Statistical NLP
Discourse (I)

M. Ostendorf
Intro to Discourse: An Analogy

• Parsing:
  – Words group together to form sentences
  – Words have functions (POS tags)
  – Relations of groups of words (syntactic structure)

• Discourse
  – Sentences group to form paragraphs
  – Sentences/clauses have functions (speech acts)
  – Relationships of groups of sentences (discourse structure)
Considerations

- Monologue/prose (single voice)
- Multi-party
  - Interactive (in person meeting, phone call)
  - Asynchronous (online discussion group)

- Multi-party discourse has turntaking, grounding, and other discourse phenomena that we won’t cover
Elements of Discourse
(Grosz & Sidner, 1986)

• Linguistic structure
  – Discourse segments (groups of sentences)
  – Embedding structure

• Intentional structure
  – Intentions & relations among sentences & segments

• Attentional structure
  – Focus of attention (local & global)
Global (Topic) Structure

• Example: encyclopedia article about cities
  – Size/location, history, culture, …
• Example: airplane crashes
  – When, where, casualties, mechanisms, …
Local Coherence

• Example 1a:
  – 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.

• Example 1b:
  – 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.
Local Coherence (cont.)

• Example 2:
  – Terry really goofs sometimes.
  – Yesterday was a beautiful day and he was excited about trying out his new sailboat.
  – He wanted Tony to join him on a sailing expedition.
  – He called him at 6AM.
  – He was sick and furious at being woken up early.
Another Example

• John went to Bob’s party and parked next to a beautiful Acura Integra.
• He went inside and talked to Bob for more than an hour.
• Bob told him that he recently got engaged.
• He also said that he bought it yesterday. (??)
Problems Related to Discourse

- Topic organization, segmentation
- Sentence ordering
- Temporal relations
- Cause/effect
- Reference resolution
Reference Resolution

• John went to Bill’s car dealership to check out an Acura Integra. He looked at it for about an hour.

• I’d like to get from Boston to San Francisco, on either December 5\textsuperscript{th} or 6\textsuperscript{th}. It’s OK if it stops in another city along the way.

• According to John, Bob bought Sue an Integra, and Sue bought Fred a Legend.
  – But that turned out to be a lie.
  – That caused Sue to become rather poor.
Applications that Build on These

- Summarization (esp. multi-document)
- Question answering
- Machine translation
- Automatic essay grading
- Text generation
- Human-computer dialog systems
- Social interaction analysis
- Emphasis prediction in speech synthesis
Who was the prime minister of Australia in 1990?

Please wait while Webclopedia answers to your question (potential answers will follow, with the top 10 below).

Current top 10 (of 109) for "Who was the prime minister of Australia in 1990?" - still finding more...

2. 334.7754 ... sound file Prime Minister RG Menzies opens 'Australia Calling' - 1939 ... 1990 BACK. image file Ian McNamara, presenter of Australia All Over - c. 1990 ... (http://www.abc.net.au/rs/gallery/gallery.htm)
3. 332.0339 John Howard is the prime minister of Australia, Hill is something he unfortunately stepped in, though he managed to scape most of it off on the curb. (http://spleenville.com/~lilith33/blog/mt-comments.cgi?entry_id=518)
4. 332.0339 Mr. Howard is the prime minister of Australia (http://www.opinionjournal.com/editorial/feature.html?Fid=11000312)
5. 347.6498 Visit to Japan by Prime Minister John Winston Howard of Australia (July 2003) (http://www.mofa.go.jp/region/asia-paci/australia/)
7. 336.8.657 Prime Minister Junichiro Koizumi held talks with Mr. John Howard, the Prime Minister of Australia, at the Prime Minister's Official Residence. (http://www.kantei.go.jp/foreign/koizumich/2003/7/16austalia_e.html)
8. 334.0130 Summit Meeting Between Prime Minister Junichiro Koizumi and Prime Minister John Winston Howard of Australia (Summary) (May 2002) (http://www.mofa.go.jp/region/asia-paci/australia/)
9. 336.3329 Prime Minister John Howard: Junichiro Koizumi's Prime Ministerial Visit to Australia (http://www.mofa.go.jp/region/asia-paci/australia/)
10. Still searching for more answers ... (397 sec used to find 116 answers so far)
Text Ordering Tasks

• Extractive summarization
  – Sentence selection & ordering

• Human articles as reference:
  – Scoring: Human order of sentences should rank higher than random permutation
  – Find the best place to insert a sentence into a document
HMM Content Model
(Barzilay & Lee, 2004)

• States = topics (learned via clustering)
• Observations = sentences
• State-dependent observation distribution = bigram language model
• Ordering experiments:
  – Metrics: % predict human ordering, Kendall’s $\tau$
  – Airplane accidents: 41%; $\tau =0.44$
    • Further improvements $\rightarrow$ 76%
  – AP articles: rank=38-96%; $\tau =0.49-0.98$
Centering
(Grosz, Joshi & Weinstein, 1995)

• Entities are key to accounting for local coherence
• Centers = entities in an utterance that link it to other utterances in the segment
• Constraints, e.g. on pronoun realization & center transitions
• Texts that adhere to more “constraints” are more coherent
Entity-Based Content Model
(Barzilay & Lapata, 2005, 2008)

• Represent text with an “entity grid”
  – Column for each entity in segment, rows for sentences
  – Entries = S (subject), O (object), X (neither), -- (not present)

• Extract entity transition frequencies (SS, SO, XS, etc.) from column subsequences

• Distinguish salient entities (most frequent)

• Rank orderings with SVM
1 [The Justice Department] is conducting an [anti-trust trial] against [Microsoft Corp.] with [evidence] that [the company] is increasingly attempting to crush [competitors].
2 [Microsoft] is accused of trying to forcefully buy into [markets] where [its own products] are not competitive enough to unseat [established brands].
4 [Microsoft] claims [its tactics] are commonplace and good economically.
5 [The government] may file [a civil suit] ruling that [conspiracy] to curb [competition] through [collusion] is [a violation of the Sherman Act].
6 [Microsoft] continues to show [increased earnings] despite [the trial].

Results on airplane accidents: 90%

<table>
<thead>
<tr>
<th></th>
<th>Department</th>
<th>Trial</th>
<th>Microsoft</th>
<th>Evidence</th>
<th>Competitors</th>
<th>Markets</th>
<th>Products</th>
<th>Brands</th>
<th>Case</th>
<th>Netscape</th>
<th>Software</th>
<th>Tactics</th>
<th>Government</th>
<th>Suit</th>
<th>Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>s o s x o</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>
<td>----</td>
<td>--------</td>
</tr>
<tr>
<td>2</td>
<td>-- o -- x s o</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>
<td>----</td>
<td>--------</td>
</tr>
<tr>
<td>3</td>
<td>-- s o</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>
<td>----</td>
<td>--------</td>
</tr>
<tr>
<td>4</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>
<td>----</td>
<td>--------</td>
</tr>
<tr>
<td>5</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>
<td>----</td>
<td>--------</td>
</tr>
<tr>
<td>6</td>
<td>- x 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>
<td>----</td>
<td>--------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>s s</th>
<th>s o</th>
<th>s x</th>
<th>s-</th>
<th>o s</th>
<th>o o</th>
<th>o x</th>
<th>o-</th>
<th>x s</th>
<th>x o</th>
<th>x x</th>
<th>x-</th>
<th>s-</th>
<th>o</th>
<th>x</th>
<th>--</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_1$</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.08</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>.09</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>.05</td>
<td>.07</td>
<td>.03</td>
<td>.59</td>
</tr>
<tr>
<td>$d_2$</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>0</td>
<td>.02</td>
<td>.14</td>
<td>.14</td>
<td>.06</td>
<td>.04</td>
<td>.03</td>
<td>.07</td>
<td>.1</td>
<td>.36</td>
</tr>
<tr>
<td>$d_3$</td>
<td>.02</td>
<td>0</td>
<td>0</td>
<td>.03</td>
<td>.09</td>
<td>.09</td>
<td>.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.05</td>
<td>.03</td>
<td>.07</td>
<td>.17</td>
<td>.39</td>
</tr>
</tbody>
</table>
Combined Topic/Entity Model  
(Elsner, Austerweil & Charniak, 2007)

- Hidden Markov topic model (like B&L ’04)
- Observation probability separately models new entities ($N_i$), known entities & their roles ($E_i, R_i$), non-entities ($W_i$)
- Naïve entity grid: drop co-reference, group entities by head nouns
Combined Topic/Entity Model
(Elsner et al., 2007)

Results on airplane accidents: 94%

Topic states

Known entities

New entity
Permutation-Based Topic Model
(Chen, Branavan, Barzilay & Karger, 2009)

• Claim: Markov topic assumption is too weak; the same topic rarely appears in disconnected places in an article

• Generative model for doc $d$ with $N_d$ paragraphs:
  – Generate bag of topics $t_d$ from multinomial
  – Topic order $\pi_d$ characterized by Generalized Mallows model (GMM)
  – Assign topic/paragraph order $z_d$ using $(t_d, \pi_d)$
  – Generate words with topic-dependent unigram
Generalized Mallows Model

- Parameterized in terms of distance from identity permutation \((1,2,3, \ldots, K)\)
- \(v_k = \#\) of elements before \(k\) that would normally follow \(k\)
  
  \((3\ 1\ 5\ 2\ 4) \rightarrow \mathbf{v} = (1\ 2\ 0\ 1)\)

- \(GMM(\mathbf{v}|r) = K \exp(-\Sigma_j r_j v_j)\)
Example

\[ t_d = \{1,1,1,1,2,4,4\} \]
\[ \pi_d = (2,4,3,1) \quad [v=(3,0,1)] \]
\[ z_d = \{2,4,4,1,1,1,1,1\} \]
## Results: Align Para’s to Topics

K=10, 20 topics, City articles, F-score

<table>
<thead>
<tr>
<th>Method</th>
<th>K=10</th>
<th>K=20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering</td>
<td>.50</td>
<td>.51</td>
</tr>
<tr>
<td>HMM</td>
<td>.31</td>
<td>.24</td>
</tr>
<tr>
<td>Most freq</td>
<td>.52</td>
<td>.49</td>
</tr>
<tr>
<td>Uniform</td>
<td>.48</td>
<td>.52</td>
</tr>
<tr>
<td>GMM</td>
<td>.57</td>
<td>.61</td>
</tr>
</tbody>
</table>