Clustering Voices in *The Waste Land*

### 1. Introduction
- Long-form modernist poetry
- Voices of differing styles throughout test, not explicitly marked
- Examples
  - *Chatty woman*
    - I can’t help it, she said, pulling a long face, it’s them pills I took, to bring it off, she said
    - [158–159]
  - Narrator
    - [97–99]
- Project goals
  - To segment according to changes in voice (Brooke et al. 2012)...
  - ... and then cluster voice segments together (the present work)
- Related work
  - Quantitative poetry analysis (Dagan 1973; Simonton 1990)
  - Clustering in literature (Luyckx, 2006; Koppel et al., 2011)
  - Stylistic inconsistency detection (Graham et al., 2005)

### 2. Automatic Segmentation
- From our earlier work (Brooke et al. 2012)
- Unsupervised model
- Consider each point in text
- Stylistic change curve based on 50-token spans on either side
- Select local maxima of curve as breakpoints
- Features
  - Readability metrics (e.g. word length, lexical density)
  - Frequency of punctuation
  - Frequency of part-of-speech
  - Frequency of line breaks
  - Sentiment metrics (Bacianella et al. 2010)
  - Formality score (Brooke et al., 2010)
  - Lexical LSA vectors from large web corpus, 20 dimensions
  - Features normalized (mean = 0, standard deviation = 1)

### 3. Clustering Method
- Same feature vector as segmentation
- Clustering with k-means
  - Randomly choose k cluster centroids
  - Assign points to cluster
  - Iterate until convergence (less than 0.0001 change)
- Differences from standard k-means
  - Centroid is weighted by span length
  - Use city-block (L1) distance instead of Euclidean
  - Based on our segmentation work
- k = 13, chosen based on expert annotation
- Non-parametric model would be preferred

### 4. Evaluation
- BCubed metrics (Bagga and Baldwin, 1999)
  - Precision: fraction of same cluster pairs also in same category
  - Recall: fraction of same cluster pairs also in same cluster
  - F-score: harmonic mean of precision and recall
- Evaluation 1: 20 artificial mixed-style poems
  - Made from 12 poems representing *The Waste Land* influences by taking 100-200 length spans from 6 of these poems
- Evaluation 2: *The Waste Land*
  - Expert annotation (not definitive)
  - Segmentation baselines
    - Even spacing
    - Gold
  - Clustering baselines
    - Initial (no clustering)
    - Random
  - Seeded k-means
    - Use longest instance of each voice as initial centroid

### 5. Results
- Similar results across both evaluations
- Though *The Waste Land* is more difficult than artificial poems
- Automatic unsupervised better than even-spacing baseline
- But not as good as suggested by segmentation metrics
- For most conditions, k-means is clearly better than baselines
- Though marginal for gold condition in *The Waste Land*
- Starting with voice seeds is very helpful
- Voices most easily distinguished:
  - *Chatty woman* (F-score 0.869)
  - *Chatty woman* (F-score 0.605)

### 6. Conclusion
- Still a long way from a potential human interpretation
- Though some correspondence between human and computer judgments of stylistic distinctiveness
- Improving segmentation seems key to future clustering gains
- Or is it possible to eliminate our separation of segmentation and clustering steps?

### References