Computational Discourse

Yejin Choi
University of Washington

Plan

1. Textual Coherence
2. Rhetorical Structure Theory (RST)
3. Penn Discourse Tree Bank (PDTB)
4. Coreference Resolution / Entity-grid Model
Textual Coherence

- John hid Bill’s car keys. He was drunk.
- John hid Bill’s car keys. He likes spinach.
Textual Coherence

- John went to his favorite music store to buy a piano.
- He had frequented the store for many years.
- He was excited that he could finally buy a piano.
- He arrived just as the store was closing for the day.

- John went to his favorite music store to buy a piano.
- It was a store John had frequented for many years.
- He was excited that he could finally buy a piano.
- It was closing just as John arrived.
Why Model Coherence

“How much wood could a woodchuck chuck if a woodchuck would chuck wood.”

It depends on whether you are talking about African or European woodchucks.

“European woodchucks”

I found 8 European restaurants fairly close to you.
Long-term Coherent Conversation
News aggregation and summary app

Target admits customer PIN data were taken in holiday shopping security breach

NASA: Crew safe as cooling system problem plagues space station

Astronauts aboard the International Space Station are said to be conserving power after one of the 15-year-old space station's two external cooling loops shut down. Officials don't know yet whether the problem resulted from a software glitch or physical failure.
Journalism: Robot or Human?

Despite an expected dip in profit, analysts are generally optimistic about Steelcase as it prepares to reports its third-quarter earnings on Monday, December 22, 2014. The consensus earnings per share estimate is 26 cents per share.

The consensus estimate remains unchanged over the past month, but it has decreased from three months ago when it was 27 cents. Analysts are expecting earnings of 85 cents per share for the fiscal year. Revenue is projected to be 5% above the year-earlier total of $784.8 million at $826.1 million for the quarter. For the year, revenue is projected to come in at $3.11 billion.

The company has seen revenue grow for three quarters straight. The less than a percent revenue increase brought the figure up to $786.7 million in the most recent quarter. Looking back further, revenue increased 8% in the first quarter from the year earlier and 8% in the fourth quarter.

The majority of analysts (100%) rate Steelcase as a buy. This compares favorably to the analyst ratings of three similar companies, which average 57% buys. Both analysts rate Steelcase as a buy.

[Forbes.com; Dec 19, 2014]
While far from op-ed, some of the formulaic news articles are now written by computers.
What is “discourse”? 

Discourse is a coherent structured group of textual units
Discourse “Relations”

- John hid Bill’s car keys. He was drunk. ➔ “Explanation” relation
- John hid Bill’s car keys. He likes spinach. ➔ ??? relation
Discourse “Relations”

- Dorothy was from Kansas. She lived on the Kansas prairies.

- The tin woodman was caught in the rain. His joints rusted.

- The scarecrow wanted some brains. The tin woodsman wanted a heart.

- Dorothy picked up the oil-can. She oiled the Tin Woodman’s joints.

- Result
  - “as a result ...”

- Occasion
  - “and then ...”

- Elaboration
  - “more specifically ...”

- Parallel
Discourse Parsing: Tree of Relations

- **Explanation**
- **Elaboration**
- **Result**
- **Parallel**
- **Occasion**

- John went to the bank to deposit the paycheck. (e1)
- He then took a train to Bill’s car dealership. (e2)
- He needed to buy a car. (e3)
- The company he works for now isn’t near any public transportation. (e4)
- He also wanted to talk to Bill about their softball league. (e5)

---

```
Occasion (e1;e2)

S1 (e1)  Explanation (e2)
     /     |
S2 (e2)  Parallel (e3;e5)

     /   |
S3 (e3)  Explanation (e3)

     /   |
S4 (e4)  S5 (e5)
```
Plan

1. Textual Coherence
2. **Theory**: Rhetorical Structure Theory (RST)
3. **Corpus**: Penn Discourse Tree Bank (PDTB)
4. Coreference Resolution
Rhetorical structure theory (RST)  
Mann and Thompson, 1987

- **Nucleus** – the central unit, interpretable independently.
- **Satellite** – interpretation depends on N

**RST relation** --- a set of constraints on the nucleus and satellite, w.r.t. the goals/beliefs/effects of the writer (W) and the reader (R)

<table>
<thead>
<tr>
<th>Relation Name:</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraints on N:</td>
<td>R might not believe N to a degree satisfactory to W</td>
</tr>
<tr>
<td>Constraints on S:</td>
<td>R believes S or will find it credible</td>
</tr>
<tr>
<td>Constraints on N+S:</td>
<td>R’s comprehending S increases R’s belief of N</td>
</tr>
<tr>
<td>Effects:</td>
<td>R’s belief of N is increased</td>
</tr>
</tbody>
</table>
Types of Schemas in RST

RST schemas := context-free rules for discourse structure
- whether or not the schema has binary, ternary, or arbitrary branching.
- whether or not the RHS has a head (called a nucleus);
- what rhetorical relation, if any, hold between right-hand side (RHS) sisters;

RST schema types in RST annotation

RST schema types in standard tree notation
RST Example

(1) George Bush supports big business.
(2) He’s sure to veto House Bill 1711.
(3) Otherwise, big business won’t support him.

Discourse structure as a tree:

- Leaf := an elementary discourse unit (a continuous text span)
- non-terminal := a contiguous, non-overlapping text span
- root := a complete, non-overlapping cover of the text
RST Example
From Theory to TreeBank

- Rhetorical Structure Theory: Mann and Thompson (1987)
- RST TreeBank: Carlson et al., (2001) defines 78 different RST relations, grouped into 16 classes.

<table>
<thead>
<tr>
<th>Relation Name:</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraints on N:</td>
<td>R might not believe N to a degree satisfactory to W</td>
</tr>
<tr>
<td>Constraints on S:</td>
<td>R believes S or will find it credible</td>
</tr>
<tr>
<td>Constraints on N+S:</td>
<td>R’s comprehending S increases R’s belief of N</td>
</tr>
<tr>
<td>Effects:</td>
<td>R’s belief of N is increased</td>
</tr>
</tbody>
</table>
Graph instead of a tree:

(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.
Discourse relations defined over "abstract objects"

Abstract Objects:
- events, states, propositions (Asher, 1993)

Example of discourse relations:
- Cause, temporal, contrast, condition

A discourse relation holds between *two and only two* AO arguments:

*She hasn’t played any music* since *the earthquake hit.*
Explicit Connectives

Explicit connectives are the lexical items that trigger discourse relations.

- **Subordinating conjunctions** (e.g., *when*, *because*, *although*, etc.)
  - The federal government suspended sales of U.S. savings bonds **because** Congress hasn't lifted the ceiling on government debt.

- **Coordinating conjunctions** (e.g., *and*, *or*, *so*, *nor*, etc.)
  - The subject will be written into the prime-time shows, **and** viewers will be given a 900 number to call.

- **Discourse adverbials** (e.g., *then*, *however*, *as a result*, etc.)
  - In the past, the socialist policies of the government strictly limited the profits businessmen could make. **As a result**, industry operated out of highly inefficient industrial units.

- **Arg2**: the argument with which connective is syntactically associated
- **Arg1**: the other argument
Explicit Connectives

Explicit connectives are the lexical items that trigger discourse relations.

- Subordinating conjunctions (e.g., when, because, although, etc.)
  - *The federal government suspended sales of U.S. savings bonds* because *Congress hasn't lifted the ceiling on government debt.*

- Coordinating conjunctions (e.g., and, or, so, nor, etc.)
  - *The subject will be written into the prime-time shows,* and *viewers will be given a 900 number to call.*

- Discourse adverbials (e.g., then, however, as a result, etc.)
  - *In the past, the socialist policies of the government strictly limited the profits businessmen could make.* As a result, *industry operated out of highly inefficient industrial units.*

- **Arg2**: the argument with which connective is syntactically associated
- **Arg1**: the other argument
Explicit Connectives

Explicit connectives are the lexical items that trigger discourse relations.

- **Subordinating conjunctions** (e.g., *when*, *because*, *although*, etc.)
  - *The federal government suspended sales of U.S. savings bonds because Congress hasn't lifted the ceiling on government debt.*

- **Coordinating conjunctions** (e.g., *and*, *or*, *so*, *nor*, etc.)
  - *The subject will be written into the prime-time shows, and viewers will be given a 900 number to call.*

- **Discourse adverbials** (e.g., *then*, *however*, *as a result*, etc.)
  - *In the past, the socialist policies of the government strictly limited the profits businessmen could make. As a result, industry operated out of highly inefficient industrial units.*

- **Arg2**: the argument with which connective is syntactically associated
- **Arg1**: the other argument
Explicit Connectives

Explicit connectives are the lexical items that trigger discourse relations.

- Subordinating conjunctions (e.g., when, because, although, etc.)
  - *The federal government suspended sales of U.S. savings bonds because* Congress hasn't lifted the ceiling on government debt.

- Coordinating conjunctions (e.g., and, or, so, nor, etc.)
  - *The subject will be written into the prime-time shows, and viewers will be given a 900 number to call.*

- Discourse adverbials (e.g., then, however, as a result, etc.)
  - *In the past, the socialist policies of the government strictly limited the profits businessmen could make. As a result, industry operated out of highly inefficient industrial units.*

- **Arg2**: the argument with which connective is syntactically associated
- **Arg1**: the other argument
Argument Labels and Order

- **Arg2** is the argument with which connective is syntactically associated.
- **Arg1** is the other argument.

Most oil companies, when they set exploration and production budgets for this year, forecast revenue of $15 for each barrel of crude produced.

The chief culprits, he says, are big companies and business groups that buy huge amounts of land "not for their corporate use, but for resale at huge profit." ... The Ministry of Finance, as a result, has proposed a series of measures that would restrict business investment in real estate even more tightly than restrictions aimed at individuals.
Argument Labels and Order

- **Arg2** is the argument with which connective is syntactically associated.
- **Arg1** is the other argument.

- Most oil companies, **when** they set exploration and production budgets for this year, **forecast revenue of $15 for each barrel of crude produced.**

- The chief culprits, he says, are big companies and business groups that buy huge amounts of land "not for their corporate use, but for resale at huge profit." ... The Ministry of Finance, **as a result**, has proposed a series of measures that would restrict business investment in real estate even more tightly than restrictions aimed at individuals.
Arg2 is the argument with which connective is syntactically associated.

Arg1 is the other argument.

- Most oil companies, when they set exploration and production budgets for this year, forecast revenue of $15 for each barrel of crude produced.

- The chief culprits, he says, are big companies and business groups that buy huge amounts of land "not for their corporate use, but for resale at huge profit." … The Ministry of Finance, as a result, has proposed a series of measures that would restrict business investment in real estate even more tightly than restrictions aimed at individuals.

Relative location of Arg1?
Finding Arg 1: Preliminary Experiment

- where to we find Arg 1 the most often?

<table>
<thead>
<tr>
<th>CONN</th>
<th>Same</th>
<th>Previous</th>
<th>Multiple Previous</th>
<th>Distant</th>
</tr>
</thead>
<tbody>
<tr>
<td>nevertheless</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>otherwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>as a result</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>instead</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Finding Arg 1: Preliminary Experiment

- where do we find Arg 1 the most often?
- which connective has highest % of “distant” arg-1?

<table>
<thead>
<tr>
<th>CONNECTIVE</th>
<th>SAME</th>
<th>PREVIOUS</th>
<th>MULTIPLE PREVIOUS</th>
<th>DISTANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>nevertheless</td>
<td>9.7%</td>
<td>54.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>otherwise</td>
<td>11.1%</td>
<td>77.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>as a result</td>
<td>4.8%</td>
<td>69.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore</td>
<td>55%</td>
<td>35%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>instead</td>
<td>22.7%</td>
<td>63.9%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Finding Arg 1: Preliminary Experiment

- where to we find Arg 1 the most often?
- which connective has highest % of “distant” arg-1?

<table>
<thead>
<tr>
<th>CONN</th>
<th>Same</th>
<th>Previous</th>
<th>Multiple Previous</th>
<th>Distant</th>
</tr>
</thead>
<tbody>
<tr>
<td>nevertheless</td>
<td>9.7%</td>
<td>54.8%</td>
<td>9.7%</td>
<td>25.8%</td>
</tr>
<tr>
<td>otherwise</td>
<td>11.1%</td>
<td>77.8%</td>
<td>5.6%</td>
<td>5.6%</td>
</tr>
<tr>
<td>as a result</td>
<td>4.8%</td>
<td>69.8%</td>
<td>7.9%</td>
<td>19%</td>
</tr>
<tr>
<td>therefore</td>
<td>55%</td>
<td>35%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>instead</td>
<td>22.7%</td>
<td>63.9%</td>
<td>2.1%</td>
<td>11.3%</td>
</tr>
</tbody>
</table>
### Hierarchy of PDTB Discourse Relations

<table>
<thead>
<tr>
<th><strong>CONTINGENCY</strong></th>
<th><strong>COMPARISON</strong></th>
<th><strong>TEMPORAL</strong></th>
<th><strong>EXPANSION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Cause</td>
<td>- Contrast</td>
<td>- Asynchronous</td>
<td>- Conjunction</td>
</tr>
<tr>
<td>- Reason</td>
<td>- Juxtaposition</td>
<td>- Synchronous</td>
<td>- Instantiation</td>
</tr>
<tr>
<td>- Result</td>
<td>- Opposition</td>
<td>- Precedence</td>
<td>- Restatement</td>
</tr>
<tr>
<td>- Condition</td>
<td>- Concession</td>
<td>- Succession</td>
<td>- Specification</td>
</tr>
<tr>
<td>- Hypothetical</td>
<td>- Expectation</td>
<td></td>
<td>- Equivalence</td>
</tr>
<tr>
<td>- ...</td>
<td>- Contra-expectation</td>
<td></td>
<td>- Generalization</td>
</tr>
<tr>
<td>- ...</td>
<td>- ...</td>
<td></td>
<td>- Exception</td>
</tr>
</tbody>
</table>

Operating revenue rose 69% to A$8.48 billion from A$5.01 billion. **But** the net interest bill jumped 85% to A$686.7 million from A$371.1 million.

The Texas oilman has acquired a 26.2% stake valued at more than $1.2 billion in an automotive lighting company, Koito Manufacturing Co. **But** he has failed to gain any influence at the company.
All WSJ sections (25 sections; 2304 texts)

100 distinct types

- Subordinating conjunctions – 31 types
- Coordinating conjunctions – 7 types
- Discourse Adverbials – 62 types

Some additional types will be annotated for PDTB-2.0.

18505 distinct tokens
Discourse analysis can help enhancing NLG. How?

- the relative **linear order** of component semantic units
- whether or not to explicitly realize discourse relations (**occurrence**), and if so, how to realize them (**lexical selection and placement**)
NLG: Preliminary Experiment 2

Question: Given a subordinating conjunction and its arguments, in what relative order should the arguments be realized? Arg1-Arg2? Arg2-Arg1?

不同模式的连词

- **When** 几乎等分布：
  - 54% (Arg1-Arg2) and 46% (Arg2-Arg1)

- **Although** 和 (even) **though** 有相反的模式：
  - Although: 37% (Arg1-Arg2) and 63% (Arg2-Arg1)
  - (Even) though: 72% (Arg1-Arg2) and 28% (Arg2-Arg1)
Question: What constrains the lexical choice of a connective for a given discourse relation? (Prasad et al., 2005)

- Testing a prediction for lexical choice rule for CAUSAL *because* and *since* (Elhadad and McKeown, 1990):
  
  - **Assumption:** New information tends to be placed at the end and *given* information at the beginning.
  - **Claim:** *Because* presents *new* information, and *since* presents *given* information
  - **Lexical choice rule:** Use *because* when subordinate clause is postposed (Arg1-Arg2); use *since* when subordinate clause is preposed (Arg2-Arg1)

Because does tend to appear with Arg1-Arg2 order (90%), but CAUSAL *since* is equally distributed as Arg1-Arg2 and Arg2-Arg1.
Sense Disambiguation of Connectives

Some discourse connectives are polysemous, e.g.,

- **While**: comparative, oppositional, concessive
- **Since**: temporal, causal, temporal/causal
- **When**: temporal/causal, conditional

Sense disambiguation is required for many applications:

- **Discourse parsing**: identification of arguments
- **NLG**: relative order of arguments
- **MT**: choice of connective in target language
Sense Disambiguation: Preliminary Experiment

- **Features (from raw text and PTB):**
  - Form of auxiliary *have* - *Has, Have, Had or Not Found.*
  - Form of auxiliary *be* – *Present (am, is, are), Past (was, were), Been, or Not Found.*
  - Form of the head - *Present (part-of-speech VBP or VBZ), Past (VBD), Past Participial (VBN), Present Participial (VBG).*
  - Presence of a modal - *Found or Not Found.*
  - Relative position of Arg1 and Arg2: preposed, postposed
  - If the same verb was used in both arguments
  - If the adverb “not” was present in the head verb phrase of a single argument

- **MaxEnt classifier (McCallum, 2002)**
- **Baseline:** most frequent sense (*CAUSAL*)
- **10-fold cross-validation**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Accuracy</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T,C,T/C)</td>
<td>75.5%</td>
<td>53.6%</td>
</tr>
<tr>
<td>({T,T/C}, C)</td>
<td>90.1%</td>
<td>53.6%</td>
</tr>
<tr>
<td>(T,{C,T/C})</td>
<td>74.2%</td>
<td>65.6%</td>
</tr>
<tr>
<td>(T,C)</td>
<td>89.5%</td>
<td>60.9%</td>
</tr>
</tbody>
</table>

T=temporal, C=causal, T/C=temporal/causal

15-20% improvement over baseline across the board, with state of the art.
Robot or Human?

Despite an expected dip in profit, analysts are generally optimistic about Steelcase as it prepares to reports its third-quarter earnings on Monday, December 22, 2014. The consensus earnings per share estimate is 26 cents per share.

The consensus estimate remains unchanged over the past month, but it has decreased from three months ago when it was 27 cents. Analysts are expecting earnings of 85 cents per share for the fiscal year. Revenue is projected to be 5% above the year-earlier total of $784.8 million at $826.1 million for the quarter. For the year, revenue is projected to come in at $3.11 billion.

The company has seen revenue grow for three quarters straight. The less than a percent revenue increase brought the figure up to $786.7 million in the most recent quarter. Looking back further, revenue increased 8% in the first quarter from the year earlier and 8% in the fourth quarter.

The majority of analysts (100%) rate Steelcase as a buy. This compares favorably to the analyst ratings of three similar companies, which average 57% buys. Both analysts rate Steelcase as a buy.
Plan

1. Textual Coherence
2. Rhetorical Structure Theory (RST)
3. Penn Discourse Tree Bank (PDTB)
4. Coreference Resolution
Discourse Coherence

Discourse is a coherent structured group of textual units

Diagram:
- Discourse Coherence
  - Reference Relations
  - Discourse Relations
    - Informational
    - Intentional
The Problem: Find and Cluster Mentions

Victoria Chen, Chief Financial Officer of Megabucks banking corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37 year old also became the Denver-based financial services company’s president. It has been ten years since she came to Megabucks from rival Lotsabucks.

[Mention Detection]

[Victoria Chen], [Chief Financial Officer of [Megabucks banking corp] since 2004], saw [[her] pay] jump 20%, to $1.3 million, as [the 37 year old] also became the [[Denver-based financial services company] ’s president]. It has been ten years since [she] came to [Megabucks] from rival [Lotsabucks].
The Problem: Find and Cluster Mentions

[Victoria Chen], [Chief Financial Officer of [Megabucks banking corp] since 2004], saw [[her] pay] jump 20%, to $1.3 million, as [the 37 year old] also became the [[Denver-based financial services company]’s president]. It has been ten years since [she] came to [Megabucks] from rival [Lotsabucks].

Co-reference chains:

1. {Victoria Chen, Chief Financial Officer...since 2004, her, the 37-year-old, the Denver-based financial services company’s president}
2. {Megabucks Banking Corp, Denver-based financial services company, Megabucks}
3. {her pay}
4. {rival Lotsabucks}
Types of Coreference (I)

- Types of coreferent phrase:
  - Referential (“semantically definite”) NPs
    - *The author of the book* walked in.
    - *His name was John Smith.*
    - *Mr. Smith* said…
  - Anaphors
    - *Mr. Smith* walked in.
    - *He* talked about *his* car.
  - Descriptive NPs
    - *The stock price* fell from *$4.02* to *$3.85.*
Types of Coreference (II)

• Types of antecedent:
  – Non-generic referring NPs
    • *Mr. Smith* likes *his* car.
  – Generic referring NPs
    • *People* like *their* cars.
  – Non-referring NPs
    • *No one* talked about *their* car.
  – Clauses
    • *Driving fast* isn’t safe, but *it*’s fun.
Coreference as Clustering

The coreference problem can be solved by assigning all NPs in the text to equivalence classes, i.e., by clustering. [Cardie and Wagstaff, 1999]

We need:

• a representation of NPs (as a set of features)
• a distance metric
• a clustering algorithm.
Representing Mentions

Each NP is represented as a set of features:

- **head noun**: last word of the NP;
- **position** in the document;
- **pronoun type**: nominative, accusative, possessive, ambiguous;
- **article**: indefinite, definite, none;
- **appositive**: based on heuristics (commas, etc.);
- **number**: plural, singular;
- **proper name**: based on heuristics (capitalization, etc.);
- **semantic class**: based on Wordnet;
- **gender**: masculine, feminine, either, neuter;
- **animacy**: based on semantic class.
## Example Mentions

<table>
<thead>
<tr>
<th>Words, Head Noun (in bold)</th>
<th>Position</th>
<th>Pronoun Type</th>
<th>Article</th>
<th>Appositive</th>
<th>Number</th>
<th>Proper Name</th>
<th>Semantic Class</th>
<th>Gender</th>
<th>Animacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Simon</td>
<td>1</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>SING</td>
<td>YES</td>
<td>HUMAN</td>
<td>MASC</td>
<td>ANIM</td>
</tr>
<tr>
<td>Chief Financial Officer</td>
<td>2</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>SING</td>
<td>NO</td>
<td>HUMAN</td>
<td>EITHER</td>
<td>ANIM</td>
</tr>
<tr>
<td>Prime Corp.</td>
<td>3</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>SING</td>
<td>NO</td>
<td>COMPANY</td>
<td>NEUTER</td>
<td>INANIM</td>
</tr>
<tr>
<td>1986</td>
<td>4</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>PIPLAR</td>
<td>NO</td>
<td>NUMBER</td>
<td>NEUTER</td>
<td>INANIM</td>
</tr>
<tr>
<td>his</td>
<td>5</td>
<td>POSS</td>
<td>NONE</td>
<td>NO</td>
<td>SING</td>
<td>NO</td>
<td>HUMAN</td>
<td>MASC</td>
<td>ANIM</td>
</tr>
<tr>
<td>pay</td>
<td>6</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>SING</td>
<td>NO</td>
<td>PAYMENT</td>
<td>NEUTER</td>
<td>INANIM</td>
</tr>
<tr>
<td>20%</td>
<td>7</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>PIPLAR</td>
<td>NO</td>
<td>PERCENT</td>
<td>NEUTER</td>
<td>INANIM</td>
</tr>
<tr>
<td>$1.3 million</td>
<td>8</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>PIPLAR</td>
<td>NO</td>
<td>MONEY</td>
<td>NEUTER</td>
<td>INANIM</td>
</tr>
<tr>
<td>the 37-year-old</td>
<td>9</td>
<td>NONE</td>
<td>DEF</td>
<td>NO</td>
<td>SING</td>
<td>NO</td>
<td>HUMAN</td>
<td>EITHER</td>
<td>ANIM</td>
</tr>
<tr>
<td>the financial-services</td>
<td>10</td>
<td>NONE</td>
<td>DEF</td>
<td>NO</td>
<td>SING</td>
<td>NO</td>
<td>COMPANY</td>
<td>NEUTER</td>
<td>INANIM</td>
</tr>
<tr>
<td>company president</td>
<td>11</td>
<td>NONE</td>
<td>NONE</td>
<td>NO</td>
<td>SING</td>
<td>NO</td>
<td>HUMAN</td>
<td>EITHER</td>
<td>ANIM</td>
</tr>
</tbody>
</table>
Clustering

Distance Metric

\[ \text{dist}(NP_1, NP_2) = \sum_{f \in F} w_f \cdot \text{incompatibility}_f(NP_1, NP_2) \]

<table>
<thead>
<tr>
<th>Feature ( f )</th>
<th>Weight</th>
<th>Incompatibility function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words</td>
<td>10.0</td>
<td>(# of mismatching words(^a)) / (# of words in the longer NP)</td>
</tr>
<tr>
<td>Head Noun</td>
<td>1.0</td>
<td>1 if the head nouns differ; else 0</td>
</tr>
<tr>
<td>Position</td>
<td>5.0</td>
<td>(difference in position) / (maximum difference in document)</td>
</tr>
<tr>
<td>Pronoun</td>
<td>( r )</td>
<td>1 if ( NP_i ) is a pronoun and ( NP_j ) is not; else 0</td>
</tr>
<tr>
<td>Article</td>
<td>( r )</td>
<td>1 if ( NP_j ) is indefinite and not appositive; else 0</td>
</tr>
<tr>
<td>Words–Substring</td>
<td>( -\infty )</td>
<td>1 if ( NP_i ) subsumes (entirely includes as a substring) ( NP_j );</td>
</tr>
<tr>
<td>Appositive</td>
<td>( -\infty )</td>
<td>1 if ( NP_j ) is appositive and ( NP_i ) is its immediate predecessor; else 0</td>
</tr>
<tr>
<td>Number</td>
<td>( \infty )</td>
<td>1 if they do not match in number; else 0</td>
</tr>
<tr>
<td>Proper Name</td>
<td>( \infty )</td>
<td>1 if both are proper names, but mismatch on every word; else 0</td>
</tr>
<tr>
<td>Semantic Class</td>
<td>( \infty )</td>
<td>1 if they do not match in class; else 0</td>
</tr>
<tr>
<td>Gender</td>
<td>( \infty )</td>
<td>1 if they do not match in gender (allows EITHER to match MASC or FEM); else 0</td>
</tr>
<tr>
<td>Animacy</td>
<td>( \infty )</td>
<td>1 if they do not match in animacy; else 0</td>
</tr>
</tbody>
</table>

compatible classes, compute transitive closure
## Pairwise Model: Features matter! [Bengston & Roth, 2008]

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mention Types</td>
<td>Mention Type Pair</td>
<td>Annotation and tokens</td>
</tr>
<tr>
<td>String Relations</td>
<td>Head Match</td>
<td>Tokens</td>
</tr>
<tr>
<td></td>
<td>Extent Match</td>
<td>Tokens</td>
</tr>
<tr>
<td></td>
<td>Substring</td>
<td>Tokens</td>
</tr>
<tr>
<td></td>
<td>Substring</td>
<td>Tokens</td>
</tr>
<tr>
<td></td>
<td>Modifiers Match</td>
<td>Tokens</td>
</tr>
<tr>
<td></td>
<td>Alias</td>
<td>Tokens and lists</td>
</tr>
<tr>
<td>Semantic</td>
<td>Gender Match</td>
<td>WordNet and lists</td>
</tr>
<tr>
<td></td>
<td>Number Match</td>
<td>WordNet and lists</td>
</tr>
<tr>
<td></td>
<td>Synonyms</td>
<td>WordNet</td>
</tr>
<tr>
<td></td>
<td>Antonyms</td>
<td>WordNet</td>
</tr>
<tr>
<td></td>
<td>Hypernyms</td>
<td>WordNet</td>
</tr>
<tr>
<td></td>
<td>Both Speak</td>
<td>Context</td>
</tr>
<tr>
<td>Relative Location</td>
<td>Apposition</td>
<td>Positions and context</td>
</tr>
<tr>
<td></td>
<td>Relative Pronoun</td>
<td>Positions and tokens</td>
</tr>
<tr>
<td></td>
<td>Distances</td>
<td>Positions</td>
</tr>
<tr>
<td>Learned</td>
<td>Anaphoricity</td>
<td>Learned</td>
</tr>
<tr>
<td></td>
<td>Name Modifiers Predicted Match</td>
<td>Learned</td>
</tr>
<tr>
<td>Aligned Modifiers</td>
<td>Aligned Modifiers Relation</td>
<td>WordNet and lists</td>
</tr>
<tr>
<td>Memorization</td>
<td>Last Words</td>
<td>Tokens</td>
</tr>
<tr>
<td>Predicted Entity Types</td>
<td>Entity Types Match</td>
<td>Annotation and tokens</td>
</tr>
<tr>
<td></td>
<td>Entity Type Pair</td>
<td>WordNet and tokens</td>
</tr>
</tbody>
</table>
Two Recent Supervised Learners

• Linear Model
  – [Bengston & Roth 2008]
  – Pairwise classification
  – Careful experimental setup with tons of features!
  – 80.8 B³ F1

• FOL-based approach
  – [Culotta et al. 2007]
  – Includes global constraints on clusters
  – 79.3 B³ F1
Multi-pass Sieve

- Basically, a ranking model with no machine learning!
  - 10 sieves, each very simple
  - Winner of CONLL 2011 competition!
### A Carefully Constructed Example

<table>
<thead>
<tr>
<th>Input:</th>
<th>John is a musician. He played a new song. A girl was listening to the song. “It is my favorite,” John said to her.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Final Output:</strong></td>
<td>[John]₁ is a musician. [He]₃ played [a new song]₄. A girl₅ was listening to [the song]₆. “It₇ is [my]₈ favorite]₉,” [John]₁₀ said to [her]₁¹.</td>
</tr>
</tbody>
</table>

| Table 1 | }
The Most Useful Sieves

• 2: Exact string match -- e.g., [the Shahab 3 ground-ground missile] and [the Shahab 3 ground-ground missile]. Precision is over 90% B3 [+16 F1]

• 5: Entity head match – The mention head word matches any head word of mentions in the antecedent entity. Also, looks at modifiers, e.g. to separate Harvard University and Yale University. [+3 F1]

• 10: Pronominal Coreference Resolution – observe constraints on number, gender, person, animacy, and NER types. Link to closest, with a maximum distance. [+10 F1]

• Most others get between 0-2 points improvement, but are cumulative
### Some Results

<table>
<thead>
<tr>
<th>System</th>
<th>MUC</th>
<th></th>
<th></th>
<th>B³</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>F1</td>
<td>R</td>
</tr>
<tr>
<td><strong>ACE2004-Culotta-Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This paper</td>
<td>70.2</td>
<td>82.7</td>
<td>75.9</td>
<td>74.5</td>
</tr>
<tr>
<td>Haghighi and Klein (2009)</td>
<td>77.7</td>
<td>74.8</td>
<td>79.6</td>
<td>78.5</td>
</tr>
<tr>
<td>Culotta et al. (2007)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>73.2</td>
</tr>
<tr>
<td>Bengston and Roth (2008)</td>
<td>69.9</td>
<td>82.7</td>
<td>75.8</td>
<td>74.5</td>
</tr>
<tr>
<td><strong>ACE2004-nwire</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This paper</td>
<td>75.1</td>
<td>84.6</td>
<td>79.6</td>
<td>74.1</td>
</tr>
<tr>
<td>Haghighi and Klein (2009)</td>
<td>75.9</td>
<td>77.0</td>
<td>76.5</td>
<td>74.5</td>
</tr>
<tr>
<td>Poon and Domingos (2008)</td>
<td>70.5</td>
<td>71.3</td>
<td>70.9</td>
<td>–</td>
</tr>
<tr>
<td>Finkel and Manning (2008)</td>
<td>58.5</td>
<td>78.7</td>
<td>67.1</td>
<td>65.2</td>
</tr>
<tr>
<td><strong>MUC6-Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This paper</td>
<td>69.1</td>
<td>90.6</td>
<td>78.4</td>
<td>63.1</td>
</tr>
<tr>
<td>Haghighi and Klein (2009)</td>
<td>77.3</td>
<td>87.2</td>
<td>81.9</td>
<td>67.3</td>
</tr>
<tr>
<td>Poon and Domingos (2008)</td>
<td>75.8</td>
<td>83.0</td>
<td>79.2</td>
<td>–</td>
</tr>
<tr>
<td>Finkel and Manning (2008)</td>
<td>55.1</td>
<td>89.7</td>
<td>68.3</td>
<td>49.7</td>
</tr>
</tbody>
</table>

Table 5
Comparison of our system with the other reported results on the ACE and MUC corpora. All these systems use gold mention boundaries.
Back to … Textual Coherence

• John went to his favorite music store to buy a piano.
• He had frequented the store for many years.
• He was excited that he could finally buy a piano.
• He arrived just as the store was closing for the day.

• John went to his favorite music store to buy a piano.
• It was a store John had frequented for many years.
• He was excited that he could finally buy a piano.
• It was closing just as John arrived.

➡ Same content, different realization through different syntactic choices
Centering Theory

- John went to his favorite music store to buy a piano.
- He had frequented the store for many years.
- He was excited that he could finally buy a piano.
- He arrived just as the store was closing for the day.

- Focus is the most salient entity in a discourse segment

- Constraints on linguistic realization of focus
  - Focus is more likely to be realized as subject or object
  - Focus is more likely to be referred to with anaphoric expression

- Constraints on the entity distribution in a coherent text
  - Transition between adjacent sentences is characterized in terms of focus switch

(Grosz et al., 1983)
Entity-grid Model

1. [Former Chilean dictator Augusto Pinochet]_{S} was arrested in [London]_{X} on [October 14th]_{X} 1998.
2. [Pinochet]_{S}, 82, was recovering from [surgery]_{X}.
3. [The arrest]_{S} was in [response]_{X} to [an extradition warrant]_{X} served by [a Spanish judge]_{S}.
4. [Pinochet]_{S} was charged with murdering [thousands]_{O}, including many [Spaniards]_{O}.
5. [He]_{S} is awaiting [a hearing]_{O}, [his fate]_{X} in [the balance]_{X}.
6. [American scholars]_{S} applauded the [arrest]_{O}.

Notation: \( S = \text{subjects}, \ O = \text{object}, \ X = \text{other} \)

2. [Pinochet]$_S$, 82, was recovering from [surgery]$_X$.


4. [Pinochet]$_S$ was charged with murdering [thousands]$_O$, including many [Spaniards]$_O$.


## Entity-grid Model

<table>
<thead>
<tr>
<th>Pinochet</th>
<th>London</th>
<th>October</th>
<th>Surgery</th>
<th>Arrest</th>
<th>Extradition</th>
<th>Warrant</th>
<th>Judge</th>
<th>Thousands</th>
<th>Spaniards</th>
<th>Hearing</th>
<th>Fate</th>
<th>Balance</th>
<th>Scholars</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>s</td>
<td>x</td>
<td>x</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td></td>
<td></td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 1 2 3 4 5 6
Comparing Grids
Quantifying Textual Coherence

- Text is encoded as a distribution over entity transition types

- Entity transition type — \( \{s, o, x, \_\}^n \)

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>o</th>
<th>x</th>
<th>i</th>
<th>s</th>
<th>o</th>
<th>o</th>
<th>x</th>
<th>i</th>
<th>s</th>
<th>o</th>
<th>x</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{i1} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>.07</td>
<td>0</td>
<td>0</td>
<td>.12</td>
<td>.02</td>
</tr>
<tr>
<td>( d_{i2} )</td>
<td>.02</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.05</td>
<td>.03</td>
</tr>
</tbody>
</table>

How to select relevant transition types?:

- Use all the unigrams, bigrams, \ldots over \( \{s, o, x, \_\} \)

- Do feature selection
Evaluation / Applications

Goal: recover the most coherent sentence ordering

Basic set-up:
– Input: a pair of a source document and a permutation of its sentences
– Task: find a source document via coherence ranking

Data: Training 4000 pairs, Testing 4000 pairs (Natural disasters and Transportation Safety Reports)
Conclusion

• Computational modeling of discourse coherence
• **Theories**: Rhetorical structure theory / Centering theory
• **Corpus**: Penn Discourse Tree Bank
• **Applications**:
  – better summarization
  – automatic ESL grading
  – better QA (sharp et al., NAACL 2010)
  – better machine translation (this workshop!)
• **Coreference Relations**
• **Entity-grid Models**