In December 1988 a Pan Am jet was blown up over Lockerbie, Scotland, killing 270. Since 1992 Libya has been under U.N. sanctions in effect until the suspects are turned over to United States or Britain. In August 1998 United States and Britain proposed a Netherlands trial. Libya asked for guarantees that the suspects would be incarcerated in Libya. Kofi Annan planned a December 1988 Libyan trip to move negotiations.

Secretary-General Kofi Annan said Wednesday that he may travel to Libya next week in hopes of closing a deal to try two Libyan suspects in the Pan Am Lockerbie bombing. The sanctions, were imposed to force Libyan leader Moammar Gadhafi to turn the men over. Louis Farrakhan, the leader of a U.S. Muslim group, congratulated on his recovery from a hip injury.

Cohesion: language devices that connect individual sentences into a unified whole

Cohesion devices: repetition, coreference, ellipsis

1. There was once a little girl and a little boy and a dog
2. And the sailor was their daddy
3. And the little doggy was white
4. And they like the little doggy
6. And they fed it
7. And they ran away
8. And then daddy had to go on a ship
9. And the children miss the 'em

Goal: induce a model that can predict the degree of text “well-formedness”

Applications: summarization, question-answering, machine-translation

– Evaluation tool
– Scoring mechanism in probabilistic generation
Linguistic Foundations: Centering

Certain connectivity patterns among text entities are characteristic of all well-formed texts (Grosz, Joshi & Weinstein, 1995)

- Unit of analysis: centers
- “Affiliation” of a center: utterance (U) and discourse segment (DS)
- Function of a center: to link between a given utterance and other utterances in discourse

Center Typology

- Types:
  - Forward-looking Centers $C_f (U, DS)$
  - Backward-looking Centers $C_b (U, DS)$
- Connection: $C_b (U_n)$ connects with one of $C_f (U_{n-1})$

Center Shifting

Shifting the center, if it is neither retained nor continued

- $C_b (U_{n+1}) \leftrightarrow C_b (U_n)$

Center Continuation

Continuation of the center from one utterance not only to the next, but also to subsequent utterances

- $C_b (U_{n+1}) = C_b (U_n)$
- $C_b (U_{n+1})$ is the most highly ranked element of $C_f (U_{n+1})$ (thus, likely to be $C_b (U_{n+2})$
Discussion on Centering

- Until now: always based on manual annotations
- Never used in applications

Does it really work?

自动连贯性评估

自动连贯性评估

Entity matrix

Coherent Discourse

Coherence is established via center continuation

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.

Our Approach

1. Construct an entity matrix that encodes distributional and syntactic information
2. Identify matrix patterns characteristic of well-formed texts

自动连贯性评估
Matrix Properties

- Dense vs. sparse columns
- Distribution of syntactic tags

<table>
<thead>
<tr>
<th></th>
<th>HRS</th>
<th>MRS</th>
<th>LRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ss</td>
<td>0.020</td>
<td>0.014</td>
<td>0.010</td>
</tr>
<tr>
<td>s o</td>
<td>0.012</td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>- -</td>
<td>0.417</td>
<td>0.433</td>
<td>0.450</td>
</tr>
</tbody>
</table>

Transformations

Goal: reduce the variability in matrix representation

<table>
<thead>
<tr>
<th>L</th>
<th>Original</th>
<th>Transformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>s o s</td>
<td>s o s</td>
</tr>
<tr>
<td>2.</td>
<td>s o s</td>
<td>s o s</td>
</tr>
<tr>
<td>3.</td>
<td>s o s</td>
<td>s o s</td>
</tr>
</tbody>
</table>

Computation of Entity matrix

- Approximate discourse entities with nouns
- Mark all the members of noun compound with the same syntactic tag

Matrix Comparison
### Vector-Based Representation

<table>
<thead>
<tr>
<th>s</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>s s s</td>
<td>1</td>
</tr>
<tr>
<td>x</td>
<td>2</td>
</tr>
<tr>
<td>o</td>
<td>3</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>x</td>
<td>2</td>
</tr>
<tr>
<td>s</td>
<td>1</td>
</tr>
</tbody>
</table>

### Generative Model

- Probability of a column is defined using an \( n \)-gram model
- The probability of a matrix is computed by multiplying together the probability for each column and normalizing the product

### Example of Transformation

<table>
<thead>
<tr>
<th>Entity Sequences</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictator</td>
<td></td>
</tr>
<tr>
<td>Augusto Pinochet</td>
<td></td>
</tr>
<tr>
<td>London</td>
<td></td>
</tr>
<tr>
<td>October Surgery</td>
<td></td>
</tr>
<tr>
<td>Arrest</td>
<td></td>
</tr>
<tr>
<td>Extradition</td>
<td></td>
</tr>
<tr>
<td>Warrant</td>
<td></td>
</tr>
<tr>
<td>Judge</td>
<td></td>
</tr>
<tr>
<td>Thousands Spaniards</td>
<td></td>
</tr>
<tr>
<td>Hearing</td>
<td></td>
</tr>
<tr>
<td>Fate</td>
<td></td>
</tr>
<tr>
<td>Balance</td>
<td></td>
</tr>
<tr>
<td>Scholars</td>
<td></td>
</tr>
</tbody>
</table>

### Discriminative Model

- Each document is described by fixed set of attributes (entity sequences) and their values (frequency)

<table>
<thead>
<tr>
<th>Texts</th>
<th>x</th>
<th>s s</th>
<th>x</th>
<th>o</th>
<th>s</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>T_2</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>T_3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

- Given a set of “good” and “bad” texts, we can learn attributes with high predictive power
Experiments: Data

<table>
<thead>
<tr>
<th></th>
<th>Humans</th>
<th>HRS</th>
<th>MRS</th>
<th>LRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.13</td>
<td>4.42</td>
<td>4.32</td>
<td>3.60</td>
<td></td>
</tr>
</tbody>
</table>

Results of Anova Analysis:
- Human summaries are more cohesive than machine generated ones
- HRS is not significantly different from MRS
- Both HRS and MRS are significantly more cohesive than LRS

Model Comparison: Baselines
- Readability Measures — a function of the average sentence length and the average number of syllables (Flesh, 1951)
- Word-based Models — the average word overlap of adjacent sentences (Foltz&Kintsch&Landauer, 1998)
- Vector-based Models — the average distances between adjacent sentences based on word distributional properties (Foltz&Kintsch&Landauer, 1998)
- Taxonomy-based Models — the average distances between adjacent sentences based on WordNet (Lin, Resnik)

Agreement

Why not to use kappa?
- Function: Upper-bound on human performance
- Procedure: Leave-one-out resampling (Weiss&Kulikowski)
- Result: Agreement = .612 (Min = .107, Max = .975, SD = .230)
Generative Model: Implementation

- Applied to 6-letter alphabet at various level of compression
- Trained on DUC human summaries
- Tested on machine summaries

Discussion: Generative Model

- No correlation for traditional cohesion model due to redundancy
- High negative correlation for Wordnet-based models!
- Best results on the transformation 3
Future directions

- Dependence on genre
- Contribution of different linguistic features
  - Preliminary results: anaphora doesn’t help
- More sophisticated model (unsupervised grammar induction, gap modeling)

Discussion: Discriminative Model

- Most predictive patterns: [s x], [x o], [s - s] and [s s s]
- Baselines: binary 67%, trinary 37.5%
- Transformation 3 is optimal in all the cases