Computational Discourse
Overview

Discourse Segmentation

Discourse Parsing

Reference Resolution
Overview

Discourse Segmentation

Discourse Parsing

Reference Resolution

Topics
Jack took out a match. He lit a candle.
What is a discourse?
Jack took out a match. He lit a candle.
Jack took out a match. The sun set.
Jack took out a match. He likes tangerines.

What is a discourse?
A discourse consists of collocated, structured, coherent groups of sentences, such as monologues & dialogues.

What is a discourse?
Goal: Discourse Parsing
Discourse Segmentation:
Segmenting a document into a linear sequence of subtopics
What?
Example
Is discourse segmentation relevant for something other than discourse parsing? Why?
Information Retrieval

VS

Information Retrieval
Text Summarization
2 major approaches:
- Unsupervised
- Supervised

How?
2 major approaches:
- Unsupervised
- Supervised
Cohesive textual units can be grouped by meaning (via linguistic devices)
Identical words:

I built a chimney, and **shingled** the sides of my house.

I have thus a tight **shingled** house.

Cohesion
Hyponym/hypernym:
Peel, core and slice the pears and the apples.
Add the fruit to the skillet.
Anaphora:
The Woodhouses were first in consequence there. All looked up to them.

Cohesion
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>D</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>E</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>F</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>G</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TextTiling

Hearst 1997
### Text Tiling

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>G</td>
</tr>
</tbody>
</table>

**Cohesion scores:**

- 8
- 3
- 9

**Depth score:**

- $y_{a1}$
- $y_{a2}$
- $y_{a3}$

---

Hearst 1997
2 major approaches:
- Unsupervised
- Supervised
Overview
Discourse Segmentation
Discourse Parsing
Reference Resolution

Topics
A discourse consists of collocated, structured, coherent groups of sentences, such as monologues & dialogues. What is a discourse?
Coherent textual units are linked by meaning relations: result, explanation, parallel, elaboration, occasion, etc.
Result:
The Tin Woodman was caught in the rain.
His joints rusted.

Coherence Relation
Elaboration:
Dorothy was from Kansas.
She lived in the midst of the great Kansas Prairies.

Coherence Relation
Discourse Structure (Hobbs)
Kevin must be here.  His car is parked outside.

**Relation Name:** Evidence  
**Constraints on N:** R might not believe N to a degree satisfactory to W  
**Constraints on S:** R believes S or will find it credible  
**Constraints on N+S:** R’s comprehending S increases R’s belief of N  
**Effects:** R’s belief of N is increased
Rhetorical Structure Theory

Marcu 2000
Cue-Phrase-based
(Typical) Discourse Parsing
Approach: SVM, Naïve-Bayes, ...
Features: Cue phrases, word-pairs, ...

Explicit: Easier (MacroF1 ~80%)
Implicit: Harder (MacroF1 ~50%)

(Typical) Discourse Parsing

Overview

Discourse Segmentation

Discourse Parsing

Reference Resolution

Topics
Cohesion vs Coherence?
Equivalent in meaning?
Vs
Collectively form a discourse?

Challenge