

Textkohärenz

Ziel: Darstellung von zwei Theorien, die entwickelt wurden, um die inhaltliche Kohärenz von Texten zu beschreiben:

- Rhetorische Strukturtheorie
- Quaestio-Theorie und struktuiierende Fragen

RST

Zielsetzung

Entwickelt aus der automatischen Textgenerierung (Vertextung der Information in natürlichsprachlichen Datenbanken).

Wichtige Werke:

Mann, William C. & Sandra A. Thompson (1988), "Rhetorical structure theory. Towards a theory of text organization", *Text* 8, 243-281.

Taboada, Maite & William C. Mann (2006), "Rhetorical Structure Theory: Looking back and moving ahead", *Discourse Studies* 8, 423-459.

Taboada, Maite & William C. Mann (2006), "Applications of Rhetorical Structure Theory", *Discourse Studies* 8, 567-588.

Stede, Manfred (2003), "Surfaces and depths in text understanding: The case of newspaper commentary", *Proc. of the HLT/NAACL Workshop on Text Meaning*,

Siehe auch Webseite der RST: <http://www.sfu.ca/rst/>

Aufgabe der RST: Systematische Untersuchung, wie gut gebaute, "kohärente" natürlichsprachliche Texte unterschiedlicher Art aufgebaut sind.

Kohärent sind Texte, für die gilt:

- Es gibt für jeden Textteil eine für den Leser evidente Funktion im Hinblick auf den gesamten Text
- Es fehlt kein Teil, der für das Verständnis des Gesamttextes wichtig ist.

Textblöcke, Nuklei und Satelliten

Texte werden in einzelne Abschnitte, “Blöcke”, zerlegt.

Dieses Verfahren kann rekursiv angewendet werden, d.h. größere Blöcke können kleinere Blöcke enthalten.

Unter den Blöcken werden zwei Typen unterschieden: zum einen “nukleare” Blöcke, zum anderen “Satelliten”, die mit solchen in Beziehung stehen unterschieden. Daneben gibt es auch sog. Schemata, die eine Folge von Blöcken zusammenfassen.

Die rhetorischen Beziehungen zwischen Satellit und Nukleus können nach verschiedenen inhaltlichen Kriterien klassifiziert werden. Beispiele für häufige Beziehungen dieser Art:

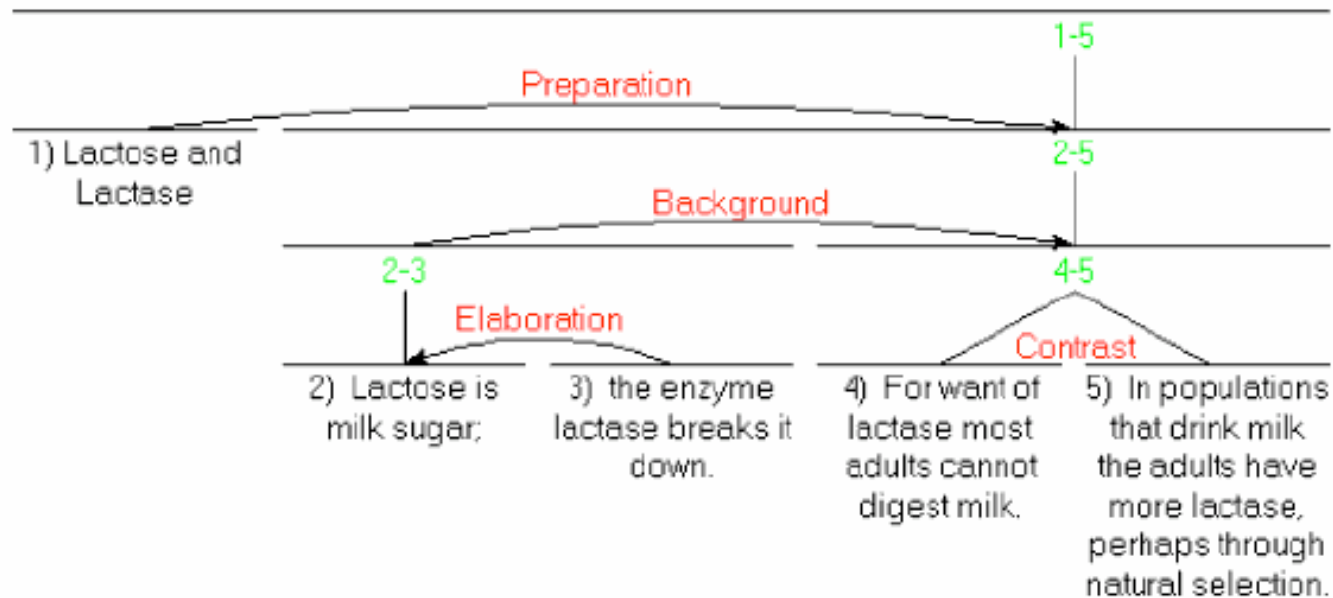
Relation	Nukleus	Satellit
Evidenz	Behauptung	Evidenz für die Behauptung
Hintergrund (background)	Text, dessen Verständnis erleichtert werden soll	Text zur Erleichterung des Verständnisses
Elaboration	elementare Information	zusätzliche Information
Vorbereitung (preparation)	Text, der präsentiert werden soll	Text, der den Leser auf N vorbereitet und den Leser hilft, N zu interpretieren.

Es gibt aber auch Sequenzen von Textteilen, die gleich wichtig sind – sogenannte multinukleare Textblöcke. Ein Beispiel hierfür ist die Kontrastrelation:

Relation	Nukleus ('span')	Nukleus ('span')
Kontrast	Alternante	Andere Alternante

Beispielanalyse

Die RST analysiert Texte auf rekursive Weise, und die rhetorischen Beziehungen werden damit ebenfalls rekursiv angewendet. Ein Beispiel:



Analyse eines entsprechenden deutschen Textes, in Tabellennotation

- (1) *(a) Lactose und Laktase. (b) Laktose ist Milchzucker, (c) das Enzym Laktase baut diesen ab. (d) Den meisten erwachsenen Menschen fehlt Laktase, (e) und daher können sie Laktose nicht abbauen. (f) In Populationen mit hohem Milchkonsum haben Erwachsene mehr Laktase. (g) Dies ist vermutlich ein Ergebnis der natürlichen Selektion.*

Vorbe- reitung>					
<i>Laktose und Lak- tase.</i>	Hintergrund>				
	<i>Laktose ist Milch- zucker,</i>	<Elaborat.	Kontrast>		<Kontrast
		<i>das Enzym Laktase baut diesen ab.</i>	Begründung>	<i>und daher können sie Laktose nicht abbauen.</i>	<i>In Poplulationen mit hohem Milchkonsum haben Erwachsene mehr Laktase.</i>
		<i>Den meisten erwachsenen Menschen fehlt Laktase</i>			<i>Dies ist vermutlich ein Ergebnis der natürlichen Selektion.</i>

Analysetiefe und Explizitheit

Soll die Textanalyse unter die Ebene von zusammengesetzten Sätzen vordringen?

Ja, denn bestimmte Beziehungen können sowohl satzintern als auch als rhetorische Relationen dargestellt werden.

- (2) a. *Die meisten Menschen können Laktose nicht abbauen, weil ihnen Laktase fehlt.*
b. *Die meisten Menschen können Laktose nicht abbauen, denn ihnen fehlt Laktase.*
c. *Die meisten Menschen können Laktose nicht abbauen. Es fehlt ihnen nämlich Laktase.*

RST-Analysen unterscheiden sich auch danach, ob ein Satz nur mit benachbarten Sätzen in Bezug gesetzt wird, oder ob der gesamte Text durchanalysiert wird.

Oft ist die Analyse des Bezugs zu Nachbarsätzen eindeutig, nicht aber der große Zusammenhang.

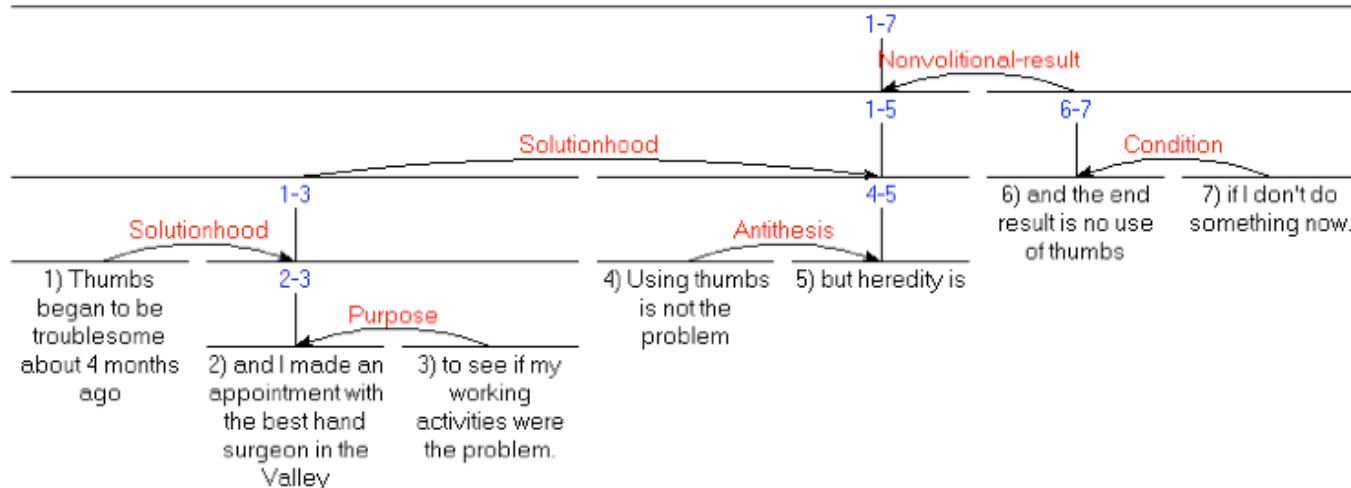
Rhetorische Relationen können explizit sein, z.B. in der Verwendung von Partikeln (*weil, denn, nämlich*), Überschriften, Interpunktion und bei gesprochenen Texten auch durch die Prosodie. Sie bleiben aber oft auch implizit bleiben:

- (3) *Die meisten Menschen können Laktose nicht abbauen. Es fehlt ihnen das Enzym Laktase.*

Weitere Beispiele für RST-Analysen

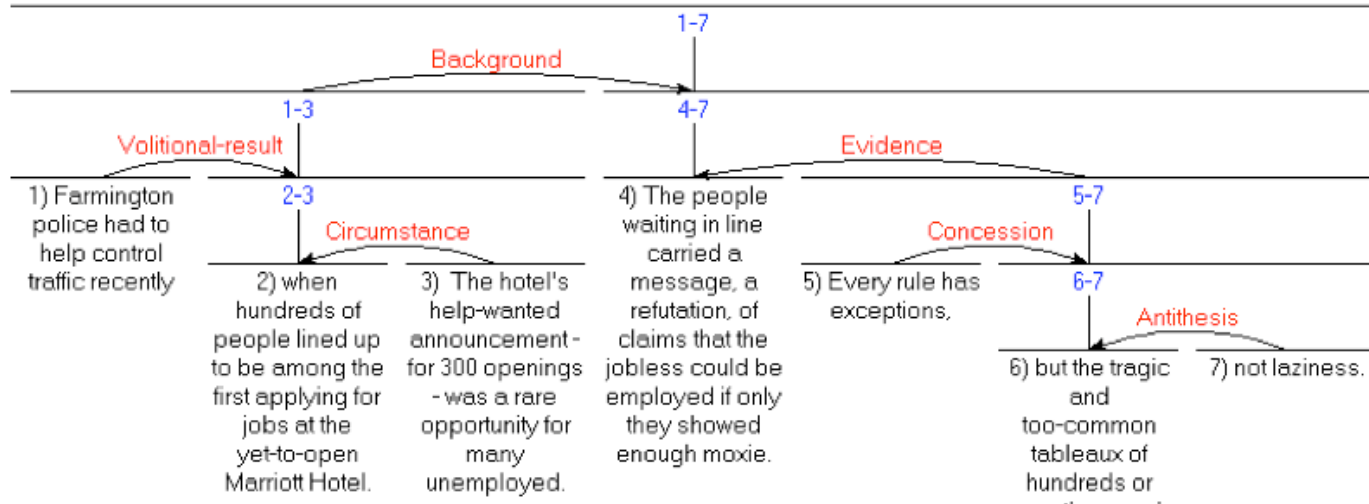
Persönlicher Brief

- (4) *1) Thumbs began to be troublesome about 4 months ago 2) and I made an appointment with the best hand surgeon in the Valley 3) to see if my working activities were the problem. 4) Using thumbs is not the problem 5) but heredity is 6) and the end result is no use of thumbs 7) if I don't do something now*



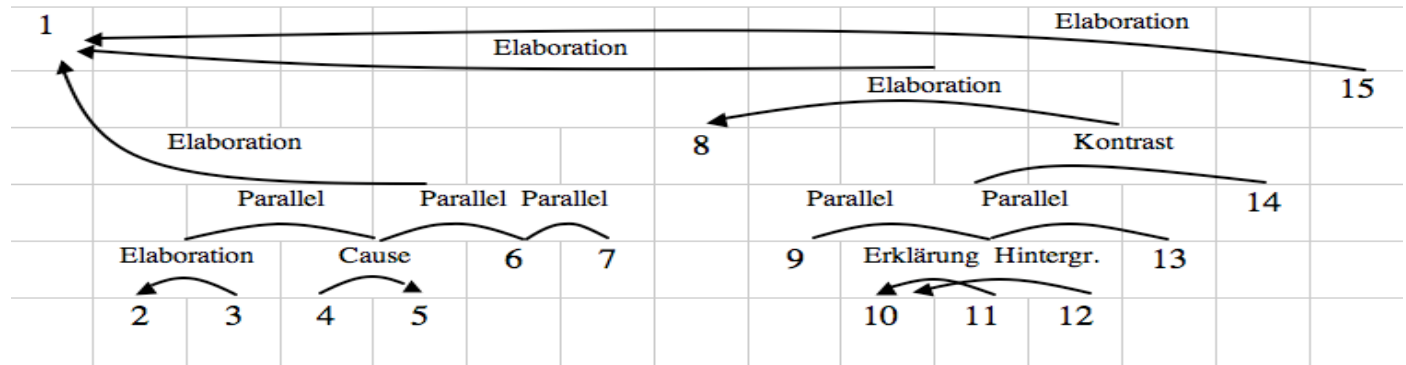
Zeitungskommentar

- (5) 1) Farmington police had to help control traffic recently 2) when hundreds of people lined up to be among the first applying for jobs at the yet-to-open Marriott Hotel. 3) The hotel's help-wanted announcement - for 300 openings - was a rare opportunity for many unemployed. 4) The people waiting in line carried a message, a refutation, of claims that the jobless could be employed if only they showed enough moxie. 5) Every rule has exceptions, 6) but the tragic and too-common tableaux of hundreds or even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs, 7) not laziness.



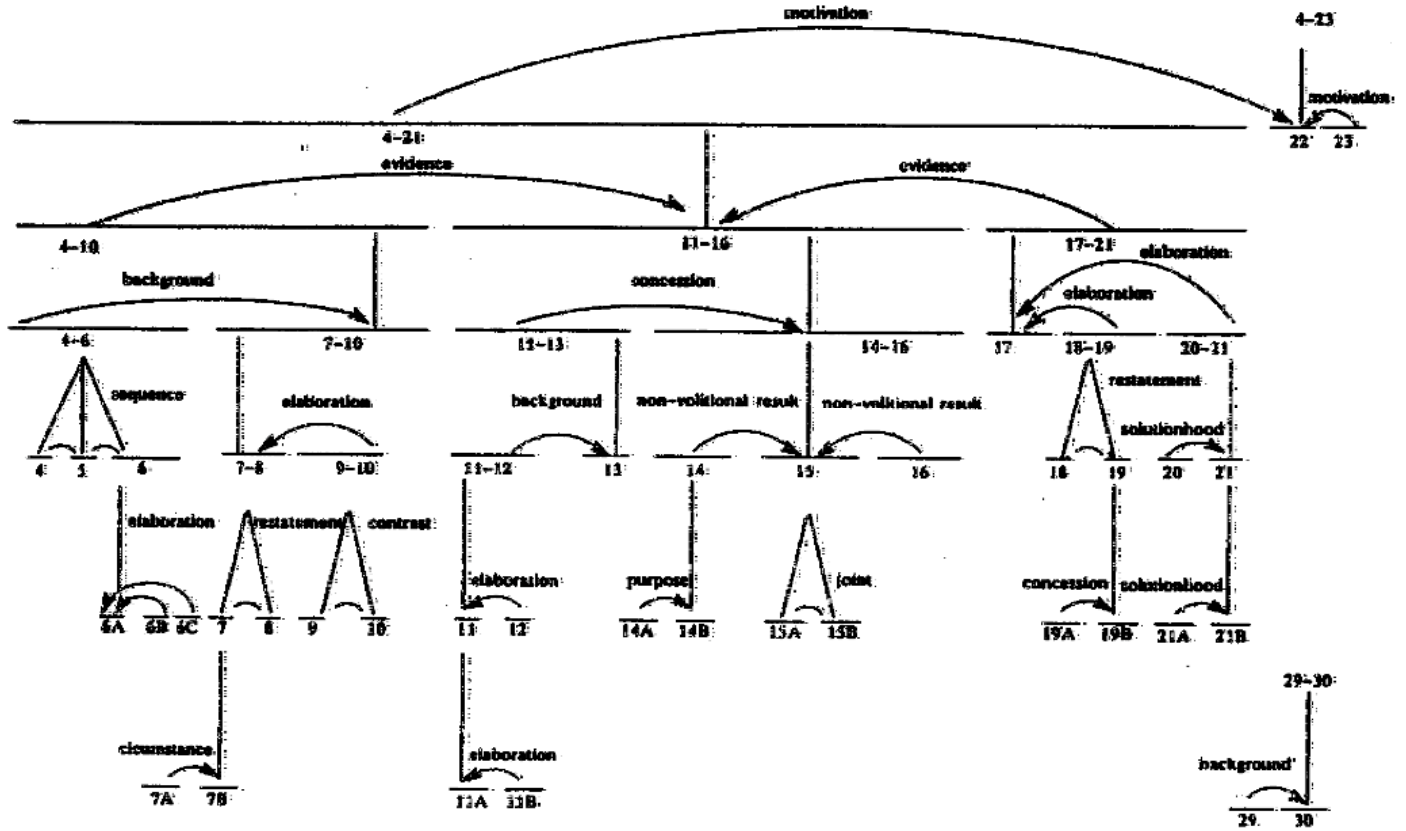
Das große Bibbern. (Aus der FAZ vom 24. 1. 2006)

(1) *Deutschland zittert wie der Osten Europas weiter unter der extremen Kälte mit zweistelligen Minusgraden.* (2) *Die eisigen Temperaturen forderten ein fünftes Opfer:* (3) *In Brandenburg starb ein Mann an den Folgen einer Unterkühlung.* (4) *Im Nachbarland Polen kamen neun weitere Menschen durch den Frost ums Leben,* (5) *so daß die Zahl der Kältetoten seit Winterbeginn auf 159 stieg.* (6) *Auch in Ungarn und Frankreich erfroren Menschen.* (7) *Die verschneite Akropolis in der griechischen Hauptstadt Athen wurde wegen Glättegefahr für Besucher geschlossen.*
 (8) *In Deutschland war es in der Nacht zum Dienstag vielerorts noch kälter als in der eisigen Nacht zuvor.* (9) *Am unbewohnten bayerischen Funtensee wurden 36,2 Grad unter Null gemessen, nach 34,8 Grad Frost in der Nacht zum Montag.* (10) *Mit einer Temperatur von minus 29,1 Grad erlebte auch die bayerische Gemeinde Haidmühle die bisher kälteste Nacht des Jahres.* (11) *Wie der Wetterdienst Meteomedia erklärte, lag die zwischen sieben und acht Uhr gemessene Temperatur ein halbes Grad unter der, die in der Nacht zum Montag festgestellt worden war.* (12) *Haidmühle, ein staatlich anerkannter Erholungsort, liegt im Dreiländereck Bayern-Tschechien-Österreich.* (13) *In Morgenröthe-Rautenkranz in Sachsen froren die Bewohner bei minus 22,9 Grad.* (14) *Am wärmsten war die Nacht mit minus 3,1 Grad in Krefeld in Nordrhein-Westfalen.* (15) *Das kräftige Hoch „Claus“ soll die Temperaturen noch bis Mittwoch tief im Keller halten.*



Bittbrief für Spenden

- (6) *4) At 7:00 a.m. on October 25, our phones started to ring. 5) Calls jammed our switchboard all day. 6A) Staffers stayed late into the night, 6B) answering questions 6C) and talking with reporters from newspapers, radio stations, wire services and TV stations in every part of the country.*
- 7A) When we released the results of ZPG's 1985 Urban Stress Test, 7B) we had no idea we'd get such an overwhelming response. 8) Media and public reaction has been nothing short of incredible! 9) At first, the deluge of calls came mostly from reporters eager to tell the public about Urban Stress Test results and from outraged public officials who were furious that we had "blown the whistle" on conditions in their cities. 10) Now we are hearing from concerned citizens in all parts of the country who want to know what they can do to hold local officials accountable for tackling population-related problems that threaten public health and well-being.*
- 11A) ZPG's 1985 Urban Stress Test, 11B) created after months of persistent and exhaustive research, 11A) is the nation's first survey of how population-linked pressures affect U.S. cities. 12) It ranks 184 urban areas on 11 different criteria ranging from crowding and birth rates to air quality and toxic wastes. 13) The Urban Stress Test translates complex, technical data into an easy-to-use action tool for concerned citizens, elected officials and opinion leaders.*
- 14A) But to use it well, 14B) we urgently need your help. 15A) Our small staff is being swamped with requests for more information 15B) and our modest resources are being stretched to the limit. 16) Your support now is critical.*
- 17) ZPG's 1985 Urban Stress Test may be our best opportunity ever to get the population message heard. 18) With your contribution, ZPG can arm our growing network of local activists with the materials they need to warn community leaders about emerging population-linked stresses before they reach crisis stage. 19A) Even though our national government continues to ignore the consequences of uncontrolled population growth, 19B) we can act to take positive action at the local level.*
- 20) Every day decisions are being made by local officials in our communities that could drastically affect the quality of our lives. 21A). To make sound choices in planning for people, 21B) both elected officials and the American public need the population-stress data revealed by our study. 22) Please make a special contribution to Zero Population Growth today. 23). Whatever you give -- \$25, \$50, \$100 or as much as you can -- will be used immediately to put the Urban Stress Test in the hands of those who need it most.*



Das Inventar der Rhetorischen Relationen

In der RST wurden etwa 20 rhetorische Relationen postuliert, die in drei Gruppen gegliedert werden:

- die präsentationellen Relationen, die mit der Darstellung von Information zu tun haben;
- die thematischen Relationen (“subject matter relations”), welche die inhaltlichen Beziehungen zwischen Textteilen betreffen,
- die multinuklearen Relationen, die nicht aus Nukleus-Satellit-Beziehungen bestehen.

Relationen werden angegeben durch Bedingungen an den Nukleus, den Satellit, die Kombination von Nukleus und Satellit und die Intentionen des Autors

Beispiele: Präsentationelle Relationen

N: Nukleus, S: Satellit, A: Autor, L: Leser

- **Antithese.** N und S stehen in einer Kontrastrelation; A hält N für positiv.
 - (7) *But I don't think endorsing a specific nuclear freeze proposal is appropriate for CCC. We should limit our involvement in defense and weaponry to matters of process, such as exposing the weapons industry's influence on the political process.*
- **Background.** L würde N nicht ohne S verstehen.
- **Concession.** A hält N für positiv. A behauptet nicht, dass S nicht zutrifft. A räumt eine tatsächliche oder potentielle Inkompatibilität zwischen N und S ein.
- **Enablement.** Verständnis von S erhöht die Fähigkeit von L, N auszuführen oder zu verstehen.
 - (8) *Tempting as it may be, we shouldn't embrace every popular issue that comes along.*
- **Evidence.** L würde N ohne S nicht für wahr halten.
 - (9) *Tempting as it may be, we shouldn't embrace every popular issue that comes along. When we do so, we use precious, limited resources where other players with superior resources are already doing an adequate job.*
- **Motivation.** Verständnis von S erhöht die Bereitschaft von L, N zu glauben oder N auszuführen.
- **Preparation.** S geht N voran; S erhöht das Interesse von L für N
- **Restatement.** S reformuliert die Bedeutung von N, wobei S und N von ähnlicher Komplexität sind.
- **Summary.** Wie bei Restatement, allerdings ist S kürzer.

Beispiele: Thematische Relationen

- **Circumstance.** S gibt einen Rahmen, innerhalb dessen N interpretiert wird.
 - (10) *Probably the most extreme case of Visitors Fever I have ever witnessed was a few summers ago when I visited relatives in the Midwest.*
- **Condition.** S stellt eine hypothetische Bedingung für die Realisierung von N vor.
 - (11) *Employees are urged to complete new beneficiary designation forms for retirement or life insurance benefits whenever there is a change in marital or family status*
- **Elaboration.** S gibt zusätzliche Information zu N.
 - (12) *Lund, in Sweden, will be the site of the 1969 International Conference on Computational Linguistics, September 1-4. It is expected that some 250 linguists will attend from Asia, West Europe, East Europe including Russia, and the United States.*
- **Non-volitional result.** N verursachte S; N ist die wichtigere Information.
 - (13) *The blast, the worst industrial accident in Mexico's history, destroyed the plant and most of the surrounding suburbs. Several thousand people were injured, 3. and about 300 are still in hospital.*
- **Solutionhood.** N ist eine Lösung des durch S gestellten Problems.
 - (14) *One difficulty ... is with sleeping bags in which down and feather fillers are used as insulation. This insulation has a tendency to slip toward the bottom. You can redistribute the filler. ...*
- **Volitional cause.** Der S-Sachverhalt bringt den Handelnden in N dazu, N auszuführen.
 - (15) *Thinking that perhaps the counsel was simplistic, coming from an unmarried person, one of them asked, "Are you married?"*

Multinukleare Relationen

- **Contrast.** N1 und N2 sind in einigen Hinsichten ähnlich, in anderen unähnlich, und werden hinsichtlich dieser Unterschiede miteinander verglichen.
(16) *Animals heal, but trees compartmentalize.*
- **Sequence.** Die Nuklei beschreiben aufeinanderfolgende Ereignisse.
(17) *Peel oranges, and slice crosswise. Arrange in a bowl and sprinkle with rum and coconut. Chill until ready to serve.*
- **Joint.** (Keine Bedingungen)
(18) *Features like our uniquely sealed jacket and protective hub ring make our discs last longer. | And a soft inner liner cleans the ultra-smooth disc surface while in use.*

Systematisierung der rhetorischen Relationen

A. Kehler (2002): *Coherence, Reference, and the Theory of Grammar*, Stanford, CSLI Press

Kehler plädiert dafür, die rhetorischen Relationen systematisch mit den drei kognitiven Relationen der Ideenassoziation zu erfassen, die David Hume 1748 in *Inquiry Regarding Human Understanding* erfasst hat. (Er folgt darin Hobbs 1990, *Literature and Cognition*).

Hume unterscheidet also zwischen den folgenden Ähnlichkeitsrelationen zwischen Begriffen:

- **Resemblance**: Ähnlichkeit, Parallelismus
- **Contiguity**: Räumliche und zeitliche Nachbarschaft
- **Cause/Effect**: Ursache/Wirking und Zweck/Handlungs-Beziehungen

Ähnlichkeitsrelationen

Ein Beispiel für die Ähnlichkeitsbeziehung von Kehler (aus dem amerik. Wahlkampf 2000):

(19) *Dick Gephardt hat Veranstaltungen für Al Gore organisiert,
und Tom Daschle hat für ihn Wahlkampfbroschüren verteilt.*

Es werden hier zwei Sätze S_1 , S_2 ausgedrückt, die Parallelen aufweisen. Beide drücken eine Form von Unterstützung für den Präsidentschaftskandidaten Al Gore aus. Sie variieren an zwei Positionen: S_1 ist eine Proposition über den demokratischen Politiker Dick Gephardt, und die Form der Unterstützung ist, Versammlungen für Gore zu organisieren. S_2 ist eine Proposition über den demokratischen Politiker Tom Daschle, und die Form der Unterstützung ist, Wahlkampfbroschüren zu verteilen.

Die Art der Ähnlichkeitsbeziehungen versucht Kehler systematisch wie folgt zu bestimmen: Die Sätze S_1 , S_2 drücken Propositionen p_1 , p_2 über Entitäten a_1, a_2, \dots, a_n bzw. b_1, b_2, \dots, b_n aus:

(20) $S_1: p_1(a_1, a_2, \dots, a_n)$
 $S_2: p_2(b_1, b_2, \dots, b_n)$

In unserem Beispiel:

(21) $S_1: \text{VERANSTALTET-VERSAMMLUNGEN-FÜR}(\text{GEPHART}, \text{GORE})$
 $S_2: \text{VERTEILT-BROSCHÜREN-FÜR}(\text{DASCHLE}, \text{GORE})$

p_1 und p_2 , a_1 und b_1 , a_2 und b_2 usw. werden **parallele Elemente** genannt.

Kehler unterscheidet eine Reihe von Untertypen (Kontrast, Exemplifikation, Generalisierung, Ausnahme, Elaboration)

Ursache-Wirkungs-Relationen

Der Adressat soll zwischen zwei von S_1 , S_2 implizierten Propositionen P_1 , P_2 eine Beziehung (Implikation, \rightarrow) aufstellen. Kehler diskutiert Untertypen wie die folgenden:

➤ **Resultat:** Erschließe P_1 aus S_1 und P_2 aus S_2 , wenn üblicherweise $P_1 \rightarrow P_2$ angenommen werden kann

(22) *Koch ist ein Politiker; man sollte ihm mißtrauen.*

➤ **Erklärung:** Erschließe P_1 aus S_1 und P_2 aus S_2 , wenn üblicherweise $P_2 \rightarrow P_1$ angenommen werden kann.

(23) *Man sollte Koch mißtrauen. Er ist ein Politiker.*

➤ **Verletzte Erwartung.** Erschließe P_1 aus S_1 und P_2 aus S_2 , obwohl üblicherweise $P_1 \rightarrow \neg P_2$ angenommen werden kann.

(24) *Koch ist ein Politiker, aber man kann ihm trauen.*

Kontiguitätsrelationen

➤ Okkasion (i): Erschließe einen Zustandswechsel für Entitäten von S_1 , wobei S_2 den finalen Zustand angibt.

➤ Okkasion (ii): Erschließe einen Zustandswechsel für Entitäten von S_2 , wobei S_1 den initialen Zustand angibt.

(25) *Merkel setzte die Brille auf. Sie begann zu lesen.*

Quaestio-Theorie

W. Klein & C. von Steutterheim (1987): "Quaestio und referentielle Bewegung in Erzählungen", *Linguistische Berichte* 109, 163-185.

W. Klein & C. von Steutterheim (2003), "How to solve a complex verbal task: Text structure, referential movement and the quaestio"

Grundannahmen

Texte sind durch implizite Fragen strukturiert. Diese Fragen ergeben sich aus dem allgemeinen Zweck des Textes und sind von bestimmten Erwartungshaltungen gesteuert.

Beispiel: Sie nehmen an einer Sommerschule teil, die mit einem Stipendium finanziert wird. Was berichten Sie einem Freund über die Sommerschule? Wie schreiben Sie einen Bericht über ihre Aktivitäten an die Organisation, die Ihre Teilnahme finanziert?

Übergeordnete/untergeordnete Quaestiones

Über diese spezifischen Dinge hinaus sagt die Quaestio-Theorie aber einiges über die Struktur von Texten im Allgemeinen:

- Es gibt für einen Text eine Haupt-Quaestio, die typischerweise durch eine Reihe von untergeordneten Quaestiones ausgefaltet wird, welche oft wiederum aus Unter-Quaestiones bestehen (hierarchische Struktur, vgl. Urlaubs-Beispiel).
- Für Äußerungen kann man sog. Referenzbereiche definieren: Zeiten, Orte, Personen, Objekte, Ereignisse, Zustände, Eigenschaften, Möglichkeiten usw.
- Die untergeordneten Quaestiones geben typischerweise Information über die Entitäten dieser Referenzbereiche ("Referentielle Bewegung": Erhalt, Verschiebung oder Wechsel der referentiellen Entitäten, über die Aussagen gemacht werden).
- Hauptstruktur: Äußerungen über die Entitäten der Referenzbereiche. Nebenstruktur: Hintergrundinformationen, wie z.B. Bewertungen (*Das hätte ich nicht tun sollen*), Kommentare (*Solche Sachen passieren halt, wenn man aufgeregt ist*) usw.

Diskursstrukturierende Fragen

Van Kuppevelt, Jan (1995), "Discourse structure, topicality, and questioning", *Journal of Linguistics* 31, 109-147.

Grundidee: Aus einer Behauptung ("Feeder" F) entsteht eine oder mehrere Fragen, deren Beantwortung wieder zu Unterfragen führen. Beispiele:

(26) *F: Students are no longer allowed to take more than 6 years over their first degree.*

Q1: Whose decision has this been?

A1: This has been the decision of the Minister of Education and Science.

Q2: What is the reason for this decision?

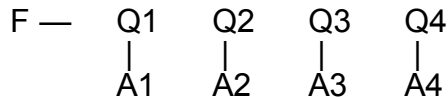
A2: It has been decided to cut the education budget drastically.

Q3: When will the measure become operative?

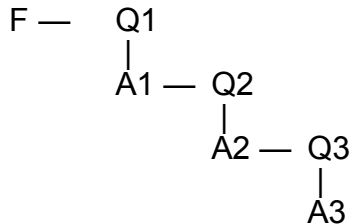
A3: It will become operative at the beginning of the new academic year.

Q4: What is expected to be the effect of this measure?

A4: One expects that in the years to come student numbers will go down.



- (27) F: *Mary is worried.*
 Q1: *Why?*
 A1: *John, her husband, wants to buy a DAT-recorder.*
 Q2: *Why is she worried about that?*
 A2: *He doesn't have enough money, so he has to borrow it.*
 Q3: *Why is that a problem?*
 A3: *He already has a lot of debts.*



Bezug auf Diskursstruktur:

- Rezenz: Jede Unterfrage Q entsteht als Resultat einer Antwort A, die eine unvollständige Antwort der vorhergehenden Frage Q ist.
- Topik-Termination: Wenn eine Frage Q hinreichend beantwortet ist, ist das ihr entsprechende Topik abgeschlossen. Der Text ist beendet, oder wendet sich der nächsten offenen Frage zu.