Upcoming...

• Final report due in a week: March 20, 2017
• Instructions up: ACL style, 5 pages (+references)
Outline

Discourse

Summarization

Wrapup
Discourse

Coreference: Resolving entities and events.

Coherence: What makes the text coherent?

Relations: Rhetorical and narrative links between units.
Discourse

Coreference

Resolving entities and events.

Coherence

What makes the text coherent?

Relations

Rhetorical and narrative links between units
Coherence

1. Sue hid Lakshmi’s keys. She was drunk.
2. ?? Sue hid Lakshmi’s keys. She likes spinach.
Coherence

1. Sue hid Lakshmi’s keys. She was drunk.
2. ?? Sue hid Lakshmi’s keys. She likes spinach.
3. Uma offered spinach to anyone who could keep Lakshmi from driving. Sue hid Lakshmi’s keys. She likes spinach.
A meaningless sentence can be grammatical..

*Colorless green ideas sleep furiously*

The discourse equivalent of grammaticality is coherence

Can a coherent text be without meaning?
Example Essay

In today’s society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world’s learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.
Example Essay

The second reason for the five-paragraph theme is that it makes you focus on a single topic. Some people start writing on the usual topic, like TV commercials, and they wind up all over the place, talking about where TV came from or capitalism or health foods or whatever. But with only five paragraphs and one topic you’re not tempted to get beyond your original idea, like commercials are a good source of information about products. You give your three examples, and zap! you’re done. This is another way the five-paragraph theme keeps you from thinking too much.
Detecting “Coherency”

In today’s society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world’s learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.
Discourse Connectors

In today’s society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world’s learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.
Lexical Chains

In today’s society, college is ambiguous. We need it to live, but we also need it to love. Moreover, without college most of the world’s learning would be egregious. College, however, has myriad costs. One of the most important issues facing the world is how to reduce college costs. Some have argued that college costs are due to the luxuries students now expect. Others have argued that the costs are a result of athletics. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants
1. In today’s society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live,
3. but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live, but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live, but we also need it to love.
3. Moreover, without college most of the world’s learning would be egregious.
4. College, however, has myriad costs.
5. One of the most important issues facing the world is how to reduce college costs.
6. Some have argued that college costs are due to the luxuries students now expect.
7. Others have argued that the costs are a result of athletics.
8. In reality, high college costs are the result of excessive pay for teaching assistants.
1. In today’s society, college is ambiguous.
2. We need it to live, but we also need it to love.
4. Moreover, without college most of the world’s learning would be egregious.
5. College, however, has myriad costs.
6. One of the most important issues facing the world is how to reduce college costs.
7. Some have argued that college costs are due to the luxuries students now expect.
8. Others have argued that the costs are a result of athletics.
9. In reality, high college costs are the result of excessive pay for teaching assistants.
Applications of Coherence

Sentence Ordering

When generating summaries, reorder till sentences are coherent.

Readability Assessment

Is a piece of text easily readable?
Discourse

Coreference

Resolving entities and events.

Coherence

What makes the text coherent?

Relations

Rhetorical and narrative links between units.
Discourse Relations

The kite was created in China, about 2,800 years ago. *Later* it spread into other Asian countries, like India, Japan and Korea. **However**, the kite only appeared in Europe by about the year 1600.

Clouds are heavy. The water in a cloud can have a mass of several million tons.
Use in Sentiment Analysis

It could have been a great movie. It could have been excellent, and to all the people who have forgotten about the older, greater movies before it, will think that as well. It does have beautiful scenery, some of the best since Lord of the Rings. The acting is well done, and I really liked the son of the leader of the Samurai. He was a likeable chap, and I hated to see him die... But, other than all that, this movie is nothing more than hidden rip-offs.
Text Summarization

Goal: produce an abridged version of a text that contains information that is important or relevant to a user.

Summarization Applications
- outlines or abstracts of any document, article, etc
- summaries of email threads
- action items from a meeting
- simplifying text by compressing sentences
What to summarize?

Single-document summarization
- Given a single document, produce
  - abstract
  - outline
  - headline

Multiple-document summarization
- Given a group of documents, produce a “gist”:
  - a series of news stories on the same event
  - a set of web pages about some topic or question
Query-focused vs Generic

Generic summarization:
◦ Summarize the content of a document

Query-focused summarization:
◦ summarize a document with respect to an information need expressed in a user query.
◦ a kind of complex question answering:
  ◦ Answer a question by summarizing a document that has the information to construct the answer
Extractive summarization & Abstractive summarization

Extractive summarization:
◦ create the summary from phrases or sentences in the source document(s)

Abstractive summarization:
◦ express the ideas in the source documents using (at least in part) different words
Summarization: Three Stages

1. content selection: choose sentences to extract from the document
2. information ordering: choose an order to place them in the summary
3. sentence realization: clean up the sentences
## Simplifying sentences

Simplest method: parse sentences, use rules to decide which modifiers to prune (more recently a wide variety of machine-learning methods)

<table>
<thead>
<tr>
<th>Concept</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>appositives</td>
<td>Rajam, 28, an artist who was living at the time in Philadelphia, found the inspiration in the back of city magazines.</td>
</tr>
<tr>
<td>attribution clauses</td>
<td>Rebels agreed to talks with government officials, international observers said Tuesday.</td>
</tr>
<tr>
<td>PPs without named entities</td>
<td>The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [PP to a sustainable number]</td>
</tr>
<tr>
<td>initial adverbials</td>
<td>“For example”, “On the other hand”, “As a matter of fact”, “At this point”</td>
</tr>
</tbody>
</table>
ROUGE (Recall Oriented Understudy for Gisting Evaluation)

Intrinsic metric for automatically evaluating summaries
- Based on BLEU (a metric used for machine translation)
- Not as good as human evaluation (“Did this answer the user’s question?”)
- But much more convenient

Given a document D, and an automatic summary X:
- Have N humans produce a set of reference summaries of D
- Run system, giving automatic summary X
- What percentage of the bigrams from the reference summaries appear in X?

\[
\text{ROUGE} - 2 = \frac{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams } i \in S} \min(\text{count}(i, X), \text{count}(i, S))}{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams } i \in S} \text{count}(i, S)}
\]
Outline

Discourse

Summarization

Wrapup
Word out of Context

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!
Words in Context

\[ P(w_i|w_1, w_2, \ldots, w_{i-1}) = \frac{\# \text{"w}_{i-1} w}{\# \text{"w}_{i-1}} \]

**HMMs/MEMMs**

**CRFs**

Locally normalized (score labels)

Globally normalized (score seqs)
Sentences

I prefer a morning flight.

United canceled the morning flights to Houston.

You can't blame the program for being unable to identify it.

∀x (Expen(x) ∧ Rest(x)) ⇒ Likes(a, x)
Information Extraction

John was born in Liverpool to Julia and Alfred Lennon.

- **John Lennon**: birthplace (Liverpool), childOf (Alfred Lennon)
- **Julia Lennon**: childOf (Alfred Lennon)
Machine Translation

"Vauquois Triangle"

Semantics

Syntax

Phrases

Words

Source

TARGET

interlingua

words

syntax

phrases

"Yo lo haré mañana, I will do it tomorrow"

\( P(Yo\lo\ haré\mañana) = 0.8 \)

English (E) | \( P(\ E \mid \ lo\ haré\ ) \)
--- | ---
will do it | 0.8
will do so | 0.2

English (E) | \( P(\ E \mid \ mañana\ ) \)
--- | ---
tomorrow | 0.7
morning | 0.3
Other “Applications”

\[ t \text{ entails } h \ (t \Rightarrow h) \text{ if humans reading } t \text{ will infer that } h \text{ is most likely true} \]
Wrapup of the Course

Other “Applications”

\( t \text{ entails } h \) if humans reading \( t \) will infer that \( h \) is most likely true
And Now!
Do research in NLP!