Qualitative Text Analysis

Course Convenor: Aude Bicquelet
(a.j.bicquelet@lse.ac.uk)
Office Hours: Thursday 11:30-13:30
CLASSICAL CONTENT ANALYSIS (I)

Conceptual and Relational Approaches

Week 4
1. Definitions and Milestones

2. Content Analysis and the scientific Method

3. Types of Content Analysis
   - Conceptual
   - Relational
   - Descriptive/Predictive
Definitions & Milestones of Content Analysis Research
Definitions

• Content Analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication (Berelson, 1952: 18).

• Content Analysis is a research technique for making replicable and valid inferences from data to their context (Krippendorff, 1980: 21).
• Content Analysis is a set of methods for systematically coding and analysing data.

• These methods are used to explore explicit and covert meanings in text.

• Whether the research task is exploratory or confirmatory, content analysis involves quantitative analysis.
CA learned its methods from:

- Classification of library books and biblical concordances.
  (i.e.; CA was used centuries ago to organize studies of the Bible through cross-listing of terms and concepts)

- Decryption of meanings of symbols.
  (i.e.; CA was used by T. Young to translate the scripts of the Rosetta Stone through quantification of the occurrences of signs on the stone)
Quantitative Newspaper Analysis
emerged at the beginning of the 20th century.

- ‘Do newspapers now give the news?’ (‘cheap yellow journalism’?)

- By measuring the column inches devoted to particular subject matters, journalists tried to reveal ‘the truth about newspapers’.

- Speed (1893) showed that between 1881 and 1893, New York newspapers dropped their coverage of religious, scientific and literary matters in favour of sport gossip and scandals.

- Other similar studies were undertaken by Wilcox (1900); Fenton (1910) and White (1924).
Propaganda Analysis: Harold Lasswell

- Most influential figure in the development of the systematic study of messages in the 20th century.

- Studied leaflets dropped from balloons and military recruitment posters in his 1927 dissertation: ‘Content Analysis of WWI propaganda techniques.

- In 1949, Lasswell et al. used quantitative methods to explore the content of political messages and the power of propaganda.
The General Inquirer Project

- Harvard in the 1960s was a centre of activity on the study of human communication project.

- An interdisciplinary group led by P. Stone set out to use computer text analysis to content analyze written messages from a linguistic, psychological, anthropological, and communication standpoints.

- They developed the first computer program designed to categorized language.

- The main dictionary was the Harvard Third Psychological Dictionary, which could code 3,564 target words with 83 different tags.
The Comparative Manifesto project

- Established by D. Robertson and I. Budge in 1979.

- Systematically content analysed party manifestos of more than 60 countries, from 1945 to 1983.

- Relies on manual coding rather than computer coding.

- Systematically classify quasi-sentences (units of analysis) into categories of arguments (57).
TYPES OF CONTENT ANALYSIS
Conceptual Analysis

- Simply establish the presence/absence of words/concepts in a text.

Example
Can help determining the authorship of disputed texts.

Mosteller and Wallace (1964) compared the used of the words common to the writings of J. Madison and Alexander Hamilton and concluded that Madison had written 12 of the *Federalist Papers*. 
Conceptual Analysis

- Can be used to trace the ebb and flow of support for political figures over time (Danielson and Lasorsa 1997; de Sola Pool 1952)
- Can be used to trace the evolution of a word/concept across time.

Example
How did the concept of political participation evolve in the media during the 1919 – 2009 period?
Caution: A concept might sometimes be defined differently by various actors or at different times.

Example (1): Participation

- ‘upstream involvement’ (policy-makers/Think Tanks)
- ‘citizens’ input in decision-making’ (NGOs, IGOs)
- ‘strong democracy’ (Academics, i.e. Barber)
- ‘people's right to have their say’ (Blogs, Journalists)
- sometimes referred to by references to ideological standpoints: ‘the populist’/‘Rousseauian view about democracy’.
Example (2): What have been the views of French Reactionaries on direct democracy, from 1789 to 2009?

- At the beginning of the 18th Century French Reactionaries fought against the idea of Universal Suffrage.

- In the 1940s, Reactionaries were against the extension of universal suffrage and fought against: women’s right to vote.

- In the 1960s, Reactionaries opposed De Gaulle’s constitutional change to directly elect the President of the Republic.

- By the end of the 1960s, Reactionaries opposed De Gaulle’s use of the referendum to settle important political issues.
Types of Content Analysis

How to deal with conceptual variations in meanings, or variation of definitions of a concept over time?

**Stratify your sampling**

Identify different key words for;

(a) different definitions of the same concept by several actors;

(b) or different definitions of a same concept across time (*counter-revolutionary* > *reactionary* > *conservative*).
How to locate the presence/absence of word/concept in a corpus?

- Word Counts
- Word Frequency

Calibrate your search

- Use STOP Lists if necessary (to avoid prepositions/conjunctions/articles…)
- Establish thresholds (< 10 >100)
- Use lemmatization/stemming techniques
- Check spelling
- Recode synonyms
### Word Frequency


<table>
<thead>
<tr>
<th>Rank</th>
<th>Term</th>
<th>Frequency</th>
<th>%</th>
<th>Term</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hair</td>
<td>68</td>
<td>46.6</td>
<td>Hair</td>
<td>102</td>
<td>43.6</td>
</tr>
<tr>
<td>2</td>
<td>Eyes</td>
<td>64</td>
<td>43.8</td>
<td>Eyes</td>
<td>92</td>
<td>39.3</td>
</tr>
<tr>
<td>3</td>
<td>Movies</td>
<td>38</td>
<td>26.0</td>
<td>Employed</td>
<td>76</td>
<td>32.5</td>
</tr>
<tr>
<td>4</td>
<td>Brown</td>
<td>37</td>
<td>25.3</td>
<td>Brown</td>
<td>72</td>
<td>30.8</td>
</tr>
<tr>
<td>5</td>
<td>Employed</td>
<td>36</td>
<td>24.7</td>
<td>Likes</td>
<td>55</td>
<td>23.5</td>
</tr>
<tr>
<td>6</td>
<td>Mom</td>
<td>34</td>
<td>23.3</td>
<td>Fishing</td>
<td>46</td>
<td>19.7</td>
</tr>
<tr>
<td>7</td>
<td>Dancing</td>
<td>30</td>
<td>20.5</td>
<td>Movies</td>
<td>43</td>
<td>18.4</td>
</tr>
<tr>
<td>8</td>
<td>Outgoing</td>
<td>27</td>
<td>18.5</td>
<td>Camping</td>
<td>43</td>
<td>18.4</td>
</tr>
<tr>
<td>9</td>
<td>Reading</td>
<td>26</td>
<td>17.8</td>
<td>Sports</td>
<td>42</td>
<td>17.9</td>
</tr>
<tr>
<td>10</td>
<td>Music</td>
<td>23</td>
<td>15.8</td>
<td>Out</td>
<td>37</td>
<td>15.8</td>
</tr>
</tbody>
</table>
Source: Why do you not have your say on European issues? (Bicquelet and Sarmiento-Mirwaldt 2014)

<table>
<thead>
<tr>
<th>Formes</th>
<th>Phi</th>
<th>Khi2</th>
<th>Effectif</th>
</tr>
</thead>
<tbody>
<tr>
<td>democrac</td>
<td>0.28</td>
<td>74</td>
<td>57</td>
</tr>
<tr>
<td>referenda</td>
<td>0.27</td>
<td>69</td>
<td>53</td>
</tr>
<tr>
<td>unit</td>
<td>0.23</td>
<td>53</td>
<td>32</td>
</tr>
<tr>
<td>example</td>
<td>0.22</td>
<td>48</td>
<td>24</td>
</tr>
<tr>
<td>represent</td>
<td>0.22</td>
<td>45</td>
<td>32</td>
</tr>
<tr>
<td>work</td>
<td>0.21</td>
<td>42</td>
<td>19</td>
</tr>
<tr>
<td>europe</td>
<td>0.20</td>
<td>38</td>
<td>64</td>
</tr>
<tr>
<td>national</td>
<td>0.19</td>
<td>35</td>
<td>29</td>
</tr>
<tr>
<td>scotland</td>
<td>0.18</td>
<td>32</td>
<td>14</td>
</tr>
<tr>
<td>kingdom</td>
<td>0.17</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>wales</td>
<td>0.17</td>
<td>28</td>
<td>10</td>
</tr>
<tr>
<td>scottish</td>
<td>0.17</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>recent</td>
<td>0.17</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>hitler</td>
<td>0.16</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>system</td>
<td>0.16</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>century</td>
<td>0.16</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>germany</td>
<td>0.16</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>society</td>
<td>0.16</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>dictator</td>
<td>0.16</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>union</td>
<td>0.16</td>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>small</td>
<td>0.15</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>number</td>
<td>0.15</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>countries</td>
<td>0.15</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>refer</td>
<td>0.14</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>ireland</td>
<td>0.14</td>
<td>21</td>
<td>10</td>
</tr>
<tr>
<td>citizen</td>
<td>0.14</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>welsh</td>
<td>0.14</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>impose</td>
<td>0.14</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>our</td>
<td>0.13</td>
<td>18</td>
<td>64</td>
</tr>
<tr>
<td>use</td>
<td>0.13</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>past</td>
<td>0.13</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>modern</td>
<td>0.13</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>wonder</td>
<td>0.13</td>
<td>17</td>
<td>6</td>
</tr>
</tbody>
</table>
Word Frequency

Source: Why do you not have your say on European issues? (Bicquelet and Sarmiento-Mirwaldt 2014)
Relational Analysis

‘You shall know a word by the company it keeps’ (J. R. Firth, 1957)

Builds on conceptual analysis by examining the relationship among words/concepts in a text.

To understand a concept, look at how it is used.

Study the meaning of each word/concept in all its contexts.
Relational Analysis

How to look for relationships between words/concepts in a corpus?

- Look for Concordances
- Use KWIC programs (Key word in Context)
  Search a text for every use of a particular word or phrase within their contexts
- Use Proximity networks.

Example
To get an understanding of how the words democracy and referenda are used in parliamentary debates about future EU enlargements we can run a simple KWIC.
Source: Why do you not have your say on European issues? (Bicquelet and Sarmiento-Mirwaldt 2014)
Source: Why do you not have your say on European issues? (Bicquelet and Sarmiento-Mirwaldt 2014)
**Descriptive Content Analysis**

Describes a particular message pool in almost archival fashion. Researchers working in this vein limit their conclusions to the content being studied.

**Example**


Findings identify a rise in sexual content, notably in shows featuring teenagers. *(Whereas the origins of the project were clearly in a concern over the effects of sexual content on viewers, the project never overstate its conclusions – they are purely content based)*
The goal of PCA is to make predictions about the outcomes of a message. By measuring key characteristics of the messages, researchers aim to predict audience responses to it.

This necessitates the merging of CA with other methods (i.e. surveys or experiments).

Example
Through a blending of content analysis and public opinion research, Hertog and Fan (1995) found that print news coverage of three potential HIV transmission routes (toilets, sneezing and insects) was significantly related to public beliefs about those routes (expressed in polls).
Many models that have been designed to undertake various types of CA, were very often influenced by particular fields of research. i.e.;

- **Relational CA** reflects the cognitive concerns of psychologists, linguists, social-psychologists.

- **Conceptual CA** reflects socio-political concerns of historians, political theorists and sociologists.

- **Predictive CA** is very often used in market research, management, Media and Communication studies
Types of Content Analysis

Word Counts are Only a Start

- Help to look at complex data, explore for central themes and make comparisons across groups.

- We may want to test hypotheses that emerge from this simple first step (Classical CA)

  sort words into categories and treat them as variables (Dictionary-based Approaches)

  examine the relationship among words across a set of texts (Semantic Network analysis)

  build codebooks for inductive approaches (Summative Content Analysis)
A large number of programs are available to facilitate enumerative Content Analysis.

- Wordstat
- T-lab
- TEXTPACK

Second-generation qualitative data management software also have content analysis capacity

- Atlas-ti
- MAXQDA
- NVivo
Useful Resources