ILLINOIS INSTITUTE OF TECHNOLOGY

Discourse

CS-585

Natural Language Processing

Derrick Higgins

- Topic segmentation
- Discourse parsing
- Summarization
- Coherence

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

TOPIC SEGMENTATION

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives. Inventing the Future. www.iit.edu

Segmentation

Words in a Sentence

The [$_{ORG}$New York Times] reported that the summit would be in [$_{PLACE}$Norway].

```
O B-ORG I-ORG I-ORG  O       O
The New York Times reported that

O     O       O   O  O  B-PLACE O
the summit would be in Norway .
```

Sentences in a Text

The New York Times reported that the summit would be in Norway.

Delegates are to begin arriving next week.

In other news, oil prices saw a sharp drop this week as GDP slowed.

OPEC nations plan an emergency meeting to agree on production targets.

In sports, the Champions League kicked off this week with a victory by German champions Dortmund over PSG.

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Topic Segmentation

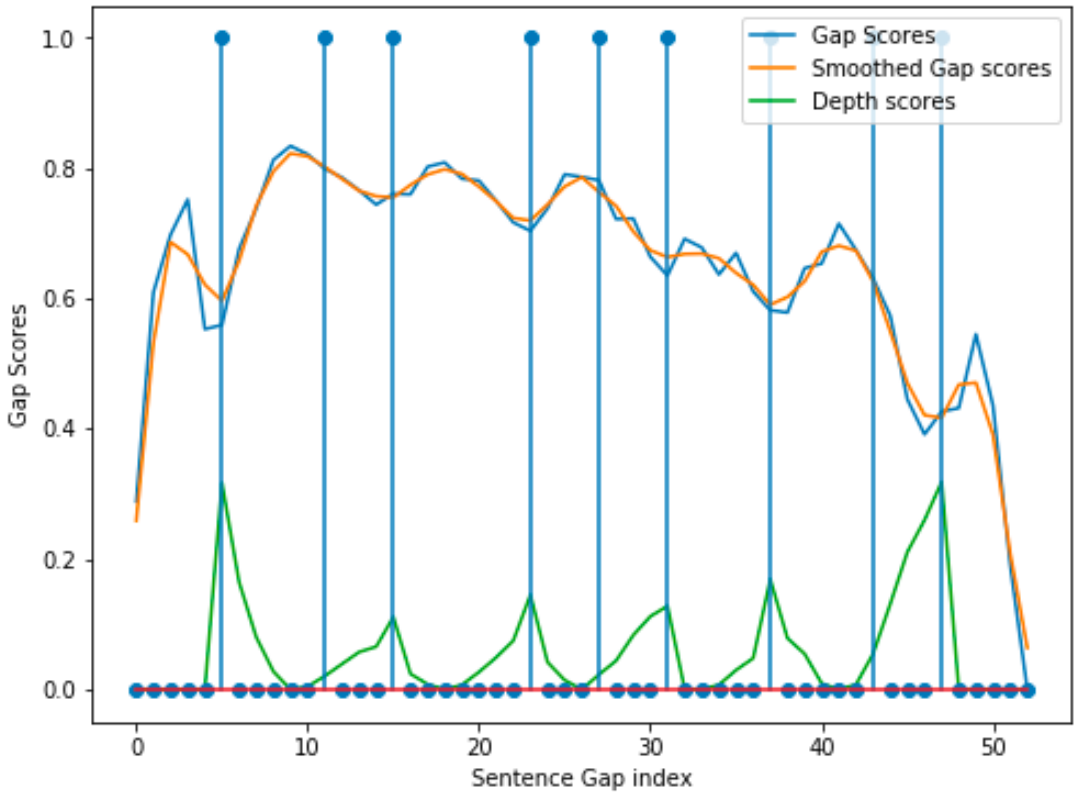
- **Topic segmentation**: dividing a text into units (groups of sentences) that are semantically related – on the same general topic
- Generally an **unsupervised task** – not common to have a list of know topics to start with and an annotated corpus to learn from
- Other possible segmentation criteria: discourse-functional, speech acts

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

TextTiling (Hearst, 1997)

A simple method for segmenting a text into segments based on topic / semantics

1. Segment text into sentences (or pseudo-sentences)
2. Calculate cosine similarity between adjacent units
3. Smooth
4. Gap scores = average distance from peaks
5. Threshold and eliminate neighbors that are too close



ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives. Inventing the Future. www.iit.edu

More complex topic segmentation

- Probabilistic models, e.g., HMM
  - Transition probabilities are from latent topics to latent topics (instead of latent tag, for example, in POS tagging)
  - Observation probabilities are models of sentence likelihoods given topic (rather than, e.g., word likelihoods given POS tag)
- Generative variants can be trained in either supervised or unsupervised fashion

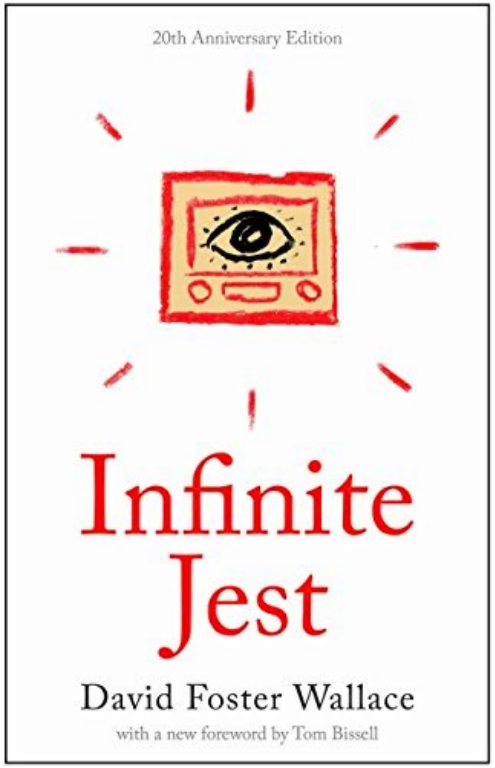
ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Topic segmentation example: Infinite Jest

Example: Wallace, B. 2012. Multiple Narrative Disentanglement: Unraveling *Infinite Jest.*

Task: recover related narrative threads in book in an unsupervised fashion

Leverages information about word usage specific to particular narratives, but also named entities (narrative *focus*)



ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Topic segmentation example: Infinite Jest

Marathe was charged with this operation's details ... A direct assault upon the Academy of Tennis itself was impossible. A.F.R.s fear nothing in this hemisphere except tall and steep hillsides. ...

1. Segment the text into *passages*
2. Extract named entities
3. Apply an LDA-like model over entities per segment to retrieve a clustering of segments into narrative threads

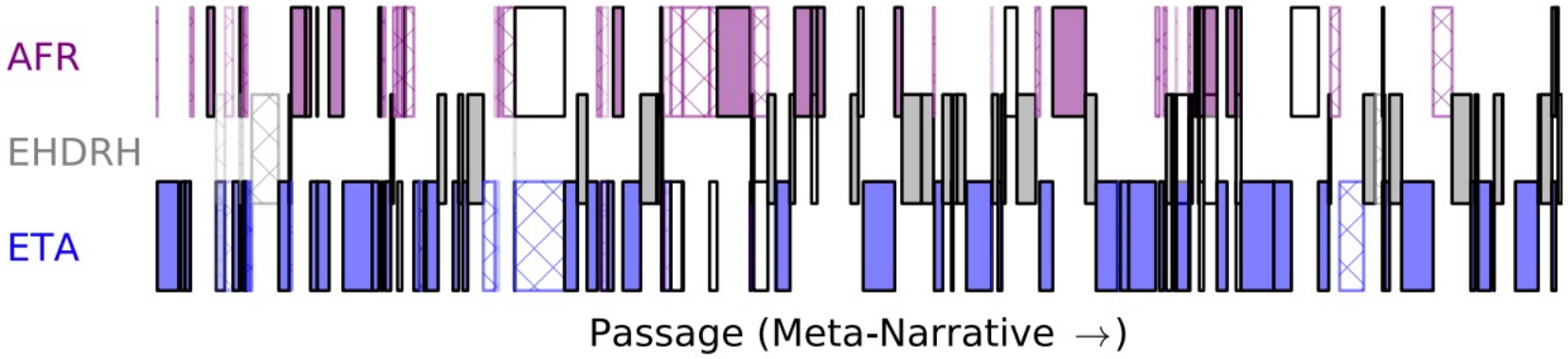


Figure 5: The unsupervised re-construction of the three main narratives using the narrative modeling approach. Hatched boxes denote false-positives (designating a passage as belonging to a narrative when it does not); empty boxes false negatives (failing to assign a passage to narrative to which it belongs).

DISCOURSE PARSING

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives. Inventing the Future. www.iit.edu

Discourse structure

- Syntactic parsing is based on the idea that words combine with one another according to language-specific rules to form higher-level structures
  - Phrase structures or dependency graphs
- Discourse parsing is based on the idea that sentences combine into higher-level units with specific functional relationships
  - E.g., temporal sequences in a narrative
  - Claims and supporting evidence in persuasive writing

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Discourse structure formalisms

- Rhetorical Structure Theory
  - Dates to 1988 (Mann & Thompson)
  - Hierarchical
  - Structure must span entire text
  - Resource: ISI RST Discourse Treebank
- Penn Discourse Treebank schema
  - Binary relations between text segments
  - Distinguishes between “implicit” and “explicit” relations
  - Resource: Penn Discourse Treebank

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Penn Discourse Treebank Relations

- TEMPORAL
  - Asynchronous
  - Synchronous: precedence, succession
- CONTINGENCY
  - Cause: result, reason
  - Pragmatic cause: justification
  - Condition: hypothetical, general, unreal present, unreal past, real present, real past
  - Pragmatic condition: relevance, implicit assertion
- COMPARISON
  - Contrast: juxtaposition, opposition
  - Pragmatic contrast
  - Concession: expectation, contra-expectation
  - Pragmatic concession
- EXPANSION
  - Conjunction
  - Instantiation
  - Restatement: specification, equivalence, generalization
  - Alternative: conjunctive, disjunctive, chosen alternative
  - Exception
  - List

Penn Discourse Treebank

```
a. ( (S
       (SBAR-TMP:ARGM-TMP (IN Until)
         (S
           (NP-SBJ-1 (DT the) (NN building) )
           (VP (VBZ is)
             (VP (VBN completed)
               (NP (-NONE- *-1) )))))
       (, ,)
       (NP-SBJ:ARG0 (NNP Exxon) )
       (VP (MD:ARGM-MOD will)
         (VP (VB rent)
           (NP:ARG1
             (NP (NN part) )
             (PP (IN of)
               (NP (DT an) (VBG existing) (NN office) (NN tower) )))))
       (. .) ))
```

b. Until (TEMPORAL.ASYNCHRONOUS.PRECEDENCE) *the building is completed,* **Exxon will rent part of an existing office tower.** [wsj_0784]

Figure 2
(a) PropBank annotation of the verb *rent*; (b) PDTB annotation of the sentence that *rent* heads.

Prasad et al. 2014. Reflections on the Penn Discourse TreeBank, Comparable Corpora, and Complementary Annotation

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Discourse parsing methods

- Non-hierarchical
  - Classification of relation presence and category for adjacent sentences/phrases
- Hierarchical
  - Modified versions of syntactic parsing algorithms (bottom-up, shift-reduce)
  - Greedy bottom-up extension of non-hierarchical methods

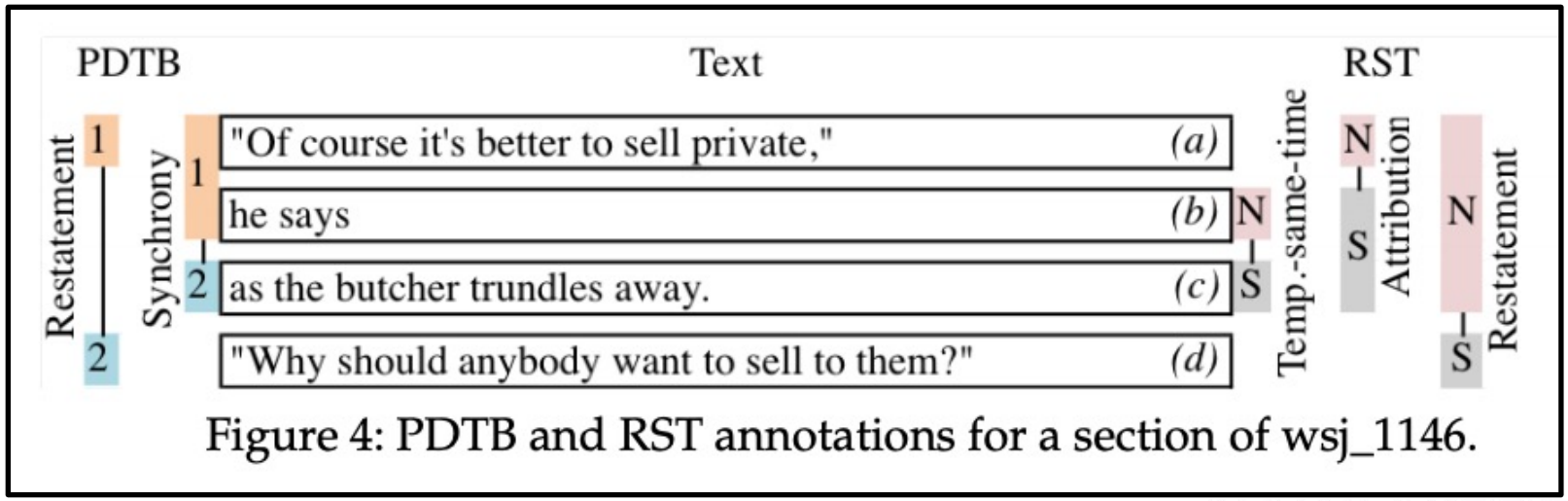


Figure 4: PDTB and RST annotations for a section of wsj_1146.

Demberg, V. et al. 2018. How compatible are our discourse annotation frameworks? Insights from mapping RST-DT and PDTB annotations

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives. Inventing the Future. www.iit.edu

SUMMARIZATION

ILLINOIS INSTITUTE
OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Summarization: an Example

Business | The Washington Post with Bloomberg

How Summly summarizes news articles

This week, 17-year-old British tech wiz Nick D'Aloisio sold his popular iPhone application, Summly to Yahoo for a reported $30 million. The app, which is currently offline, used an algorithm to scan articles and identify the most important information. Summly versions of articles were no more than 400 characters long. Here are some examples of news it summarized:

ORIGINAL NEWS ARTICLE

Source: Back-channel talks but no US-Iran deal on one-to-one nuclear meeting

Andrea Mitchell, NBC News. Published Oct. 20, 2012

A senior administration official told NBC on Saturday that there have been back-channel talks between the U.S. and Iran about meeting bilaterally on the Iranians' nuclear program – but that no meeting has been agreed to.

Expanding on a statement issued by the White House after The New York Times reported that there was an agreement, the official says that the backchannel talks have been done in full consultation with the allies – the P5 + 1 and Israel.

The official pointed out that there have been bilateral talks in the past – but that Iran refused to even meet with the P5 +1 during the recent United Nations meetings. He said the Iranians know there will be no agreement unless they give up their nuclear program.

Asked about the impact on Monday's foreign policy debate between President Barack Obama and Republican nominee Mitt Romney, the official said the administration is not happy that the story came out before the debate, but said the American people might be happy to know the administration is willing to explore all possibilities to get Iran to give up its nuclear program.

... Read the rest of the original 3,499 character story

SUMMLY

Extractive vs. Abstractive Summarization

Extractive

Across the country, college campuses have become ghost towns. Students and professors are hunkered down inside, teaching and learning online. University administrators are tabulating the financial costs of the Covid-19 pandemic, which already exceed the CARES Act's support for higher education. The toll of this pandemic is high and will continue to rise. But another crisis looms for students, higher education and the economy if colleges and universities cannot reopen their campuses in the fall.

Abstractive

Campuses have closed, but another crisis looms for universities and colleges if they cannot reopen in the fall.

Methods for Extractive Summarization (1): Sentence Scoring

Unsupervised Approaches

What makes a good summary?

- *Frequency*: Topic/information gets mentioned a lot
  - For example, assign each sentence a score based on the average probability of its content words
- *Centrality*: Topic/information is close to the center of the semantic space represented by the document
  - For example, sentence scores based on similarity with other sentences in the same document (higher is better)

Supervised Approaches

We can train a model based on human-created (extractive) summaries: which sentences are *actually* included in the summary?

Possibly also label sentences for relative importance

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Methods for Extractive Summarization (2): Sentence Selection

- Naive approach: choose top N sentences by score to include in summary
  - Could end up with very redundant summary
- Improved method: add sentences iteratively, considering score as well as how much new information is added
  - Maximum Marginal Relevance:

$$MMR(s_k, S, q) = \lambda Sim(s_k, q) - (1 - \lambda) \max_{s_i \in S} Sim(s_i, s_k)$$

- Also possible to combine sentence scoring and selection into a combined objective function that incorporates centrality and redundancy and optimize directly

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Methods for Abstractive Summarization

- Sequence-to-sequence neural network models have great potential for summarization
- Just as in MT, seq-to-seq models can learn from bilingual corpora, abstractive neural summarization models aim to learn from collections of text with summaries (or abstracts, for example)
- Challenges
  - Training data not as readily available
  - Need to process long documents (and consequent issues with memory capacity, vanishing gradients)
  - Problems with OOV words
  - Capturing global information about salience and topicality

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Evaluation of Text Summarization Models

- Similar to machine translation, in that ground-truth evaluation involves manual review and is very costly
- Multiple criteria:
  - Coverage: Including all key information
  - Diversity: Lack of redundancy across sentences included
  - Fluency: Readability as a standalone text
  - Brevity: Omission of information that is not critical to summary
- Similar to MT (BLEU score), researchers use easily-calculated metrics to approximate true evaluations, given a human-created summary (or set of summaries)
  - ROUGE-N: Precision, Recall and F-measures for ngram overlap between system and reference summaries
  - Slightly different from MT/BLEU – brevity penalty does not apply

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Summarization: related tasks

- Text Simplification
  - Produce a version of a text with simpler, more accessible language
  - Cf., e.g., "Simple Wikipedia"
  - Useful applications: educational support, simplifying legal jargon
- Multi-document summarization
  - Given a *set* of documents about the same document/event, produce a summary that incorporates key information from all of them

Standard deviation

From Wikipedia, the free encyclopedia

Simple English WIKIPEDIA

Standard deviation is a number used to tell how measurements for a group are spread out from the average (mean), or expected value. A low standard deviation means that most of the numbers are close to the average. A high standard deviation means that the numbers are more spread out.[1][2]

Standard deviation

From Wikipedia, the free encyclopedia

WIKIPEDIA The Free Encyclopedia

For other uses, see Standard deviation (disambiguation).

In statistics, the **standard deviation** is a measure of the amount of variation or dispersion of a set of values.[1] A low standard deviation indicates that the values tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the values are spread out over a wider range.

COHERENCE

ILLINOIS INSTITUTE
OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Text Coherence

- Coherence is a property of a text as a whole
  - Related to segmentation, because it has to do with how well segments of a text flow together to create a reader-friendly whole
- Used for
  - Scoring of student writing (and feedback)
  - Clinical applications (coherence of patients' speech, writing)
  - Evaluation of machine-generated text

The staff recommended great restaurants with very reasonable prices within walking distance.
The Paris hop on bus stops nearby.
The Gare l'Est is within 3 blocks.
We paid 75 euro per nite excluding breakfast but paid for breakfast one day and found it very good and reasonably priced.
The rooms are clean and bathrooms ensuite .

Low coherence

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

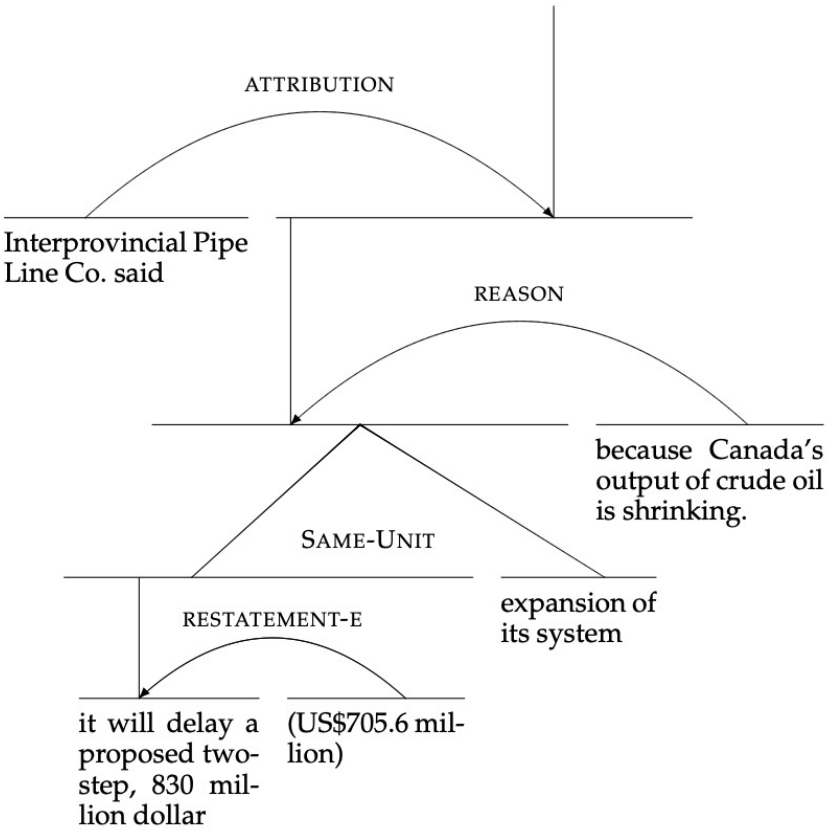
Text Coherence: Sequential Methods

- Sequential modeling of coherence:
  - Idea: coherent texts are ones in which each sentence/segment is closely enough related to the previous one that the reader can follow along
  - Minimal semantic distance between adjacent segments
  - Early approaches: Coherence ∝ average cosine similarity between adjacent segments (LSA)
  - Similar criterion to segmentation method of TextTiling

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Text Coherence: Hierarchical Methods

- But is coherence of a text really just about "one thing following from another"?
- As we saw previously, Rhetorical Structure Theory treats discourse structure as hierarchical
- An alternative to sequential models of text coherence is to try to parse the text into a discourse tree, and then use the relationships between segments that are "close" in the tree to determine coherence (instead of "close" in linear order)



Moret et al. 2018. A Dependency Perspective on RST Discourse Parsing and Evaluation

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Centering Theory and Coherence

- Approaches to coherence discussed thus far have used very simple measures of meaning – vector similarities between segments
- Centering Theory involves a more refined notion of discourse semantics
  - Entities (people and things) referenced in a discourse have greater or lesser *salience* (centrality to the discussion)
  - Part of what makes a discourse coherent is whether entities follow conventional patterns of transition in salience
- In NLP, we approximate entity salience and transitions through the use of an *entity grid*
  - We can use entity grid features in different types of coherence models

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives.Inventing the Future.www.iit.edu

Centering Theory and Coherence

	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

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 & Lapata. 2007. Modeling Local Coherence: An Entity-Based Approach

ILLINOIS INSTITUTE OF TECHNOLOGY
Transforming Lives. Inventing the Future. www.iit.edu

THANK YOU!

ILLINOIS INSTITUTE
OF TECHNOLOGY
Transforming Lives. Inventing the Future. www.iit.edu