



# TEXT ANALYSIS: PRINCIPLES

Miriam Matejova, PhD

November 19, 2024

# Agenda

- 1) Documents as sources of data
- 2) Interpreting documents: content analysis, semiotics, hermeneutics, discourse analysis, rhetorical analysis

# Sources of data

- Letters, diaries, newspapers, magazines, photographs, etc.
- They are “out there” waiting to be assembled and analyzed
  - Not from the process of field work

# How to evaluate documents?

- **Authenticity:** Is the evidence genuine and of unquestionable origin?
- **Credibility:** Is the evidence free from error and distortion?
- **Representativeness:** Is the evidence typical of its kind, and if not, is the extent of its untypicality known?
- **Meaning:** Is the evidence clear and comprehensible?

# Personal documents

- Diaries, letters, autobiographies
- Potential issues:
  - Authenticity
  - *Credibility*: factual accuracy, whether they reflect the true feelings of writer
  - *Representativeness*: the literate and the middle class; women under-represented
  - *Meaning*: abbreviations and codes used by author; damage to documents

# Visual objects

- Photographs, other?
- What do pictures reveal about families?
  - A record of ceremonial occasions and recurring events (weddings, Christmas, etc.)
  - Idealizations, natural portrayals, demystification
  - **Need to probe beneath the surface of an image**
  - Representativeness: what is not photographed?

# Official documents from the state

- State as a source of statistical information, textual material
- Issues with state documents:
  - *Credibility*: is the source biased?
  - *Representativeness*: these documents are often unique

# Official documents from private sources

- Company documents (annual reports, mission statements, etc.)
- Authors may have a particular point of view – depends on their position in the company
- Cannot be treated as fully objective and need to be examined in the context of other sources of data
- *Representativeness*: do we have access to the full set of documents?



# Mass-media outputs

- Newspapers, magazines, TV programs, films, etc.
- Authenticity often difficult – outputs likely genuine but authorship may be unclear (e.g., editorials)
- May also need to be aware of context

# Virtual documents

- Anything on the Internet...
- *Authenticity*: anyone could set up a page
- Credibility?
- Constant flux – can never know how representative the websites are on certain topics
- May need inside knowledge to understand what is being said

# Interpreting documents

- Qualitative content analysis, semiotics, hermeneutics, discourse analysis
- **Qualitative content analysis:** a search for underlying themes
- **Semiotics:** “science of signs”, analysis of symbols to uncover hidden meanings
  - Signifier = points to an underlying meaning; signified = the meaning to which the signifier points
- **Hermeneutics:** about interpretation of text; bringing out the meanings of a text from the perspective of its author.
  - Brings attention to social and historical context within which the text was produced

# Discourse analysis

- An approach to language that can be applied to forms of communication other than talk (i.e., texts)
- Michael Foucault – discourse denotes the way in which a particular set of linguistic categories relating to an object and the ways of depicting it frame the way we comprehend that object
  - The version of an object comes to constitute it
  - A discourse is constitutive of the social world that is a focus of interest or concern

# Rhetorical analysis

- A mode of analysis often used in its own right
- Emphasis on the ways in which arguments are constructed either in speech or in written texts
- Role that various linguistic devices (metaphor, analogy, irony) play in the formulation of arguments