

3. Diskurs-Analyse

Das Gebiet der Diskurs-Analyse umfasst viele Teilbereiche:

vgl. <http://bank.rug.ac.be/da/da.htm>

a. Analytical philosophy

Speech act theory

Principles of information exchange

b.Linguistics

Structuralist linguistics

Register studies and stylistics

Text Linguistics

Pragmatics

Presuppositions

Face and politeness

Reference

c. Linguistic Anthropology

Ethnography of speaking

Ethnopoetics

Indexicality

Interactional Sociolinguistics

Natural Histories of Discourse

d. New Literacy Studies

e. Post-structuralist theory

M.M. Bakhtin

f. Semiotics and cultural studies

Semiotics and communication studies

Cultural studies

g. Social Theory

Pierre Bourdieu

Michel Foucault

Jürgen Habermas

h. The sociology of order in interaction

Erving Goffman

Interaction order

Frame analysis

Footing

Face

Conversation analysis

Ethnomethodology

Michael Stubbs: Discourse Analysis - The Sociolinguistics Analysis of Natural Language. Oxford: Blackwell, 1983.

Hier Beschränkung auf Linguistik

Hintergrund-Theorie: Pragmatik

Anwendungsgebiet: Computerlinguistik

Ziel: Entwicklung von maschinellen Verfahren zur Annotation von Diskursen (Texten) und zur Auswertung der annotierten Texte.

potentielle Anwendungen:

- Informations-Extraktion (*information extraction*)
- inhaltliche Zusammenfassungen (*summarization*)
- Verbesserung der Generierung (*generation systems*)
- Schreibkontrollsysteme

Älterer Ansatz zur Beschreibung von Diskurs- (Text-) Strukturen:

W. Lenders: Semantische und Argumentative Textdeskription. Hamburg: Buske, 1975.

Hier Vorschlag einer Argumentativen Textanalyse.

Ausgang: Annahme von Superstrukturen

In Philosophischen Texten: Sprechakttypus der Argumentation

Ausgangspunkte:

- Argumentative Texte weisen eine **argumentative Struktur** auf.
- Argumentative Texte enthalten **Konstituenten**, die den Fortgang der Argumentation tragen.
- Argumentative Konstituenten haben eine **argumentative Rolle/Funktion**.
- Argumentative Konstituenten werden durch natürlichsprachliche **argumentative Operatoren** verknüpft.
- Die Art der Verknüpfung (des argumentativen Operators) bestimmt die argumentative Rolle.
- Argumentative Konstituenten bilden **argumentative Muster**.

Klassen argumentativer Operatoren:

Aus der grammatischen Klasse der Konjunktionen

- in kopulativer Funktion: ja, noch, sowohl, überdem, und, weder
- in disjunktiver Funktion: entweder, oder
- in adversativer Funktion: aber, allein, dagegen, dennoch, doch, gleichwohl, indessen, sondern
- in temporaler Funktion: als, nun
- in modaler Funktion: indem, wie

- in kausaler Funktion

in rein kausaler Funktion: da, denn, nämlich, weil

in konsekutiver Funktion: also, darum, demnach, folglich, mithin

in konditionaler Funktion: sonst, wenn

in konzessiver Funktion: zwar

Phrasen als argumentative Operatoren:

- in konsekutiver Funktion:

daraus folgt

es ergibt sich

hieraus läßt sich schließen

- Äquivalenzbeziehungen anzeigend:

das ist

welches eben so viel heißt als

Argumentative Muster

Beispieltext:

Das: Ich denke
muss alle meine Vorstellungen begleiten können;
denn sonst würde etwas (α) in mir vorgestellt werden,
was gar nicht gedacht werden könnte
welches eben so viel heißt, als
die Vorstellung würde entweder unmöglich, oder wenigstens für mich nichts
sein.

Diejenige Vorstellung (β),
die vor allem Denken gegeben sein kann,
heißt Anschauung.

...

Diese Vorstellung ...

Analyseschritte:

morphologische Analyse: Wortklassen, Flexionsformen

syntaktische Analyse: Ermittlung der syntaktischen Konstituenten

Auflösung von Relativsätzen

etwas, was gar nicht gedacht werden könnte (α)

diejenige Vorstellung, die vor allem Denken gegeben sein kann (β)

Auflösung von Referenzidentitäten

diese Vorstellung \longrightarrow die Vorstellung

semantische Analyse: 1. relationale semantische Tiefenstruktur der Sätze

begleiten (x,y) \longrightarrow R1 (x,y)

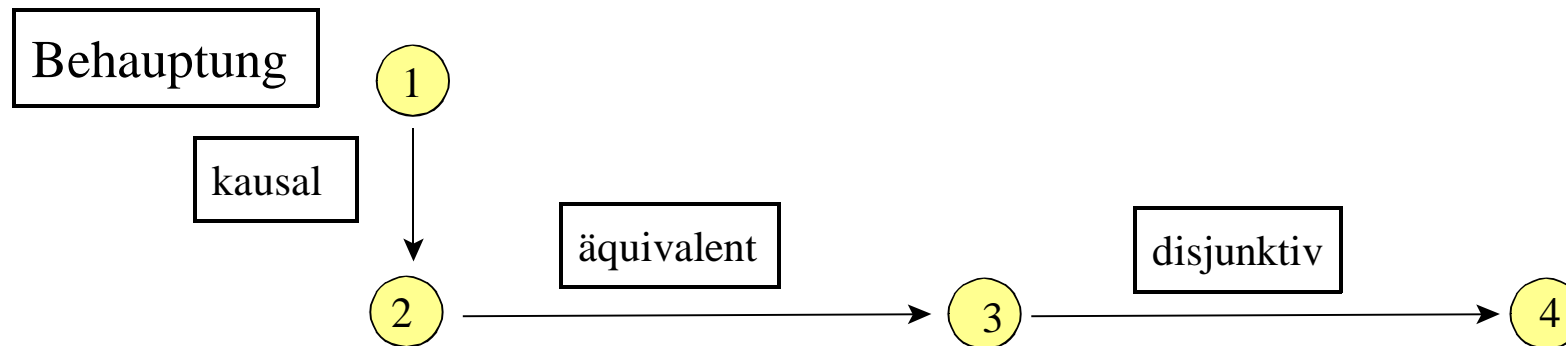
vorstellen (x,y) \longrightarrow R2 (x,y)

nennen (x,y) \longrightarrow Rdef (x,y,z)

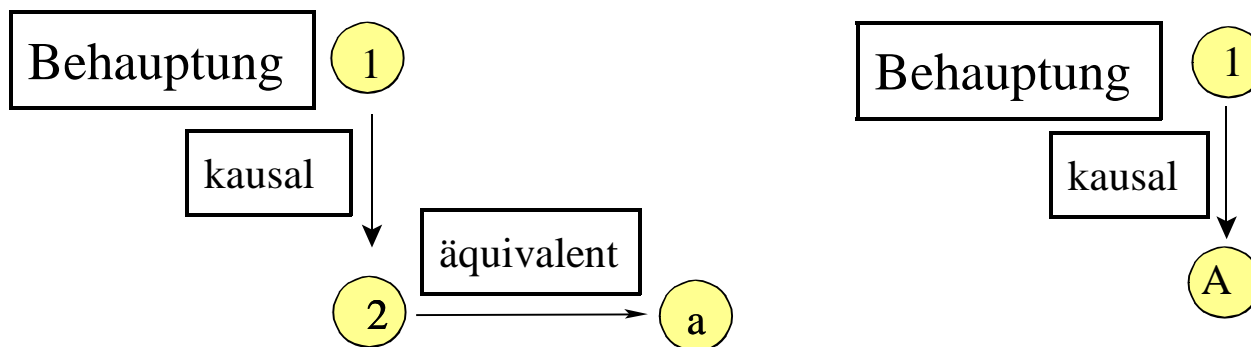
2. Identifikation der präsupponierten Entitäten

(Ich, das: Ich denke; meine Vorstellungen; ...)

Ermittlung **argumentativer Konstituenten** anhand argumentativer Operatoren:



Darstellung **Argumentativer Muster** durch sukzessive Reduktion auf solche Konstituenten, die selbst keine argumentativen Konstituenten enthalten:



Rhetorical Structure Theory

Mann, William C. and Sandra A. Thompson (1983): Relational Propositions in Discourse, ISI: Information Sciences Institute of University of Southern California, Los Angeles, ISI/RR-83-115, 1-28.

Mann, William C. and Sandra A. Thompson (1986) Rhetorical Structure Theory: description and construction of text structures, Information Sciences Institute, Nijmegen, The Netherlands, ISI/RS-86-174, 1-15.

Mann, William C./Matthiessen, Christian M.I.M./Thompson, Sandra A. (1992): Rhetorical structure theory and text analysis. In: Mann, William C./Thompson, Sandra A.: Discourse Description: Diverse Analyses of a Fund Raising Text. Amsterdam: Benjamins, 39-78.

Weiterentwicklung der RST durch Daniel Marcu und andere

Rezeption weltweit

Neuste Entwicklung: Konstruktion eines diskursanalytischen Textkorpus auf der Basis von RST

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski: Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory.

Warum RST:

RST ermöglicht es, intentionale, semantische und textuale Information zu berücksichtigen.

Methode:

- Texte werden manuell annotiert
- Ergebnis sind RST-Bäume

Anwendungsgebiete:

- Summarization
- Verbesserung der MÜ
- Feedback bzgl. des Aufbaus von Texten beim Schreiben

Was ist RST?

Theorie, nach der ein Text durch **vier** Aspekte repräsentiert werden kann:

- Es entstehen Bäume, deren Blätter **minimalen Texteinheiten** entsprechen (*elementar discourse units*)
- Die Knoten des Baumes entsprechen benachbarten *Text Spans*.
- Jeder Knoten ist als **Nukleus** charakterisiert, der einen wesentlichen Teil der Information enthält. Er kann von einem **Satelliten** begleitet sein, der Hintergrund-Information anzeigt.
- Jeder Knoten ist durch eine **rhetorische Relation** (*rhetorical relation*) charakterisiert, der zwei oder mehr sich nicht überlappende benachbarte *text spans* verbindet

Die Relationen sind

- intentional, d.h. sie beziehen sich auf die Absichten des Sprechers
- semantisch, d.h. sie ergeben sich aus den Bedeutungen der Wörter oder
- textual, d.h. sie bezeichnen Beziehungen zwischen Ausdrücken auf der sprachlichen Oberfläche.

Schritte:

Erster Schritt: Festlegung der Diskurs-Einheiten:

Entscheidung (Marcu et al.): Jede isolierbare *Clause* ist ein Diskurseinheit. Sätze jedoch, die als Subjekt, Objekt eines Verbs fungieren, sind keine eigenen Diskurseinheiten.

Beispiele als dem Korpus von Marcu et al.:

5. [**Although** Mr. Freeman is retiring,] [he will continue to work as a consultant for American Express on a project basis.]wsj_1317
6. [Bond Corp., a brewing, property, media and resources company, is selling many of its assets] [**to reduce** its debts.]wsj_0630

Sätze, die als Subjekte, Objekte oder Verbcomplemente dienen, sind keine Diskurs-Einheiten:

7. [**Making computers smaller** often means **sacrificing memory.**]wsj_2387

8. [The company's current management found itself **locked into this,**" he said.]wsj_1103

Relativsätze dagegen sind eigene Diskurs-Einheiten:

9. [The results underscore Sears's difficulties] [**in implementing the "everyday low pricing" strategy...**]wsj_1105

10. [The Bush Administration,]1 [**trying to blunt growing demands from Western Europe for a relaxation of controls on exports to the Soviet bloc ,**] [is questioning...]wsj_2326

Schließlich: Einige Phrasale Diskurs-Einheiten, die am Beginn durch sog. Diskurs-Marker angezeigt sind:

Diskurs-Marker: *because, in spite of, as a result of, according to* .

Zweiter Schritt: Aufbau der Diskurs-Struktur:

Zwischen den Diskurs-Einheiten werden Relationen bestimmt, die den Diskurs in Text-Spans gliedern:

Zwei Typen von Relationen: zwischen Text-Spans:

- *Mono-Nukleus-Relations*
- *Multi-Nukleus-Relations*

Unterscheidung zwischen *Nukleus* und *Satellite*:

Nucleus: Text-Span, die wichtige, zentrale Information trägt

Satellite: Text-Span, die unterstützende Information trägt.

Mono-Nuclear relations: nur ein *nucleus* ist beteiligt, das andere Fundament der Relation ist ein *satellite*.

In dieser Konstellation ist der Nukleus wichtig für die Diskurs-Struktur, während der Satellit eher unterstützende Information trägt.

Multi-nuclear relations zwei oder mehr *spans*, die *nuclei* sind, sind beteiligt. Jede von diesen ist gleich gewichtig in der Diskurs-Struktur.

Für das Tagging des RST-Corpus wurden 53 Mononukleare und 25 Multinukleare Relationen verwendet.

Die Relationen werden durch Untersuchung des Korpus gewonnen, ihre Zahl kann je nach Textsorte verändert werden.

78 Relationen, in 16 Klassen gegliedert:

- *Attribution*: attribution, attribution-negative
- *Background*: background, circumstance
- *Cause*: cause, result, consequence
- *Comparison*: comparison, preference, analogy, proportion
- *Condition*: condition, hypothetical, contingency, otherwise
- *Contrast*: contrast, concession, antithesis
- *Elaboration*: elaboration -additional, elaboration -general-specific, elaboration -part-whole, elaboration-process-step, elaboration -object-attribute, elaboration -set-member, example, definition
- *Enablement*: purpose, enablement
- *Evaluation*: evaluation, interpretation, conclusion, comment
- *Explanation*: evidence, explanation-argumentative, reason
- *Joint*: list, disjunction
- *Manner-Means*: manner, means

- *Topic-Comment*: problem -solution, question -answer, statement -response, topic -comment, comment-topic, rhetorical-question
- *Summary*: summary, restatement
- *Temporal*: temporal-before, temporal-after, temporal-same-time, sequence, inverted-sequence
- *Topic Change*: topic-shift, topic-drift

Frühere Ansätze (Mann und Thompson) hatten andere Klassen:

Relationen

Table 1: Organization of the Relation Definitions

Circumstance	Antithesis and Concession
Solutionhood	Antithesis
Elaboration	Concession
Background	Condition and Otherwise
Enablement and Motivation	Condition
Enablement	Otherwise
Motivation	Interpretation and Evaluation
Evidence and Justify	Interpretation
Evidence	Evaluation
Justify	Restatement and Summary
Relations of Cause	Restatement
Volitional Cause	Summary
Non-Volitional Cause	Other Relations
Volitional Result	Sequence
Non-Volitional Result	Contrast
Purpose	

Für das Tagging des Corpus wurden Personen speziell in RST trainiert und mit einem RST-Tagging-Tool vertraut gemacht.

Wichtig sind die Beschreibungen der Relationen, weil daraus die Kriterien ihrer Anwendbarkeit hervorgehen:

Beispiele aus einer deutschen Anwendung der RST:

Strube, M. /U. Hahn (1996): Functional Centering. In: Proc. ACL, 270-277.

Rösner, Dietmar/Stede, Manfred (1992): Customizing RST for the Automatic Production of Technical Manuals. In: Dale, Robert/ Mellish, Chris/ Zock, Michael: Berlin und Heidelberg: Springer, 199-214.

Relationsdefinitionen für DB-Text

Relationsname:	CONTRAST
Bedingungen an N:	multinuklear
Bedingungen an die Kombination der Nuklei:	Nicht mehr als zwei Nuklei; die in diesen beiden Nuklei präsentierte Situation wird a) als in vieler Hinsicht identisch verstanden, b) als in einigen Aspekten differierend, und c) verglichen in Beziehung zu einem oder mehreren dieser Unterschiede
Effekt:	L erkennt die Vergleichbarkeit und die durch den Vergleich deutlich gemachten Unterschiede
Ort des Effekts:	multiple Nuklei

Relationsdefinitionen für DB-Text

Relationsname:	ELABORATION		
Bedingungen an N:	keine		
Bedingungen an S:	keine		
Bedingungen an die Kombination von N + S:	S präsentiert zusätzliche Detailinformation über die Situation oder ein Element des Gegenstandsbereichs, die in N präsentiert werden oder durch N inferentiell zugänglich sind, auf eine oder mehrere der unten genannten Arten. In der folgenden Zusammenstellung gilt: Wenn für N das erste Element eines Paares zutrifft, ist für S dann das zweite zutreffend:		
		1. Menge:	Element
		2. Abstrakt:	Instanz
		3. Ganzes:	Teil
		4. Prozeß:	Prozeßschritt
		5. Objekt:	Attribut
		6. Generalisierung:	Spezifische Information
	Effekt:	Leser erkennt, daß die in S präsentierte Situation zusätzliche Details für N zur Verfügung stellt. L identifiziert das Element des Gegenstandsbereichs, für das zusätzliche Details verfügbar gemacht werden.	
	Ort des Effekts:	N + S	

Relationsdefinitionen für DB-Text

Relationsname: SEQUENCE

Bedingungen an N: multinuklear

Bedingungen an die Kombination der Nuklei:

Zwischen den in den Nuklei präsentierten Situationen muß eine NachfolgerRelation bestehen.

Effekt: L erkennt die Nachfolgerbeziehungen zwischen den Nuklei

Ort des Effekts: multiple Nuklei

Prüfung der Qualität der Kodierung durch Messung der Intercoder-Reliabilität

Handbuch:

<http://www.isi.edu/~marcu/discourse>.

Die Qualität des Korpus wurde durch mehrere Methoden validiert:

Durchsicht durch mehre Kodierer auf Validität

Consistenzmessung (Intercoder-reliability) mit dem Kappa-Maß.

Beispiel:

[Still, analysts don't expect the buy -back to significantly affect per -share earnings in the short term.]¹⁶

[“The impact won’t be that great,”] 17

[said Graeme Lidgerwood of First Boston Corp.] 18

[This is in part because of the effect] 19 [

of having to average the number of shares outstanding,] 20

[she said.]21

[In addition,] 22

[Mrs. Lidgerwood said,] 23

[Norfolk is likely to draw down its cash initially] 24

[to finance the purchases]25

[and thus forfeit some interest income.] 26

wsj_1111 (*Kennung*)

RST-Diskurs-Struktur dieses Textes:

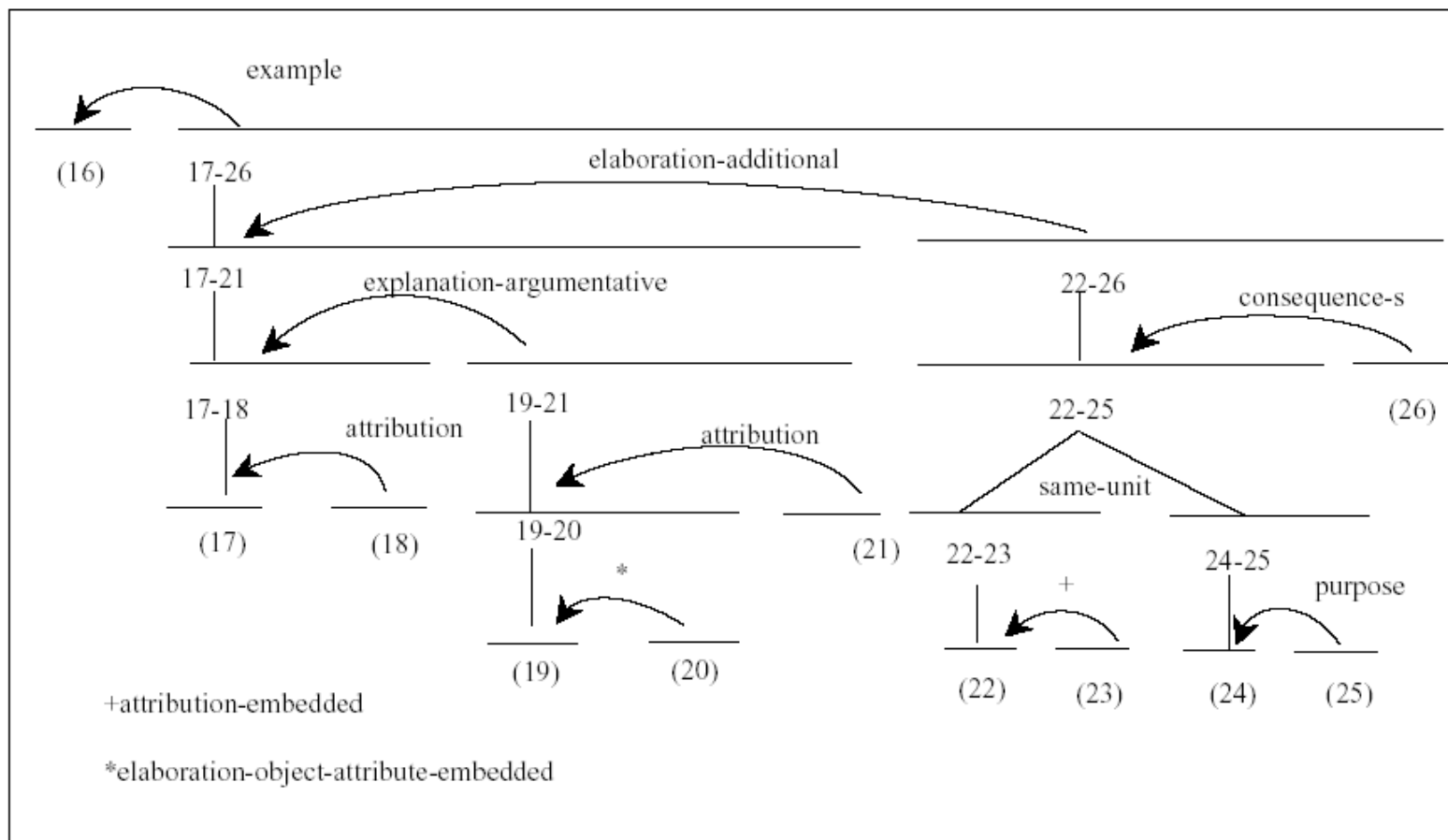


Figure 1. Discourse sub-tree for multiple sentences

Zwei mögliche Abläufe des Kodierprozesses:

- Der Kodierer überblickt das gesamte Textstück und entscheidet, dass es sich zwischen Einheit 16 und dem Rest, der 17 - 26 überspannt, um die *example*-Relation handelt (daher: *text span*).

Der Kodierer kann später seine Entscheidungen rvidieren.

- Der Kodierer betrachtet jeden einzelnen Satz und bildet zunächst ‚subtrees‘:

Ermittlung der *spans* [16], [17 -18], [19 -21], [22 -26].

Anschließend Verbindund des 3. und 4. subtrees durch die Relation *explanation-argumentative*

danach Verbindung des 4. sub-tree mit diesem durch die Relation *elaboration additional*

Schließlich Zuordnung der Spanne [17 -26] dem Knoten [16] als *example*.

Man sieht hier, dass in der Struktur schematische Teilstrukturen enthalten sind.

Mann und Thompson haben fünf Basis-Schemata unterschieden, zwei mononukleare und drei multinukleare Schemata:

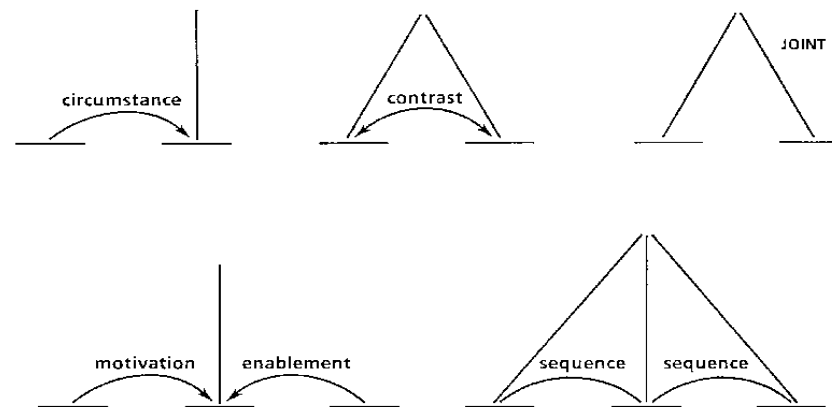
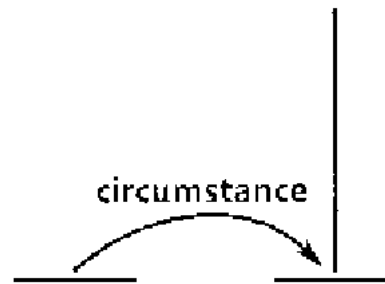
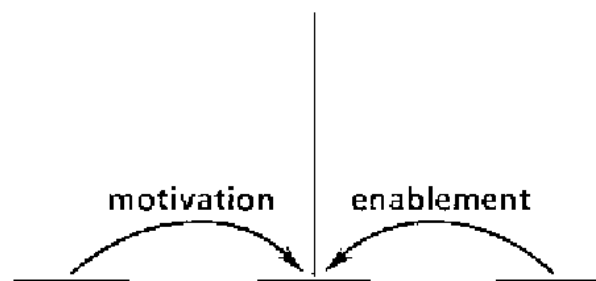


Figure 1: Examples of the Five Schema Types

Mononukleare Schemata

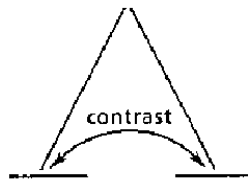


Der am häufigsten vorkommende Typ besteht aus einer einzigen Relation mit Nukleus und Satellit.

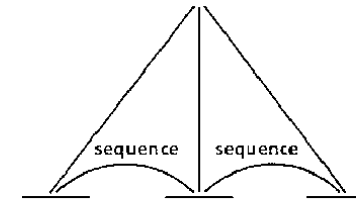


Ein weiteres Schema ist für den Fall vorgesehen, dass ein Textsegment als Nukleus gleichzeitig mit Satelliten aus unterschiedlichen Relationen verbunden ist.

Multinukleare Schemata



Das CONTRAST-Schema hat immer exakt zwei Nuklei.



SEQUENCE hat unbestimmt viele Nuklei, wobei zwischen den Inhalten benachbarter Nuklei im Gegenstandsbereich eine Nachfolgerrelation besteht.



JOINT hat unbestimmt viele Nuklei ohne korrespondierende Satelliten.

JOINT ist eine "inhaltsleere" Relation und immer dann anzusetzen, wenn zwei Textsegmente lediglich benachbart in einem text vorkommen, aber durch keine andere Relation verbunden sind.

Beispiele für unterschiedliche Textstrukturen:

Beispiel für ein stark strukturiertes Dokument mit öfteren Topics-Wechseln

Beispiel für eine Erzählung

RST Structurer

File Display Print Save Help

Auto Link Disconnect Modify Change Relation Change Schema Undo Reduce Enlarge Show Text Sent

advertising the promotion.

/home/nac/lmcarls/TREES-ANALYSIS/ws_j_1193.out

```

Good grief!<1> Charlie Brown is selling out.<2>
Those Metropolitan Life ads were bad enough.<3>
But now, Charlie Brown is about to start pitching everything from Chex Party Mix to light bulbs.<4>
Why is he cashing in now?<5>
Turns out that next year, Charlie Brown, Snoopy and the gang turn 40 <6>-- and Scripps Howard's United Media unit, the syndicator and licensing agent for Charles Schulz's comic strip, sees a bonanza in licensing the cartoon characters to a bevy of advertisers for ads, tie-ins and promotions.<7>
*Peanuts has become a major part of American culture,* <8>says Peter Shore, United Media's vice president of marketing and licensing.<9>
The comic strip *has a magical, everlasting quality about it.<10>
Our plan is to honor Charles Schulz and the strip all year long.*<11>
The effort will make the Peanuts gang very familiar pitchmen in 1990.<12>
General Electric plans to use the characters <13>to plug its Miser light bulb.<14>
Teleflora will run TV ads at Valentine's Day <15>promoting its *Snoopy's Love Bouquet.*<16> Ralston Purina will

```

Beispiel für ein stark strukturiertes Dokument.

Mehrere kurze thematisch abgrenzbare News mit klaren Überschriften.

Diese thematischen Bereiche sind durch die Relation TOPIC-SHIFT verknüpft.

Mehrere Schemata werden durch eine Relation mit dem Namen TEXTUAL-ORGANIZATION verknüpft.

RST Structurer

File Display Print Save Help

Auto Link Disconnect Modify Change Relation Change Schema Undo Reduce Enlarge Show Text Sent

elling on the spot market for around \$1.47 a thousand cubic feet, down 13% from \$1.69 a thousand cubic feet a year
/home/nac/1mcar1s/TREES-ANALYSIS/ws_j_2325.out

The oil industry's middling profits could persist through the rest of the year.<1>

Major oil companies in the next few days are expected to report much less robust earnings <2>than they did for the third quarter a year ago,<3> largely reflecting deteriorating chemical prices and gasoline profitability.<4>

The gasoline picture may improve this quarter,<5> but chemicals are likely to remain weak,<6> industry executives and analysts say,<7> reducing chances <8>that profits could equal their year-earlier performance.<9>

The industry is "seeing a softening somewhat in volume and certainly in price in petrochemicals," <10>Glenn Cox, president of Phillips Petroleum Co., said in an interview.<11> "That change will obviously impact third and fourth quarter earnings" for the industry in general,<12> he added.<13> He didn't forecast Phillips's results.<14>

But securities analysts say <15>Phillips will be among the companies hard-hit by weak chemical prices<16> and will probably post a drop in third-quarter earnings.<17> So, too,<18> many analysts predict,<19> will Exxon Corp., Chevron Corp. and Amoco Corp.<20>

Dieses Dokument zeigt die typische journalistische Praxis, eines Artikels mit einer einleitenden Zusammenfassung und nachfolgenden Details.

RST Structurer

File Display Print Save Help

Auto Link Disconnect Modify Change Relation Change Schema Undo Reduce Enlarge Show Text

where the roadway collapsed.

/home/nac/lmcarls/TREES-ANALYSIS/file3

```

She lives in Oakland, a community<214> hit hard by the earthquake.<215>
She didn't have hot water for five days.<216>
The apartment<217> she shares with a 12-year-old daughter and her sister
<218>was rattled,<219> books and crystal hit the floor,<220> but nothing was severely
damaged.<221>
Her sister, Cynthia, wishes<222> Toni had a different job.<223>
*We worry about her out there,* <224>Cynthia says.<225>
Last Sunday, Ms. Johnson finally got a chance <226>to water her plants,<227> but stopped abruptly.<228>
*I realized<229> I couldn't waste this water<230> when there are people
in Watsonville <231>who don't have fresh water<232> to drink.*<233>
She hasn't played any music <234>since the earthquake hit,<235> out of
respect for those <236>who died on Interstate 880 <237>where the roadway collapsed.<238>

```

Dieses Dokument zeigt den TOPIC-DRIFT style:
Mischung von beschreibendem und narrativem Stil:

„A human-interest story is woven around an expository text detailing the impact of the San Francisco earthquake on the insurance trade. The text begins with an on-the-scene description of an insurance claims adjuster assessing damages to a house destroyed by the San Francisco earthquake, switches to a general description of the impact on the insurance trade, and then returns to the personal experience of the adjuster.”