Day 0: Touching Base

Kenneth Benoit

Quantitative Analysis of Textual Data

September 16, 2014
Targets

- Whom this class is for

- Learning objectives
  - fundamentals
  - availability and consequences of *choices*
  - practical ability to work with texts
  - issues of text for social science

- Prerequisites
  - quantitative methods
  - familiarity with R
  - ability to use a text editor
  - (optional) ability to process text files in a programming language such as Python
When I presented the supplementary budget to this House last April, I said we could work our way through this period of severe economic distress. Today, I can report that notwithstanding the difficulties of the past eight months, we are now on the road to economic recovery.

In this next phase of the Government’s plan we must stabilise the deficit in a fair way, safeguard those worst hit by the recession, and stimulate crucial sectors of our economy to sustain and create jobs. The worst is over.

This Government has the moral authority and the well-grounded optimism rather than the cynicism of the Opposition. It has the imagination to create the new jobs in energy, agriculture, transport and construction that this green budget will incentivise. It has the...
What role for “qualitative” analysis in QTA?

- Ultimately all reading of texts is qualitative, even when we count elements of the text or convert them into numbers.
- QTA may involve human judgment in the construction of the feature-document matrix.
- But quantitative text analysis differs from more qualitative approaches in that it:
  - Involves large-scale analysis of many texts, rather than close readings of few texts.
  - Requires no interpretation of texts in a non-positivist fashion.
  - Does not explicitly concern itself with the social or cultural predispositions of the analysts (not critical or constructivist).
- Uses a variety of statistical techniques to extract information from the document-feature matrix.
Key feature of quantitative text analysis (cont.)

- Conversion of textual features into a quantitative matrix. Features can mean:
  - A quantitative or statistical procedure to extract information from the quantitative matrix
  - Summary and interpretation of the quantitative results
When I presented the supplementary budget to this House last April, I said we could work our way through this period of severe economic distress. Today, I can report that notwithstanding the difficulties of the past eight months, we are now on the road to economic recovery.

In this next phase of the Government’s plan we must stabilise the deficit in a fair way, safeguard those worst hit by the recession, and stimulate crucial sectors of our economy to sustain and create jobs. The worst is over.

This Government has the moral authority and the well-grounded optimism rather than the cynicism of the Opposition. It has the imagination to create the new jobs in energy, agriculture, transport and construction that this green budget will incentivise. It has the...
LOGISTICS
quanteda: R package
Course resources

- **Syllabus**: describes class, lists readings, links to reading, and links to exercises and datasets
- **Web page** on http://www.kenbenoit.net/nyu2014qta
  - Contains course handout
  - Slides from class
  - In-class exercises and supporting materials
  - Texts for analysis
  - (links to) Software tools and instructions for use
- **Main readings**
  - Lots of articles
  - Some other texts or on-line articles linked to the course handout (downloadable online)
EXAMPLES
Amicus Curiae Textscores by Party
Using Litigants' Briefs as Reference Texts
(Set Dimension: Petitioners = 1, Respondents = 5)

(from Evans et. al. 2007)
Document classification: “Naive Bayes” classifier
Government v. Opposition in yearly budget debates

Average position among cabinet ministers, government backbenchers, and opposition members, 1983–2013

(from Herzog and Benoit EPSA 2013)