Discourse Structure

Ling575
Discourse & Dialogue
April 13, 2011

Roadmap

- Project discussion
- Discourse structure
  - Definition & Motivation
- Discourse Models & Resources
  - Rhetorical Structure Theory (RST)
    - RST Treebank
  - Linguistic Discourse Model
  - Discourse Graphbank
  - D-LTAG & the Penn Discourse Treebank

Why Model Discourse Structure? (Theoretical)

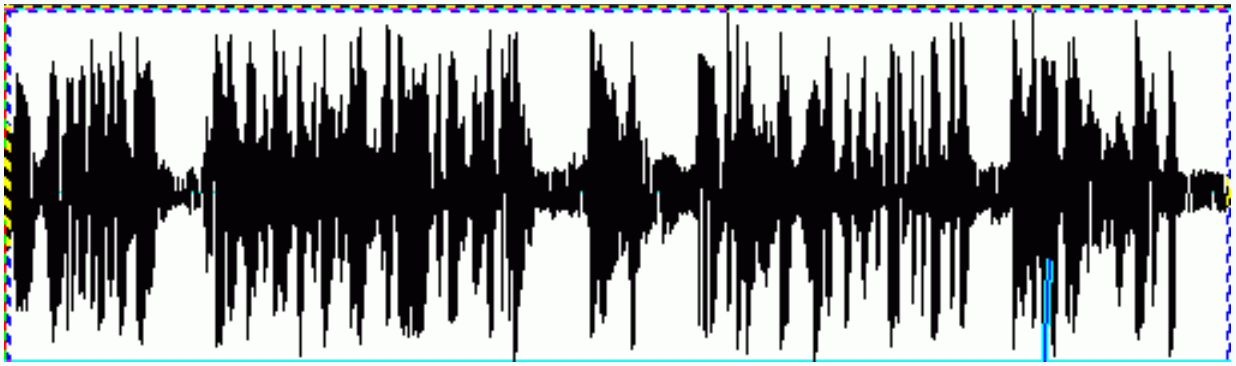
- Discourse: not just constituent utterances
  - Create joint meaning
  - Context guides interpretation of constituents
  - How????
    - What are the units?
    - How do they combine to establish meaning?
      - How can we derive structure from surface forms?
    - What makes discourse coherent vs not?
    - How do they influence reference resolution?

Why Model Discourse Structure?(Applied)

- Design better summarization, understanding
- Improve speech synthesis
  - Influenced by structure
- Develop approach for generation of discourse
- Design dialogue agents for task interaction
- Guide reference resolution

Discourse Topic Segmentation

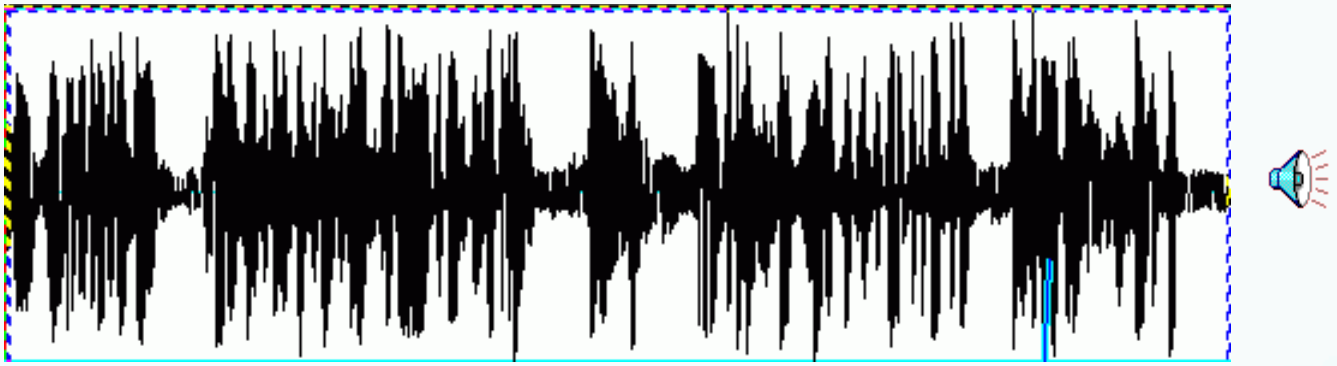
- Separate news broadcast into component stories
  - Necessary for information retrieval



On "World News Tonight" this Thursday, another bad day on stock markets, all over the world global economic anxiety. Another massacre in Kosovo, the U.S. and its allies prepare to do something about it. Very slowly. And the millennium bug, Lubbock Texas prepares for catastrophe, Banglaore in India sees only profit.

Discourse Topic Segmentation

- Separate news broadcast into component stories



On "World News Tonight" this Thursday, another bad day on stock markets, all over the world global economic anxiety. ||
Another massacre in Kosovo, the U.S. and its allies prepare to do something about it. Very slowly. ||
And the millennium bug, Lubbock Texas prepares for catastrophe, Bangalore in India sees only profit.||

Discourse Segmentation

- Basic form of discourse structure
  - Divide document into linear sequence of subtopics

- Many genres have conventional structures:

Discourse Segmentation

- Basic form of discourse structure
  - Divide document into linear sequence of subtopics

- Many genres have conventional structures:
  - Academic: Into, Hypothesis, Methods, Results, Concl.

Discourse Segmentation

- Basic form of discourse structure
  - Divide document into linear sequence of subtopics

- Many genres have conventional structures:
  - Academic: Into, Hypothesis, Methods, Results, Concl.

  - Newspapers: Headline, Byline, Lede, Elaboration

Discourse Segmentation

- Basic form of discourse structure
  - Divide document into linear sequence of subtopics

- Many genres have conventional structures:
  - Academic: Into, Hypothesis, Methods, Results, Concl.
  - Newspapers: Headline, Byline, Lede, Elaboration
  - Patient Reports: Subjective, Objective, Assessment, Plan

Discourse Segmentation

- Basic form of discourse structure
  - Divide document into linear sequence of subtopics
- Many genres have conventional structures:
  - Academic: Into, Hypothesis, Methods, Results, Concl.
  - Newspapers: Headline, Byline, Lede, Elaboration
  - Patient Reports: Subjective, Objective, Assessment, Plan
- Can guide: summarization, retrieval

Cohesion

- Use of linguistics devices to link text units
  - Lexical cohesion:
    - Link with relations between words
      - Synonymy, Hypernymy
      - *Peel, core and slice the pears and the apples. Add the fruit to the skillet.*

Cohesion

- Use of linguistics devices to link text units
  - Lexical cohesion:
    - Link with relations between words
      - Synonymy, Hypernymy
      - *Peel, core and slice the pears and the apples. Add the fruit to the skillet.*
  - Non-lexical cohesion:
    - E.g. anaphora
      - *Peel, core and slice the pears and the apples. Add them to the skillet.*

Cohesion

- Use of linguistics devices to link text units
  - Lexical cohesion:
    - Link with relations between words
      - Synonymy, Hypernymy
      - *Peel, core and slice the pears and the apples. Add the fruit to the skillet.*
  - Non-lexical cohesion:
    - E.g. anaphora
      - *Peel, core and slice the pears and the apples. Add them to the skillet.*
  - Cohesion chain establishes link through sequence of words
  - Segment boundary = dip in cohesion

Coherence

- *First Union Corp. is continuing to wrestle with severe problems. According to industry insiders at PW, their president, John R. Georgius, is planning to announce his retirement tomorrow.*
- Summary*:*
- *First Union President John R. Georgius is planning to announce his retirement tomorrow.*

- Inter-sentence coherence relations:

Coherence

- *First Union Corp. is continuing to wrestle with severe problems. According to industry insiders at PW, their president, John R. Georgius, is planning to announce his retirement tomorrow.*
- Summary*:*
- *First Union President John R. Georgius is planning to announce his retirement tomorrow.*

- Inter-sentence coherence relations:
  - Second sentence: main concept (nucleus)

Coherence

- *First Union Corp. is continuing to wrestle with severe problems. According to industry insiders at PW, their president, John R. Georgius, is planning to announce his retirement tomorrow.*
- Summary*:*
- *First Union President John R. Georgius is planning to announce his retirement tomorrow.*

- Inter-sentence coherence relations:
  - Second sentence: main concept (nucleus)
  - First sentence: subsidiary, background

Discourse
Cohesion & Coherence

- Mechanisms that holds discourse together
  - Derive meaning of discourse from components

Discourse
Cohesion & Coherence

- Mechanisms that holds discourse together
  - Derive meaning of discourse from components
- Depends on:
  - Reference relations: last class
  - Discourse relations: today

Discourse
Cohesion & Coherence

- Mechanisms that holds discourse together
  - Derive meaning of discourse from components
- Depends on:
  - Reference relations: last class
  - Discourse relations: today
- Discourse relations can be: (Moore & Pollock 1992)
  - Intentional: related to the goals, plans of participants
    - Complex issues of planning, goal, belief inference

Discourse
Cohesion & Coherence

- Mechanisms that holds discourse together
  - Derive meaning of discourse from components
- Depends on:
  - Reference relations: last class
  - Discourse relations: today
- Discourse relations can be: (Moore & Pollock 1992)
  - Intentional: related to the goals, plans of participants
    - Complex issues of planning, goal, belief inference
  - Informational: related the semantic content
    - Will focus on these

Discourse Relations

- Establish links between sentences in discourse
- Can be annotated fairly reliably
  - Yield a range of corpus resources
- Enable the applications discussed earlier

Dimensions of Discourse Structure

- Discourse relations:

Dimensions of Discourse Structure

- Discourse relations:
  - What are the relations?
    - Dominance and precedence; elaboration, sequence, etc..

Dimensions of Discourse Structure

- Discourse relations:
  - What are the relations?
    - Dominance and precedence; elaboration, sequence, etc..
  - How many relations are there?
    - 2? 10? 400?

Dimensions of Discourse Structure

- Discourse relations:
  - What are the relations?
    - Dominance and precedence; elaboration, sequence, etc..
  - How many relations are there?
    - 2? 10? 400?
  - How are relations structured?
    - Symmetric? Asymmetric?

Dimensions of Discourse Structure

- Discourse relations:
  - What are the relations?
    - Dominance and precedence; elaboration, sequence, etc..
  - How many relations are there?
    - 2? 10? 400?
  - How are relations structured?
    - Symmetric? Asymmetric
- Discourse structures:
  - What are the legal structures produced by relations?
    - Trees?, Graphs?, Other?
    - Binary? N-ary?

Dimensions of Discourse Structure

- Units:
  - What are the basic units of discourse structure?
    - Phrases?
    - Prosodic units?
    - Intention-based units?
    - Clauses?
    - Sentences?

Dimensions of Discourse Structure

- Units:
  - What are the basic units of discourse structure?
    - Phrases?
    - Prosodic units?
    - Intention-based units?
    - Clauses?
    - Sentences?
  - How are larger segments structured?
    - Overlapping?
    - Non-overlapping?

Dimensions of Discourse Structure

- Discourse relation triggers:
  - Structure:
    - Relations hold between sequentially or structurally adjacent spans

Dimensions of Discourse Structure

- Discourse relation triggers:
  - Structure:
    - Relations hold between sequentially or structurally adjacent spans
  - Lexical elements:
    - Relations are lexically cued, may act on non-adjacent elements

Dimensions of Discourse Structure

- Discourse relation triggers:
  - Structure:
    - Relations hold between sequentially or structurally adjacent spans
  - Lexical elements:
    - Relations are lexically cued, may act on non-adjacent elements
  - Lexical elements & structure: Both

Text Coherence

- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse

Text Coherence

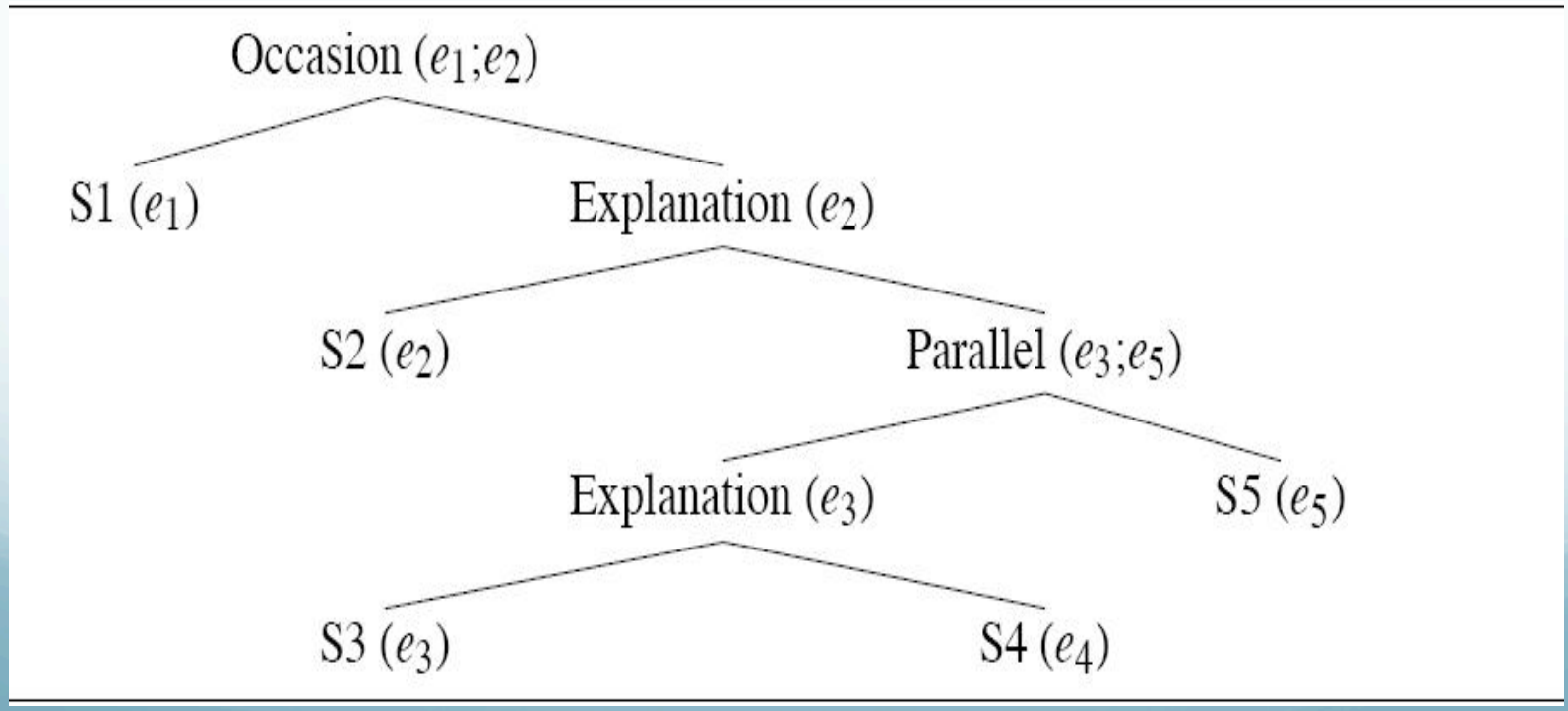
- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse
  - Examples:
    - **Result:** Infer state of S_0 cause state in S_1
      - The Tin Woodman was caught in the rain. His joints rusted.

Text Coherence

- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse
  - Examples:
    - **Result:** Infer state of S_0 cause state in S_1
      - The Tin Woodman was caught in the rain. His joints rusted.
    - **Explanation**: Infer state in S_1 causes state in S_0
      - John hid Bill's car keys. He was drunk.

Coherence Analysis

S1: John went to the bank to deposit his paycheck.
S2: He then took a train to Bill's car dealership.
S3: He needed to buy a car.
S4: The company he works now isn't near any public transportation.
S5: He also wanted to talk to Bill about their softball league.



Identifying Segments & Relations

- Key source of information:

Identifying Segments & Relations

- Key source of information:
  - Cue phrases
    - Aka discourse markers, cue words, clue words

Identifying Segments & Relations

- Key source of information:
  - Cue phrases
    - Aka discourse markers, cue words, clue words
  - Typically connectives
    - E.g. conjunctions, adverbs
  - Clue to relations, boundaries

Identifying Segments & Relations

- Key source of information:
  - Cue phrases
    - Aka discourse markers, cue words, clue words
  - Typically connectives
    - E.g. conjunctions, adverbs
  - Clue to relations, boundaries
  - Although, but, for example, however, yet, with, and….
    - John hid Bill's keys **because** he was drunk.

Cue Phrases

- Issues:
  - Ambiguity:

Cue Phrases

- Issues:
  - Ambiguity: discourse vs sentential use
    - **With** its distant orbit, Mars exhibits frigid weather.
    - We can see Mars **with** a telescope.
  - Disambiguate?

Cue Phrases

- Issues:
  - Ambiguity: discourse vs sentential use
    - **With** its distant orbit, Mars exhibits frigid weather.
    - We can see Mars **with** a telescope.
  - Disambiguate?
    - Rules (regexp): sentence-initial; comma-separated, ...
    - WSD techniques...
  - Ambiguity:

Cue Phrases

- Issues:
  - Ambiguity: discourse vs sentential use
    - **With** its distant orbit, Mars exhibits frigid weather.
    - We can see Mars **with** a telescope.
    - Disambiguate?
      - Rules (regexp): sentence-initial; comma-separated, ...
      - WSD techniques...
  - Ambiguity: cue multiple discourse relations
    - Because: CAUSE/EVIDENCE; But: CONTRAST/CONCESSION

Cue Phrases

- Last issue:
  - Insufficient:

Cue Phrases

- Last issue:
  - Insufficient:
    - Not all relations marked by cue phrases
    - Only 15-25% of relations marked by cues

Rhetorical Structure Theory

Mann & Thompson (1987)

Dimensions of RST

- Discourse relations:
  - 78 detailed informational relations; mostly asymmetric
- Discourse structures:
  - Trees: predominantly binary, some n-ary (schemas)
- Discourse units:
  - Clauses
- Discourse Segments:
  - Non-overlapping
- Discourse Relation Triggers:
  - Structure

Components of RST

- Schemas:
  - Grammar of legal relations between text spans
  - Define possible RST text structures
    - Most common: N + S, others involve two or more nuclei

Components of RST

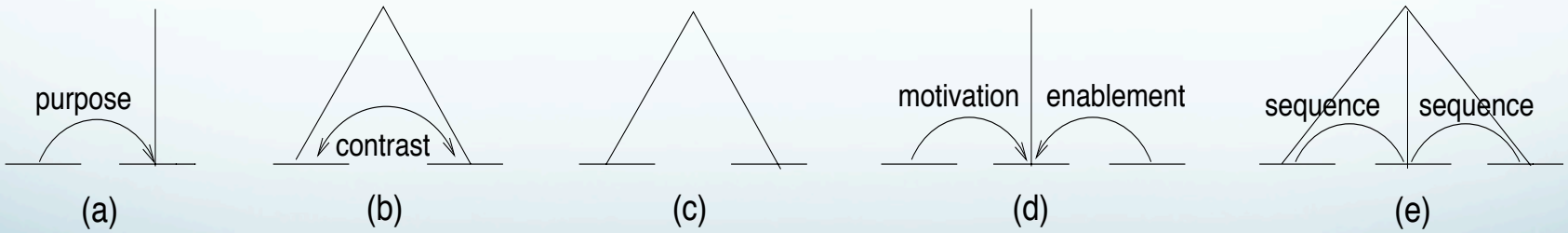
- Schemas:
  - Grammar of legal relations between text spans
  - Define possible RST text structures
    - Most common: N + S, others involve two or more nuclei
- Relations:
  - Hold b/t two text spans, nucleus and satellite
    - Constraints on each, between
    - Effect: why the author wrote this

Components of RST

- Schemas:
  - Grammar of legal relations between text spans
  - Define possible RST text structures
    - Most common: N + S, others involve two or more nuclei
- Relations:
  - Hold b/t two text spans, nucleus and satellite
    - Constraints on each, between
    - Effect: why the author wrote this
- Structures:
  - Using clause units, complete, connected, unique, adjacent

Schemas

- Schemas differ in:
  - A/Symmetry of relations
  - Brancing (arity) of relations
  - Relations between sisters



RST Relations

- Core of RST
  - RST analysis requires building tree of relations
  - Circumstance, Solutionhood, Elaboration. Background, Enablement, Motivation, Evidence, Justify, Vol. Cause, Non-Vol. Cause, Vol. Result, Non-Vol. Result, Purpose, Antithesis, Concession, Condition, Otherwise, Interpretation, Evaluation, Restatement, Summary, Sequence, Contrast

Nuclearity

- Many relations between pairs asymmetrical
  - One is incomprehensible without other
  - One is more substitutable, more important to W

Nuclearity

- Many relations between pairs asymmetrical
  - One is incomprehensible without other
  - One is more substitutable, more important to W
- Deletion of all nuclei creates gibberish
  - Deletion of all satellites is just terse, rough

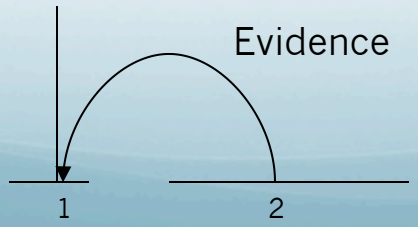
Nuclearity

- Many relations between pairs asymmetrical
  - One is incomprehensible without other
  - One is more substitutable, more important to W
- Deletion of all nuclei creates gibberish
  - Deletion of all satellites is just terse, rough
- Demonstrates role in coherence

RST Relations

- Evidence
  - Effect: Evidence (Satellite) increases R's belief in Nucleus
    - The program really works. (N)
    - I entered all my info and it matched my results. (S)

Relation Name:	Evidence
Constraints on N:	R might not believe N to a degree satisfactory to W
Constraints on S:	R believes S or will find it credible
Constraints on N+S:	R's comprehending S increases R's belief of N
Effects:	R's belief of N is increased



RST Relations

- Justify
  - Effect: Justify (Satellite) increases R's willingness to accepts W's authority to say Nucleus
    - The next music day is September 1.(N)
    - I'll post more details shortly. (S)

RST Relations

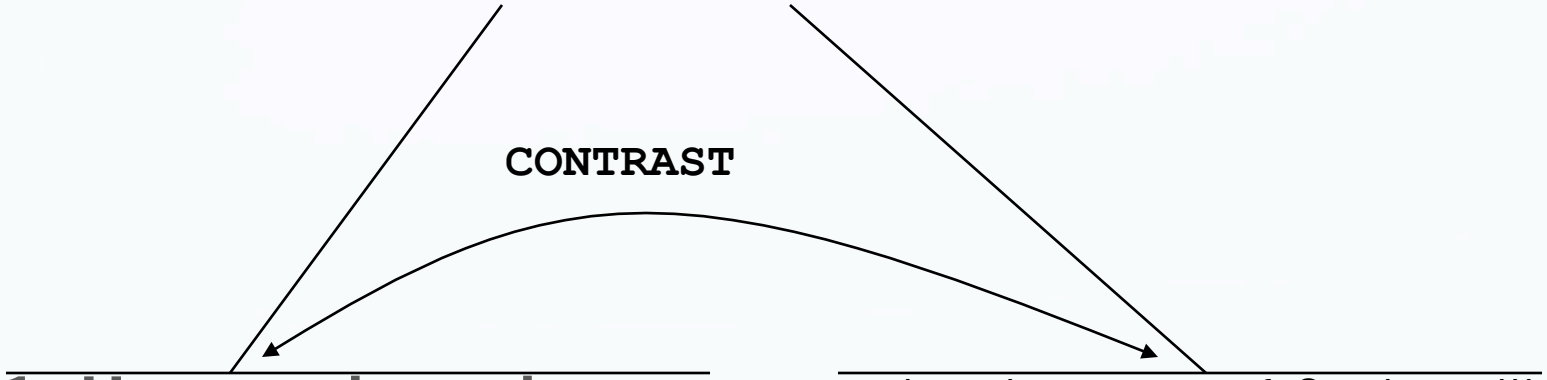
- Concession:
  - Effect: By acknowledging incompatibility between N and S, increase Rs positive regard of N
    - Often signaled by “although”
      - Dioxin: Concerns about its health effects may be misplaced.(N1) Although it is toxic to certain animals (S), evidence is lacking that it has any long-tern effect on human beings.(N2)

RST Relations

- Concession:
  - Effect: By acknowledging incompatibility between N and S, increase Rs positive regard of N
    - Often signaled by “although”
      - Dioxin: Concerns about its health effects may be misplaced.(N1) Although it is toxic to certain animals (S), evidence is lacking that it has any long-tern effect on human beings.(N2)
- Elaboration:
  - Effect: By adding detail, S increases Rs belief in N
- Etc

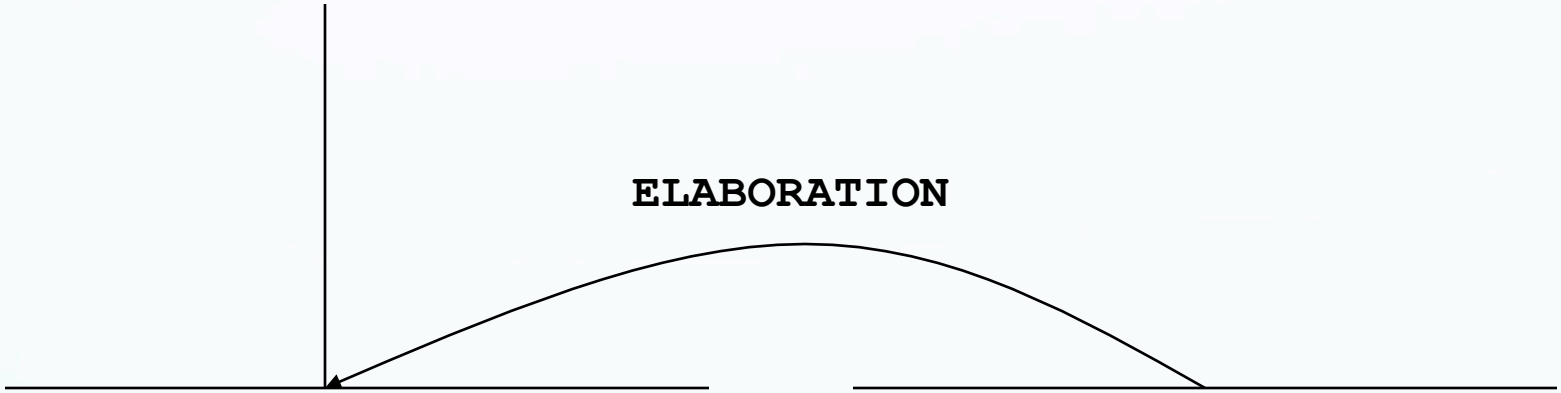
RST-relation example (1)

Symmetric (multiple nuclei) Relation:



RST-relation example (2)

Asymmetric (nucleus-satellite) Relation:



1. In other parts of Spain, still hot, dry weather with temperatures up to 35 degrees Celcius.

2. In Cadiz, the thermometer might rise as high as 40 degrees.

RST Annotation Procedure

- Step 1: Annotated elementary discourse units (EDUs)

RST Annotation Procedure

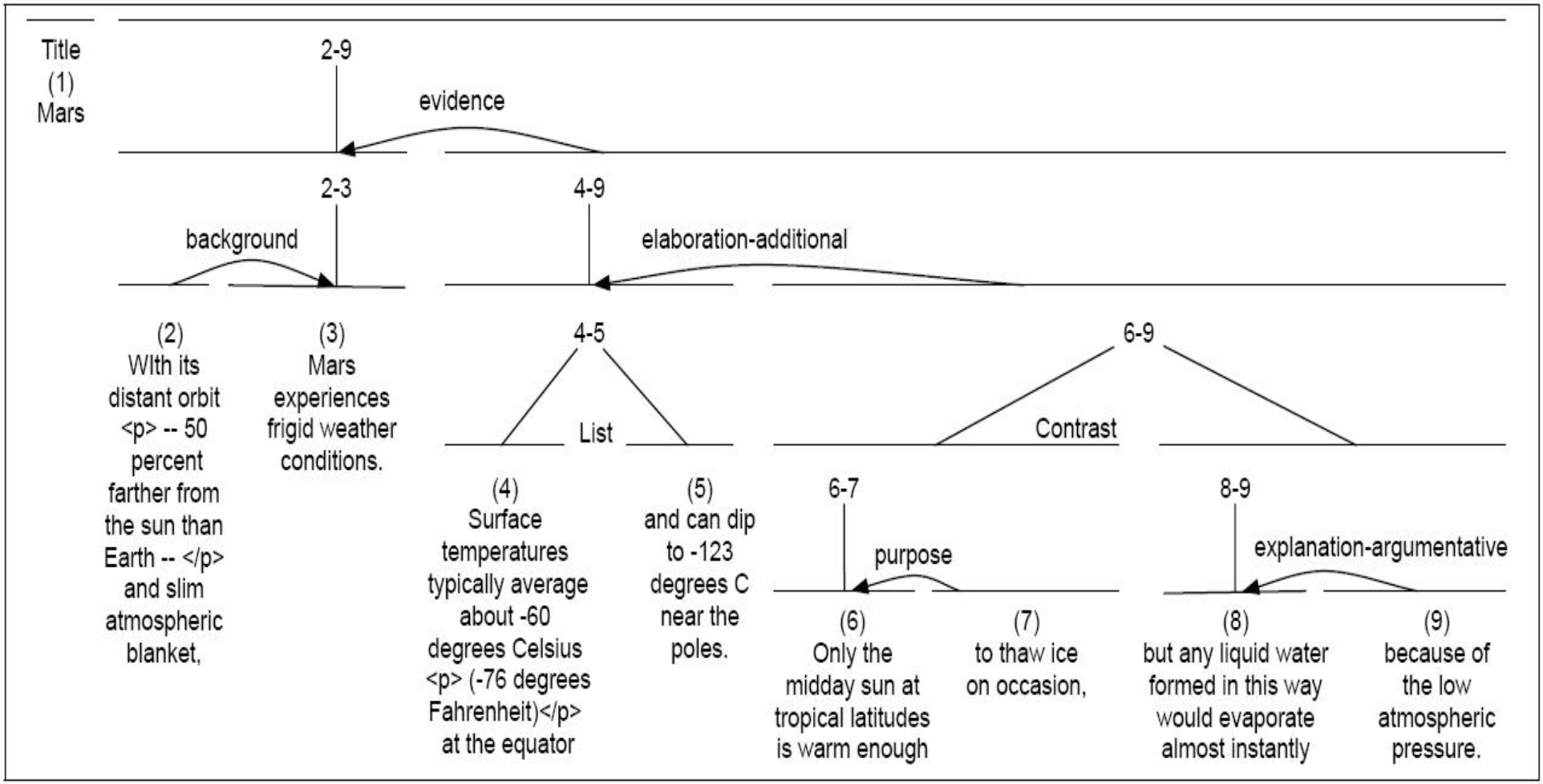
- Step 1: Annotated elementary discourse units (EDUs)
- Step 2: Connect units, tag as N(ucleus) or S(atellite)

RST Annotation Procedure

- Step 1: Annotated elementary discourse units (EDUs)
- Step 2: Connect units, tag as N(ucleus) or S(atellite)
- Step 3: Assign relation

RST Annotation Procedure

- Step 1: Annotated elementary discourse units (EDUs)
- Step 2: Connect units, tag as N(ucleus) or S(atellite)
- Step 3: Assign relation
- Finished when complete, singly-rooted, spanning tree
- RST Discourse Treebank (Carlson et al, LDC)


Title
(1)
Mars
2-9
evidence
2-3
background
4-9
elaboration-additional
(2)
With its distant orbit -- 50 percent farther from the sun than Earth -- and slim atmospheric blanket,
(3)
Mars experiences frigid weather conditions.
4-5
List
6-9
Contrast
(4)
Surface temperatures typically average about -60 degrees Celsius (-76 degrees Fahrenheit) at the equator
(5)
and can dip to -123 degrees C near the poles.
6-7
purpose
8-9
explanation-argumentative
(6)
Only the midday sun at tropical latitudes is warm enough
(7)
to thaw ice on occasion,
(8)
but any liquid water formed in this way would evaporate almost instantly
(9)
because of the low atmospheric pressure.

Linguistic Discourse Model

LDM (Polanyi 1988; Polanyi et al 2004)

Dimensions of LDM

- Discourse relations:
  - Viewed outside of theory: discourse interpretation
- Discourse structures:
  - Trees: predominantly binary, some n-ary : context free rules
- Discourse units:
  - Clauses (event and infinitive),
  - Subordinating/co-ordinating conjunctions
- Discourse Segments:
  - Non-overlapping
- Discourse Relation Triggers:
  - Structure (vacuously)

Discourse Structure Rules

- Discourse coordination: lists, narratives
  - N-ary branching
  - Semantic compositions (SC) rule:
    - Parent is information common to its children

Discourse Structure Rules

- Discourse coordination: lists, narratives
  - N-ary branching
  - Semantic compositions (SC) rule:
    - Parent is information common to its children
- Discourse subordination:
  - Binary branching; subordination child elaborates dominant
  - SC rule: Parent receives interpretation of dominant child

Discourse Structure Rules

- Discourse coordination: lists, narratives
  - N-ary branching
  - Semantic compositions (SC) rule:
    - Parent is information common to its children
- Discourse subordination:
  - Binary branching; subordination child elaborates dominant
  - SC rule: Parent receives interpretation of dominant child
- Logical/rhetorical relation:
  - N-ary branching: Relation holds among children
  - SC rule: Parent inherits interpretation of rel'n over children

LDM Annotation

- Identify basic discourse units:
  - Event clauses, infinitive clauses, sub/co-ordinating conj

Examples from Joshi, Prasad, Webber, Discourse Annotation Tutorial 2006

LDM Annotation

- Identify basic discourse units:
  - Event clauses, infinitive clauses, sub/co-ordinating conj
    - [Though] [these methods are applicable to general media,] [we concentrate here on audio.]

Examples from Joshi, Prasad, Webber, Discourse Annotation Tutorial 2006

LDM Annotation

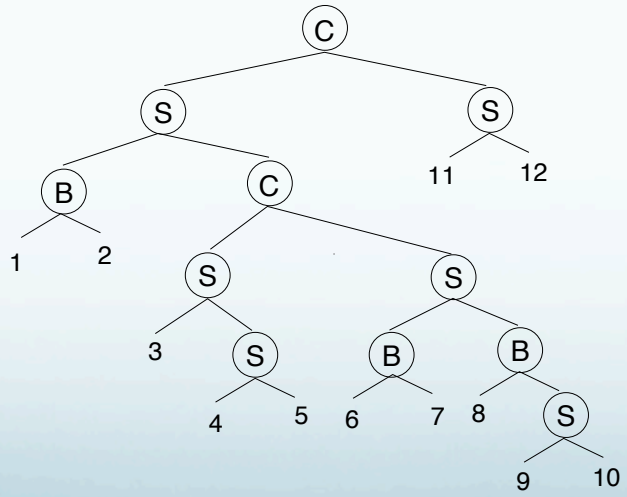
- Identify basic discourse units:
  - Event clauses, infinitive clauses, sub/co-ordinating conj
    - [Though] [these methods are applicable to general media,] [we concentrate here on audio.]

- Incrementally attach units to tree, start to end
  - Identify node to attach next unit as right child
  - Identify attachment rule: coord, subord, relation

Examples from Joshi, Prasad, Webber, Discourse Annotation Tutorial 2006

Example LDM Annotation

- [$_1$ Whatever advances we may have seen in knowledge management,] [$_2$ knowledge sharing remains a major issue.] [$_3$ A key problem is] [$_4$ that documents only assume value] [$_5$ when we reflect upon their content.] [$_6$ Ultimately,] [$_7$ the solution to this problem will probably reside in the documents themselves.] [$_8$ In other words,] [$_9$ the real solution to the problem of knowledge sharing involves authoring,] [$_{10}$ rather than document management.] [$_{11}$ This paper is a discussion of several new approaches to authoring and opportunities for new technologies] [$_{12}$ to support those approaches.]



C: Discourse coordination

S: Discourse subordination

B: Binary construction

Discourse Graphbank

Wolf & Gibson 2005

Dimensions of DG

- Discourse relations:
  - 11 relations: cause-effect, elaboration, condition, etc
  - Symmetric and Asymmetric; binary or n-ary
- Discourse structures:
  - Arbitrary Graphs
- Discourse units:
  - Clauses
- Discourse Segments:
  - Basic units - Non-overlapping, or groups of segments
- Discourse Relation Triggers:
  - Structure and Lexical

Annotation in DG

- Identify basic segments:
  - Clauses by punctuation, or conjunctions

➢ The economy,

➢ according to some analysts,

is expected to improve by early next year.

[Wolf & Gibson 2005, p.255]

Annotation in DG

- Create *groupings* of segments, if they are:
  - Also in quotations
  - In a common attribution
  - In the same sentence
  - On a common topic
-

Annotation in DG

- Create *groupings* of segments, if they are:
  - Also in quotations
  - In a common attribution
  - In the same sentence
  - On a common topic
- 1. $_{a}$ [Difficulties have arisen] $_{b}$ [in enacting the accord for the independence of Namibia]
- 2. for which SWAPO has fought many years,

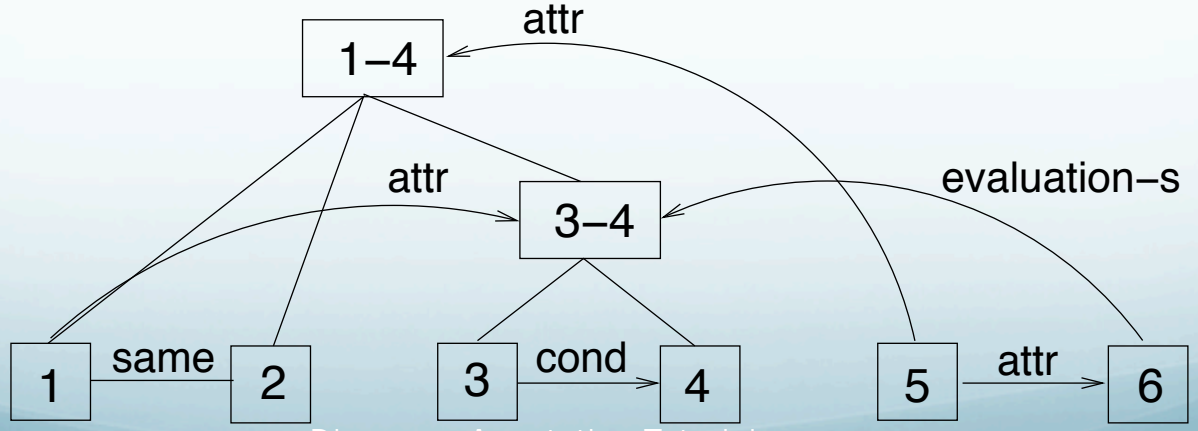
Annotation in DG

- Proceed through discourse from beginning to end:
  - For each segment or grouping
    - For each previous segment or grouping
      - Check if a relation holds
      - If a relation holds, create a node that is parent to both
  - Note: Allows crossing dependencies, multiple parents

Example Discourse GraphBank Analysis

- (1) The administration should now state
  (2) that
  (3) if the February election is voided by the Sandinistas
  (4) they should call for military aid,
  (5) said former Assistant Secretary of State Elliot Abrams.
  (6) In these circumstances, I think they'd win.

[Wolf and Gibson, 2005, Example 26]



Observations

- This is really, really complicated
- Also, debated
  - http://itre.cis.upenn.edu/~myl/languagelog/archives/000541.html
- Available as a corpus from the LDC

Models of Discourse Informational Structure

- Create structural analysis of discourse
  - Based on information relations
  - Composed of elementary units
  - Linking pairs or groups of units
  - Some hierarchical structure
  - Exploit cue words

Models of Discourse Structure

- Differ in small and large ways:
- Smaller:
  - Slight differences in minimal units
  - Similar branching structure (binary, nary)
- Moderate:
  - Differences in relation inventory
  - Grouping of units
- Major:
  - Fundamental structure: Tree vs graph

Similar Challenges

- Reliable segmentation of units
- Consistent linkage of constituents
- Determination of correct relations
  - Especially in absence of explicit cue words
- Automatic recognition – next time!