NLP

Discourse Analysis

Coherence

Coherence

- Examples

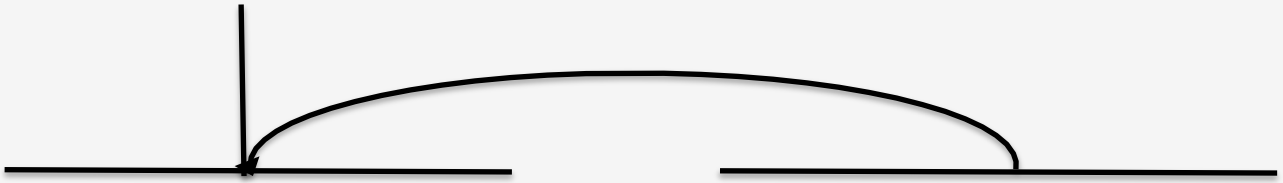
  I saw Mary in the street. She was looking for a bookstore.

  ? I saw Mary in the street. She has a cat.

  ?? I saw Mary in the street. The Pistons won.

- Rhetorical Structure Theory (Mann and Thompson 1988)

Nucleus and Satellite



The carpenter was tired.

He had been working all day.

Nucleus and Satellite

- The satellite increases the belief in the relation described in the nucleus
- Some relations have only a nucleus, others have two nuclei, yet others have one nucleus and one satellite

Coherence Relations

- Result
  - The carpenter worked all day. The new cabinet was ready in the evening.
- Explanation
  - The carpenter was tired. He had spent the entire day building a new cabinet.
- Parallel
  - The carpenter worked all day. The upholsterer took the day off.
- Elaboration
  - The carpenter built a cabinet. The cabinet had four drawers and an oversized rear panel.
- Other relations
  - Nucleus+satellite: circumstance, volitional cause, purpose, interpretation, restatement, summary
  - Multi-nuclear: sequence, contrast, joint

[Mann and Thompson 1988]

Sample Rhetorical Relations

Relation	Nucleus	Satellite
Antithesis	ideas favored by the author	ideas disfavored by the author
Background	text whose understanding is being facilitated	text for facilitating understanding
Concession	situation affirmed by author	situation which is apparently inconsistent but also affirmed by author
Elaboration	basic information	additional information
Purpose	an intended situation	the intent behind the situation
Restatement	a situation	a reexpression of the situation
Summary	text	a short summary of that text

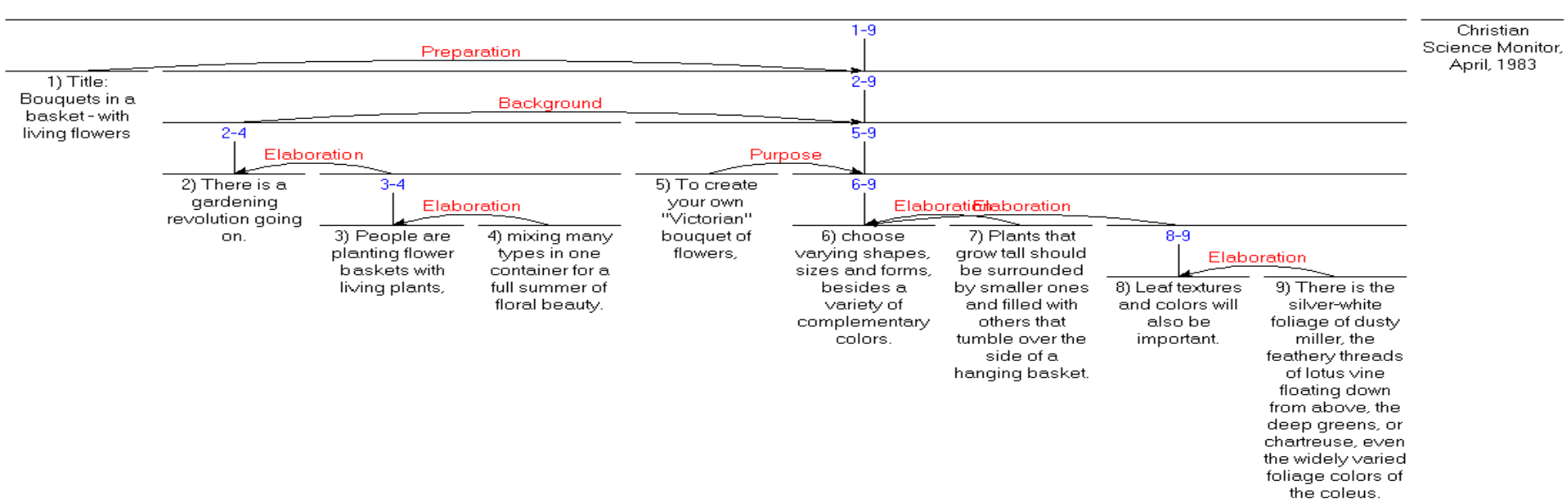
Example

1) Title: Bouquets in a basket – with living flowers
2) There is a gardening revolution going on.
3) People are planting flower baskets with living plants,
4) mixing many types in one container for a full summer of floral beauty.
5) To create your own "Victorian" bouquet of flowers,
6) choose varying shapes, sizes and forms, besides a variety of complementary colors.
7) Plants that grow tall should be surrounded by smaller ones and filled with others that tumble over the side of a hanging basket.
8) Leaf textures and colors will also be important.
9) There is the silver-white foliage of dusty miller, the feathery threads of lotus vine floating down from above, the deep greens, or chartreuse, even the widely varied foliage colors of the coleus.

Christian Science Monitor, April, 1983

from Mann/Matthiessen/Thompson

Example (cont'd)



http://www.sfu.ca/rst/

Discourse Parsing

- Four RST relations: contrast, cause-explanation-evidence, condition, elaboration + non-relation
- Up to 4M automatically labeled examples per relation
- Naïve Bayes
- Word co-occurrence features

[Marcu and Echihabi 2002]

Centering

- Goal: understand the local coherence of discourse
- Why some texts are considered more coherent
- Inference load associated with badly chosen referring expressions
- Too much focus shift makes the text hard to understand.

Centering

- Every utterance U_n has a backwards looking center Cb, which connects U_n with the previous utterance U_{n-1}.
- Every utterance also has a partially ordered set of forward looking centers C_f related to the next utterance U_{n+1}. The order depends on syntax (e.g., subject>object)
- The preferred center C_p is the highest ranking element of C_f.

Cross-document Structure (CST)

Number	Relationship type	Level	Description
1	Identity	Any	The same text appears in more than one location
2	Equivalence (paraphrasing)	S, D	Two text spans have the same information content
3	Translation	P, S	Same information content in different languages
4	Subsumption	S, D	One sentence contains more information than another
5	Contradiction	S, D	Conflicting information
6	Historical background	S	Information that puts current information in context
7	Cross-reference	P	The same entity is mentioned
8	Citation	S, D	One sentence cites another document
9	Modality	S	Qualified version of a sentence
10	Attribution	S	One sentence repeats the information of another while adding an attribution
11	Summary	S, D	Similar to Summary in RST: one sentence summarizes another

S=Sentence, P=Paragraph, D=document

Cross-document Structure (CST)

Number	Relationship type	Level	Description
12	Follow-up	S	Additional information which reflects facts that have happened since the previous account
13	Elaboration	S	Additional information that wasn't included in the last account
14	Indirect speech	S	Shift from direct to indirect speech or vite-versa
15	Refinement	S	Additional information that is
16	Agreement	S	One source expresses agreement with another
17	Judgement	S	A qualified account of a fact
18	Fulfilment	S	A prediction turned true
19	Description	S	Insertion of a description
20	Reader profile	S	Style and background-specific change
21	Contrast	S	Contrasting two accounts of facts
22	Parallel	S	Comparing two accounts of facts
23	Generalization	S	Generalization
24	Change of perspective	S,D	The same source presents a fact in a different light

Argumentative Zoning

- Aim
  - research goal of the paper
- Textual
  - statements about section structure
- Own
  - description of the authors' work (methodology, results, discussion)
- Background
  - generally accepted scientific background
- Contrast
  - comparison with other work
- Basis
  - statements of agreement with other work
- Other
  - description of other researchers' work

[Teufel and Moens 2002]

Local Entity Coherence

Table 2
Summary augmented with syntactic annotations for grid computation.

1 [The Justice Department]$_S$ is conducting an [anti-trust trial]$_O$ against [Microsoft Corp.]$_X$ with [evidence]$_X$ that [the company]$_S$ is increasingly attempting to crush [competitors]$_O$.
2 [Microsoft]$_O$ is accused of trying to forcefully buy into [markets]$_X$ where [its own products]$_S$ are not competitive enough to unseat [established brands]$_O$.
3 [The case]$_S$ revolves around [evidence]$_O$ of [Microsoft]$_S$ aggressively pressuring [Netscape]$_O$ into merging [browser software]$_O$.
4 [Microsoft]$_S$ claims [its tactics]$_S$ are commonplace and good economically.
5 [The government]$_S$ may file [a civil suit]$_O$ ruling that [conspiracy]$_S$ to curb [competition]$_O$ through [collusion]$_X$ is [a violation of the Sherman Act]$_O$.
6 [Microsoft]$_S$ continues to show [increased earnings]$_O$ despite [the trial]$_X$.

[Barzilay and Lapata 2008]

Local Entity Coherence

Table 1
A fragment of the entity grid. Noun phrases are represented by their head nouns. Grid cells correspond to grammatical roles: subjects (S), objects (O), or neither (X).

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	–	–	–	–	–	–	–	–	–	–	1
2	–	–	O	–	–	X	S	O	–	–	–	–	–	–	–	2
3	–	–	S	O	–	–	–	–	S	O	O	–	–	–	–	3
4	–	–	S	–	–	–	–	–	–	–	–	S	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	S	O	–	5
6	–	X	S	–	–	–	–	–	–	–	–	–	–	–	O	6

- 6 sentences
- S=subject, O=object, X=neither

[Barzilay and Lapata 2008]

Local Entity Coherence

Table 3
Example of a feature-vector document representation using all transitions of length two given syntactic categories S, O, X, and –.

	S S	S O	S X	S –	O S	O O	O X	O –	X S	X O	X X	X –	– S	– O	– X	– –
d_1	.01	.01	0	.08	.01	0	0	.09	0	0	0	.03	.05	.07	.03	.59
d_2	.02	.01	.01	.02	0	.07	0	.02	.14	.14	.06	.04	.03	.07	0.1	.36
d_3	.02	0	0	.03	.09	0	.09	.06	0	0	0	.05	.03	.07	.17	.39

[Barzilay and Lapata 2008]

NLP