The Problem

- How to create discourse which seems natural

- Natural being defined as being human-like
Coherence

- Uses the idea of discourse relations
- Also involves the application of anaphors and deixis
- Concerns what makes a document coherent
- Different theories try to determine what is the smallest unit that is combined to make coherent discourse/documents
- Some theories use sentences as the smallest unit, others use clauses
Barzilay and Lapata’s View

- Barzilay and Lapata view the smallest unit to be a sentence

- They only consider chains of n sentences for coherence

- Since they are only considering sentence to sentence coherence, their system is heavily influenced by centering theory

- This view also allows coherence to be a ranking problem, i.e. which sets of sentences in which order are most coherent
Related Work

- Previous work tried to implement all the different aspects of centering theory

- Many projects required hand annotated corpora

- This one doesn’t - this is a benefit
The model

- They turn documents into grids with coreference chains as the columns and sentences as the rows.

- If someone wanted to recreate this using the Stanford CoreNLP, it would be pretty easy.

- At each point in the grid, whether an object in that coreference chain was in that sentence, and what grammatical position it was, is stored.

- This grid is used to rank sentence groupings based on their transitions and the grammatical positions of coreferent participants.
The math

- Chose to go with ranking over classification since coherence is inherently scalar

- The grid is turned into a feature vector, $\Phi(x_{ij}) = (p_1(x_{ij}), p_2(x_{ij}), \ldots, p_m(x_{ij}))$, where $m$ is the number of all pre-defined entity transitions

- Using this feature vector, the goal is to find a parameter vector $w$ such that $w^*(\Phi(x_{ij}) - \Phi(x_{ik})) > 0$ for all $j, i, k$ where $j$ is more coherent than $k$ and $j$ and $k$ are renderings (permutations) of the same document $i$
Text ordering evaluation

- First test involved ranking an original document and various permutations of it

- 100 documents for training, 20 random permutations of each

- Same number for testing

- When compared to a model using LSA (Latent Semantic Analysis), Barzilay and Lapata’s method had a **15.2 increase in accuracy** for newspaper articles and an **18.3 increase in accuracy** for accident reports written by government officials

- For the newspaper articles, the biggest increase was from salience feature (only **4.4% increase without it**), followed by coreference (only **11.3% increase without**), then syntax (only **14.8% increase without**)

- For the accident reports, the biggest increase was from syntax (**16.2% without**), then salience (**16.7% without**), then coreference (**17.6% without**)

- Best performance when using all features
Text summarization evaluation

- Second test involved permutations of automatic multi-document summaries

- These permutations were rated by humans, these ratings were then used to train the system

- The human rankings were evaluated to make sure they agree

- Best results were from just syntax and salience (56.3% increase in accuracy), followed by coreference and syntax (50% increase), then coreference, syntax, and salience (43.8% increase) with coreference and salience in last (37.5% increase)

- For text summarization, it appears that coreference decreases accuracy, and syntax has the greatest benefit