Participant Subjectivity and Involvement as a Basis for Discourse Segmentation
Theory

- Uses a theory similar to Waletzky
- Segment splits recognizable by the speakers relation to the sentence
The problem and related work

- The general problem is how to automatically segment stories
- Passonneau and Litman 1997 - used coreference to segment
The algorithm

1. Pre-process
2. Extract features
3. Determine similarity measures
4. Assign boundaries
Pre-processing

- Remove disfluencies
- Parse
- Generate grammatical dependencies
- Tag for tense and aspect
Feature extraction

- Very simple feature set
- Creates features for each prosodic phrase
- Checks in order:
  1. First person subject or object
  2. Second person subject or object
  3. Third person progressive (-ing form)
  4. Third person event (other forms)
  5. Past/present
Similarity measure

- Sums up the similarity between a sentence and the previous sentences
- Only looks at the $w$ previous sentences, where $w$ is the mean segment length/2
- Similarity measures are weighted such that closer sentence similarities are weighted more heavily
Boundary assignment

- Looks at the similarity measure for each phrase
- Phrases with the lowest similarity measure are assigned as boundaries
- "Lowest" defined as being less than the first $1/l$ quantile for the discourse
- $l = \text{mean segment length}$
Results

- Performs better than P&L with computer co-reference
- Worse than P&L with human co-reference
- Only ~20% as effective as a human annotator
- Lots of work left to be done
Problems

- This system can't compete with the P&L system when it has the appropriate input
- When being compared to topic segmentation models, if everything is compared to human annotations based on participant relation models, the topic segmentation models are going to do worse inherently
Thank you

hope none of you booed. Because I'm assuming nobody did. But, if anybody did, I'm rescinding my thanks from them specifically.