “*der Präsident*” als Subjekt des Hauptsatzes die Anaphern-NP “*dem Amtskollegen*”. Die Anwendungsvoraussetzungen von [E3a] sind somit erfüllt, was zum Ausschluß der entsprechenden Koindexierung führt. (Regelschema [E3b] deckt den symmetrischen Fall ab, der etwa dann gegeben wäre, wenn für das besprochene Beispiel eine kataphorische Bezugnahme angenommen, d.h. “*der Präsident*” als anaphorisches Vorkommen betrachtet würde.) Der Hintergrund der Regeln [E3a] und [E3b] besteht in der in jedem Fall gegebenen Verletzung des Bindungsprinzips C der Anapher. Auch Regel [E4] soll anhand eines Belegs illustriert werden, für den eine strukturelle Ambiguität einer PP angenommen wird.

(12) *Der Präsident rasiert sich mit dem Messer.*

In diesem Fall ist die Anapher im dominierenden Fragment enthalten. Die per Unterstreichung hervorgehobene Antezedenszuordnung kann in diesem Fall ausgeschlossen werden, weil die Kandidaten-NP - gleich welche PP-Interpretation gewählt wird - die NP der Anapher nicht k-beherrscht. In Regel [E4] wird diese Bedingung durch die Enthaltenseins-Bedingung des nächstverzweigenden Knotens des Kandidaten im subordinierten Fragment schematisiert. Bindungsprinzip A der Anapher, das eine (lokale) Bindungsbeziehung fordert, kann somit in derartigen Konfigurationen niemals erfüllt sein.

Die Tabelle in Abbildung 10.6 faßt den bindungstheoretischen Hintergrund der acht Regelschemata zusammen. Für jede Regel ist angegeben, auf der Grundlage von welchem Bindungsprinzip und in bezug auf welche(s) Vorkommen die jeweilige Bewertung erfolgt (mittlere Spalte). In der rechten Spalte wird aufgeschlüsselt, wie die Entscheidung zustandekommt.¹⁷

Auch für den i-über-i-Filter können einige nichtheuristische Regelschemata für den fragmentarischen Fall formuliert werden. Ohne eine feinere Differenzierung nach den bestehenden Anknüpfungsoptionen kann zumindest in einigen Fällen, in denen Fragment-Wurzeln mit den NPs

¹⁷Bindungstheoretische Bedingungen, die aufgrund von Dominanzrelationen zwischen den Fragmenten erfüllt sind, bleiben hierbei unerwähnt ([E1a], [E1b]).

$$\begin{array}{l}
[\text{IF}] \quad * \quad \{ \dots F_i = [\gamma(\dots)], \dots, F_j = [\alpha(\dots)] \dots \} \\
[\text{IEa}] \quad * \quad \{ \dots F_d = [\gamma(\dots)], \dots, F_e = [\dots \alpha(\dots) \dots] \dots \} \\
[\text{IEb}] \quad * \quad \{ \dots F_d = [\alpha(\dots)], \dots, F_e = [\dots \gamma(\dots) \dots] \dots \}
\end{array}$$

Abbildung 10.7: Regelschemata des i-über-i-Filters auf fragmentarischer Syntax

der Anapher oder des Kandidaten zusammenfallen, eine definitiv negative Entscheidung getroffen werden. Erneut bietet sich eine Unterscheidung danach an, ob Einbettungs-Information zur Verfügung steht (IEa)/(IEb) oder nicht (IF). In Abbildung 10.7 werden die entsprechenden Schemata beschrieben. Die angegebenen Restriktionen verstehen sich als *Ergänzung* der in Abbildung 10.5 beschriebenen Regelschemata der Bindungsprinzipien: Die positiven Vorhersagen der Regeln [F1], [E1a] und [E1b] kommen ausschließlich dann zum Tragen, wenn keine i-über-i-Konfiguration vorliegt.¹⁸

10.3.3 Eine robuste Version des Basisalgorithmus

Auf der Grundlage des beschriebenen Ensembles von Regelschemata ist es nunmehr möglich, eine Version des Basisalgorithmus zu formulieren, durch die eine verfeinerte Form der Robustheit gegenüber fragmentarischen oberflächenstrukturellen Beschreibungen erreicht wird. Auch für bestimmte Konstellationen des Falles, in dem Anapher und Antezedenskandidat in unterschiedlichen oberflächenstrukturellen Fragmenten vorkommen, wird die Verifikation der bindungstheoretischen Bedingungen in nichtheuristischer Form bewerkstelligt. Folglich steigt das erreichte Ausmaß an Monotonie - der Ansatz erweist sich somit als graduell robuster als die in Abschnitt 10.3.1 skizzierte elementare Lösung. In bezug auf die Verifikation der bindungstheoretischen Bedingungen läßt sich das Verfahren folgendermaßen beschreiben:

Kommen Anapher α und Antezedenskandidat γ in demselben oberflächenstrukturellen Fragment vor, so verifiziere die bindungstheoretischen Bedingungen wie im Falle von nichtfragmentarischer Syntax. Kommen Anapher α und Antezedenskandidat γ in unterschiedlichen Fragmenten vor, so prüfe, ob eines der in Abbildung 10.5 beschriebenen Regelschemata anwendbar ist. Falls ja, so entscheide gemäß (einer) dieser Regel(n).¹⁹ Falls nein, so betrachte die Koindexierung von α und γ als bindungstheoretisch zulässig, jedoch reduziere die Plausibilitätsbewertung der möglichen Entscheidung $\alpha \rightarrow \gamma$.

¹⁸Hinsichtlich der Zielsetzung einer vollständig nichtheuristischen *Bestätigung* der Zulässigkeit von Koindexierungen im Fragmentierungsfall erweist es sich hierbei als problematisch, daß die Nichtanwendbarkeit der angegebenen i-über-i-Regelschemata in keinem Fall die definitive *Erfülltheit* der i-über-i-Bedingungen impliziert, da es sich ausschließlich um *Filter-Regeln* (mit negativen Vorhersagen) handelt. Die entstehende Unschärfe erweist sich jedoch in der Praxis als allenfalls marginal relevant, zumal der Skopus des i-über-i-Filters in der syntaktischen Konfiguration ohnehin begrenzt zu sein scheint.

¹⁹Die Menge der Regelschemata ist *konsistent*, sind also auf eine bestimmte Fragmentierungs-Konfiguration mehrere Regeln gleichzeitig anwendbar - etwa [F1] und [E1a]/[E1b] (vgl. o.) - so spielt es keine Rolle, welche Regel angewendet wird, da die Vorhersage in jedem Fall dieselbe ist.

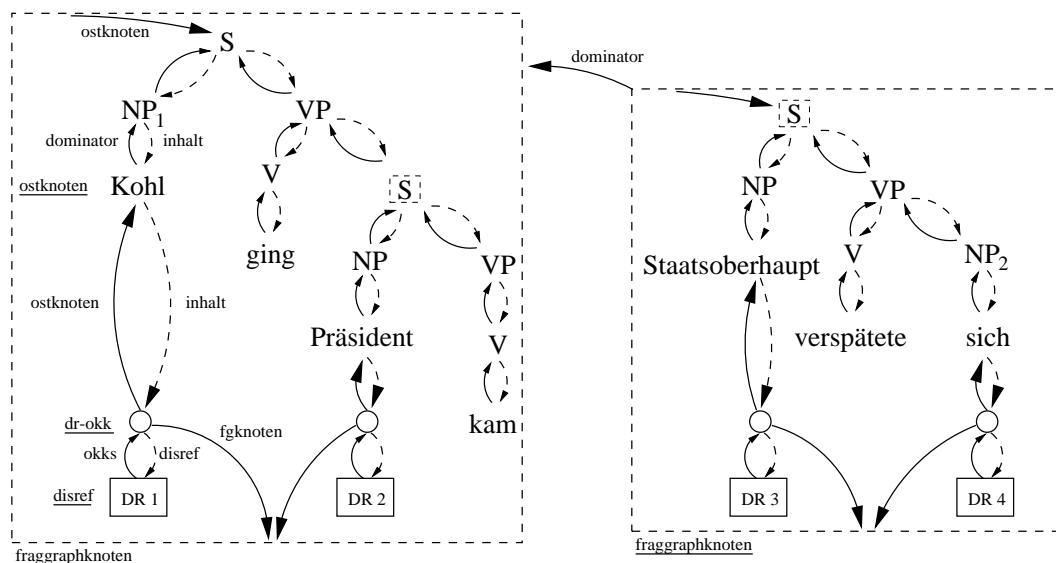


Abbildung 10.8: Repräsentation der Fragmentierungs-Information für Beleg (6)

Um diese robuste Version F-SYNKONFIG der Strategie SYNKONFIG im Rahmen des Basisalgorithmus zu operationalisieren, sind die in Abschnitt 9.3.2 beschriebenen Datenstrukturen in geeigneter Weise zu verfeinern und zusätzliche Verifikationsfunktionen für die Regelschemata einzuführen.

Verfeinerung der Datenstrukturen

Anhand der Skizze des Verfahrens zur robusten Verifikation der Bindungsbedingungen lassen sich zusätzliche Anforderungen an die Datenstrukturen ableiten. Grundvoraussetzung ist eine geeignete Repräsentation der oberflächenstrukturellen Fragmentierung, in der festzuhalten ist, welche Fragmente existieren und welche Dominanzbeziehungen bestehen. Um an der relevanten Stelle entscheiden zu können, ob der fragmentarische oder nichtfragmentarische Fall vorliegt und - ggf. - welches Regelschema anwendbar ist, sollte diese Information von den Beschreibungen der Okkurrenzen aus zugreifbar sein.

In Abbildung 10.8 ist eine Verfeinerung der für den Basisalgorithmus entwickelten Datenstrukturen (vgl. Abbildung 9.5) am Beispiel von Beleg (6) skizziert, die diesen Anforderungen genügt. Die Fragmentierungsinformation wird durch Entitäten eines neuen Datentyps **fraggraphknoten** modelliert, die durch das zeigerwertige Attribut **dominator** entsprechend der bestehenden Dominanzbeziehungen partiell verküpft sind. Die Struktur, die sich auf dieser zusätzlichen Ebene der Repräsentation ergibt, ist somit eine Menge (ein "Wald") von Fragment-Hierarchien, die sich vermöge der Dominanzrelationen ergeben. Unter einem weiteren Merkmal **ostknoten** wird ein Rückverweis auf den Wurzelknoten des Fragments in der oberflächenstrukturellen Beschreibung verwaltet; diese Information erweist sich für die Operationalisierung der Regelschemata als nützlich (vgl. u.). Die Zugreifbarkeit der neuen Repräsentationsebene von Seiten der Vorkommen wird durch die Einführung eines zusätzlichen Attributs **fgknoten** in den in Abbildung 9.2 beschriebenen Datentyp **dr-okk** hergestellt, durch das auf das **fraggraphknoten**-Element des

syntaktischen Fragments verwiesen wird, dem die NP des jeweiligen Vorkommens angehört.

Operationalisierung der Regelschemata

In Ergänzung der in Abbildung 9.6 beschriebenen Funktionen für die Verifikation der Bindungsprinzipien im nichtfragmentarischen Fall verbleibt die Operationalisierung der in Abbildung 10.5 formal dargestellten Regelschemata zu beschreiben. Je nachdem, ob Anapher und Antezedens-kandidat demselben Fragment oder unterschiedlichen Fragmenten der oberflächenstrukturellen Beschreibung entspringen (Entscheidungsgrundlage ist ein Vergleich der Werte des jeweiligen Attributs **fgknoten**), kommt entweder das bisherige oder das zur Operationalisierung ausstehende neue Verifikationsverfahren zur Anwendung.

In Abbildung 10.9 werden zwei Funktionen beschrieben, die die Regelschemata [E1a]/[E1b], [E2], [E3a]/[E3b] und [E4] bzw. [F1] und [F2] operationalisieren. Die Funktion **E-definitiv?** ist anzuwenden, wenn Dominanzinformation bezüglich der Fragmente von Anapher und Kandidat verfügbar ist; Funktion **F-definitiv?** kommt im entgegengesetzten Fall zum Einsatz. Welche dieser beiden Funktionen aufzurufen ist, wird unter Rückgriff auf die verfügbare Dominanzinformation entschieden.²⁰

Die Funktion **E-definitiv?** wird mit drei Input-Argumenten aufgerufen - anaphorischem Vorkommen (**ana-okk**), Kandidaten-Okkurrenz (**kandi-okk**) und Richtung der bestehenden Einbettungsrelation (**dominanzrichtung**). Falls das Fragment des Kandidaten das Fragment der Anapher dominiert, so wird auf die Anwendbarkeit der Regeln [E1a], [E2] und [E3a] getestet (Randmarkierung (1a)). Liegt Dominanz in umgekehrter Richtung vor, so wird versucht, eines der Schemata [E1b], [E3b], [E4] anzuwenden (Randmarkierung (1b)). Die feinere Unterscheidung, welche der jeweiligen Regeln potentiell anwendbar ist, geschieht in beiden Fällen anhand des bindungstheoretischen Typs von Anapher und Kandidat.

Liegt etwa in ersterem der beiden Fälle eine Anapher vom Typ B vor, so kommt (ausschließlich) Regel [E1a] in Frage (Randmarkierung (E1a)). Als weitere Bedingung verbleibt zu testen, ob das Fragment der Anapher deren Bindende Kategorie enthält. Dies wird per Aufruf der Testfunktion **bindende-kategorie-in-fragment?** bewerkstelligt, die die entsprechende Suche in der oberflächenstrukturellen Beschreibung durchführt. Ist auch diese Kondition erfüllt, so werden zwei Werte zurückgegeben (Randmarkierung (3)). Das erste Ergebnis - der aussagenlogische Wahrheitswert **t (rue)** - signalisiert, daß eines der Regelschemata anwendbar ist und folglich die *heuristische* Akzeptanz der Koindexierung entfallen kann. Das zweite Ergebnis - hier ebenfalls der Wahrheitswert **t** - beschreibt das eigentliche Resultat der Regelanwendung - in diesem Fall die Vorhersage von [E1a], daß die Koindexierung von Anapher und Kandidat in der vorliegenden Fragment-Konfiguration definitiv zulässig ist.

Die Verifikation der Bedingungen der übrigen Regeln geschieht analog. Die Suche nach dem nächstverzweigenden Knoten (Regeln [E3a]/[E3b] und [E4]) wird durch eine zweite Testfunktion **naechst-verzweigender-knoten** bewerkstelligt, die - soweit im Fragment der jeweils betrachteten Okkurrenz existent - die entsprechende Entität zurückgibt. Ist keine der Regeln anwendbar, so wird als erster Wert das Ergebnis **nil** (entsprechend dem Wahrheitswert **false**) zurückge-

²⁰Die Operationalisierung eines entsprechenden Tests ist angesichts der Verfügbarkeit der benötigten Information über das Attribut **dominator** der **fraggraphknoten**-Entitäten elementar. Ausgehend vom **fraggraphknoten**-Element, das das syntaktische Fragment der einen Okkurrenz beschreibt, wird so lange den **dominator**-Zeigern gefolgt, bis entweder die **fraggraphknoten**-Entität des Fragments der anderen Okkurrenz erreicht wird (Dominanz-Fall) oder die Wurzel; in letzterem Fall ist darüberhinaus (in analoger Weise) auf die umgekehrte Dominanz-Konstellation zu testen.

```

(defun F-definitiv? (ana-okk kandi-okk)
  (let ((ana-bttyp (dr-okk-bt-klasse ana-okk))
        (kandi-bttyp (dr-okk-bt-klasse kandi-okk))
        (ana-NP-ostkn (ostknoten-dominator (dr-okk-ostknoten ana-okk)))
        (kandi-NP-ostkn (ostknoten-dominator (dr-okk-ostknoten kandi-okk))))
    (if (eq ana-bttyp 'A) ;;; (F2)
        ;; => auf Anwendbarkeit der Regel [F2] zu testen
        (if (and (bindende-kategorie-in-fragment? ana-NP-ostkn)
                 (naechst-verzweigender-knoten kandi-NP-ostkn))
            (values t nil) ;;; (3)
            (values nil nil)) ;;; (4)
        (if (and (eq ana-bttyp 'B) ;;; (F1)
                 (eq kandi-bttyp 'B))
            ;; => auf Anwendbarkeit der Regel [F1] zu testen
            (bindende-kategorie-in-fragment? ana-NP-ostkn)
            (bindende-kategorie-in-fragment? kandi-NP-ostkn))
            (values t t) ;;; (3)
            (values nil nil)))) ;;; (4)
    );;F-definitiv?

(defun E-definitiv? (ana-okk kandi-okk dominanzrichtung)
  (let ((ana-bttyp (dr-okk-bt-klasse ana-okk))
        (ana-NP-ostkn (ostknoten-dominator (dr-okk-ostknoten ana-okk)))
        (kandi-NP-ostkn (ostknoten-dominator (dr-okk-ostknoten kandi-okk))))
    (if (eq dominanzrichtung 'ana-in-kandi)
        ;; => Fragment der Anapher subordiniert, Test auf [E1a], [E2], [E3a] ;;; (1a)
        (case ana-bttyp ;;; (2)
          (A (if (bindende-kategorie-in-fragment? ana-NP-ostkn) ;;; (E2)
                 (values t nil) ;;; (3)
                 (values nil nil))) ;;; (4)
          (B (if (bindende-kategorie-in-fragment? ana-NP-ostkn) ;;; (E1a)
                 (values t t) ;;; (3)
                 (values nil nil))) ;;; (4)
          (C (if (eq (naechst-verzweigender-knoten kandi-NP-ostkn) ;;; (E3a)
                    (fraggraphknoten-ostknoten (dr-okk-fgknoten kandi-okk)))
                 (values t nil) ;;; (3)
                 (values nil nil)))) ;;; (4)
        ;; => Fragment des Kandidaten subordiniert, Test auf [E1b], [E3b], [E4] ;;; (1b)
        (if (and (eq ana-bttyp 'A) ;;; (E4)
                 (naechst-verzweigender-knoten kandi-NP-ostkn))
            (values t nil) ;;; (3)
            (case (dr-okk-bt-klasse kandi-okk) ;;; (2)
              (B (if (and (not (eq ana-bttyp 'A)) ;;; (E1b)
                          (bindende-kategorie-in-fragment? kandi-NP-ostkn))
                     (values t t) ;;; (3)
                     (values nil nil))) ;;; (4)
              (C (if (eq (naechst-verzweigender-knoten ana-NP-ostkn) ;;; (E3b)
                        (fraggraphknoten-ostknoten (dr-okk-fgknoten ana-okk)))
                     (values t nil) ;;; (3)
                     (values nil nil))) ;;; (4)
                  (t (values nil nil)))))) ;;; (4)
    );;E-definitiv?

```

Abbildung 10.9: Bindungsprinzip-Verifikation auf fragmentarischer Oberflächenstruktur

geben (Randmarkierungen (4)); hiermit wird signalisiert, daß für das betrachtete Vorkommenspaar keine definiten Entscheidungskriterien zur Verfügung stehen.

Das Regelpaar [E3a]/[E3b] ist nur partiell operationalisiert: Die Bedingung der K-Herrschaft des einen Vorkommens über das jeweils andere “*unabhängig von der Entscheidung der strukturellen Ambiguität*” wird bislang ausschließlich per Test auf den Sonderfall umgesetzt, in dem eine der Okkurrenzen - etwa als Subjekt eines finiten Satzes - den gesamten Rest ihres Fragments k-beherrscht; dies ist genau dann der Fall, wenn der nächstverzweigende Knoten dieses Vorkommens gleich der Wurzel des Fragments ist.²¹ Eine Verfeinerung der Umsetzung der Schemata [E3a]/[E3b] ist prinzipiell möglich, sie bedingt jedoch eine Bezugnahme auf detailliertere Information über die für das subordinierte Fragment bestehenden Interpretations-Optionen. Ob sich dieser Aufwand lohnt, kann erst unter Betrachtung der durch einen konkreten Parser typischerweise gelieferten Ergebnisstrukturen entschieden werden.

Die Operationalisierung der Regelschemata [F1] und [F2] durch die Funktion `F-definitiv?` bedient sich derselben Elemente und soll deshalb nicht näher diskutiert werden. Nach demselben Muster verläuft die Umsetzung der in Abbildung 10.7 beschriebenen elementaren Regelschemata des i-über-i-Filters, auf deren Besprechung daher ebenfalls verzichtet wird.

Verfeinerung des Basisalgorithmus

Im Unterschied zu der in Abbildung 9.6 beschriebenen Verifikation der Bindungsprinzipien im nichtfragmentarischen Fall erfolgt für ein gegebenes Paar von Anapher und Antezedenskandidat stets nur *ein* Funktionsaufruf. Ist eines der Regelschemata anwendbar, das die *Zulässigkeit* der Koindexierung attestiert ([F1] sowie [E1a]/[E1b]), so garantiert dies bereits die Erfüllung der Bindungsprinzipien sowohl der Anapher als auch des Kandidaten.

Somit ergibt sich folgendes Bild bezüglich der Verfeinerung des in Abbildung 9.9 beschriebenen Basisalgorithmus. Zum einen ist Schritt 1b, in dem die originäre Verifikation der Bindungsbedingungen stattfindet, per Einführung einer Fallunterscheidung zu verfeinern, durch die das eingangs informell beschriebene robuste Verfahren realisiert wird: Im fragmentarischen Fall ist zunächst auf die Anwendbarkeit der Regelschemata zu prüfen; erst falls dies scheitert, ist das zur Entscheidung ausstehende Vorkommenspaar auf heuristischer Basis zu akzeptieren. Die Verifikation der Regelschemata umfaßt die Bedingungen sowohl der Bindungsprinzipien (Abbildung 10.5) als auch des i-über-i-Filters (Abbildung 10.7). Eine definitiv (nichtheuristisch) positive Entscheidung auf der Grundlage der Regeln der Bindungsprinzipien wird hierbei (marginal vereinfachend) bereits dann als *definitiv* positives *Gesamtergebnis* angenommen, falls die Regeln des i-über-i-Filters *nicht anwendbar* sind.²²

Zum anderen verbleibt natürlich weiterhin auf mögliche Entscheidungsinterdependenzen zu testen; auch dieser Schritt ist in geeigneter Weise auf fragmentarische Syntax zu adaptieren. Wie

²¹Strenggenommen fallen die durch dieses Vorkommen selbst dominierten oberflächenstrukturellen Knoten nicht unter diese Bedingung, denn eine K-Herrschafts-Relation zwischen zwei Knoten setzt voraus, daß keiner der beiden Knoten den anderen dominiert. Dennoch ist die Vorhersage der Unzulässigkeit der Koindexierung uneingeschränkt korrekt, denn im Falle, daß das zu subordinierende Fragment an diese Knoten angeknüpft wird, greift der i-über-i-Filter.

²²Andernfalls ergäben sich auf der Basis der Kombination dieser beiden Regelschema-Mengen ausschließlich definitiv *negative* Entscheidungen. Die beschriebene Verfahrensweise scheint aber unter sowohl theoretischen als auch anwendungspraktischen Gesichtspunkten gerechtfertigt: Einerseits sind Gültigkeitsbereich und Reichweite der i-über-i-Kondition in der Theorie nicht unkontrovers - insofern ist bereits mit zweifelhaften Vorhersagen zu rechnen; andererseits wird die i-über-i-Bedingung in den weitaus meisten Fällen, in denen die Regeln 10.7 nicht anwendbar sind, erfüllt sein. Aus diesen Gründen wird an dieser Stelle auf ein Plausibilitätsdekrement verzichtet.

1. Für jede anaphorische Okkurrenz Y , ermittle die Menge der möglichen Antezedenten X :
 ...
 (b) (F-SYNKONFIG) Falls der Antezedens kandidat X intrasententiell ist, verifiziere, daß die Bindungsbedingungen von Y und X erfüllt sind.
 - Falls X und Y in demselben oberflächenstrukturellen Fragment vorkommen, so verfare gemäß Schritt 1b des Basisalgorithmus
 - Falls X und Y in unterschiedlichen Fragmenten vorkommen, so
 - i. versuche, die *Bindungsprinzipien* von X und Y durch Anwendung eines der Regelschemata definitiv zu verifizieren:
 - Ist eine Einbettungsbeziehung zwischen den beiden Fragmenten bekannt, so versuche Regelschemata [E1a], [E1b], [E2], [E3a], [E3b] und [E4];
 - ansonsten versuche Regeln [F1] und [F2].
 Falls keines der Regelschemata anwendbar ist, so *nimm heuristisch an*, daß die Bindungsprinzipien von X und Y *erfüllt* sind;
 - ii. verifiziere ferner, daß keine (definitive) *i-über-i-Konfiguration* vorliegt:
 - im Falle einer Einbettungsbeziehung über die Regelschemata [IEa], [IEb];
 - ansonsten über Regel [IF].
- ...
 2. Bewertung und Sortierung nach Plausibilität:
 ...
 (a') (Plausibilitätsbewertung ff.) Für jedes Paar (Y_i, X_j) von Anapher und Antezedens kandidat, für das in Schritt 1(b)i die Erfüllung der Bindungsprinzipien *heuristisch angenommen* wurde: Reduziere die Plausibilitätsbewertung $v(Y_i, X_j)$.
 ...
 3. Antezedensauswahl: Betrachte Anaphern Y und jeweilige Antezedens kandidaten $X(Y)$ in derselben Reihenfolge wie in Schritt 3 des Basisalgorithmus. Wähle $X(Y)$, falls keine Interdependenz besteht, d.h. genau dann wenn
 ...
 (b) (F-SYNKONFIG) für alle Okkurrenzen Z_1 und Z_2 , die dem Diskursreferenten zu $X(Y)$ bzw. Y im *aktuellen* Lauf der Analyse zugeordnet wurden, gilt: Die Koindexierung von Z_1 und Z_2 , die transitiv aus der Wahl von $X(Y)$ als Antezedens für Y resultiert, *verletzt nicht* die Bindungsprinzipien und fällt nicht unter den i-über-i-Filter -
 - falls Z_1 und Z_2 in demselben oberflächenstrukturellen Fragment vorkommen, so verfare gemäß der näheren Spezifikation von Schritt 3b des Basisalgorithmus (nicht-konstruktive Verifikation der Bindungsprinzipien qua Ausblendung von BP A);
 - falls X und Y in unterschiedlichen Fragmenten vorkommen, so verfare wie in den Schritten 1(b)i und 1(b)ii, wobei jedoch die Regelschemata [F2], [E2] und [E4], die BP A konstruktiv verifizieren, nicht angewendet werden.
 (Der Spezialfall $Z_1 = X(Y) \wedge Z_2 = Y$ braucht nicht erneut betrachtet zu werden.)
 ...

Abbildung 10.10: Verfeinerung des Basisalgorithmus für fragmentarische Syntax

im bindungstheoretischen Interdependenz-Testschritt 3b des Basisverfahrens ist im entsprechenden Schritt des verfeinerten Algorithmus auf eine konstruktive Anwendung von Bindungsprinzip A zu verzichten, um - im Einklang mit der bindungstheoretischen Vorhersage - zusätzliche nicht-bindende Koindexierungen von Typ-A-Pronomen zuzulassen.²³ Im fragmentarischen Fall wird

²³vgl. Abschnitt 9.3.2, Diskussion zu Beleg (2)

dies per Verzicht auf die Anwendung der Regelschemata [F2], [E2] und [E4] erreicht, deren Ausschlußkriterium auf BP A zurückzuführen ist (vgl. Abbildung 10.6).

In Abbildung 10.10 sind die Verfeinerungen des Basisalgorithmus zusammengefaßt. Schritt 1b operationalisiert die im Rahmen dieses Abschnitts erarbeitete Verifikation der Bindungsprinzipien, die robust gegenüber fragmentarischer Oberflächenstruktur ist; Schritt 3b bewerkstelligt Entsprechendes im Rahmen des Interdependenz-Testschritts. Die nunmehr verfeinerte Interpretationsstrategie der syntaktisch-konfiguralen Restriktionen läuft unter neuem Namen F-SYNKONFIG.

10.4 Ein vollständiger Algorithmus

Der in Abschnitt 10.2 entwickelte Algorithmus zur Verifikation der Bindungsbedingungen erreicht einen höheren Grad an Robustheit gegenüber fragmentarischen oberflächenstrukturellen Beschreibungen dadurch, daß nichtheuristische Entscheidungen auch in bestimmten Konfigurationen getroffen werden, in denen Anapher und Kandidat in unterschiedlichen Fragmenten vorkommen. Allen durch die Anwendung der Regelschemata sowie der Verifikationsfunktionen des Basisalgorithmus getroffenen bindungstheoretischen Vorhersagen ist gemein, daß deren Gültigkeit unabhängig von den Entscheidungen einer weitergehenden strukturellen Disambiguierung gewährleistet ist - *auf lesarten-spezifische syntaktisch-konfigurationale Evidenz wird nicht zurückgegriffen*. Im folgenden soll ein Algorithmus erarbeitet werden, in dem zur Verifikation der Bindungsbedingungen auch solche konfigurationale Evidenz ausgenutzt wird, die nur in einer *Teilmenge* der syntaktischen Lesarten gültig ist. Auf der Grundlage eines Protokolls der Lesarten-Abhängigkeiten der Einzelentscheidungen wird es möglich werden, die Ergebnisse der referentiellen Interpretation zu einer strukturellen Disambiguierung der oberflächenstrukturellen Beschreibung heranzuziehen. Die solchermaßen ermöglichte *Interaktion* zwischen den Teilprozessen der referentiellen und der strukturellen Interpretation erlaubt die Operationalisierung eines Verfahrens, das im Sinne der Festlegungen in Abschnitt 10.1 als *ultimativ robust* bezeichnet werden kann.

10.4.1 *Packed Shared Forests* zur Repräsentation struktureller Ambiguität

Im Hinblick auf die erweiterten Zielsetzungen des im folgenden zu entwickelnden Verfahrens stellen die in Abschnitt 10.2 als Eingabe zugrundegelegten fragmentarischen syntaktischen Beschreibungen mit (ggf. partieller) Einbettungsinformation nicht mehr länger den geeigneten Ausgangspunkt dar. Es soll nunmehr von einem Repräsentationsformat ausgegangen werden, mit Hilfe dessen die unresolvierten strukturellen Mehrdeutigkeiten des Syntaxanalyse-Ergebnisses vollständig beschrieben werden können.²⁴ Eine Enumeration der Syntaxbäume ist zu vermeiden, da das Problem der kombinatorischen Explosion besteht (vgl. Abschnitt 5.4.1).

Ein Standard-Repräsentationsschema, das den identifizierten Anforderungen genügt, sind sog. "*Packed Shared Forests*" (PSFs) (vgl. Tomita, [Tomi85]). In diesem Format werden oberflächenstrukturelle Mehrdeutigkeiten durch das "*Packing*" der unterschiedlichen Ableitungs-Varianten

²⁴U.a. erfordert dies eine geeignete Abbildung *bedingter* struktureller Rekombinationsmöglichkeiten, die in dem bislang zugrundegelegten partiellen Beschreibungsformat außen vor blieben.

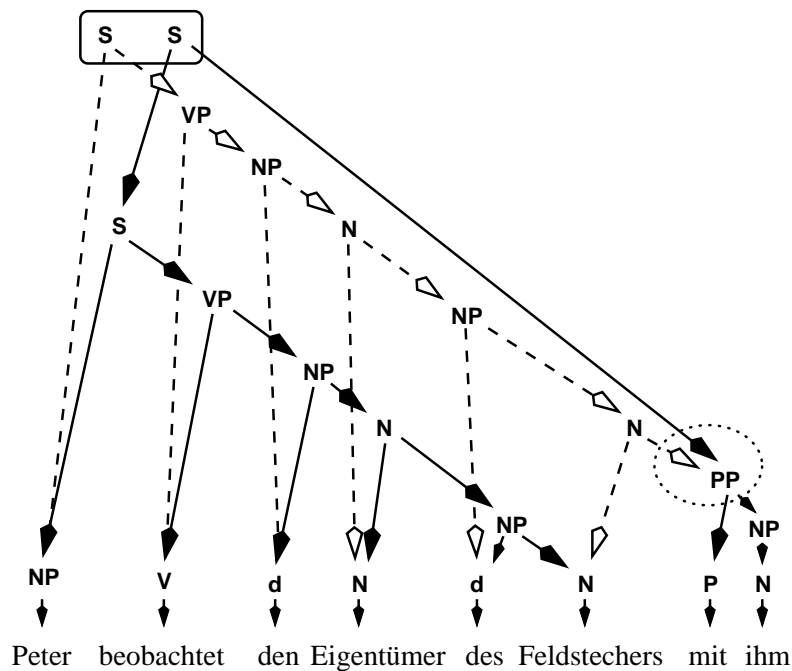


Abbildung 10.11: Repräsentation oberflächenstruktureller Ambiguität in PSFs

ein und desselben Ausschnitts des analysierten Satzes in einen *einzelnen* inneren Knoten des Repräsentationsgraphen abgebildet; die notwendige Beschreibungsökonomie wird durch multiplen Verweis auf identische Teilanalysen (“*Sharing*”) erreicht. Abbildung 10.11 zeigt ein elementares Beispiel. Gesetzt den Fall, die PP “*mit ihm*” wird als ambig bezüglich der Alternativen adverbiale vs. attributive Interpretation analysiert. Entsprechend ergeben sich für den Knoten S an der Wurzel des Syntaxbaums zwei Möglichkeiten, die durch ein *Packing* der adverbialen und der attributiven Ableitungs-Variante (unterbrochene bzw. durchgezogene Linien) abgebildet werden. Da die strukturell ambige PP in beide Lesarten in identischer Form einfließt, ergibt sich eine (nichttriviale) Instanz des *Sharing*, die sich in *zwei* auf den PP-Knoten zeigenden Pfeilen niederschlägt (unterbrochenes Oval).²⁵

Unter formalem, graphentheoretischem Gesichtspunkt verkörpert ein PSF einen *gerichteten azyklischen Graphen* mit eindeutig bestimmtem Wurzelement (dem obersten S-Knoten); die Blätter entsprechen den Wörtern des Satzes, dessen oberflächenstrukturelle Interpretationsalternativen beschrieben werden. Sei T ein PSF mit internen (nichtterminalen) Knoten $V = \{v_1, \dots, v_k\}$, seien ferner $P(v_1), \dots, P(v_k)$ die Mengen der für die internen Knoten bestehenden (“gepackten”) Ableitungs-Alternativen. Demnach entspricht die Anzahl $|P(v_i)|$ der Elemente einer Menge $P(v_i)$ der Anzahl von Ableitungsvarianten des Knotens v_i . (*Packing* im eigentlichen Sinne ist genau dann gegeben, wenn $|P(v_i)| > 1$ ist.) Unter Bezugnahme auf diese allgemeine, formale Charakterisierung von PSFs kann nun eine obere Schranke für die Anzahl n der Lesarten, die ein PSF T in kompakter Form repräsentiert, gegeben werden. Im Extremfall multiplizieren sich die strukturellen Ambiguitäten, d.h. es gilt

²⁵Darüberhinaus gibt es im Falle von Mehrdeutigkeit stets das (triviale) *Sharing* der lexikalischen Kategorien (der Präterminale, die unmittelbar die Wörter der Eingabe dominieren) des Syntaxbaums.

$$n = \prod_{i=1}^k |P(v_i)|$$

Folglich können Mengen von Lesarten durch Bitvektoren der Länge n repräsentiert werden.

10.4.2 Dominanzrelationen in PSFs

Die Operationalisierung der bindungstheoretischen Restriktionen basiert in wesentlichen Teilen auf den Dominanzbeziehungen, die zwischen den Knoten des Oberflächenstrukturbaums bestehen. Während diese konfigurationsalen Relationen in unambigen bzw. fragmentarischen Beschreibungen (fragment-lokal) eindeutig bestimmt sind, besteht im Falle von PSF-Repräsentationen struktureller Mehrdeutigkeit aufgrund des *Packings* eine *Lesarten-Abhängigkeit*: Die Gültigkeit einer Dominanzrelation kann auf eine bestimmte *Teilmenge* aller Lesarten beschränkt sein. In Abbildung 10.11 beispielsweise dominiert der NP-Knoten mit Kopf “*Feldstecher*” nur in einer der beiden Lesarten den NP-Knoten des Ausdrucks “*ihm*”. Es ist somit notwendig, die Dominanzbeziehung zwischen zwei Knoten v_i und v_j nicht länger durch einen einfachen aussagenlogischen Wert - v_i dominiert bzw. dominiert nicht v_j - zu charakterisieren, sondern als Bitvektor der Länge n mit folgender Bedeutung aufzufassen:²⁶

$$\vec{d}_k(v_i, v_j) = \left\{ \begin{array}{l} 1 \Leftrightarrow v_i \text{ dominiert } v_j \text{ in Lesart } k \\ 0 \Leftrightarrow v_i \text{ dominiert } v_j \text{ in Lesart } k \text{ nicht} \end{array} \right\}, 1 \leq k \leq n$$

Die im Rahmen der Verifikation der Bindungsprinzipien zentrale Unterscheidung zwischen lokaler und nichtlokaler Bindung macht eine feinere *Differenzierung zwischen lokaler und nichtlokaler Dominanz* notwendig. Das Interesse konzentriert sich ferner auf die Dominanzbeziehungen, die zwischen bestimmten inneren Knoten v_i des Syntaxbaums - den nächstverzweigenden Knoten zu vorkommens-induzierenden Knoten - und vorkommens-induzierenden Knoten v_o bestehen. Eine geeignete Grundlage für eine lesarten-sensitive Operationalisierung der bindungstheoretischen Bedingungen wird somit dadurch geschaffen, daß für relevante Knotenpaare (v_i, v_o) *Dominanzvektoren* $\vec{ld}(v_i, v_o)$ und $\vec{ad}(v_i, v_o)$ berechnet werden: $\vec{ld}(v_i, v_o)$ kennzeichnet die Menge der Lesarten, in denen der Knoten v_i den Knoten v_o im Sinne des bindungstheoretischen Lokalitätsbegriffs *lokal* dominiert; entsprechend charakterisiert $\vec{ad}(v_i, v_o)$ die Lesarten, in denen eine *arbiträre* (nicht notwendig lokale) Dominanzrelation besteht. Diese lesarten-qualifizierte Information fließt in den Prozeß der Bindungsbedingungs-Verifikation in noch näher zu beschreibender Weise ein. Zunächst ist jedoch ein Algorithmus zu beschreiben, mit Hilfe dessen die Vektoren $\vec{ld}(v_i, v_o)$ und $\vec{ad}(v_i, v_o)$ berechnet werden können. Eine Möglichkeit, diese Aufgabe zu bewerkstelligen, ergibt sich vermöge einer Generalisierung der u.a. von Correa ([Corr88]) sowie Ingria und Stallard ([InSt89]) vorgeschlagenen Technik einer vollständigen *Vorab-Berechnung* dieser Information per rekursiver Traversierung des PSFs.²⁷ Ausgangspunkt sind Knotenpaare (v_p, v_o) , wobei v_o ein vorkommens-induzierender Knoten und v_p ein präterminaler, d.h. unmittelbar die Wörter des

²⁶Zur Notation: $\vec{d}(v_i, v_j) \in \{0, 1\}^n$ ist ein n -stelliger boolescher (aussagenlogischer) Vektor und $\vec{d}_k(v_i, v_j)$ dessen Wert ($\in \{0, 1\}$) an der Stelle k .

²⁷Alternativ könnten die Dominanz-Vektoren *anforderungs-gesteuert* berechnet werden, d.h. es würden nur diejenigen Vektoren berechnet, die für die bindungstheoretische Bewertung der tatsächlich zur Entscheidung ausstehenden Okkurrenz-Paare unmittelbar oder mittelbar benötigt werden. Dies entspräche der Vorgehensweise, die

beschriebenen Satzes dominierender Knoten ist. Als *vorkommens-induzierend* werden die dem *lexikalischen Kopf* der Vorkommens-NP entsprechenden präterminalen Knoten - in Abbildung 10.11 beispielsweise der N-Knoten unmittelbar oberhalb des Worts “*Feldstechers*” - und nicht etwa die vollständigen NPs der Vorkommen selbst angesehen, da sich für letztere bereits lesarten-spezifische Dominanzvektoren ergeben, die rekursiv zu berechnen sind. Für die Gesamtheit der präterminalen Knoten wird ferner (vereinfachend) angenommen, daß sie vermöge *Sharing* in *alle* Lesarten einfließen. (Diese Annahme ist genau dann nicht erfüllt, wenn Mehrdeutigkeiten bezüglich der Wortarten bestehen. In diesem Fall bedarf das im folgenden beschriebene Berechnungsschema einer Verfeinerung in Form einer *lesarten-qualifizierten Initialisierung* der Dominanzvektoren $\vec{ad}(v_p, v_o)$ und $\vec{ld}(v_p, v_o)$; die benötigte Information wird durch eine elementare *Top-Down*-Traversierung des PSF gewonnen.)²⁸

Als Verankerung des rekursiven Berechnungs-Schemas werden jedem Paar (v_p, v_o) folgende zwei Dominanz-Vektoren zugeordnet:

$$\vec{ad}(v_p, v_o) = \vec{ld}(v_p, v_o) := \begin{cases} (1, \dots, 1), & v_p = v_o \\ (0, \dots, 0), & v_p \neq v_o \end{cases}$$

Durch die angegebenen Initialisierungs-Vektoren werden reflexive (triviale) Dominanzbeziehungen für die - zur Verifikation der Bindungsprinzipien relevanten - vorkommens-induzierenden präterminalen Knoten festgelegt.

Im weiteren Verlauf der *aufsteigend-rekursiven Berechnung* werden diese Vorgaben schrittweise auf die übrigen internen Knoten v_i und damit insbesondere auf die ebenfalls bindungstheoretisch relevanten nächstverzweigenden Knoten der Okkurrenz-NPs propagiert.²⁹ In einem bestimmten Rekursionsschritt werden jeweils die Dominanzvektoren eines noch nicht betrachteten Knotens v_i berechnet, für den bereits alle Kinder (unmittelbar dominierte Knoten) in allen Ableitungs-Varianten der Menge $P(v_i)$ betrachtet wurden. Da PSFs *azyklische* Graphen sind, ist die Existenz eines solchen Knotens bis zum abschließenden Erreichen der Wurzel stets garantiert. Die Berechnungen der Vektoren $\vec{ad}(v_i, v_o)$ und $\vec{ld}(v_i, v_o)$ unterscheiden sich insofern, als im letzteren Fall zu berücksichtigen ist, ob die Grenze der lokalen Bindungsdomäne des Knotens v_o überschritten wird und demnach keine *lokale* Dominanzbeziehung mehr besteht.

Abbildung 10.12 zeigt die beiden Rekursionsformeln, mit Hilfe derer sich die Dominanzvektoren für die internen Knoten schrittweise berechnen lassen. (Die binären Operatoren \wedge und \vee stehen für die bitweise *Konjunktion* bzw. *Disjunktion* von Bitvektoren $\in \{0, 1\}^n$; Querstriche über den Bitvektoren spezifizieren den unären Operator der bitweisen *Komplementbildung*.) Das Grundschema der Berechnung läßt sich anhand der Formel für die Berechnung der Vektoren

in den oben beschriebenen Versionen des Basisalgorithmus gewählt wurde.

Die im Rahmen dieses Abschnitts entwickelte Operationalisierungs-Variante ist jedoch laufzeit-effizienter - durch die Darlegung des Alternativ-Ansatzes wird insofern auch eine Explikations-Lücke der Ausführungen in Abschnitt 9.4.3 geschlossen. Nicht zuletzt erscheint die hier bewerkstelligte “*Bottom-Up*”-Entwicklung des Verfahrens konzeptuell klarer.

²⁸An dieser Stelle bleibt ferner außer Betracht, daß damit auch die Eigenschaft eines Knotens, ein Vorkommen zu induzieren, lesarten-spezifisch ist. Auch für dieses Problem existiert eine theoretisch befriedigende Lösung: Unter der Annahme, daß der (zu beschreibende) Algorithmus zur referentiellen Interpretation ein solches potentielles Vorkommen entweder als (resolvierte) Anapher oder als (erfolgreichen) Antezedens-Kandidaten ansieht, sind nur noch diejenigen Lesarten relevant, in denen ein okkurrenz-induzierendes Präterminal vorliegt.

²⁹Die Möglichkeit einer gezielten Beschränkung der Berechnungen auf Knoten v_i , die nächstverzweigende Knoten sind, besteht i.a. nicht, da die Information bezüglich der Ableitungsvarianten von internen Knoten, die durch nächstverzweigende Knoten dominiert werden, für die Lesarten-Qualifizierung der Dominanzvektoren relevant und somit in adäquater Weise zu berücksichtigen sind.

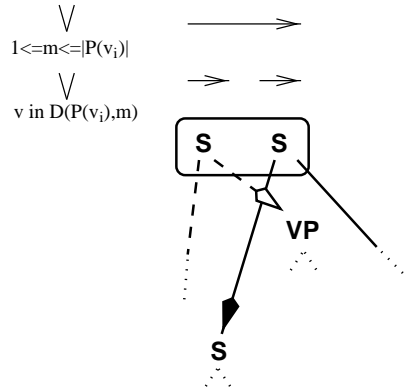
$$\vec{ad}(v_i, v_o) := \bigvee_{1 \leq m \leq |P(v_i)|} (\vec{p}(P(v_i), m) \wedge (\bigvee_{v_d \in D(P(v_i), m)} \vec{ad}(v_d, v_o)))$$

$$\vec{ld}(v_i, v_o) := \bigvee_{1 \leq m \leq |P(v_i)|} (\vec{p}(P(v_i), m) \wedge (\bigvee_{v_d \in D(P(v_i), m)} (\vec{ld}(v_d, v_o) \wedge \overline{bk}(v_d, v_o))))$$

Abbildung 10.12: Rekursionsformeln für die Berechnung der Dominanzvektoren

$\vec{ad}(v_i, v_o)$, v_o präterminal und vorkommens-induzierend, erläutern. Die äußere Disjunktion summiert über alle Ableitungsvarianten, die sich vermöge *Packing* für Knoten v_i ergeben. Bitvektor $\vec{p}(P(v_i), m)$ verkörpert einen Filter, der die Teilmenge aller Lesarten charakterisiert, in denen die m -te Ableitungsvariante von v_i , $1 \leq m \leq |P(v_i)|$, gültig ist. (Ausgehend von der Festlegung eines spezifischen Codierschemas zur Charakterisierung von Lesarten-Teilmengen des PSF durch Bitvektoren ist die Berechnung der Vektoren $\vec{p}(P(v_i), m)$ mit Hilfe einer elementaren Formel möglich.) Für jede Ableitungs-Variante m existiert eine Menge $D(P(v_i), m)$ derjenigen Knoten (Kinder) v_d , die v_i in dieser Variante *unmittelbar dominiert*.³⁰ Die Knoten, die durch die Kinder v_d dominiert werden, werden transitiv durch v_i selbst dominiert. Somit ergeben sich die Vektoren $\vec{ad}(v_i, v_o)$ für v_i durch rekursives Aufsummieren der Vektoren $\vec{ad}(v_d, v_o)$ aller Kinder v_d (innere Disjunktion), wobei die notwendige Lesarten-Qualifizierung per Konjunktion mit dem Vektor $\vec{p}(P(v_i), m)$ der jeweiligen Ableitungs-Variante bewerkstelligt wird.

Folgende Skizze illustriert das beschriebene Iterationsschema anhand des gepackten Wurzelknotens des PSFs aus Abbildung 10.11. Zwei Ableitungs-Varianten sind zu betrachten (oberer Pfeil); entsprechend finden zwei lokale Summierungen über die jeweiligen Kinder statt (untere Pfeile).



Die Berechnung der Vektoren $\vec{ld}(v_i, v_o)$ geschieht auf analogem Wege. Der im zusätzlichen lokalen Konjunkt verwendete Vektor $\overline{bk}(v_d, v_o)$ ist folgendermaßen definiert:

$$\overline{bk}_k(v_d, v_o) = \begin{cases} 1 & \Leftrightarrow v_d \text{ ist Bindende Kategorie von } v_o \text{ in Lesart } k \\ 0 & \Leftrightarrow \text{sonst} \end{cases}, \quad 1 \leq k \leq n$$

³⁰Unter Rekurs auf die dem PSF zugrundeliegende formale (i.a. kontextfreie) Grammatik läßt sich $D(P(v_i), m)$ als die Menge aller Symbole (Nichtterminale) auf der *rechten Seite der Ersetzungsregel* beschreiben, die in Variante m von $P(v_i)$ zur Anwendung kommt.

(Die Berechnung von $\vec{b}\vec{k}(v_d, v_o)$ verläuft parallel zur Berechnung der Dominanz-Vektoren.)³¹

Durch die Spezifikation des bitweisen Komplements $\vec{b}\vec{k}(v_d, v_o)$ dieses Vektors wird folglich bewirkt, daß die Bits genau derjenigen Lesarten, in denen ein Knoten v_d überschritten wurde, der eine Bindende Kategorie des betrachteten Vorkommens-Induktors v_o verkörpert, auf 0 gesetzt werden. Somit operationalisiert die angegebene Rekursionsformel in Übereinstimmung mit der Definition von $\vec{l}\vec{d}$ die Berechnung des Vektors, der genau die Lesarten charakterisiert, in denen eine lokale Dominanzbeziehung zwischen v_i und v_o besteht.

10.4.3 Verifikation der Bindungsbedingungen auf PSFs

Auf der Basis der Vektoren $\vec{a}\vec{d}(v_i, v_o)$ und $\vec{l}\vec{d}(v_i, v_o)$ können nun für ein gegebenes Vorkommenspaar bestehend aus Anapher α und Antezedenskandidat γ die bindungstheoretischen Bedingungen verifiziert werden. Entsprechend den Vorgaben der Bindungstheorie ist eine konstruktive bzw. nichtkonstruktive Verifikation der Bindungsprinzipien zu bewerkstelligen. Hierzu ist ein Bitvektor $\vec{r}(\alpha, \gamma)$ zu bestimmen, der diejenigen Lesarten des PSFs charakterisiert, in denen die Koindexierung von α und γ konfigural zulässig ist. Unter Rückgriff auf die (noch zu definierenden) Vektoren $\vec{b}\vec{p}\vec{k}$ bzw. $\vec{b}\vec{p}\vec{n}$ der Bindungsprinzip-Verifikation kann $\vec{r}(\alpha, \gamma)$ auf folgende Weise berechnet werden:

$$\vec{r}(\alpha, \gamma) := \left(\bigvee_{v \in nvk(\gamma)} (\vec{p}(P(v), m) \wedge \vec{b}\vec{p}\vec{k}(v, \alpha)) \right) \wedge \left(\bigvee_{v \in nvk(\alpha)} (\vec{p}(P(v), m) \wedge \vec{b}\vec{p}\vec{n}(v, \gamma)) \right)$$

Hierbei verkörpert $nvk(x)$ die Menge der Dominatoren eines Knotens x , die *nächstverzweigende Knoten* im Sinne der K-Herrschafts-Definition sind (vgl. auch Fußnote 32).

Konjunkt 1 (erste Zeile der Formel) spezifiziert den Bitvektor, der die Lesarten des PSF charakterisiert, in denen das Bindungsprinzip der Anapher α (konstruktiv) erfüllt ist. In analoger Weise beschreibt Konjunkt 2 (zweite Zeile der Formel) diejenigen Lesarten, in denen dem Bindungsprinzip des Kandidaten γ (nichtkonstruktiv) Genüge getan ist. In beiden Fällen sind *nächstverzweigende Knoten* zu betrachten, da diese entsprechend den Vorgaben der K-Herrschafts-Definition die Ausgangspunkte für die Suche nach Dominanzbeziehungen sind. Folglich ist über die Menge aller nichtverzweigenden Knoten zu summieren (Disjunktionen). Die Eigenschaft eines Knotens v , nächstverzweigender Knoten eines anderen Knotens x zu sein, kann *lesarten-abhängig*, d.h. beschränkt auf eine Teilmenge der Ableitungsvarianten $P(v)$ sein. Somit bedürfen die einzelnen Disjunkte einer Lesarten-Qualifizierung durch die bereits beschriebenen Vektoren $\vec{p}(P(v), m)$, $v \in nvk(x)$.³²

Die eigentliche Verifikation der Bindungsprinzipien von Anapher α und Kandidat γ erfolgt per

³¹ Auch $\vec{b}\vec{k}(v_d, v_o)$ liefert u.U. Vorhersagen, die sich für die Ableitungsvarianten von v_d unterscheiden können. Die Berechnung von $\vec{b}\vec{k}(v_d, v_o)$ hat daher gemäß einem Rekursionsschema zu erfolgen, das simultan zur Berechnung der Dominanzvektoren verläuft und anhand der Ableitungsvarianten sowie ferner des jeweiligen Ableitungs-Zweiges, in dem v_o vorkommt, die Lesarten ermittelt, in denen v_d Bindende Kategorie zu v_o ist.

³² Hierbei ist ferner zu berücksichtigen, daß ein Knoten v u.U. mehrere Summanden zur Disjunktion beiträgt. Dies ist genau dann der Fall, wenn zu einem bestimmten Knoten x v nächstverzweigender Knoten in mehr als einer Ableitungsvariante ist. Insofern sind die Mengen $nvk(x)$ als *Multi-Mengen* aufzufassen, in denen *Duplikate möglich* sind.

Konjunktion mit den Vektoren $\vec{bpk}(v, \alpha)$ bzw. $\vec{bpn}(v, \gamma)$. Die Definition dieser Vektoren geschieht unter Rückgriff auf die in Abschnitt 10.4.2 operationalisierten Dominanzvektoren:

$$\vec{bpk}(v, \alpha) := \begin{cases} \vec{ld}(v, \alpha) & \text{falls } \text{typ}(\alpha) = A \\ \overline{\vec{ld}(v, \alpha)} & \text{falls } \text{typ}(\alpha) = B \\ \overline{\vec{ad}(v, \alpha)} & \text{falls } \text{typ}(\alpha) = C \end{cases}$$

$$\vec{bpn}(v, \gamma) := \begin{cases} (1, \dots, 1) & \text{falls } \text{typ}(\gamma) = A \\ \overline{\vec{ld}(v, \gamma)} & \text{falls } \text{typ}(\gamma) = B \\ \overline{\vec{ad}(v, \gamma)} & \text{falls } \text{typ}(\gamma) = C \end{cases}$$

Der bindungstheoretische Hintergrund dieser Festlegungen ist elementar. Welche Formel anwendbar ist, entscheidet sich anhand des bindungstheoretischen Typs $\text{typ}(x) \in \{A, B, C\}$ des betrachteten Vorkommens x , der das zu verifizierende Bindungsprinzip determiniert. Beispielsweise wird Bindungsprinzip B von Anapher α (bzw. Kandidat γ) dadurch operationalisiert, daß genau die Lesarten *ausgeblendet* werden, in denen der nächstverzweigende Knoten v von γ (bzw. α) ein lokaler Dominator von α (bzw. γ) ist. Dies geschieht durch die Bildung des bitweisen Komplements des Vektors $\vec{ld}(v, \alpha)$ (bzw. $\vec{ld}(v, \gamma)$).³³ Der einzige Unterschied zwischen der konstruktiven und der nichtkonstruktiven Verifikation der Bindungsprinzipien besteht für Vorkommen, die Bindungsprinzip A unterliegen (Reflexiva, Reziproka) - die bereits beschriebene Zulässigkeit *zusätzlicher* nichtlokaler Koindexierungen schlägt sich in der Definition von \vec{bpn} per Spezifikation des Vektors $(1, \dots, 1)$ nieder, der die lesarten-unabhängige nichtkonstruktive Erfüllung von BP A zum Ausdruck bringt.

Auch der bislang noch nicht berücksichtigte i-über-i-Filter ist unter Rekurs auf die lesartenqualifizierte Dominanzinformation operationalisierbar. Ausgangspunkte für die Anwendung der Dominanzvektoren sind hierbei nicht die nächstverzweigenden Knoten, sondern die Menge aller NP-Knoten von Anapher α bzw. Kandidat γ .³⁴

$$\vec{if}(\alpha, \gamma) := \bigwedge_{\delta \in np(\alpha)} \overline{\vec{ad}(\delta, \gamma)} \wedge \bigwedge_{\delta \in np(\gamma)} \overline{\vec{ad}(\delta, \alpha)}$$

Ergebnis ist ein Bitvektor, der - entsprechend den Vorgaben der i-über-i-Bedingung - alle und nur die Lesarten repräsentiert, in denen keiner der jeweiligen NP-Knoten der beiden Okkurrenzen die jeweils andere NP dominiert.

Anhand des Beispiel-PSFs in Abbildung 10.11 läßt sich die Notwendigkeit der Berücksichtigung von *Mengen* von NP-Knoten verdeutlichen. U.a. ist der NP-Knoten des Vorkommens-Induktors "Feldstechers" lesarten-spezifisch. Die i-über-i-Bedingung schließt eine Koindexierung mit dem

³³Auf den ersten Blick erscheint es notwendig, seitens des auf Dominiertheit zu überprüfenden Vorkommens (gegeben in Form eines präterminalen Knotens $v_o \in \{\alpha, \gamma\}$ als rechtem Argument der Vektoren \vec{ad} bzw. \vec{ld}) von der Menge der NP-Knoten $np(v_o)$ auszugehen, da in PSFs der NP-Knoten zu v_o nicht notwendig eindeutig bestimmt ist (vgl. Abbildung 10.11). Es erweist sich jedoch als hinreichend, unmittelbar von v_o auszugehen, da durch die Definitionen der Berechnungsschemata für die Dominanzvektoren gewährleistet ist, daß alle relevanten NP-Knoten $\delta \in np(v_o)$ Berücksichtigung finden (vgl. Abbildung 10.12).

³⁴vgl. hierzu erneut die Bemerkung in Fußnote 33

(lesartenunabhängigen) NP-Knoten des Pronomens “*ihm*” nur in der durch unterbrochene Linien gekennzeichneten Lesart aus; in der zweiten Lesart besteht die Dominanzbeziehung, die diesem Ausschluß zugrundeliegt, nicht.

10.4.4 Verfeinerung des Basisalgorithmus

Auch für das in diesem Abschnitt spezifizierte Verfahren zur lesarten-sensitiven Verifikation der Bindungsbedingungen auf *Packed Shared Forests* soll eine Integration in den in Abbildung 9.9 beschriebenen Basisalgorithmus bewerkstelligt werden. Die spezifizierten Formeln zur Verifikation der Bindungsbedingungen kommen wiederum in den zwei relevanten Phasen des Algorithmus zur Anwendung. Während der initialen Phase der Restriktions-Anwendung wird für jedes betrachtete Paar von Anapher α und Kandidat γ der Vektor $\vec{r}(\alpha, \gamma)$ entsprechend der Definition in Abschnitt 10.4.3 berechnet. Die Koindexierung eines Paares $((\alpha, \gamma))$ wird genau dann als konfigurationsal zulässig angesehen, wenn $\vec{r}(\alpha, \gamma)$ nicht bezüglich aller Lesarten 0 ist. An dieser Stelle wird der Unterschied zu dem in Abschnitt 10.3 entwickelten Ansatz der referentiellen Interpretation auf fragmentarischer Syntax deutlich: Während dort ausschließlich solche Koindexierungen akzeptiert werden, die in *allen* Lesarten den bindungstheoretischen Vorgaben entsprechen, werden hier bereits solche Koindexierungen als konfigurationsal zulässig angesehen, für die *zumindest eine* Lesart existiert.

Während der abschließenden Auswahlphase wird auch hier u.a. verifiziert, daß die transitiv entstehenden Koindexierungen den bindungstheoretischen Bedingungen entsprechen. Wie in den beiden zuvor entwickelten Algorithmen kommt dabei zur Bindungsprinzip-Verifikation ausschließlich das *nichtkonstruktive* Prädikat \vec{bpn} zum Einsatz. Zusätzlich zur transitiven Verträglichkeit der Einzelentscheidungen ist in dem hier beschriebenen Verfahren ferner zu gewährleisten, daß *zumindest eine* Lesart existiert, in der die Bindungsbedingungen aller - unmittelbar bzw. transitiv resultierender - Koindexierungen *gleichzeitig* erfüllt sind. Der Vektor \vec{r}^* , der die globale Lesarten-Abhängigkeit charakterisiert, ist folgendermaßen definiert:

$$\vec{r}^* = \bigwedge_{\text{ante}(\gamma, \alpha)} (\vec{r}(\alpha, \gamma) \wedge \vec{iff}(\alpha, \gamma)) \wedge \bigwedge_{\text{trans}(\eta_1, \eta_2)} (\vec{r}'(\eta_1, \eta_2) \wedge \vec{iff}(\eta_1, \eta_2))$$

Ersteres Konjunkt spezifiziert die Lesarten-Abhängigkeit der getroffenen Antezedens-Entscheidungen, letzteres die der transitiv entstehenden Koindexierungen. Vektor \vec{r}' ist definiert wie \vec{r} mit dem einzigen Unterschied, daß auch für das erste Argument das nichtkonstruktive Prädikat \vec{bpn} angewendet wird.

Somit ergibt sich als weitere Restriktion, daß Vektor \vec{r}^* nicht vollständig 0 wird.

Abbildung 10.13 faßt die beschriebenen Verfeinerungen des Basisalgorithmus zusammen. In Schritt 1b findet der einzelentscheidungsbezogene Test der Bindungsbedingungen unter Rekurs auf die lesarten-sensitiven Prädikate \vec{r} und \vec{iff} statt. Der Test auf bindungstheoretische Entscheidungs-Interdependenz 3b umfaßt nunmehr zwei Komponenten. Schritt 3(b)i verifiziert die Zulässigkeit der transitiv resultierenden Koindexierungen. In Schritt 3(b)ii wird die zuvor identifizierte zusätzliche Bedingung der Existenz (mindestens) einer globalen, den konfigurationsalen Bedingungen aller Koindexierungen zugleich genügenden Lesart überprüft. Dies geschieht per inkrementeller Berechnung des oben definierten Vektors \vec{r}^* . Falls sich herausstellt, daß letzteres Kriterium unter Wahl eines der für die aktuell betrachtete Anapher verfügbaren Antezedenskandidaten nicht erfüllbar ist, kann es notwendig werden, bereits vollzogene Entscheidungen

1. Für jede anaphorische Okkurrenz α , ermittle die Menge der möglichen Antezedenten γ :
 ...
 (b) (P-SYNKONFIG) Falls der Antezedenskandidat γ intrasententiell ist, verifiziere, daß die *Bindungsprinzipien* und die *i-über-i-Bedingungen* einer Koindexierung von α und γ in mindestens einer Lesart des PSF erfüllt sind, d.h.:

$$\vec{r}(\alpha, \gamma) \wedge \vec{iif}(\alpha, \gamma) \neq (0, \dots, 0)$$

...

2. Bewertung und Sortierung nach Plausibilität:
 ...

3. Antezedenauswahl: Betrachte Anaphern α und jeweilige Antezedenskandidaten $\gamma(\alpha)$ in derselben Reihenfolge wie in Schritt 3 des Basisalgorithmus. Wähle $\gamma(\alpha)$, falls keine Interdependenz besteht, d.h. genau dann wenn folgende Tests erfolgreich verlaufen:

...

- (b) (P-SYNKONFIG) Initialisierung $\vec{r}^* := (1, \dots, 1)$.

- i. Für alle Okkurrenzen η_1 und η_2 , die dem Diskursreferenten zu $\gamma(\alpha)$ bzw. α im *aktuellen* Lauf der Analyse zugeordnet wurden, gilt: Es existiert mindestens eine Lesart des PSF, in der die Koindexierung von η_1 und η_2 , die transitiv aus der Wahl von $\gamma(\alpha)$ als Antezedens für α resultiert, sowohl die *Bindungsprinzipien nicht verletzt* als auch die *i-über-i-Bedingungen erfüllt*, d.h.:

$$\vec{r}(\eta_1, \eta_2) \wedge \vec{iif}(\eta_1, \eta_2) \neq (0, \dots, 0)$$

(Der Spezialfall $\eta_1 = \gamma(\alpha) \wedge \eta_2 = \alpha$ braucht nicht erneut betrachtet zu werden.)

- ii. Die unmittelbaren und transitiv entstehenden Koindexierungen des aktuell betrachteten Paares (α, γ) sind mit der Gesamtheit der Koindexierungen der bislang getroffenen Entscheidungen kompatibel, d.h. es existiert mindestens eine *gemeinsame* PSF-Lesart, in der alle bindungstheoretischen Bedingungen erfüllt sind:

$$\begin{aligned} \vec{r}^* &:= \vec{r}^* \wedge \vec{r}(\alpha, \gamma) \wedge \vec{iif}(\alpha, \gamma) \\ &\quad \wedge \bigwedge_{\eta_1, \eta_2 \text{ wie oben}} (\vec{r}(\eta_1, \eta_2) \wedge \vec{iif}(\eta_1, \eta_2)) \\ &\neq (0, \dots, 0) \end{aligned}$$

...

Abbildung 10.13: Verfeinerung des Basisalgorithmus für PSFs

zurückzunehmen. Um dieses Problem zu lösen, könnte Auswahlsschritt 3 des Algorithmus mit einer sog. *Backtracking*-Strategie ausgestattet werden.³⁵

Abschließend soll die asymptotische Laufzeitkomplexität des beschriebenen Algorithmus analysiert werden. Sei $m := |V|$ die Anzahl der Knoten des PSF, $q \leq m$ die Anzahl der Vorkommens-Kontributoren, $b := \max(\{|P(v)| \mid v \in V\})$ die maximale Anzahl von Ableitungs-Alternativen ($b \leq m$). Eine Analyse der Rekursionsformeln zur Berechnung der Vektoren $\vec{ad}(v_i, v_o)$, $\vec{ld}(v_i, v_o)$ und $\vec{bk}(v_i, v_o)$ zeigt auf, daß $O(b)$ Bitvektorverknüpfungen (Konjunktionen, Disjunktionen, Kom-

³⁵Die Notwendigkeit einer Rücknahme von Entscheidungen kann strenggenommen bereits im Rahmen des gewöhnlichen Interdependenz-Testschritts bestehen. In praktischen Anwendungen der Algorithmen sollten sich derartige Situationen allerdings eher selten ergeben, da oftmals satzübergreifende Alternativen der Antezedens-Wahl existieren, die den beschriebenen Bedingungen genügen.

plementbildungen) je zu betrachtendem Paar (v_i, v_o) durchzuführen sind. Da es $O(qm)$ Paare gibt, entstehen Gesamtkosten $O(bqm)$. Seien ferner m_{nvk} und m_{np} die Maximalzahl nächstverzweigender Knoten bzw. NP-Knoten eines Vorkommens-Kontributors v_o . Die Verifikation der Bindungsbedingungen für $O(q^2)$ Vorkommens-Paare verursacht Gesamtkosten von $O(m_{nvk}q^2)$ (Vektoren $\vec{r}(\alpha, \gamma)$) bzw. $O(m_{np}q^2)$ (Vektoren $\vec{iif}(\alpha, \gamma)$). Da die Operationen des Sortierschritts sowie die Berechnungen der Vektoren \vec{r}^j und \vec{r}^* während der Auswahlphase nicht teurer sind, wird die Laufzeitkomplexität des Algorithmus beschrieben durch

$$O(bqm) + O(m_{nvk}q^2) + O(m_{np}q^2) = O(qm(b + q))$$

insgesamt also $O(m^3)$ (ohne Backtracking), wobei zu berücksichtigen ist, daß letztere Abschätzung vergleichsweise grob ist, da die Werte b , m_{nvk} und m_{np} im Durchschnitt wesentlich kleiner ausfallen als m .

Unter der fernerer Berücksichtigung, daß jede der hier gezählten Operationen n (= Anzahl Lesarten) Bitverknüpfungen beinhaltet und daß w Bitoperationen (w = Wortlänge des Prozessors) parallel ausgeführt werden, ergibt sich eine *Worst-Case*-Zeitkomplexität von $O(\frac{n}{w}m^3)$. Somit hängt das Laufzeitverhalten des Algorithmus von der Anzahl der Lesarten ab, die im ungünstigsten Fall der kombinatorischen Explosion exponentiell in der Anzahl der NPs ist.

10.4.5 Strukturelle Disambiguierung durch referentielle Evidenz

Der in der Auswahlphase des Algorithmus schrittweise berechnete Vektor \vec{r}^* charakterisiert die Teilmenge der Lesarten, in denen die Kombination der getroffenen Entscheidungen bindungstheoretisch zulässig ist. Diese Information kann nun als zusätzliche Evidenz für die Disambiguierung der als PSF gegebenen oberflächenstrukturellen Beschreibung herangezogen werden. Hierzu ist der PSF folgendermaßen zu reduzieren:

Für alle Knoten $v \in V$: Eliminiere die m -te Ableitungsvariante, $1 \leq m \leq |P(v)|$, genau dann, wenn die Bedingung

$$\vec{p}(P(v), m) \wedge \vec{r}^* = (0, \dots, 0)$$

erfüllt ist, d.h. wenn deren charakterisierender Vektor $\vec{p}(P(v), m)$ orthogonal zu \vec{r}^ ist.*

Nunmehr wird deutlich, daß durch das beschriebene Verfahren die Voraussetzungen für ein paralleles Verarbeitungsmodell geschaffen werden, das eine *bidirektionale* Verschränkung der Prozesse der oberflächenstrukturellen und referentiellen Interpretation bewerkstelligt und damit das Merkmal einer graduell robusteren Verarbeitung per *Interaktion* aufweist. Anhand des Beispiel-PSFs in Abbildung 10.11 läßt sich das Potential eines solchen Vorgehens veranschaulichen: Je nachdem, welche Lesart gewählt wird, ist die Koindexierung der NPs zu “*Feldstechers*” und “*ihm*” entweder zulässig oder nicht (i-über-i-Bedingung). Anstatt hier inadäquaterweise davon auszugehen, daß die oberflächenstrukturelle Konfiguration (unidirektional) die bindungstheoretischen Antezedens-Optionen determiniert und daher aufgrund von struktureller Ambiguität

keine definitive Entscheidung möglich ist, wird eine verfeinerte Sichtweise eingenommen: Trifft der Teilprozeß der referentiellen Interpretation die entsprechende Antezedens-Entscheidung, so ergibt sich umgekehrt Evidenz für die strukturelle Disambiguierung. Nicht zuletzt entspricht das parallele Verarbeitungsmodell den theoretischen Vorgaben der GB-Theorie, denn eine isolierte Operationalisierung der bindungstheoretischen Bedingungen läßt außer Acht, daß deren beschriebenes Potential zur strukturellen Disambiguierung dort implizit mitmodelliert ist.

Bislang ungeklärt ist hingegen die Frage nach der genaueren Ausgestaltung der Interaktion in einem entsprechenden Prozeßmodell. Neben der oben skizzierten Strategie, derzufolge erst das Endergebnis der Referenzanalyse zur strukturellen Disambiguierung herangezogen wird, besteht ferner die Möglichkeit einer feineren Verzahnung der beiden Teilprozesse, indem bereits Einzelentscheidungen der referentiellen Interpretation berücksichtigt werden. Ob dies sinnvoll ist, hängt letztlich davon ab, wie groß der referentielle Beitrag zur oberflächenstrukturellen Disambiguierung in der Praxis ist. Eine Beantwortung dieser Frage kann somit erst auf der Basis von Implementierung und Evaluation des beschriebenen Algorithmus erfolgen.

10.5 Diskussion

In Rahmen dieses Kapitels wurden zwei Algorithmen zur referentiellen Interpretation entwickelt, die die gewünschte Eigenschaft der Robustheit gegenüber unvollständigen, d.h. ambigen bzw. partiellen oberflächenstrukturellen Beschreibungen aufweisen. Das in Abschnitt 10.3 beschriebene, auf fragmentarischer Syntax arbeitende Verfahren bewerkstelligt die Verifikation der bindungstheoretischen Bedingungen unter weitestgehender Bezugnahme auf konfigurationale Relationen, die - unabhängig von weiteren Entscheidungen der strukturellen Disambiguierung - in *jeder* Lesart des Syntaxanalyse-Ergebnisses gültig sind; ausschließlich in Fällen, in denen keines der Regelmuster zur definitiven Entscheidung anwendbar ist, wird die Koindexierung des betrachteten Vorkommens-Paars auf heuristischem Wege akzeptiert. Durch eine weiterreichende Ausnutzung der fragmentarischen konfiguralen Information weist dieser Ansatz im Vergleich zu Elementarlösungen - gemessen an der Monotonieeigenschaft im Sinne der Definition Menzels - einen höheren Grad an Robustheit auf. Basierend auf einem adäquaten (vollständigen und ökonomischen) Repräsentationsformat für ambige Syntaxanalyse-Ergebnisse, sog. *Packed Shared Forests*, wurde in Abschnitt 10.4 ein zweiter Algorithmus erarbeitet, der eine weitere Erhöhung der Robustheit erzielt, indem die Bindungsbedingungen in lesarten-sensitiver Weise operationalisiert werden. Koindexierungen werden bereits dann als akzeptabel angesehen, wenn die konfiguralen Restriktionen in einer (nichtleeren) Teilmenge aller Lesarten erfüllt sind. Durch Protokollierung der Lesarten-Abhängigkeiten der Menge aller unmittelbaren und transitiven Koindexierungs-Entscheidungen werden die Voraussetzungen für eine *Interaktion zwischen referentieller und oberflächenstruktureller Disambiguierung* - und somit für eine erneute Verbesserung der Robustheit im Sinne der zugrundegelegten Definition - geschaffen.

Unter anwendungspraktischen Gesichtspunkten ist jedoch noch nicht entschieden, welcher der beiden Ansätze der vorteilhaftere ist. Der aus theoretischer Perspektive zu bevorzugende vollständige, PSF-basierte Algorithmus läßt folgende Fragen offen:

1. Wie sollte die bidirektionale Verknüpfung der Teilprozesse der referentiellen und oberflächenstrukturellen Disambiguierung ausgestaltet werden? Wie groß ist der *praktische Nutzwert* der auf diese Weise bewerkstelligten Interaktion?

2. In den Ausführungen in Abschnitt 10.4 wurde ausschließlich der Fall oberflächenstruktureller Mehrdeutigkeit betrachtet. Wie ist jedoch für Eingaben zu verfahren, für die *keine syntaktische Lesart* gefunden wird?
3. Ist das *praktische Laufzeitverhalten* des Algorithmus auch im Falle einer großen Anzahl n von Lesarten akzeptabel?

10.5.1 Der ultimative Algorithmus?

Unter der Voraussetzung, daß ein geeignetes Verarbeitungsmodell gefunden wird, für das die Antwort auf die unter Punkt 1 genannte Relevanzfrage positiv ausfällt, bietet sich folgendes *Hybridverfahren* zur Lösung der unter den Punkten 2 und 3 genannten Probleme an. Die Anwendung des PSF-basierten Algorithmus geschieht nurmehr *lokal*, d.h. für den Umgang mit Instanzen struktureller Ambiguität, die innerhalb bestimmter, adäquat zu wählender Ausschnitte des vollständigen PSF (auffaßbar als *künstliche Fragmente*) lokalisiert sind; der Algorithmus für fragmentarische Syntax behandelt die ausschnitt-übergreifenden Fälle. Mögliche und augenscheinlich sinnvolle Rahmen für den Einsatz des vollständigen Verfahrens stellen PSF-Ausschnitte dar, die Haupt- und Nebensätzen entsprechen (Fragmente der Kategorie S). Die in der Praxis am häufigsten anzutreffenden Instanzen struktureller Mehrdeutigkeit - anknüpfungsambige Präpositionalphrasen - sind in der Regel innerhalb eines spezifischen Satzfragments lokalisierbar; unter dem beschriebenen Analysemodell würden diese Fälle durch den vollständigen Algorithmus bearbeitet. Die übrigen Instanzen struktureller Mehrdeutigkeit - anknüpfungsambige Satz-Fragmente usw. - sowie die durch die beschriebene Aufteilung ggf. künstlich geschaffene Fragmentierung würden durch den Algorithmus für fragmentarische Syntax behandelt.

Durch die lokal begrenzten Anwendungen des PSF-Algorithmus wird die Anzahl n der jeweils (lokal) zu handhabenden Lesarten reduziert, was sich in einem akzeptablen Laufzeitverhalten auch in denjenigen Fällen niederschlagen sollte, in denen das vollständige Satzgefüge hochgradig ambig ist. Darüberhinaus erlaubt das Hybridverfahren die Verarbeitung fragmentarischer Ergebnisse von Analysen, in denen *keine* Lesart gefunden wurde: Für die auf diese Weise bedingte Fragmentierung gelangt das Verfahren für fragmentarische Syntax zur Anwendung, wobei auch hier die Option zur fragmentlokalen Anwendung des PSF-Ansatzes besteht.

10.5.2 Ausblick

Unter der Voraussetzung, daß ein geeigneter Parser zur Verfügung steht, der oberflächenstrukturelle Beschreibungen im jeweils zugrundegelegten Format - fragmentarische Syntax bzw. PSFs - generiert, sind nunmehr alle Bedingungen für die *Implementierung* der in den Abschnitten 10.3 bzw. 10.4 beschriebenen Algorithmen zur robusten referentiellen Interpretation erfüllt - ein geeigneter Ausgangspunkt für eine tatsächlich operationale Anapherninterpretation ist somit geschaffen. Einer näheren Untersuchung bedarf hingegen das Ausmaß der Beeinträchtigung der Ergebnisqualität, die durch die Unvollständigkeit des syntaktischen Inputs bedingt ist. Reicht die verfügbare konfigurationale Information aus, um die in idealisierenden Szenarien auf vollständigen und eindeutigen oberflächenstrukturellen Beschreibungen erzielte Ergebnisqualität zumindest annäherungsweise zu reproduzieren? Insoweit ein unmittelbarer Vergleich überhaupt

möglich ist: Ergeben sich Anhaltspunkte dafür, daß die Ergebnisse derjenigen Ansätze übertroffen werden, die auf dem Modell der seichten Voranalyse (vgl. Definition 10.1) basieren, d.h. die - wie insbesondere das in Abschnitt 8.7 analysierte Verfahren von Kennedy und Boguraev - auf eine unrestringierte syntaktische Analyse verzichten?

Diese Fragestellungen sind nur auf der Grundlage einer konkreten *Systemimplementierung* sowie einer *formalen Evaluation* des Systems auf *relevanten Textkorpora* unter *Anwendungsbedingungen* beantwortbar. Ausschließlich im Rahmen eines solchen Programms können aussagefähige Daten gewonnen werden. Im nun folgenden Kapitel 11 der Arbeit soll eine derartige Untersuchung per Implementierung und Evaluation des *Softwaresystems ROSANA* durchgeführt werden - einem operationalen Verfahren zur referentiellen Interpretation, dem das Modell der potentiell fragmentarischen Voranalyse (Definition 10.2) zugrundeliegt, und das auf der in Abschnitt 10.3 erarbeiteten Strategie F-SYNKONFIG zur robusten Verifikation der Bindungsbedingungen basiert. Es wird sich zeigen, daß die Antworten auf die zentralen Fragen positiv ausfallen.

Kapitel 11

Implementierung und Evaluation

Eine Bewertung der Leistungsfähigkeit computergestützter Textanalyse-Verfahren, die wissenschaftlichen Ansprüchen genügt, ist ausschließlich auf der Grundlage einer formalen Evaluation von Implementierungen der zugrundeliegenden Algorithmen möglich. Die in der computerlinguistischen Literatur weit verbreitete informelle Diskussion von Ansätzen anhand von Einzelbeispielen erweist sich als unzureichend, da eventuelle Explikationslücken nicht offen zutage treten und darüberhinaus keine aussagekräftigen Ergebnisse betreffend die anwendungspraktische Tragweite gewonnen werden.¹ Bezogen auf das im Rahmen der vorliegenden Arbeit behandelte Problem der algorithmischen referentiellen Interpretation heißt dies, daß der Nachweis der Leistungsfähigkeit der in Kapitel 10 entwickelten Algorithmen unter *Anwendungsbedingungen*, d.h. unter Rekurs auf verfügbare Voranalyse-Systeme sowie ferner in bezug auf die Verarbeitung anwendungsrelevanter, nichttrivialer Texte zu erbringen ist. Auf diese Weise werden darüberhinaus die Voraussetzungen zur Evaluation der Relevanz der in die Algorithmen einfließenden Einzelstrategien geschaffen; diese Frage ist in noch nicht allen Fällen abschließend geklärt (vgl. Abbildung 7.7, S. 146). Nicht zuletzt besteht die Gelegenheit zu einer feineren Aufschlüsselung und Kategorisierung typischer Analysefehler, woraus sich Anhaltspunkte für eine weitere Verfeinerung der zugrundeliegenden Strategien ergeben.

Vergleichsmaßstab sind Analyseergebnisse, die durch menschliche Leser auf *intellektuellem* Wege erzeugt werden. Somit läßt sich die Evaluations-Aufgabe folgendermaßen umreißen:

Für ein gegebenes, anwendungstypisches Korpus von Texten: Ermittle den Grad der Übereinstimmung zwischen den auf algorithmischem Wege erzeugten Inhaltsbeschreibungen und den auf intellektuellem Wege generierten Inhaltsbeschreibungen.

Unter der zusätzlichen Annahme, daß der Abgleich der Inhaltsbeschreibungen auf der Basis formaler Evaluationsmaße und per Computer erfolgt, ergibt sich das in Abbildung 11.1 skizzierte generische *Evaluationsszenario*.

Anhand dieser ersten, vergleichsweise allgemeinen Charakterisierung lassen sich die *zentralen Aufgaben* identifizieren, die im Rahmen eines solchen Programms anfallen:

¹Ob dies im jeweiligen Fall in vollem Bewußtsein oder auf fahrlässige Weise geschieht, sei dahingestellt. Unabhängig hiervon stellen Einzelbeispiele natürlich eine adäquate Grundlage zur *Erläuterung* von Phänomenen bzw. Strategien dar.

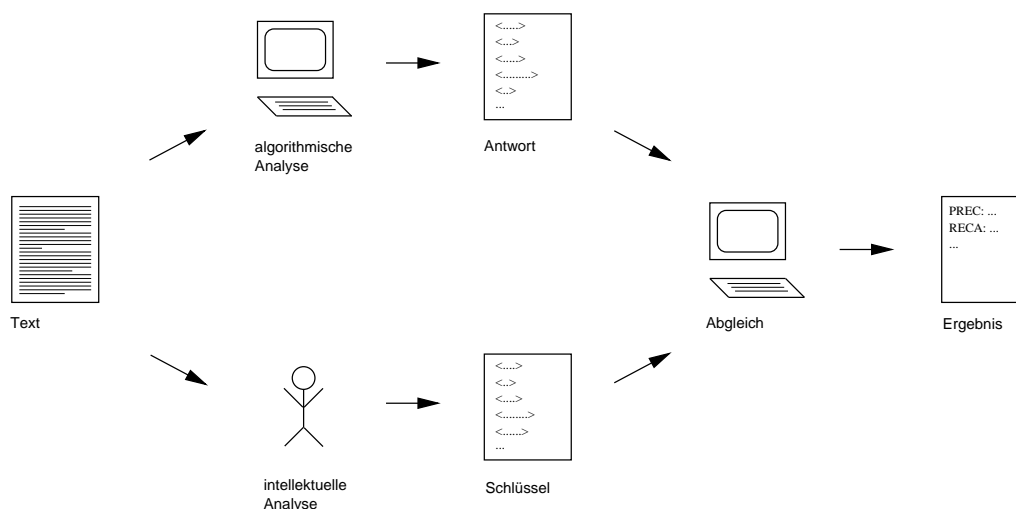


Abbildung 11.1: Szenario der Evaluation von Systemen der algorithmischen Inhaltserschließung

1. möglichst präzise Definition relevanter *Teildisziplinen* der Inhaltserschließung,
2. Definition adäquater formaler *Evaluationsmaße*,
3. Festlegung geeigneter, anwendungstypischer *Textkorpora*,
4. Konstruktion eines *Referenzkorpus* per intellektueller Annotation,
5. Durchführung der *Evaluation*.

Der unter Punkt 1 genannten Zielsetzung kommt grundlegende Bedeutung zu. Zunächst einmal sind geeignete Inhaltserschließungs-Aufgaben als *Teildisziplinen* der Evaluation festzulegen, deren Betrachtung einen Rückschluß auf die Brauchbarkeit der Analyse-Ergebnisse auch unter Anwendungsgesichtspunkten erlaubt. Darüberhinaus sollten *möglichst feinkörnige Definitionen* für jede der Einzelaufgaben erarbeitet werden, denn nur, falls es gelingt, die betrachteten Teildisziplinen der Inhaltserschließung in einer für menschliche Textinterpreten hinreichend eindeutigen Weise zu explizieren, macht es überhaupt Sinn, die per intellektueller Analyse erzeugbaren Referenzdaten als Maßstab für die Leistungsfähigkeit algorithmischer Ansätze zugrundezulegen. Von ähnlicher Wichtigkeit ist die *formale Definition von Evaluationsmaßen* (Punkt 2), denn erst auf dieser Grundlage sind die Voraussetzungen für eine graduelle Bewertung nebst Vergleich der Analysegröße von Textanalyse-Systemen gegeben. Für die im Rahmen dieser Arbeit betrachtete Inhaltserschließungs-Aufgabe der referentiellen Interpretation werden sich die genannten Teilprobleme als durchweg nichttrivial erweisen.

Darüberhinaus sind zwei weitere Vorbedingungen zu erfüllen, die mit vergleichsweise hohem Aufwand verbunden sind. Zum einen sind, basierend auf der Auswahl anwendungstypischer Textkorpora, *Referenzdaten* für die einzelnen Inhaltserschließungsaufgaben zu erstellen (Punkt 4), die als Vergleichsmaßstab in die eigentliche Evaluation (Punkt 5) einfließen. Diese Aufgabe ist per intellektueller Annotation von Exemplaren der gewählten Textkorpora mit der relevanten Information zu bewerkstelligen. Zum anderen ist der zu evaluierende Algorithmus zu *implementieren*, wobei die noch bestehenden Explikationslücken zu schließen sind (vgl. Abschnitt 9.5.1) sowie

eine Adaption an die spezifischen Gegebenheiten tatsächlich verfügbarer Voranalyse-Systeme (morphologische und syntaktische Analysatoren) vorzunehmen ist. Die Aufgabe der programmiertechnischen Umsetzung umfaßt ferner die Realisierung eines Softwaresystems zum automatischen Abgleich der Ergebnisse der algorithmischen und der intellektuellen Inhaltserschließung sowie die Berechnung der Werte für die einzelnen Evaluationsmaße.

Die identifizierten Aufgaben sollen in drei Schritten gelöst werden: Eingrenzung des Evaluationsszenarios (Abschnitt 11.1), Erarbeitung einer Implementierung des in Abschnitt 10.3 entwickelten Algorithmus (Abschnitt 11.2) und Durchführung der Evaluation (Abschnitt 11.3).

11.1 Eingrenzung des Evaluationsszenarios

Zielsetzung der durchzuführenden Evaluation ist die Bewertung der Leistungsfähigkeit von Software-Systemen, die Algorithmen zur referentiellen Interpretation implementieren. Somit ist für dieses spezifische Problem der algorithmischen Inhaltserschließung zu explizieren, wie das in Abbildung 11.1 skizzierte generische Evaluationsszenario auszugestalten ist. Als Ausgangspunkt sollen daher zunächst *drei Teildisziplinen* identifiziert werden, die im Rahmen einer Evaluation von Systemen der referentiellen Interpretation unterschieden werden sollten. Des Weiteren sind geeignete Textkorpora festzulegen.

Bislang wurde eine vergleichsweise abstrakte Eingrenzung des Untersuchungsgegenstands zugrundegelegt, derzufolge Relationen der *Ko*-Spezifikation zwischen sprachlichen Ausdrücken aufzudecken sind, die sich auf *Objekte* der projizierten Welt beziehen (vgl. Abschnitt 6.4). Eine genauere Analyse zeigt auf, daß sich diese Definition im Hinblick auf die Vielfältigkeit des sprachlichen Phänomens der Referenz als hochgradig unterspezifiziert erweist. Ein weiterer zentraler Punkt besteht daher in der Präzisierung der inhaltlichen Vorgaben für die Teildisziplinen.

11.1.1 Teildisziplinen der Evaluation

Auf den ersten Blick erscheint die Definition eines geeigneten Evaluationskriteriums elementar:

Ermittle die Prozentzahl der für die anaphorischen Vorkommen mit Objektbezug getroffenen korrekten Antezedens-Entscheidungen!

Eine nähere Betrachtung legt jedoch offen, daß diese erste Spezifikation einer Vorgehensweise nicht hinreichend präzise ist. Folgendes Beispiel illustriert die wesentlichen Probleme:

- (1) *Gropius besuchte Behrens, mit dem er die Entwürfe besprechen wollte.
Dem Leiter des Bauhauses war es klar, daß er noch viel zu tun bekommen würde.*

Eine erste Schwierigkeit besteht darin, daß der korrekte Vorgänger i.d.R. *nicht eindeutig bestimmt* ist. Für das zweite pronominale Vorkommen “*er*” existieren etwa die Antezedens-Optionen “*Leiter des Bauhauses*”, “*er*” und “*Gropius*”, die allesamt Okkurrenzen ein und desselben Diskursreferenten, d.h. Mitglieder einer bestimmten Äquivalenzklasse der Koreferenzrelation sind. Entsprechend den Festlegungen in Abschnitt 6.2 besteht die Aufgabenstellung der

referentiellen Interpretation ja nicht in der Aufdeckung von Wiederaufgriffsbeziehungen auf der Ebene der sprachlichen Oberfläche; vielmehr geht es um die hierdurch zum Ausdruck gebrachten Kospezifikations-Relationen und damit letztlich um das zugrundeliegende semantische Phänomen der Koreferenz. Folglich sind *alle* genannten Vorkommen als korrekte Antezedenten der betrachteten Anapher “*er*” anzusehen. In der *originären Evaluationsdisziplin* für Systeme der Koreferenz-Resolution ist somit auf *Äquivalenzklassen* von Vorkommen Bezug zu nehmen.

Ein zweites Problem entsteht durch die im Rahmen des Bewertungskriteriums implizit getroffene Annahme, daß die spezifizierenden Vorkommen bereits bekannt sind. Wie aus den Untersuchungen in Abschnitt 6.1.2 jedoch hervorgeht, ist diese Aufgabe in nicht allen Fällen algorithmisch trivial. Ein Beispiel hierfür verkörpert die nichtreferentielle, *expletive* Verwendungsweise des Vorkommens “*es*” in Satz 2 des obigen Beispiels, deren Erkennung u.a. davon abhängt, ob der Parser die Information zur Verfügung stellt, daß es sich um ein formales Subjekt handelt. Selbst unter der Annahme, daß diese Voraussetzung erfüllt ist, ergeben sich weitere Schwierigkeiten alleine dadurch, daß die syntaktische Analyse der typischen Vorkommens-Kontributoren - Nominalphrasen - nicht immer korrekt bzw. vollständig vonstatten gehen wird. Es stellt sich somit die Frage, wie die Fehler auf der Ebene der Vorkommensermittlung im Rahmen der Evaluation zu erfassen sind. Als Grundlage für eine präzisere Lokalisierung bestehender Mängel der Algorithmen zur referentiellen Interpretation bietet es sich an, das Problem der *Bestimmung spezifizierender Vorkommen als eigenständige Evaluationsdisziplin* zu betrachten.

Wenn somit die primäre Leistung von Algorithmen zur referentiellen Interpretation unter Rekurs auf *Äquivalenzklassen* von Vorkommen gemessen wird, dann bleibt ein wichtiges Problem offen. Aus der Perspektive einer Anwendung im Rahmen der klassischen, *lexikalischen* Computergestützten Inhaltsanalyse besteht die eigentliche Aufgabe darin, *nichtpronominaler Substitute* für pronominal-anaphorische Vorkommen zu ermitteln. Die Güte, mit der dies bewerkstellbar ist, wird in den bisher definierten Evaluationsdisziplinen nicht ermittelt. Die originäre, äquivalenzklassenbezogene Evaluation würde in Beispiel (1) die Wahl des Pronomens “*er*” (Satz 1) als Antezedens der Pronominalanapher “*er*” (Satz 2) als korrekt bewerten. Der Performanzwert, der für das Problem der Ermittlung eines *beliebigen* kospezifizierenden, zur selben Äquivalenzklasse gehörigen Antezedens erzielt wird, kann jedoch nicht einfach auf das Problem der Ermittlung nichtpronominaler, *konzeptueller* Anker für Pronomen übertragen werden. Vielmehr ergeben sich aus theoretischer Sicht Anhaltspunkte dafür, daß letztere Aufgabe schwieriger ist. Entsprechend den Vorhersagen der in den Abschnitten 7.1.4 und 7.2.1 beschriebenen Theorien der lokalen Fokussierung beziehen sich Pronominalanaphern tendenziell auf Diskursreferenten, die in der jeweiligen Diskursrepräsentation *fokussiert* sind, da in pronominalen Ausdrücken i.d.R. zu wenig Information enthalten ist, um ein Objekt außerhalb des Fokus eindeutig zu spezifizieren. Unter der Prämisse kohärenten Diskurses ist ferner davon auszugehen, daß der Fokus des Diskurses nicht zu häufig wechselt. Folglich steht zu erwarten, daß die Entscheidung für ein pronominales, alle Restriktionen erfüllendes Antezedens mit höherer Wahrscheinlichkeit korrekt ist als die Entscheidung für eine nichtpronominalen Vorgänger.² Als dritte, durch die Anforderungen potentieller Anwendungen motivierte Evaluationsdisziplin ist daher das Problem der *Ermittlung lexikalischer Substitute* zu betrachten.

Somit sollen drei Teildisziplinen der Evaluation von Systemen zur Kospezifikations-Interpretation betrachtet werden:

1. algorithmische Bestimmung objektspezifizierender Vorkommen (*OV-Aufgabe*),

²Die Ergebnisse der Evaluation werden diese Einschätzung bestätigen.

2. algorithmische Bestimmung der Äquivalenzklassen kospezifizierender Vorkommen (*KV-Aufgabe*),
3. algorithmische Bestimmung lexikalischer Substitute für pronominale Vorkommen (*PS-Aufgabe*).

Die Definition der entsprechenden formalen Evaluationsmaße wird erst in Abschnitt 11.1.4 erfolgen, da sie auf zusätzlichen inhaltlichen Vorgaben aufbaut (Abschnitt 11.1.3).

11.1.2 Wahl geeigneter Textkorpora

Im Hinblick auf die Zielsetzungen der Evaluation ist ein Textkorpus zu wählen, das mehreren grundsätzlichen Anforderungen genügt. Erstens sollte das Textgenre³ von grundsätzlicher *Anwendungsrelevanz* sein, d.h. es sind Textkorpora zu bevorzugen, die hinsichtlich der Relevanz ihres Inhalts sowie der typischerweise anfallenden Menge erste Kandidaten für den Einsatz von Software-Werkzeugen der computergestützten (nicht notwendig sozialwissenschaftlichen) Inhaltsanalyse darstellen. Zweitens sollte die Textsammlung eine gewisse *Bandbreite* hinsichtlich der spezifischen inhaltlichen Domänen aufweisen, da es um die Evaluation von Systemen geht, denen keine domänenspezifische Wissensbasis zugrundeliegt und die auf ihre *allgemeine* Tauglichkeit hin untersucht werden sollen. Drittens sollten die Texte ein Mindestmaß an verarbeitungstechnischer *Komplexität* aufweisen, die sie als ideale Kandidaten für die Evaluation *robuster* Verfahren der algorithmischen Textinterpretation erscheinen läßt.

Textkorpus “*Presse-Meldungen*”

Alle drei genannten Bedingungen werden durch *Presse-Meldungen von Nachrichtenagenturen* erfüllt. Täglich wird eine beinahe unüberschaubare Menge derartiger Textdokumente produziert, die die geforderte große thematische Bandbreite aufweisen und deren automatische, computergestützte Verarbeitung überdies für die unterschiedlichsten Anwendungsgebiete potentiell relevant ist. Auch der Anforderung an die verarbeitungstechnische Komplexität scheint Genüge getan, da Nachrichtenagentur-Meldungen (wie sich noch zeigen wird) die typischen Eigenschaften von Texten aufweisen, die unter Zeitdruck und mit allenfalls rudimentären Anforderungen an die sprachliche Realisierung der Inhalte verfaßt werden. Orthographische und grammatische Fehler sind vergleichsweise häufig; Einschübe wie etwa Zitate in direkter Rede oder nachgestellte Erklärungen in Klammern kommen ebenso vor wie sonstige (unspezifische) Abfolgen von Ausdrücken, die keine vollständigen Sätze konstituieren. Die potentiellen Defizite der zu analysierenden Texte stellen somit einen weiteren wichtigen Anlaß für die Forderung einer robusten Verarbeitung dar, die bislang v.a. im Hinblick auf den i.a. partiellen Abdeckungsgrad der verfügbaren Verarbeitungsressourcen gestellt wurde.

Die Evaluation soll daher auf der Basis eines Korpus von Presse-Meldungen erfolgen. Aus

³Entsprechend der allgemein üblichen Praxis soll unterschieden werden zwischen dem *Genre* eines Textes und dessen spezieller inhaltlicher *Domäne*, also etwa *Zeitungsnachricht* auf der allgemeinen und *Sportmeldung* auf der spezifischen Ebene.

Wörter ⁴	Sätze	Texte	PE-3	PE12	PO-3	PO12	REFL	RELA	PRON
24712	1093	66	406	97	246	17	13	172	$\Sigma = 951$

Wörter/Satz	Wörter/Text	Sätze/Text	Pronomen/Satz	Pronomen/Text
22.61	374.42	16.56	0.87	14.41

Abbildung 11.2: Quantitative Charakteristika des Textkorpus “*Presse-Meldungen*”

Verfügbarkeitsgründen wird vornehmlich auf Berichte der Themengebiete Innenpolitik, Außenpolitik, nationale und internationale Krisen und Konflikte, Wirtschaftspolitik und sonstige Wirtschaftsmeldungen zurückgegriffen. Die hohe Anwendungsrelevanz dieses Teil-Genres liegt auf der Hand; auch die inhaltliche Bandbreite erscheint aufgrund der Heterogenität der Texte in hinreichendem Maße gewährleistet. Aus Gründen der Parser-Verfügbarkeit (vgl. u.) wird auf *Texte in englischer Sprache* zurückgegriffen.

Das der Evaluation zugrundegelegte Korpus entstammt einer Sammlung englischsprachiger Presse-Meldungen internationaler Nachrichtenagenturen. Die quantitativen Daten dieser Textsammlung sind in Abbildung 11.2 zusammengefaßt. Der linke Teil der oberen Tabelle beschreibt die allgemeinen Eigenschaften der Dokumentensammlung; im rechten Teil wird die Anzahl der Pronomen je Typ aufgeschlüsselt: PE-3/PE12 = Personalpronomen in dritter bzw. erster/zweiter Person, PO-3/PO12 = Possessivpronomen in dritter bzw. erster/zweiter Person, REFL = Reflexiv- und Rezipropronomen, RELA = Relativpronomen, PRON = Gesamtanzahl objektspezifischer Pronomen.⁵ Die untere Tabelle zeigt einige relevante Durchschnittswerte. Mit insgesamt 951 objektspezifischen Pronomen und durchschnittlich 0.87 Pronomen je Satz ist eine geeignete Grundlage für die Evaluation der Systemperformanz in der Ermittlung lexikalischer Substitute gegeben. Die durchschnittliche Anzahl von Wörtern je Satz ist mit 22.61 vergleichsweise groß; dies illustriert die verarbeitungstechnische Komplexität des Textkorpus, das somit als idealer Prüfstein für die Evaluation robuster Analyseverfahren der referentiellen Interpretation anzusehen ist.

Der inhaltliche Teil der Nachrichtenagentur-Texte wird der Evaluation *in unveränderter Form* zugrundegelegt: Auf die Korrektur orthographischer, grammatischer oder sonstiger Fehler wird verzichtet, da die Evaluation unter *Anwendungsbedingungen* ablaufen soll. Ggf. a priori ediert bzw. entfernt werden hingegen bestimmte Dokument-Komponenten wie Über- und Unterschriften sowie sonstige Hinzufügungen ohne unmittelbaren Inhaltsbezug wie etwa Autoren- und Copyrightvermerke und interne Kategorisierungsmarken. Dies erscheint gerechtfertigt, da diese Aufgaben der Vorverarbeitung zumindest auf der Basis verbindlicher Dokumentstruktur-Definitionen von konventionellen Softwarewerkzeugen ohne Textinhaltserschließung gelöst werden können und somit nicht in den Skopus der vorliegenden Arbeit fallen. Die im Hinblick auf die Zielsetzungen der Anwendungstauglichkeit wichtige Algorithmisierbarkeit der hier per Hand durchgeführten Vorarbeiten auf dem gegenwärtigen Stand der Technik ist gewährleistet.

Das Textkorpus “*Presse-Meldungen*” bildet die Grundlage sowohl der Feinabstimmung als auch der formalen Evaluation des betrachteten Softwaresystems zur referentiellen Interpretation. Um zu aussagefähigen Ergebnissen zu gelangen, wird die Textsammlung in zwei Teile mit in etwa der gleichen Anzahl von Pronomen unterteilt. Die Dokumente 1 bis 31 bilden das *Trainings-*

⁴Die Zählung erfolgte mit dem UNIX-Werkzeug *wc* (*word count*). Diese Zahl ist geringer als die Anzahl sog. *Token* (parsing-relevante Wörter), die durch den Präprozessor des verwendeten Parsers generiert werden (29566), da u.a. *wc* nur solche Satzzeichen zählt, die von den benachbarten Wörtern durch Leerzeichen getrennt sind.

⁵Pronomen mit relationalem Bezug sind nicht evaluationsrelevant und werden somit - da sie keine relevanten Vorkommen verkörpern - nicht mitgezählt (vgl. Abschnitt 11.1.3).

Wörter	Sätze	Texte	PE-3	PE12	PO-3	PO12	REFL	RELA	PRON
2522	131	3	121	0	87	0	7	21	$\Sigma = 236$

Wörter/Satz	Wörter/Text	Sätze/Text	Pronomen/Satz	Pronomen/Text
19.25	840.67	43.67	1.80	78.67

Abbildung 11.3: Quantitative Charakteristika des Textkorpus “*Mozart-Opern*”

korpus (471 Pronomen); die Dokumente 32-66 konstituieren das eigentliche *Evaluationskorpus* (479 Pronomen). Durch die Trennung in Trainings- und Evaluationsdaten wird gewährleistet, daß die ermittelten Werte nicht alleine die (prinzipiell beliebig optimierbare) Interpretationsleistung des Systems auf einem *spezifischen* Set von Texten widerspiegeln, sondern die tatsächliche Leistungsfähigkeit der Algorithmen in bezug auf das Textgenre der Presse-Meldungen charakterisieren.

Textkorpus “*Mozart-Opern*”

Für den Nachweis der prinzipiellen Domänen-Unabhängigkeit der Interpretationsleistungen des zu evaluierenden Systems bietet sich eine Gegenprobe auf Texten an, die sowohl einem anderen Genre als auch einer anderen Domäne angehören. Als Grundlage sollen englischsprachige Inhaltsbeschreibungen von drei Mozart-Opern dienen.⁶ Diese Textbasis erweist sich als besonders geeignet, da sie sich unter sowohl inhaltlichen als auch quantitativen Gesichtspunkten grundlegend vom Korpus der Presse-Meldungen unterscheidet.

Abbildung 11.3 veranschaulicht, daß die Texte eine wesentlich höhere Dichte von durchschnittlich 1.8 Pronomen je Satz aufweisen. Dies ist unmittelbar auf die Domäne der Beschreibung von Handlungsabläufen zurückzuführen, in die i.d.R. mehrere Personen verwickelt sind. Somit steht zu erwarten, daß die referentielle Interpretation schwieriger wird, da der Diskurs an einer bestimmten Stelle meist mehr als eine inhaltliche Entität fokussiert. Andererseits wird die Aufgabe dadurch erleichtert, daß die fokussierten Elemente i.d.R. ein natürliches Geschlecht besitzen, das sich in einer entsprechenden Differenzierung des Genus der Pronominalanaphern niederschlägt.

11.1.3 Präzisierung der inhaltlichen und formalen Vorgaben

Für die nunmehr festgelegte Textsammlung verbleibt ein Referenzkorpus zu konstruieren, das den Bezugspunkt für die Bewertung der Systemleistungen in den Teildisziplinen der Evaluation bildet. Dies kann nur auf der Grundlage einer präziseren inhaltlichen Eingrenzung der zu bewerkstellenden Aufgaben geschehen. Es soll daher eine nähere Charakterisierung der zentralen Begriffe *Vorkommen* bzw. *Ko-Spezifikation* erfolgen. Das Ziel besteht darin, die Freiheitsgrade der bisherigen Definitionen unter Wahrung von inhaltlicher Adäquatheit und Reichweite möglichst weitgehend zu beseitigen. Denn nur falls es gelingt, die inhaltlichen Vorgaben kommunikabel zu gestalten, ist die im Hinblick auf die Aussagefähigkeit der Evaluationsergebnisse

⁶Die Texte wurden den Anlagen der drei CDs “*Don Giovanni*” (Philips: 438 494-2; Textautor: Lionel Salter), “*Le Nozze di Figaro*” (Philips: 416 870-2; Wolfgang Glaser) und “*Così fan tutte*” (Deutsche Grammophon GmbH: 427 680-2; Karl Dietrich Gräwe) entnommen.

zu fordernde Intercoder-Reliabilität, d.h. die hinreichend gute Reproduzierbarkeit der per intellektueller Annotation zu erzeugenden Referenzdaten, gewährleistet. Die verfeinerte Definition fließt natürlich auch als Zielvorgabe in die Entwicklung der Software-Systeme ein.

Ausgangspunkt ist die in Abschnitt 6.4 beschriebene Eingrenzung des Untersuchungsgegenstands, derzufolge auf die detaillierteren Vorgaben der “*Coreference Task Definition*” (CTD) der “*Message Understanding*”-Konferenzen zurückgegriffen werden soll ([CoTD97]). Die grundlegenden Annahmen dieser Aufgabenbeschreibung decken sich mit den Rahmenbedingungen der Evaluation in der vorliegenden Arbeit - Beschränkung auf Instanzen von *Ko*-spezifikation, Verzicht auf die Behandlung relationaler Bezugnahmen. Obwohl die auf der Basis der MUC-Definitionen erstellten Referenzdaten den zu stellenden Anforderungen an die intersubjektive Reproduzierbarkeit nicht in vollem Umfang genügen (vgl. Abschnitt 6.4), soll die CTD dennoch als Anknüpfungspunkt dienen, da dies aus u.a. folgenden Gründen gerechtfertigt erscheint:

- Für die in bezug auf Anwendungszielsetzungen besonders relevanten Personalpronomen ist der Grad an Übereinstimmung wesentlich höher (vgl. Sundheim: [Sund96], S. 20);
- die Werte sind bereits hoch genug, so daß die Ergebnisse zumindest in eingeschränktem Umfang Aussagekraft besitzen;
- die Spezifikation der als relevant anzusehenden Kospezifikations-Relationen ist bereits relativ elementar ausgestaltet und kann ohne wesentliche Abstriche bezüglich der inhaltlichen Reichweite nicht noch weiter vereinfacht werden;
- es gibt bislang keine bessere Alternative, d.h. es handelt sich um einen *De-facto-Standard*, dessen Zugrundelegung u.a. die Vergleichbarkeit der Evaluations-Ergebnisse gewährleistet.

Die CTD bezieht sich auf die Gegebenheiten der englischen Sprache und ist somit auf die durchzuführende Evaluation unmittelbar anwendbar (vgl. Abschnitt 11.1.2).

Im folgenden wird eine Übersicht zentraler Vorgaben der CTD bezüglich der zu berücksichtigenden Vorkommen bzw. Kospezifikations-Relationen gegeben. Darüberhinaus sollen die wesentlichen Fälle genannt und begründet werden, in denen von den CTD-Spezifikationen abgewichen wird.

Eingrenzung der relevanten Vorkommen

Die CTD definiert die zu berücksichtigenden Vorkommen als sog. “*markables*”. Entsprechend der Zielsetzung der Evaluation von Systemen zur Interpretation objekt-spezifischer sprachlicher Ausdrücke finden ausschließlich sprachliche Ausdrücke Berücksichtigung, die NPs bzw. Nomen (lexikalischen Kategorien N) entsprechen. Dies beinhaltet, wie oben dargelegt, die zentralen Typen anaphorischer Vorkommen wie Possessiv-, Reflexiv- und sonstige Personalpronomen erster, zweiter und dritter Person sowie ferner definite nichtpronominale NP, Namen, Datumsangaben, Geldbeträge und sonstige numerische Ausdrücke. An der sprachlichen Oberfläche leere Ausdrücke (Ellipsen) finden keine Berücksichtigung; ferner als irrelevant angesehen werden “*wh*”-NP in interrogativer Funktion wie beispielsweise “*what*”.

In Abweichung von den Vorgaben der CTD sollen *Relativpronomen als Vorkommens-Kontributoren* angesehen werden. Aus der CTD geht nicht hervor, *warum* die Kospezifikationen von Relativpronomen mit der jeweiligen syntaktischen Bezugs-NP als irrelevant angesehen werden. Aus

der Perspektive der Zielsetzung, inhaltsnähere Repräsentationen des Textinhaltes als Ausgangspunkt für spezifische Anwendungen zu schaffen, erscheint es jedoch ungerechtfertigt, Vorkommen von Relativpronomen anders zu behandeln als Vorkommen sonstiger Personalpronomen. In der durchzuführenden Evaluation sollen daher auch die Kospezifikations-Relationen von Relativpronomen einbezogen werden. Da die Antezedenten der Relativpronomen jedoch in vielen Fällen durch die verfügbare oberflächenstrukturelle Information eindeutig bestimmt sind und deren Interpretation somit erwartet einfacher ist, wird eine feinere Aufschlüsselung der Evaluationsergebnisse nach Vorkommens-Typen vorgenommen werden (s.u.).

Die Differenzierung der CTD zwischen “*Present Participles*” und “*Gerunds*” soll hingegen übernommen werden, da eine im allgemeinen nachvollziehbare Unterscheidung zwischen relevanter Objekt- und irrelevanter Relationen-Spezifikation zugrundeliegt (Beispiele nach [CoTD97], S. 6):

- (2a) *The slowing of the economy is supported by some Fed officials; it is repudiated ...*
 (2b) *Slowing the economy is supported by some Fed officials; it is repudiated ...*

Demnach verkörpert “*slowing*” in Beleg (2a) als substantivisch gebrauchtes Partizip ein (abstraktes) *Objekt* und ist somit Vorkommens-Stifter, wohingegen “*Slowing*” in Beleg (2b) als Verbalsubstantiv (Gerundium) einen *Handlungsvorgang* bezeichnet und damit keine evaluationsrelevante Entität beisteuert. Die inhaltliche Nähe der beiden unterschiedlich zu behandelnden Belege verdeutlicht, daß diese Vorgabe der CTD nicht unproblematisch ist.⁷

Eine weitere Fallklasse, für die von der CTD abgewichen werden soll, wird durch sog. “*prenominal modifiers*” konstituiert, die eine in der englischen Sprache zentrale Form der Nominalkomposita-Bildung verkörpern. Entsprechend der CTD-Spezifikation sind in NPs, die aus mehreren Nomen zusammengesetzt sind, auch diejenigen Nomen als Vorkommen zu erfassen, die nicht Kopf der NP sind, für die jedoch im Text ein weiteres ko-spezifizierendes Vorkommen existiert, das als Kopf einer NP bzw. als Name realisiert ist. Die CTD nennt folgendes Beispiel (ibid., S. 7):

- (3) *He was accused of money laundering and drug trafficking.
 However, the trade in drugs ...*

in dem der *prenominal modifier* “*drug*” in “*drug trafficking*” zu markieren ist, da eine weitere, als kospezifizierend aufzufassende Kopf-Okkurrenz existiert. Diese Vorgabe erscheint jedoch in mehrfacher Hinsicht als problematisch. Erstens werden zwei Aufgabenstellungen - Vorkommensidentifikation und referentielle Interpretation - miteinander vermengt, die sowohl unter Evaluationsgesichtspunkten (vgl. o., Abschnitt 11.1.1) als auch aus inhaltlichen Erwägungen auseinandergelassen werden sollten - die Einstufung eines sprachlichen Ausdrucks als *referentiell* sollte nicht davon abhängig gemacht werden, ob weitere Spezifikationen desselben Diskursreferenten existieren. Zweitens sprechen technische Gründe dagegen, da nicht festgelegt ist, wie groß der Abstand zu der entsprechenden Kopf-Realisierung an der textuellen Oberfläche sein darf; da i.a. eine Kospezifikation auf der Konzeptebene, d.h. eine “*Identity Of Sense*”-Beziehung vorliegen dürfte, kann der Wiederaufgriff über viele Sätze hinweg erfolgen, was den Annotationsvorgang erheblich verkompliziert und gegen die Zielsetzung einer guten Reproduzierbarkeit der Evaluationsergebnisse verstößt. Nicht zuletzt ist eine Inflation von Vorkommensbeschreibungen die Folge,

⁷ Als (nichterschöpfende) Kriterien zur Erkennung der zu markierenden “*Present Participles*” werden genannt: Vorhandensein von Modifikatoren (“*excessive spending*”), Artikeln (“*the slowing*”) oder attributive “*of*”-PPs (“*slowing of the economy*”).

da die beschriebene Form der Kompositabildung im Englischen häufig anzutreffen ist. Überdies geht bei einem generellen Verzicht auf die Vorkommens-Markierung pränominaler Modifizierer für die unter Anwendungssichtspunkten zentrale Evaluation der Performanz der referentiellen Interpretation von *Pronomen* nicht viel verloren, da Pronomen i.d.R. *transitiv* - qua Antezedenswahl einer als Kopf realisierten Okkurrenz - mit Vorkommen des nunmehr unberücksichtigten Typs kospezifizieren.

Auch *Konjunktionen von NPs* sollen abweichend von den Spezifikationen der CTD behandelt werden. Während die CTD in Belegen wie

(4) Behrens und Gropius trafen sich zu einer Besprechung.

stets ein Vorkommen für das aus der Summation entstehende komplexe Objekt ansetzt, jedoch Okkurrenzen für die Individuen ausschließlich dann berücksichtigt, wenn entsprechende Kospezifikationen vorliegen, soll im Rahmen der vorliegenden Arbeit genau umgekehrt verfahren werden - nur im Falle der Existenz eines kospezifizierenden anaphorischen Wiederaufgriffs wird der Konjunktionsausdruck als Vorkommens-Kontributor angesehen. Für diese abgewandelte Vorgabe spricht die im Vergleich zu Kospezifikationen mit den Individuen wesentlich geringere Häufigkeit des Wiederaufgriffs des komplexen Objekts;⁸ somit liegt es aus Vereinheitlichungs- und damit Vereinfachungs-Gründen nahe, ohne Weiteres nur die Okkurrenzen der Individuen-NP zu annotieren.

Im Rahmen der intellektuellen Verarbeitung des Evaluationskorpus wurden einige Spezialfälle von NP bzw. Nomen identifiziert, die entsprechend den Vorgaben der CTD offenbar als Vorkommensstifter zu berücksichtigen sind, die jedoch aus inhaltlichen Erwägungen als *nichtreferentiell* behandelt werden sollten. Dies gilt insbesondere für *NP in idiomatischer Funktion* ((5a)), invariable Bestandteile von Anreden, Titeln und Eigennamen ((5b)) sowie bestimmte Verwendungsweisen der Ausdrücke “*it*” ((5c), Korrelat bzw. Expletivum)⁹ bzw. “*there*” ((5d)):

(5a) *He took office. / It took place. / The case came to light.*

(5b) *Your Excellency / the Joint Chiefs of Staff / the lower house of Parliament*

(5c) *more than 4000 have made it to safety / It's common ground now that ...*

(5d) *There will be a private religious ceremony for the families.*

Da zumindest die primäre Funktion dieser Ausdrücke nicht in der Spezifikation konkreter Objekte besteht, sollen sie hier nicht als Vorkommens-Kontributoren betrachtet werden. Die CTD schreibt ferner vor, auch solche Pronomen als Okkurrenzen aufzufassen, die kein Objekt, sondern eine relationale Entität spezifizieren; erst die entsprechende Spezifikationsrelation wird als irrelevant klassifiziert. Auch von dieser Vorgabe wird im Rahmen der vorliegenden Arbeit aus Konsistenzgründen abgewichen: Ausdrücke mit relationalem Bezug sollen *unabhängig von der spezifischen Oberflächenform*, d.h. nicht nur im Falle einer Realisierung durch einen nichtpronominalen Ausdruck, als irrelevant angesehen werden.

Um die Aufgabenstellung der “*Markables*”-Identifikation von der vollständigen syntaktischen Analyse von Nominalphrasen weitestgehend zu entkoppeln, läßt die CTD in den technischen

⁸Dies geht aus den Ergebnissen der intellektuellen Annotation der zugrundegelegten Evaluationskorpora hervor.

⁹vgl. Abschnitt 6.1.2, Belege (11)

Vorgaben für die zu bewerkstelligende Annotation vorkommens-induzierender Ausdrücke einen großen Spielraum. Im Falle herkömmlicher NP ist mindestens der Kopf, maximal die vollständige NP inklusive aller Attribute (Appositionen, Relativsätze, PP) anzugeben; Namen sowie die unterschiedlichen Spielarten numerischer Ausdrücke werden entsprechend den Zielsetzungen der übrigen MUC-Tasks als (weitgehend) atomar angesehen und sind entsprechend dieser Vorgabe *vollständig* auszuzeichnen ([CoTD97], S. 8). Um den Annotationsaufwand zu begrenzen und einen einfacheren Bewertungsalgorithmus zu ermöglichen, wird diese Vorgabe im Rahmen der vorliegenden Arbeit weiter abgespeckt, indem die Evaluation der Vorkommens-Identifikation ausschließlich an der Bestimmung der entsprechenden NP-Köpfe bzw. Einzelnomen der Kategorie N festgemacht wird. Für die im Rahmen der CTD als atomar angesehenen zusammengesetzten Ausdrücke werden einige naheliegende ergänzende Vorgaben definiert, denenzufolge z.B. der zu annotierende Kopf eines zusammengesetzten Personennamens stets der Nachname ist usw. Für die unter der Zielsetzung einer unmittelbaren Verfeinerung der klassischen, lexikalischen Computergestützten Inhaltsanalyse zentrale Identifikation nichtpronominaler Substitute sind ohnehin primär die lexikalischen Köpfe relevant. Darüberhinaus wird sich zeigen, daß der vollständige Ausdruck im Falle von verteilten Nominal-Komposita (z.B. *“rice assistance”*) oder Namen (*“William Jefferson Clinton”*) auf der Grundlage der Ergebnisse des zugrundegelegten Voranalyse-systems ohnehin mit hoher Güte identifiziert werden kann, da eine uneingeschränkte syntaktische Analyse durchgeführt wird, die insbesondere die interne Struktur der NP verlässlich aufschlüsselt. Im Hinblick auf die primären Zielsetzungen der Evaluation erscheint die mit erheblichen technischen Erleichterungen einhergehende ausschließliche Betrachtung von NP-Köpfen somit vertretbar.

Eingrenzung der relevanten Kospezifikations-Relationen

Die CTD-Definition der zu berücksichtigenden Relationen unterliegt einer methodologischen Rahmen-Richtlinie, derzufolge die Reproduzierbarkeit der Annotations-Ergebnisse Vorrang hat vor dem Grad der Abdeckung von Kospezifikations-Phänomenen ([CoTD97], S. 2):

“[...] it is more important to preserve high interannotator agreement than to capture every possible phenomenon that could fall under the heading of ‘coreference’.”

Dennoch sind die Vorgaben relativ weit gefaßt. Entsprechend der in Abschnitt 6.2 beschriebenen Differenzierung zwischen *“Identity of Reference”*- und *“Identity Of Sense”*-Beziehungen umfaßt die CTD-Definition die Kospezifikationen auf der Ebene sowohl der *Instanzen* eines bestimmten Typs als auch des *Typs* bzw. der *Funktion*, d.h. des durch den vorkommens-induzierenden Ausdruck spezifizierten *Konzepts*. Hierdurch kann es zu Konflikten während der Annotation kommen, denn in bestimmten Fällen konkurrieren die beiden Interpretationsebenen - Kospezifikation der Extension vs. Kospezifikation der Intension -, die jedoch in der Annotation aus Konsistenzgründen nicht vermengt werden dürfen. In der CTD wird u.a. folgender Beleg angegeben ([CoTD97], S. 13):

(6) The temperature is 90. ... The temperature is rising.

Auf der *konzeptuellen Ebene* der Temperatur als *Funktion* wäre eine Kospezifikation der beiden unterstrichenen Vorkommen anzusetzen. Wird jedoch, wie in den CTD-Spezifikationen für Prädikatsnominative vorgegeben, die erste Okkurrenz von “*temperature*” auf der *Objekt-Ebene* mit der Wertangabe “90” identifiziert, so ergäbe sich mittelbar ein Konflikt, da zwei referentiell zu unterscheidende Entitäten als kospezifizierend markiert würden - das zweite Vorkommen von “*temperature*” bezieht sich auf einen Vorgang *steigender* Temperatur, der nicht mit der für das erste Vorkommen angesetzten Extension identifiziert werden darf (“*collapsing coreference chains*”, ibd.). Offenbar entsteht der Konflikt durch die in den technischen Vorgaben zur Annotation implizit fixierte Rahmenbedingung, derzufolge jeder Ausdruck den Kopf maximal einer Okkurrenz verkörpert, was im vorliegenden Fall augenscheinlich zu Problemen führt. Die CTD schreibt für derartige Fälle vor, dem *Erhalt der extensionalen Konsistenz* Vorrang einzuräumen vor der Erfassung möglichst vieler Kospezifikations-Relationen der unterschiedlichen Ebenen ([CoTD97], S. 3, Hervorhebung im Original):

“We propose to place highest priority on preserving reasonable semantics for the equivalence classes. This means that two values (or instances) that are clearly distinct should NOT be allowed to merge into an equivalence class, even if this means not being able to mark all of the function/value or type/instance relations we might want to mark.”

In bezug auf Beleg (6) bedeutet dies, daß für das zweite Vorkommen eine unterschiedliche Kospezifikations-Klasse anzusetzen ist.

Im Rahmen der vorliegenden Arbeit sollen diese Vorgaben der CTD unmittelbar übernommen werden. Ebenfalls zugrundegelegt werden die CTD-Spezifikationen bestimmter stereotyper sprachlicher Phänomene, die als Induktoren von Kospezifikations-Beziehungen vordefiniert sind, wie u.a. die auf der syntaktischen Ebene reflektierten Fälle von *Apposition* ((7a)) und *Prädikatsnominativ* ((7b)):

- (7a) *Clinton, the President of the USA. / Julius Caesar, a well-known emperor*
 (7b) *Clinton is the President of the USA. / Clinton is one of the most powerful men.*

Realisierungsvarianten mit vergleichbarer Semantik sind ebenfalls zu berücksichtigen.

- (8a) *Clinton was named to be President of the USA.*
 (8b) *Itamar Franco dropped his role as ambassador to Portugal.*

Beleg (8b) illustriert darüberhinaus, daß die Identitätsbeziehung nicht zeitlich unbegrenzt zu bestehen braucht (Konzept der “*time-dependent identity*”). Keine Ko-Spezifikation ist hingegen in Fällen anzunehmen, in denen nur die *Möglichkeit* einer Identität besteht ((9a)/(9b)/(9c)) bzw. wenn die Identität negiert wird ((9d)):

- (9a) *The President of the USA should be a well-educated person.*
 (9b) *His death has twice been judged a suicide.*
 (9c) *the Presidents of the USA, often well-educated persons*
 (9d) *the President of the USA, not a well-educated person*

Auch anaphorische Beziehungen, die nicht unmittelbar zwischen zwei Ausdrücken bestehen, sondern nur mittelbar auf der Repräsentationsebene der Quantifikations-Struktur¹⁰ identifizierbar sind, werden gemäß CTD als Kospezifikations-Relationen zwischen den entsprechenden Okkurrenzen berücksichtigt:

(10) *Most farmers beat their own donkeys.*

Auch Instanzen von *Metonymie* werden als potentiell relevant angesehen. In folgendem Beispiel ist die Bedingung der Gleichheit auf der extensionalen Ebene zweifelsfrei erfüllt, so daß von Kospezifikation auszugehen ist.

(11) *The White House sent a proposal to the congress.
The Clinton administration intended to ...*

In Beleg (12) hingegen sind die beiden Vorkommen des Ausdrucks “*New York Times*” aufgrund von Metonymie *nicht* zu identifizieren, da ersteres Vorkommen ein *Exemplar* einer Ausgabe der Zeitung, letzteres hingegen das *Verlagshaus*, d.h. das *Unternehmen* spezifiziert.

(12) *I bought the New York Times this morning.
I read that the editor of the New York Times is resigning.*

In der CTD wird das Phänomen der Metonymie als zentrales Problem identifiziert, für das die Erarbeitung verfeinerter Vorgaben noch aussteht.

Als Grundlage für einen adäquaten Umgang mit Zweifelsfällen beliebiger Art stellen die CTD-Spezifikationen ein Beschreibungsmittel zur Verfügung, das die referenzkorpuseitige Kennzeichnung von Kospezifikations-Beziehungen als *optional* erlaubt. Ein anaphorisches Vorkommen, dessen Wiederaufgriffsrelation derart gekennzeichnet ist, wird dann und nur dann in der Evaluation berücksichtigt, wenn das zu bewertende Computersystem der Okkurrenz ein Antezedens zuordnet; bleibt das Vorkommen hingegen unresolviert, schlägt sich dies nicht in einer Verschlechterung der Bewertung nieder (vgl. u.).¹¹

Im Rahmen der Präparation des Referenzkorpus für die durchzuführende Evaluation erwies sich die Möglichkeit der Annotation optionaler Kospezifikation insbesondere für Zweifelsfälle der Metonymie-Interpretation als nützlich. Für folgenden Beleg etwa wurde die Beziehung zwischen den beiden Okkurrenzen von “*United States*” - Regierung vs. geographisches Gebiet - entsprechend gekennzeichnet, da sich in der CTD keine eindeutige Vorgabe findet:

¹⁰vgl. Abschnitt 7.1.3

¹¹Es fällt auf, daß die CTD - obwohl prinzipiell äquivalenzklassen-orientiert - ein *sequentielles Beschreibungsformat* für Kospezifikations-Beziehungen zugrundelegt, in dem Ausdrücke an Ausdrücke und nicht an Kospezifikationsklassen (Diskursreferenten) anknüpfen (vgl. [CoTD97], S. 3 f.). Im Rahmen der vorliegenden Arbeit soll eine sequentielle Notation ausschließlich zur Spezifikation optionaler Relationen herangezogen werden. Dies erscheint adäquat, da sich Zweifelsfälle regelmäßig anhand *spezifischer Erwähnungen* der entsprechenden Diskursreferenten manifestieren.

Es wird sich zeigen, daß eine adäquate Berücksichtigung der als optional gekennzeichneten Kospezifikations-Beziehungen vergleichsweise komplexe Bewertungs-Prozeduren bedingt (vgl. Abschnitt 11.1.4).

- (13) *The United States said on Tuesday it had reached no new agreements with Cuba on immigration policy but would keep open a dialogue with Havana to ensure safe, legal and orderly migration of Cubans to the United States.*

Die Inhalte von eingeklammerten und zitierten Textstellen werden nicht anders behandelt als der umgebende Text. Im Textkorpus der Evaluation finden sich einige Belege direkter Rede mit Pronomen in erster bzw. zweiter Person, die mit Okkurrenzen außerhalb der direkter Rede kospezifizieren und dort sogar lexikalisch (*konzeptuell*) verankert sind:

- (14) *“Thank you for this. I will return to my duties soon.” said Yeltsin.*

In Ergänzung der CTD-Vorgaben werden auch solche Kospezifikationsbeziehungen als annotations-relevant angesehen.

Da die Zielsetzung in einer Evaluation von Textanalyse-Systemen *unter Anwendungs-Bedingungen* besteht, wurden auch Beziehungen annotiert, die durch orthographisch inkorrekte Ausdrücke induziert werden, insofern nur die Beziehung für den Annotator rekonstruierbar ist. Entsprechend wurde für folgenden Beleg auf eine Korrektur - “*their*” anstelle von “*there*” - verzichtet, die Kospezifikationsbeziehung zwischen den unterstrichenen Vorkommen aber dennoch berücksichtigt:

- (15) *Groups are having to realize how careful they will have to be to make sure that there (!) actions do not have a negative impact [...].*

11.1.4 Definition formaler Evaluationsmaße

Ausgehend von der abstrakten Beschreibung der Evaluationsdisziplinen und der präziseren Eingrenzung der formalen und inhaltlichen Vorgaben für Referenzkorpus und Systemleistung verbleiben geeignete formale Evaluationsmaße zu definieren, auf deren Grundlage die numerische Bewertung der Systemperformanz vollzogen werden kann. Für jede der drei Teildisziplinen sind geeignete Instanzen der in Abschnitt 5.4.2 generisch definierten Evaluationsmetriken *Precision* und *Recall* anzugeben. Entsprechend Definition 5.2 ist jeweils zu bewerkstelligen:

- die formale Charakterisierung des *Ergebnisraums* G der möglichen Gesamtergebnisse, dem sowohl der aus dem Referenzkorpus ableitbare Ergebnisschlüssel R^k als auch das durch das zu evaluierende System generierte Ergebnis R^s angehören,
- die formale Beschreibung eines *Zerlegungsverfahrens*, mit Hilfe dessen die Elemente des Ergebnisraums G auf geeignete Weise in *Einzelergebnisse* zerlegt werden können, auf deren Grundlage die Bestimmung des *Grads der Übereinstimmung* zwischen R^k und R^s entsprechend den Definitionen der *Precision*- und *Recall*-Metriken erfolgen kann, indem Größen E und K (Gesamtanzahl bzw. Anzahl korrekter Einzelergebnisse von R^s) sowie A (Gesamtanzahl per definitionem korrekter Einzelergebnisse von R^k) ermittelt werden.

Textwörter	R_{ov}^k	R_{ov}^s	$R_{ov}^s \cap R_{ov}^k$
1 Es		+	
2 war			
3 Behrens	+	+	+
4 ,			
5 der	+		
6 gähnte			
7 .			
$\Rightarrow A_{ov} = 2 \quad E_{ov} = 2 \quad K_{ov} = 1$			

Abbildung 11.4: Berechnung der Evaluationsmaße der OV-Aufgabe

Evaluationsmaße für die OV-Aufgabe

Aus den formalen Vorgaben für die Aufgabenstellung der Identifikation objektspezifizierender Vorkommen geht hervor, daß die zu bestimmenden Entitäten Köpfe von NP bzw. sonstige sprachliche Einzelausdrücke, d.h. in jedem Fall *Wörter* i.S.v. Zeichenketten an bestimmten Textpositionen sind. Somit ist der Raum der möglichen Gesamtergebnisse definierbar als die *Menge* G_{ov} aller Teilmengen von *Wörtern* des jeweils betrachteten Texts. Die Ergebnisse der computergestützten Interpretation und die Ergebnisvorgaben des Referenzkorpus sind Teilmengen von *Wörtern*, d.h. es gilt $R_{ov}^s, R_{ov}^k \in G_{ov}$.

Im vorliegenden Fall ist die Charakterisierung des Zerlegungsverfahrens und die Definition geeigneter Größen E_{ov} , K_{ov} und A_{ov} elementar. Ein Einzelergebnis besteht in einem Einzelwort, das (intellektuell bzw. maschinell) als relevantes Vorkommen angesehen wird. Somit ergibt sich die Gesamtanzahl E_{ov} der systemgenerierten Einzelergebnisse als die Anzahl der ermittelten okkurrenzinduzierenden Wörter, d.h. $E_{ov} := |R_{ov}^s|$. Analog ergibt sich schlüsselseitig die Festlegung $A_{ov} := |R_{ov}^k|$ entsprechend der Anzahl tatsächlicher Vorkommen. Die Anzahl korrekter Systemergebnisse entspricht der Kardinalität der Schnittmenge des System-Gesamtergebnisses und der Ergebnisvorgabe des Schlüssels, m.a.W. $K_{ov} := |R_{ov}^s \cap R_{ov}^k|$.

Folglich gilt

$$P_{ov} := \frac{K_{ov}}{E_{ov}} = \frac{|R_{ov}^s \cap R_{ov}^k|}{|R_{ov}^s|} = \frac{\text{Anzahl algorithmisch ermittelter tatsächlicher Vorkommen}}{\text{Gesamtanzahl algorithmisch ermittelter Vorkommen}}$$

$$R_{ov} := \frac{K_{ov}}{A_{ov}} = \frac{|R_{ov}^s \cap R_{ov}^k|}{|R_{ov}^k|} = \frac{\text{Anzahl algorithmisch ermittelter tatsächlicher Vorkommen}}{\text{Gesamtanzahl tatsächlicher Vorkommen}}$$

Abbildung 11.4 zeigt ein elementares Beispiel, für das sich die Werte $P_{ov} = 0.5$ und $R_{ov} = 0.5$ ergeben. Den intuitiven Erwartungen wird hiermit entsprochen, da 50 % aller systemgenerierten Okkurrenzen korrekt sind und die Schlüsselvorgabe zu 50 % reproduziert wird.

Evaluationsmaße für die KV-Aufgabe

In bezug auf die Teildisziplin der Identifikation der Äquivalenzklassen der Kospezifikationsrelation gestaltet sich die Definition geeigneter Evaluationsmaße erheblich schwieriger. Der Raum G_{kv} der möglichen Gesamtergebnisse läßt sich definieren als die *Menge aller Äquivalenzrelationen* über der Menge der Okkurrenzen des betrachteten Texts. Jede Äquivalenzrelation induziert

eine *Partition* auf der Menge der Vorkommen, d.h. eine Zerlegung in paarweise disjunkte Teilmengen; jede Teilmenge verkörpert eine Äquivalenzklasse der Kospezifikations-Relation, d.h. die in der Teilmenge enthaltenen Vorkommen werden als paarweise kospezifizierend aufgefaßt. Die Betrachtungen sind von vornherein auf die Menge der sowohl im Referenzkorpus spezifizierten als auch systemseitig identifizierten Okkurrenzen beschränkt, um die Aussagefähigkeit der im folgenden definierten Evaluationsmaße zu gewährleisten. Da die durch das zu evaluierende System erzeugte Lösungsrelation R_{kv}^s i.d.R. auch Vorkommen umfaßt, die nicht im Schlüssel spezifiziert sind (*Precision-Fehler* der OV-Aufgabe), ist die systemgenerierte Ergebnispartition a priori zu filtern, indem Okkurrenzen ohne Pendant im Referenzkorpus aus den die Partition konstituierenden Mengen eliminiert werden; Äquivalenzklassen, die ausschließlich aus derartigen Vorkommen bestehen, werden gänzlich entfernt. In entsprechender Weise sind die durch die Schlüsselrelation R_{kv}^k induzierten Äquivalenzklassen um Okkurrenzen zu reduzieren, die ohne Gegenstück im Systemergebnis sind (*Recall-Fehler* der OV-Aufgabe). Durch das beschriebene Verfahren werden OV- und KV-Aufgabe konzeptuell getrennt. Wie sich unter Betrachtung der im folgenden definierten Evaluationsmaße nachweisen läßt, würden sich andernfalls Fehler auf der Ebene der OV-Aufgabe in uneinheitlicher Weise in den Ergebnissen der KV-Aufgaben-Evaluation niederschlagen und somit die Aussagefähigkeit der Ergebnisse potentiell beeinträchtigen.

Es stellt sich die zentrale Frage, wie ein adäquates Zerlegungsverfahren zur Bestimmung von Einzelergebnissen aussieht, auf dessen Basis geeignete Größen E_{kv} , K_{kv} und A_{kv} anhand der Schlüssel- bzw. Ergebnispartitionen R_{kv}^s und R_{kv}^k ermittelt werden können. Die (indirekte) Verankerung an den individuellen Anknüpfungsentscheidungen des Systemoutputs greift zu kurz, da das Ergebnis auf dieser Ebene Redundanzen aufweisen kann, die nicht zu einer Erhöhung des *Recall* bzw. zu einer Verminderung der *Precision* führen sollten.¹² Eine adäquate Lösung erschließt sich somit nur unter Rekurs auf die Beschreibungsebene der *Äquivalenzklassen*, auf deren Grundlage von den Einzelentscheidungen abstrahiert wird. Vilain, Burger, Aberdeen, Connolly und Hirschman skizzieren ein geeignetes Verfahren, das in der Evaluation der *Coreference Task* der MU-Konferenzen zugrundegelegt wird und auch hier zum Einsatz gelangen soll ([Vil+96]). Der Grundgedanke dieses Ansatzes erschließt sich auf der Basis eines Perspektivenwechsels: Wie sehen die *Fehler* aus, die zu einer *Verminderung* der *Precision* bzw. des *Recall* führen? *Precision-Fehler* manifestieren sich offenbar dadurch, daß die Elemente (Vorkommen) einer bestimmten Kospezifikations-Klasse der systemgenerierten Partition R_{kv}^s nicht in ein und derselben Kospezifikations-Klasse der Schlüssel-Vorgabe R_{kv}^k liegen. Somit ergibt sich die *Anzahl* der *Precision-Fehler* einer bestimmten Äquivalenzklasse von R_{kv}^s als die Anzahl unterschiedlicher Schlüsselklassen der Elemente minus 1. Formal gesprochen: Für jede systemgenerierte Kospezifikations-Klasse C^s ergibt sich die Anzahl der *Precision-Fehler* als die Kardinalität der Partition (Anzahl der Teilmengen) minus 1, die auf C^s durch die Äquivalenzrelation R_{kv}^k des Schlüssels induziert wird. D.h. jede systemerzeugte Klasse C^s kann zwischen 0 und $|C^s| - 1$ *Precision-Fehler* beisteuern, wobei die Obergrenze der *Minimalanzahl* sequentieller Anknüpfungsentscheidungen entspricht, die benötigt wird, um eine Klasse der Größe $|C^s|$ zu konstruieren. Jedoch spielen die tatsächlichen Anknüpfungsentscheidungen - und damit eventuelle Redundanzen - keine Rolle, da die Definition entsprechend den Anforderungen ausschließlich in den Kospezifikations-Klassen verankert ist.

Analog läßt sich die Anzahl der *Recall-Fehler* einer Kospezifikations-Klasse C^k der Schlüsselpartition definieren als die Kardinalität der Partition minus 1, die auf C^k vermöge der systemgenerierten Kospezifikations-Relation R_{kv}^s entsteht. Der inhaltliche Hintergrund dieser Festlegung ergibt sich auf symmetrische Weise.

¹²Eine solche Redundanz liegt insbesondere dann vor, wenn eine aus einer systemgenerierten sequentiellen Wiederaufgriffsrelation resultierende Kospezifikationsrelation bereits transitiv aus anderen Entscheidungen folgt.

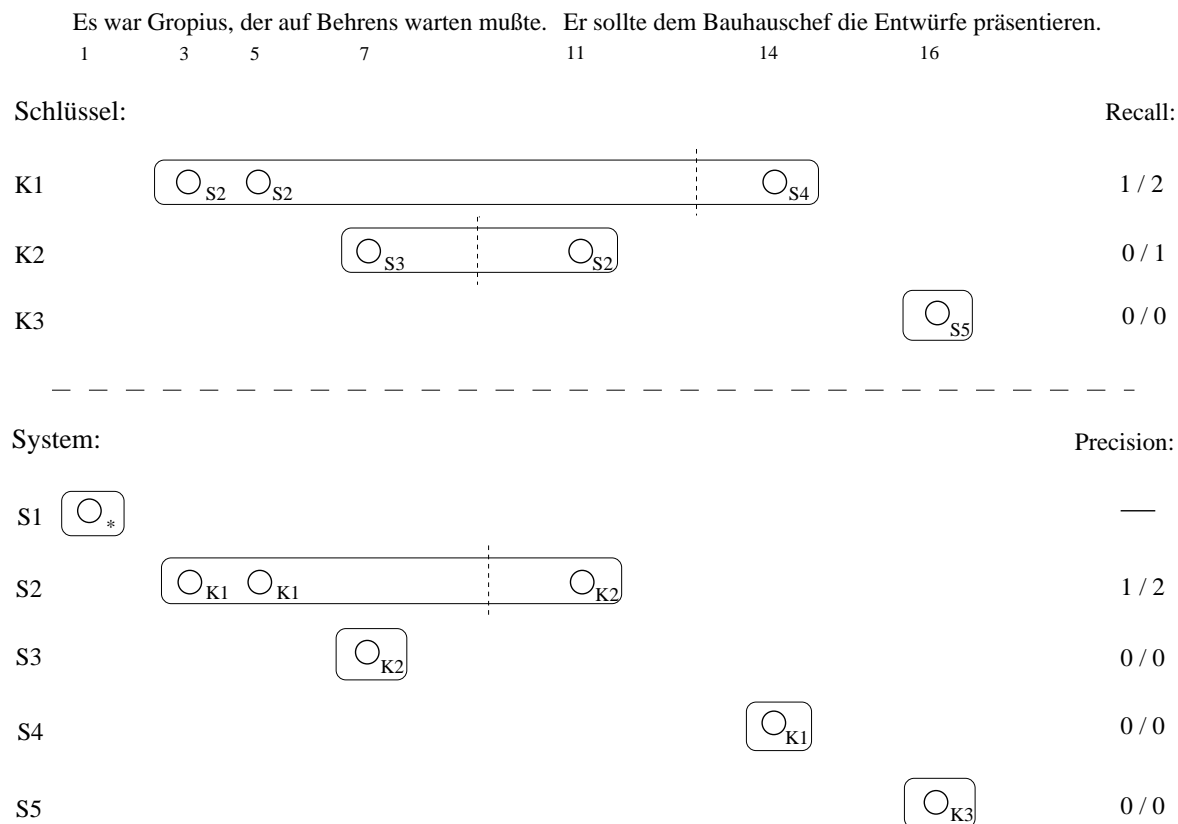


Abbildung 11.5: Berechnung der Evaluationsmaße der KV-Aufgabe

Abbildung 11.5 zeigt ein Beispiel. Für den angegebenen Minitext sind im Schlüssel (oberer Teil der Grafik) drei Koreferenz-Klassen (längliche Kästen) angegeben, die insgesamt sechs objektspezifisierende Okkurrenzen (Kreise) umfassen. (Die Zuordnung zwischen Vorkommen und entsprechenden sprachlichen Ausdrücken ergibt sich spaltenweise.) Klasse K1 entspricht dem Diskursreferenten des Objekts “Gropius”, das insgesamt dreimal an der sprachlichen Oberfläche realisiert ist; K2 repräsentiert den Diskursreferenten der Entität “Behrens” (zwei Erwähnungen); K3 steht für das nur an einer Stelle realisierte Objekt “Entwürfe”.

Es sei nun angenommen, daß ein Sprachverarbeitungssystem zur Kospezifikations-Interpretation die fünf im unteren Teil der Grafik skizzierten Äquivalenzklassen ermittelt hat. Entsprechend der oben motivierten Vorgehensweise wird das inkorrekt als Okkurrenz identifizierte nicht-referentielle, thematische “Es”, das keine Entsprechung im Schlüssel hat, aus dem Ergebnis eliminiert; mit ihm entfällt die nunmehr leere Kospezifikations-Klasse S1. Die Ermittlung der Precision- und Recall-Fehler geschieht nun per wechselseitigem Abgleich der verbleibenden Äquivalenzklassen S2, S3, S4 und S5 von R_{kv}^s bzw. K1, K2 und K3 von R_{kv}^k durch das beschriebene Verfahren der klassenlokalen Partitionierung. Für Klasse K1 ergibt sich ein Recall-Fehler: Da deren Elemente in zwei unterschiedlichen Klassen S2 und S4 (vermerkt durch Subskripte an den Okkurrenzen) der systemgenerierten Äquivalenzrelation R_{kv}^s vorkommen, zerfällt K1 in zwei Teilmengen (gekennzeichnet durch die vertikale Trennlinie); dies findet seine inhaltliche Entsprechung darin, daß im Systemoutput eine entsprechende Einzelentscheidung - Anknüpfung von Bauhauschef an eines der beiden übrigen Elemente - fehlt, die die beiden Diskursreferenten

unifizieren würde. Von den insgesamt notwendigen zwei korrekten Einzelentscheidungen wurde somit nur eine getroffen, womit sich der K1-lokale *Recall* von 0.5 ergibt (rechte Spalte). Auch für K2 liegt ein *Recall*-Fehler vor; da die Klasse aus nur zwei Elementen besteht, ist der K2-lokale *Recall* gleich 0. K3 wurde (trivialerweise) vollständig rekonstruiert entsprechend einem K3-lokalen *Recall* von 1.

In analoger Weise werden die lokalen *Precision*-Werte der Klassen S2, S3, S4 und S5 berechnet als 0.5, 1, 1 bzw. 1. Der einzige *Precision*-Fehler wird durch das falsch interpretierte Personalpronomen "Er" in S2 beigesteuert.

Es verbleiben geeignete formale *Precision*- und *Recall*-Maße P_{kv} bzw. R_{kv} zu definieren, die die Systemleistung insgesamt charakterisieren. Seien $[R_{kv}^s]$ und $[R_{kv}^k]$ die Mengen der Äquivalenzklassen der Äquivalenzrelationen R_{kv}^s bzw. R_{kv}^k , seien ferner $\Phi(C^k, R_{kv}^s)$ und $\Phi(C^s, R_{kv}^k)$ die Äquivalenzrelationen, die sich vermöge Einschränkung der Äquivalenzrelationen R_{kv}^s bzw. R_{kv}^k auf die Elemente der Äquivalenzklassen $C^k \in [R_{kv}^k]$ bzw. $C^s \in [R_{kv}^s]$ ergeben¹³ und $[\Phi(C^k, R_{kv}^s)]$ bzw. $[\Phi(C^s, R_{kv}^k)]$ die Mengen der Äquivalenzklassen der eingeschränkten Relationen. Folgende Definitionen formalisieren das zuvor informell beschriebene Berechnungsverfahren:

$$P_{kv} := \frac{K'_{kv}}{E_{kv}} = \frac{\sum_{C^s \in [R_{kv}^s]} (|C^s| - |[\Phi(C^s, R_{kv}^k)]|)}{\sum_{C^s \in [R_{kv}^s]} (|C^s| - 1)}$$

$$R_{kv} := \frac{K''_{kv}}{A_{kv}} = \frac{\sum_{C^k \in [R_{kv}^k]} (|C^k| - |[\Phi(C^k, R_{kv}^s)]|)}{\sum_{C^k \in [R_{kv}^k]} (|C^k| - 1)}$$

Die Kardinalitäten E_{kv} und A_{kv} der Grundgesamtheit möglicher *Precision*- bzw. *Recall*-Fehler, die der wertungsrelevanten Maximalanzahl potentiell korrekter Entscheidungen entspricht, ergibt sich per Summation der Kardinalitäten der Äquivalenzklassen von Schlüssel bzw. Systemergebnis in den Nennern der Brüche, reduziert jeweils um 1. In den Zählern werden die jeweils tatsächlich gemachten *Precision*- bzw. *Recall*-Fehler abgezogen, deren Bestimmung entsprechend den obigen Ausführungen qua Ermittlung der Kardinalitäten der jeweiligen klassen-lokalen Partitionierungen erfolgt; entsprechend ist zwischen zwei unterschiedlichen Größen K'_{kv} und K''_{kv} - der Anzahl *Precision*-korrekter bzw. *Recall*-korrekter Einzelentscheidungen - zu differenzieren. Somit ergeben sich für das Beispiel in Abbildung 11.5 folgende Werte:

$$P_{kv} := \frac{1 + 0 + 0 + 0}{2 + 0 + 0 + 0} = \frac{1}{2}, \quad R_{kv} := \frac{1 + 0 + 0}{2 + 1 + 0} = \frac{1}{3}$$

Nunmehr wird auch deutlich, warum Okkurrenzen ignoriert werden sollten, die nicht sowohl im Schlüssel als auch im systemgenerierten Ergebnis spezifiziert sind. Würden derartige Vorkommen im Rahmen der Ermittlung der *Precision*- bzw. *Recall*-Fehler mitberücksichtigt, so entstünde der jeweilige negative Beitrag dann und nur dann, wenn die Okkurrenzen in Äquivalenzklassen vorkommen, die weitere Vorkommen enthalten, die sowohl im Schlüssel als auch im Systemoutput enthalten sind - Äquivalenzklassen, die ausschließlich derartige Vorkommen enthalten, werden trivialerweise nicht lokal partitioniert. Es erscheint somit angezeigt, das von Vilain, Burger, Aberdeen, Connolly und Hirschman angegebene Verfahren in der beschriebenen Weise zu verfeinern, um die Aussagefähigkeit des Evaluationsergebnisses zu gewährleisten. Eine weitere Verfeinerung ist bezüglich der evaluationstechnischen Behandlung individueller

¹³Es gilt $\Phi(C^k, R_{kv}^s) = R_{kv}^s \cap (C^k \times C^k)$ und $\Phi(C^s, R_{kv}^k) = R_{kv}^k \cap (C^s \times C^s)$.

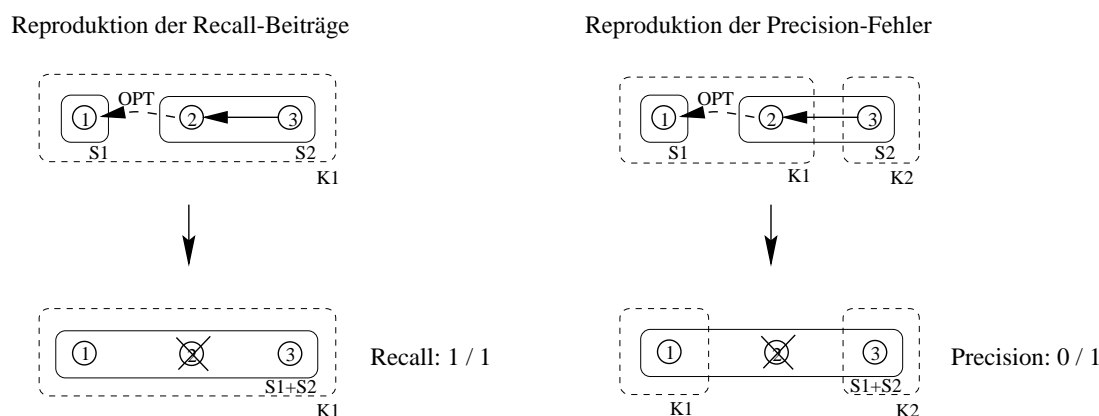


Abbildung 11.6: evaluationstechnische Behandlung optionaler Wiederaufgriffsbeziehungen

Wiederaufgriffs-Beziehungen zu leisten, die im Schlüssel als *optional* markiert sind. Entsprechend den Ausführungen in Abschnitt 11.1.3 wird in diesem Fall von der äquivalenzklassenbezogenen Sichtweise abgerückt, indem der Schlüssel für eine *bestimmte* anaphorische Okkurrenz o_2 ein *bestimmtes* Antezedens o_1 spezifiziert. Diese Angaben werden ausschließlich dann relevant, wenn die Anapher o_2 im Systemergebnis *unresolved* bleibt, d.h. auf der Ebene der Einzelentscheidungen leer ausgeht. Da die im Schlüssel für o_2 spezifizierte Kospezifikations-Relation als optional markiert ist, sollte die Anapher in geeigneter Weise aus der Bewertung herausgenommen werden. Anhand von Abbildung 11.6 wird deutlich, daß eine einfache Entfernung der Okkurrenz o_2 aus den entsprechenden Äquivalenzklassen von Schlüssel (gekennzeichnet durch unterbrochene Linien) und Systemergebnis (durchgezogene Linien) nicht hinreichend ist: Im linken Beispiel entstünde ein *Recall* von 0 anstelle von 1; im rechten Beispiel entstünde eine *Precision* von 1 anstelle von 0. Eine genauere Untersuchung führt zu dem Ergebnis, daß *zwei Schritte* zu erfolgen haben: (1) die *systemergebnis-seitige* Unifikation der Diskursreferenten (Äquivalenzklassen) der durch die im Schlüssel spezifizierte optionale Wiederaufgriffsrelation verbundenen Vorkommen o_1 und o_2 ,¹⁴ sowie (2) die Elimination der Anapher o_2 aus der entsprechenden Äquivalenzklasse sowohl des Schlüssels als auch des Systemergebnisses. Wie die Beispiele aufzeigen, sorgt Schritt 1 dafür, daß der *Recall*-Beiträge bzw. die *Precision*-Fehler des Systemergebnisses in der jeweils entstehenden Vereinigungs-Klasse $K1+K2$ reproduziert werden und die inhaltlich korrekten Ergebnisse 1/1 bzw. 0/1 entstehen.

Evaluationsmaße für die PS-Aufgabe

Die entscheidungsrelevanten Entitäten der PS-Aufgabe sind pronominale Okkurrenzen, für die geeignete lexikalische (konzeptuelle) Substitute zu bestimmen sind. Im Sinne der Evaluation als *geeignet*, d.h. korrekt angesehen werden sollen *beliebige nichtpronominale Vorkommen*, die entsprechend der Schlüssel-Spezifikation *derselben Kospezifikations-Klasse* angehören. Diese Vorgabe ist nicht unproblematisch, denn nicht in jedem Fall ist die Substitution eines Pronomens durch die lexikalische Form einer *beliebigen* kospezifizierenden nichtpronominalen Okkurrenz

¹⁴Es ist durchaus möglich, daß die beteiligten Vorkommen o_1 und o_2 bereits ein und demselben Diskursreferenten angehören; in diesem Fall bleibt die Operation wirkungslos.

inhaltsadäquat. In der Sprachphilosophie werden Beispiele sog. *opaker* bzw. *intensionaler Kontexte*¹⁵ diskutiert, in denen die Substitution extensionsgleicher (referenzidentischer) Ausdrücke mit unterschiedlichem *Sinn* (im Sinne von Frege, [Freg94]) den Wahrheitswert beeinflussen kann. Anhand von folgendem Beleg wird deutlich, daß diese Feinheit auch für die Problemstellung der PS-Aufgabe potentiell relevant ist.

- (16) *Behrens begutachtet die Entwürfe des anonymen Autors.
Behrens glaubt, daß er unbegabt ist.*

Wird für das Pronomen “er” der Ausdruck “*anonymer Autor*” substituiert, so reproduziert dies sicherlich den textuellen Inhalt. Anders könnte es sich verhalten, wenn aus dem weiteren Kontext bekannt ist, daß der unbekannte Autor mit einer Person extensionsgleich ist, deren fachliche Kompetenz für Behrens außer Frage stehen dürfte - etwa der als Leiter des Staatlichen Bauhauses zu Weimar bekannte Walter Gropius. Obwohl eine Kospezifikationsrelation besteht, ist die Substitution dieser nichtpronominalen Realisierungsalternative unter inhaltlichen Gesichtspunkten problematisch ([*]):

- (17) [*] *Behrens glaubt, daß Walter Gropius unbegabt ist.*

Das oben beschriebene Evaluationskriterium greift demnach aus theoretischer Sicht zu kurz, weil es ein *beliebiges* kospezifizierendes Substitut lizensiert, ohne die *Intension* des durch den spezifischen sprachlichen Ausdruck transportierten Begriffs einzubeziehen. Potentielle Induktoren des beschriebenen Problems sind *Kontexte des Wissens und Glaubens* (induziert durch Verben wie “*wissen*” und “*glauben*”) sowie *modale Kontexte* etwa der Form “*Es ist notwendig, daß*”.

Dennoch gibt es überzeugende Gründe für eine Beibehaltung des vereinfachten Evaluationskriteriums. Das primäre Argument ist kommunikationspragmatisch motiviert. Unter Wahl einer geeigneten Auswahlstrategie, durch die an der Oberfläche bzw. in den Fokusstrukturen “nahe” lexikalische Substitute bevorzugt werden, dürften inadäquate Entscheidungen eine seltene Ausnahme bilden. Die Argumentation verläuft auf einer ähnlichen Ebene wie in der Diskussion der unterschiedlichen Reichweiten von grammatischem Genus bzw. ontologischem Sexus in bezug auf die zwischen Pronomen und Antezedens zum Tragen kommende Kongruenzbedingungen:¹⁶ je näher das nichtpronominale Antezedens, desto präsenter der konzeptuelle Inhalt, der durch die lexikalische Form des *spezifischen* Ausdrucks transportiert wird. Auch der menschliche Interpret würde potentiell fehlgeleitet, wenn die Kospezifikationsbeziehung per Aufgriff eines unmittelbar vorausgehenden Antezedens vermittelt wird, dessen Konzept sich bezüglich des intendierten Kommunikationsinhalts jedoch als inkompatibel zum Kontext des Pronomens erweist. Konversationspragmatische Gründe - Stichpunkt: Grice’sche Konversationspostulate ([Gric68]) - sprechen somit für die Irrelevanz des beschriebenen Problems opaker Kontexte unter Verwendung adäquater Auswahl-Algorithmen.

Eine zweiter inhaltlicher Punkt ergibt sich in bezug auf die im Rahmen typischer Anwendungen zu generierenden Wissensstrukturen. Die Substitution findet in aller Regel nicht, wie etwa in Beispiel (17) im Hinblick auf eine sich anschließende *intellektuelle* Interpretation supponiert, im Ausgangstext statt, sondern *in den Datenstrukturen der Anwendung*, in deren Weiterverarbei-

¹⁵Die Bezeichnungen gehen auf Quine ([Quin60]) bzw. Frege ([Freg94]) zurück.

¹⁶vgl. Abschnitt 7.1.1, S. 111

Menge	Klassifikation	Definition
o_{++}	korrekt	P und L gehören im Schlüssel derselben Kospezifikations-Klasse an
o_{+-}	inkorrekt	P und L gehören unterschiedlichen Schlüssel-Klassen an
$o_{+?}$	inkorrekt	P , nicht jedoch L entspricht einem Vorkommen im Schlüssel
o_{+-}	leer	P entspricht einem Vorkommen im Schlüssel, kein Substitut L ermittelt
o_{+*}	leer	P entspricht einem Vorkommen im Schlüssel, kein Substitut L ermittelt, Kospezifikation von P im Schlüssel als optional vorgegeben
$o_{?+}$	inkorrekt	P hat keine Entsprechung im Schlüssel
$o_{?-}$	leer	P hat keine Entsprechung im Schlüssel, kein Substitut L ermittelt

Abbildung 11.7: Klassifikation der Ergebnisse der PS-Aufgabe

tion etwaige Subtilitäten opaker Kontexte ohnehin nicht zum Tragen kommen.¹⁷ Der potentielle Nutzen der lexikalischen Substitution von Pronomen sollte die vergleichsweise selten in Erscheinung tretenden Probleme kontextuell inadäquater Substitute bei weitem übertreffen.

Schließlich sprechen auch methodische Gründe für eine vereinfachte Ausgestaltung der Evaluation. Der zusätzliche Anspruch einer Spezifikation *intensional geeigneter* lexikalischer Substitute in den ohnehin selten anzutreffenden Fällen opaker Kontexte würde die Anforderungen an die Ersteller des Referenzkorpus erhöhen und die ohnehin nicht sehr hohe Intercoder-Reliabilität weiter beeinträchtigen.

Ausgehend von den Grundannahmen des skizzierten vereinfachten Evaluationsszenarios sollen nun geeignete formale Evaluationsmaße für die PS-Aufgabe definiert werden. Der Lösungsraum G_{ps} besteht aus einer Menge von Paaren (P, L) , wobei P ein Pronomen und L eine nichtpronominale Okkurrenz ist. G_{ps} erfüllt ferner die Bedingung, daß für jedes Pronomen P *maximal ein Paar* (P, L) enthalten ist; demnach können auch Pronomen ohne Zuordnung existieren. Das systemgenerierte Ergebnis R_{ps}^s und die Vorgaben des Referenzkorpus R_{ps}^k unterscheiden sich insofern, als letztere - entsprechend der obigen Festlegung - i.a. eine *Menge* von Elementen aus G_{ps} als korrekt spezifizieren, da die Kospezifikationsklassen, auf die schlüsselseitig Bezug genommen wird, i.d.R. Freiraum in bezug auf die Wahl eines korrekten lexikalischen Substituts lassen. Somit gilt $R_{ps}^s \in G_{ps}$ und $R_{ps}^k \subseteq G_{ps}$.

Ausgangspunkt für die Definition adäquater *Precision*- und *Recall*-Maße ist eine Kategorisierung des systemgenerierten Ergebnisses entsprechend den Spezifikationen des Schlüssels (vgl. Abbildung 11.7). Bezugspunkt der Klassifikation sind Okkurrenzen P , die vom System als *pronominale* und somit als Gegenstand der PS-Aufgabe erkannt werden. Die Einordnung in die angegebenen Mengen geschieht in Abhängigkeit davon, ob P im Referenzkorpus als Vorkommen spezifiziert ist, ob eine nichtleere Zuordnung (P, L) vorgenommen wurde sowie (ggf.) ob auch das vorgeschlagene Substitut L Vorkommensstatus im Schlüssel hat. Entsprechend den obigen Festlegungen wird eine Zuordnung (P, L) genau dann als *korrekt* angesehen, wenn P und L im Referenzkorpus als kospezifizierend klassifiziert sind, wodurch impliziert wird, daß sowohl P als auch L als Okkurrenzen im Schlüssel vorgegeben sind. Im Hinblick auf die zu formulierenden Definitionen von *Precision*- und *Recall*-Maßen wird ferner zwischen *inkorrekten* und *leeren* Zuordnungen unterschieden, wobei jeweils feiner danach differenziert wird, ob die beteiligten Vorkommen Pendants im Schlüssel haben oder nicht.

Anhand der Aufschlüsselung wird deutlich, daß auch hier Fehler in der Bestimmung objektspezifizierender Vorkommen, die bereits in den Evaluations-Maßen der OV-Aufgabe erfaßt werden,

¹⁷Mittelbar gilt dies insbesondere in bezug auf Anwendungen zur Verfeinerung der klassischen lexikalischen Computergestützten Inhaltsanalyse der Sozialwissenschaften. Zwar wird eine Ersetzung im Originaltext vorgenommen, jedoch erfolgt die anschließende wörterbuchbasierte Verarbeitung derart grobkörnig, daß opake Kontexte eine allenfalls marginale Rolle spielen.

die Qualität der Ergebnisse des Prozesses der Bestimmung nichtpronominaler Substitute potentiell beeinträchtigen. Somit stellt sich erneut die Frage, ob und ggf. in welcher Weise die beiden Evaluationsdisziplinen voneinander abgegrenzt werden sollten und wie demnach die Definition geeigneter Größen E_{ps} , K_{ps} und A_{ps} vonstatten gehen kann. Unter der Festlegung, daß alleine die Systemleistung in bezug auf die PS-Aufgabe sowie unter Ausblendung eventueller OV-Fehler - d.h. aufgaben-lokal, "netto" - zu bewerten ist, erscheint folgende Vorgehensweise als adäquat. Die Betrachtungen werden eingeschränkt auf die eigentlichen entscheidungsrelevanten Entitäten, d.h. Pronomen P , die (auch) im Schlüssel als Okkurrenzen ausgezeichnet sind; relevant sind somit ausschließlich die Mengen o_{++} , o_{+-} , $o_{+?}$, o_{+} und o_{+*} . Auf der Grundlage inhaltlich adäquat erscheinender Definitionen der Bezugsmengen $K_{ps} := o_{++}$, $E_{ps} := o_{++} \cup o_{+-} \cup o_{+?}$ und $A_{ps} := o_{++} \cup o_{+-} \cup o_{+?} \cup o_{+}$ (die unresolvierten Vorkommen der Menge o_{+*} , deren Kospezifikation schlüsselseitig als optional markiert ist, werden auch hier nicht gezählt) ergibt sich folgende Festlegung der Evaluations-Maße:

$$P_{ps} := \frac{K_{ps}}{E_{ps}} = \frac{|o_{++}|}{|o_{++}| + |o_{+-}| + |o_{+?}|}$$

$$R_{ps} := \frac{K_{ps}}{A_{ps}} = \frac{|o_{++}|}{|o_{++}| + |o_{+-}| + |o_{+?}| + |o_{+}|}$$

Aufgrund der Gegebenheit, daß die Grundmenge der entscheidungsrelevanten Entitäten in Gestalt der systemseitig ermittelten pronominalen Okkurrenzen mit Schlüssel-Entsprechung fixiert ist und ausschließlich die Qualität der für die Instanzen dieser Menge zu treffenden Entscheidungen zur Evaluation ansteht, ergibt sich in bezug auf die angegebenen Definitionen die Besonderheit, daß stets $P_{ps} \geq R_{ps}$ ist.¹⁸ Die beschriebene Unterscheidung zwischen *Precision* und *Recall* erscheint aber dennoch angemessen, da ein *inneres* (auf die beschriebene Grundmenge eingegrenztes) Austauschverhältnis besteht, das durch die Beziehung der Mengen o_{+} auf der einen und den Mengen o_{++} , o_{+-} und $o_{+?}$ auf der anderen Seite determiniert wird. Auf der Grundlage der definierten Maße ist somit noch immer die Untersuchung der zentralen Frage möglich, ob durch einen gezielten Verzicht auf bestimmte, als unsicher angesehene Entscheidungen eine Erhöhung der *Precision* erreicht werden kann.

Aus der Anwendungsperspektive könnte es darüberhinaus naheliegen, eine Version der Evaluationsmaße zu definieren, in die die *kumulierten Effekte* von Fehlern auf den Ebenen der OV- und PS-Aufgaben einfließen. Unter Rekurs auf die Klassifikation in Abbildung 11.7 bedeutet dies für die *Precision*-Definition, daß zusätzlich (Fehl-)Entscheidungen zu berücksichtigen sind, die für *fälschlicherweise* als pronominal objektspezifizierende Okkurrenzen aufgefaßte Okkurrenzen getroffen wurden; somit ergibt sich als erweiterte Grundgesamtheit die Menge $E_{ps}^a := E_{ps} \cup o_{?+}$ ("a" wie Anwendung). In analoger Weise ist im Rahmen einer entsprechend erweiterten *Recall*-Definition die Menge o_i der fälschlicherweise *nicht* systemgenerierten, jedoch im Schlüssel spezifizierten pronominalen Vorkommen zu erfassen: $A_{ps}^a := A_{ps} \cup o_i$. Hiermit ergeben sich fol-

¹⁸Gleichheit ergibt sich genau dann, wenn für alle entscheidungsrelevanten Entitäten eine Entscheidung getroffen wurde, d.h. falls $o_{+-} = \emptyset$ gilt.

gende kombinierte, aus Anwendungssicht ultimativ aussagekräftige Maße:¹⁹

$$P_{ps}^a := \frac{K_{ps}}{E_{ps}^a} = \frac{|o_{++}|}{|o_{++}| + |o_{+-}| + |o_{+?}| + |o_{?+}|}$$

$$R_{ps}^a := \frac{K_{ps}}{A_{ps}^a} = \frac{|o_{++}|}{|o_{++}| + |o_{+-}| + |o_{+?}| + |o_{+-}| + |o_i|}$$

11.2 Implementierung: das ROSANA-System

Nachdem die Parameter des Evaluationsszenarios fixiert wurden, soll nun zunächst die Implementierung eines Softwaresystems diskutiert werden, das die drei in Abschnitt 11.1.1 identifizierten Teilprobleme der referentiellen Interpretation löst. Die Kernkomponente dieses Systems, das sog. *ROSANA-System*, basiert auf dem in Abschnitt 10.3 entwickelten Algorithmus zur Verifikation bindungstheoretischer Bedingungen auf fragmentarischer Syntax. Das Akronym ROSANA steht für

Robuste syntaxbasierte Interpretation anaphorischer Ausdrücke

Das ROSANA-System bearbeitet die drei oben definierten Teilaufgaben der Kospezifikations-Interpretation. Es soll Gegenstand einer Evaluation sein, der die entsprechenden formalen Evaluationsmaße zugrundeliegen.

11.2.1 Übersicht der Systemarchitektur

Abbildung 11.8 gibt eine datenflußorientierte Darstellung der Architektur des zu evaluierenden Systems einschließlich der Bewertungskomponente. Die Ausgangsdaten, Zwischenresultate und Endergebnisse werden in *Dateien* mit identischem Basisnamen abgelegt; entsprechend der üblichen Praxis erfolgt die Unterscheidung über die dreibuchstabile Dateinamen-Extension. Die referentiell zu interpretierenden Textkorpora liegen in ASCII-Dateien mit Extension `.txt` ab. In diesem Originalformat fließen die Textdokumente ausschließlich in die der eigentlichen referentiellen Interpretation vorgelagerten Prozesse der morphologischen und syntaktischen Analyse ein.

¹⁹ Der Rückgriff auf die Menge o_i ausschließlich schlüsselseitig spezifizierter *pronominaler* Okkurrenzen ist in evaluationstechnischer Hinsicht problematisch, da an das Referenzkorpus die zusätzliche Anforderung gestellt wird, daß Information bezüglich des *Typs* der spezifizierten Vorkommen zur Verfügung steht. Da diese Bedingung durch die Referenzkorpora der im folgenden durchgeführten Evaluation nicht erfüllt wird, werden die entsprechenden Werte per Hand ermittelt. Dies erweist sich jedoch als wenig problematisch, da die *Gesamtheit* der *Recall*-Fehler, aus der sich die Elemente der Menge o_i rekrutieren, überschaubar ist und automatisch generiert werden kann. Ein ähnlicher Einwand betrifft die bislang stillschweigend getroffene Annahme, daß die systemseitig getroffene Entscheidung, welche Okkurrenz *pronominal* ist, mit hundertprozentiger Genauigkeit getroffen werden kann - auf der Grundlage der Vorgaben des Schlüssels ist jedoch lediglich die Überprüfung möglich, ob es sich bei einer systemgenerierten Entität um ein Vorkommen *beliebigen Typs* handelt. Das hieraus potentiell entstehende Problem hat jedoch allenfalls theoretischen Status, da, wie anhand einer näheren Untersuchung der Systemergebnisse deutlich wird, die als Okkurrenzen identifizierten Elemente hinsichtlich der Kategorien *pronominal* vs. *nichtpronominal* fehlerfrei klassifiziert werden können.

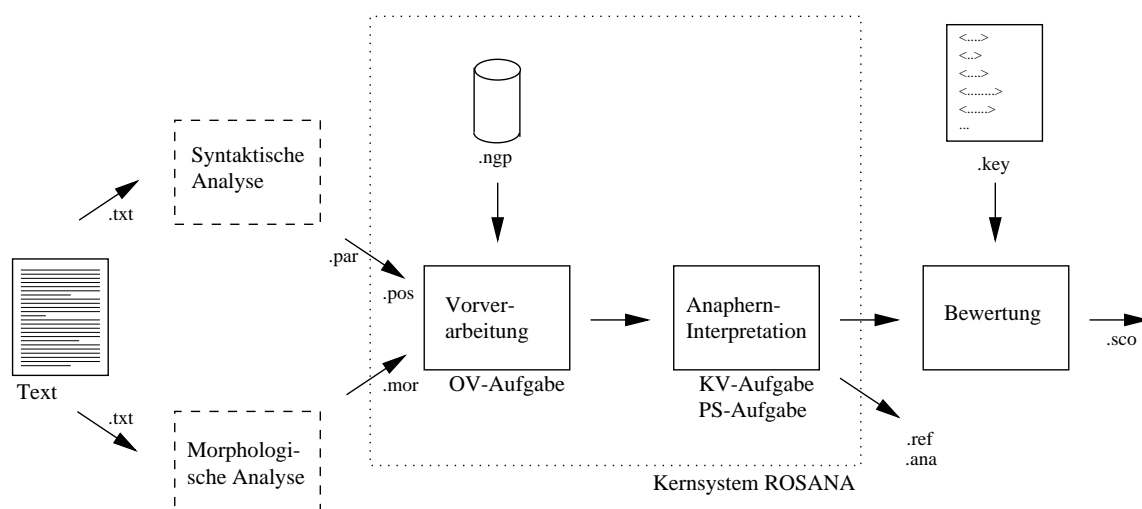


Abbildung 11.8: Teilsysteme: syntaktische/morphologische Analyse, ROSANA, Bewertung

Die morphologische und syntaktische Verarbeitung findet *extern*, d.h. "offline" statt, da die entsprechenden Software-Module auf dem für die Evaluation verwendeten Computersystem nicht zur Verfügung stehen (vgl. Abschnitt 11.2.2). Die Ergebnisse dieser beiden Teilprozesse werden in Dateien mit Endungen `.mor` bzw. `.par/.pos` (zusätzlich mit fortlaufenden Wortnummern) verwaltet, die die beiden zentralen Eingaben des Kernsystems der referentiellen Interpretation bilden.

Das Kernsystem ROSANA, das die drei beschriebenen Teilaufgaben der referentiellen Interpretation löst, läßt sich in zwei wesentliche Komponenten zerlegen. Auf der Stufe der *Vorverarbeitung* werden die in den `.mor`- und `.pos`-Dateien zur Verfügung stehenden Ergebnisse der morphologischen bzw. syntaktischen Analyse eingelesen und in systeminternen Datenstrukturen abgelegt. U.a. wird auf dieser Ebene bereits das Problem der Bestimmung der objektspezifizierenden Vorkommen, d.h. die *OV-Aufgabe*, gelöst; zusätzliche allgemeine sowie ggf. korpuspezifische Lexika mit Informationen über Numerus und Genus bestimmter Stammformen (insbesondere für Namen) erleichtern die Anlegung der korrekten morphologischen Kongruenzmerkmale in den Beschreibungen der Okkurrenzen (`.ngp`-Dateien). Ein zweites zentrales Ziel dieses Verarbeitungsschritts besteht in der möglichst umfassenden Rekonstruktion der potentiell fragmentarischen syntaktischen Oberflächenstruktur aus den Ergebnissen der syntaktischen Analyse; da die durch den verwendeten Parser generierten Abhängigkeitsstrukturen nicht das benötigte phrasenstrukturelle Format aufweisen, sind nichttriviale Transformationen zu leisten, die von einem gesonderten Systemmodul, dem sog. *Oberflächenstruktur-Konstruktor*, durchgeführt werden (vgl. Abschnitt 11.2.3). Die Ergebnisse der Vorverarbeitung - Oberflächenstrukturbäume, die entsprechend den Ausführungen in Abschnitt 10.3.3 mit den Vorkommensbeschreibungen verknüpft sind - werden unmittelbar dem Prozeß der Anapherninterpretation zugeführt, der die Teilprobleme der Kospezifikationsklassen-Ermittlung sowie der Bestimmung nichtpronominaler Substitute für Pronomen (*KV*- bzw. *PS-Aufgabe*) löst. Die Ergebnisse der referentiellen Interpretation werden in Dateien mit Endungen `.ana` (klassenweise Aufschlüsselung der Kospezifikationsbeziehungen) und `.ref` (allgemeine, bezüglich Inhalt und Detailliertheit konfigurierbare Beschreibung der Ergebnisse für die drei Teildisziplinen) dokumentiert.

Die Bewertung der Interpretationsleistung des Systems für die drei Teildisziplinen geschieht

auf der Grundlage der internen Repräsentation der Analyseergebnisse. Sie wird durch ein eigenständiges Modul bewerkstelligt, das die in Abschnitt 11.1.4 hergeleiteten formalen Evaluationsmaße implementiert. Vergleichsmaßstab sind intellektuell erstellte Referenzdaten, die in einer separaten Datei (Endung *.key*) abliegen. Als zentraler Bezugspunkt für den Abgleich des systemgenerierten Ergebnisses mit den Vorgaben des Referenzkorpus dienen gemeinsame *Wortnummern*, die sowohl in den Eingabedateien der computergestützten Referenzanalyse als auch in der Schlüsseldatei zur Verfügung stehen.

Bevor nun zur Evaluation geschritten wird, sollen die zentralen Aspekte des in Abbildung 11.8 skizzierten Systems näher erörtert werden. Von besonderem Interesse ist die genauere Ausgestaltung des zugrundeliegenden Algorithmus zur referentiellen Interpretation hinsichtlich der betreffend die Operationalisierung der Einzelstrategien bestehenden Freiheitsgrade. Ebenfalls einer näheren Betrachtung unterzogen werden sollen die Systeme der externen Voranalyse sowie die im Rahmen der Vorverarbeitung zur Anwendung gelangenden Strategien bzw. Heuristiken für den Umgang mit partieller Information; anhand der Diskussion werden erneut die Schwierigkeiten deutlich werden, die sich in bezug auf den Einsatz massendatentauglicher, robuster Systemkomponenten ergeben.

11.2.2 Syntaktische und morphologische Analyse

Die syntaktische Analyse der referentiell zu interpretierenden Textdokumente erfolgt durch den *Dependency Parser for English*, der an der *Research Unit For Multilingual Language Technology* der *University of Helsinki* entwickelt wurde und von LINGSOFT, Helsinki, vertrieben wird (Järvinen und Tapanainen, [JäTa97]). Die zentralen Merkmale dieses Parsers entsprechen den Anforderungen einer robusten referentiellen Interpretation, der das in Abschnitt 10.1.3 beschriebene Modell der potentiell fragmentarischen Voranalyse - vgl. Definition 10.2 - zugrundeliegt. Ungrammatische Eingaben werden *partiell*, d.h. *so weit wie möglich* syntaktisch analysiert (ibid., S. 5):

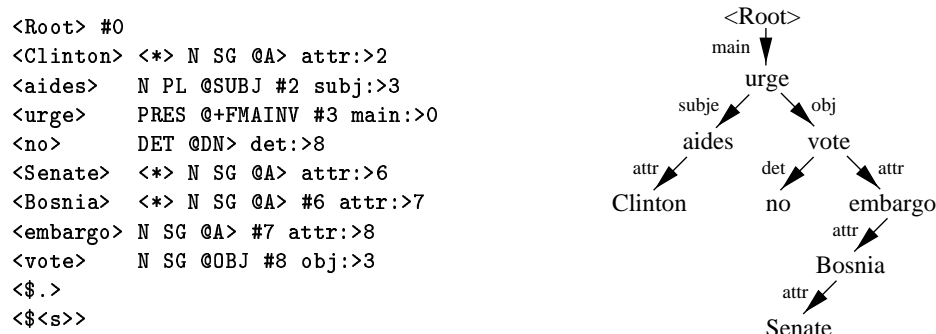
“The grammar is not generative. This means that the parser accepts every input sentence, and returns an analysis even for ungrammatical sentences to the extent that the structure is recoverable.”

Auf die Anwendung tiefergehender semantischer Kriterien wird verzichtet; strukturelle Ambiguitäten bleiben ggf. unresolviert. Die Ausgabe des Syntaxanalyse-Prozesses ist somit *potentiell fragmentarisch*.

Auch gegenüber der Normativität der Grammatik wird der Robustheits-Eigenschaft Vorrang eingeräumt: Auf Kongruenztests zwischen Subjekt und finitem Verb wird verzichtet, da es zahlreiche Ausnahmen gibt, die nicht zu einem (lokalen) Scheitern der Analyse führen sollten. Die implementierten grammatischen Restriktionen sind somit eher zu schwach als zu stark.

Die von Järvinen und Tapanainen angegebene Analysegeschwindigkeit von 350 Wörtern pro Sekunde auf einem PC (Pentium-Prozessor, 166 MHz) unter dem Betriebssystem LINUX genügt den Anforderungen der Verarbeitung umfangreicher Textkorpora. Die praktische Robustheit des Softwaresystems wurde per Analyse umfangreicher Korpora mit insgesamt mehr als 30 Millionen Wörtern (Stand: März 1997) nachgewiesen.

Die Ausführungen im Vorspann zu Kapitel 10 betreffend die Probleme einer robusten Algorith-

Abbildung 11.9: Syntaxanalyse-Ergebnis des *Dependency Parser for English* für Beleg (18)

misierung des Syntaxanalyse-Prozesses werden somit durch die Gegebenheiten in einem realen System untermauert. Mit den beschriebenen Eigenschaften erweist sich der *Dependency Parser for English* als idealer Kandidat für den Einsatz als syntaktischer Voranalyseprozeß eines Anaphernresolutions-Systems, das auf einer uneingeschränkten, jedoch potentiell fragmentarischen Syntaxanalyse basiert; er erfüllt alle Anwendungsvoraussetzungen des in Abschnitt 10.3 beschriebenen Algorithmus zur robusten referentiellen Interpretation. Mit der Entscheidung, den *Dependency Parser for English* zu verwenden, geht auch die Wahl des Englischen als Sprache der Evaluationskorpora einher; dies stellt jedoch keine prinzipielle Einschränkung des Verfahrens dar, da mit der mittelfristigen Entwicklung gleichwertiger Parser für weitere Sprachen (insbesondere für Deutsch) gerechnet werden kann.²⁰

Da der *Dependency Parser for English* nicht auf dem Computersystem installiert ist, auf dem die Evaluation erfolgt, stellt dieser keine Komponente des realisierten Softwaresystems (umfassend ROSANA und Bewertung) dar. Die syntaktischen Analysen der Evaluationskorpora entstammen einer *externen* Voranalyse und liegen in Form von ASCII-Dateien vor.²¹ Die Anwendungstauglichkeit des Gesamtsystems ist dennoch gewährleistet, da die Integration des Syntaxanalysators mit dem ROSANA-System eine softwaretechnische Aufgabe darstellt, die nicht mit prinzipiellen (und damit diskussionswürdigen) Problemen verbunden ist.

Abbildung 11.9 zeigt das Ergebnis der Syntaxanalyse für folgenden Beleg:

(18) *Clinton aides urge no Senate Bosnia embargo vote.*

Im linken Teil der Abbildung sind die wortbezogenen Ausgaben des Parsers zeilenweise dargestellt.²² Die syntaktisch-strukturelle Information wird unter Angabe der vom jeweiligen Wort eingenommenen syntaktischen Funktion sowie des dependentiellen Dominators repräsentiert; beispielsweise codiert *obj:>3* in der Zeile von “*vote*”, daß der sprachliche Ausdruck mit diesem Kopf - “*no Senate Bosnia embargo vote*” - die Rolle des direkten Objekts des finiten Verbs “*urge*” (Oberflächen-Position #3) einnimmt. Der rechte Teil der Abbildung gibt eine graphische Veranschaulichung des auf diese Weise repräsentierten Syntaxbaums.

²⁰Der *Dependency Parser for English* basiert in Teilen auf dem in Abschnitt 5.3 erwähnten robusten “*Part-Of-Speech*”-Tagger *English Constraint Grammar*, für den u.a. auch ein deutschsprachiges Pendant verfügbar ist. Unabhängig hiervon bietet die Zugrundelegung der englische Sprache den Vorteil, daß die Aufgaben-Spezifikationen der MU-Konferenzen unmittelbar anwendbar sind und daß ferner die Vergleichbarkeit der Evaluationsergebnisse mit den MUC-Resultaten gewährleistet ist.

²¹Einer der Entwickler des Parsers, Pasi Tapanainen, stellte die Syntaxanalyse-Ergebnisse freundlicherweise auf elektronischem Wege (per Email, Pasi.Tapanainen@ling.Helsinki.fi) zur Verfügung.

²²Um Platz zu sparen, wurden die eigentlich je Wort zweizeiligen Ergebnisse auf eine Zeile zusammengezogen.

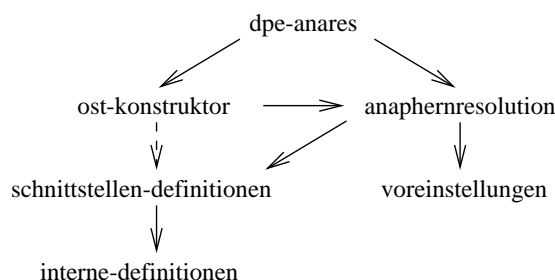


Abbildung 11.10: Modulstruktur des Kernsystems ROSANA

Das Syntaxanalyse-Ergebnis enthält noch weitere Information, die sich für die unterschiedlichen Aufgabenstellungen der referentiellen Interpretation als relevant erweist. In der als Beispiel betrachteten Zeile der Analyseeinheit “*vote*” sind dies Angaben bezüglich Wortart (N wie *Nomen*) und morphologischer Flektion (S wie *Singular*). Nicht in der Parsing-Ausgabe enthalten ist hingegen die Spezifikation der jeweiligen *Stammform*. Da diese Information ebenfalls von zentraler Bedeutung ist, wird zusätzlich auf die Ergebnisse einer *morphologischen Analyse* zurückgegriffen. Für das Wort “*aides*” in Beleg (18) etwa liefert diese Verarbeitungsstufe das Ergebnis

"<aides>" "aide" N NOM PL

aus dem insbesondere die entsprechende Stammform - das Nomen “*aide*” - hervorgeht. Aus denselben Gründen wie zuvor wird auch die morphologische Analyse extern bewerkstelligt. Das hierzu herangezogene Software-System - die *morphologische Analysekomponente ENGTWOL* der *English Constraint Grammar ENGCG* - genügt hinsichtlich der Robustheit denselben hohen Ansprüchen wie der *Dependency Parser for English*, so daß sich auch hierdurch keine Einschränkungen betreffend die Anwendungstauglichkeit des Gesamtsystems ergeben.²³

11.2.3 Das Kernsystem ROSANA

Abbildung 11.10 zeigt die Abhängigkeitshierarchie der “*Common Lisp*”-Module, die das Kernsystem ROSANA implementieren. In den einzelnen Modulen sind die Datenstrukturen und Prozeduren (LISP-Funktionen) zur Bearbeitung der drei Teilaufgaben der referentiellen Interpretation codiert. Das Design der Modulstruktur reflektiert die üblichen softwaretechnischen Zielsetzungen der Wartbarkeit, Adaptierbarkeit und Wiederverwertbarkeit.

Modul *dpe-anares* realisiert den ersten Teil des *Vorverarbeitungs-Schritts* (vgl. Abbildung 11.8)

²³Die morphologischen Analysen für die relevanten Wörter des Evaluations-Textkorpus wurden über die im Internet zur Verfügung stehende Demonstrations-Version von ENGTWOL/ENGCG akquiriert (Adresse: <http://www.lingsoft.fi/cgi-pub/{engtwol,engcg}>, Stand: März 1998). Aus Praktikabilitätsgründen wurde nicht unmittelbar auf die morphologische Analyse des Softwaresystems ENGTWOL zurückgegriffen, da dessen Demonstrations-Version lediglich die Analyse einzelner Wörter unterstützt. Die Verarbeitung der Wörter durch ENGCG geschah in einer syntaktisch dekontextualisierten Weise, so daß die erzeugten Analyseergebnisse letztlich äquivalent zu denen einer ausschließlichen ENGTWOL-Analyse sind.

Die Verarbeitungsstufe einer separaten morphologischen Analyse könnte entfallen, wenn der *Dependency Parser for English*, der ohnehin auf einer ENGTWOL-Voranalyse basiert, diese Information im Ergebnis weiterreichen würde.

Modul	#Zeilen	Inhalt
dpe-anares	3164	Einlesen Morphologie u. Syntax; Ermittlung der Okkurrenzen
ost-konstruktor	1386	Aufbau phrasenstruktureller Oberflächenstruktur-Beschreibungen
anaphernresolution	2179	Kospezifikations-Interpretation; Ermittlung lexikalischer Substitute
schnittstellen-def...	513	globale Strukturen/Konstanten d. Schnittstelle: Vorkommen (dr-okk), Dependenzbaum (depknoten), morpholog. Kongruenzmerkmale (ngp)
interne-def...	195	sonst. globale Strukturen/Konstanten: Diskursreferent (disref), evtl. fragmentar. Syntaxbaum (ostknoten , fraggraphknoten)
voreinstellungen	196	unterschiedliche Modi betreffend die zu generierenden Ergebnisse

Abbildung 11.11: Module des Kernsystems ROSANA

in Form von Datenstrukturen und Funktionen für das Einlesen der syntaktischen und morphologischen Analyseergebnisse in den Beschreibungsformaten, die durch die zugrundegelegten Systeme *Dependency Parser for English* bzw. *ENGTWOL* erzeugt werden. Die Datenstrukturen, in denen die Resultate dieses Einleseprozesses repräsentiert werden, sind *standardisiert*. Hierdurch entsteht der gewichtige Vorteil einer vereinfachten Adaptierbarkeit des ROSANA-Systems an andere syntaktische bzw. morphologische Analysensysteme mit abweichendem Ergebnisformat, da die entsprechenden Adaptionen auf das Modul **dpe-anares** begrenzt bleiben. Zentrale Bestandteile des anzulegenden standardisierten Zwischenformats sind die Beschreibungen der objektspezifizierenden Okkurrenzen; **dpe-anares** löst somit bereits die OV-Aufgabe.

Der zweite Schritt der Vorverarbeitung wird durch das Modul **ost-konstruktor**, dem sog. *Oberflächenstruktur-Konstruktor*, realisiert. Die in standardisierter Form vorliegenden dependenzstrukturellen Ergebnisse der syntaktischen Analyse werden auf phrasenstrukturelle Beschreibungen der syntaktischen Oberflächenstruktur abgebildet, die entsprechend den Spezifikationen in Abschnitt 10.3.3 mit den Vorkommensbeschreibungen vernetzt sind. Die Ergebnisse dieses Verarbeitungsschritts bilden die geeignete Grundlage für die eigentliche Anapherninterpretation, die durch das Modul **anaphernresolution** geleistet wird, das die beiden Kernaufgaben - die Kospezifikations-Interpretation und die Ermittlung lexikalischer Substitute für Pronominalanaphern - löst.

In den beiden Modulen **schnittstellen-definitionen** und **interne-definitionen** sind Konstanten, Variable und Datenstruktur-Definitionen mit modul-übergreifender Relevanz codiert. Die Unterscheidung der beiden Module reflektiert erneut die Intention der Vereinfachung einer Adaption an andere Voranalyse-Systeme: **schnittstellen-definitionen** spezifiziert die hierfür relevanten standardisierten Datenstrukturen; **interne-definitionen** beschreibt die übrigen, ausschließlich die Interna von ROSANA betreffenden Entitäten mit nichtlokalem Skopus. Einige Variable und Funktionen für die System-Konfiguration (zu generierende Ergebnisse etc.) sind in die Komponente **voreinstellungen** ausgelagert.

In Abbildung 11.11 sind die wesentlichen Charakteristika der ROSANA-Module einschließlich deren Größe²⁴ zusammengefaßt. Es sollen nun einige zentrale Verarbeitungsschritte näher expliziert werden.

Ermittlung objektspezifizierender Vorkommen

Einer der wesentlichen Bestandteile des durch das Modul **dpe-anares** realisierten Einlesens der syntaktischen und morphologischen Information besteht in der Bearbeitung der OV-Aufgabe als

²⁴Angegeben sind die Anzahl der Zeilen (kommentierten) LISP-Codes des jeweiligen Moduls.

erster Teildisziplin der referentiellen Interpretation. Der Ansatz zur Lösung dieser spezifischen inhaltlichen Aufgabe soll exemplarisch für die Arbeitsweise dieses Moduls beschrieben werden, die insgesamt auf ähnlichen Techniken beruht.

Die Identifikation objektspezifischer Vorkommen geschieht in zwei Schritten: (1) Ermittlung sprachlicher Ausdrücke, die potentielle Vorkommens-Induktoren verkörpern; (2) Klassifikation nach Vorkommens-Typ und Anlegung entsprechender Beschreibungen vom Datentyp **dr-okk**²⁵. Schritt (1) geschieht unter Rekurs auf die Wortarten-Angaben im Ergebnis der syntaktischen Analyse; Schritt (2) erfolgt unter Betrachtung der Zeichenkette des jeweiligen Ausdrucks.

Abbildung 11.12 beschreibt die Datenstrukturen, auf deren Grundlage dieses Verfahren vollzogen wird. In Schritt (1) wird auf die Liste **Okk-WAs** Bezug genommen, die die Wortarten potentieller Vorkommens-Induktoren spezifiziert. Die Identifikation vorkommens-relevanter Nomen geschieht *sensitiv zum syntaktischen Kontext*, so daß (entsprechend der Aufgaben-Definition) nur eine Teilmenge aller Ausdrücke der angegebenen Wortarten - solche als Kopf *in der syntaktischen Funktion eines Mitspielers* - Berücksichtigung findet. Des Weiteren werden Ausdrücke in der Rolle eines Artikelworts als relevant betrachtet, insofern sie in der Liste **Possessivpronomen** spezifiziert sind. Die interne Vorkommens-Typisierung in Schritt (2) wird u.a. (für Pronomen) über die Zuordnungstabelle **Pronomen-Typisierung** bewerkstelligt, die jeden pronominalen Ausdruck auf einen entsprechenden internen Typ abbildet.²⁶ In analoger Weise spezifiziert die Tabelle **Pronomen-NGP** die Kongruenzmerkmale Numerus, Genus und Person für die unterschiedlichen Pronomen, wodurch die Basis für die Konstruktion der Okkurrenz-Beschreibungen Basis geschaffen wird.

Vergleichbare Datenstrukturen existieren für Namen und sonstige Nomen. **Morph-Tabelle** stellt die Ergebnisse der morphologischen Analyse zur Verfügung. Über die Struktur **NGP-Lexikon** können morphologische Kongruenzmerkmale für Eigennamen und sonstige Nomen vorgegeben werden, für die die morphologische Analyse unzureichende bzw. fehlerhafte Analyseergebnisse liefert.²⁷ ROSANA bietet daher die Möglichkeit, ergänzende Vorgaben einzubringen (**ngp**-Dateien, vgl. Abbildung 11.8), die den Inhalt von **NGP-Lexikon** konstituieren. Eine vergleichbare Zielsetzung liegt der Struktur **Namens-Lexikon** zugrunde, über die Flexionstabellen (Stammformen-Zuordnungen) für häufig wiederkehrende, systemseitig unbekannte Namen spezifiziert werden können.

Abbildung 11.13 expliziert einen Ausschnitt der in der ROSANA-Implementierung spezifizierten Vorgaben, anhand dessen sich die Verarbeitung des Pronomens "she" erläutern läßt. In den Analysen des *Dependency Parser for English* ist "she" von der Wortart PRON und somit gemäß der Definition von **Okk-WAs** ein potentieller Vorkommens-Induktor. Da "she" nicht zu denjenigen PRON-klassifizierten Ausdrücken gehört, die im Hinblick auf die Spezifikation der OV-Aufgabe als nicht-spezifizierend zu interpretieren sind (Liste **PRONs-ohne-Okk**), wird zunächst der Okkurrenztyp ermittelt, der in Tabelle **Pronomen-Typisierung** als ***Personalpronomen-3*** beschrieben ist. Analog ergeben sich per Nachschlagen in **NGP-Lexikon** die grammatischen Kongruenzmerkmale **sg** (Singular), **fem** (Femininum), **3** (dritte Person).

Der vollständige Algorithmus zu Schritt (2) ist in Abbildung 11.14 beschrieben. Eine spezielle Heuristik kommt in den Schritten zur Bestimmung der morphologischen Kongruenzmerkmale im Falle nichtpronominaler Vorkommen zum Einsatz. Insofern sich keine entsprechende Spezifikation in der Struktur **NGP-Lexikon** vorfindet, werden für definite NP ausschließlich Merkmals-

²⁵ vgl. die Definition der Datenstruktur **dr-okk** auf S. 166, Abbildung 9.2

²⁶ Im Falle des hinsichtlich possessiver versus nichtpossessiver Lesart ambigen Pronomens "her" fließt zusätzlich der syntaktische Kontext ein.

²⁷ Dies erweist sich insbesondere deshalb als wichtig, weil das Lexikon der morphologischen Analyse keine hundertprozentige Abdeckung aufweist. Im Speziellen gilt dies für Eigennamen.

Datenstruktur	Beschreibung
Okk-WAs	parserseitig generierte Wortarten potentieller Vorkommens-Induktoren
Possessivpronomen	Artikelwörter, die als Possessivpronomen zu interpretieren sind
Pronomen-Typisierung	interne Typen für pronominale Wörter
Pronomen-NGP	morpholog. Kongruenzmerkmale (Num, Gen, Per) für pronominale Wörter
PRONs-ohne-Okk	als PRON klassifizierte Wörter, die keine Vorkommen induzieren
Morph-Tabelle	Ergebnisse der morphologischen Analyse: Stammformen für flektierte Formen
NGP-Lexikon	morphologische Kongruenzmerkmale (Num, Gen, Per) für Namen/Nomen
Namens-Lexikon	Stammformen für Namen
Nomen-ohne-Okk	als N/NUM/ABBR klassifizierte Wörter, die keine Vorkommen induzieren

Abbildung 11.12: Vorgaben für die Vorkommens-Ermittlung des ROSANA-Algorithmus

```
(defconstant Okk-WAs '(PRON N NUM ABBR))
(defconstant PRONs-ohne-Okk
  '("less" "more" "other" "such" "what" "whichever" "whoever"))
(defconstant Pronomen-Typisierung
  (list .. (list "he" "she" "it" "him" "her" "they" "them" *Personalpronomen-3*) ..))
(defconstant Pronomen-NGP
  (list .. (list "she" "her" "herself" (list (cons-ngp 'sg 'fem '3))) ..))
```

Abbildung 11.13: Ausschnitt der Vorgaben (Pronomen “she”)

- Falls der *Dependency Parser for English* dem potentiell vorkommens-induzierenden Ausdruck (Kopf) w die Wortart PRON zugeordnet hat:
 - Falls $w \in \text{PRONs-ohne-Okk}$, so betrachte w als nichtspezifizierend.
 - Sonst: Ermittle den internen Okkurrenztyp $t(w)$ über die Tabelle Pronomen-Typisierung.
 - * Falls $t(w)$ definiert ist:
 1. Lege eine Vorkommens-Beschreibung $\text{dr-okk}(w)$ für w mit Typ $t(w)$ an;
 2. ermittle die morphologischen Kongruenzmerkmale für $\text{dr-okk}(w)$ durch Nachschlagen in der Tabelle Pronomen-NGP.
 - * Falls $t(w)$ nicht definiert ist, so lege eine Default-Beschreibung $\text{dr-okk}(w)$ an.
- Sonst ($\Rightarrow w$ ist Eigenname oder sonstiges Nomen):
 - Falls $w \in \text{Namens-Lexikon}$ oder falls w im Ergebnis Morph-Tabelle der morphologischen Analyse als *Eigenname* klassifiziert ist:
 1. Lege eine Vorkommens-Beschreibung $\text{dr-okk}(w)$ für w vom Typ *Name* an;
 2. ermittle die morphologischen Kongruenzmerkmale für $\text{dr-okk}(w)$:
 - * durch Nachschlagen in der Tabelle NGP-Lexikon (sofern enthalten),
 - * oder per Ergänzung der im Parse-Ergebnis verfügbaren Numerus-Information.
 - Sonst ($\Rightarrow w$ ist sonstiges Nomen):
 - * Falls $w \in \text{Nomen-ohne-Okk}$, so betrachte w als nichtspezifizierend.
 - * Sonst:
 1. Lege eine Vorkommens-Beschreibung $\text{dr-okk}(w)$ für w vom Typ *Nomen* an;
 2. ermittle die morphologischen Kongruenzmerkmale für $\text{dr-okk}(w)$:
 - durch Nachschlagen in der Tabelle NGP-Lexikon (sofern enthalten),
 - oder per Ergänzung der im Parse-Ergebnis verfügbaren Numerus-Information.

Abbildung 11.14: ROSANA-Algorithmus zur Vorkommens-Klassifikation

kombinationen mit dem grammatischen Geschlecht Neutrum erzeugt, wohingegen für Namen alle drei Genus Berücksichtigung finden. Zwar wird hiermit nicht in allen Fällen die korrekte Entscheidung getroffen; jedoch erweist sich diese Heuristik im Vergleich am erfolgreichsten.

Rekonstruktion phrasenstruktureller Oberflächenstruktur-Beschreibungen

Die Ergebnisse der Analysen des *Dependency Parser for English* liegen noch nicht in dem phrasenstrukturellen Format vor, das für die Verifikation der bindungstheoretischen Bedingungen benötigt wird. Bestimmte Bestandteile der oberflächenstrukturellen Konfiguration, auf die in den Definitionen der K-Herrschafts- und damit der Bindungs-Relation sowie der Bindungsprinzipien Bezug genommen wird, sind in den dependentiellen Beschreibungen nicht *strukturell* expliziert, sondern indirekt über die syntaktischen Funktionen sowie über die Oberflächenanordnung repräsentiert. Die Aufgabe des Moduls *ost-konstruktor* besteht darin, die phrasenstrukturelle Repräsentation der syntaktischen Oberflächenstruktur insoweit aus den Dependenzbeschreibungen zu rekonstruieren, daß die notwendigen Details zur Verifikation der syntaktisch-konfiguralen Restriktionen *konfigural* zur Verfügung stehen.

Zu den phrasenstrukturellen Feinheiten, die in dependentiellen Syntaxbeschreibungen nicht erfaßt werden, jedoch hinsichtlich der Zielsetzung einer Anwendung der Bindungstheorie für die Kospezifikations-Interpretation objekt-spezifischer Vorkommen relevant sind, zählen in erster Linie die konfiguralen *Asymmetrien zwischen Subjekt und Objekt* sowie ferner bestimmte konfigurale Differenzierungen zwischen *extraponierten und nichtextraponierten Komplementsätzen*. Folgende Belege illustrieren den ersteren Fall (vgl. auch Abschnitt 7.1.2):

- (19a) * He_i shaved the client_j before the barber_i left.
 (19b) The barber_i shaved him_j before the client_j left.

Die Unterschiede in der Akzeptabilität der Koindexierungen in den Sätzen (19a) bzw. (19b) werden ausschließlich in *phrasenstrukturellen* Repräsentationen der syntaktischen Oberflächenstruktur konfigural expliziert: Nur in ersterem, nicht jedoch in letzterem Beleg ist das Bindungsprinzip C des nichtpronominalen Antezedens verletzt, da zwar das Subjekt, nicht jedoch das Objekt des Hauptsatzes den Inhalt des Adverbialsatzes k-beherrscht. Dependenzstrukturelle Repräsentationen bilden diesen konfiguralen Unterschied zwischen Subjekt und Objekt (bzw. - allgemeiner - dem Inhalt der VP) nicht ab (vgl. Abbildung 11.15).

Ansätze zur Adaption der bindungstheoretischen Restriktionen auf die konfiguralen Gegebenheiten dependentieller Beschreibungen der syntaktischen Oberflächenstruktur, die ausschließlich auf einer ergänzenden Anwendung von Anordnungsinformation basieren und die syntaktische Funktion außer Betracht lassen, greifen somit zu kurz.²⁸ Das Modul *ost-konstruktor* bewerkstelligt daher die notwendigen Transformationen - insoweit die Information verfügbar ist - unter zusätzlichem Rekurs auf die syntaktischen Rollen des Ergebnisses der Dependenzanalyse. Eventuelle dependentielle Fragmentierungen des Parsing-Ergebnisses werden unmittelbar in entsprechende phrasenstrukturelle Fragmentierungen übersetzt, wobei die Information über die

²⁸Vgl. etwa die Arbeit von Strube und Hahn, in der ein dependenzstrukturelles Analogon zu den Bindungsprinzipien B und C vorgeschlagen wird, das auf die Anordnung von Anapher und Antezedens Bezug nimmt ([StHa95]). Da keine Unterscheidung zwischen Subjekt und VP-Inhalt getroffen wird, unterscheidet die Definition nicht zwischen den Koindexierungen der Belege (19a) und (19b) und erklärt fälschlicherweise *beide* als unzulässig.

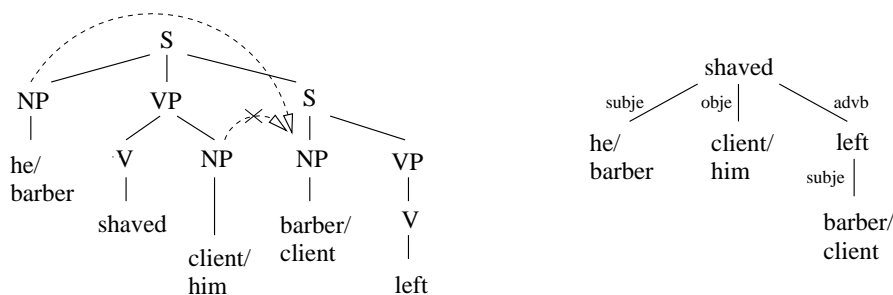


Abbildung 11.15: Vergleich von Phrasen- und Dependenzstruktur der Belege (19a)/(19b)

zwischen den Fragmenten bestehenden Dominanzbeziehungen mitberücksichtigt wird. Als Ergebnis des Abbildungsprozesses liegen syntaktische Beschreibungen in dem in Abbildung 10.8, Abschnitt 10.3.3 skizzierten Format vor.

Die Implementierung des Abbildungsprozesses ist relativ aufwendig, doch sie bietet den Vorteil, daß die relevanten konfiguralen Feinheiten weitestgehend und mit hoher Transparenz, da *strukturell expliziert*, zur Verfügung stehen. Da das Modul *ost-konstruktor* Dependenzbeschreibungen verarbeitet, die im *standardisierten*, internen Repräsentationsformat vorliegen, besteht zudem der Vorteil, daß eine Portierung von ROSANA auf abweichende (dependenz-)syntaktische oder morphologische Voranalyse-Systeme keine Reimplementierung dieser Systemkomponente bedingt.

Anaphorische und antezedenstaugliche Okkurrenzen

Die beiden zentralen Problemstellungen der Kospezifikations-Interpretation - KV-Aufgabe und PS-Aufgabe - konstituieren den Zuständigkeitsbereich des Moduls *anaphernresolution*, das somit die *Kernkomponente des ROSANA-Systems* verkörpert. Die Implementierung basiert auf dem in Abschnitt 10.3.3 beschriebenen Verfahren zur robusten Anapherninterpretation auf fragmentarischer Syntax (Abbildung 10.10, S. 207) sowie auf dem zugrundeliegenden Basisalgorithmus (Abbildung 9.9, S. 179). Es soll nun näher beschrieben werden, in welcher Weise diese Verfahrensspezifikationen hinsichtlich bestehender Freiheitsgrade ausgestaltet sind.

Der erste Punkt betrifft die konkrete Ausgestaltung der in der bisherigen Formulierung des Algorithmus stillschweigend vorausgesetzten Klassifikation der identifizierten objektspezifizierenden Vorkommen als *anaphorisch* bzw. *nichtanaphorisch*. Primärer Bezugspunkt ist die *bindungs-theoretische Klasse* (A, B, C), die den einzelnen Okkurrenzen in Abhängigkeit des internen Vorkommens-Typs (vgl. o.) nach folgendem generellen Schema zugeordnet wird: Nichtpronominale Nomen und Namen sind vom Typ C, Reflexiva vom Typ A und alle übrigen Pronomen vom Typ B. Als anaphorisch angesehen werden alle Okkurrenzen der Typen A und B sowie ferner diejenigen Okkurrenzen des Typs C, die gemäß der im *dr-okk*-Attribut *quantor* vorliegenden Information als *definit* aufzufassen sind.²⁹ Von dieser Regel abweichend als *nichtanaphorisch*

²⁹Da die Unterscheidung zwischen definiten und indefiniten, einen neuen Diskursreferenten einführenden Vorkommen unter algorithmischen Gesichtspunkten nicht unproblematisch ist (vgl. die Diskussion in Abschnitt 6.1.2, Belege (12a)/(12b)), wird in der gegenwärtigen Implementierung auf eine entsprechende Differenzierung anhand ggf. vorliegender Artikelwörter verzichtet und *alle* nichtpronominalen Nomen als anaphorisch eingeordnet.

Typ der Anapher	Kongruenzbedingungen
Nomen	Kandidat vom Typ *Nomen* , Zeichenkettengleichheit der Stammform, Kongruenz der morphologischen Merkmale Numerus und Genus
Name	Kandidat vom Typ *Name* , vollständige Übereinstimmung oder partielle Übereinstimmung mit Nachnamen eines mehrteiligen Kandidaten-Namens
Personalpronomen-3 *Possessivpronomen-3*	Kongruenz der morphologischen Merkmale Numerus und Person, Genuskongruenz fakultativ, Zeichenkette des Kandidaten ist kein Zahl
Relativpronomen	ausschließlich satzinterne Kandidaten, Kandidat geht Anapher voraus, Kongruenz der morphologischen Merkmale Numerus, Genus und Person
Personalpronomen-1 *Possessivpronomen-1*	entweder: Kandidat vom Typ *Name* oder *Nomen* , morphologische Kongruenz in Numerus und Genus, Numerus Singular, Genus Maskulinum oder Femininum, Kandidat in Subjekt-Rolle, Zeichenkette des Kandidaten ist kein Zahl, oder: pronominaler Kandidat, Kongruenz in Numerus, Genus und Person
Reflexivpronomen *Personalpronomen-2* *Possessivpronomen-2*	Kongruenz der morphologischen Merkmale Numerus, Genus und Person

Abbildung 11.16: Kongruenzbedingungen für die unterschiedlichen Anapherotypen

behandelt werden solche Relativpronomen, die bereits während der syntaktischen Analyse qua Anknüpfung des Relativsatzes interpretiert wurden, sowie die ebenfalls auf der Grundlage der syntaktischen Struktur bereits resolvierten Nomen in Appositionen;³⁰ dies gilt ferner für die per Nominal-Konjunktion induzierten komplexen Vorkommen, die ebenfalls ausschließlich als Antezedens-*Kandidat* vom Typ C Berücksichtigung finden (vgl. Beleg (4)).

In analoger Weise werden die Vorkommen a priori hinsichtlich ihrer *Antezedens-Tauglichkeit* klassifiziert: Auf die Berücksichtigung von Typ-A-Pronomen wird verzichtet, da stets eine alternative lokale Realisierung des Diskursreferenten mit i.d.R. höherer Prominenz existiert.

Typspezifische Kongruenzbedingungen

Weitere Freiheitsgrade bestehen in bezug auf die Umsetzung des Algorithmus selbst. Im Hinblick auf die verfügbaren Ressourcen wird auf die Implementierung der Restriktion SEMROLLEN und des Präferenzkriteriums SEMPARALL verzichtet (Abbildung 9.9, Schritte 1c bzw. 2a). Des Weiteren verbleibt die typabhängige Anwendung geeigneter Kongruenzrestriktionen der Strategien MORKONGRU und ZEICHKETT auszugestalten (Schritt 1a). Als Bezugspunkt einer adäquaten Unterscheidung dient in diesem Fall der *Typ* des zu interpretierenden anaphorischen Vorkommens (**dr-okk**-Attribut **typ**). Abbildung 11.16 gibt eine Übersicht der für die unterschiedlichen Anapherotypen implementierten Bedingungen.

In den Bedingungen für Anaphern des Typs ***Nomen*** wird die Genuskongruenz gefordert, um eine inkorrekte Identifikation von Homonymen mit unterschiedlichem grammatischem Geschlecht zu unterbinden; die Kongruenz im Merkmal Person ergibt sich implizit und bedarf daher keiner Überprüfung. Die Verifikation der Kongruenzbedingungen für anaphorische Okkurrenzen des Typs ***Name*** wird durch die interne Verwaltung eines Verzeichnisses bekannter Namensvorkommen unterstützt, das eine beschleunigte Auffindung geeigneter Antezedenskandidaten ermöglicht. Als heuristisch kompatibel angesehen werden bereits Kandidaten, deren *Nachname* - formal definiert als kürzestes Suffix, das vom übrigen Teil der Namenszeichenkette durch ein Leerzeichen getrennt ist - mit der Zeichenkette der Anapher übereinstimmt:

³⁰Diese Antezedens-Entscheidungen werden bereits durch die Systemkomponente **dpe-anares** getroffen.

(20) *Peter Behrens besuchte Gropius. Behrens kannte den Bauhaus-Chef gut.*

Für Pronomen in dritter Person (Typen **Personalpronomen-3**, **Possessivpronomen-3**) findet keine Verifikation der Genuskongruenz statt, da Ausnahmen existieren, die nicht a priori blockiert werden sollen (vgl. Abschnitt 7.1.1, S. 111). Gemäß einer zusätzlichen Heuristik werden Antezedens-Kandidaten verworfen, deren Ausdruck eine Zahl repräsentiert (z.B. eine Uhrzeit "12:45") - im Rahmen der Evaluation wurde deutlich, daß ein pronominaler Wiederaufgriff derartiger Okkurrenzen äußerst selten ist. Vorkommen vom Typ **Relativpronomen**, die nicht bereits syntaktisch zugeordnet sind, werden ausschließlich satzintern und nichtkataphorisch sowie ferner unter der Bedingung der *uneingeschränkten* morphologischen Kongruenz resolviert, da eine Abweichung des Kandidaten in bezug auf den Genus aufgrund der Lokalität des Wiederaufgriffs ausgeschlossen werden kann.³¹ Am komplexesten gestalten sich die Restriktionen für pronominalen Okkurrenzen in erster Person (Typen **Personalpronomen-1**, **Possessivpronomen-1**). Neben uneingeschränkt morphologisch kongruenten pronominalen Kandidaten sind auch Dritte-Person-Vorkommen der Typen **Name** und **Nomen** zulässig, die in der syntaktischen Funktion des Subjekts realisiert sind und das Geschlecht Maskulinum oder Femininum besitzen. Diese Heuristik erweist sich in bezug auf typische Belege eingeschobener direkter Rede als nützlich, in denen das Antezedens eines Erste-Person-Pronomens außerhalb der direkten Rede in Gestalt einer natürlichen Person in Subjekt-Rolle realisiert ist (vgl. Abschnitt 11.1.3, Beleg (14)). Da sich die Gegebenheiten für plurale Erste-Person-Pronomen i.d.R. komplizierter gestalten, gelangt diese Teilstrategie ausschließlich im Falle des Singular zur Anwendung.

Typspezifische Prominenzfaktoren

Auch bezüglich der anzuwendenden Präferenzkriterien bedarf der Algorithmus einer näheren Ausgestaltung, denn die bisherige Beschreibung läßt offen, welche der implementierten Faktoren im jeweiligen Fall zur Anwendung gelangen sollen (Abbildung 9.9, Schritt 2a). Es ist eine Zuordnung zu treffen, durch die die spezifischen referentiellen Dispositionen der unterschiedlichen Typen anaphorischer Vorkommen möglichst gut abgebildet werden. Abbildung 11.17 zeigt die Festlegung der Faktorenmengen für die einzelnen Okkurrenztypen, die im ROSANA-System implementiert wurde. SYR steht für syntaktische Rollenträgheit (Strategie SYNPARALL), EEP für Präferenz der Existenz-Emphase³² (TOPIKALIS), SUP, PGP, DOP, IOP und APP für die hierarchisch abgestufte Präferenz von Subjekt, Possessivpronomen/sächsischem Genitiv/Genitivattribut (als logischem NP-Subjekt), direktem Objekt, indirektem Objekt und adverbialer (nicht nominalattributiver) Präpositionalphrase (SUBJEKT, SYNROLLE), KAM für den Malus kataphorischer Bezugnahmen (KATAMALUS), SDM und WDM für Satz- bzw. Wortdistanz an der Oberfläche (DISTANZ) - vgl. Abschnitt 7.2.2. Der Faktor PKM (Partialkongruenz-Malus) induziert die Abwertung von Antezedenskandidaten mit nichtgegebener Genus-Übereinstimmungen für diejenigen Anapherntypen, für die die entsprechende Kongruenzbedingung fakultativ ist; da ihm somit die Rolle einer Heuristik zur stringenten Strategie MORKONGRU zukommt, verkörpert er keinen Prominenzfaktor im engeren Sinne. Eine vergleichbare Funktion

³¹Vgl. jedoch die literarische Ausnahmeerscheinung des Dürrenmatt-Zitats in Fußnote 3, S. 111!

³²Die Bezeichnung wird von Lappin und Leass übernommen ("*existential emphasis*"), deren Untersuchungen betreffend den idealisierenden, nichtfragmentarischen Fall den Ausgangspunkt für die Wahl der Prominenzfaktoren von ROSANA darstellen (vgl. [LaLe94]).

Typ der Anapher	SYR	EEP	SUP	PGP	DOP	IOP	APP	KAM	SDM	WDM	PKM
Nomen								+	+		
Name								+			
Pers..pronomen-3	+	+	+	+	+	+	+	+	+		+
Pers..pronomen-2	+	+	+	+	+	+	+	+	+		
Relativpronomen										+	
Poss..pronomen-3	+		+					+	+		+
Poss..pronomen-2	+		+					+	+		
Reflexivpronomen											
Pers..pronomen-1			+					+		+	
Poss..pronomen-1			+					+		+	

Abbildung 11.17: Prominenzfaktoren für die unterschiedlichen Anapherntypen

erfüllt die Heuristik FRM der Plausibilitätsreduktion im Falle nicht-definitiver Verifikationen der Strategie SYNCONFIG im Fragmentfall (Fragmentmalus; Abbildung 10.10, Schritt 2a’); da dieser Faktor für alle Anapherntypen gleichermaßen relevant ist, wird er in der Tabelle nicht ausgewiesen.

Die Zuordnung der Faktoren zu den unterschiedlichen Anapherntypen sowie die Festlegung der entsprechenden numerischen Gewichte wurde auf der Grundlage von Testdurchläufen auf dem Trainingskorpus optimiert. Für den Grundtypus des Personalpronomen in dritter Person (***Personalpronomen-3***) kommt das Gros der Faktoren zur Anwendung; diese Festlegung, die dem Ansatz von Lappin und Leass zugrundeliegt ([LaLe94]), erweist sich auch im Kontext potentiell fragmentarischer syntaktischer Voranalyse-Ergebnisse des hier verwendeten Parsers als sinnvoll. Possessivpronomen in dritter Person (***Possessivpronomen-3***) werden abweichend behandelt, da in den weitaus meisten Fällen ein geeignetes Antezedens in der Rolle des Subjekts oder ebenfalls des Possessiv-Pronomen verfügbar ist; dementsprechend ist die Anwendung der Faktoren SUP bzw. SYR hinreichend. Entsprechend der für Pronomen in erster Person implementierten Kongruenzheuristik (vgl. Tabelle 11.16) wird die Anknüpfung an (üblicherweise den Sprecher der direkten Rede realisierenden) Subjekt-Okkurrenzen der Typen Name und Nomen bevorzugt, um eine nichtpronominale Verankerung außerhalb der direkten Rede zu fördern. Da Relativpronomen stets ein Bezugsnomen in der unmittelbaren Nachbarschaft attribuieren, ist WDM der einzige für den Typ ***Relativpronomen*** relevante Faktor. Die Klasse ***Reflexivpronomen*** stellt einen Sonderfall dar, in dem ganz auf die Anwendung von Faktoren verzichtet werden kann, da das Antezedens i.d.R. alleine aufgrund der bindungstheoretischen Lokalitätsbedingung eindeutig bestimmt ist.³³

Festlegung der Faktoren-Gewichte

Für den Einsatz der Faktoren zur *Bewertung* der Plausibilität von Antezedenskandidaten verbleiben *numerische Gewichte* entsprechend der relativen Vorhersagekraft der unterschiedlichen Heuristiken festzulegen. Einer der Ausgangspunkte ist die Vorgabe der Heuristik SYNROL-

³³Dies gilt für sämtliche Typ-A-Vorkommen des Evaluationskorpus. Jedoch lassen sich Belege angeben, in denen unter alleinigem Rekurs auf die in ROSANA realisierten Strategien ein Ambiguitätsfall entsteht. Obwohl solche Fälle von primär theoretischer Bedeutung zu sein scheinen, bietet es sich somit an, auch für Reflexiva/Reziproka die Strategie der Subjektpräferenz (SUBJEKT, Faktor SUP) bzw. - allgemeiner - der syntaktischen Rollenhierarchie (SYNROLLE) anzuwenden (vgl. auch die Arbeit von Primus zu den unterschiedlichen Faktoren, die die Akzeptabilität der Reflexivierung determinieren: [Prim89]).

Größenverhältnis	Beschreibung/Begründung
$SDM > SYR$	Syntaktische Parallelität bildet ausschließlich lokale Präferenzen; weiter zurückliegende Kandidaten werden nicht gegenüber näheren bevorzugt.
$SYR > SUP$	Kandidaten mit identischer syntaktischer Rolle sind plausibler als die Subjekt-Okkurrenz desselben Satzes.
$SYR, SUP > FRM$	Die Abwertung bei heuristischer Verifikation der Bindungsbedingungen fällt - verglichen mit den zentralen Präferenzheuristiken - gering aus.
$SUP \geq DOP \geq IOP \geq APP$	Vorgabe der Strategie SYNROLLE, grammatische Rollenhierarchie.
$EEP \geq SUP$	Topikalisierung ist das zentrale syntaktische Ausdrucksmittel der Fokussierung.

Abbildung 11.18: Bedingungen für die relative Größe der numerischen Gewichte

LE, derzufolge die Gewichte der Faktoren SUP, DOP, IOP und APP in absteigender Größe zu vergeben sind. Weitere Anhaltspunkte für die Ausgestaltung der Größenverhältnisse der Faktoren-Gewichte ergeben sich anhand der Vorarbeiten von Lappin und Leass für den nicht-fragmentarischen Fall ([LaLe94]). Die Optimierung der Gewichtszuordnungen geschieht auf der Grundlage des Trainingskorpus, wobei die Evaluationsergebnisse sowohl der KV- als auch der PS-Aufgabe als Zielgrößen dienen. Als *optimal* sollen alle Ergebniskombinationen gelten, bei denen ein Precision-Wert nur unter gleichzeitiger Verschlechterung des entsprechenden Recall-Werts erhöht werden kann bzw. umgekehrt; ein spezifischer Schwerpunkt hinsichtlich Precision oder Recall wird zunächst nicht gesetzt.

Es ist vorab zu betonen, daß es sich hierbei nicht um eine „*Optimierung*“ der algorithmischen Strategien zur Kospezifikations-Interpretation im engeren Sinne handelt. Die zum Einsatz gelangenden Auswahlheuristiken, auf die sich die getroffene Wahl der Prominenzfaktoren begründet, verkörpern *Ersatzverfahren* für inhaltlich adäquate Entscheidungsprozeduren, für die die geeigneten Verarbeitungsressourcen (noch) nicht zur Verfügung stehen. Entsprechend gestaltet sich der *Skopus* der durchzuführenden Optimierung. In Abwesenheit des *unmittelbaren* inhaltlichen Bezugs der Faktoren geht es weniger um die Ermittlung des - relativ zum Trainingskorpus - globalen Optimums, sondern primär um die empirische Bestimmung bzw. Bestätigung einiger heuristischer Regeln für die Ausgestaltung der Größenverhältnisse der numerischen Gewichte. Diese allgemeine Charakterisierung wird der Ausgangspunkt für die eigentliche Wahl absoluter Größen sein, wobei ein Ergebnis als *hinreichend gut* angesehen wird, wenn es keinen größeren Spielraum für weitere Verbesserungen läßt.

Abbildung 11.18 faßt die wesentlichen Größenverhältnisse zusammen, die sich unter Berücksichtigung theoretischer bzw. empirischer, anhand des Trainingskorpus validierter Gesichtspunkte ergeben. Die Relationen wurden durch gezielte Variationen der zugeordneten absoluten Gewichte verifiziert. Vergleichsweise eindeutig bestätigt wurde die Beziehung zwischen Fragmentmalus (FRM) und Subjektpräferenz (SUP) bzw. syntaktischer Rollenübereinstimmung (SYR) - wird für FRM ein Gewicht gewählt, das größer als sowohl SUP als auch SYR ist, so führt dies zu einer wesentlichen Verschlechterung der Ergebnisse. Vergleichbares gilt für das Größenverhältnis zwischen SYN und SUP. Problematisch gestaltet sich einzig die empirische Validierung der Relation $EEP \geq SUP$, die aufgrund der Seltenheit von Instanzen syntaktischer Topikalisierung weder bestätigt noch widerlegt werden konnte und deshalb heuristisch als $EEP = SUP$ umgesetzt wird. Für die grammatische Rollenhierarchie (inklusive dem Faktor PGP für logische NP-Subjekte) erweist sich die Implementierung mit $SUP > PGP > DOP > IOP = APP$ als optimal.

Es verbleiben die konkreten Gewichte festzulegen, die den identifizierten Größenrelationen genügen und zu einem (im obigen Sinne) hinreichend guten Ergebnis führen. Eine Werte-Zuordnung, die diesen Anforderungen entspricht und im ROSANA-System implementiert ist, wird in Ab-

SYR	EEP	SUP	PGP	DOP	IOP	APP	KAM	SDM	WDM	PKM	FRM
20	15	15	13	10	5	5	125	25	25	40	10

Abbildung 11.19: numerische Gewichte für die Prominenzfaktoren

bildung 11.19 gezeigt. Für den Kataphernmalus KAM zeigte es sich, daß dieser einfach nur hinreichend groß gewählt werden sollte. Hingegen sollte der Partialkongruenzmodus PKM nicht zu groß ausgestaltet sein, denn ansonsten hochplausible Kandidaten, denen im Rahmen der OV-Aufgabe auf heuristischem Wege ein inkorrektes Genusattribut zugewiesen wurde, sollten nicht zugunsten beliebig weit zurückliegender Alternativen verworfen werden.

Festlegung der nichtlokalen Plausibilitätsordnung

Auf der Basis der bisher diskutierten Faktoren wird eine interne Rangordnung der für anaphorische Okkurrenzen bestehenden Antezedens-Optionen hergestellt, die die Grundlage der lokalen Sortierung in Schritt 2b des Basisalgorithmus bildet (vgl. Abbildung 9.9, S. 179). Gemäß den Ausführungen in Abschnitt 9.3.2 besteht die Kernstrategie des Algorithmus zur heuristischen Auflösung von Entscheidungsinterdependenzen in der Bevorzugung der global plausibleren Entscheidungsalternative; die entsprechende globale Sortierung geschieht in Schritt 2c. Somit ist dafür zu sorgen, daß die kumulierte Plausibilitätsbewertung für solche Entscheidungsinstanzen (anaphorische Vorkommen) am höchsten ausfällt, die mit der größten Wahrscheinlichkeit korrekt resolviert werden.

Eine Analyse der Evaluationsergebnisse für das Trainingskorpus zeigt auf, daß sich eine adäquate Rangordnung unter Rekurs auf den *Typ* der zu resolvierenden Anaphern treffen läßt: Die größte Wahrscheinlichkeit einer korrekten Entscheidung ergibt sich für Reflexivpronomina und Namen; das Schlußlicht bilden nichtpronominale Nomen. Die Idee besteht nun darin, entsprechende *Basis-Plausibilitäten* für die unterschiedlichen Typen von Anaphern zu definieren, die für die interne Ordnung der jeweiligen Kandidaten ohne Auswirkungen bleiben, jedoch die globale Entscheidungsreihenfolge in der gewünschten Weise determinieren. Folgende Festlegungen erweisen sich als angemessen:

Faktor	Gewicht	Anapherntypen
REFLB	220	*Reflexivpronomen*
NAMVB	200	*Name*
NAMNB	180	*Name*
PO12B	160	*Possessivpronomen-1*, *Possessivpronomen-2*
APRNB	100	*Personalpronomen-3*, *Possessivpronomen-3*, *Relativpronomen*, *Personalpronomen-2*, *Personalpronomen-1*
NOMB	80	*Nomen*

Die feinere Unterscheidung für die Anaphernklasse NAME geschieht in Abhängigkeit der Entscheidungsgrundlage: Kandidaten, deren Zeichenkette vollständig mit der der Anapher übereinstimmen, werden als grundsätzlich plausibler (NAMVB) eingestuft als Kandidaten, die auf Basis der Nachnamen-Heuristik akzeptiert werden (NAMNB).

Schranke	# Sätze	Anapherntypen
P1-RS	7	*Personalpronomen-1*, *Possessivpronomen-1*
P2-RS	6	*Personalpronomen-2*, *Possessivpronomen-2*
P3-RS	5	*Personalpronomen-3*, *Possessivpronomen-3*
NO-RS	20	*Nomen*

Abbildung 11.20: typspezifische Begrenzung der satzübergreifenden Antezedenssuche

Schranken für die satzübergreifende Antezedenssuche

Aus praktischen Gründen erscheint es notwendig, den Skopus der Suche nach intersententiellen Antezedenskandidaten zu begrenzen - andernfalls würde der kumulierte Suchaufwand mit der Länge des Texts quadratisch wachsen.³⁴ Empirische Untersuchungen anhand des Trainingskorpus zeigen auf, daß der Verlust für praktische Anwendungen allenfalls marginal ausfällt: Für den weitaus größten Anteil der Personal- und Possessivpronomen in dritter Person existiert satzintern oder im unmittelbar vorausgehenden Satz ein geeignetes Antezedens; nur in wenigen Ausnahmefällen ist es notwendig, zwei Sätze zurückzuschauen. Aufgrund der Besonderheiten betreffend die Verarbeitung von eingeschobener direkter Rede gilt diese Aussage nicht für Pronomen in erster oder zweiter Person, doch auch hier ist letztlich zu beobachten, daß eine Begrenzung der Rückschau auf wenige Sätze bereits zu guten Precision-Recall-Kombinationen führt. Am kritischsten gestaltet sich die Suchraumbegrenzung im Falle nichtpronominaler Nomen bzw. Namen, für die auch größere Schranken zu einer (wenn auch geringfügigen) Reduktion des Recall - allerdings zugunsten der Precision - führen.

Versuche anhand des Trainingskorpus mit unterschiedlichen Schranken für die anapherntypspezifische Suche nach satzübergreifenden Antezedenten zeigen auf, daß ein Austauschverhältnis zwischen den jeweils erzielten Precision- und Recall-Werte besteht; insbesondere betrifft dies die Grenzwerte für die nichtpronominalen Okkurrenzen der Klassen *Nomen* und *Name*. Die Wahl einer bestimmten Kombination von Schranken ist somit i.d.R. nicht in jeder Hinsicht optimal, sondern geht mit einer bestimmten Ausrichtung der Interpretationsleistung einher. Nimmt man die Summe der in der Disziplin KV (Bestimmung der Kospezifikationsklassen) erzielten Precision- und Recall-Werte als Zielgröße, so führen die in Abbildung 11.20 spezifizierten Suchgrenzen zum maximalen Ergebnis. Für Reflexiv- und Relativpronomen sind keine Schranken festgelegt, da die Suche bereits durch die übrigen Restriktionen auf intrasententielle Kandidaten begrenzt wird. Auch für Namen wird der Suchraum nicht eingeschränkt; die potentiellen Antezedenten - ebenfalls Namen - werden in einem gesonderten Verzeichnis verwaltet, das eine effiziente, diskursglobale Suche unterstützt. Da in den zu interpretierenden Dokumenten i.d.R. nur eine vergleichsweise kleine Anzahl von Namen vorkommt, die oftmals mit weit entfernten Antezedenten kospezifizieren, die mit hoher Genauigkeit bestimmt werden können, erscheint diese Vorgehensweise sowohl unter Effizienzgesichtspunkten als auch aus inhaltlichen Gründen gerechtfertigt.

³⁴Daß diese Vorgehensweise unter theoretischen Gesichtspunkten inadäquat ist, wird mit Blick auf die in Abschnitt 7.1.4 diskutierte Theorie der globalen Fokussierung deutlich, derzufolge das Antezedens eines Pronomens - entsprechend dem vom attentionalen Zustand gesteckten Rahmen - prinzipiell beliebig weit zurückliegen kann. Dies scheint jedoch nur in Einzelfällen tatsächlich der Fall zu sein.

- Sei P eine pronominale Okkurrenz, für die ein nichtpronominales Substitut zu bestimmen ist; sei ferner $K(P)$ die Kospezifikationsklasse zu P , die als Ergebnis der Bearbeitung der KV-Aufgabe vorliegt, und A das spezifische Antezedens, über den der Anschluß an $K(P)$ hergestellt wurde.
 - Falls das durch P aufgegriffene Antezedens A nichtpronominal, d.h. vom Typ `*Nomen*` oder `*Name*` ist, so wähle A als Substitut.
 - Sonst: Wähle - sofern existent - diejenige nichtpronominale Okkurrenz L der Kospezifikationsklasse $K(P)$ als Substitut, die den geringsten Wortabstand zu P hat.

Abbildung 11.21: ROSANA-Algorithmus zur Ermittlung nichtpronominaler Substitute

Eine Strategie zur Ermittlung lexikalischer Substitute

Das Modul `anaphernresolution` stellt auch eine Funktion zur Lösung der PS-Aufgabe zur Verfügung, die auf den Kospezifikationsklassen aufbaut, die als Ergebnis der Bearbeitung der KV-Aufgabe vorliegen. Da eine Kospezifikationsklasse mehr als einen nichtpronominalen Repräsentanten aufweisen kann, ist das lexikalische Substitut eines Pronomens nicht in jedem Fall eindeutig bestimmt. Somit ist eine geeignete Auswahlstrategie zu formulieren.

Das ROSANA-System bedient sich des in Abbildung 11.21 dargestellten Auswahlalgorithmus. Erster Kandidat ist das für das jeweilige Pronomen gewählte Antezedens, über den der Anschluß zu der entsprechenden Kospezifikationsklasse hergestellt wurde. Ist diese Okkurrenz nichtpronominal - d.h. vom Typ `*Nomen*` oder `*Name*` -, so wird sie als Substitut des Pronomens gewählt.³⁵ Wird hingegen ein Pronomen wiederaufgegriffen, so werden die übrigen Vertreter der Klasse in den Auswahlprozeß einbezogen; in diesem Fall gewinnt der nichtpronominale Kandidat mit dem *geringsten Oberflächenabstand* (gemessen in der Anzahl intervenierender Wörter) zum zu resolvierenden Pronomen. In problematischen Fällen, in denen eine Klasse keinen nichtpronominalen Vertreter hat, wird keine Zuordnung vorgenommen.

Die Motivation dieser Strategie ergibt sich aus Plausibilitätserwägungen. Je weiter das gewählte Substitut in der der Kospezifikationsklasse zugrundeliegenden Wiederaufgriffskette zurückliegt, desto größer ist das ausschlaggebende *kumulierte* Risiko einer Fehlentscheidung. Am plausibelsten ist demnach das unmittelbare Antezedens des betrachteten Pronomens; für die übrigen Kandidaten ergibt sich eine angemessene heuristische Rangordnung in Gestalt der Oberflächenabstand. Das beschriebene Vorgehen kann dazu führen, daß verschiedene Pronomen derselben Klasse unterschiedliche Substitute zugewiesen bekommen. Im Hinblick auf das in Abschnitt 11.1.4 definierte Evaluationsszenario für die PS-Aufgabe sollte die beschriebene Strategie jedoch zu den besten Ergebnissen führen.

11.2.4 Bewertung der systemgenerierten Ergebnisse

Ein gesondertes Modul `bewerter` (2556 Zeilen LISP-Code) bewerkstelligt die Berechnung der in Abschnitt 11.1.4 für die drei Evaluationsdisziplinen definierten Precision- und Recall-Maße. Die systemgenerierten Ergebnisse werden auf der Grundlage ihrer internen Repräsentationen

³⁵Als zulässige Substitute angesehen werden auch *Konjunktionen* der angegebenen Vorkommens-Typen (vgl. o., Beleg (4)), die als Okkurrenzen eines gesonderten Typs `*Koord*` verwaltet werden.

verarbeitet. Die Vorgaben des Ergebnisschlüssels liegen in einer `.key`-Datei vor, in der die zu generierenden Kospezifikationsklassen sowie die relevanten Okkurrenzen spezifiziert sind; optionale Kospezifikationsbeziehungen sind in geeigneter Weise ausgezeichnet.

Es existieren unterschiedliche Konfigurationsoptionen, durch die die durchzuführenden Evaluationen sowie der Detailliertheitsgrad der zu erzeugenden Ausgaben spezifiziert werden. Das Bewertungsmodul realisiert somit nicht alleine die Berechnung der einschlägigen numerischen Gütemaße, sondern es ermöglicht darüberhinaus die Generierung ausführlicherer Beschreibungen, auf deren Grundlage die bestehenden Problemfälle der Interpretationsalgorithmen identifiziert und analysiert werden können. In Ergänzung der beschriebenen Evaluationsdisziplinen wird beispielsweise eine nähere Aufschlüsselung der Precision ermöglicht, mit der die individuellen Anknüpfungsentscheidungen für die einzelnen Anapherntypen getroffen werden. Dies erlaubt die Identifikation interpretationstechnisch problematischer Okkurrenztypen - eine Aufgabe, die alleine auf der Basis der für die KV- und PS-Aufgaben ermittelten Evaluationsergebnisse nicht lösbar wäre. Somit verkörpert das Modul **bewertung** darüberhinaus ein nützliches Instrument für die Suche nach Ansätzen zur Verfeinerung der implementierten Algorithmen.

11.3 Evaluation

Das System ROSANA soll nun auf den in Abschnitt 11.1.2 beschriebenen Textsammlungen evaluiert werden. Den geeigneten Ausgangspunkt bilden Schlüssel-Spezifikationen, die für die beiden Textkorpora *“Presse-Meldungen”* und *“Mozart-Opern”* entsprechend dem in Abbildung 11.1 skizzierten Szenario per intellektueller Annotation erstellt wurden.³⁶ Die Ergebnisse der Annotation fließen in Form von `.key`-Dateien in den Evaluationsprozeß ein (vgl. Abbildung 11.8).

11.3.1 Überblick der Ergebnisse für das Korpus *“Presse-Meldungen”*

Abbildung 11.22 zeigt die Ergebnisse, die ROSANA auf dem Evaluationskorpus (Dokumente 32 bis 66) der Textsammlung *“Presse-Meldungen”* in den drei Evaluationsdisziplinen OV, KV und PS erzielt. (Zur Bearbeitung der OV-Aufgabe wurde ein NGP-Lexikon mit insgesamt 79 Einträgen, umfassend 63 Ländernamen und 16 sonstige Entitäten, sowie ein Namens-Lexikon mit 7 Einträgen (flektierte Formen einiger Namen prominenter Politiker) zugrundegelegt (vgl. S. 248 ff.)) Demnach wird die Aufgabe der Bestimmung objektspezifizierender Vorkommen mit einer Precision P_{ov} von 0.9404 und einem Recall R_{ov} von 0.9623 gelöst. Die Identifikation der Äquivalenzklassen der Kospezifikationsrelation (KV-Aufgabe) geschieht mit $P_{kv} = 0.8081$ und $R_{kv} = 0.6845$. Für die Aufgabe der Bestimmung nichtpronominaler Substitute (PS-Aufgabe) lauten die Werte $P_{ps} = 0.7009$ und $R_{ps} = 0.6532$ (Durchschnittswerte); die Tabelle schlüsselt die Ergebnisse in bezug auf die unterschiedlichen Typen von Pronomen näher auf.³⁷ In Anhang A wird in bezug auf ein charakteristisches Dokument des Evaluationskorpus ausführlich erläutert,

³⁶Die Annotation geschah mit Hilfe eines auf dem Texteditor EMACS basierenden Software-Werkzeugs, das für diese Aufgabe eigens entwickelt wurde.

³⁷Vgl. S. 226: PE-3/PE12 = Personalpronomen in dritter bzw. erster/zweiter Person, PO-3/PO12 = Possessivpronomen in dritter bzw. erster/zweiter Person, REFL = Reflexiv- und Rezipropronomen, RELA = Relativpronomen.

EVALUATIONSERGEBNIS OKKURRENZEN:

- NUR ANA: 243
 - NUR KEY: 150
 - ANA UND KEY: 3831
 => PRECISION: 0.9404
 => RECALL: 0.9623

A. R. -ERGEBNIS-PARTITIONIERUNG

- SCHNITTE: 256
 - MOEGLICH: 1334
 => PRECISION: 0.8081

KEY-PARTITIONIERUNG

- SCHNITTE: 496
 - MOEGLICH: 1572
 => RECALL: 0.6845

DURCHSCHNITT JE DOKUMENT

=> PRECISION: 0.8191
 => RECALL: 0.6659

LEXIKALISCHE SUBSTITUTION

	PRECIS	RECALL	++	+ -	+ ?	+ _	+ *	? +	? _
PE-3	0.6766	0.6667	136	54	11	3	0	18	0
PE12	0.9091	0.3846	10	1	0	15	6	0	0
PO-3	0.6641	0.6641	87	39	5	0	0	0	0
PO12	1.0000	0.5000	2	0	0	2	1	0	0
REFL	1.0000	0.7500	3	0	0	1	0	0	0
RELA	0.7667	0.6832	69	18	3	11	0	7	4

DURCHSCHNITT JE PRONOMEN

=> PRECISION: 0.7009
 => RECALL: 0.6532

Abbildung 11.22: Ergebnisse für das Evaluationskorpus der Textsammlung “Presse-Meldungen”

wie die Ergebnisse der drei Evaluationsdisziplinen durch ROSANA per satzweiser Analyse generiert werden, welche Gestalt sie haben und wie die zugehörigen Precision- und Recall-Werte berechnet werden. Im Speziellen werden die Ausführungen die Arbeitsweise des zugrundeliegenden Algorithmus (Abbildung 10.10) illustrieren.

Gegenüber den Ergebnissen, die auf dem Trainingskorpus (Dokumente 1 bis 31) erzielt werden, ergeben sich für die Aufgabenstellungen OV und KV nur marginale Abweichungen: $P_{ov} +0.26$ Prozentpunkte (Trainingskorpus: 0.9378), $R_{ov} -0.35$ (TK: 0.9658), $P_{kv} -0.08$ (TK: 0.8089), $R_{kv} -0.53$ (TK: 0.6898). Deutlich schlechter ist hingegen der Precision-Wert für die PS-Disziplin: $P_{ps} -4.73$ (TK: 0.7482), $R_{ps} -0.49$ (TK: 0.6581). Angesichts der weitgehenden Übereinstimmung der KV-Werte für die beiden Teilkorpora und der vergleichsweise schlichten Ausgestaltung des PS-Algorithmus (Abbildung 11.21), der keiner (potentiell korpuspezifischen) Feinabstimmung unterworfen wurde, sondern auf den für die KV-Disziplin generierten Ergebnissen unmittelbar

aufbaut, läßt diese Abweichung den Schluß zu, daß ein systematischer Unterschied zwischen Trainings- und Evaluationskorpus besteht, durch den sich die PS-Aufgabe in bezug auf letzteren Korpus schwieriger gestaltet.

Die Streuung der dokumentenindividuellen Ergebnisse ist für die KV-Aufgabe geringer als für die PS-Aufgabe und für die Precision-Maße geringer als für die entsprechenden Recall-Maße (vgl. u.).

Das ROSANA-System wurde in *ALLEGRO Common Lisp, Version 4.3 for LINUX* implementiert. Die Laufzeit für die Analyse des Evaluationskorpus beträgt auf einem PC (Pentium-Prozessor, 133 MHz) ca. 78 Sekunden, falls ausschließlich die anwendungsrelevanten Ergebnisse (Anaphern und zugehörige Antezedenten, Äquivalenzklassen der Kospezifikationsrelation sowie nichtpronominale Substitute für Pronomen) generiert werden und kein Abgleich mit dem Schlüssel stattfindet. Hiervon entfallen ca. 48.9 % auf die Vorverarbeitung (16.6 % für Einlesen, 31.1 % für Extrahieren der Ergebnisse der vorgelagerten morphologischen und syntaktischen Analyse (.mor- und .pos- Dateien) inkl. OV-Aufgabe, 1.2 % für Rekonstruktion der phrasenstrukturellen Oberflächenstruktur-Beschreibung) und ca. 51.1 % auf die eigentliche Anapherninterpretation (11.7 % KV-Aufgabe, ≤ 4.1 % PS-Aufgabe, ≥ 35.3 % Schreiben der Ergebnisse). Mit ≥ 51.9 % ist somit das Lesen der Eingaben sowie das Schreiben der Ergebnisse der bestimmende Faktor. Entscheidend ist die insgesamt hohe Geschwindigkeit der Analyse. Auf dem Evaluationskorpus, der 12904 Wörter umfaßt, wird eine *Verarbeitungsrate von 165 Wörtern pro Sekunde* erzielt. Die Verarbeitungsgeschwindigkeit liegt somit annähernd in der Größenordnung des DPE-Parsers, womit die Voraussetzungen zur Verarbeitung umfangreicherer Korpora als erfüllt zu betrachten sind. Die Geschwindigkeit sollte sich noch weiter steigern lassen, da das Potential für die Optimierung der zugrundeliegenden Datenstrukturen und bestimmter Komponenten-Algorithmen nicht vollständig ausgeschöpft wurde.

Entscheidend in bezug auf die Anwendungstauglichkeit ist jedoch primär die Qualität der Interpretationsergebnisse. Im folgenden sollen daher die inhaltlichen Gesichtspunkte der Evaluation der drei Einzeldisziplinen in den Vordergrund rücken. Wie *gut* sind die Ergebnisse der algorithmischen referentiellen Interpretation durch ROSANA? Welche Fehler werden gemacht? An welchen Stellen besteht Raum für Verfeinerungen? Welche Ergebnisse sind auf der Basis der verwendeten Ressourcen prinzipiell erzielbar?

11.3.2 Diskussion der Ergebnisse für die OV-Aufgabe

Mit einer Precision von 0.9404 und einem Recall von 0.9623 liegen die Evaluationsergebnisse im erwarteten hohen Bereich. Gleichwohl besteht jedoch ein Spielraum für weitere Verfeinerungen, so daß ein genauerer Blick auf die Ursachen der entsprechenden Fehlentscheidungen lohnt.

Eine Analyse der *Precision-Fehler* des ROSANA-Systems bezüglich der Disziplin der Bestimmung objektspezifizierender Vorkommen zeigt auf, daß ca. 50 % der Okkurrenzen ohne Schlüsselentsprechung aufgrund von Fehlern in den vorgelagerten Verarbeitungsschritten der syntaktischen oder morphologischen Analyse generiert werden; hierbei überwiegen die Fälle, in denen eine (im Hinblick auf den syntaktischen Kontext) falsche Wortart - insbesondere Nomen anstelle von Adjektiv - zugeordnet wird oder das Parsing zusammengesetzter NP scheitert. Weitere 40 % der *Precision-Fehler* lassen sich auf Unvollständigkeiten in den Algorithmen des für die OV-Aufgabe zuständigen Vorverarbeitungsmoduls *dpe-anares* von ROSANA zurückführen. Etwa die Hälfte dieser Fehler könnte durch geeignete Verfeinerungen dieses Programmteils eliminiert werden. Problematisch sind NP in quasi-nichtreferentieller Verwendungsweise (z.B. "office" in

“*he took office*”), da ROSANA nicht über die notwendigen Ressourcen für deren Erkennung verfügt. Immerhin ca. 10 % aller Fälle sind auf irrtümliche Auslassungen in der Schlüssel-Spezifikation zurückzuführen und somit keine *Precision-Fehler* im eigentlichen Sinne.

Die *Recall-Fehler* von ROSANA sind sogar zu ca. 80 % den genannten typischen Fehlern der syntaktischen oder morphologischen Voranalyse zuzuschreiben. Die übrigen Probleme lassen sich in ROSANA lokalisieren; meist kann jedoch über eine elementare Verfeinerung der Algorithmen Abhilfe geschaffen werden.

Somit folgt, daß ein gewichtiger Anteil der Problemfälle seinen Ursprung außerhalb von ROSANA hat. Möglichkeiten zur Verbesserung der Systemperformanz bestehen primär auf der Seite der Precision; daß die Precision-Leistung hinter der Recall-Leistung zurückbleibt, ist auf die implementierten wortartenbezogenen Strategien zur Erkennung von Vorkommens-Induktoren zurückzuführen, die tendenziell *übergenerieren* und noch um geeignete Ausnahmeregeln zu erweitern sind, die sich erst auf der Basis der Analyse größerer Textmengen schrittweise vervollständigen lassen.

Aus der Perspektive typischer Anwendungen interessiert in erster Linie die Performanz von ROSANA in bezug auf die *Extraktion pronominaler Okkurrenzen*. Mit einer Precision von $\frac{477}{477+29} = 0.9427$ und einem Recall von $\frac{477}{477+2} = 0.9958$ (479 gesuchte Pronomen, 29 Precision- und 2 Recall-Fehler) wird diese Teilaufgabe besser bewältigt als die generelle Problemstellung der OV-Aufgabe. Speziell bezüglich des Recall scheint das Optimum bereits erreicht: Die beiden Fehler entstehen durch zwei Vorkommen des *Relativpronomens* “*that*”, die in der syntaktischen Voranalyse fälschlicherweise als Einleiter von Komplementsätzen kategorisiert und folglich von ROSANA als vorkommens-irrelevant interpretiert werden. Seitens der Precision-Fehler ergibt sich ein vielfältigeres Bild. Die übergenerierten Pronomen fallen in die zwei Klassen “*it*”-Pronomen (18 irrelevante Okkurrenzen) und Relativpronomen (11 irrelevante Okkurrenzen).³⁸ Von den 18 “*it*”-Vorkommen sind 15 referentiell irrelevante Korrelate oder Expletiva (vgl. S. 230, Beispiele (5c)), deren Elimination sich schwierig gestaltet, da diese nur auf der Basis einer detaillierteren Analyse des syntaktischen Kontexts geschehen kann,³⁹ für die übrigen 3 Fälle von “*it*”-Okkurrenzen mit relationalem Bezug existiert bislang ebenfalls noch kein formales Erkennungskriterium in ROSANA. Von den 11 als inkorrekt erkannten Relativpronomen sind 6 auf syntaktische Fehlanalysen zurückzuführen, 4 spezifizieren relationale Entitäten und eines entspricht einer irrtümlichen Auslassung im Schlüssel. Ein Spielraum für Verbesserungen erscheint zumindest für den Problemfall nichtreferentieller “*it*”-Verwendungsweisen zu bestehen, womit sich ca. 50 % der Precision-Fehler beseitigen ließen. Die Systemleistung für die pronomenbezogene OV-Aufgabe läge damit dicht unter dem auf der Grundlage der verfügbaren Voranalyse-Ressourcen erzielbaren Optimum.

11.3.3 Diskussion der Ergebnisse für die KV-Aufgabe

Mit einer Precision von 0.8081 und einem Recall von 0.6845 erfüllt die Performanz des ROSANA-Systems auch in der Disziplin der Bestimmung der Kospezifikations-Äquivalenzklassen die Erwartungen. Eine Gegenüberstellung mit den Werten, die die besten der in Abschnitt 8.8 diskutierten Teilnehmersysteme der “*Coreference Task*” von MUC-6 erzielen - (*P, R*)-Kombinationen

³⁸Vgl. Abbildung 11.22, Tabelle, Spalten ?+ und ?-.

³⁹Zwar vergibt der DPE-Parser eine spezielle Markierung für derartige *formale* Subjekte, jedoch zumindest für die entsprechenden “*it*”-Vorkommen mit unzureichendem Erfolg; nichtreferentielle “*there*”-Okkurrenzen in entsprechender Rolle können hingegen mit hundertprozentigem Erfolg eliminiert werden.

ANKNUEPFUNGS-ENTSCHEIDUNGEN

		PRECIS	++	+-	+?	+ ₋	++	?+	? ₋
PRON	PE-3	0.7143	145	48	10	1	0	18	0
	PE12	0.9474	18	1	0	7	6	0	0
	P0-3	0.7634	100	28	3	0	0	0	0
	P012	1.0000	3	0	0	1	1	0	0
	REFL	1.0000	3	0	0	1	0	0	0
	RELA	0.7789	74	18	3	6	0	7	4
		0.7555	343	95	16	16	7	25	4
NOMN	VNOM	0.7014	357	136	16	1973	43	31	133
	NAME	0.9390	308	15	5	368	5	5	28
		0.7945	665	151	21	2341	48	36	161

DURCHSCHNITT JE ANAPHER

=> PRECISION: 0.7808

Abbildung 11.23: nach Okkurrenztypen aufgegliederte Precision-Werte der Einzelentscheidungen

von (0.63, 0.63), (0.72, 0.59) bzw. (0.63, 0.55) / (0.72, 0.63) -, ergibt einen ersten Anhaltspunkt für die vergleichsweise hohe Qualität der erzielten Interpretationsleistung (0.81, 0.68) und somit für die Richtigkeit der dem Kernalgorithmus von ROSANA zugrundeliegenden Designentscheidung, auf einer möglichst vollständigen, jedoch potentiell fragmentarischen syntaktischen Analyse aufzubauen.⁴⁰

Die *Streuung* der dokumentenbezogenen Einzelwerte des Maßes P_{kv} fällt wesentlich geringer aus als die Streuung des Maßes R_{kv} . Ersteres nimmt auf dem Evaluationskorpus Werte zwischen 0.7037 und 1.0 ein. Letzteres bewegt sich in einem breiteren Korridor zwischen 0.4286 und 0.8333, wobei allerdings mehr als 80 % der Werte oberhalb der Marke 0.6 liegen; die Ausreißer entstehen für kurze Dokumente mit einem vergleichsweise geringen Anteil an Pronomen - ein Indiz dafür, daß Recall-Fehler vornehmlich für nichtpronominale Anaphern begangen werden.

Eine genauere Analyse der Systemperformanz sowie der typischen Precision- und Recall-Fehler, die ROSANA im Rahmen der Kospezifikationsklassen-Konstruktion unterlaufen, ist erst auf der Grundlage einer näheren Aufschlüsselung der Qualität der individuellen Antezedensentscheidungen für die unterschiedlichen Typen anaphorischer Okkurrenzen möglich. Abbildung 11.23 zeigt eine entsprechende, durch das Bewertungsmodul automatisch generierte Aufstellung. In den Spalten der Tabelle erfolgt eine Klassifikation der Entscheidungen gemäß den Festlegungen in Abbildung 11.7, wobei zusätzlich nichtpronominale Okkurrenzen (Namen (NAME) und gewöhnliche NP (VNOM)) Berücksichtigung finden.

Die Berechnung des Precision-Werts für die individuellen Antezedensentscheidungen geschieht gemäß der Spezifikation des Maßes P_{ps} als $\frac{|o_{++}|}{|o_{++}|+|o_{+-}|+|o_{+?}|}$. Anhand der Aufschlüsselung wird deutlich, daß Reflexiv- und Reziprokpronomina (Klasse REFL), für die Bindungsprinzip A eine stringente syntaktische Grenze für die Antezedenssuche absteckt, mit optimaler Genauigkeit

⁴⁰Ein direkter Vergleich der gemessenen Werte mit den Ergebnissen der MUC-Teilnehmer erscheint aus mehreren Gründen inadäquat: (1) anderes Evaluationskorpus, (2) geringfügig abweichende Aufgabedefinition, (3) separate Evaluation der OV-Aufgabe in der hier durchgeführten Evaluation, was zu einer geringfügigen Verbesserung der Ergebnisse für die KV-Aufgabe führt.

resolviert werden; ähnlich erfolgreich erweisen sich die Strategien für Pronomen in erster und zweiter Person, was u.a. darauf zurückzuführen ist, daß in ein und demselben Kontext niemals mehr als ein Bezugsobjekt mit dem entsprechenden Kongruenzmerkmal spezifiziert wird. Auch Namen werden mit sehr guter Präzision (94 %) zugeordnet, womit ein entsprechendes Ergebnis der MUC-6-Evaluation untermauert wird. Am unteren Ende der Skala, jedoch noch immer oberhalb von 70 Prozent, liegen Possessivpronomen und sonstige Pronomen in dritter Person, für die in der Regel eine höhere Entscheidungsambiguität besteht, da die auswahlrelevante Menge der Antezedenskandidaten größer ist, sowie gewöhnliche NP, deren Attributinformation derzeit nicht zur näheren Eingrenzung der korrekten Antezedenten herangezogen wird.

Erste Anhaltspunkte für Recall-Fehler ergeben sich anhand der Zahlen in den Spalten +-, +? und +_. Die auf den ersten Blick hohen Werte für *nichtpronominale* Okkurrenzen (VNOM, NAME) bedürfen jedoch einer sensiblen Interpretation. Jede Koreferenzklasse hat stets eine erste, in der Regel *nichtpronominale* Okkurrenz, die den Diskursreferenten *einführt*. Somit sind die insgesamt 2513 entsprechenden Vorkommen ohne korrektes Antezedens keineswegs allesamt als Induktoren von KV-Recall-Fehlern einzustufen. Da der Schlüssel für das Evaluationskorpus ca. 2300 Diskursreferenten spezifiziert, bewegt sich diese Zahl zumindest annähernd in der richtigen Größenordnung; die für die Klasse VNOM verbleibenden Recall-Fehler erklären sich u.a. dadurch, daß ROSANA gegenwärtig ausschließlich einen vergleichsweise elementaren Kongruenztest anwendet, durch den Instanzen von Wiederaufgriffen durch nicht zeichenkettenidentische Ausdrücke nicht erkannt werden. Andererseits führt jede inkorrekte Zuordnung einer nicht diskursreferent-einführenden Okkurrenz zu einem Recall-Fehler da jeweils nur *eine* Anknüpfungsentscheidung getroffen wird; ist diese falsch, so wird auch die entsprechende Schlüssel-Klasse unvollständig reproduziert. Somit führen praktisch alle Fehlentscheidungen und Auslassungen für pronominale Vorkommen zu einer Verminderung des Recall.

Detailstudien zu inkorrekten ROSANA-Entscheidungen für *Pronomen* zeigen auf, daß ein relativ großer Anteil der Fehler (ca. 30 %) auf die heuristische Zuordnung inkorrektur Genusmerkmale während des Anlegens der Vorkommensbeschreibungen zurückzuführen ist und durch entsprechende Erweiterung des NGP-Lexikons eliminiert werden könnte. Weitere 30 % der Fehlentscheidungen konstituieren die Klasse derjenigen Fälle, die den Horizont der faktorenbasierten Auswahlheuristiken von ROSANA übersteigen und deren Resolution einen Rekurs auf semantisches oder pragmatisches Hintergrundwissen bedingt (vgl. das Beispiel in Anhang A.2.6, S. 295). Die übrigen Fälle sind entweder auf irrtümliche Auslassungen im Schlüssel zurückzuführen (d.h. die Entscheidung ist eigentlich korrekt), oder sie entstehen aufgrund des nicht hundertprozentigen Abdeckungsgrads des Vorverarbeitungsmoduls bzw. aufgrund einer inkorrekten syntaktischen Analyse. In ca. 80 % aller Fälle steht das gesuchte Antezedens in der Auswahlliste der Kandidaten an Position Zwei. Für *nichtpronominale Okkurrenzen* sind fehlerhafte bzw. nichtvollzogene Entscheidungen praktisch ausschließlich auf die Nichtverfügbarkeit von Hintergrundwissen bzw. auf die Nichtrealisierung algorithmischer Heuristiken zur Berücksichtigung referentiell differenzierender Nominalattribute (Adjektive, Possessivmarker, PPs) in ROSANA zurückzuführen.

Aus der Diskussion läßt sich ableiten, daß eine Performanzerhöhung in der KV-Aufgabe primär für Pronomen in dritter Person sowie für nichtpronominale Nomen (Klasse VNOM) erstrebenswert ist. Anhand der Ergebnisse wird jedoch ferner deutlich, daß die Möglichkeiten der gegenwärtig in ROSANA eingesetzten Techniken begrenzt sind (vgl. u.). Andererseits erreicht die Leistung von ROSANA mit einer durchschnittlichen Präzision der Antezedensentscheidungen von 0.7808 auf einem vergleichsweise schwierigen Textkorpus bereits ein hohes Niveau; eine Gegenüberstellung mit der im Rahmen der MUC-6-Konferenz bezüglich der "Coreference Task"-Annotation ermittelten Intercoder-Übereinstimmung von ca. 81 % untermauert diese Einschätzung.

11.3.4 Diskussion der Ergebnisse für die PS-Aufgabe

Auf der Basis eines Vergleich der Precision-Werte der PS-Disziplin für die unterschiedlichen Pronomentypen (vgl. die Tabelle in Abbildung 11.22) mit den Precision-Werten der entsprechenden Einzelentscheidungen (Abbildung 11.23) ergeben sich Anhaltspunkte dafür, daß die Aufgabe der Bestimmung nichtpronominaler Substitute schwieriger ist als die Wahl eines (beliebigen) korrekten Antezedens. Für Possessivpronomen in dritter Person (Klasse PO-3) fällt die Differenz mit ca. 10 Prozentpunkten (0.6641 gegenüber 0.7634) am deutlichsten aus. Diese Tendenz läßt sich durch folgende Gegenüberstellung untermauern: Für 306 Pronomen sind sowohl die Anknüpfungsentscheidung als auch das ermittelte lexikalische Substitut korrekt; jedoch existieren 21 weitere Fälle, in denen zwar ein korrektes Antezedens, nicht jedoch ein korrektes Substitut ermittelt wird, und in nur einem Fall erweist sich zwar das Substitut, nicht jedoch das unmittelbare Antezedens als richtig.⁴¹

Mit Blick auf die Entscheidungsgrundlage des in Abbildung 11.21 beschriebenen Algorithmus zur Ermittlung nichtpronominaler Substitute wird der beschriebene Sachverhalt erklärbar. Die systemgenerierte Kospezifikationsrelation zwischen Pronomen und Substitut, auf die Bezug genommen wird, kommt oftmals durch eine *Kette* von Antezedensentscheidungen zustande, die eine Länge größer als 1 hat; die Gegebenheit, daß das *unmittelbare* Antezedens eines Pronomens nichtpronominal ist, stellt einen Spezialfall dar. Somit steigt die Wahrscheinlichkeit, daß der ermittelte nichtpronominale Anker inkorrekt ist, da es für einen Fehler hinreichend ist, wenn *irgendeine* der einfließenden Entscheidungen inkorrekt ist. Im folgenden ist ein entsprechendes Fallbeispiel aus Dokument Nr. 32 skizziert, in dem nur die erste der vier Einzelentscheidungen, stets jedoch das ermittelte Substitut inkorrekt ist, das sich zu "*commentator philippe alexandre*" ergibt:

(21) *commentator philippe alexandre* \leftarrow^- *he* \leftarrow^{++} *he* \leftarrow^{++} *him* \leftarrow^{++} *him*

Eine zweite Ursache läßt sich durch eine weitere Aufschlüsselung der Precision der Einzelentscheidungen für Pronomen hinsichtlich des Typs des jeweils gewählten Antezedens ausmachen: Während sich pronominalen Antezedenten zu 85.6 % als korrekt erweisen, wird für Antezedenten der Klassen VNOM und NAME nur eine Genauigkeit von 71.3 % erreicht. Einerseits erklärt sich dies durch die teils inkorrekte, da heuristische Zuordnung morphologischer Kongruenzmerkmale, von der ausschließlich die nichtpronominalen Okkurrenzen betroffen sind. Ein weiterer Grund läßt sich auf der fokustheoretischen Ebene ausmachen. Die Diskursreferenten von pronominalen Okkurrenzen sind bereits fokussiert - sonst wären sie nicht pronominal realisiert -; ein nachfolgendes Pronomen instanziiert somit mit hoher Wahrscheinlichkeit denselben Diskursreferenten (vgl. Abschnitt 7.1.4, R1 der Centering-Theorie). Im Rahmen der Auswahl unter mehreren möglichen *nichtpronominalen* Antezedenskandidaten steht dieses Kriterium nicht zur Verfügung.

Die durchschnittlichen Precision- und Recall-Werte betragen 0.7009 % bzw. 0.6532. Die dokumentenspezifische Streuung ist größer als im Falle der KV-Aufgabe: P_{ps} liegt zwischen 0.3 und 1.0 mit mehr als zwei Drittel aller Werte oberhalb von 0.6; auch R_{ps} nimmt Werte zwischen 0.3 und 1.0 an, wobei jedoch lediglich 60 % aller Ergebnisse oberhalb der Marke 0.6 lokalisiert sind. Für weniger als 10 % bzw. 15 % der Dokumente wird die 50%-Marke unterschritten.

Abbildung 11.24 zeigt die Ergebnisse für die unter Anwendungsgesichtspunkten aussagekräftigen

⁴¹Letzterer Ausnahmefall entsteht durch eine Kette falscher Antezedensentscheidungen, deren negative Effekte sich hinsichtlich der PS-Aufgabe gegenseitig neutralisieren.

	A-PREC	A-RECA
PE-3	0.6210	0.6667
PE12	0.9091	0.3846
PO-3	0.6641	0.6641
PO12	1.0000	0.5000
REFL	1.0000	0.7500
RELA	0.7113	0.6699
DURCHSCHNITT JE PRONOMEN		
=> PRECISION: 0.6631		
=> RECALL: 0.6504		

Abbildung 11.24: Evaluationsergebnisse für die Anaphernresolution im engeren Sinne

Precision- und Recall-Maße P_{ps}^a bzw. R_{ps}^a , in die die Fehler auf den Ebenen sowohl der OV- als auch der PS-Aufgabe einfließen.⁴² Demnach löst ROSANA das Problem der Anaphernresolution im engeren Sinne mit einer Precision- und Recall-Performanz oberhalb von 65 %.

11.3.5 Überblick der Ergebnisse für das Korpus “Mozart-Opern”

Abbildung 11.25 faßt die Ergebnisse zusammen, die ROSANA auf den drei Texten des Korpus “Mozart-Opern” erzielt. (Zur Bearbeitung der OV-Aufgabe wurde ein NGP-Lexikon, umfassend die 27 (Vor-)Namen der Figuren der drei Opern (sowie ferner, wie oben, 63 Ländernamen), zugrundegelegt. Im Übrigen kommt ROSANA in derselben Konfiguration wie für die Textsammlung “Presse-Meldungen” zur Anwendung.) Mit Werten von $P_{ov} = 0.9530$ und $R_{ov} = 0.9759$ (OV-Aufgabe), $P_{kv} = 0.8798$ und $R_{kv} = 0.8117$ (KV-Aufgabe) sowie $P_{ps} = 0.7468$ und $R_{ps} = 0.7436$ bzw. $P_{ps}^a = 0.7250$ und $R_{ps}^a = 0.7373$ (PS-Aufgabe) liegt die Performanz von ROSANA in allen drei Disziplinen oberhalb der entsprechenden Ergebnisse für die Textsammlung “Presse-Meldungen”.

Anhang B zeigt ein Dokument der Textsammlung “Mozart-Opern” und die entsprechenden, ROSANA-generierten Interpretationsergebnisse (Antezedenten, nichtpronominale Substitute) sowie deren individuelle Bewertung. Auf den ersten Blick erscheint der Text hinsichtlich der Problemstellungen der KV- und PS-Disziplinen komplexer, da die Dichte der pronominalen Wiederaufgriffe höher ist und oftmals in ein und demselben lokalen Kontext mehrere Diskursreferenten gleichzeitig pronominal realisiert und somit fokussiert sind, was sich in entsprechenden Entscheidungsambiguitäten niederschlagen sollte. Die vergleichsweise hohe Precision- und Recall-Performanz in den unterschiedlichen Evaluationsdisziplinen bestätigt diese Einschätzung jedoch nicht. Eine Detailanalyse zeigt auf, daß in vielen Fällen eine hochwirksame Vorauswahl auf der Grundlage der morphologischen Kongruenzrestriktion erfolgt, denn im Unterschied zum Korpus “Presse-Meldungen” besitzen die meisten pronominalen Okkurrenzen einen natürlichen Genus (Maskulinum, Femininum), auf dessen Grundlage die Entscheidungsambiguität stark reduziert wird. Zumindest für die Interpretation nichtpossessiver Pronomen (Klasse PE-3) leistet darüberhinaus die Präferenzstrategie der syntaktischen Rollenträgheit (SYNPARALL) einen wichtigen Beitrag zur Wahl eines korrekten Antezedenten. Die guten Ergebnisse von ROSANA

⁴²Vgl. Fußnote 19, S. 243; die Recall-Werte wurden per Hand ermittelt.

EVALUATIONSERGEBNIS OKKURRENZEN:

- NUR ANA: 44
 - NUR KEY: 22
 - ANA UND KEY: 892
 => PRECISION: 0.9530
 => RECALL: 0.9759

A.R.-ERGEBNIS-PARTITIONIERUNG

- SCHNITTE: 59
 - MOEGLICH: 491
 => PRECISION: 0.8798

KEY-PARTITIONIERUNG

- SCHNITTE: 100
 - MOEGLICH: 531
 => RECALL: 0.8117

DURCHSCHNITT JE DOKUMENT

=> PRECISION: 0.8840
 => RECALL: 0.8132

LEXIKALISCHE SUBSTITUTION

	PRECIS	RECALL	++	+-	+?	+ ₋	+*	?+	? ₋
PE-3	0.7025	0.7025	85	33	3	0	0	6	0
PE12	--	--	0	0	0	0	0	0	0
PO-3	0.7558	0.7558	65	18	3	0	0	0	0
PO12	--	--	0	0	0	0	0	0	0
REFL	0.8333	0.8333	5	1	0	0	0	0	0
RELA	0.9500	0.9048	19	0	1	1	0	1	0

DURCHSCHNITT JE PRONOMEN

=> PRECISION: 0.7468
 => RECALL: 0.7436

	A-PREC	A-RECA
PE-3	0.6693	0.7025
PE12	--	--
PO-3	0.7558	0.7471
PO12	--	--
REFL	0.8333	0.7143
RELA	0.9048	0.9048

DURCHSCHNITT JE PRONOMEN

=> PRECISION: 0.7250
 => RECALL: 0.7373

Abbildung 11.25: Ergebnisse für die Textsammlung "Mozart-Opern"

auf den pronomerenreichen Operntexten bestätigen das Ergebnis aus Abschnitt 11.3.4, derzufolge die Wahl pronominaler Antezedenten für Pronominalanaphern mit höherer Genauigkeit erfolgt als die Wahl nichtpronominaler Vorgänger.

Die hohe Qualität der Interpretationsleistung auch auf der Textsammlung “*Mozart-Opern*” untermauert darüberhinaus die Domänenunabhängigkeit der verwendeten Algorithmen und stellt somit ein wichtiges Indiz für die generelle Verwendbarkeit von ROSANA dar. Weniger die inhaltliche Domäne, sondern vielmehr oberflächlich-formale Charakteristika der zu verarbeitenden Dokumente wie etwa die relative Häufigkeit von Pronomen scheinen ausschlaggebend für die Qualität der erzielten Ergebnisse. Den Evaluationsergebnissen zufolge sind die Mozart-Texte *leichter* als die Pressemeldungen.

11.3.6 Stellenwert und Skopus der Prominenzfaktoren und Gewichte

Die Auswahl der Prominenzfaktoren für die unterschiedlichen Typen anaphorischer Ausdrücke sowie die Festlegung der entsprechenden numerischen Gewichte wurde auf der Basis von Testdurchläufen auf dem Trainingskorpus der Textsammlung “*Presse-Meldungen*” optimiert (vgl. S. 254 ff.). Dies wirft die Frage auf, welche *Tragweite* die solchermaßen ermittelte Konfiguration der Auswahlheuristiken von ROSANA hat. Sind die getroffenen Festlegungen auch für die Dokumente der Evaluationskorpora optimal oder wurde ROSANA nur in bezug auf die Trainingsdokumente “optimiert”? Ergeben sich Unterschiede in Bezug auf die Evaluationstexte der Korpora “*Presse-Meldungen*” und “*Mozart-Opern*”?

Ein geeigneter Ausgangspunkt für die Beantwortung dieser Fragen ergibt sich auf der Grundlage entsprechender Variationen der Konfiguration von ROSANA. Indem Versionen von ROSANA evaluiert werden, in denen einzelne Präferenzstrategien *deaktiviert* sind, werden deren individuelle Beiträge zum Gesamtergebnis und somit deren *relativer Stellenwert* für die betrachteten Korpora ermittelt. Durch eine Aufschlüsselung der Auswirkungen für unterschiedliche Anapherentypen ergeben sich Anhaltspunkte dafür, inwieweit die typabhängige Zuordnung der Strategien einer (ggf. korpuspezifischen) Verfeinerung bedarf.

In der Tabelle in Abbildung 11.26 sind die Evaluationsergebnisse für eine Serie von fünf entsprechenden Experimenten zusammengefaßt: (1) ohne Strategie SYNPARALL (Faktor SYR), (2) ohne Strategien SUBJEKT und SYNROLLE (Faktoren SUP, DOP, IOP, APP, EEP, PGP), (3) ohne Strategien SYNPARALL, SUBJEKT und SYNROLLE (Faktoren SYR, SUP, DOP, IOP, APP, EEP, PGP), (4) ohne Strategie DISTANZ (Faktor SDM), (5) ohne Strategie KATAMALUS (Faktor KAM). Die oberste Zeile zeigt die Performanzwerte der Originalversion von ROSANA als Referenzdaten. Betrachtet werden die Evaluationsdisziplinen KV (Kospezifikationsklassen) und PS (nichtpronominale Substitute). Da die auf syntaktischer Information aufbauenden Strategien SYNPARALL, SUBJEKT und SYNROLLE ausschließlich für Pronominalanaphern zur Anwendung gelangen (vgl. Abbildung 11.17), werden zusätzlich die Precision-Werte der Güte der Antezedenzzuordnung für die beiden zentralen Typen von Personalpronomen in dritter Person - PE-3 (Nichtpossessiva) und PO-3 (Possessiva) - angegeben.

Anhand der Ergebnisse für die unterschiedlichen Experimente wird zunächst ersichtlich, daß die Evaluationsergebnisse P_{kv} und R_{kv} der Disziplin KV eine vergleichsweise geringe Sensitivität gegenüber den Änderungen der Präferenzstrategien von ROSANA aufweisen. Dies läßt sich mit Blick auf die relative Häufigkeit der unterschiedlichen Anapherentypen erklären: Sowohl für das Korpus “*Presse-Meldungen*” als auch für die Textsammlung “*Mozart-Opern*” sind die Okkurrenzen der Typen VNOM und NAME in der klaren Überzahl (vgl. etwa Abbildung 11.23).

Experiment	Evaluationskorpus "Presse-Meldungen"						Korpus "Mozart-Opern"					
	P_{kv}	R_{kv}	P_{ps}	R_{ps}	PE-3	PO-3	P_{kv}	R_{kv}	P_{ps}	R_{ps}	PE-3	PO-3
ROSANA (Orig.)	0.81	0.68	0.70	0.65	0.71	0.76	0.88	0.81	0.75	0.74	0.79	0.77
(1) –SYR	0.80	0.68	0.68	0.63	0.70	0.73	0.89	0.82	0.76	0.75	0.77	0.80
(2) –SUP,...	0.80	0.68	0.67	0.62	0.69	0.73	0.87	0.80	0.68	0.68	0.74	0.67
(3) –SYR, –SUP,...	0.78	0.66	0.58	0.54	0.56	0.69	0.87	0.80	0.70	0.70	0.73	0.79
(4) –SDM	0.78	0.66	0.60	0.56	0.63	0.60	0.84	0.77	0.55	0.55	0.55	0.50
(5) –KAM	0.80	0.68	0.66	0.61	0.65	0.77	0.88	0.81	0.71	0.71	0.75	0.67

Abbildung 11.26: Variation der Präferenzstrategien von ROSANA

Die Modifikation der Auswahlheuristiken betrifft jedoch primär die Entscheidungsfindung für *pronominale* Vorkommen; entsprechend schlagen sich die Änderungen in erster Linie in den Performanzwerten P_{ps} und R_{ps} der Disziplin PS nieder. Als erstes Nebenprodukt der Untersuchung ergibt sich somit die Aussage, daß sich die MUC-Evaluationsmaße P_{kv} und R_{kv} auch aus Häufigkeitsgründen als wenig geeignet zur Bewertung der Leistungsfähigkeit von Systemen zur Interpretation *pronominaler* Ausdrücke erweisen. Die Definition der zusätzlichen Maße P_{ps} und R_{ps} sowie die Analyse der Performanz auf der Ebene der individuellen Antezedensentscheidungen erweist sich somit unter einem weiteren Gesichtspunkt als gerechtfertigt.

Anhand der Ergebnisse im linken Teil der Tabelle ist ersichtlich, daß sich die mit Blick auf das Trainingskorpus getroffenen Vereinbarungen der Faktoren und Gewichte auch für das Evaluationskorpus der Textsammlung "Presse-Meldungen" als adäquat erweisen. Jede der betrachteten Variationen schlägt sich in einer merklichen Verminderung der Werte P_{ps} und R_{ps} nieder. Die Resultate der Interpretationsgenauigkeit für Personal- bzw. Possessivpronomina (PE-3 bzw. PO-3) zeigen auf, daß der Beitrag der einzelnen Strategien in Abhängigkeit des Typs der Anapher unterschiedlich ausfällt: Mit –3 Prozentpunkten wirkt sich die Deaktivierung der Heuristik SYNPARALL (Faktor SYR, Experiment (1)) für Possessivpronomina deutlicher aus als für Nichtpossessiva (–1 Prozentpunkte). Für Nichtpossessiva scheinen die Strategien SYNPARALL und SYNROLLE (Experiment (2)) in vergleichsweise vielen Fällen zur identischen (korrekten) Entscheidung zu führen: Die Deaktivierung einer der beiden Strategien führt nur zu einer vergleichsweise geringen Reduktion der PE-3-Precision um 1 bzw. 2 Prozentpunkte; unter Verzicht auf beide Präferenzstrategien (Experiment (3)) ergibt sich ein Rückgang von 15 Prozentpunkten. Eine nähere Analyse der Texte des Korpus "Presse-Meldungen" gibt Aufschluß über die Hintergründe des scheinbaren Substitutionsverhältnisses der Strategien SYNPARALL und SYNROLLE. Die Pronomendichte in dieser Textsammlung ist relativ gering; die meisten PE-3-Pronomen kommen in der syntaktischen Funktion des Subjekts vor, woraus folgt, daß beide Heuristiken zu derselben Vorhersage kommen und somit die beobachtete Redundanz aufweisen.

Als wichtigste Einzelstrategie erweist sich die Heuristik DISTANZ, deren Deaktivierung (Faktor SDM, Experiment (4)) zu einem deutlichen Rückgang der pronomenenbezogenen Performanzwerte P_{ps} , R_{ps} , PE-3- und PO-3-Precision führt (≥ 8 Prozentpunkte). Auch die Präferenzheuristik KATAMALUS hat einen entscheidenden Anteil am Gesamtergebnis für Nichtpossessiva (Faktor KAM, Experiment (5)). Für Possessivpronomenen kommt es hingegen zu einer (marginalen) *Verbesserung*; dies ist ein Indiz für eine entsprechende Verfeinerung der Präferenzstrategie-Mixes für Possessiva, wobei allerdings zu berücksichtigen ist, daß sich die Strategie KATAMALUS für die Possessivpronomenen des Trainingskorpus als erfolgreich erweist.

Anhand der Resultate für die Textsammlung "Mozart-Opern" ergeben sich weitere Anhaltspunkte für eine Relativierung bzw. Verfeinerung der Präferenzstrategien. Zunächst einmal bestätigen die in Experiment (5) gewonnenen Ergebnisse den potentiellen Nutzen der Strategie KATA-

MALUS für Possessiva, der hier bei immerhin 10 Prozentpunkten liegt und den Beitrag für PE-3-Pronomen (4 Prozentpunkte) deutlich übersteigt. Mit mindestens 20 Prozentpunkten fällt der Nutzwert der Heuristik DISTANZ (Experiment (4)) sogar noch wesentlich deutlicher aus als für das Korpus “*Presse-Meldungen*”. Eine mögliche Erklärung läßt sich unter Rekurs auf die *spezifische Kohäsionsstruktur* der Mozart-Texte geben: Der lokale Fokus wird durch die Protagonisten der jeweils beschriebenen Opernszene konstituiert; wechselt die Szene - was häufig geschieht -, so verschiebt sich der Fokus und damit der bevorzugte referentielle Anknüpfungspunkt. Unter Verzicht auf die Heuristik DISTANZ führen die fokuspräferenz-approximierenden Heuristiken SUBJEKT, SYNROLLE und SYNPARALL somit häufiger zu einem falschen Ergebnis als für das Korpus “*Presse-Meldungen*”, in dem der lokale Fokus weniger häufig wechselt. Ein weiteres wichtiges Ergebnis ergibt sich betreffend den Nutzwert der Rollenträgheitsstrategie SYNPARALL, die sich - entgegen den Resultaten für das Korpus “*Presse-Meldungen*” - für die Interpretation von Possessivpronomen als kontraproduktiv erweist: Die Deaktivierung der Strategie führt zu einem *Ansteigen* des PO-3-Werts um 3 Prozentpunkte (Experiment (1)). Auch diese Abweichung ist durch die spezifischen Charakteristika der “*Mozart-Opern*”-Texte bedingt, die eine Anzahl von Sätzen mit zwei oder mehreren Possessiva enthalten, die *unterschiedliche* Diskursreferenten spezifizieren. Satz 1 des Texts “*Don Giovanni*” fällt in diese Klasse:

“On a dark night in Seville, Leporello is keeping watch, grumbling, outside a house in which his master Don Giovanni is engaged in his latest amorous pursuit.”

Für das unterstrichene Possessivpronomen “*his*” führt SYNPARALL zu einem inkorrekten Ergebnis, da das zweite Vorkommen “*his*” eine andere Person - Leporello anstelle von Don Giovanni - spezifiziert; nach Deaktivierung der Rollenträgheitsheuristik wird die richtige Entscheidung getroffen, da die Strategie SUBJEKT (zusammen mit einer nachrangigen Nähe-Heuristik) zur Wahl des korrekten Antezedens, dem Subjekt “*Don Giovanni*” des Relativsatzes, führt. Daß sich für die Texte des Korpus “*Presse-Meldungen*” umgekehrt sogar ein positiver Effekt der Strategie SYNPARALL beobachten läßt, liegt somit erneut in der unterschiedlichen Kohäsionsstruktur begründet: Possessiva in ein und demselben Kontext spezifizieren hier in der Regel *denselben* Diskursreferenten.

Auch die für das Korpus “*Presse-Meldungen*” beobachtete Substitutionsbeziehung der Strategien SYNPARALL und SYNROLLE für Nichtpossessiva tritt bezüglich der Mozart-Texte nicht in Erscheinung. Die Deaktivierung einer einzelnen der beiden Strategien (Experimente (1) bzw. (2)) führt zu einer deutlicheren Reduktion der Performanz (2 bzw. 5 Prozentpunkte) als im Falle der Presse-Texte. Werden hingegen beide Strategien ausgeblendet, so entspricht der Verlust von 6 Prozentpunkten in etwa der Summe der Einzelbeiträge der beiden Strategien (Experiment (3)); daß der Verlust an Precision im Falle der Deaktivierung beider syntaktischer Fokussierungsheuristiken wesentlich moderater ausfällt als für das Korpus “*Presse-Meldungen*”, ist auf unabhängige Faktoren wie insbesondere der Verfügbarkeit differenzierender Genusinformation in den Beschreibungen der einschlägigen Vorkommen der Mozart-Texte zurückzuführen, die in der Regel Objekte mit natürlichem Geschlecht (Personen) spezifizieren.

Die Ergebnisse der durchgeführten Untersuchungen ermöglichen eine differenzierte Beantwortung der Frage nach dem Skopus der in bezug auf das Trainingskorpus optimierten Auswahlstrategien. Für das Evaluationskorpus “*Presse-Meldungen*” erscheinen die getroffenen Festlegungen adäquat, wohingegen für die Mozart-Texte eine Verfeinerungsmöglichkeit zu bestehen scheint - die Elimination bzw. Abschwächung der Rollenträgheitsheuristik im Strategiemix für Possessivpronomen. Anhand der Betrachtungen wird ferner deutlich, daß der an dieser Stelle zum Tragen

kommende Unterschied zwischen den beiden Textsammlungen erneut weniger auf der Ebene der zugrundeliegenden inhaltlichen Domänen - Welt der Mozart-Opern vs. Welt der Politik - lokalisiert ist, sondern vielmehr auf Unterschiede in der typischen referentiellen Kohäsionsstruktur der jeweiligen Texte zurückführbar ist. Hieraus ergibt sich ein interessanter Anhaltspunkt für eine *automatische Ermittlung korpusadäquater Strategiekombinationen*, die auf der Basis einer computergestützten Analyse struktureller Eigenschaften der Kohäsionsstruktur referentiell annotierter Korpora bewerkstelligt werden könnte. Letztlich darf hierbei jedoch nicht außer Betracht gelassen werden, daß es sich lediglich um die Optimierung eines *Ersatzverfahrens* handelt, das eine tiefergehende semantische bzw. pragmatische Analyse niemals vollständig ersetzen kann. An den prinzipiellen Grenzen der Techniken, auf denen ROSANA sowie etliche andere Systeme bislang basieren, ändert sich hierdurch nichts.

11.4 Diskussion

Die Werte, die ROSANA in der Disziplin der Ermittlung der Kospezifikationsklassen auf den beiden Evaluationskorpora erzielt, geben einen Anhaltspunkt für die hohe Qualität der Interpretationsleistung. Mit $P_{kv} = 0.8081$ (Presse-Meldungen) bzw. 0.8798 (Mozart-Opern) und $R_{kv} = 0.6845$ bzw. 0.8117 liegen die Ergebnisse mindestens in der Größenordnung der besten Teilnehmersysteme von MUC-6 (wobei allerdings die nicht unmittelbare Vergleichbarkeit zu berücksichtigen ist, vgl. Abschnitt 11.3.3). Den geeigneteren Maßstab bilden die niedrigeren Ergebnisse für das Korpus "*Presse-Meldungen*", da die in den MU-Konferenzen zugrundegelegten Textsammlungen - i.d.R. ebenfalls Presse-Meldungen, jedoch aus anderen Domänen - eine vergleichbare analysetechnische Komplexität aufweisen. Im Vergleich mit den Resultaten der MUC-Teilnehmer ist ferner zu berücksichtigen, daß durchaus noch Spielraum für die Erhöhung der domänenspezifischen Interpretationsleistung von ROSANA besteht, das bislang lediglich auf rudimentäre Weise für das Anwendungsgebiet "*(politische) Presse-Meldungen*" konfiguriert wurde.⁴³

11.4.1 Vergleich mit anderen Systemen zur Interpretation von Pronomen

Die Evaluation hat jedoch auch gezeigt, daß eine Bewertung der Leistungsfähigkeit von Systemen zur Kospezifikations-Analyse auf dem zentralen Gebiet der Interpretation *pronominaler* Anaphern einer verfeinerten Betrachtung bedarf, da die in der KV-Disziplin erzielten Ergebnisse diesbezüglich wenig aussagekräftig sind. Die für die in den Abschnitten 8.6 und 8.7 diskutierten Algorithmen von Lappin und Leass bzw. Kennedy und Boguraev genannten Erfolgsquoten von 86 % bzw. 75 % in der Interpretation von Dritte-Person-Pronomen sind mit der Prozentzahl korrekter *Antezedensentscheidungen* von ROSANA für die Pronomentypen PE-3, PO-3, REFL und RELA zu vergleichen. Für das Korpus "*Presse-Meldungen*" ergibt sich ein Wert von $\frac{322}{432} = 0.75$ (vgl. Abbildung 11.23); auf der Textsammlung "*Mozart-Opern*" wird sogar ein Resultat von $\frac{185}{233} = 0.79$ erzielt. Vergleichsmaßstab ist das Ergebnis des robusten, anwendungstauglichen Verfahrens von Kennedy und Boguraev, dessen Performanz von durchschnittlich 75 % auf einem

⁴³ Als *domänenspezifisch* sind v.a. die (bislang vergleichsweise kleinen) Lexika für Namen und morphologische Merkmale zu bezeichnen, die der Konstruktion der Vorkommensbeschreibungen zugrundegelegt werden.

breitgefächerten Evaluationskorpus mit Texten unterschiedlicher Genres und Domänen ermittelt wurde. ROSANA erzielt eine Interpretationsgenauigkeit von 75 % bereits für die vergleichsweise schwierigen Dokumente des Korpus *“Presse-Meldungen”*. Folglich steht zu erwarten, daß der durchschnittliche Wert in bezug auf eine Textsammlung breiteren Zuschnitts um einige Prozentpunkte höher ausfällt. Eine abschließende Beurteilung ist nicht möglich, da die Resultate auf unterschiedlichen Korpora erzielt wurden und die von Kennedy und Boguraev genannte Zahl nicht nach Textsorte aufgeschlüsselt ist.

Eine interessante Alternative zu der im vorliegenden Fall problembehafteten unmittelbaren Gegenüberstellung der numerischen Evaluationsergebnisse besteht in einer qualitativen Aufschlüsselung der Ursachen inkorrekt er Antezedensentscheidungen. Sowohl ROSANA als auch dem System von Kennedy und Boguraev liegt der Anspruch zugrunde, die für die Interpretation pronominaler Ausdrücke als zentral bekannten bindungstheoretischen Koindexierungs-Restriktionen auf der Basis der Ergebnisse einer uneingeschränkt robusten, anwendungstauglichen Voranalyse zu bewerkstelligen. Der auf dem Modell der seichten Voranalyse basierende Ansatz von Kennedy und Boguraev verfügt in zumindest einigen Fällen nicht über die notwendige Information, um konfigural unzulässige Antezedenskandidaten zu eliminieren: 2 (von insgesamt 75) Fehlentscheidungen seien auf diese Ursache zurückzuführen; weitere Fehler (Zahlen werden nicht genannt) entstünden durch eine inkorrekte heuristische Rekonstruktion der syntaxbasierten Prominenzfaktoren. Für die im Fragmentfall zum Teil ebenfalls heuristische Verifikation der Bindungsbedingungen von ROSANA, das auf dem Modell der potentiell fragmentarischen Voranalyse basiert, wurden *keine entscheidungsrelevanten Fehler* ermittelt. In allen Fällen, in denen die Strategie F-SYNKONFIG inkorrektweise entweder ein gesuchtes, konfigural zulässiges Antezedens eliminiert oder einen konfigural unzulässigen Kandidaten nicht eliminiert und infolge dessen auswählt (6 bzw. 1 von insgesamt 246 Fehlentscheidungen), liegen Fehler im Ergebnis der syntaktischen Voranalyse des DPE-Parsers vor, die ursächlich für das Versagen von F-SYNKONFIG sind; selbige Ursache liegt 3 weiteren inkorrekten Antezedensentscheidungen zugrunde, in denen die falsche Auswahl durch Fehlinstanzierungen syntaktischer Prominenzfaktoren bedingt ist, die zu einer Bevorzugung des falschen bzw. Zurückstufung des richtigen Kandidaten führt.

Im Vergleich mit dem System von Kennedy und Boguraev neigt ROSANA somit eher zur Übergenerierung denn zur Untergenerierung syntaktisch-konfiguraler Restriktionen, was jedoch in allen beobachteten Fällen auf Fehler in der syntaktischen Voranalyse zurückzuführen ist. Dies verdeutlicht erneut die *Grenzen des zugrundeliegenden sequentiellen Verarbeitungsmodells*. Durch eine Verzahnung der referentiellen und der syntaktischen Analyse könnten die Antezedens-Präferenzen der Anapherninterpretation in einem Teil dieser Fälle dazu herangezogen werden, die Anzahl der Fehlentscheidungen im Rahmen der syntaktisch-strukturellen Disambiguierung zu reduzieren. Dies gilt genau dann, wenn der bevorzugte Antezedenskandidat nur in einer Teilmenge der oberflächenstrukturellen Lesarten konfigural zulässig ist; die Präferenz könnte als Entscheidungskriterium der referentiellen Interpretationsebene in die strukturelle Interpretation einfließen, auf dessen Grundlage die übrigen Lesarten eliminiert werden (vgl. Kap. 10). Eine weitere Verbesserung der Performanz von ROSANA ist jedoch auch bereits unter Beibehaltung des sequentiellen Analysemodells möglich. Die tendenzielle Übergenerierung bindungstheoretischer Restriktionen legt es nahe, mit *defensiveren Parsing-Strategien* zu experimentieren, die Disambiguierungsentscheidungen ausschließlich dann treffen, wenn deren Korrektheit aus strukturellen Gründen garantiert ist.⁴⁴ In der gegenwärtigen Konfiguration des DPE-Parsers entstehen

⁴⁴Anhand der Interpretationsergebnisse für die Evaluationskorpora ist ersichtlich, daß insbesondere im Rahmen der Interpretation strukturell ambiger Relativsätze vergleichsweise viele Fehlentscheidungen getroffen werden, die immerhin 4 der 6 problematischen Fälle mit Übergenerierung syntaktisch-konfiguraler Information bedingen.

durchschnittlich 2.61 oberflächenstrukturelle Fragmente je Satz; es steht zu erwarten, daß eine geringfügige Erhöhung der Fragmentierungsquote nicht mit einer die Anapherninterpretation tangierenden Untergenerierung syntaktisch-konfiguraler Evidenz einhergeht.

11.4.2 Möglichkeiten zur Verfeinerung von ROSANA

Ein wesentlich größerer Spielraum zur Verbesserung der Performanz von ROSANA in der Interpretation von Pronomen besteht jedoch an anderen Stellen. Sogar betreffend die intrasententiellen Antezedensentscheidungen sind nur ca. 8 % der Fehler auf Fehlvorhersagen der Restriktionen F-SYNKONFIG zurückzuführen. Mehr als 25 % der Problemfälle sind durch inkorrekte Kongruenzmerkmale in den Vorkommensbeschreibungen bedingt, die zu falschen Vorhersagen der Strategie MORKONGRU führen; weitere 25 % der Fehler entstehen durch die begrenzte Leistungsfähigkeit der Auswahlheuristiken. In bezug auf die Gesamtheit der Fehlentscheidungen kommt den beiden zuletztgenannten Fehlerklassen sogar ein Anteil von je 30 % zu (vgl. o.).

Den größten Zugewinn bei vergleichsweise geringstem Aufwand verspricht die (ggf. anwendungsdomänen-orientierte) Erweiterung des morphologischen Lexikons, wobei der Genus-Information die zentrale Rolle zukommt. Bezüglich der Fehlvorhersagen der Auswahlheuristiken ergibt sich ein uneinheitliches Bild: Während zumindest ein Teil der inkorrekten Entscheidungen auf der Grundlage automatisch akquirierter selektionaler Information vermieden werden könnte (Strategie SEMROLLEN),⁴⁵ basiert die Mehrzahl dieser Fälle auf Subtilitäten, die auf der inhaltlichen Ebene angesiedelt sind und mit den gegenwärtig zur Verfügung stehenden Ressourcen nicht in adäquater Weise abgedeckt werden können.

Eine erfolgversprechende Möglichkeit zur Verfeinerung von ROSANA wird durch die in Abschnitt 11.3.6 diskutierten Experimente aufgezeigt, derzufolge die Ausgestaltung der jeweils anzuwendenden Kombinationen von Präferenzfaktoren unter Berücksichtigung korpuspezifischer Kriterien erfolgen sollte. Durch eine computergestützte Analyse der referentiellen Kohäsionsstruktur intellektuell annotierter Beispieltex-te könnte die entsprechende Information automatisch gewonnen werden.

Unter Ausschöpfung der identifizierten Spielräume sollte sich die Performanz von ROSANA bezüglich der Antezedensermittlung für Dritte-Person-Pronomen auch für die schwierigen Texte des Typs *“Presse-Meldungen”* auf Werte oberhalb von 80 % steigern lassen. In der Disziplin der Ermittlung nichtpronominaler Substitute, die entsprechend den Ergebnissen von Abschnitt 11.3.4 schwieriger zu bearbeiten ist, liegen Werte oberhalb von 75 % im Bereich des Möglichen, wobei allerdings weiterhin das Problem einer vergleichsweise breiten, von der Reihenfolge pronominaler und nichtpronominaler sprachlicher Ausdrücke in den Sequenzen anaphorischer Wiederaufgriffe abhängigen Streuung der erzielten Ergebnisse besteht (vgl. Beleg (21)).

Geht es primär um die Verbesserung der Performanz in der Koreferenzklassen-Disziplin KV, so sind ferner Techniken für die Interpretation von VNOM-Anaphern interessant, die über das Elementarkriterium der lexikalischen Rekurrenz hinausgehen und eine Identifikation auch solcher kospezifizierender nichtpronominaler Okkurrenzen erlauben, deren NPs im lexikalischen Kopf nicht übereinstimmen. Die Forschung auf diesem Gebiet befindet sich jedoch erst am Anfang (vgl. z.B. Vieira und Poesio, [ViPo96]); es ist noch nicht abzusehen, welche Steigerungsmöglichkeit sich unter den Rahmenbedingungen einer robusten Interpretation unrestringierter Korpora bietet.

⁴⁵Lappin und Leass ermittelten eine Zahl von ca. 2.5 Prozentpunkten; vgl. Abschnitt 8.6.

Kapitel 12

Zusammenfassung und Ausblick

Die wörterbuchbasierte Computergestützte Inhaltsanalyse, die in Abschnitt 3.3 beschrieben wurde, steht in unmittelbarem Zusammenhang mit der klassischen inhaltsanalytischen Methode Berelsons, die auf die Erstellung einer quantitativen Deskription des “manifesten” Textinhalts abzielt. Die Ausführungen in Kapitel 2 haben jedoch verdeutlicht, daß eine scharfe Abgrenzung zwischen Qualitativer und Quantitativer Analyse inadäquat ist, da in jedem Fall qualitative Verarbeitungsschritte anfallen. Neben den Aufgaben der Entwicklung eines adäquaten Kategoriensystems und der Interpretation der Inhaltsanalyse-Ergebnisse betrifft dies im Speziellen auch die auf eine Maximierung der Objektivität der Analyse abzielende Operationalisierung des Codierschritts, dessen ultimative Reproduzierbarkeit etwa im Rahmen der klassischen CGI unter Rekurs auf elementare Zeichenketten-*Algorithmen* hergestellt wird. Auch wenn es den Anschein hat, daß das Objektivitätsziel unter Ausblendung intellektuell-interpretativer Elemente aus dem Codiervorgang erreicht wird - mit Blick auf die Zielsetzung der Vorab-Definition von “*Durchschnittsbedeutungen*” (Fühlau: [Fühl82], S. 188) ohne Anschauung des Textmaterials, die ja in der klassischen (Berelsonschen) Inhaltsanalyse als Grundlage der Erschließung des sog. manifesten Inhalts angenommen wird, wird ersichtlich, daß eben *doch* subjektive Elemente - und zwar diejenigen des Forschers anstelle derjenigen des Codierers - im Spiel sind. Es stellt sich unmittelbar die Frage nach der Bevorrechtigung dieser Form von Vorab-Interpretation durch den (Lisch und Kriz: [LiKr78], S. 45 f.)

“[...] präpotenten Soziologen, der unabhängig von der Kenntnis der Texte durchschnittliche Bedeutungen definieren kann, während die Kodierer, die ja die Texte lesen und außerdem die Adressaten dieser Texte sehr gut substituieren könnten, die spezifische Bedeutung nicht erkennen dürfen.”

Während sich der Begriff “manifeste Inhalt” somit als Schimäre erweist, für den in der einschlägigen Literatur zahlreiche mehr oder weniger obskure Definitionsvorschläge unterbreitet werden, stellt sich andererseits die Frage nach der Qualität der inhaltsanalytischen Ergebnisse, die auf diese Weise erzielbar sind. Durch den bewußten Verzicht auf subjektive Entscheidung, so Lisch und Kriz weiter, könne das “*Verständnis der Inhaltsanalyse als rekonstruierte Interaktion*” letztlich nicht eingelöst werden (ibid., S. 46). Statt also weiterhin von einer - theoretisch unhaltbaren - Unterscheidung zwischen manifestem und latentem Inhalt auszugehen, so wird

argumentiert, solle der Objektivitätszielsetzung durch eine möglichst weitreichende *Explikation* der subjektiven Elemente des Interpretationsvorgangs Rechnung getragen werden, die die Kommunizierbarkeit der persönlichen Entscheidungen und damit letztlich deren *intersubjektive Reproduzierbarkeit* herstellt.

Ausgangspunkt der Arbeit war die Fragestellung, inwieweit Computer bestimmte Interpretationsaufgaben in einer derartigen - qualitativ erweiterten - Inhaltsanalyse übernehmen können. Fühlau mahnt einerseits eine textlinguistische Fundierung der klassischen Inhaltsanalyse an ([Fühl82], S. 190)

“In der Texttheorie vorliegendes Wissen über Textkonstitution, Textkohärenz und Textfunktionen ist gewiß nützlicher für eine Textanalyse als das bloße Auszählen von Einzelwörtern.”

sieht jedoch gleichzeitig keine Möglichkeit, diesen Anspruch im Rahmen einer *maschinellen* Analyse einzulösen (ibid., S. 188):

“In der für den Computer erforderlichen Einwortanalyse ist [...] entschieden keine Möglichkeit enthalten, kommunikatives Wissen über spezifische Textproduzenten und Rezipienten, deren Interessen und Absichten, über Kommunikationssituationen und -kontexte einzubringen.”

Die Ergebnisse der Dissertation legen indes eine Revision dieser Einschätzung nahe. Auf der Grundlage der Resultate läßt sich die Gültigkeit der eingangs postulierten zentralen Thesen folgern (vgl. Abschnitt 1.4).

12.1 Neubewertung der Computergestützten Inhaltsanalyse

Die in Kapitel 4 entwickelte textlinguistische Fallstudie belegt zunächst einmal erwartungsgemäß die Aussage des ersteren Fühlau-Zitats, derzufolge das im Rahmen einer Computergestützten Inhaltsanalyse klassischen Zuschnitts bewerkstelligte “Auszählen von Einzelwörtern” zu kurz greift und einer texttheoretischen Verfeinerung bedarf. Die Studie geht jedoch über eine bloße Kritik erheblich hinaus, indem sie einerseits offenlegt, warum die wortorientierte CGI in zumindest rudimentärer Weise funktioniert und darüberhinaus aufzeigt, welche Möglichkeiten einer insbesondere auf der Erschließung linguistischen Wissens aufbauenden Verfeinerung existieren. Die Operationalisierung des Kategorisierungsschritts auf der Basis von Wörterbüchern entspricht einer elementaren, auf der lexikalischen Ebene angesiedelten *Themenanalyse* entlang der vom Forscher als klassifikationsrelevant angesehenen *Isotopieebenen*; allgemein jedoch können Themen - definiert als textuell etablierte und mehrfach instantiierte Kontexte - auf ausdrucksseitig beliebig komplexe Weise konstituiert werden. Gelänge es, zusätzliche Ebenen der textuellen Themenentwicklung algorithmisch zu erschließen, so wäre dies als ein wichtiger Schritt hin zu einer verfeinerten CGI zu werten.

Ein wichtiges Nebenprodukt der Fallstudie besteht in der Identifikation der Untauglichkeit “verfeinerter” Beschreibungskonzepte, die für einige der klassischen CGI-Softwaresysteme entwickelt

wurden, durch die jedoch die Ebene einer einzelwortbasierten Analyse nicht in entscheidendem Ausmaß transzendiert wird (Abschnitt 4.5). Da keine theoretisch fundierte bzw. zumindest empirisch abgesicherte Erschließung der kommunizierten Inhalte aus den sprachlichen Ausdrücken bewerkstelligt wird, bewegt sich die Steigerung des Nutzwerts der dergestalt erweiterten Spezifikationsprachen in engen Grenzen; die Erfahrungen aus praktischen Anwendungen bestätigen diese theoretische Einschätzung. Folglich gilt die zentrale These 1 der Arbeit.

Somit untermauert die Fallstudie die zitierte Einschätzung Fühlaus bezüglich der unzureichenden textlinguistischen Verankerung der klassischen CGI auf qualifizierte Weise; zugleich jedoch verkörpert sie den Ausgangspunkt einer Bestandsaufnahme neuerer computerlinguistischer Verfahren, anhand derer deutlich wird, daß die computergestützte Textinhaltserschließung nicht notwendig auf die Ebene der Analyse einzelner Wörter beschränkt ist. Die modernen Techniken des *Information Extraction*, die im Rahmen der "Message Understanding"-Konferenzen entwickelt wurden, ermöglichen eine *inhaltsorientierte* Erschließung nichtatomarer thematischer Entitäten eines Texts, die - qua unterliegender linguistischer Theorie - eine erheblich stärkere Unabhängigkeit von der konkreten Oberflächenrealisierung des zu erschließenden Inhalts erreicht als die theoretisch unfundierten und damit letztlich untauglichen Ad-Hoc-Reparaturversuche der "verfeinerten" klassischen CGI. In Abschnitt 5.4.2 wurden Vorschläge für einen Einsatz von "Information Extraction"-Techniken im Rahmen einer verfeinerten CGI erarbeitet. Der Blick auf die zentralen Eigenschaften der Teilnehmer-Systeme der MUC-6-Evaluation - hohe Ergebnisgüte, robuste anwendungstaugliche Analyse, Verwendung linguistisch fundierter Generalisierungen für die Faktenextraktion, Erhöhung der Portabilität per Separation linguistischen und domänenspezifischen Wissens - verdeutlicht, daß die Entwicklung in die richtige Richtung geht. Die aktuell verfügbaren computerlinguistischen Analyseverfahren erfüllen hiermit die wesentlichen Anforderungen, die in bezug auf eine Verfeinerung der maschinellen Themen-Analyse zu stellen sind. Die Prämisse des letzteren Fühlau-Zitats, derzufolge von einer "für den Computer erforderlichen Einwortanalyse" auszugehen ist, ist somit im Hinblick auf die neueren Forschungsergebnisse der Computerlinguistik als widerlegt einzustufen, woraus sich auf die Gültigkeit der zentralen These 2 schließen läßt.

Folglich läßt sich im Rahmen einer Computergestützten Inhaltsanalyse jenseits aller Scheingehefte der Philosophien "qualitativer" und "quantitativer" Analysen deutlich mehr leisten als eine einzelwortorientierte Kategorisierung. Ob die *Ergebnisse* des algorithmisierten Codiervorgangs letztlich durch Auszählen, d.h. quantitativ, oder per intellektueller Interpretation, d.h. qualitativ, ausgewertet werden, ist in diesem Zusammenhang (zunächst) ebensowenig von Bedeutung wie die Scheinfrage, ob der Computer "manifeste" oder "latente" Inhalte ermittelt. Entscheidend ist vielmehr die entsprechend den zitierten Ausführungen von Lisch und Kriz an Stelle dessen zu fordernde möglichst weitreichende *Explikation* der Analyse, die nunmehr auch für (bestimmte) Themenanalysen jenseits einer reinen Einwort-Analyse per Rekurs auf Algorithmen in ultimativer Form gewährleistet werden kann. Daß das Einlösen dieses Anspruchs keine Frage der Anwendung quantitativer - i.S. numerischer - Auswerte-Verfahren, sondern vielmehr der *Formalisierung* einer *inhärent qualitativen Operation* ist, wird bereits von Lisch und Kriz unter Verweis auf die Mathematik als Wissenschaft des *Qualitativen* in den Vordergrund gestellt ([LiKr78], S. 49):

"Genausowenig [...] hat die Frage nach der Formalisierung etwas mit 'Qualität' versus 'Quantität' zu tun. Mathematik ist bekanntlich nicht die Wissenschaft vom Quantitativen, sondern vom Operieren mit bestimmten Symbolen nach explizit und eindeutig festgelegten Regeln. Diese Symbole sind nichts anderes als die Abbilder oder Repräsentanten von Relationen - man könnte auch sagen: von Qualitäten."

Auf der Grundlage der Ergebnisse der vorliegenden Arbeit läßt sich diese Aussage in Richtung auf einen Grundbegriff der *Informatik* präzisieren: *Algorithmen* ermöglichen eine *ultimative* Formalisierung des (qualitativen) Codierschritts der Kategorienanalyse, wobei die lexikalische Ebene der einzelwortbezogenen Zeichenkettenvergleiche in entscheidendem Maße transzendierbar ist. Mit anderen Worten: Der Computer bewerkstelligt eine *qualitative* Analyse, durch die auch *komplexe* inhaltliche Entitäten erschlossen werden können.

In bezug auf die Frage der prinzipiell erzielbaren Leistung einer maschinellen qualitativen Textinterpretation wird keine theoretisch begründbare Schranke betreffend die algorithmische Explizierbarkeit des impliziten Vorverständnisses der menschlichen Textinterpretation gesehen. Sicherlich ist die Kritik Fühlaus in bezug auf die Verfahren der klassischen Computergestützten Inhaltsanalyse gerechtfertigt, in der sowohl der intra- als auch der extratextuelle Kontext (die Kommunikationssituation) nicht bzw. allenfalls marginal berücksichtigt werden. Die Folgerung, daß eine *prinzipielle* Grenze der Explizierbarkeit des impliziten Verstehenshintergrunds existiert, über den somit nur der Forscher, nicht jedoch die Maschine verfügen kann, erscheint indes nicht nachvollziehbar. Im Gegenteil: Die Untersuchungen aktueller computerlinguistischer Ansätze sowohl des *Information Extraction* als auch der referentiellen Analyse haben gezeigt, daß nicht-triviale Probleme der Textinterpretation durch Computersysteme gelöst werden können. Wird ein *ergebnisorientierter* Bewertungsmaßstab angelegt - durch den die laut Lisch und Kriz zu fordernde Intersubjektivität (Explikation) ungeachtet obskurer "latenter" Inhaltsebenen bereits ultimativ eingelöst wird -, so verbleibt nur zu konzedieren, daß bereits entscheidende Schritte hin zu einer maschinellen Sprachinterpretation vollzogen sind (vgl. Abschnitt 5.5). Die Frage nach der Machbarkeit einer algorithmischen Textinhaltserschließung kann somit nicht unter Rekurs auf eine hermeneutische Argumentation a priori verneint werden. Eine von vornherein negative Aussage bezüglich der Tauglichkeit jedweder Computergestützten Inhaltsanalyse ist folglich zu verwerfen.

(Spätestens) mit der algorithmischen Transzendierung des simplen Zeichenketten-Vergleichs der klassischen CGI ergibt sich der Anlaß für eine *differenzierte* Neubewertung. Ob die jeweiligen inhaltsanalytischen Zielsetzungen auf der Basis einer maschinellen Bewerkstelligung des Codierschritts erreicht werden können, bedarf einer sensiblen Analyse der spezifischen Rahmenbedingungen von

- zu beantwortender *Forschungsfrage*,
- vorliegender *Vertextung* des Untersuchungsgegenstands,
- den für eine gegenstandsadäquate *Entwicklung des Kategoriensystems* in Frage kommenden Verfahren.

Genausowenig, wie es gerechtfertigt ist, die Objektivität, d.h. die hundertprozentige Reproduzierbarkeit der Ergebnisse einer klassischen (ggf. computergestützten) Inhaltsanalyse à la Berelson zum alleinigen Maß zu erheben, erscheint es umgekehrt adäquat, den Anspruch jedweder maschinellen Inhaltsanalyse per Postulierung der inhärenten Ungeeignetheit des Werkzeugs Computer a priori zu verwerfen. In Abhängigkeit vom konkreten inhaltsanalytischen Problem ist vielmehr vor dem Hintergrund der gegenwärtig verfügbaren computerlinguistischen Analyseverfahren abzuwägen, ob eine computerbasierte Analyse *machbar* ist und, falls ja, ob der hierbei entstehende *Aufwand vertretbar* ist. Die Grundvoraussetzung besteht darin, daß die entsprechend der Forschungsfrage relevanten Inhalte im vorliegenden Textmaterial in einer Form sprachlich

realisiert sind, die eine Erschließung mit den Techniken des computerlinguistischen State-of-the-Art Erfolg versprechend erscheinen läßt. Aus der hermeneutischen Theorie läßt sich zwar ableiten, daß die algorithmische Explikation der Inhaltserschließung *möglicherweise* schwierig ist; das Postulat eines hermeneutischen *Circulus vitiosus*, der das Unterfangen einer maschinellen Inhaltsanalyse unbeachtet des konkreten Falls zum Scheitern verurteilt, kann indes nicht nachvollzogen werden. Somit läßt sich auf die Gültigkeit des Kernpostulats der Arbeit (zentrale These 5) schließen.

Die Rolle des Computers im Rahmen der maschinellen Textinhaltsanalyse soll dennoch völlig unpräventiös in der eines *Werkzeugs* zum hundertprozentig reliablen Vollzug *bestimmter* symbolischer Transformationen gesehen werden, das den Forscher von *bestimmten* qualitativen Operationen bei gleichzeitig maximaler, da algorithmischer Explikation der gewählten Vorgehensweise entlastet. Die Interpretation der generierten Beschreibungen vor dem Hintergrund der Forschungsfrage verbleibt auch hier letztlich dem Forscher überlassen. Keineswegs erscheint es somit in bezug auf eine solchermaßen verstandene Computergestützte Inhaltsanalyse weiterhin gerechtfertigt, etwa von einer umfassenden "*Instrumentalisierung und Vorschaltung des Verstehensprozesses*" (Fühlau: [Fühl82], S. 189) durch den Computer zu sprechen, deren Aufhebung zu fordern sei, da der Computer nicht über das notwendige kommunikations-kontextuelle Wissen verfüge, das jedoch für eine Inhaltsanalyse, die ja "*auf wissenschaftliche Art und Weise Handlungen verstehen*" (ibid., S. 188) wolle, unabdingbar sei. Die interpretative Interaktion des Forschers mit dem Untersuchungsgegenstand, die zu einem *verstehenden* Erschließen der Inhalte unabdingbar sei (ibid., S. 189, Anführungszeichen und Hervorhebung im Original)

"Handlungen kommt 'kein Sinn-an-sich' zu, sie bekommen erst Sinn im Verstehen anderer Menschen."

soll keineswegs eliminiert werden. Der Computereinsatz betrifft ausschließlich den Codierschritt, von dessen intellektueller Durchführung der Forscher *unter bestimmten Voraussetzungen* befreit werden kann. Die Entwicklung des Kategoriensystems ist nicht an eine - erkenntnistheoretisch ohnehin uneinlösbare - Fixierung sogenannter "Durchschnittsbedeutungen" ohne Betrachtung von Textmaterial und Kontext gebunden, sondern kann - und sollte - unter sensibler Berücksichtigung der inhaltsanalytischen Rahmenbedingungen erfolgen; die geforderte Explikation ist hier in aller Regel (im Hinblick auf den computerlinguistischen State-of-the-Art) nicht auf algorithmischem Wege herzustellen, sondern unter Rekurs auf Verfahren der *intellektuellen* qualitativen Analyse, die eine wenigstens hohe intersubjektive Reproduzierbarkeit gewährleisten. Selbiges gilt für die Interpretation der Codierungs-Ergebnisse vor dem Hintergrund der jeweiligen Forschungsfrage.

Wenn auch keine prinzipielle Grenze für die Tragweite von Computer-Interpretationen ausgemacht werden kann - schlußendlich fallen die wesentlichen Aufgaben der Sinnzuweisung des inhaltsanalytischen Verstehensprozesses weiterhin in den Zuständigkeitsbereich des Forschers.

12.2 Algorithmische Lösbarkeit des Pronomen-Problems

Die somit postulierte Neubewertung der Tragfähigkeit einer maschinellen qualitativen Textanalyse gilt im Speziellen für ein zentrales Problem der klassischen Computergestützten Inhaltsanalyse

- die Verarbeitung pronominaler Ausdrücke -, dessen approximative algorithmische Lösbarkeit in Teil II der Arbeit ganz konkret per Entwicklung und Implementierung des Softwaresystems ROSANA nachgewiesen wurde. Auch bezüglich des Pronomen-Problems greift die Prognose Fühlaus zu kurz (ibid., S. 171)

“[...] strebt KWIC nur die Disambiguierung von Einzelwörtern, nicht aber von Sätzen oder Texten an, daher entziehen sich z.B. die - oft satzübergreifenden - Verweisungen der Pronomina der Möglichkeit automatischer Bearbeitung und müssen als Teil der Textaufbereitung von Hand substituiert werden.”

da diese Aussage ausschließlich in bezug auf das von Stone entwickelte elementare statistische KWIC-Verfahren gilt - das tatsächlich als untauglich zur Lösung des Pronomen-Problems einzustufen ist -, jedoch die Möglichkeiten einer linguistisch-symbolischen, nicht auf die lexikalisch-statistische Ebene beschränkten Analyse außer Betracht läßt.

Gemäß den Ergebnissen der formalen Evaluation in Abschnitt 11.3 liegt die erwartete Leistung von ROSANA in der Disziplin der Interpretation von Pronomen (PS-Aufgabe) bei ca. 70 Prozent, wobei die tatsächlich erzielte Leistung vom Genre der verarbeiteten Texte abhängt und die Möglichkeit zur Erhöhung der Genauigkeit durch eine textspezifische Konfiguration der Auswahlheuristiken gegeben ist. Zu berücksichtigen ist ferner, daß nicht die Hundert-Prozent-Marke, sondern die entsprechend den MUC-Ergebnissen wesentlich geringere Intercoder-Übereinstimmung intellektuell generierter Referenzdaten den Vergleichsmaßstab bildet; insofern ist die Analysequalität bereits relativ nahe am prinzipiell auf manuellem Wege Erreichbaren. Ob die erzielbare Güte den jeweiligen Anforderungen genügt, kann erneut ausschließlich unter Berücksichtigung der konkreten inhaltsanalytischen Rahmenbedingungen beantwortet werden. Jedenfalls ist die generelle Aussage, derzufolge eine maschinelle Interpretation pronominaler Ausdrücke einen prinzipiell uneinlösbaren Anspruch verkörpere, vor dem Hintergrund der Evaluationsergebnisse als widerlegt anzusehen. Da nachweislich erfolgreiche Algorithmen angegeben werden können, ist vielmehr auf die Gültigkeit der zentralen These 3 der Arbeit zu schließen.

Folglich ist in bezug auf das Pronomen-Problem der klassischen Computergestützten Inhaltsanalyse exemplarisch der konkrete Nachweis erbracht worden, daß ein unvoreingenommener und differenzierter interdisziplinärer Blick auf aktuelle Forschungsarbeiten zur algorithmischen Textinhaltserschließung in Computerlinguistik und Informatik lohnenswert ist. Die algorithmische Lösbarkeit dieses spezifischen Problems qualitativer Analyse verkörpert zusätzliche wertvolle Evidenz für die Gültigkeit des Kernpostulats (zentrale These 5) der Arbeit. Wie im Rahmen der Evaluation des Systems ROSANA ferner deutlich wurde, bedarf die adäquate Übertragung der Ergebnisse jedoch einer sensiblen Vorgehensweise, die den spezifischen Anforderungen einer verfeinerten CGI Rechnung trägt. So erwies sich eine unmittelbare Beurteilung der Tauglichkeit von Computersystemen zur referentiellen Interpretation für eine Anwendung im Rahmen einer verfeinerten CGI auf der Grundlage der Evaluationsergebnisse in der den MUC-Evaluationen entlehnten KV-Disziplin u.a. deshalb als unmöglich, weil die Resultate auch die Systemleistung in der Interpretation der anteilig wesentlich häufiger vertretenen *nichtpronominalen* spezifizierenden Ausdrücke (Namen, gewöhnliche nichtpronominale Nomen) widerspiegeln. Da ferner das (erste) Ziel nicht etwa darin besteht, das “referentielle Netz” des Textes zu rekonstruieren¹, son-

¹Eine solche erweiterte Zielsetzung würde jedoch spätestens im Rahmen einer verfeinerten CGI relevant werden, die die von Fühlaus stillschweigend supponierte elementare, lexikalische Spielart der Themenanalyse transzendiert und verfeinerte thematische Entitäten einbezieht, deren Kohäsionsbeziehungen auch auf der referentiellen Ebene begründet werden. (Vgl. die Ergebnisse der Fallstudie, Abschnitt 4.4.)

dern konkrete nichtpronominale Substitute zu ermitteln, durch die das Pronomenproblem *aus Sicht der klassischen CGI* gelöst wird, erweist sich eine einfache Bewertung der Antezedensentscheidungen für pronominale Okkurrenzen ebenfalls als unzureichend. Im Rahmen der Untersuchungen in Kapitel 11 wurde der entscheidende Schritt per Einführung der zusätzlichen Disziplin PS vollzogen. Die entsprechende Evaluation führte zu dem fokustheoretisch begründbaren Ergebnis, daß die Problemstellung der Bestimmung nichtpronominaler Substitute *schwieriger* ist als die Ermittlung eines *beliebigen* kospezifizierenden Antezedens. Somit besteht ein weiteres zentrales Resultat von Teil II der Arbeit in der Feststellung, daß einerseits die in der computerlinguistischen Literatur genannten guten Evaluationsergebnisse für Pronomeninterpretationssysteme nicht unmittelbar für eine (elementar) verfeinerte CGI zum Tragen kommen, daß jedoch andererseits auch in der schwierigeren PS-Disziplin noch relativ gute Ergebnisse erzielt werden, deren Qualität wiederum vom Textgenre sowie von bestimmten formalen Charakteristika der Realisierung der referentiellen Kohäsionsstruktur auf der Ausdrucksseite determiniert wird.

Aus den Ergebnissen der Evaluation ist ferner zu schließen, daß auf der Basis der gegenwärtig verfügbaren robusten (anwendungstauglichen) Vorverarbeitungsressourcen zumindest betreffend das Problem der referentiellen Interpretation *pronominaler* Okkurrenzen keine entscheidenden Möglichkeiten einer Verfeinerung des Kernsystems ROSANA bestehen. Zentraler Schritt und wesentliches computerlinguistisches Forschungsergebnis ist die in Abschnitt 10.3 erarbeitete Generalisierung der bindungstheoretischen Restriktionen auf potentiell fragmentarische syntaktische Strukturen, dem charakteristischen Ausgabeformat robuster Parser. Detailuntersuchungen im Rahmen der Evaluation belegten, daß die Verluste gegenüber einer Operationalisierung der Bindungsbedingungen auf eindeutigen oberflächenstrukturellen Beschreibungen ausschließlich auf (ROSANA-externe) Parsing-Fehler zurückzuführen sind und somit minimal ausfallen.

Die Resultate bewegen sich nahe am derzeit ohne einen umfassenden Bestand an semantischem Hintergrundwissen erzielbaren Optimum; die Möglichkeiten einer primär auf morphologischer und syntaktischer Information basierenden algorithmischen referentiellen Interpretation scheinen somit nahezu ausgeschöpft. Die größte Steigerungspotential existiert bezüglich der externen morphologischen Analyse, deren unvollständige bzw. inkorrekte Genus-Ergebnisse einen vergleichsweise hohen Anteil von etwa 30 Prozent der Fehlentscheidungen von ROSANA bedingen. Rechnet man diesen - prinzipiell einfach eliminierbaren - Fehler aus den Resultaten heraus, so gelangt man zu dem Ergebnis, daß in allen Evaluationsdisziplinen, die Pronomen betreffen, mit Werten deutlich oberhalb von 70 Prozent auch im Falle von vergleichsweise schwierigen Texten gerechnet werden kann.

Ein weiteres zentrales Ergebnis der computerlinguistischen Untersuchungen zeigt auf, daß das in ROSANA und allen sonstigen bisherigen Ansätzen zugrundeliegende sequentielle Prozeßmodell prinzipiell durch ein theoretisch adäquates Verarbeitungsmodell ersetzbar ist, in dem syntaktische und referentielle Analyse interagieren (Abschnitt 10.4). Der in diesem Zusammenhang entwickelte Algorithmus löst den Anspruch einer robusten Sprachverarbeitung in erhöhtem Maße ein, indem die Ergebnisse der referentiellen Interpretation für weitere Verarbeitungsschritte der Syntaxanalyse - insbesondere der strukturellen Disambiguierung - herangezogen werden. Somit ergibt sich die aus computerlinguistischer Sicht bedeutsame Aussage, daß die Ansprüche einer robusten, anwendungstauglichen Analyse nicht in einem unvereinbaren Gegensatz zu der von der Theorie geforderten, in algorithmischen Analyseverfahren oft jedoch mehr oder weniger stillschweigend übergangenen Verschränkungen unterschiedlicher Beschreibungsebenen stehen. Betreffend die Teilprozesse der referentiellen und syntaktischen Interpretation belegt der beschriebene Algorithmus die informatische Einlösbarkeit der Zielsetzung einer integrierten Verarbeitung und somit die Gültigkeit der zentralen These 4 der Arbeit. Folglich wurde der konstruktive Nachweis erbracht, daß dem von Karlsson und Karttunen bezüglich jedweder massen-

datentauglichen Verarbeitung angemahnte ([CMU+95], S. 115)

“global view of how the various processing components relate to one another”

in bezug auf die Problemstellung einer integriert syntaktisch-referentiellen Analyse prinzipiell Rechnung getragen werden kann. Aufgrund der Symmetrie des Verarbeitungsmodells wird hierdurch nicht nur dem Anspruch einer robusten referentiellen Interpretation, sondern gleichzeitig dem einer robusten syntaktischen Analyse Rechnung getragen (vgl. die Ausführungen in Abschnitt 5.4.1, S. 69).

12.3 Der nächste Schritt: anwendungsbezogene Studien

Wenn somit das zentrale Ergebnis der Untersuchungen in einem Plädoyer für eine sensible Neubewertung der Tragfähigkeit einer algorithmischen Textinhaltserschließung unter Berücksichtigung der spezifischen inhaltsanalytischen Zielsetzungen und Rahmenbedingungen besteht, so stellt sich unmittelbar die Frage nach einer näheren Charakterisierung der Bedingungen, unter denen sich der Rekurs auf maschinelle Verfahren anbietet. Diese überaus komplexe Problemstellung läßt sich wie folgt paraphrasieren:

Welche konkreten inhaltsanalytischen Forschungsfragen lassen sich unter Verfügbarkeit welcher Form von Vertextung des Untersuchungsgegenstands unter Anwendung welcher computerlinguistischer Analysetechniken mit einem wie hohen Aufwand auf der Basis eines computergestützten Kategorisierungsvorgangs lösen?

Die Ergebnisse der vorliegenden Arbeit sind als geeigneter Anlaß zu werten, diese grundlegende Fragestellung auf kontextsensitive Weise neu aufzurollen.

Eine qualifizierte Beantwortung dieser Frage kann nur auf der Grundlage einer umfassenden Untersuchung typischer inhaltsanalytischer Anwendungen erfolgen. Die vorrangigen Aufgaben betreffen die Schnittstelle zwischen den vorgelagerten intellektuellen Verarbeitungsstufen und der maschinellen Vercodung. Auf welche Weise sollte die subjektiv-interpretative Entwicklung des Kategoriensystems mit den spezifischen Anforderungen der Algorithmisierung abgestimmt werden? Können Erhebungsverfahren formuliert werden, die den Untersuchungsgegenstand in einer Weise vertexten, die den Anforderungen einer maschinellen Weiterverarbeitung in besonderem Maße entspricht? Bereits elementare Zusatzbedingungen wie etwa die Vorschrift, eine Vertextung in grammatisch vollständigen Sätzen zu bewerkstelligen, sollten sich als nutzbringend erweisen.

Von zentraler Relevanz ist die Frage nach dem im jeweiligen Fall zu treibenden Aufwand. Gemäß einem zentralen Ergebnis der vorliegenden Arbeit wird zwar keine prinzipielle, theoretisch begründbare Schranke bezüglich der Algorithmisierbarkeit von Aufgaben der Textinhaltserschließung gesehen; stets jedoch verbleibt abzuwägen, ob sich der Aufwand einer algorithmischen Formalisierung im Hinblick auf die unter Rekurs auf gegenwärtig verfügbare computerlinguistische Analyseansätze erzielbaren Ergebnisse lohnt. Hierbei handelt es sich um ein Problem, dessen Diskussion vor dem Hintergrund u.a. der Anwendungsportabilität der zur Verfügung stehenden Analyseverfahren zu führen ist. Die im Rahmen der Analyse aktueller Verfahren des

Information Extraction identifizierten Forschungsfragen (vgl. S. 83 f.) bilden einen geeigneten Ausgangspunkt für entsprechende Untersuchungen.

Unter angemessener Würdigung dieser offenen Folgefragen lassen sich die Ergebnisse der Untersuchungen und Experimente der vorliegenden Arbeit in Anlehnung an ein zentrales Zitat der Arbeit Fühlaus (vgl. o., S. 37) folgendermaßen zusammenfassen:

Für die Protagonisten einer hermeneutischen Kritik mag die prinzipielle Unmöglichkeit einer algorithmischen Textinhaltserschließung endgültig und zufriedenstellend nachgewiesen sein - zu diskutieren ist die Adäquatheit von Computeranalysen vor dem Hintergrund konkreter inhaltsanalytischer Fragestellungen und Rahmenbedingungen sowie mit Blick auf die gegenwärtig verfügbaren computerlinguistischen Analyseverfahren unter sorgfältiger Abwägung des jeweils entstehenden Spezifikations- und Implementierungsaufwands, nicht jedoch per Verweis auf theoretisch nicht begründbare Schranken der Explizierbarkeit des impliziten Vorverständnisses der intellektuellen Textinterpretation.

Zwar befindet sich die Suche nach angemessenen und charakteristischen Anwendungen für eine linguistisch verfeinerte Computergestützte Inhaltsanalyse erst in ihrer initialen Phase - eine undifferenzierte, grundsätzliche Ablehnung der maschinellen Textinhaltserschließung sowie der algorithmischen Explikation qualitativer Prozesse schlechthin erscheint indes bereits heute nicht länger vertretbar.²

²Die Inhalte der vorliegenden Arbeit wurden maschinell generiert und sind daher im Hinblick auf die ultimative Objektivität des verwendeten Verfahrens ohne Unterschrift gültig.

Anhang A

ROSANA-Interpretation einer Pressemeldung

Anhand eines Texts des Korpus “*Presse-Meldungen*” soll die Arbeitsweise des ROSANA-Systems sowie der Ablauf der Bewertung der Interpretationsleistung im Detail diskutiert werden.

A.1 Pressemeldung Nr. 57

Der Beispieltext hat mehreren Anforderungen zu genügen. Er soll es einerseits ermöglichen, die wesentlichen Aspekte der Arbeitsweise des ROSANA-Systems sowie dessen prinzipielle Grenzen zu erläutern. Andererseits soll der Text charakteristisch für die Dokumente des Textkorpus “*Presse-Meldungen*” sein und dabei dennoch (aus expositorischen Gründen) eine überschaubare Größe aufweisen. Im Spannungsfeld dieser teils entgegengesetzten Bedingungen erweist sich Pressemeldung Nr. 57 des Evaluationskorpus (Texte 32 bis 66) als vergleichsweise am besten geeignet.

Im folgenden wird der Text in der Originalfassung wiedergegeben, in der er der algorithmischen referentiellen Interpretation durch ROSANA zugeführt wurde (vgl. Abbildung 11.8).

A Punch is Sometimes as Good as a Speech in Taiwanese Democracy.
Taiwanese legislators are finding that
good oratory is not the only skill needed to survive in Taiwan's
blooming democracy: a powerful right-hook also helps.

Opposition lawmaker Huang Chao-hui was back at work Wednesday
after being felled by a punch in the latest of the legislative
brawls that have marked the island's transition from virtual
dictatorship to democracy since 1987.

Huang, of the Democratic Progressive Party, had sought hospital
treatment for suspected concussion Tuesday after Lin Ming-yi of the
ruling Nationalists punched him during a debate.

Lin said he could not stomach Huang's taunts that Nationalist

lawmakers have recently started attending more legislative sessions only to try to ensure victory in elections later this year.

Lin later apologized for his violent outburst, and Huang was released from a hospital with bruises on his head.

On Monday, more than 10 lawmakers traded punches during a brawl started when opposition New Party legislator Ju Gau-jeng, nicknamed ‘‘Rambo’’ by Taiwanese newspapers, jumped onto the legislative speaker’s desk.

Tensions are running high in the legislature because lawmakers are debating a bill to govern a presidential election next March.

The vote is seen as a milestone in Taiwan’s march to democracy because it will mark the first time that the island’s president is elected by universal suffrage.

A.2 Satzweise Interpretation (OV- und KV-Aufgabe)

Die Interpretation des Beispieltexts durch ROSANA soll nun schrittweise - satzorientiert - nachvollzogen werden. Die Diskussion basiert auf den Zwischen- und Endergebnissen, die im Rahmen der Verarbeitung von Pressemeldung Nr. 57 durch ROSANA erzeugt werden. Die betrachteten Ausgaben lassen sich folgenden fünf Gruppen zuordnen.

- (1) Ergebnisse der syntaktischen Analyse (Ausgabe des Moduls *ost-konstruktor*): ggf. fragmentarische Beschreibung der *syntaktischen Oberflächenstruktur* des jeweiligen Satzes; Grundlage insbesondere der Verifikation der bindungstheoretischen Restriktionen.
- (2) Details und Ergebnisse der *Verifikation der bindungstheoretischen Restriktionen* für einzelne Paare bestehend aus Anapher und Antezedenskandidat: Bindungsprinzipien, i-über-i-Filter (vgl. Algorithmus in Abbildung 10.10, Schritt 1b).
- (3) Anaphern und zugehörige Listen bereits plausibilitäts-bewerteter und -sortierter, den Restriktionen genügenden *Antezedens-Kandidaten*; ggf. Details zum Beitrag der unterschiedlichen Prominenzfaktoren (Ergebnisse Schritt 2).
- (4) Details und Ergebnisse der Tests auf *Entscheidungs-Interdependenz* (Schritt 3).
- (5) *Ergebnis* der referentiellen Interpretation, Resultat des Auswahlprozesses: Zuordnung von Antezedenten zu den anaphorischen Vorkommen (Ergebnisse Schritt 3).

Für jeden der 9 Sätze werden zumindest die bezüglich der KV-Aufgabe getroffenen Interpretationsentscheidungen angegeben (Gruppe (5)), anhand derer auch die Ergebnisse der vorgelagerten Bestimmung der objektspezifizierenden Vorkommen (OV-Aufgabe) ersichtlich sind. Darüberhinaus erfolgt in bezug auf geeignete Beispiele exemplarisch eine nähere Aufschlüsselung der Arbeitsweise von ROSANA. Im Speziellen betrifft dies die Verifikation der bindungstheoretischen Bedingungen (2) sowie die Anwendung der Heuristiken zur Auswahl von Antezedenten (3). Anhand weiterer Fallbeispiele werden die prinzipiellen Grenzen von ROSANA erörtert.

Da die Interpretationsentscheidungen für die PS-Aufgabe erst auf der Basis der referentiellen

Analyse des gesamten Texts getroffen werden, sollen diese erst im nachfolgenden Abschnitt betrachtet werden.

A.2.1 Satz 1

A Punch is Sometimes as Good as a Speech in Taiwanese Democracy.

(5) Die Analyse des diskursinitialen Satzes bringt keine (notwendig intrasententiellen) referentiellen Wiederaufgriffsbeziehungen zutage. Die folgende Liste leerer Antezedenszuordnungen beschreibt somit ausschließlich die von ROSANA erkannten objektspezifizierenden Vorkommen:

AUSWAHL:
 1 25872 democracy
 1 25869 speech
 1 25866 good
 1 25862 punch

Die inkorrekte Identifikation des Adverbs *good* als Vorkommens-Induktor beruht auf einer Fehlklassifikation der syntaktischen Voranalyse, die die syntaktische Kategorie N (entsprechend der *nominalen* Lesart “*das Gut*”) zuordnet.

A.2.2 Satz 2

Taiwanese legislators are finding that good oratory is not the only skill needed to survive in Taiwan’s blooming democracy: a powerful right-hook also helps.

(3) Die Suche nach Antezedenskandidaten, die den Restriktionen genügen, führt zu einem einzigen Vorschlag für das durch die NP *Taiwan’s blooming democracy* induzierten Vorkommen: die Okkurrenz zu *Taiwanese democracy* in Satz 1, die sich durch die identische Stammform *democracy* des Kopf-Nomen qualifiziert (“lexikalische Rekurrenz”, vgl. Abbildung 11.16):

KANDIDATEN:
 Anapher: 2 25894 democracy
 & Kandidat: 1 25872 democracy, Begründung: LEXIKALISCHE_REKURRENZ
 Plausibilitaet: 55
 Anapher: 2 25898 right-hook
 Anapher: 2 25892 taiwan
 Anapher: 2 25887 skill
 Anapher: 2 25882 oratory
 Anapher: 2 25877 legislator

(5) Da keine Entscheidungs-Interdependenzen bestehen, wird dieser Kandidat auch gewählt:

AUSWAHL:

```

2 25894 democracy      ---> 1 25872 democracy
2 25898 right-hook
2 25892 taiwan
2 25887 skill
2 25882 oratory
2 25877 legislator

```

A.2.3 Satz 3

Opposition lawmaker Huang Chao-hui was back at work Wednesday after being felled by a punch in the latest of the legislative brawls that have marked the island's transition from virtual dictatorship to democracy since 1987.

(1) Die syntaktische Analyse generiert u.a. folgendes oberflächenstrukturelles Fragment:

```
--> OST-Fragment [Quelle] :: Nr. 2:
```

```

(P (NP brawl
  (S (NP that)
    (VP (NP transition
      (S (NP island)
        (NP (PP (NP dictatorship)))))))
    (PP (NP 1987))))))

```

Die referentielle Interpretation des Relativpronomens **that** ist somit unter alleinigem Rekurs auf die syntaktische Beschreibung möglich: Als Antezedens wird die Okkurrenz des Bezugsnomens **brawl** gewählt. Ausschließlich die übrigen objektspezifizierenden Vorkommen sind somit Gegenstand der weiteren Verarbeitung durch ROSANA.

(3) Wie in Satz 2 ergeben sich die Antezedenskandidaten der übrigen (nichtpronominalen) Okkurrenzen vermöge Zeichenkettenidentität der jeweiligen Kopfnomen:

KANDIDATEN:

```

Anapher: 3 25936 democracy
  & Kandidat: 2 25894 democracy, Begründung: LEXIKALISCHE_REKURRENZ
                Plausibilitaet: 55
  & Kandidat: 1 25872 democracy, Begründung: LEXIKALISCHE_REKURRENZ
                Plausibilitaet: 30

Anapher: 3 25918 punch
  & Kandidat: 1 25862 punch, Begründung: LEXIKALISCHE_REKURRENZ
                Plausibilitaet: 30

Anapher: 3 25938 1987
Anapher: 3 25934 dictatorship
Anapher: 3 25931 transition
Anapher: 3 25930 island
Anapher: 3 25925 brawl
Anapher: 3 25912 work wednesday
Anapher: 3 25907 opposition lawmaker huang chao-hui

```

Aufgrund der Heuristik der Bevorzugung weniger weit zurückliegender Antezedenten, die sich im Plausibilitätswert der Kandidaten niederschlägt, wird für das Vorkommen **democracy** der entsprechende Kandidat in Satz 2 bevorzugt.¹

(5) Da keine Entscheidungs-Interdependenzen bestehen, werden die jeweils plausibelsten Kandidaten gewählt:

```
AUSWAHL:
  3 25936 democracy      --->  2 25894 democracy
  3 25918 punch          --->  1 25862 punch
  3 25938 1987
  3 25934 dictatorship
  3 25931 transition
  3 25930 island
  3 25925 brawl
  3 25912 work wednesday
  3 25907 opposition lawmaker huang chao-hui
```

Anhand der Okkurrenz **opposition lawmaker huang chao-hui** wird ferner deutlich, daß RO-SANA auch über (syntaxbasierte) Heuristiken zur Extraktion zusammengesetzter Namen und NP verfügt. Die inkorrekte Identifikation eines Vorkommens **work wednesday** ist unmittelbar auf einen entsprechenden Fehler der syntaktischen Analyse zurückzuführen, die für die beiden Nomen auf eine *zusammengesetzte* NP erkannt hat; dies wird sich in einem Recall-Fehler (!) der OV-Aufgabe niederschlagen, da für das im Schlüssel spezifizierte Vorkommen **work** kein Pendant im systemgenerierten Ergebnis existiert.

A.2.4 Satz 4

Huang, of the Democratic Progressive Party, had sought hospital treatment for suspected concussion Tuesday after Lin Ming-yi of the ruling Nationalists punched him during a debate.

(1) Die syntaktische Analyse generiert folgende drei oberflächenstrukturellen Fragmente:

```
--> OST-Fragment :: Nr. 1:

(NP tuesday)

--> OST-Fragment :: Nr. 2:
```

¹Im vorliegenden Beispiel spielt es keine Rolle, welcher der beiden Kandidaten gewählt wird, da beide bereits demselben Diskursreferenten angehören. Prinzipiell könnten die Listen von Antezedenskandidaten für anaphorische Okkurrenzen auf jeweils denjenigen Kandidaten ein und desselben Diskursreferenten mit der höchsten Plausibilität reduziert werden. Da diese Operation jedoch zusätzliche Kosten verursacht, wird auf eine entsprechende Filterung verzichtet; ein A-Priori-Verzicht auf die Betrachtung mehrerer Kandidaten ein und desselben Diskursreferenten ist im Hinblick auf eine adäquate Verarbeitung von Pronomen des bindungstheoretischen Typs A (Reflexiva, Reziproka) nicht möglich.

```
(S (NP huang
    (NP party))
  (VP (NP hospital treatment))
  (S' (S (NP lin ming-yi
          (NP (PP (NP ruling nationalist))))
        (VP (NP him))
        (PP (NP debate))))))
```

--> OST-Fragment :: Nr. 3:

```
(NP concussion)
```

Nr. 2 ist das Kernfragment, das sowohl den Hauptsatz als auch den subordinierten Adverbialsatz (*after* Lin Ming-yi ...) beschreibt. Die Nichtzuordnung der beiden übrigen Fragmente, die den Phrasen *Tuesday* bzw. *for suspected concussion* entsprechen, ist unmittelbar auf entsprechende Unvollständigkeits der syntaktischen Analyse des DPE-Parsers zurückzuführen.

(2) Am Beispiel des Personalpronomens *him* vom bindungstheoretischen Typ B soll die Verifikation der syntaktisch-konfiguralen Restriktionen der Bindungstheorie erörtert werden. Hinsichtlich der bereits a priori verifizierten morphologischen Kongruenzbedingung sind u.a. sowohl *huang* als auch *lin ming-yi* zulässige *satzinterne* Typ-C-Antezedenten, da beiden - als *unbekannten* Lexemen - heuristisch die Genusattribute Maskulinum und Femininum zugeordnet werden. Es verbleibt deren syntaktisch-konfigurale Zulässigkeit zu überprüfen. Die Verifikation kann nichtheuristisch bewerkstelligt werden, da die jeweils betrachteten Vorkommenspaare ein und demselben oberflächenstrukturellen Fragment (Nr. 2) angehören (Algorithmus in Abb. 10.10, Schritt 1b, erster Unterpunkt). Bezüglich *huang* verlaufen die Tests erfolgreich, da sowohl die jeweiligen Bindungsprinzipien als auch die i-über-i-Bedingung erfüllt sind.

BT-Test fuer:

```
Anapher: 4 25966 him
Kandidat: 4 25942 huang
```

```
potentieller Binder: 4 25942 huang
potentiell Gebundenes: 4 25966 him
OST: (S (NP [Binder: huang]
            (NP party))
      (VP (NP hospital treatment))
      (S' (S (NP lin ming-yi
              (NP (PP (NP ruling nationalist))))
            (VP (NP [Ziel: him]))
            (PP (NP debate))))))
```

---> BP B erfuehlt

```
potentieller Binder: 4 25966 him
potentiell Gebundenes: 4 25942 huang
OST: (S (NP [Ziel: huang]
            (NP party))
      (VP (NP hospital treatment))
      (S' (S (NP lin ming-yi
              (NP (PP (NP ruling nationalist))))
            (VP (NP [Binder: him]))
            (PP (NP debate))))))
```

---> BP C erfuehlt

---> I-UEBER-I-BEDINGUNG erfuehlt

Umgekehrt verhält es sich für den Kandidaten *lin ming-yi*, der eine lokale Bindung der Pronominalanapher induzieren würde und somit auf der Basis von Bindungsprinzip B ausgeschlossen werden kann:

```
BT-Test fuer:
  Anapher:      4 25966 him
  Kandidat:     4 25960 lin ming-yi

  potentieller Binder:  4 25960 lin ming-yi
  potentiell Gebundenes: 4 25966 him
  OST: (S (NP huang
          (NP party))
        (VP (NP hospital treatment))
        (S' (S (NP [Binder: lin ming-yi]
                  (NP (PP (NP ruling nationalist))))
                (VP (NP [Ziel: him]))
                (PP (NP debate))))))
  ---> BP B verletzt
```

(3) Als Ergebnis der Anwendung der stringenten Strategien liegen folgende Listen potentieller Antezedenskandidaten vor:

```
KANDIDATEN:
  Anapher:      4 25966 him
    & Kandidat:  4 25942 huang, Begrueudung: KONGRUENZ
                  Plausibilitaet: 115
    & Kandidat:  4 25957 tuesday, Begrueudung: KONGRUENZ
                  Plausibilitaet: 90
    & ...
  Anapher:      4 25969 debate
  Anapher:      4 25964 ruling nationalist
  Anapher:      4 25960 lin ming-yi
  Anapher:      4 25957 tuesday
  Anapher:      4 25956 concussion
  Anapher:      4 25953 hospital treatment
  Anapher:      4 25942 huang
```

Die Prominenzheuristik SUBJEKT (Subjektpräferenz) führt in diesem Fall zur korrekten Entscheidung: *huang* wird gegenüber dem (in unspezifischer Rolle vorliegenden) Nichtsubjekt *tuesday* bevorzugt. Anhand des Kandidaten *tuesday* wird ferner die begrenzte Tragweite der zuvor genannten Heuristik zur Zuweisung von Genus-Attributen deutlich: Auch *tuesday* wird (inkorrekterweise) als Maskulinum/Femininum klassifiziert und somit - bei uneingeschränkter Kongruenz - als vergleichsweise plausibel bewertet.

Eine Verminderung des KV-Recall ergibt sich aufgrund der unterbliebenen Erkennung der Okkurrenz *opposition lawmaker huang chao-hui* aus Satz 3 als Antezedenskandidat von *huang* im vorliegenden Satz: Die gegenwärtig implementierte Heuristik zum Test auf Nachnamens-Übereinstimmung erweist sich in bezug auf chinesische Namen als unzureichend.

(4) Für den Antezedenskandidaten *huang* wird aus syntaktischen Gründen auf eine weitere satzinterne Koindexierung befunden, da der Parser die NP *the Democratic Progressive Party* inkorrekterweise als Apposition zu *huang* interpretiert hat, deren Vorliegen - entsprechend den

Vorgaben der KV-Aufgaben-Definition - heuristisch als Kospezifikations-Induktor interpretiert wird. Folglich ist zu verifizieren, daß die transitiv resultierenden Koindexierungen von **party** und **him** miteinander verträglich sind (Algorithmus, Schritt 3). Im vorliegenden Fall verläuft der Test auf Entscheidungs-Interdependenz erfolgreich:

```

Interdependenztest fuer:
  Anapher:      4 25966 him
  Kandidat:     4 25942 huang

- Kongruenz: VOLL

- BT:

  potentieller Binder:  4 25948 party
  potentiell Gebundenes: 4 25966 him
  OST: (S (NP huang
          (NP [Binder: party]))
        (VP (NP hospital treatment))
        (S' (S (NP lin ming-yi
                (NP (PP (NP ruling nationalist))))
              (VP (NP [Ziel: him]))
              (PP (NP debate))))))
  ---> BP B erfuehlt

  potentieller Binder:  4 25966 him
  potentiell Gebundenes: 4 25948 party
  OST: (S (NP huang
          (NP [Ziel: party]))
        (VP (NP hospital treatment))
        (S' (S (NP lin ming-yi
                (NP (PP (NP ruling nationalist))))
              (VP (NP [Binder: him]))
              (PP (NP debate))))))
  ---> BP C erfuehlt

  ---> I-UEBER-I-BEDINGUNG erfuehlt

```

(5) Da somit keine Entscheidungs-Interdependenzen bestehen, wird der plausibelste Kandidat gewählt:

```

AUSWAHL:
  4 25966 him          ---> 4 25942 huang
  4 25969 debate
  4 25964 ruling nationalist
  4 25960 lin ming-yi
  4 25957 tuesday
  4 25956 concussion
  4 25953 hospital treatment
  4 25942 huang

```

A.2.5 Satz 5

Lin said he could not stomach Huang's taunts that Nationalist lawmakers have recently started attending more legislative sessions only to try to ensure victory in elections later this year.

(1) Die Fragmentierung der oberflächenstrukturellen Beschreibung ist teils auf die Unvollständigkeit bereits des DPE-Ergebnisses - fehlende Anknüpfung des Fragments Nr. 1 des Objektsatzes mit Korrelat (*taunts that ...*) - und teils auf die eingeschränkte Regelbasis des Abbildungsmoduls *ost-konstruktor* - Fragment Nr. 3, Objektsatz der indirekten Rede ([*that*] *he could not stomach ...*) mit fehlendem *that* - zurückzuführen:

--> OST-Fragment :: Nr. 1:

```
(S (NP nationalist lawmaker)
  (VP (S (VP (NP session)
            (VP (S (VP (S (VP (NP victory))
                        (PP (NP election))
                        (NP year))))))))))
```

--> OST-Fragment :: Nr. 2:

```
(S (NP lin)
  (VP ))
```

--> OST-Fragment :: Nr. 3:

```
(S (NP he)
  (VP (NP taunt
        (S (NP huang))))))
```

Zusätzlich ist aus den Strukturen des DPE-Ergebnisses bekannt, daß Fragment Nr. 3 durch Fragment Nr. 2 *dominiert* wird.

(2) Auf der Basis der verfügbaren Fragmentierungs- und Dominanzinformation ergibt sich für das Personalpronomen *he* und den Antezedenskandidaten *lin* eine Konstellation, in der eine definitive Verifikation der Bindungsbedingungen trotz deren Lokalisation in unterschiedlichen oberflächenstrukturellen Fragmenten möglich ist. Die Grundlage hierfür bildet das in Abbildung 10.5 beschriebene Regelschema [E1a], dessen Anwendungsvoraussetzungen erfüllt sind, da die Anapher vom Typ B ist, das Fragment der Anapher (Nr. 2) deren Bindende Kategorie (den obersten S-Knoten) enthält und ferner dem Fragment des Kandidaten subordiniert ist. Folglich gelangt ROSANA zu dem Ergebnis, daß die Bindungsprinzipien der beteiligten Okkurrenzen erfüllt sind:

BT-Test fuer:

```
Anapher: 5 25975 he
Kandidat: 5 25973 lin
```

.... in unterschiedlichen Fragmenten.
Definitives Kriterium:

[E1a](-[IEa]): Typ-B-Anapher in lokaler Domaene, Fragment eingebettet
 ---> BPs erfuehlt

Das *Nicht*vorliegen einer lokalen i-über-i-Konfiguration, die gemäß einer abgeschwächten Regel [IEa] mit begrenztem Skopus auszuschließen wäre, ist bereits mitverifiziert worden. Unter bindungstheoretischen Gesichtspunkten ist die Koindexierung somit definitiv zulässig.

(3) Als Ergebnis der Anwendung der stringenten Strategien liegen folgende Listen potentieller Antezedenskandidaten vor.

KANDIDATEN:
 Anapher: 5 25979 huang
 & Kandidat: 4 25942 huang, Begrueudung: BEKANNTER_NAME
 Plausibilitaet: 200
 Anapher: 5 25975 he
 & Kandidat: 5 25973 lin, Begrueudung: KONGRUENZ
 Plausibilitaet: 135
 & Kandidat: 4 25960 lin ming-yi, Begrueudung: KONGRUENZ
 Plausibilitaet: 110
 & Kandidat: 4 25942 huang, Begrueudung: KONGRUENZ
 Plausibilitaet: 110
 & ...
 Anapher: 5 26001 year
 Anapher: 5 25998 election
 Anapher: 5 25996 victory
 Anapher: 5 25990 session
 Anapher: 5 25983 nationalist lawmaker
 Anapher: 5 25980 taunt
 Anapher: 5 25973 lin

Erneut ist es die Subjektpräferenz SUBJEKT, die in Verbindung mit der Strategie DISTANZ (Faktor SDM) zur Bevorzugung eines korrekten Kandidaten führt.

(5) Da keine Entscheidungs-Interdependenzen bestehen, werden die jeweils plausibelsten Kandidaten gewählt:

AUSWAHL:
 5 25979 huang ---> 4 25942 huang
 5 25975 he ---> 5 25973 lin
 5 26001 year
 5 25998 election
 5 25996 victory
 5 25990 session
 5 25983 nationalist lawmaker
 5 25980 taunt
 5 25973 lin

A.2.6 Satz 6

Lin later apologized for his violent outburst, and Huang was released from a hospital with bruises on his head.

(2) Am Beispiel des letzteren Vorkommens des Pronomens *his* läßt sich die Arbeitsweise des *i-über-i-Filters* veranschaulichen, der insbesondere die Koindexierung von Possessiva mit dem jeweils dominierenden Bezugs-Nomen blockiert.

```

BT-Test fuer:
  Anapher:    6 26009 his
  Kandidat:   6 26011 outburst

  potentieller Binder:    6 26011 outburst
  potentiell Gebundenes:  6 26009 his
  OST: (S (NP lin)
        (VP )
        (PP (NP [Binder: outburst]
                (S (NP [Ziel: his])))))
  ---> BP B erfuehlt

  potentieller Binder:    6 26009 his
  potentiell Gebundenes:  6 26011 outburst
  OST: (S (NP lin)
        (VP )
        (PP (NP [Ziel: outburst]
                (S (NP [Binder: his])))))
  ---> BP C erfuehlt

  ---> I-UEBER-I-BEDINGUNG verletzt

```

Die Bindungsprinzipien führen nicht zum (erwünschten) Ausschluß der Koindexierung, da die K-Herrschafts- und somit die Bindungsbeziehung auf Fälle beschränkt ist, in denen keine der betrachteten NP die jeweils andere dominiert; die Restriktion ergibt sich erst auf der Basis des *i-über-i-Filters*.

(3) Bezüglich des zweiten Vorkommens des Possessivpronomens *his* ergibt sich erstmals die Situation, daß die anzuwendenden Prominenzfaktoren nicht zur Bevorzugung eines korrekten Antezedenskandidaten führen. Es soll daher eine nähere Aufschlüsselung der Faktoren-Beiträge für die beiden plausibelsten Kandidaten betrachtet werden:

```

Anapher:    6 26023 his
  & Kandidat:    6 26009 his, Begrueudung: KONGRUENZ
                  Plausibilitaet: 110
                  - BT fragment-heuristisch: JA
                  - Kongruenz: VOLL
                  - Rollentraegheit: SYNTAKTISCH
                  - Subjekt: NEIN
                  - Satzdistanz: 0
                  - Richtung: ANAPHORISCH
  & Kandidat:    6 26014 huang, Begrueudung: KONGRUENZ
                  Plausibilitaet: 105
                  - BT fragment-heuristisch: JA
                  - Kongruenz: VOLL
                  - Rollentraegheit: NEIN
                  - Subjekt: JA
                  - Satzdistanz: 0
                  - Richtung: ANAPHORISCH

```

Aufgrund des für Possessivpronomina spezifizierten Rollenträgheitskriteriums SYNPARALL wird das erste *his*-Vorkommen gegenüber dem in Subjektfunktion realisierten korrekten Kandidaten *huang* bevorzugt. Für ersteren Kandidaten kommt eine kumulierte Plausibilität von $APRN + SYR - FRM = 100+20-10 = 110$ zustande gegenüber $APRN + SUP - FRM = 100+15-10 = 105$ für letzteren Kandidaten (vgl. Abbildung 11.19, S. 257).

Das Beispiel illustriert die Grenzen der Auswahlstrategie von ROSANA, die eben nur ein (wenn auch erfolgreiches) heuristisches Substitut, nicht jedoch einen vollständigen Ersatz für eine auf tiefergehenden semantischen und pragmatischen Kriterien basierende Entscheidungsfindung darstellt. Die durch die Wahl der Faktorengewichte spezifizierte Präferenz eines rollenträgen Wiederaufgriffs gegenüber der Resumption eines Subjekts für Possessiva hat sich in bezug auf die kumulierten Evaluationsergebnisse des Trainingskorpus bewährt; eine Garantie der Korrektheit der Strategie in bezug auf Einzelfälle geht hiermit jedoch nicht einher.

Für die übrigen anaphorischen Okkurrenzen werden jeweils die korrekten Kandidaten bevorzugt:

```
KANDIDATEN:
Anapher: 6 26005 lin
  & Kandidat: 5 25973 lin, Begründung: BEKANNTER_NAME
  Plausibilitaet: 200

Anapher: 6 26014 huang
  & Kandidat: 5 25979 huang, Begründung: BEKANNTER_NAME
  Plausibilitaet: 200

Anapher: 6 26009 his
  & Kandidat: 6 26005 lin, Begründung: KONGRUENZ
  Plausibilitaet: 115
  & Kandidat: 5 25979 huang, Begründung: KONGRUENZ
  Plausibilitaet: 95
  & ...

Anapher: 6 26024 head
Anapher: 6 26021 bruise
Anapher: 6 26019 hospital
Anapher: 6 26011 outburst
```

(5) Da keine Entscheidungs-Interdependenzen bestehen, werden die jeweils plausibelsten Kandidaten gewählt:

```
AUSWAHL:
6 26005 lin      ---> 5 25973 lin
6 26014 huang   ---> 5 25979 huang
6 26009 his     ---> 6 26005 lin
6 26023 his     ---> 6 26009 his
6 26024 head
6 26021 bruise
6 26019 hospital
6 26011 outburst
```

A.2.7 Satz 7

On Monday, more than 10 lawmakers traded punches during a brawl started when opposition New Party legislator Ju Gau-jeng, nicknamed ‘‘Rambo’’ by Taiwanese newspapers, jumped onto the legislative speaker’s desk.

(5) ROSANA trifft keine Antezedensentscheidungen. Dies erweist sich als korrekt, da die objektspezifizierenden Vorkommen von Satz 7 ohne Ausnahme neue Diskursreferenten einführen. (Die Kohäsion des Texts wird auf anderen Ebenen als der der unmittelbaren Kospezifikation etabliert - insbesondere Isotopie kommt zum Tragen.)

Das einzige Problem besteht in bezug auf die identifizierte Okkurrenz *opposition*, die - als eigentlichem Prämodifizierer zu *party legislator ju gau-jeng* - entsprechend der Spezifikation der OV-Aufgabe irrelevant ist und somit zu einem Precision-Fehler führt. (Dieser Fehler ist unmittelbar auf einen Defekt des Ergebnisses der externen DPE-Voranalyse zurückzuführen und demnach kein unmittelbarer Mangel von ROSANA.)

AUSWAHL:

7 26046 party legislator ju gau-jeng
 7 26061 desk
 7 26060 speaker
 7 26054 newspaper
 7 26050 rambo
 7 26041 opposition
 7 26038 brawl
 7 26035 punch
 7 26033 10 lawmaker
 7 26028 monday

A.2.8 Satz 8

Tensions are running high in the legislature because lawmakers are debating a bill to govern a presidential election next March.

(5) Auch hier trifft ROSANA in Übereinstimmung mit dem Schlüssel keine Antezedensentscheidungen. Die Menge der identifizierten Vorkommen erweist sich als korrekt und vollständig.

AUSWAHL:

8 26084 march
 8 26082 election
 8 26077 bill
 8 26073 lawmaker
 8 26071 legislature
 8 26065 tension

A.2.9 Satz 9

The vote is seen as a milestone in Taiwan's march to democracy because it will mark the first time that the island's president is elected by universal suffrage.

(3) Für die Pronominalanapher *it* führen die Prominenzfaktoren SYNROLLE (grammatische Rollenhierarchie) und ferner SYNPARALL zur Bevorzugung des richtigen Kandidaten *vote* gegenüber (u.a.) dem als sächsischen Genitiv realisierten Vorkommen *taiwan*:

Anapher: 9 26100 it
 & Kandidat: 9 26088 vote, Begründung: KONGRUENZ
 Plausibilität: 135
 - BT fragment-heuristisch: NEIN
 - Kongruenz: VOLL
 - Rollentragheit: SYNTAKTISCH
 - Subjekt: JA
 - PossGenAttr: NEIN
 - exist. Emphase: NEIN
 - dir. Objekt: NEIN
 - indir. Objekt: NEIN
 - adv.PP-Objekt: NEIN
 - Satzdistanz: 0
 - Richtung: ANAPHORISCH

& Kandidat: 9 26095 taiwan, Begründung: KONGRUENZ
 Plausibilität: 113
 - BT fragment-heuristisch: NEIN
 - Kongruenz: VOLL
 - Rollentragheit: NEIN
 - Subjekt: NEIN
 - PossGenAttr: JA
 - exist. Emphase: NEIN
 - dir. Objekt: NEIN
 - indir. Objekt: NEIN
 - adv.PP-Objekt: NEIN
 - Satzdistanz: 0
 - Richtung: ANAPHORISCH

& ...

Interessant ist darüberhinaus die Betrachtung der NP *Taiwan's march*, für das der inkorrekte Kandidat *next march* aus dem vorangegangenen Satz vorgeschlagen wird:

Anapher: 9 26096 march
 & Kandidat: 8 26084 march, Begründung: LEXIKALISCHE_REKURRENZ
 Plausibilität: 55

Die Konfusion der Homonyme *march* im Sinne von *Marsch* und *march* im Sinne von *März* verdeutlicht die Grenzen des zeichenkettenbasierten Kongruenztests für Anaphern des Typs **Nomen** (vgl. Abbildung 11.16, S. 253).

Für die übrigen Okkurrenzen werden die richtigen Kandidaten bevorzugt.

(5) Somit ergeben sich die folgenden Antezedensentscheidungen:

AUSWAHL:

9 26095 taiwan	---	2 25892 taiwan
9 26100 it	---	9 26088 vote
9 26096 march	---	8 26084 march
9 26098 democracy	---	3 25936 democracy
9 26108 island	---	3 25930 island
9 26114 suffrage		
9 26109 president		
9 26105 first time		
9 26093 milestone		
9 26088 vote		

A.3 Bestimmung nichtpronominaler Substitute (PS-Aufgabe)

Auf der Grundlage der als Ergebnis der satzweisen Analyse von Pressemeldung Nr. 57 vorliegenden referentiellen Textrepräsentation ermittelt der in Abbildung 11.21, S. 259 beschriebene Algorithmus folgende lexikalische Substitute für die pronominalen Okkurrenzen: (Die Zahl in eckigen Klammern ist die Nummer des zu dem jeweiligen Vorkommen gehörenden Diskursreferenten, die i.d.R. mit der Nummer der Okkurrenz übereinstimmt, durch die der Diskursreferent erstmals realisiert wird.)

LEXIKALISCHE SUBSTITUTION DER PRONOMEN:

3	25926	:	[25925]	that	==	3	25925	:	brawl
4	25966	:	[25942]	him	==	4	25942	:	huang
5	25975	:	[25973]	he	==	5	25973	:	lin
6	26009	:	[25973]	his	==	6	26005	:	lin
6	26023	:	[25973]	his	==	6	26005	:	lin
9	26100	:	[26088]	it	==	9	26088	:	vote

Die nichtpronominalen Vorkommen auf der rechten Seite gehören jeweils demselben Diskursreferenten wie das zu ersetzende Pronomen an. Der einzige Fehler, der unmittelbar auf die entsprechende inkorrekte Antezedensentscheidung während der Analyse von Satz 6 zurückzuführen ist, entsteht für die zweite Okkurrenz des Possessivums *his*; korrekt wäre ein beliebiges *huang*-Vorkommen.

A.4 Bewertung der Interpretationsleistung

Abschließend sollen die Ergebnisse der Bewertung der Interpretationsleistung von ROSANA auf Pressemeldung Nr. 57 betrachtet werden. Die Bewertung wird softwaregestützt durch das Modul **bewerter** vorgenommen, das die ROSANA-Resultate für die drei Teildisziplinen OV, KV und PS mit den Vorgaben des für den Evaluationskorporus intellektuell erstellten Schlüssels abgleicht (vgl. Abschnitt 11.2.4). Formale Grundlage sind die in Abschnitt 11.1.4 definierten Evaluationsmaße.

A.4.1 Bestimmung objektspezifizierender Vorkommen (OV-Aufgabe)

Insgesamt existieren zwei systemseitig generierte Okkurrenzen, die nicht im Schlüssel spezifiziert sind, sowie zwei im Schlüssel vorgegebene Vorkommen, die nicht von ROSANA erkannt werden. Somit liegen zwei Precision- und zwei Recall-Fehler vor. Der Abgleich ergibt ferner, daß 70 Vorkommen in der Schnittmenge von Systemergebnis und Schlüsselvorgabe liegen. Im Ergebnis der softwaregestützten Bewertung liest sich dies folgendermaßen:

ZUORDNUNG OKKURRENZEN:

- NUR ANA:
25866: C good
26041: C opposition

- NUR KEY:
25911: work
25921: latest

EVALUATIONSERGEBNIS OKKURRENZEN:

- NUR ANA: 2
- NUR KEY: 2
- ANA UND KEY: 70
=> PRECISION: 0.9722
=> RECALL: 0.9722

Im oberen Teil der Ausgabe sind die Precision- bzw. Recall-Fehler näher aufgeschlüsselt. Beispielsweise steht die Zeile 25866: C good für das Vorkommen good in Satz 1, das aufgrund eines Fehlers der syntaktischen Voranalyse irrtümlicherweise angelegt wurde und somit zu einem Precision-Fehler führt; umgekehrt induziert die in Satz 3 fälschlicherweise *nicht* erkannte Okkurrenz 25911: work einen Recall-Fehler.

Die berechneten Precision- und Recall-Werte entstehen durch Anwendung der auf S. 235 definierten Formeln für P_{ov} bzw. R_{ov} :

$$P_{ov} = \frac{70}{70 + 2} = 0.9722, \quad R_{ov} = \frac{70}{70 + 2} = 0.9722$$

A.4.2 Bestimmung der Kospezifikations-Äquivalenzklassen (KV-Aufgabe)

Die Bewertung der Interpretationsleistung von ROSANA in bezug auf die KV-Aufgabe geschieht auf der Basis des auf S. 235 ff. skizzierten Verfahrens des wechselseitigen Abgleichs systemseitig generierter und schlüsselseitig spezifizierter Kospezifikationsklassen. Precision- und Recall-Fehler sind unter Rekurs auf die Partitionierungen zu bestimmen, die die Kospezifikationsrelation des Schlüssels auf den systemgenerierten Klassen bzw. die systemgenerierte Kospezifikationsrelation auf den Schlüssel-Klassen induziert.

Im Rahmen des computergestützten Bewertungsvorgangs lassen sich die Klassen sowie die entstehenden Partitionierungen näher aufschlüsseln. Die insgesamt drei Fehlentscheidungen von ROSANA im Rahmen der Antezedensauswahl (vgl. o.) schlagen sich in entsprechenden nicht-trivialen Partitionierungen der generierten Koreferenzklassen nieder (Die langen Trennstriche grenzen die systemgenerierten Klassen ab; die kurzen Trennstriche visualisieren die durch die Schlüsselvorgaben induzierten Partitionierungen.):

```
-----
25942: huang
25966: him
25979: huang
26014: huang
-----
25948: party
-----
1 / 4
```

```

-----
25973: lin
25975: he
26005: lin
26009: his
-----
26023: his
-----
1 / 4

-----
26084: march
-----
26096: march
-----
1 / 1

```

Beispielsweise gehören die Okkurrenzen der zuerst beschriebenen Klasse zwei unterschiedlichen Klassen des Schlüssels an, wodurch ein Precision-Fehler (von insgesamt vier möglichen: 1 / 4) entsteht (vgl. Abbildung 11.5); dieser Fehler entsteht im Rahmen der Interpretation von Satz 4, in der irrtümlicherweise auf Koreferenz der Vorkommen 25942: **huang** und 25948: **party** entschieden wird (vgl. o.). Entsprechend der Definition von P_{kv} (vgl. S. 238) ergibt sich das über alle Klassen kumulierte Ergebnis

```

A.R. -ERGEBNIS-PARTITIONIERUNG
- SCHNITTE: 3
- MOEGLICH: 17
=> PRECISION: 0.8235

```

$$(P_{kv} = \frac{17-3}{17} = 0.8235)$$

Die Ermittlung der Recall-Fehler geschieht symmetrisch über die Betrachtung der Partitionierung der Koreferenzklassen des Schlüssels. In insgesamt vier Fällen kommt es hierbei zu nicht-trivialen Partitionierungen, die allesamt *Auslassungen* von ROSANA in bezug auf die Rekonstruktion der Schlüsselklassen entsprechen:

```

-----
25862: Punch
25918: punch
-----
25898: right-hook
-----
1 / 2

-----
25892: Taiwan's
26095: Taiwan's
-----
25930: island's
26108: island's
-----
1 / 3

```

```

-----
25907: Chao-hui
-----
25942: Huang
25966: him
25979: Huang's
26014: Huang
-----
26023: his
-----
2 / 5

-----
25960: Ming-yi
-----
25973: Lin
25975: he
26005: Lin
26009: his
-----
1 / 4

```

Beispielsweise entstehen die Recall-Fehler für die drittgenannte Klasse durch die fehlende korrekte Zuordnung des Pronomens 26023: his in Satz 6 sowie ferner durch die für chinesische Namen unzureichenden Namensbehandlungs-Heuristiken, die eine Identifikation der Okkurrenzen Huang Chao-hui und Huang (noch) nicht bewerkstelligen (vgl. Satz 3).

Insgesamt liegen 5 von 19 möglichen Partitionierungen vor; der kumulierte Recall R_{kv} (vgl. S. 238) beträgt somit

```

KEY-PARTITIONIERUNG
- SCHNITTE: 5
- MOEGLICH: 19
=> RECALL: 0.7368

```

$$(R_{kv} = \frac{19-5}{19} = 0.7368)$$

A.4.3 Bestimmung lexikalischer Substitute für Pronomen (PS-Aufgabe)

Die Bewertung der ROSANA-Leistung in bezug auf die Identifikation nichtpronominaler Substitute basiert auf einer Einordnung der systemgenerierten Lösungen in die in Abbildung 11.7 definierten Mengen, deren Kardinalitäten in die Evaluationsmaße einfließen. Für die durch ROSANA für die Pronomen der betrachteten Pressemeldung ermittelten Substitute ergibt sich folgendes Bild (in der linken Spalte ist das Subskript der jeweiligen Menge angegeben):

LEXIKALISCHE SUBSTITUTION

```

++ 3 25926: that == 3 25925: brawl
++ 4 25966: him == 4 25942: huang
++ 5 25975: he == 5 25973: lin
++ 6 26009: his == 6 26005: lin
+- 6 26023: his == 6 26005: lin
++ 9 26100: it == 9 26088: vote

```

Eine einzige Entscheidung wird als fehlerhaft klassifiziert; da das entsprechende Substitut im Schlüssel als Okkurrenz ausgezeichnet ist und das Pronomen darüberhinaus nicht als *optional kospesifizierend* markiert ist, erfolgt die Einordnung in die Menge o_{+-} .

Entsprechend den Definitionen der Evaluationsmaße P_{ps} und R_{ps} (vgl. S. 242) ergeben sich folgende Precision- bzw. Recall-Werte für die unterschiedlichen Typen von Pronomen² bzw. im Durchschnitt:

	PRECIS	RECALL	++	+-	+?	+ ₋	++	?+	? ₋
PE-3	1.0000	1.0000	3	0	0	0	0	0	0
PE12	--	--	0	0	0	0	0	0	0
PO-3	0.5000	0.5000	1	1	0	0	0	0	0
PO12	--	--	0	0	0	0	0	0	0
REFL	--	--	0	0	0	0	0	0	0
RELA	1.0000	1.0000	1	0	0	0	0	0	0

DURCHSCHNITT JE PRONOMEN

=> PRECISION: 0.8333

=> RECALL: 0.8333

Beispielsweise ergibt sich für Possessivpronomen eine Precision P_{ps} von

$$P_{ps} = \frac{|o_{++}|}{|o_{++}| + |o_{+-}| + |o_{+?}|} = \frac{1}{1 + 1 + 0} = 0.5$$

Der Recall-Wert R_{ps} ist identisch, da die Menge o_{+-} leer ist.

²Vgl. S. 226: PE-3/PE12 = Personalpronomen in dritter bzw. erster/zweiter Person, PO-3/PO12 = Possessivpronomen in dritter bzw. erster/zweiter Person, REFL = Reflexiv- und Rezipropronomen, RELA = Relativpronomen.

Anhang B

ROSANA-Interpretation eines “*Mozart-Opern*”-Texts

Für das Dokument “*Don Giovanni*” des Korpus “*Mozart-Opern*” werden vollständige Listen der Antezedenten sowie der nichtpronominalen Substitute angegeben, die ROSANA für die anaphorischen bzw. pronominalen Okkurrenzen dieses Texts ermittelt. Um einen Überblick der Qualität der Interpretationsleistung herzustellen, wird jede Entscheidung hinsichtlich ihrer Korrektheit klassifiziert. Darüberhinaus werden die Precision- und Recall-Werte für die entsprechenden Interpretationsdisziplinen KV bzw. PS angegeben.

B.1 Text “*Don Giovanni*”

Das Dokument wurde der Anlage der CD “*Don Giovanni*” (Philips: 438 494-2; Textautor: Lionel Salter) entnommen. Mit 898 Wörtern handelt es sich um das längste Dokument des Korpus “*Mozart-Opern*”.

Act I.

On a dark night in Seville, Leporello is keeping watch, grumbling, outside a house in which his master Don Giovanni is engaged in his latest amorous pursuit. His target is Donna Anna, daughter of the Commendatore; but she, discovering only that the intruder is not her fiancée Don Ottavio, rouses the household and hampers Giovanni’s escape. Her father appears and challenges him to a duel which he reluctantly accepts. They fight, the old man is killed, and Giovanni makes off. Anna has meantime fetched Ottavio, and finding her father lying dead makes Ottavio swear revenge on the unknown murderer.

As Giovanni and Leporello make their way through the deserted

streets, they perceive a young woman in distress. Seizing his opportunity, Giovanni goes to console her, only to find that it is Donna Elvira whom he has already seduced and deserted. She turns on him furiously, but he pushes Leporello forward to distract her attention while he makes his escape. Leporello half censoriously, half admiringly, reads her the extensive catalogue of his master's conquests.

In the next scene, set in the countryside, villagers celebrate the forthcoming wedding of Masetto and the pretty Zerlina, on whom Giovanni's roving eye immediately alights. He invites everyone to his villa nearby for refreshments, ordering Leporello to take special care of Masetto, who, though demurring at leaving his bride alone with the handsome nobleman, is dragged away. Giovanni exercises all his charm on Zerlina, who is ready to succumb to his advances when Elvira suddenly appears and, warning Zerlina of her peril, sweeps her away. Donna Anna comes by with Ottavio and, recognising Giovanni only as a nobleman of her acquaintance, asks his help in discovering her father's murderer. While he is promising to do anything he can, Elvira returns and denounces him. Embarrassed, he tries to pass this off as an outburst of hysteria; but their suspicions have been aroused, and when Anna is alone with Ottavio she affirms that she now recognises Giovanni as the would-be seducer who killed her father, and reminds Ottavio of his oath. Meanwhile, not in the least deterred, Giovanni orders Leporello to prepare a grand supper party for all the villagers at his home, where he is confident of increasing the score in Leporello's catalogue.

The villagers begin to arrive in Giovanni's garden, among them Zerlina, trying to soothe her jealous Masetto. Giovanni renews his attentions to her, and when Masetto inconveniently intervenes he coolly leads them into the house to join the dancing. Now Elvira, Anna and Ottavio enter, wearing elegant masks and set on vengeance. At Giovanni's bidding, Leporello invites the newcomers inside. During the dancing Giovanni manages to inveigle Zerlina into an adjoining room; but her alarmed cries for help bring everyone to her aid. Giovanni tries to bluff it out by pretending he was rescuing her from Leporello, but no one believes him, and when his three new guests unmask and face him there is general confusion, during which he somehow manages to slip away.

Act II.

Elvira's maid has attracted Giovanni's eye, and after buying off Leporello's reluctance to continue in his service, Giovanni forces him to exchange clothes with him so as to carry out his latest design. Standing in the dusk under Elvira's balcony as she laments her lot, he lures her down into the street, and she, mistaking the servant for the master, is led off by Leporello, leaving the field clear for Giovanni to serenade his latest fancy to the accompaniment

of a mandolin. But at this moment along comes Masetto with a party of his friends, all armed with sticks and looking for the man who had tried to seduce Zerlina. Giovanni, whom they mistake for Leporello, gives them a description of the clothes Leporello now has on and sends them off in various directions to search for him. When they have gone, he disarms Masetto and gives him a drubbing. Zerlina, coming to look for her bridegroom, finds him nursing his bruises, and she comforts him.

Anna and Ottavio, as well as Masetto and Zerlina, are hot on the supposed Giovanni's trail, and coming across Leporello in the darkness with Elvira are about to attack him. But when Elvira implores them to spare her husband, Leporello reveals his true identity - to everyone's amazement - and manages to run away. Ottavio now feels he has enough proof of Giovanni's guilt to go to the civil authorities. Left alone Elvira confesses that she feels pity for Giovanni and fear for his peril in spite of the wrongs he has done her.

In a churchyard to which Giovanni and Leporello have escaped, Giovanni's ribald laughter over his recent escapades is interrupted by a ghostly voice from the statue over the Commendatore's grave, warning him of a terrible retribution for his misdeeds. Giovanni's nonchalant reply is to order the trembling Leporello to invite it to supper that night: to Leporello's horror, the statue seems to nod its head in acceptance.

In Anna's house Ottavio, seeking to console her, suggests that they should marry straightaway. When she refuses and he charges her with cruelty, she protests that she does love him but is still too grief-stricken at her father's death, and begs him to have patience.

The finale is set in Giovanni's villa. He is at supper while a wind band plays popular operatic airs of the day. Elvira bursts in with a last appeal to him to mend his ways; but he merely mocks her, and she leaves in despair. As she departs, she is heard to scream outside: Leporello, sent to investigate, returns in terror to stammer that the Commendatore's statue is approaching. There is heavy knocking at the door, and the statue enters to accept Giovanni's invitation to supper. By the rules of hospitality, it says, will Giovanni now dine with him? Giovanni boldly accepts, but when, obliged to offer his hand in agreement, he feels it in an icy grip, a shudder of fear passes through him. Still, however, he rejects the statue's urgings that he should repent; and then, before Leporello's horrified gaze, the earth opens up and his stone guest drags Giovanni down into the flames of hell. When the other characters enter with ministers of justice to arrest Giovanni, Leporello tells them what has happened. In a final sextet they point the moral - that sinners cannot finally escape their deserts.

B.2 Antezedensentscheidungen

Die folgende Aufstellung schlüsselt die Antezedensentscheidungen auf, die ROSANA für die erkannten und als anaphorisch klassifizierten Okkurrenzen trifft. Um einen Überblick der Qualität des Interpretationsergebnisses zu geben, sind die einzelnen Zuordnungen entsprechend den Festlegungen in Abbildung 11.7, S. 241 hinsichtlich ihrer Korrektheit klassifiziert (linke Spalte).

A.R. -ENTSCHEIDUNGEN

+	?	2	25: which	---	>	2	19: grumble
+	+	2	26: his	---	>	2	14: leporello
+	-	2	33: his	---	>	2	26: his
+	+	3	40: his	---	>	2	33: his
+	+	3	52: she	---	>	3	44: donna anna
+	+	3	61: her	---	>	3	52: she
+	+	4	76: her	---	>	3	61: her
+	-	4	81: him	---	>	3	64: fiancee don ottavio
+	+	4	85: which	---	>	4	84: duel
+	+	4	86: he	---	>	4	81: him
+	-	5	92: they	---	>	4	85: which
+	+	5	102: giovanni	---	>	3	71: hampers giovanni
+	+	6	108: anna	---	>	3	44: donna anna
+	+	6	112: ottavio	---	>	3	64: fiancee don ottavio
+	+	6	116: her	---	>	6	108: anna
+	+	6	117: father	---	>	4	77: father
+	+	6	121: ottavio	---	>	3	64: fiancee don ottavio
+	+	7	132: giovanni	---	>	5	102: giovanni
+	+	7	134: leporello	---	>	2	14: leporello
+	+	7	136: their	---	>	7	133: and
+	+	7	143: they	---	>	7	133: and
+	-	8	154: his	---	>	7	134: leporello
+	+	8	157: giovanni	---	>	7	132: giovanni
+	-	8	161: her	---	>	8	155: opportunity
?	+	8	167: it	---	>	8	155: opportunity
+	+	8	171: whom	---	>	8	170: donna elvira
+	+	8	172: he	---	>	8	157: giovanni
+	+	9	181: she	---	>	8	171: whom
+	+	9	184: him	---	>	8	172: he
+	+	9	188: he	---	>	8	172: he
+	+	9	190: leporello	---	>	7	134: leporello
+	+	9	194: her	---	>	9	181: she
+	+	9	197: he	---	>	9	188: he
+	+	9	199: his	---	>	9	197: he
+	-	9	200: escape	---	>	3	72: escape
+	+	10	204: leporello	---	>	9	190: leporello
?	+	10	208: half	---	>	10	205: half
+	+	10	212: her	---	>	9	181: she
+	+	10	217: his	---	>	10	204: leporello
+	+	11	246: whom	---	>	11	243: zerlina
+	+	11	247: giovanni	---	>	8	157: giovanni
+	+	12	255: he	---	>	11	247: giovanni
+	+	12	259: his	---	>	12	255: he
+	+	12	266: leporello	---	>	10	204: leporello
+	+	12	272: masetto	---	>	11	239: masetto
+	+	12	274: who	---	>	12	272: masetto
+	-	12	280: his	---	>	12	259: his
+	+	13	294: giovanni	---	>	11	247: giovanni

++	13	297: his	---	13	294: giovanni
++	13	300: zerlina	---	11	243: zerlina
++	13	302: who	---	13	300: zerlina
++	13	308: his	---	13	297: his
++	13	311: elvira	---	8	170: donna elvira
+ -	13	319: her	---	13	311: elvira
++	13	323: her	---	13	319: her
++	14	329: donna anna	---	3	44: donna anna
++	14	333: ottavio	---	6	121: ottavio
++	14	337: giovanni	---	13	294: giovanni
++	14	341: nobleman	---	12	286: nobleman
++	14	343: her	---	14	329: donna anna
++	14	347: his	---	13	308: his
++	14	351: her	---	14	343: her
++	14	352: father	---	6	117: father
++	14	353: murderer	---	6	127: murderer
++	15	358: he	---	14	347: his
++	15	364: he	---	15	358: he
++	15	367: elvira	---	13	311: elvira
++	15	371: him	---	15	364: he
++	16	377: he	---	15	364: he
+ -	16	390: their	---	13	309: advance
++	16	398: anna	---	14	329: donna anna
++	16	402: ottavio	---	14	333: ottavio
++	16	403: she	---	16	398: anna
++	16	406: she	---	16	403: she
++	16	409: giovanni	---	14	337: giovanni
++	16	414: who	---	16	409: giovanni
+ -	16	416: her	---	16	414: who
++	16	417: father	---	14	352: father
++	16	421: ottavio	---	14	333: ottavio
+ -	16	423: his	---	16	414: who
++	17	436: giovanni	---	16	409: giovanni
++	17	438: leporello	---	12	266: leporello
++	17	448: villager	---	11	233: villager
++	17	450: his	---	17	436: giovanni
++	17	453: where	---	17	451: home
++	17	454: he	---	17	436: giovanni
++	17	462: leporello	---	12	266: leporello
++	17	463: catalogue	---	10	215: catalogue
++	18	468: villager	---	17	448: villager
++	18	473: giovanni	---	17	436: giovanni
++	18	477: them	---	18	468: villager
++	18	478: zerlina	---	13	317: warning zerlina
++	18	483: her	---	18	478: zerlina
++	18	485: masetto	---	12	272: masetto
++	19	489: giovanni	---	18	473: giovanni
++	19	491: his	---	19	489: giovanni
++	19	498: masetto	---	18	485: masetto
+ -	19	501: he	---	19	498: masetto
+ -	19	504: them	---	19	492: attention
+ -	19	507: house	---	2	23: house
++	20	516: elvira	---	15	367: elvira
++	20	518: anna	---	16	398: anna
++	20	520: ottavio	---	16	421: ottavio
++	21	534: giovanni	---	19	489: giovanni
++	21	537: leporello	---	17	462: leporello
++	22	547: dance	---	19	511: dance
++	22	548: giovanni	---	21	534: giovanni

++	22	552: zerlina	---	18	478: zerlina
++	22	559: her	---	22	552: zerlina
+ -	22	563: help	---	14	348: help
++	22	567: her	---	22	559: her
++	23	572: giovanni	---	22	548: giovanni
?+	23	576: it	---	22	568: aid
++	23	580: he	---	23	572: giovanni
++	23	583: her	---	22	552: zerlina
++	23	585: leporello	---	21	537: leporello
++	23	591: him	---	23	580: he
++	23	595: his	---	23	580: he
++	23	602: him	---	23	580: he
++	23	609: which	---	23	606: confusion
++	23	610: he	---	23	580: he
++	25	624: elvira	---	20	516: elvira
++	25	628: giovanni	---	23	572: giovanni
++	25	629: eye	---	11	249: eye
++	25	635: leporello	---	23	585: leporello
++	25	640: his	---	25	635: leporello
++	25	643: giovanni	---	23	572: giovanni
+ -	25	645: him	---	25	643: giovanni
++	25	650: him	---	25	643: giovanni
+ -	25	654: his	---	25	640: his
++	26	665: elvira	---	25	624: elvira
++	26	668: she	---	26	665: elvira
++	26	670: her	---	26	665: elvira
+ -	26	673: he	---	25	643: giovanni
++	26	675: her	---	26	668: she
++	26	682: she	---	26	668: she
++	26	689: master	---	10	218: master
++	26	695: leporello	---	25	635: leporello
++	26	702: giovanni	---	25	643: giovanni
+ -	26	705: his	---	26	673: he
++	27	723: masetto	---	19	498: masetto
++	27	728: his	---	27	723: masetto
++	27	740: who	---	27	739: man
++	27	745: zerlina	---	22	552: zerlina
++	28	749: giovanni	---	26	702: giovanni
++	28	751: whom	---	28	749: giovanni
+ -	28	752: they	---	27	734: stick
++	28	755: leporello	---	26	695: leporello
++	28	758: them	---	28	752: they
++	28	770: them	---	28	752: they
++	28	778: him	---	28	755: leporello
++	29	783: they	---	28	752: they
++	29	787: he	---	28	749: giovanni
++	29	789: masetto	---	27	723: masetto
++	29	792: him	---	29	789: masetto
++	30	798: zerlina	---	27	745: zerlina
++	30	804: her	---	30	798: zerlina
++	30	808: him	---	29	789: masetto
++	30	810: his	---	30	808: him
++	30	814: she	---	30	798: zerlina
++	30	816: him	---	30	808: him
++	31	820: anna	---	20	518: anna
++	31	822: ottavio	---	20	520: ottavio
++	31	825: masetto	---	29	789: masetto
++	31	827: zerlina	---	30	798: zerlina
++	31	834: giovanni	---	28	749: giovanni


```

++ 31 840: leporello ---> 28 764: clothes leporello
++ 31 845: elvira ---> 26 665: elvira
+- 31 850: him ---> 30 816: him
++ 32 856: elvira ---> 31 845: elvira
+- 32 858: them ---> 30 811: bruise
++ 32 861: her ---> 32 856: elvira
++ 32 864: leporello ---> 31 840: leporello
++ 32 866: his ---> 32 864: leporello
++ 33 882: ottavio ---> 31 822: ottavio
++ 33 885: he ---> 33 882: ottavio
++ 33 890: giovanni ---> 31 834: giovanni
++ 34 903: elvira ---> 32 856: elvira
++ 34 906: she ---> 34 903: elvira
++ 34 910: giovanni ---> 33 890: giovanni
++ 34 914: his ---> 34 910: giovanni
++ 34 919: he ---> 34 914: his
++ 34 922: her ---> 34 906: she
++ 35 931: giovanni ---> 34 910: giovanni
++ 35 933: leporello ---> 32 864: leporello
++ 35 937: giovanni ---> 34 910: giovanni
++ 35 941: his ---> 35 937: giovanni
++ 35 955: commendatore ---> 3 49: commendatore
++ 35 959: him ---> 35 941: his
+- 35 965: his ---> 35 955: commendatore
++ 36 970: giovanni ---> 35 937: giovanni
++ 36 978: leporello ---> 35 933: leporello
+- 36 981: it ---> 36 972: reply
++ 36 988: leporello ---> 35 933: leporello
++ 36 992: statue ---> 35 952: statue
++ 36 996: its ---> 36 992: statue
++ 37 1004: anna ---> 31 820: anna
++ 37 1011: her ---> 37 1004: anna
+- 37 1015: they ---> 36 997: head
++ 38 1023: she ---> 37 1004: anna
++ 38 1026: he ---> 37 1006: anna's house ottavio
++ 38 1028: her ---> 38 1023: she
++ 38 1032: she ---> 38 1023: she
++ 38 1035: she ---> 38 1032: she
++ 38 1038: him ---> 38 1026: he
++ 38 1045: her ---> 38 1035: she
++ 38 1051: him ---> 38 1038: him
++ 39 1063: giovanni ---> 36 970: giovanni
++ 40 1068: he ---> 39 1063: giovanni
+- 40 1071: supper ---> 36 983: supper
++ 41 1086: elvira ---> 34 903: elvira
++ 41 1094: him ---> 40 1068: he
++ 41 1097: his ---> 41 1094: him
++ 41 1101: he ---> 40 1068: he
++ 41 1104: her ---> 41 1086: elvira
++ 41 1107: she ---> 41 1086: elvira
++ 42 1115: she ---> 41 1107: she
++ 42 1118: she ---> 42 1115: she
++ 42 1125: leporello ---> 36 988: leporello
++ 42 1138: commendatore ---> 35 955: commendatore
++ 42 1139: statue ---> 36 992: statue
++ 43 1155: statue ---> 42 1139: statue
++ 43 1159: giovanni ---> 39 1063: giovanni
++ 43 1162: supper ---> 40 1071: supper
++ 44 1172: it ---> 43 1155: statue

```

++	44	1176: giovanni	--->	43	1159: giovanni
++	44	1180: him	--->	44	1172: it
++	45	1184: giovanni	--->	44	1176: giovanni
++	45	1194: his	--->	45	1184: giovanni
++	45	1199: he	--->	45	1184: giovanni
++	45	1201: it	--->	45	1195: hand
??	45	1210: fear	--->	34	912: fear
++	45	1213: him	--->	45	1199: he
++	46	1221: he	--->	45	1199: he
++	46	1224: statue	--->	43	1155: statue
?+	46	1226: that	--->	46	1225: urging
++	46	1227: he	--->	46	1221: he
++	46	1235: leporello	--->	42	1125: leporello
++	46	1244: his	--->	46	1227: he
++	46	1248: giovanni	--->	45	1184: giovanni
++	47	1269: giovanni	--->	46	1248: giovanni
++	47	1271: leporello	--->	46	1235: leporello
++	47	1273: them	--->	47	1261: character
++	48	1284: they	--->	47	1261: character
++	48	1295: their	--->	48	1290: sinner

B.3 Nichtpronominale Substitute

ROSANA schlägt folgende nichtpronominale Substitute für die pronominalen Okkurrenzen vor, die für den Text "*Don Giovanni*" ermittelt werden: (Auch hier sind die angegebenen Ergebnisse bereits gemäß Abbildung 11.7, S. 241 bewertet.)

LEXIKALISCHE SUBSTITUTION

??	2	25: which	==	2	19: grumble
++	2	26: his	==	2	14: leporello
+--	2	33: his	==	2	14: leporello
+--	3	40: his	==	2	14: leporello
++	3	52: she	==	3	44: donna anna
++	3	61: her	==	3	44: donna anna
++	4	76: her	==	3	44: donna anna
+--	4	81: him	==	3	64: fiancee don ottavio
++	4	85: which	==	4	84: duel
+--	4	86: he	==	3	64: fiancee don ottavio
+--	5	92: they	==	4	84: duel
++	6	116: her	==	6	108: anna
++	7	136: their	==	7	133: and
++	7	143: they	==	7	133: and
+--	8	154: his	==	7	134: leporello
+--	8	161: her	==	8	155: opportunity
?+	8	167: it	==	8	155: opportunity
++	8	171: whom	==	8	170: donna elvira
++	8	172: he	==	8	157: giovanni
++	9	181: she	==	8	170: donna elvira
++	9	184: him	==	8	157: giovanni
++	9	188: he	==	8	157: giovanni
++	9	194: her	==	8	170: donna elvira

++	9	197: he	==	8	157: giovanni
++	9	199: his	==	8	157: giovanni
++	10	212: her	==	8	170: donna elvira
++	10	217: his	==	10	204: leporello
++	11	246: whom	==	11	243: zerlina
++	12	255: he	==	11	247: giovanni
++	12	259: his	==	11	247: giovanni
++	12	274: who	==	12	272: masetto
+-	12	280: his	==	13	294: giovanni
++	13	297: his	==	13	294: giovanni
++	13	302: who	==	13	300: zerlina
++	13	308: his	==	13	294: giovanni
+-	13	319: her	==	13	311: elvira
+-	13	323: her	==	13	311: elvira
++	14	343: her	==	14	329: donna anna
++	14	347: his	==	14	337: giovanni
++	14	351: her	==	14	329: donna anna
++	15	358: he	==	14	337: giovanni
++	15	364: he	==	14	337: giovanni
++	15	371: him	==	14	337: giovanni
++	16	377: he	==	16	409: giovanni
+-	16	390: their	==	13	309: advance
++	16	403: she	==	16	398: anna
++	16	406: she	==	16	398: anna
++	16	414: who	==	16	409: giovanni
+-	16	416: her	==	16	409: giovanni
+-	16	423: his	==	17	436: giovanni
++	17	450: his	==	17	436: giovanni
++	17	453: where	==	17	451: home
++	17	454: he	==	17	436: giovanni
++	18	477: them	==	18	468: villager
++	18	483: her	==	18	478: zerlina
++	19	491: his	==	19	489: giovanni
+-	19	501: he	==	19	498: masetto
+-	19	504: them	==	19	492: attention
++	22	559: her	==	22	552: zerlina
++	22	567: her	==	22	552: zerlina
??	23	576: it	==	22	568: aid
++	23	580: he	==	23	572: giovanni
++	23	583: her	==	22	552: zerlina
++	23	591: him	==	23	572: giovanni
++	23	595: his	==	23	572: giovanni
++	23	602: him	==	25	628: giovanni
++	23	609: which	==	23	606: confusion
++	23	610: he	==	25	628: giovanni
++	25	640: his	==	25	635: leporello
+-	25	645: him	==	25	643: giovanni
++	25	650: him	==	25	643: giovanni
+-	25	654: his	==	25	635: leporello
++	26	668: she	==	26	665: elvira
++	26	670: her	==	26	665: elvira
+-	26	673: he	==	25	643: giovanni
++	26	675: her	==	26	665: elvira
++	26	682: she	==	26	665: elvira
++	26	705: his	==	26	702: giovanni
++	27	728: his	==	27	723: masetto
++	27	740: who	==	27	739: man
++	28	751: whom	==	28	749: giovanni
+-	28	752: they	==	27	734: stick

+-	28	758: them	==	27	734: stick
+-	28	770: them	==	27	734: stick
++	28	778: him	==	28	755: leporello
+-	29	783: they	==	27	734: stick
++	29	787: he	==	28	749: giovanni
++	29	792: him	==	29	789: masetto
++	30	804: her	==	30	798: zerlina
++	30	808: him	==	29	789: masetto
++	30	810: his	==	31	825: masetto
++	30	814: she	==	30	798: zerlina
++	30	816: him	==	31	825: masetto
+-	31	850: him	==	31	825: masetto
+-	32	858: them	==	30	811: bruise
++	32	861: her	==	32	856: elvira
++	32	866: his	==	32	864: leporello
++	33	885: he	==	33	882: ottavio
++	34	906: she	==	34	903: elvira
++	34	914: his	==	34	910: giovanni
++	34	919: he	==	34	910: giovanni
++	34	922: her	==	34	903: elvira
+_	35	930: which	==	?????	
++	35	941: his	==	35	937: giovanni
++	35	959: him	==	36	970: giovanni
+-	35	965: his	==	35	955: commendatore
+-	36	981: it	==	36	972: reply
++	36	996: its	==	36	992: statue
++	37	1011: her	==	37	1004: anna
+-	37	1015: they	==	36	997: head
++	38	1023: she	==	37	1004: anna
++	38	1026: he	==	37	1006: anna's house ottavio
++	38	1028: her	==	37	1004: anna
++	38	1032: she	==	37	1004: anna
++	38	1035: she	==	37	1004: anna
++	38	1038: him	==	37	1006: anna's house ottavio
++	38	1045: her	==	37	1004: anna
++	38	1051: him	==	37	1006: anna's house ottavio
++	40	1068: he	==	39	1063: giovanni
++	41	1094: him	==	39	1063: giovanni
++	41	1097: his	==	39	1063: giovanni
++	41	1101: he	==	39	1063: giovanni
++	41	1104: her	==	41	1086: elvira
++	41	1107: she	==	41	1086: elvira
++	42	1115: she	==	41	1086: elvira
++	42	1118: she	==	41	1086: elvira
++	44	1172: it	==	43	1155: statue
++	44	1180: him	==	43	1155: statue
++	45	1194: his	==	45	1184: giovanni
++	45	1199: he	==	45	1184: giovanni
++	45	1201: it	==	45	1195: hand
++	45	1213: him	==	45	1184: giovanni
++	46	1221: he	==	46	1248: giovanni
?+	46	1226: that	==	46	1225: urging
++	46	1227: he	==	46	1248: giovanni
++	46	1244: his	==	46	1248: giovanni
++	47	1273: them	==	47	1261: character
++	48	1284: they	==	47	1261: character
++	48	1295: their	==	48	1290: sinner

B.4 Bewertung der Interpretationsleistung

Abbildung B.1 zeigt die Evaluationsergebnisse von ROSANA in den Disziplinen KV (Ermittlung der Kospezifikationsklassen) und PS (Ermittlung nichtpronominaler Substitute) auf dem Text *“Don Giovanni”*. Zusätzlich werden die Precision-Werte der Antezedensentscheidungen für die einzelnen Anapherarten aufgeschlüsselt (Tabelle ANKNUEPFUNGS-ENTSCHEIDUNGEN).

A. R. -ERGEBNIS-PARTITIONIERUNG

- SCHNITTE: 22
- MOEGLICH: 240
=> PRECISION: 0.9083

KEY-PARTITIONIERUNG

- SCHNITTE: 42
- MOEGLICH: 259
=> RECALL: 0.8378

ANKNUEPFUNGS-ENTSCHEIDUNGEN

		PRECIS	++	+-	+?	+ ₋	+*	?+	? ₋
PRON	PE-3	0.8462	66	12	0	0	0	2	0
	PE12	--	0	0	0	0	0	0	0
	P0-3	0.7826	36	10	0	0	0	0	0
	P012	--	0	0	0	0	0	0	0
	REFL	--	0	0	0	0	0	0	0
	RELA	0.9091	10	0	1	1	0	1	0
		0.8296	112	22	1	1	0	3	0
NOMN	VNOM	0.7619	16	4	1	135	2	1	9
	NAME	1.0000	85	0	0	13	0	0	0
		0.9528	101	4	1	148	2	1	9

DURCHSCHNITT JE ANAPHER

=> PRECISION: 0.8838

LEXIKALISCHE SUBSTITUTION

	PRECIS	RECALL	++	+-	+?	+ ₋	+*	?+	? ₋
PE-3	0.7821	0.7821	61	17	0	0	0	2	0
PE12	--	--	0	0	0	0	0	0	0
P0-3	0.7826	0.7826	36	10	0	0	0	0	0
P012	--	--	0	0	0	0	0	0	0
REFL	--	--	0	0	0	0	0	0	0
RELA	0.9091	0.8333	10	0	1	1	0	1	0

DURCHSCHNITT JE PRONOMEN

=> PRECISION: 0.7926
=> RECALL: 0.7868

Abbildung B.1: Evaluationsergebnisse für den Text *“Don Giovanni”*

Literaturverzeichnis

- [Alsh87] Hiyan Alshawi. *Memory and Context for Language Interpretation*. Cambridge University Press, 1987.
- [AsWa88] Nicholas Asher, Hajime Wada. *A Computational Account of Syntactic, Semantic and Discourse Principles for Anaphor Resolution*. *Journal of Semantics* 6, 1988, 309-344.
- [BaGo82] Friedrich L. Bauer, Gerhard Goos. *Informatik. Eine einführende Übersicht. Erster Teil*. Springer Verlag, Berlin Heidelberg New York, 1982.
- [BaLa55] Allen H. Barton, Paul F. Lazarsfeld. *The Concept of Property-Space in Social Research*. In: Paul F. Lazarsfeld, Morris Rosenberg (eds.). *The Language of Social Research*, Glencoe, 1955, 40ff.
- [Bere52] Bernard Berelson. *Content Analysis in Communication Research*. Free Press, New York, 1952.
- [BFP87] Susan E. Brennan, Marilyn W. Friedman und Carl J. Pollard. *A Centering Approach to Pronouns*. In: *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*. Stanford, Juli 1987, 155-162.
- [BoDö95] Jürgen Bortz, Nicola Döring. *Forschungsmethoden und Evaluation*. 2. Auflage, Springer-Verlag, Berlin/Heidelberg, 1995.
- [Bött96] Martin Böttcher. *Unifikation disjunktiver Attributterme. Auswertalgorithmen und Anwendungen in einer Valenzgrammatik zur Textanalyse*. Reihe "Dissertationen zur Künstlichen Intelligenz" (DISKI), Nr. 118, E. Hundt, 1996.
- [BoTa89] Wilfrid Bos, Christian Tarnai (Hrsg.) *Angewandte Inhaltsanalyse in Empirischer Pädagogik und Psychologie*. Waxmann, Münster/New York, 1989.
- [BrBe83] Michael Brady, Robert C. Berwick (eds). *Computational Models of Discourse*. M.I.T. Press, Cambridge, MA, 1983.
- [Buß90] Hadumod Bußmann. *Lexikon der Sprachwissenschaft*. Alfred Kröner Verlag, Stuttgart, 1990.
- [CaBr88] Jaime G. Carbonell, Ralf D. Brown. *Anaphora Resolution: A Multi-Strategy Approach*. *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988, 96-101.

- [Cart87] David Carter. *Interpreting Anaphors in Natural Language Texts*. Ellis Horwood Series in Artificial Intelligence, Chichester, 1987.
- [Chom81] Noam Chomsky. *Lectures on Government and Binding*. Foris Publications, Dordrecht, 1981.
- [Chom82] Noam Chomsky. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass., 1982.
- [Chom86] Noam Chomsky. *Knowledge of Language*. Praeger, New York, 1986.
- [CMU+95] Roland A. Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, Victor Zue (eds.). *Survey of the State of the Art in Human Language Technology*. Online-Dokument: <http://www.cse.ogi.edu/CSLU/HLTsurvey/HLTsurvey.html> (Internetseite des Center for Spoken Language Understanding), 21. November 1995.
- [Cook88] Vivian J. Cook. *Chomsky's Universal Grammar*. Basil Blackwell, Oxford, 1988.
- [Corb79] G. G. Corbett. *The Agreement Hierarchy*. In: *Journal of Linguistics*, Vol. 15, 1979, 203-224.
- [Corr88] Nelson Correa. *A Binding Rule for Government-binding Parsing*. In: *Proceedings of the 12th International Conference on Computational Linguistics, COLING*, Vol. I, Budapest, 1988, 123-129.
- [CoTD97] *MUC-7 Coreference Task Definition, Version 3.0, 13.7.97*. Bezogen über: Nancy Chinchor, SAIC (chinchor@gso.saic.com). Auch in: *Proceedings of the Seventh Message Understanding Conference (MUC-7)*, 1998.
- [DaIt90] Ido Dagan, Alon Itai 1990. *Automatic Processing of Large Corpora for the Resolution of Anaphoric References*. In: *Proceedings of the 13th International Conference on Computational Linguistics, COLING*, Vol. 3, Helsinki, 1990, 330-332.
- [Dane70] František Daneš. *Zur linguistischen Analyse der Textstruktur*. In: *Folia Linguistica* 4, 1970.
- [DAvi94] Winfried D'Avis. *Können Computer denken? Eine bedeutungs- und zeittheoretische Analyse von KI-Maschinen*. Campus Verlag, Frankfurt/New York, 1994.
- [DeSt75] Alexander Deichsel, Philip Stone. *Newspaper headlines: A multinational content analysis project on textual indicators from mass media*. In: *Social Science Information*, 14/1/75, 112-116.
- [Dürr58] Friedrich Dürrenmatt. *Das Versprechen*. Verlag der Arche, Zürich, 1958.
- [Evan80] G. Evans. *Pronouns*. In: *Linguistic Inquiry*, Vol. 11, 1980.
- [FiLe91] Nigel G. Fielding, Raymond M. Lee (eds.). *Using Computers in Qualitative Research*. SAGE Publications, 1991.
- [Fill68] Charles J. Fillmore. *The Case for Case*. In: E. Bach, R.T. Harms (eds.). *Universals in Linguistic Theory*. New York, 1968, 1-88.
- [Freg94] Gottlob Frege. *Funktion, Begriff, Bedeutung*. Vandenhoeck & Ruprecht, Göttingen, 1994.

- [Früh91] Werner Früh. *Inhaltsanalyse. Theorie und Praxis*. 3., überarbeitete Auflage, Verlag Ölschläger, München, September 1991.
- [Fühl82] Ingunde Fühlau. *Die Sprachlosigkeit der Inhaltsanalyse. Linguistische Bemerkungen zu einer sozialwissenschaftlichen Methode*. Gunter Narr Verlag, Tübingen, 1982.
- [Gada60] Hans-Georg Gadamer. *Wahrheit und Methode*. Tübingen, 1960.
- [GaHu96] Robert Gaizauskas, Kevin Humphreys. *Quantitative Evaluation of Coreference Algorithms in an Information Extraction System*. Department of Computer Science, University of Sheffield. Manuskript zu Vortrag gleichen Titels auf: *DAARC96 - Discourse Anaphora and Anaphor Resolution Colloquium*, Lancaster University, Juli 1996.
- [Geor59] Alexander L. George. *Quantitative and Qualitative Approaches to Content Analysis*. In: Ithiel de Sola Pool (ed). *Trends in Content Analysis*. Urbana, 1959, 7-32.
- [GGG93] Peter C. Gordon, Barbara J. Grosz, Laura A. Gilliom. *Pronouns, Names, and the Centering of Attention in Discourse*. In: *Cognitive Science*. Vol. 17(3), 1993, 311-348.
- [GJW83] Barbara J. Grosz, Aravind K. Joshi, Scott Weinstein. *Providing a unified Account of definite Noun Phrase in Discourse*. In: *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*. Cambridge, Mass., Juni 1983, 44-50.
- [GJW95] Barbara J. Grosz, Aravind K. Joshi, Scott Weinstein. *Centering: A Framework for Modeling the Local Coherence of Discourse*. In: *Computational Linguistics*. Vol. 21, Nr. 2, Juni 1995. 203-225.
- [GPS90] Alessandra Giorgi, Fabio Pianesi, Giorgio Satta. *A Computational Approach to Binding Theory*. In: *Proceedings of the 13th International Conference on Computational Linguistics, COLING*, Vol. 3, Helsinki, 1990, 120-125.
- [Grew88] Günter Grewendorf. *Aspekte der deutschen Syntax. Eine Rektions-Bindungs-Analyse*. Gunter Narr Verlag Tübingen, 1988.
- [Gric68] H. Paul Grice. *Logic and Conversation*. In: P. Cole, J. L. Morgan (eds.). *Syntax and Semantics. Vol. 3, Speech Acts*. New York, 1975, 41-58.
- [Gros77] Barbara J. Grosz. *The Representation and Use of Focus in a System for Dialogue Understanding*. Technical Report 151, Artificial Intelligence Center, SRI International, Menlo Park, CA, July 1977.
- [GrPo71] John T. Grinder und Paul Martin Postal. *Missing Antecedents*. In: *Linguistic Inquiry*, Vol. 2(3), Summer 1971, 269-312.
- [GrSi86] Barbara J. Grosz, Candace L. Sidner. *Attention, Intentions, and the Structure of Discourse*. In: *Computational Linguistics*, Vol. 12, Number 3, July-September 1986, 175-204.
- [GrSu96a] Ralph Grishman und Beth Sundheim. *Message Understanding Conference - 6: A Brief History*. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Vol. I, August 1996, Kopenhagen, 466-471.

- [GrSu96b] Ralph Grishman und Beth Sundheim. *Design of the MUC-6 Evaluation*. In: *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann, San Francisco, 1996, 1-11.
- [Habe86] Christopher Habel. *Prinzipien der Referentialität. Untersuchungen zur propositionalen Repräsentation von Wissen*. Springer-Verlag, 1986.
- [Haeg91] Liliane Haegeman. *Introduction to Government and Binding Theory*. Basil Blackwell, Oxford, 1991.
- [Haen95] Karin Haenelt. *Das KONTEXT-Modell. Verarbeitung natürlichsprachiger Texte auf der Basis eines Textmodells*. Habilitationsschrift, Universität Heidelberg, 1995.
- [HaHa76] M.A.K. Halliday, Ruqaiya Hasan. *Cohesion in English*. Longman, 1976.
- [HaMa94] Mariikka Haapalainen, Ari Majorin. *GERTWOL: Ein System zur automatischen Wortformerkennung deutscher Wörter*. Lingsoft, Inc., Helsinki, Finnland, September 1994.
- [Haus94] R. Hausser. *The Coordinators final Report on the first Morpholympics*. LDV-Forum 11(1), 1994, 54-64.
- [Heid79] Martin Heidegger. *Sein und Zeit*. Tübingen, 1979.
- [Heim82] Irene Heim. *The Semantics of definite and indefinite NPs*. Ph.D. dissertation, University of Amherst, 1982.
- [Hein87] Th. Heinze. *Qualitative Sozialforschung: Erfahrungen, Probleme und Perspektiven*. Westdeutscher Verlag, Opladen, 1987.
- [Hell84] Peter Hellwig. *Grundzüge einer Theorie des Textzusammenhangs*. In: Anneli Rothkegel und Barbara Sandig (Hrsg.). *Text - Textsorten - Semantik Linguistische Modelle und maschinelle Verfahren*. Helmut Buske Verlag Hamburg, 1984.
- [Hirs81] Graeme Hirst. *Anaphora in Natural Language Understanding: A Survey*. Lecture Notes in Computer Science 119, Springer-Verlag, 1981.
- [HKRS81] Jürgen Höhe, Hans-Dieter Klingemann, Klaus Rademacher, C. Schickle. *TEXT-PACK Version IV*. Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, 1982.
- [Hobb78] Jerry R. Hobbs. *Resolving Pronoun References*. In: *Lingua*, Vol. 44, 1978, 311-338.
- [Hobb79] Jerry R. Hobbs. *Coherence and Coreference*. In: *Cognitive Science*, Vol. 3(1), 1979, 67-82.
- [Hols69] Ole R. Holsti. *Content Analysis for the Social Sciences and Humanities*. Reading/Mass. 1969.
- [HoUl79] John E. Hopcroft, Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.
- [Hube89] Günter L. Huber. *Qualität versus Quantität in der Inhaltsanalyse*. In: [BoTa89], 32-47.

- [Hube92a] Günter L. Huber (Hrsg.). *Qualitative Analyse. Computereinsatz in der Sozialforschung*. R. Oldenbourg Verlag, München, Wien, 1992.
- [Hube92b] Günter L. Huber. *Qualitative Analyse mit Computerunterstützung*. In: [Hube92a], 115-175.
- [InSt89] Robert J. P. Ingria, David Stallard. *A Computational Mechanism for Pronominal Reference*. In: *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*. Vancouver, Juni 1989.
- [JäTa97] Timo Järvinen, Pasi Tapanainen. *A Dependency Grammar for English*. Technical Reports, No. TR-1, Department of General Linguistics, University of Helsinki, March 1997.
- [Kamp81] Hans Kamp. *A Theory of Truth and Semantic Representation*. In: Groenendijk, Janssen, Stockhof (eds). *Formal Methods in the Study of Language*. Mathematisch Centrum, Amsterdam, 1981, 1-41.
- [KaRe93] Hans Kamp, Uwe Reyle. *From Discourse to Logic. Introduction to the Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers, Dordrecht, 1993.
- [Kart76] Lauri Karttunen. *Discourse Referents*. In: *Syntax and Semantics*, Vol. 7, J. McCawley (ed), Academic Press, New York, 1976, 363-385.
- [KeBo96] Christopher Kennedy, Branimir Boguraev. *Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser*. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Vol. I, August 1996, Kopenhagen, 113-118.
- [Kell95] Udo Kelle (ed.). *Computer-Aided Qualitative Data Analysis. Theory, Methods, and Practice*. SAGE Publications, London, 1995.
- [Kell96] Udo Kelle. *Computer-Aided Qualitative Data Analysis. An Overview*. In: [ZHH96], 33-64.
- [Klei95] Harald Klein. *Computerunterstützte Inhaltsanalyse mit INTEXT*. Münster, Hamburg, 1995.
- [Klei97] Harald Klein. *INTEXT - Text Analyse Software*. Kurzdokumentation zur Version INTEXT 4.0. Online-Dokument: <http://www.soziologie.uni-jena.de/home/klein/intextd.html>, 4.6.1997.
- [Klin84] Hans-Dieter Klingemann (Hrsg.). *Computerunterstützte Inhaltsanalyse in der empirischen Sozialforschung*. Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim. Monographien Sozialwissenschaftliche Methoden. Campus Verlag, Frankfurt/New York, 1984.
- [Kosk83] Kimmo Koskenniemi. *Two-Level Morphology: a General Computational Model for Word-Form Recognition and Production*. University of Helsinki, Department of General Linguistics, Publication No. 11, 1983.

- [Krac52] Siegfried Kracauer. *The Challenge of Qualitative Content Analysis*. In: *Public Opinion Quarterly*, 16, 1952, 631-642.
- [Krip80] Klaus Krippendorff. *Content Analysis. An Introduction to its Methodology*. SAGE Publications, Beverly Hills/London, 1980.
- [Kuno87] Susumu Kuno. *Anaphora and Discourse Principles*. In: Makoto Nagao (ed). *Language and Artificial Intelligence*. Elsevier, 1987.
- [Kunz91] Jürgen Kunze. *Kasusrelationen und Semantische Emphase*. studia grammatica XX-XII, Akademie Verlag, 1991.
- [KVHA95] Fred Karlsson, Atro Voutilainen, Juha Heikkilä, and Arto Antilla. *Constraint Grammar: A language-independent system for parsing free text*. Mouton de Gruyter, Berlin/New York, 1995.
- [LaCo90a] Shalom Lappin, Michael McCord. *A Syntactic Filter on Pronominal Anaphora for Slot Grammar*. In: *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*. 1990, 135-142.
- [LaCo90b] Shalom Lappin, Michael McCord. *Anaphora Resolution in Slot Grammar*. In: *Computational Linguistics*, 16 (4), December 1990, 197-212.
- [LaLe94] Shalom Lappin and Herbert J. Leass. *An Algorithm for Pronominal Anaphora Resolution*. In: *Computational Linguistics*, 20 (4), 1994, 535-561.
- [Lamp83] Vittorio Magnago Lampugnani (Hrsg.). *Lexikon der Architektur des 20. Jahrhunderts*. Hatje-Verlag, Stuttgart, 1983.
- [Lass27] Harold D. Lasswell. *Propaganda-Technique in the World War*. New York, 1927.
- [Lass65] Harold D. Lasswell. *Language of Politics. Studies in Quantitative Semantics*. 2. Auflage, Cambridge, 1965.
- [Leas91] Herbert Leass. *Anaphora Resolution for Machine Translation: A Study* IWBS-Report 187, Wissenschaftliches Zentrum - Institut für Wissensbasierte Systeme, IBM Heidelberg, Oktober 1991.
- [LiKr78] Ralf Lisch, Jürgen Kriz. *Grundlagen und Modelle der Inhaltsanalyse*. Rowohlt, Reinbek bei Hamburg, 1978.
- [LS91] Herbert Leass, Ulrike Schwall. *An Anaphora Resolution Procedure for Machine Translation*. IWBS-Report 172, Wissenschaftliches Zentrum - Institut für Wissensbasierte Systeme, IBM Heidelberg, Mai 1991.
- [Marc97] Daniel Marcu. *The Rhetorical Parsing of Natural Language Texts*. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*. Madrid, Spain, July 1997, 96-103.
- [MaTo87] William C. Mann, Sandra A. Thompson. *Rhetorical Structure Theory. A Theory of Text Organization*. Marina del Rey, California. University of Southern California, 1987.

- [Mayr96] Philipp Mayring. *Einführung in die qualitative Sozialforschung*. 3. überarbeitete Auflage, Psychologie VerlagsUnion, Weinheim 1996
- [Menz95] Wolfgang Menzel. *Robust Processing of Natural Language*. Fachbereich Informatik, Universität Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Deutschland, 1995.
- [Mert95] Klaus Merten. *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis*. 2., verbesserte Auflage, Westdeutscher Verlag, 1995.
- [Mill64] G. A. Miller. *Computers, Communication, and Cognition*. In: G. A. Miller. *The Psychology of Communication*. Penguin, 1968, 95-123.
- [Mitk97] Ruslan Mitkov. *Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches*. In: *Proceedings of the ACL'97 / EACL'97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, July 1997, Madrid, Spain, 14-21.
- [Mohl89] Peter Ph. Mohler. *Computergestützte Inhaltsanalyse: Überblick über die linguistischen Leistungen*. In: Batori, Lenders (Hrsg). *HSK 4. Handbuch zur Sprach- und Kommunikationswissenschaft*. 1989, 580-585.
- [MoZü95] Peter Ph. Mohler, Cornelia Züll. *TEXTPACK PC Release 5.0, Short Description*. Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, Juli 1995.
- [MUC696] *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann, San Francisco, 1996.
- [OAKK79] Ulrich Oevermann, T. Allert, E. Konau, J. Krambeck. *Die Methodologie einer "objektiven Hermeneutik" und ihre allgemeine forschungslogische Bedeutung in den Sozialwissenschaften*. In: H.-G. Soeffner (Hrsg.). *Interpretative Verfahren in den Sozial- und Textwissenschaften*, Metzler, Stuttgart, 352-434.
- [OgRi74] C. K. Ogden und I. A. Richards. *Die Bedeutung der Bedeutung*. Suhrkamp, Frankfurt, 1974.
- [PoSa92] Carl Pollard, Ivan A. Sag. *Anaphors in English and the Scope of Binding Theory*. In: *Linguistic Inquiry*, Vol. 23, 2, Spring 1992.
- [Prim89] Beatrice Primus. *Parameter der Herrschaft: Reflexivpronomen im Deutschen*. In: *Zeitschrift für Sprachwissenschaft* 8, 1, 1989, 53-88.
- [Quin60] Willard Van Orman Quine. *Word and Object*. Cambridge, Mass., 1960.
- [Ragi87] C. C. Ragin. *The Comparative Method. Moving beyond Qualitative and Quantitative Strategies*. University of California Press, Berkeley, 1987.
- [Ragi95] C. C. Ragin. *Using Qualitative Comparative Analysis to Study Configurations*. In: [Kell95].
- [Rein76] Tanya Reinhart. *The syntactic Domain of Anaphora*. MIT-Diss., Cambridge, Mass., 1976.
- [Rein83] Tanya Reinhart. *Anaphora and Semantic Representation*. Croom Helm, 1983.

- [Rohl77] Lutz Rohland. *Die Technik der elektronischen Inhaltsanalyse mit dem Programmsystem EVA*. Universität Hamburg, Seminar für Sozialwissenschaften, ZVV, 1977.
- [Rust80] Holger Rust. *Qualitative Inhaltsanalyse - begriffslose Willkür oder wissenschaftliche Methode?* In: *Publizistik*, 25:5-23.
- [SBNO62] Philip J. Stone, Robert F. Bales, J. Zvi Namenwirth, Daniel M. Ogilvie. *The General Inquirer: A Computer System for Content Analysis and Retrieval Based on the Sentence as a Unit of Information*. Behavioral Science 7, 1962, 484-498.
- [SDSO66] Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, Daniel M. Ogilvie. *The General Inquirer. A Computer Approach to Content Analysis*. The M.I.T. Press, Cambridge, Mass., 1966.
- [Sidn79] Candace L. Sidner. *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse*. M.I.T. Artificial Intelligence Laboratory, TR-537, 1979.
- [Sidn81] Candace L. Sidner. *Focusing for Interpretation of Pronouns*. American Journal of Computational Linguistics, Vol.7, 1981, 217-231.
- [Sidn83] Candace L. Sidner. *Focusing in the Comprehension of Definite Anaphora*. In: [BrBe83], 267-330.
- [SmCh96] Ronald H. Smith, Graig G. Chambers. *Parallelism Effects on Pronoun Resolution in Discourse Contexts*. In: *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium (DAARC96)*, UCREL Technical Papers Series, Vol. 8, Lancaster University, Juli 1996, 268-280.
- [Smit94] Ronald H. Smith. *Grammatical Determinants of Ambiguous Pronoun Resolution*. In: *Journal of Psycholinguistic Research*, Vol. 23, 1994, 197-229.
- [SP88] Remko Scha, Livia Polanyi. *An Augmented Context Free Grammar for Discourse*. In: *Proceedings of the 12th International Conference on Computational Linguistics, COLING*, Vol. II, Budapest, 1988, 573-577.
- [STa89] Hansjörg Scheerer, Christian Tarnai. *Computerunterstützte Inhaltsanalyse von Verbalbeurteilungen in der Grundschule*. In: [BoTa89], 286-302.
- [StHa95] Michael Strube, Udo Hahn. *ParseTalk about Sentence- and Text-Level Anaphora*. In: *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics, EACL*, 237-244. Dublin, March 28-31, 1995.
- [Ston69] Philip J. Stone. *Improved Quality of Content-Analysis Categories: Computerized Disambiguation Rules for High Frequency English Words*. In: G. Gerbner, O. L. Holsti, K. Krippendorff, W. J. Paisley, P. Stone. *The Analysis of Communication Content*, 1969, 199-221.
- [StSt88] Arnim von Stechow, Wolfgang Sternefeld. *Bausteine syntaktischen Wissens*. Westdeutscher Verlag, 1988.

- [Stu96a] Roland Stuckardt. *An Interdependency-Sensitive Approach to Anaphor Resolution*. In: *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium (DAARC96)*, UCREL Technical Papers Series, Vol. 8, Lancaster University, Juli 1996, 400-413.
- [Stu96b] Roland Stuckardt. *Anaphor Resolution and the Scope of Syntactic Constraints*. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Vol. II, August 1996, Kopenhagen, 932-937.
- [Stu96c] Roland Stuckardt. *Centering und Anaphernresolution*. GMD - Forschungszentrum Informationstechnik GmbH, Gruppe KONTEXT, Darmstadt, Oktober 1996.
- [Stu97a] Roland Stuckardt. *Algorithmische Interpretation anaphorischer Ausdrücke*. Arbeitspapiere der GMD 1045, GMD - Forschungszentrum Informationstechnik GmbH, Sankt Augustin, Januar 1997.
- [Stu97b] Roland Stuckardt. *Resolving Anaphoric References on Deficient Syntactic Descriptions*. In: *Proceedings of the ACL'97 / EACL'97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, July 1997, Madrid, Spain, 30-37.
- [StuH94] Roland Stuckardt, Karin Haenelt. *A Text-Model based Conception of Focus*. GMD - Forschungszentrum Informationstechnik GmbH, Gruppe KONTEXT, Darmstadt, Juli 1994.
- [Sund96] Beth Sundheim. *Overview of Results of the MUC-6 Evaluation*. In: *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann, San Francisco, 1996, 13-31.
- [Te92] Renata Tesch. *Verfahren der computerunterstützten qualitativen Analyse*. In: [Hube92a], 43-69.
- [Tesn59] Lucien Tesnière. *Éléments de syntaxe structurale*. Paris, 1959.
- [Teub79] Wolfgang Teubert. *Valenz des Substantivs. Attributive Ergänzungen und Angaben*. Pädagogischer Verlag Schwann, Düsseldorf, 1979.
- [Tomi85] Masaru Tomita. *Efficient Parsing for Natural Language*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1985.
- [Vil+96] Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, Lynette Hirschman. *A Model-Theoretic Coreference Scoring Scheme*. In: *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann Publishers, 1996.
- [ViPo96] Renata Vieira, Massimo Poesio. *Processing Definite Descriptions in Corpora*. In: Simon Botley, Tony McEnery (Eds.) *Corpus-Based and Computational Approaches to Discourse Anaphora*. UCL Press, London, 1996.
- [Walk89] Marilyn A. Walker. *Evaluating Discourse Processing Algorithms*. In: *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*. Vancouver, Juni 1989, 251-161.

- [Webb88] Bonnie Lynn Webber. *Discourse Deixis: Reference to Discourse Segments*. In: *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics, ACL*, 1988, 113-121.
- [Webb89] Bonnie Lynn Webber. *Deictic Reference and Discourse Structure*. Department of Computer and Information Science, University of Pennsylvania, Philadelphia PA 19104-6389, September 11, 1989.
- [Weiz65] Joseph Weizenbaum. *ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine*. In: *Communications of the Association for Computing Machinery*, 9 (1965), 36-45.
- [Weiz76] Joseph Weizenbaum. *Computer Power and Human Reason. From Judgement to Calculation*. Freeman, 1976.
- [WiFl89] Terry Winograd, Fernando Flores. *Erkenntnis Maschinen Verstehen*. Rotbuch Verlag Berlin, 1989.
- [Wilk75a] Yorick Wilks. *An Intelligent Analyzer and Understander of English*. Communications of the ACM, Vol.18, Number 5, May 1975, 264-274.
- [Wilk75b] Yorick Wilks. *A Preferential, Pattern-Seeking, Semantics for Natural Language Inference*. AI 6 (1975), 53-74.
- [Wilk75c] Yorick Wilks. *Preference Semantics*. In: E. Keenan (ed.). *The Formal Semantics of Natural Language*. Cambridge University Press, 1975, 329-348.
- [Wino72] Terry Winograd. *Understanding Natural Language*. Academic Press, New York, 1972.
- [ZHH96] Cornelia Züll, Janett Harkness, Jürgen H. P. Hoffmeyer-Zlotnik (eds.). *Text Analysis and Computers*. (Reihe ZUMA-Nachrichten Spezial). Zentrum für Umfragen, Methoden und Analysen (ZUMA), Mannheim, Mai 1996.

