

CS4705: Natural Language Processing

Discourse: Structure and Coherence
Kathy McKeown

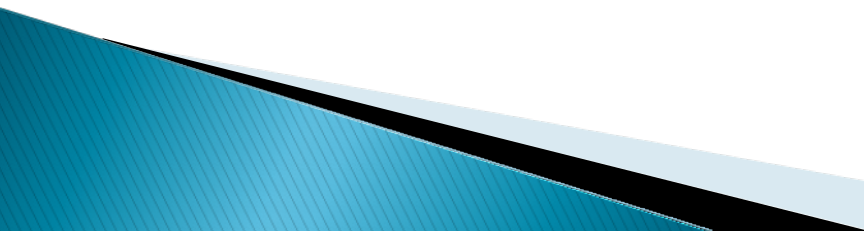
Thanks to Dan Jurafsky, Diane Litman, Andy Kehler, Jim Martin




Homework questions?

▶ HW4:

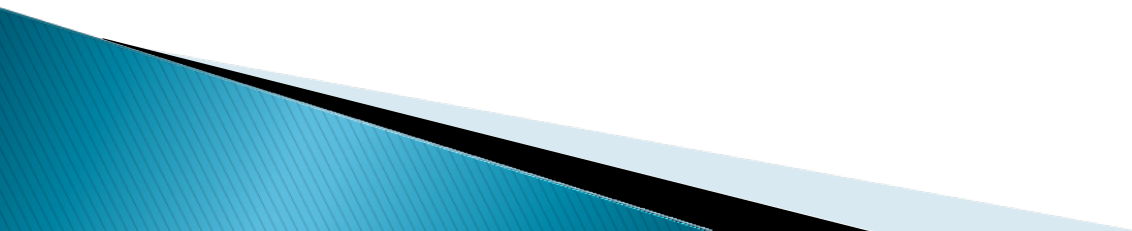
- For HW3 you experiment with different features (at least 3) and different learning algorithms (at least 2) but you turn in your best model
- For HW4 you are asked to write up your findings from your experiments in HW3
 - What features did you experiment with and why?
 - How did each individual feature contribute to success vs. the combination? (show the evaluation results)
 - Why do you think the features worked this way?
 - How do the different machine learning algorithms compare?
 - What features did you try but throw out?
 - You should provide charts with numbers both comparison of feature impact and learning algorithm impact

- Evaluation: How would your system fare if you used the pyramid method rather than precision and recall? Show how this would work on one of the test document sets. That is, for the first 3 summary sentences in the four human models, show the SCUs, the weights for each SCU, and which of the SCUs your system got.
 - If you could do just one thing to improve your system, what would that be? Show an example of where things went wrong and say whether you think there is any NL technology that could help you address this.
 - Your paper should be between 5–7 pages.
 - Professor McKeown will grade the paper
- 

Class Wrap-Up

- ▶ Final exam: December 17th, 1:10–4:00 here
 - ▶ Pick up your work: midterms, past assignments from me in my office hours or after class
 - ▶ HW2 grades will be returned the Thurs after Thanksgiving
 - ▶ Interim class participation grades will be posted on courseworks the week after Thanksgiving
- 

What is a coherent/cohesive discourse?



Summarization, question answering, information extraction, generation

- ▶ Which are more useful where?
 - Discourse structure: subtopics
 - Discourse coherence: relations between sentences
 - Discourse structure: rhetorical relations

Outline

- ▶ Discourse Structure
 - Textiling
- ▶ Coherence
 - Hobbs coherence relations
 - Rhetorical Structure Theory

Part I: Discourse Structure

- ▶ Conventional structures for different genres
 - Academic articles:
 - Abstract, Introduction, Methodology, Results, Conclusion
 - Newspaper story:
 - inverted pyramid structure (lead followed by expansion)

Discourse Segmentation

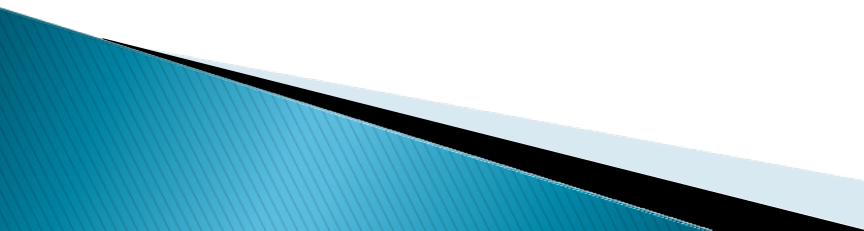
- ▶ Simpler task
 - Discourse segmentation
 - Separating document into linear sequence of subtopics

Unsupervised Discourse Segmentation

- ▶ Hearst (1997): 21–pgraph science news article called “Stargazers”
- ▶ Goal: produce the following subtopic

1-3	Intro - the search for life in space
4-5	The moon’s chemical composition
6-8	How early earth-moon proximity shaped the moon
9-12	How the moon helped life evolve on earth
13	Improbability of the earth-moon system
14-16	Binary/trinary star systems make life unlikely
17-18	The low probability of nonbinary/trinary systems
19-20	Properties of earth’s sun that facilitate life
21	Summary

Applications

- ▶ Information retrieval:
 - automatically segmenting a TV news broadcast or a long news story into sequence of stories
 - ▶ Text summarization: ?
 - ▶ Information extraction:
 - Extract info from inside a single discourse segment
 - ▶ Question Answering?
- 

TileBars: Term Distribution in Information Access

Min Overlap Span

	Min Hits	Min Distribution (%)
Term Set 1: patient medicine medical		
Term Set 2: test scan cure diagnosis		
Term Set 3: software program		

Documents Within Constraints

- Massive storage: big disks for the iMac. (includes related article summarizin
- "TextPert. (Software Review) (evaluation)"
- "PC printers gain 3287 power with protocol converters. (Hardware
- "VA automation means faster admissions. (US Department of Veterans Affair
- "Better ADP could cut VA delays 40%, officials say. (automatic data process
- "Card smarts. (smart cards)"
- "It's hard to ghostbust a network with current diagnostic tools managers say
- "Army tests prototype battlefield information system.&O "
- "Lack of imagination stalls optical-disk applications.&O "
- "The electric cadaver. (computerized anatomy lessons and digital dissection
- "Interesting new things. (monitoring and testing equipment) (buyers g
- "MegaDrive 20 is reliable, has excellent software: 'nice adjunct to standard h

Key intuition: cohesion

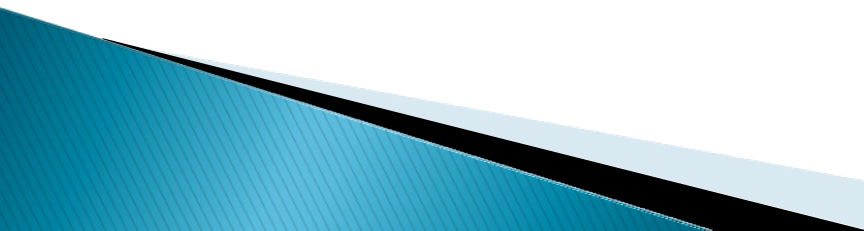
- ▶ Halliday and Hasan (1976): “The use of certain linguistic devices to link or tie together textual units”
- ▶ Lexical cohesion:
 - Indicated by relations between words in the two units (identical word, synonym, hypernym)
 - Before winter I built a chimney, and shingled the sides of my house.
I thus have a tight shingled and plastered house.
 - Peel, core and slice the pears and the apples. Add the fruit to the skillet.

Key intuition: cohesion

- ▶ Non-lexical: anaphora
 - The Woodhouses were first in consequence there. All looked up to them.

- ▶ Cohesion chain:
 - Peel, core and slice the pears and the apples. Add the fruit to the skillet. When they are soft...

Intuition of cohesion-based segmentation

- ▶ Sentences or paragraphs in a subtopic are cohesive with each other
 - ▶ But not with paragraphs in a neighboring subtopic
 - ▶ Thus if we measured the cohesion between every neighboring sentences
 - We might expect a ‘dip’ in cohesion at subtopic boundaries.
- 

tence:	05	10	15	20	25	30	35	40	45	50	55	60	65	
form	1	111	1	1						1	1	1	1	
scientist				11		1	1				1		1	
space	11	1	1											
star	1			1								11	22	111
binary												11	1	
trinary												1	1	
astronomer	1			1								1	1	
orbit	1				1								12	
pull					2		1	1						1
planet	1	1		11			1			1			21	11
galaxy	1											1		
lunar			1	1	1		1							
life	1	1	1					1	11	1	11	1		1
moon		13	1111	1	1	22	21	21			11	1		
move									1	1	1			
continent									2	1	1	2	1	
shoreline											12			
time				1				1	1	1		1		
water							11				1			
say							1	1		1		11		
species									1	1	1			

tence:	05	10	15	20	25	30	35	40	45	50	55	60	65
--------	----	----	----	----	----	----	----	----	----	----	----	----	----



TextTiling (Hearst 1997)

1. Tokenization

- Each space-delimited word
- Converted to lower case
- Throw out stop list words
- Stem the rest
- Group into pseudo-sentences of length $w=20$

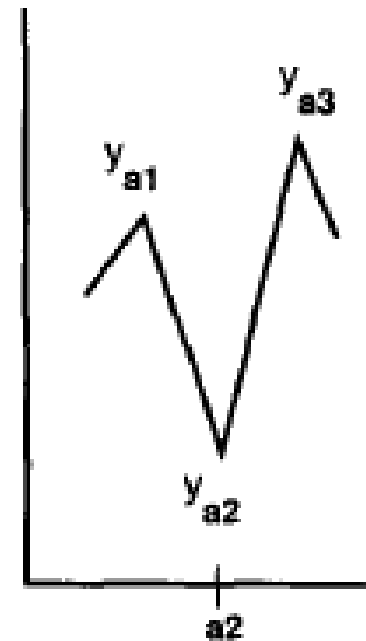
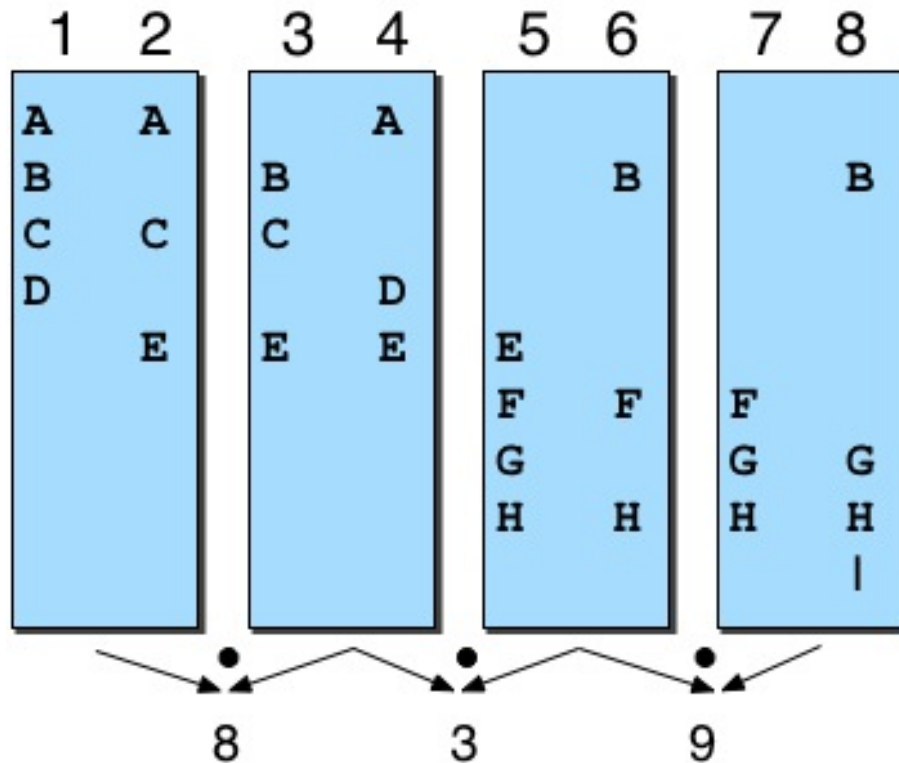
2. Lexical Score Determination: cohesion score

1. Three part score including

- Average similarity (cosine measure) between gaps
- Introduction of new terms
- Lexical chains

3. Boundary Identification

TextTiling algorithm



Cosine

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

Vector Space Model

- ▶ In the vector space model, **both** documents and queries are represented as vectors of numbers.
 - For textiling: both segments are represented as vectors
 - For categorization, both documents are represented as vectors
- ▶ The numbers are derived from the words that occur in the collection

Representation

- ▶ Start with bit vectors $\vec{d}_j = (t_1, t_2, t_3, \dots, t_N)$
- ▶ This says that there are N word types in the collection and that the representation of a document consists of a 1 for each corresponding word type that occurs in the document.
- ▶ We can compare two docs or a query and a doc by summing the bits they have in common

$$\text{sim}(\vec{q}_k, \vec{d}_j) = \sum_{i=1}^N t_{i,k} \times t_{i,j}$$

Term Weighting

- ▶ Bit vector idea treats all terms that occur in the query and the document equally.
- ▶ Its better to give the more important terms greater weight.
 - Why?
 - How would we decide what is more important?

Term Weighting

- ▶ Two measures are used
 - **Local weight**
 - How important is this term to the meaning of this document
 - Usually based on the frequency of the term in the document
 - **Global weight**
 - How well does this term discriminate among the documents in the collection
 - The more documents a term occurs in the less important it is; The fewer the better.

Term Weights

- ▶ Local weights
 - Generally, some function of the frequency of terms in documents is used
- ▶ Global weights
 - The standard technique is known as inverse document frequency

$$idf_i = \log\left(\frac{N}{n_i}\right)$$

N = number of documents; n_i = number of documents with term i

TFxIDF Weighting

- ▶ To get the weight for a term in a document, multiply the term's frequency derived weight by its inverse document frequency.

Back to Similarity

- ▶ We were counting bits to get similarity

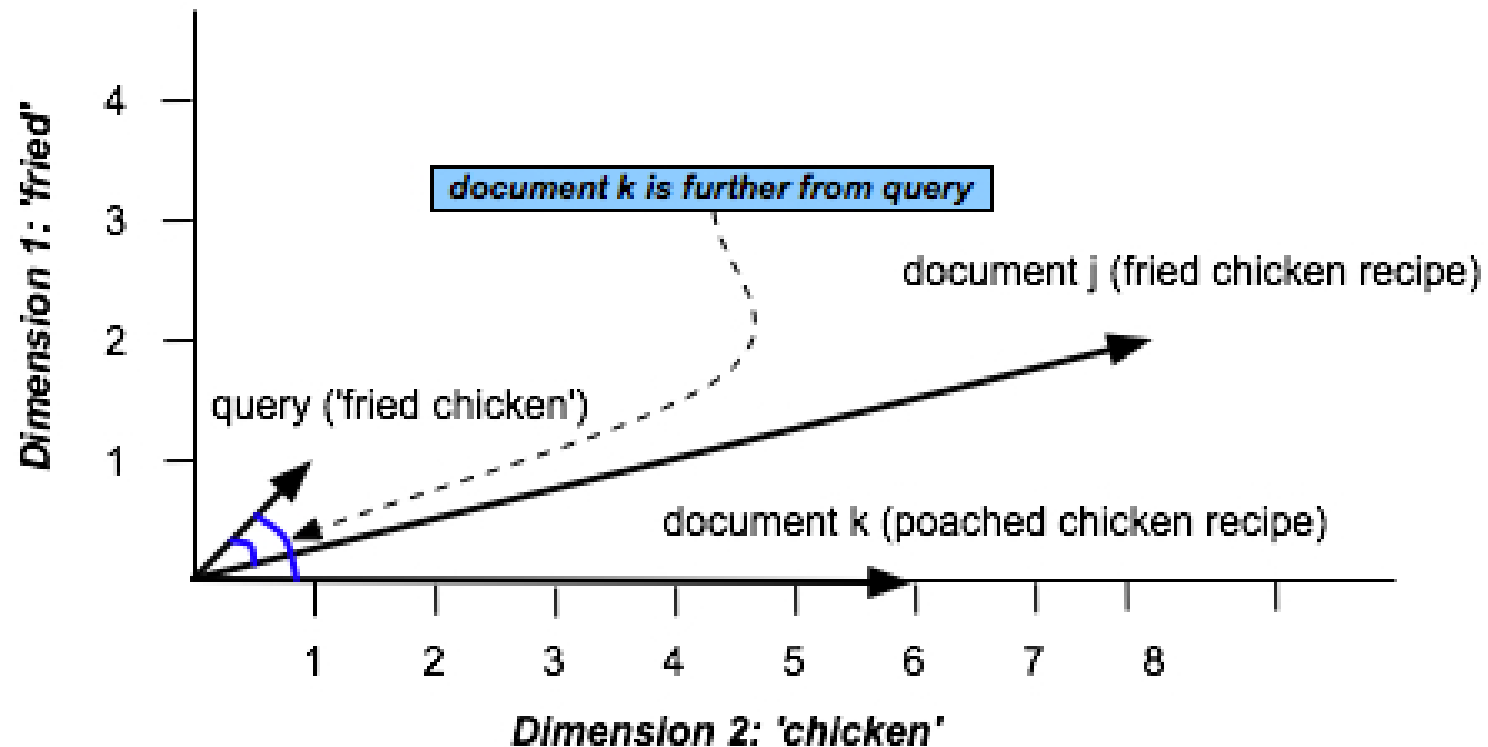
$$\text{sim}(\vec{q}_k, \vec{d}_j) = \sum_{i=1}^N t_{i,k} \times t_{i,j}$$

- ▶ Now we have weights

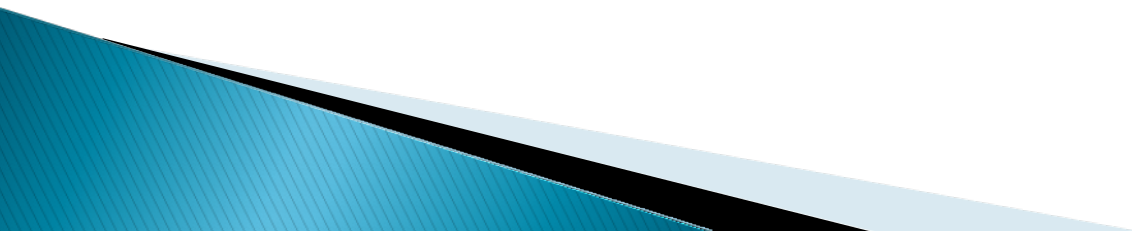
- ▶ But that favors long documents over shorter ones

$$\text{sim}(\vec{q}_k, \vec{d}_j) = \sum_{i=1}^N w_{i,k} \times w_{i,j}$$

Similarity in Space (Vector Space Model)



Similarity

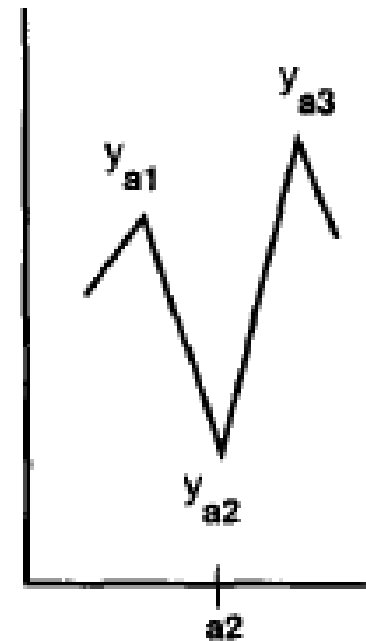
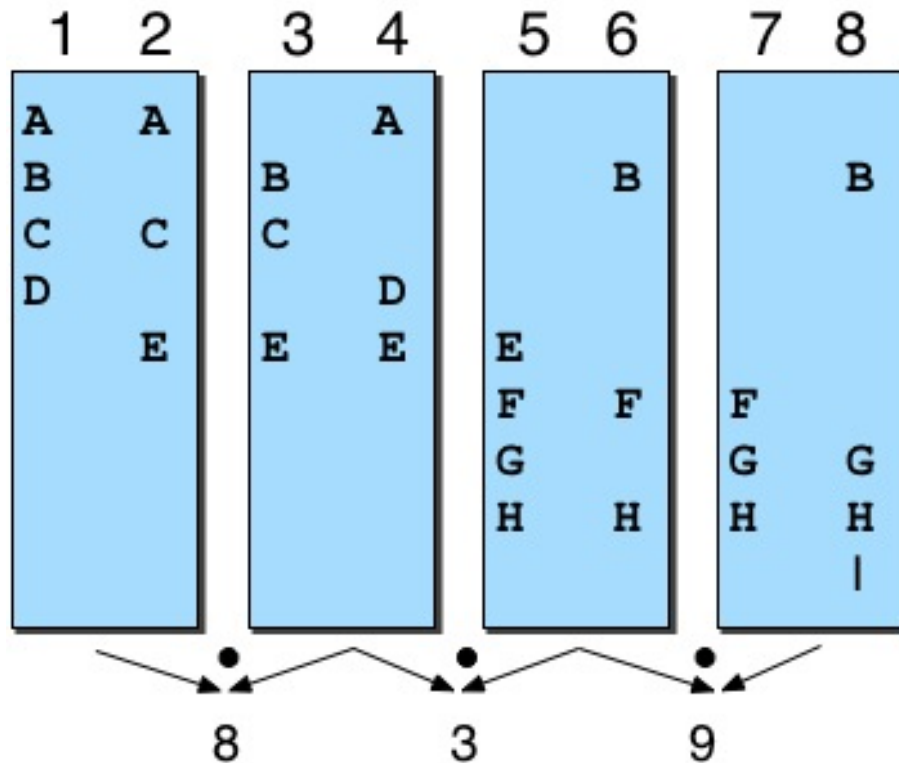
- ▶ View the document as a vector from the origin to a point in the space, rather than as the point.
 - ▶ In this view it's the **direction** the vector is pointing that matters rather than the exact position
 - ▶ We can capture this by normalizing the comparison to factor out the length of the vectors
- 

Similarity

- ▶ The cosine measure

$$\text{sim}(q_k, d_j) = \frac{\sum_{i=1}^N w_{i,k} \times w_{i,j}}{\sqrt{\sum_{i=1}^N w_{i,k}^2} \times \sqrt{\sum_{i=1}^N w_{i,j}^2}}$$

TextTiling algorithm



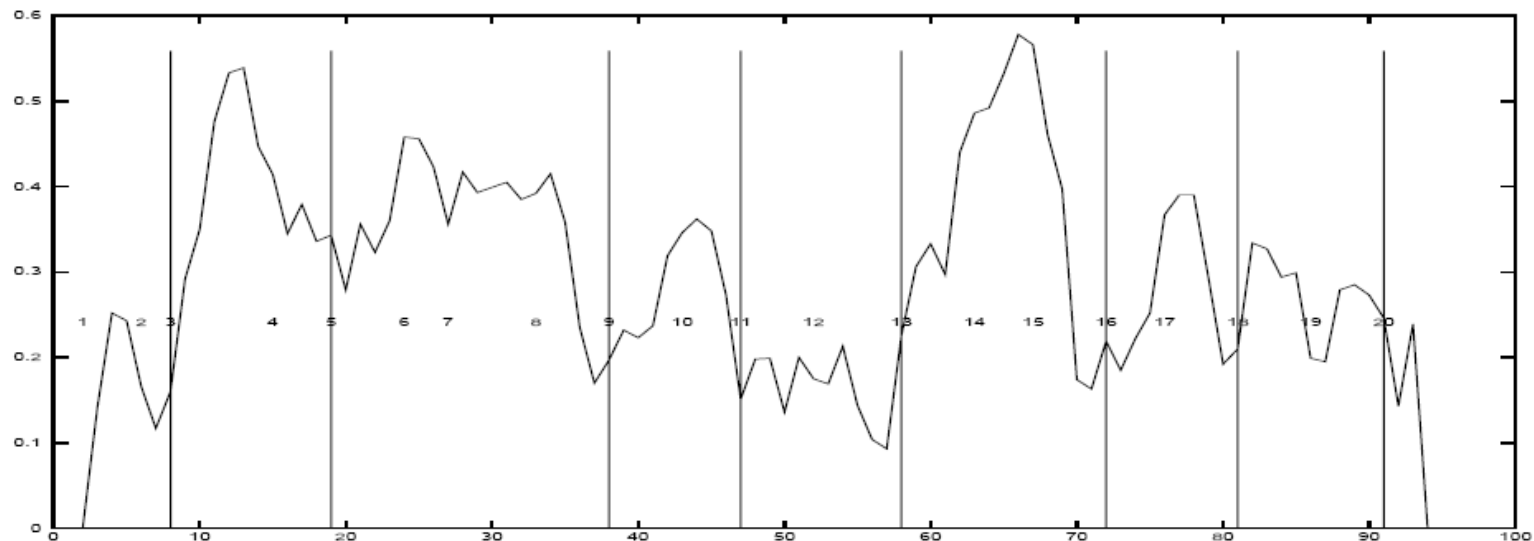
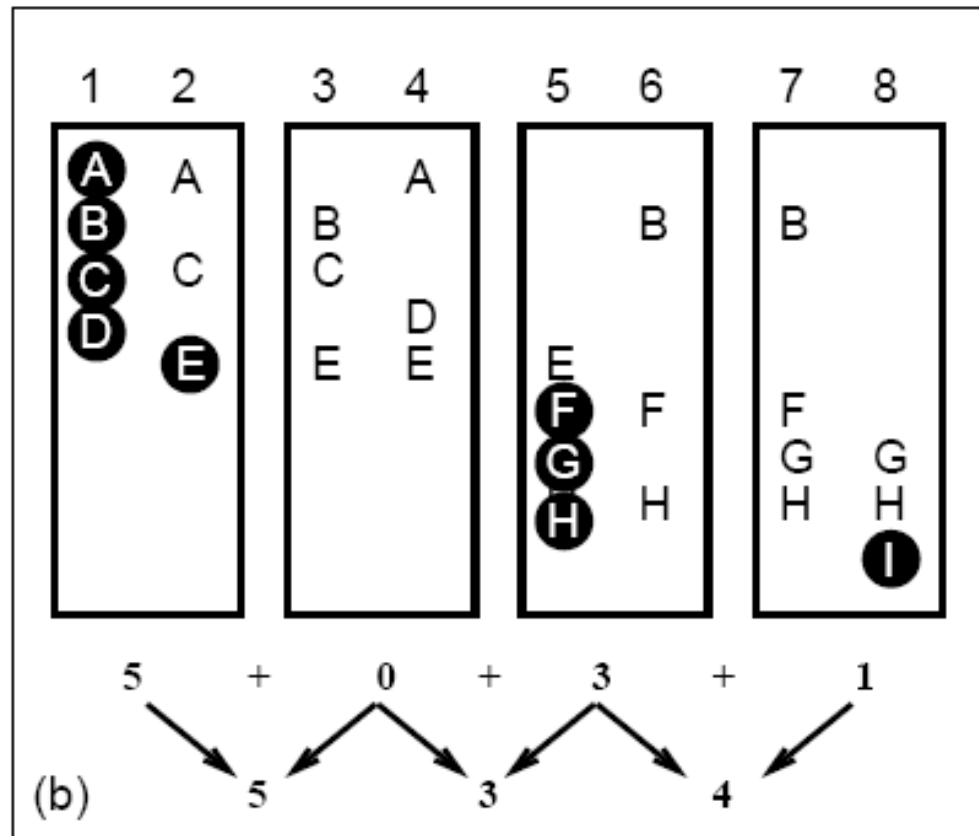


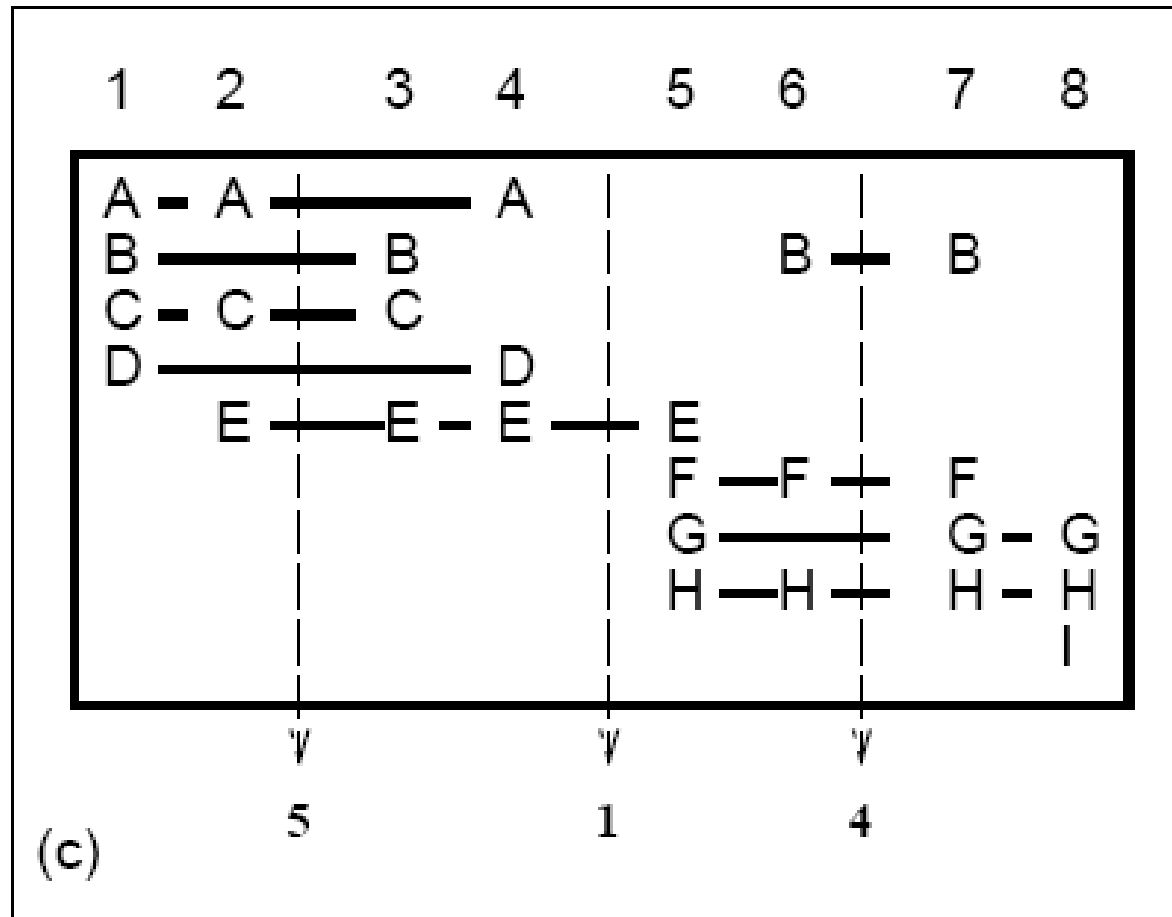
Figure 4: Results of the block similarity algorithm on the *Stargazer* text. Internal numbers indicate paragraph numbers, x-axis indicates token-sequence gap number, y-axis indicates similarity between blocks centered at the corresponding token-sequence gap. Vertical lines indicate boundaries chosen by the algorithm; for example, the leftmost vertical line represents a boundary after paragraph 3. Note how these align with the boundary gaps of Figure 3 above.

Lexical Score Part 2: Introduction of New Terms

TABLE 7.1 (continued)



Lexical Score Part 3: Lexical Chains



Supervised Discourse segmentation

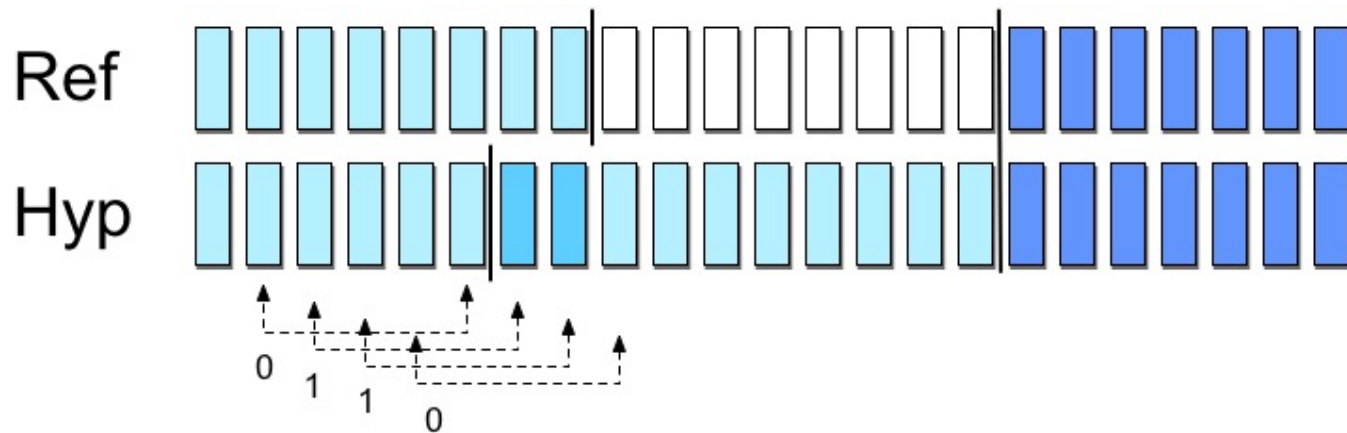
- ▶ Discourse markers or cue words
 - Broadcast news
 - Good evening, I'm <PERSON>
 - ...coming up....
 - Science articles
 - “First,....”
 - “The next topic....”

Supervised discourse segmentation

- ▶ Supervised machine learning
 - Label segment boundaries in training and test set
 - Extract features in training
 - Learn a classifier
 - In testing, apply features to predict boundaries

Supervised discourse segmentation

- ▶ Evaluation: WindowDiff (Pevzner and Hearst 2000)
assign partial credit



Summarization, Question answering, Information extraction, generation

- ▶ Which are more useful where?
 - Discourse structure: subtopics
 - Discourse coherence: relations between sentences
 - Discourse structure: rhetorical relations

Part II: Text Coherence

What makes a discourse coherent?

The reason is that these utterances, when juxtaposed, will not exhibit coherence. Almost certainly not. Do you have a discourse? Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book.

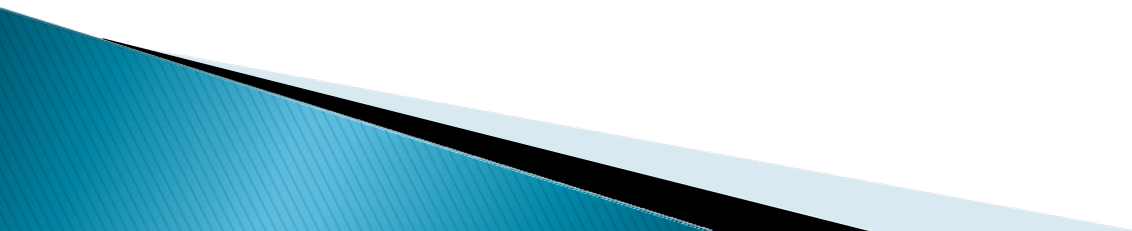
Better?

Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book. Do you have a discourse? Almost certainly not. The reason is that these utterances, when juxtaposed, will not exhibit coherence.

Coherence

- ▶ John hid Bill's car keys. He was drunk.
- ▶ ??John hid Bill's car keys. He likes spinach.

What makes a text coherent?

- ▶ Appropriate use of coherence relations between subparts of the discourse -- **rhetorical structure**
 - ▶ Appropriate sequencing of subparts of the discourse -- **discourse/topic structure**
 - ▶ Appropriate use of **referring expressions**
- 

Hobbs 1979 Coherence Relations

- ▶ “ Result ” :
- ▶ Infer that the state or event asserted by S0 causes or could cause the state or event asserted by S1.
 - The Tin Woodman was caught in the rain. His joints rusted.

Hobbs: “Explanation”

- ▶ Infer that the state or event asserted by S1 causes or could cause the state or event asserted by S0.
 - John hid Bill’s car keys. He was drunk.

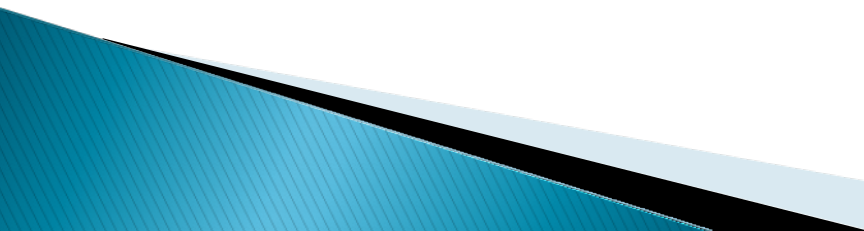
Hobbs: “Parallel”

- ▶ Infer $p(a_1, a_2..)$ from the assertion of S_0 and $p(b_1, b_2...)$ from the assertion of S_1 , where a_i and b_i are similar, for all i .
 - The Scarecrow wanted some brains. The Tin Woodman wanted a heart.

Hobbs “Elaboration”

- ▶ Infer the same proposition P from the assertions of S_0 and S_1 .
 - Dorothy was from Kansas. She lived in the midst of the great Kansas prairies.

Summarization, question answering, information extraction, generation

- ▶ Which are more useful where?
 - Discourse structure: subtopics
 - Discourse coherence: relations between sentences
 - Discourse structure: rhetorical relations
- 

Coherence relations impose a discourse structure

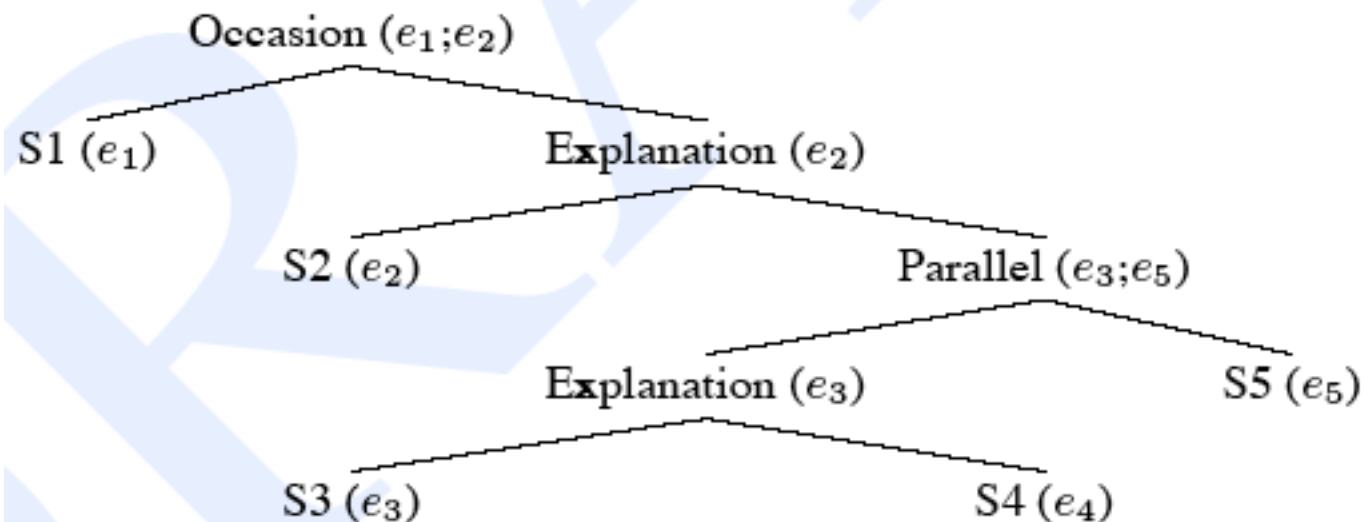
John went to the bank to deposit his paycheck. (S1)

He then took a train to Bill's car dealership. (S2)

He needed to buy a car. (S3)

The company he works for now isn't near any public transportation. (S4)

He also wanted to talk to Bill about their softball league. (S5)

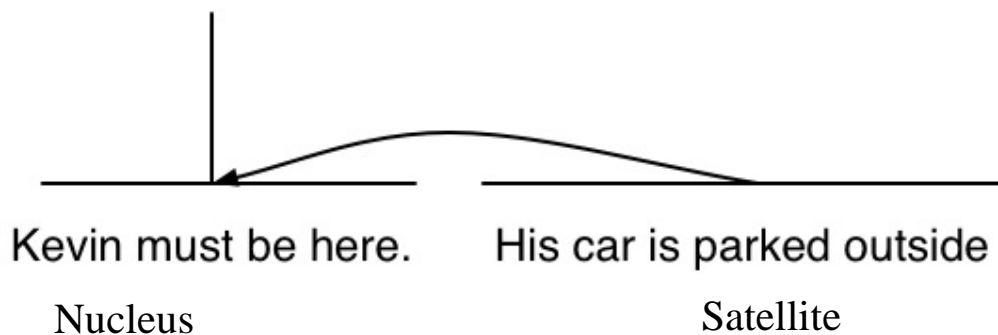


Rhetorical Structure Theory

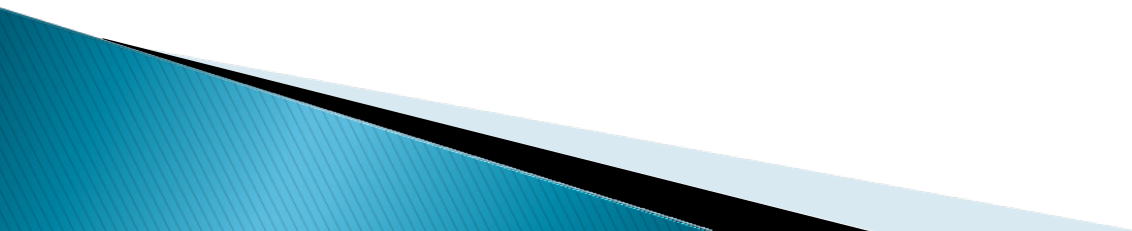
- ▶ Another theory of discourse structure, based on identifying relations between segments of the text
 - **Nucleus/satellite** notion encodes asymmetry
 - Nucleus is thing that if you deleted it, text wouldn't make sense.
 - Some **rhetorical relations**:
 - **Elaboration**: (set/member, class/instance, whole/part...)
 - **Contrast**: multinuclear
 - **Condition**: Sat presents precondition for N
 - **Purpose**: Sat presents goal of the activity in N

One example of rhetorical relation

- ▶ A sample definition
 - Relation: **Evidence**
 - Constraints on N: H might not believe N as much as S think s/he should
 - Constraints on Sat: H *already believes or will believe* Sat
 - Effect: H's belief in N is increased
- ▶ An example:
 - Kevin must be here.**
 - His car is parked outside.**




Automatic Rhetorical Structure Labeling

- ▶ Supervised machine learning
 - Get a group of annotators to assign a set of RST relations to a text
 - Extract a set of surface features from the text that might signal the presence of the rhetorical relations in that text
 - Train a supervised ML system based on the training set
- 

Features: cue phrases

- ▶ Explicit markers: *because, however, therefore, then, etc.*
- ▶ Tendency of certain syntactic structures to signal certain relations:
Infinitives are often used to signal purpose relations: *Use rm to delete files.*
- ▶ Ordering
- ▶ Tense/aspect
- ▶ Intonation

Some Problems with RST

- ▶ How many Rhetorical Relations are there?
 - ▶ How can we use RST in dialogue as well as monologue?
 - ▶ RST does not model overall structure of the discourse.
 - ▶ Difficult to get annotators to agree on labeling the same texts
- 

Summarization, Question answering, Information Extraction, generation

- ▶ Which are more useful where?
 - Discourse structure: subtopics
 - Discourse coherence: relations between sentences
 - Discourse structure: rhetorical relations