

# Discourse: Structure

Ling571  
Deep Processing Techniques for NLP  
March 7, 2011

# Roadmap

- Reference Resolution Wrap-up
- Discourse Structure
  - Motivation
    - Theoretical and applied
  - Linear Discourse Segmentation
  - Text Coherence
    - Rhetorical Structure Theory
    - Discourse Parsing

# Reference Resolution Algorithms

- Hobbs algorithm:
  - Syntax-based: binding theory+recency, role
- Many other alternative strategies:
  - Linguistically informed, saliency hierarchy
    - Centering Theory
  - Machine learning approaches:
    - Supervised: Maxent
    - Unsupervised: Clustering
  - Heuristic, high precision:
    - Cogniac

# Reference Resolution: Agreements

- Knowledge-based
  - Deep analysis: full parsing, semantic analysis
  - Enforce syntactic/semantic constraints
  - Preferences:
    - Recency
    - Grammatical Role Parallelism (ex. Hobbs)
    - Role ranking
    - Frequency of mention
- Local reference resolution
- Little/No world knowledge
- Similar levels of effectiveness

# Questions

- 80% on (clean) text. What about...
  - Conversational speech?
    - Ill-formed, disfluent
  - Dialogue?
    - Multiple speakers introduce referents
  - Multimodal communication?
    - How else can entities be evoked?
    - Are all equally salient?

# More Questions

- 80% on (clean) (English) text: What about..
  - Other languages?
    - Salience hierarchies the same
      - Other factors
    - Syntactic constraints?
      - E.g. reflexives in Chinese, Korean,...
    - Zero anaphora?
      - How do you resolve a pronoun if you can't find it?

# Reference Resolution: Extensions

- Cross-document co-reference
  - (Baldwin & Bagga 1998)
  - Break “the document boundary”
  - Question: “John Smith” in A = “John Smith” in B?
  - Approach:
    - Integrate:
      - Within-document co-reference
    - with
      - Vector Space Model similarity

# Cross-document Co-reference

- Run within-document co-reference (CAMP)
  - Produce chains of all terms used to refer to entity
- Extract all sentences with reference to entity
  - Pseudo per-entity summary for each document
- Use Vector Space Model (VSM) distance to compute similarity between summaries

# Cross-document Co-reference

- Experiments:
  - 197 NYT articles referring to “John Smith”
    - 35 different people, 24: 1 article each
  - With CAMP: Precision 92%; Recall 78%
  - Without CAMP: Precision 90%; Recall 76%
  - Pure Named Entity: Precision 23%; Recall 100%

# Conclusions

- Co-reference establishes coherence
- Reference resolution depends on coherence
- Variety of approaches:
  - Syntactic constraints, Recency, Frequency, Role
- Similar effectiveness - different requirements
- Co-reference can enable summarization within and across documents (and languages!)

# Why Model Discourse Structure? (Theoretical)

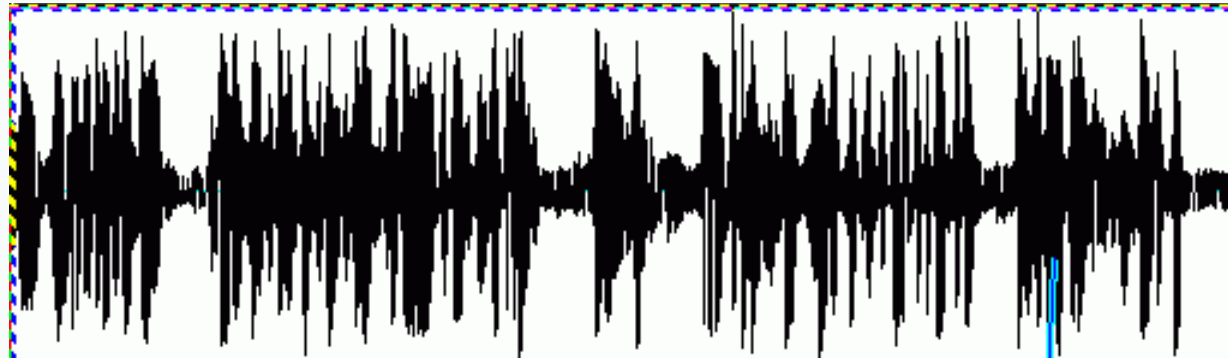
- Discourse: not just constituent utterances
  - Create joint meaning
  - Context guides interpretation of constituents
  - How????
  - What are the units?
  - How do they combine to establish meaning?
    - How can we derive structure from surface forms?
  - What makes discourse coherent vs not?
  -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

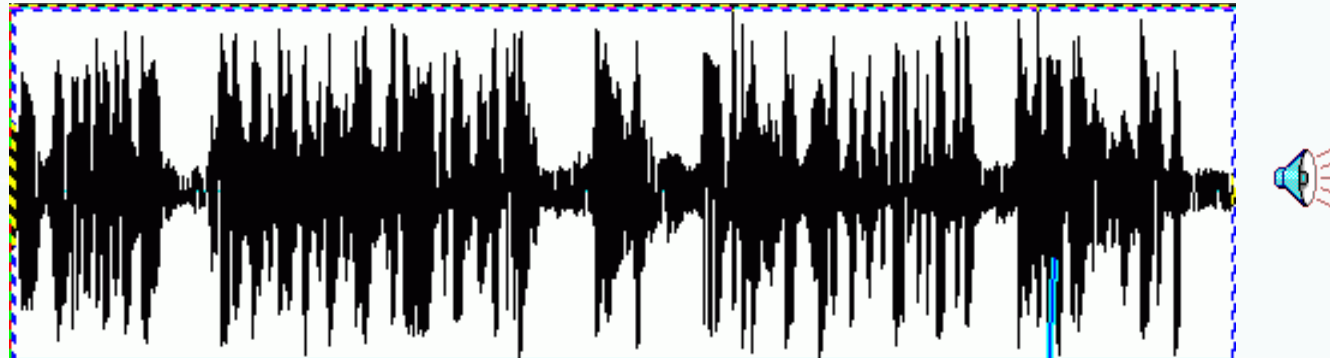
- Separate news broadcast into component stories



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 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:
  - Academic: Intro, 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.*
  - 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

# TextTiling (Hearst '97)

- Lexical cohesion-based segmentation
  - Boundaries at dips in cohesion score
  - Tokenization, Lexical cohesion score, Boundary ID
- Tokenization
  - Units?
    - White-space delimited words
    - Stopped
    - Stemmed
    - 20 words = 1 pseudo sentence

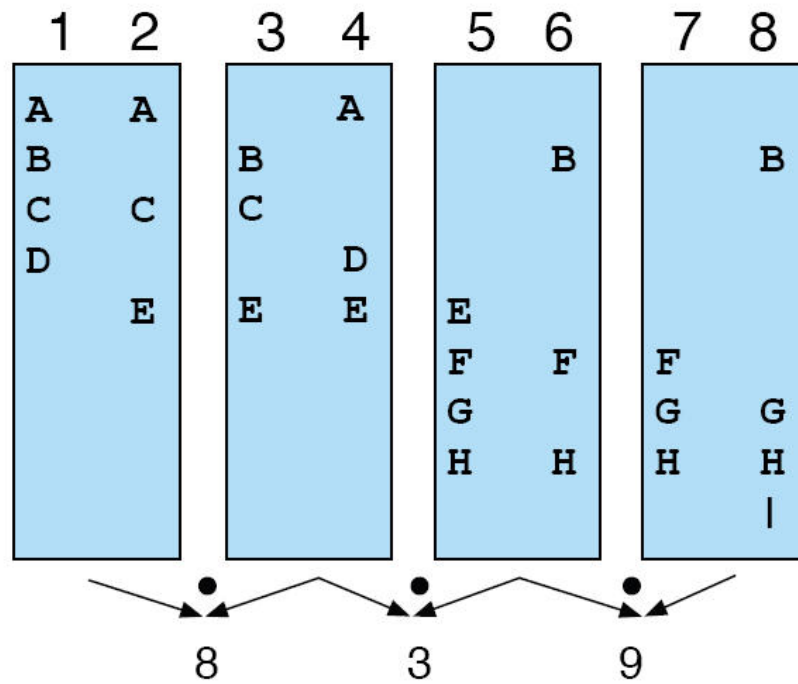
# Lexical Cohesion Score

- Similarity between spans of text
  - b = 'Block' of 10 pseudo-sentences before gap
  - a = 'Block' of 10 pseudo-sentences after gap
  - How do we compute similarity?
    - Vectors and cosine similarity (again!)

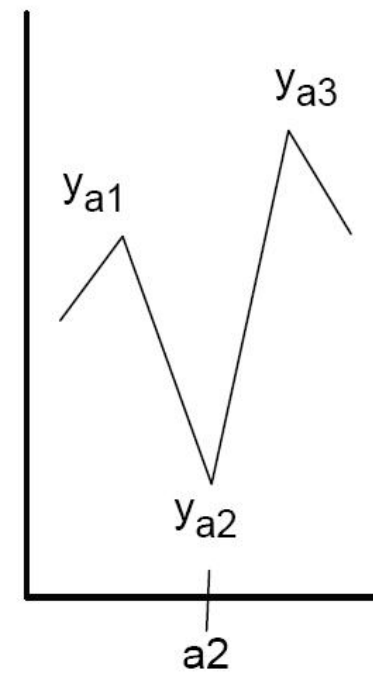
$$sim_{\cosine}(\vec{b}, \vec{a}) = \frac{\vec{b} \cdot \vec{a}}{|\vec{b}| |\vec{a}|} = \frac{\sum_{i=1}^N b_i \times a_i}{\sqrt{\sum_{i=1}^N b_i^2} \sqrt{\sum_{i=1}^N a_i^2}}$$

# Segmentation

- Depth score:
  - Difference between position and adjacent peaks
  - E.g.,  $(y_{a1} - y_{a2}) + (y_{a3} - y_{a2})$



(a)



(b)

# Evaluation

- Contrast with reader judgments
  - Alternatively with author or task-based
- 7 readers, 13 articles: “Mark topic change”
  - If 3 agree, considered a boundary
- Run algorithm – align with nearest paragraph
  - Contrast with random assignment at frequency
- Auto: 0.66, 0.61; Human: 0.81, 0.71
  - Random: 0.44, 0.42

# Discussion

- Overall: Auto much better than random
  - Often “near miss” – within one paragraph
    - 0.83,0.78
- Issues: Summary material
  - Often not similar to adjacent paras
- Similarity measures
  - Is raw tf the best we can do?
  - Other cues??
- Other experiments with TextTiling perform less well – Why?

# 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

# Early Discourse Models

- Schemas & Plans
  - (McKeown, Reichman, Litman & Allen)
  - Task/Situation model = discourse model
    - Specific->General: “restaurant” -> AI planning
- Topic/Focus Theories (Grosz 76, Sidner 76)
  - Reference structure = discourse structure
- Speech Act
  - single utt intentions vs extended 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.
    - **Explanation:** Infer state in  $S_1$  causes state in  $S_0$ 
      - John hid Bill's car keys. He was drunk.
    - **Elaboration:** Infer same prop. from  $S_0$  and  $S_1$ .
      - Dorothy was from Kansas. She lived in the great Kansas prairie.
  - Pair of locally coherent clauses: discourse segment

# Coherence Analysis

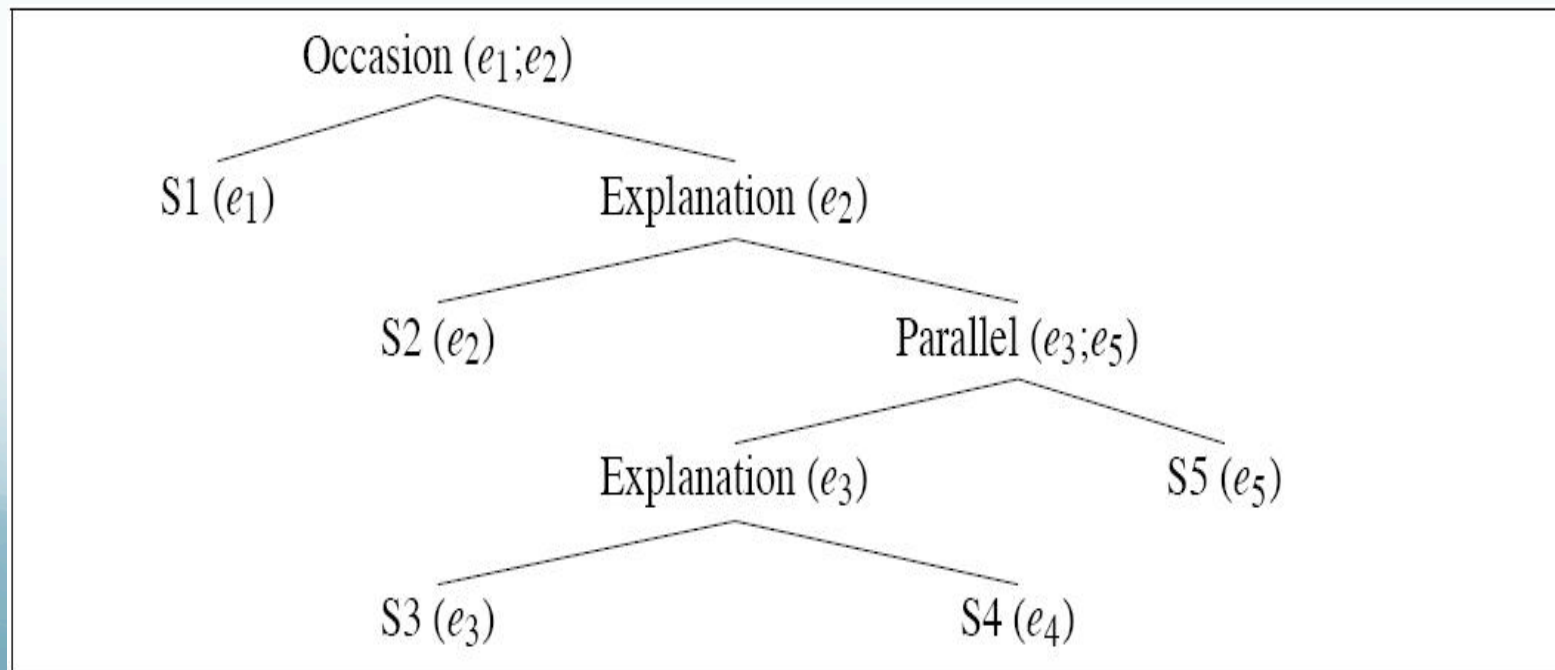
S1: John went to the bank to deposit his paycheck.

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.



# Rhetorical Structure Theory

- Mann & Thompson (1987)
- Goal: Identify hierarchical structure of text
  - Cover wide range of TEXT types
    - Language contrasts
  - Relational propositions (intentions)
- Derives from functional relations b/t clauses

# Components of RST

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

# 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

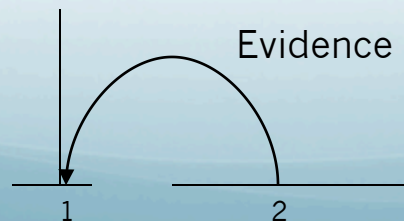
# Nuclearity

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

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



# RST Relations

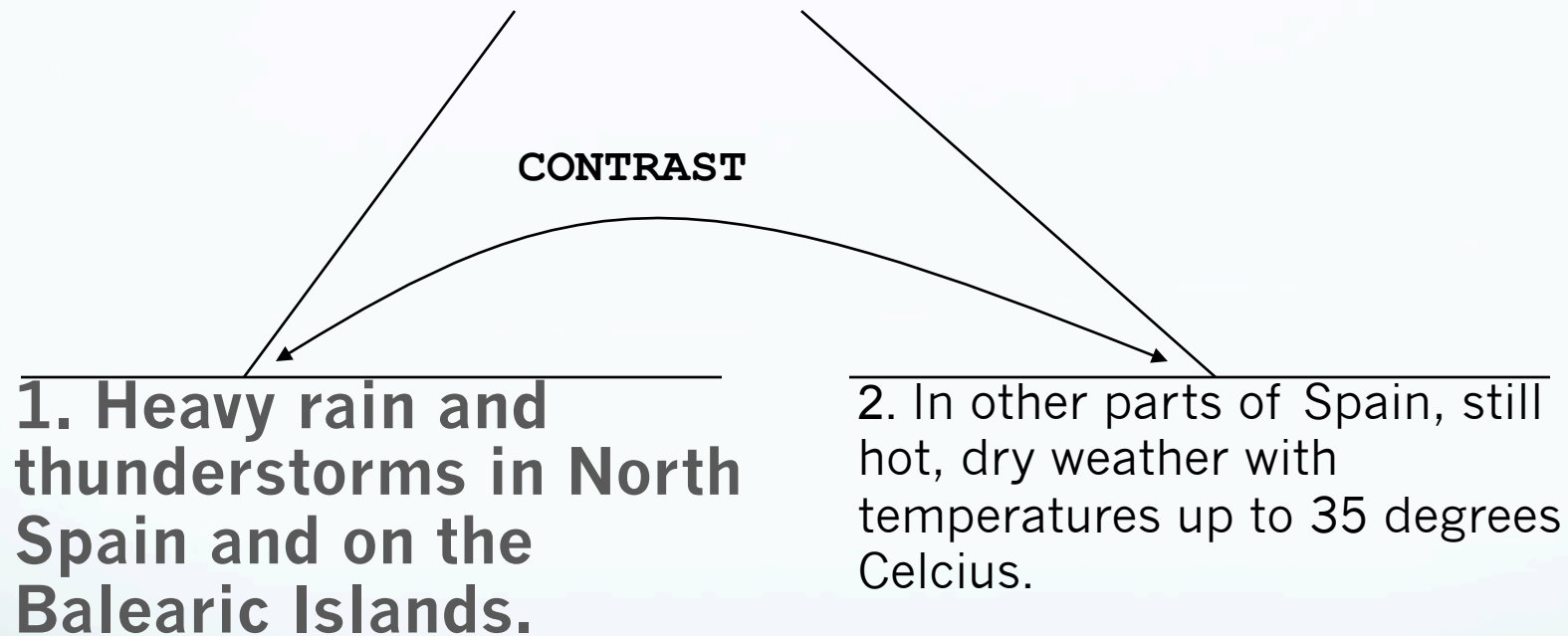
- Effect: Justify (Satellite) increases R' s willingness to accepts W' s authority to say Nucleus
  - The next music day is September 1.(N)
  - I' ll post more details shortly. (S)

# 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-term effect on human beings.(N2)
- Elaboration:
  - Effect: By adding detail, S increases Rs belief in N

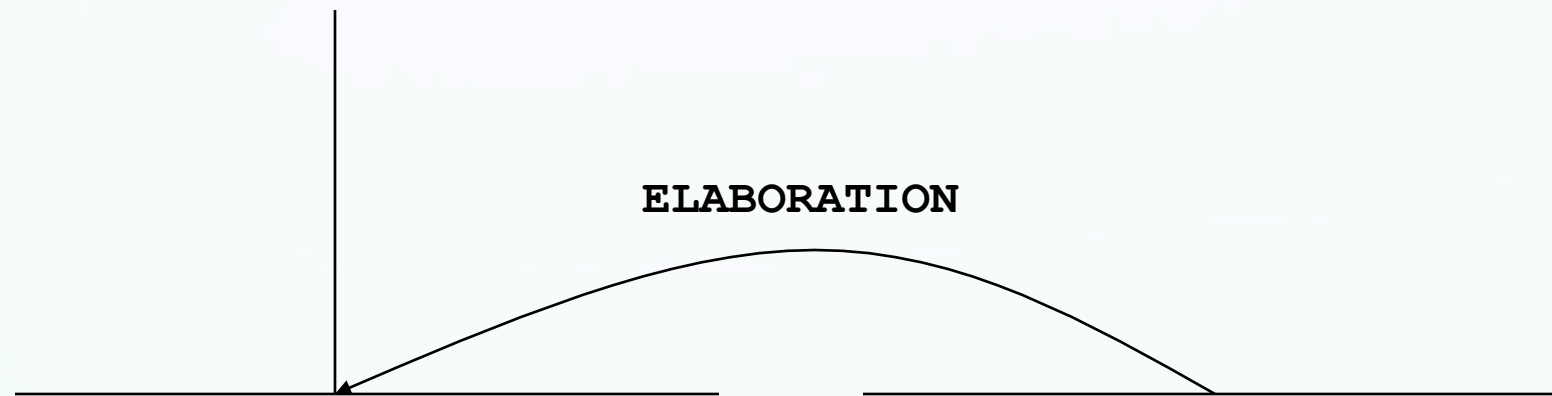
# RST-relation example (1)

Symmetric (multiple nuclei) Relation:



# RST-relation example (2)

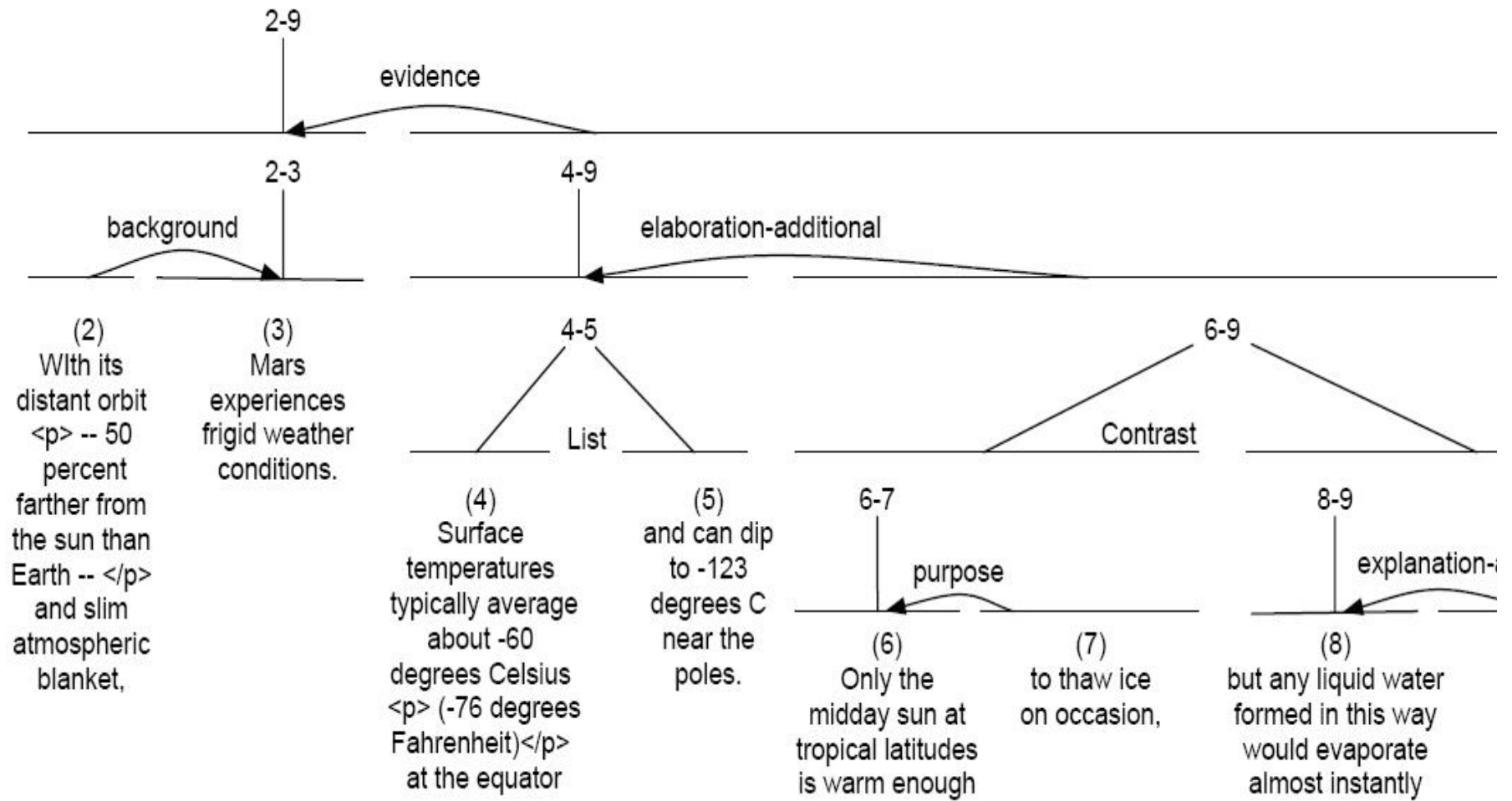
Asymmetric (nucleus-satellite) Relation:



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.**

Title  
(1)  
Mars



# RST Parsing (Marcu 1999)

- Learn and apply classifiers for
  - Segmentation and parsing of discourse
- Assign coherence relations between spans
- Create a representation over whole text => parse
- Discourse structure
  - RST trees
    - Fine-grained, hierarchical structure
      - Clause-based units
      - Inter-clausal relations: 71 relations: 17 clusters
    - Mix of intention, informational relations

# Corpus-based Approach

- Training & testing on 90 RST trees
  - Texts from MUC, Brown (science), WSJ (news)
- Annotations:
  - Identify “edu”s – elementary discourse units
    - Clause-like units – key relation
    - Parentheticals – could delete with no effect
  - Identify nucleus-satellite status
  - Identify relation that holds – i.e. elab, contrast...

# 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: 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:
    - Not all relations marked by cue phrases
    - Only 15-25% of relations marked by cues

# Learning Discourse Parsing

- Train classifiers for:
  - EDU segmentation
  - Coherence relation assignment
  - Discourse structure assignment
    - Shift-reduce parser transitions
- Use range of features:
  - Cue phrases
  - Lexical/punctuation in context
  - Syntactic parses

# Evaluation

- Segmentation:
  - Good: 96%
    - Better than frequency or punctuation baseline
- Discourse structure:
  - Okay: 61% span, relation structure
- Relation identification: poor

# Discussion

- Noise in segmentation degrades parsing
  - Poor segmentation -> poor parsing
- Need sufficient training data
  - Subset (27) texts insufficient
    - More variable data better than less but similar data
- Constituency and N/S status good
  - Relation far below human

# Issues

- Goal: Single tree-shaped analysis of all text
  - Difficult to achieve
    - Significant ambiguity
    - Significant disagreement among labelers
    - Relation recognition is difficult
      - Some clear “signals”, i.e. although
      - Not mandatory, only 25%