

Translating Texts into Interpretations and Numbers

Department of Government
London School of Economics and Political Science

1 Texts as Sources

2 Content Analysis

3 Practicalities/Reflection

Preview

- Two weeks on data collection
 - Text analysis
 - Interviewing methods
- Problem Set 3
 - Due November 22
- Problem Set 4
 - Discuss in class next 2 weeks
 - Due December 13

1 Texts as Sources

2 Content Analysis

3 Practicalities/Reflection

What counts as text?

- Primary sources
 - Raw, original evidence
- Secondary sources
 - Interpretations of raw evidence
- Tertiary sources
 - Compendia or indices of two other types of sources

Use of Texts

- Text as description
 - Rely on text in lieu of direct observation
 - What do we gain? What do we lose?

Use of Texts

- Text as description
 - Rely on text in lieu of direct observation
 - What do we gain? What do we lose?
- Text as DSOs
 - Treat texts as units of analysis
 - What do we gain? What do we lose?

Challenges of Text

- 1 Source “Quality”
- 2 Subjectivity and differing perspectives
- 3 Historiography
- 4 Selection and confirmation bias

Challenges of Text

- 1 Source “Quality”
- 2 Subjectivity and differing perspectives
- 3 Historiography
- 4 Selection and confirmation bias

But these are really the challenges of *any* research!

1 Texts as Sources

2 Content Analysis

3 Practicalities/Reflection

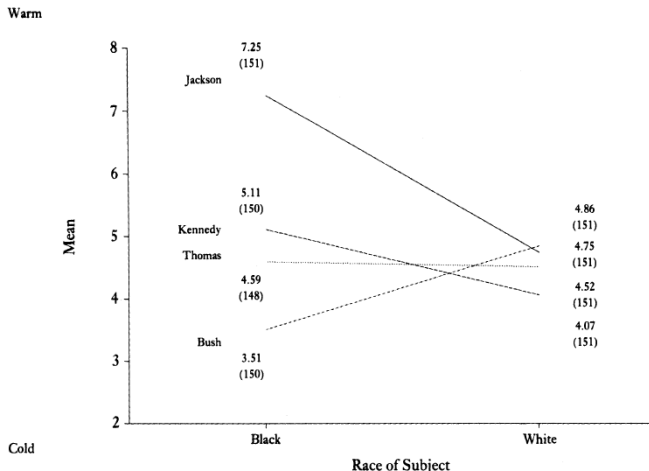
But the biggest issue is that
once we have some texts, our
interpretations are inherently
subjective!

Consider the following quote from a U.S. politician:

African Americans must stop making excuses and rely much more on themselves to get ahead in society.

What does it mean? Is it positive or negative? Is it prejudicial?

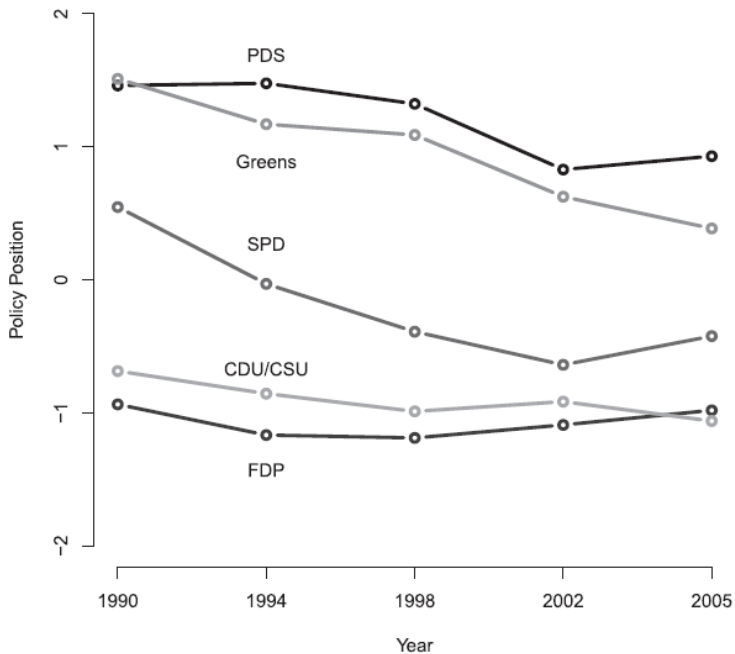
WARMTH TOWARD POLITICAL LEADER BY RACE OF SUBJECT (N IN PARENTHESES)

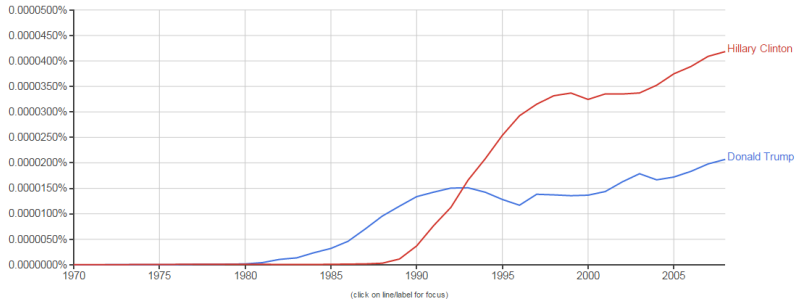


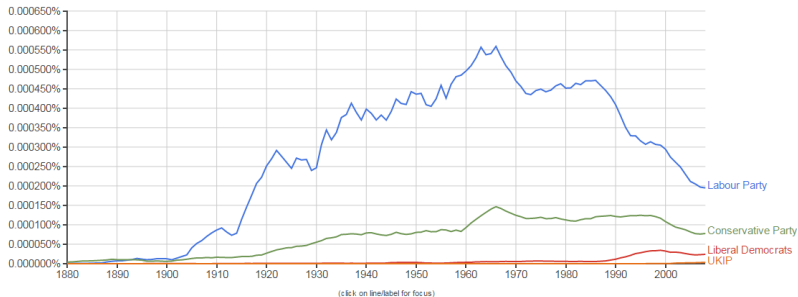
Kuklinski & Hurley. 1994. "On Hearing and Interpreting Political Messages: A Cautionary Tale of Citizen Cue-Taking." *Journal of Politics* 56(3): 729-751.

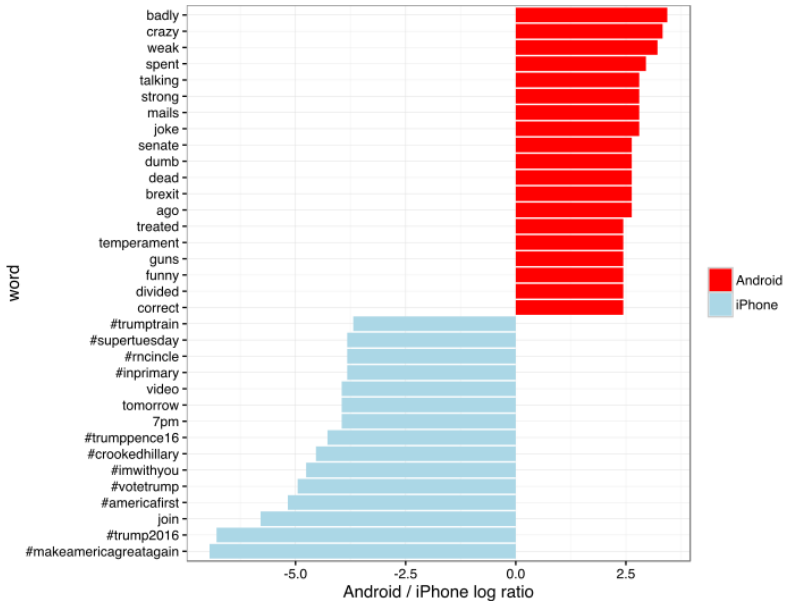
Content Analysis

- Definition: “systematic description of the contents of a communication”
- Treat texts as DSOs or a source of multiple DSOs
 - Data set observation: score for a case on a variable
- Technique can be applied to any kind of document

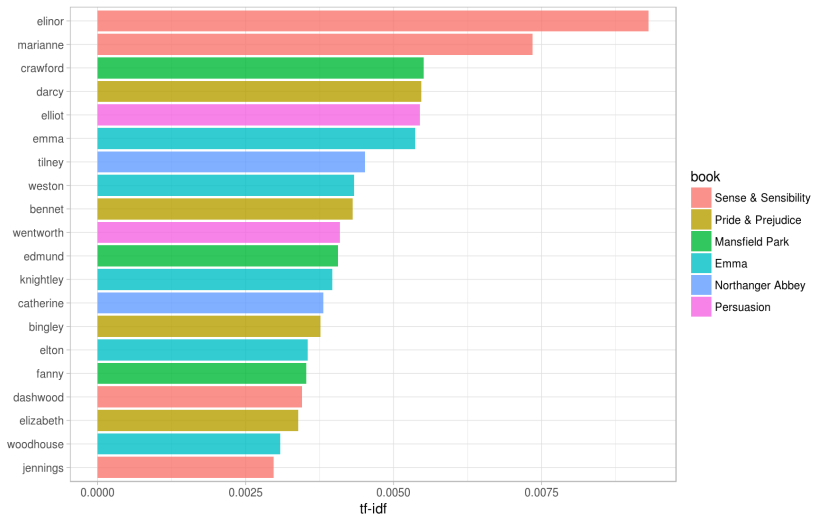








Highest tf-idf words in Jane Austen's Novels



Content Analysis Process

- 1 Define a research question

Content Analysis Process

- 1 Define a research question
- 2 Define concepts

Content Analysis Process

- 1 Define a research question
- 2 Define concepts
- 3 Identify population (sample) of texts

Content Analysis Process

- 1 Define a research question
- 2 Define concepts
- 3 Identify population (sample) of texts
- 4 Create a *codebook* that describes how to operationalize concepts using text

Content Analysis Process

- 1 Define a research question
- 2 Define concepts
- 3 Identify population (sample) of texts
- 4 Create a *codebook* that describes how to operationalize concepts using text
- 5 Assign scores to text units based on codebook

Content Analysis Process

- 1 Define a research question
- 2 Define concepts
- 3 Identify population (sample) of texts
- 4 Create a *codebook* that describes how to operationalize concepts using text
- 5 Assign scores to text units based on codebook
- 6 Aggregate and summarize results

Example 1: Categorization

- Common RQ relates to categorization:
What is this text saying?
- Example RQs:
 - What arguments are raised in this text?
 - Are certain ideas more prevalent in some texts than others?
 - How is an issue discussed in politics?

Example 1: Categorization

- Common RQ relates to categorization:
What is this text saying?
- Example RQs:
 - What arguments are raised in this text?
 - Are certain ideas more prevalent in some texts than others?
 - How is an issue discussed in politics?
- Running ex.: UKIP's 2015 Manifesto

Example 1 (Frames/Discourses)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 1 (Frames/Discourses)

If only all politicians could **believe in Britain** as UKIP does. If only they could share our **positive vision of Britain** as a **proud**, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 1 (Frames/Discourses)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, **independent sovereign nation**, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 1 (Frames/Discourses)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, **independent sovereign nation**, a **country respected on the world stage**, a **major player in global trade**, with **influence and authority** when it comes to tackling the pressing international issues of the day.

Example 2 (Mentions)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 2 (Mentions)

If only all **politicians** could believe in **Britain** as **UKIP** does. If only they could share our positive vision of **Britain** as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, **independent sovereign nation**, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, **independent sovereign nation**, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, **independent sovereign nation**, a country respected on the world stage, a major player in **global trade**, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, **independent sovereign nation**, a country respected on the world stage, a major player in **global trade**, with influence and authority when it comes to tackling the pressing international issues of the day.

Example 3 (Issues/Priorities)

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, **independent sovereign nation**, a country respected on the world stage, a major player in **global trade**, with influence and authority when it comes to tackling the **pressing international issues** of the day.

Developing Codebooks I

- Developed *deductively*
 - 1 Start with set of concepts
 - 2 Describe how those concepts might be expressed in text
 - 3 Create a “coding scheme” by which to score a case on a variable

Developing Codebooks II

- Developed *inductively*
 - 1 Examine texts for key ideas and expressions
 - 2 Create a set of categories that summarize those expressions
 - 3 Examine new texts for expressions of those categories
 - 4 Expand and clarify the codebook as new concepts or expressions emerge

Codebook Example I

- 1 Assign 1 if the text mentions “Europe”

Codebook Example I

- 1 Assign 1 if the text mentions “Europe”
- 2 Count times “Europe” is mentioned

Codebook Example I

- 1 Assign 1 if the text mentions “Europe”
- 2 Count times “Europe” is mentioned
- 3 More complex example:
 - Assign +1 if mentions “Europe” positively
 - Assign -1 if mentions “Europe” negatively
 - Assign 0 if text does not mention “Europe”

Codebook Example I

- 1 Assign 1 if the text mentions “Europe”
- 2 Count times “Europe” is mentioned
- 3 More complex example:
 - Assign +1 if mentions “Europe” positively
 - Assign -1 if mentions “Europe” negatively
 - Assign 0 if text does not mention “Europe”
- 4 List all frames used in the text

Example Scoring Sheet (Paragraph-level)

	Mention	Count	Affect	Frame
UKIP				
Conservative				
Labour				
LibDems				
SNP				
Greens				

Example Scoring Sheet (Sentence-level)

	Sentence	Mention	Affect	Frame
UKIP	1			
UKIP	2			
UKIP	3			
UKIP	4			
UKIP	5			

Example 2: Dictionary Methods

- Rather than rely on subjective interpretation, dictionary methods create a “dictionary” of terms or expressions that operationalize a concept
- Score is based on word presence of words or their frequencies
- Often, goal is to categorize the text based on how it scores relative to other texts

Example 2: Dictionary Methods

- Rather than rely on subjective interpretation, dictionary methods create a “dictionary” of terms or expressions that operationalize a concept
- Score is based on word presence of words or their frequencies
- Often, goal is to categorize the text based on how it scores relative to other texts
 - Sentiment (positivity/negativity)
 - Ideology (left/right)

Affect/Sentiment

If only all politicians could believe in Britain as UKIP does. If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Affect/Sentiment

If only all politicians could believe in Britain as UKIP does. If only they could share our **positive** vision of Britain as a **proud**, independent sovereign nation, a country **respected** on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Affect/Sentiment

If only all politicians could *believe* in Britain as UKIP does. If only they could *share* our **positive** vision of Britain as a **proud**, *independent sovereign* nation, a country **respected** on the world stage, a *major* player in global trade, with **influence** and *authority* when it comes to tackling the pressing international issues of the day.

Unit of Analysis

It can be difficult to decide what the correct unit of analysis is. Possibilities include:

- Author/Source
- Text as a whole
- Section
- Page
- Paragraph
- Sentence
- Sentence fragment
- Word

Unit of Analysis

It can be difficult to decide what the correct unit of analysis is. Possibilities include:

- Author/Source
- Text as a whole
- Section
- Page
- Paragraph
- Sentence
- Sentence fragment
- Word

Smaller units can always be *aggregated* later, but larger units cannot be *disaggregated* later

Sentence

If only all politicians could believe in Britain as UKIP does.

If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

Sentence fragments

If only all politicians could believe in Britain as UKIP does.

If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

“Bag of Words” Approaches I

If only all politicians could believe in Britain as UKIP does . If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

“Bag of Words” Approaches I

If only all politicians could believe in Britain as UKIP does . If only they could share our positive vision of Britain as a proud, independent sovereign nation, a country respected on the world stage, a major player in global trade, with influence and authority when it comes to tackling the pressing international issues of the day.

“Bag of Words” Approaches II

only all politicians could believe Britain UKIP
does only they could share our positive vision
Britain proud independent sovereign nation
country respected world stage major player
global trade influence authority comes
tackling pressing international issues day

“Bag of Words” Approaches III

independent sovereign nation country international
world global

positive proud respected major influence authority

politicians player Britain Britain UKIP

believe share tackling pressing

could could does comes

vision stage day trade issues

only only all they our

N-grams and “words”

- “Bag of words” usually uses *unigrams*

N-grams and “words”

- “Bag of words” usually uses *unigrams*
- Unigrams are often “stemmed”
 - family, families, familiarity, and familial all become *famil*.

N-grams and “words”

- “Bag of words” usually uses *unigrams*
- Unigrams are often “stemmed”
 - family, families, familiarity, and familial all become *famil*.
- Can also be bigrams, trigrams, *n*-grams

N-grams and “words”

- “Bag of words” usually uses *unigrams*
- Unigrams are often “stemmed”
 - family, families, familiarity, and familial all become *famil*.
- Can also be bigrams, trigrams, *n*-grams
- These allow for *multi-word expressions*
 - “If only” has a negative connotation even though “if” and “only” are basically neutral words on their own

1 Texts as Sources

2 Content Analysis

3 Practicalities/Reflection

Content Analysis in Practice

- Inductive processes generally require a “training set” and “test set”
 - Establish coding scheme on “training set”
 - Apply it to broader “test set”

Content Analysis in Practice

- Inductive processes generally require a “training set” and “test set”
 - Establish coding scheme on “training set”
 - Apply it to broader “test set”
- Be cautious about scope conditions

Content Analysis in Practice

- Inductive processes generally require a “training set” and “test set”
 - Establish coding scheme on “training set”
 - Apply it to broader “test set”
- Be cautious about scope conditions
- Develop a precise and reliable codebook

Intercoder Reliability

- Definition: “extent to which two (or more) coders agree on scores”

Intercoder Reliability

- Definition: “extent to which two (or more) coders agree on scores”
- Have multiple coders analyze same text using coding scheme
- Assess % of time they assign same scores
- Increase specificity of codebook if reliability is low (less than 80%)

Manual versus Automated Techniques

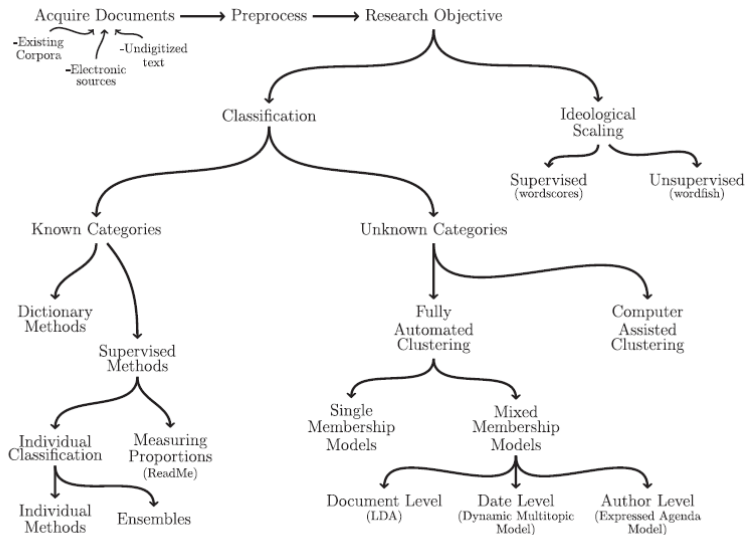
- Text analysis methods vary along a continuum from fully manual to fully automated

Manual versus Automated Techniques

- Text analysis methods vary along a continuum from fully manual to fully automated
- Automated methods can be computationally intensive

Manual versus Automated Techniques

- Text analysis methods vary along a continuum from fully manual to fully automated
- Automated methods can be computationally intensive
- Automated methods make assumptions, but so do manual methods



“Quant” or “Qual”?

Given what you know now, in what ways is content analysis a “qualitative” versus “quantitative” research method?

Hands-On Practice

- There is no explicit lab for this topic
- If you are interested, consider reading the Recommended text for this week, which includes hands-on exercises in R:
<http://tidytextmining.com/>

