**Tracking Literary Reputation with Text Analysis Tools**

### Introduction

This project marriage two different research tracks

- Literary reputation
- How is reputation made or lost?
- Sentiment extraction
- How can computational tools calculate the sentiment expressed in a document?

### Literary reputation

- “Why does some literature supposedly transcend the ages and so constitute ‘culture’ while other once-popular books languish in disuse?” (Tuchman & Ford 1989: 1)
- Can we correlate what is written about an author and his/her work to the author’s reputation and subsequent career?

#### Goals of the project

- Examine the critical reviews of six authors writing in the first half of the 20th century
- These are no longer part of the canon, although they were once considered important

#### Examples

- John Galsworthy (1867–1931)
- Marie Corelli (1855–1919)
- D. H. Lawrence (1885–1930)
- Virginia Woolf (1882–1941)
- T. S. Eliot (1888–1965)
- Arnold Bennett (1867–1931)

- Map information contained in the critical texts to the authors’ reputation

### Sentiment extraction

- Disclose whether a text is expressing positive or negative sentiment about its topic
- Employ information retrieval and text classification methods
- **Current state of the art**
  - Text is treated as a bag of words
  - No consideration is given to
    - Where positive and negative words occur
    - Structural/functional within the text (e.g., introduction, conclusion)
- **Proposed improvement**: Make full use of the structure of the text by developing a discourse parsing tool

### Discourse parsing

- In this project, based on Rhetorical Structure Theory (Mann & Thompson 1988, Taboada and Mann 2006)
- Rhetorical relations are the building blocks of text
- They help explain coherence
- Examples: Cause, Concession, Condition, Elaboration, Summary
- Review texts tend to have a typical rhetorical structure
  - List of pros and cons (performance reviews)
  - Opinions usually summarized at the end
  - Frequent use of concessive relations
  - Elaborations sometimes tangential
- Automated discourse parsing
  - We are developing a parsing method for literary reviews, based on our own data

#### Example: After discourse parsing

We must not, however, discuss that aspect of the problem further, but hasten to acknowledge the worth of Mr. Galsworthy’s character-drawing. His women are as good as his men, and we cannot single out any one of them for special praise. His editor and journalist help to sweeten callings which have a tendency to embitter his characters. His women are as good as his men, and we cannot single out any one of them for special praise. His editor and journalist help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters.

#### Evaluation and results

- Preliminary results based on 10 texts: qualitative evaluation of individual tools
- **Different between discourse- and content-based methods are not significant yet**

#### Contribution

- A large body of data about six authors
- Will be coded in XML and made available
- A set of tools for text analysis, usable for other tasks
- Parallel project on extracting semantic orientation from present-day movie and book reviews and consumer products

### Methods

- Tag documents with parts of speech (Brill 1995)
- Develop a dictionary for literary discourse
- Adaptation of taggers developed for present-day text to early-20th-century British and American texts
- Extract relevant words (positive and negative)
- Aggregate words’ semantic orientation
- Novel or basic method: using keywords
- Need to take into account intensifiers (very good) and negation (not very good)
- Performance of similar methods on present-day movie reviews is about 68% accurate
- Taking text structure into account will enhance performance
- Use discourse parsing to determine
  - Subjective and objective sentences
  - Trivial sentences
- Value

#### Example: Using keywords

We must not, however, discuss that aspect of the problem further, but hasten to acknowledge the worth of Mr. Galsworthy’s character-drawing. His women are as good as his men, and we cannot single out any one of them for special praise. His editor and journalist help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters. His editors and journalists help to sweeten callings which have a tendency to embitter his characters.

#### References & Acknowledgements


---

*Acknowledgements:*

This project is supported by the Natural Sciences and Engineering Research Council of Canada, and Simon Fraser University, under a SSHRC Institutional Grant. Thanks to Julian Brooke and Kimberly Voll for their work on the semantic orientation calculator.