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

Lecture Outline

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).

Definitions

- 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.

Milestones of Content Analysis Research

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)

Milestones of Content Analysis Research

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.

Milestones of Content Analysis Research

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

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

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?

Conceptual Analysis

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'.

Conceptual Analysis

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>*).

Types of Content Analysis

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

Example: What Women and Men *Offer* in Personal Ads - Source: Russell and Ryan (2010:202)

Rank	Women			Men		
	Term	**Frequency**	**%**	**Term**	**Frequency**	**%**
1	Hair	68	46.6	Hair	102	43.6
2	Eyes	64	43.8	Eyes	92	39.3
3	Movies	38	26.0	Employed	76	32.5
4	Brown	37	25.3	Brown	72	30.8
5	Employed	36	24.7	Likes	55	23.5
6	Mom	34	23.3	Fishing	46	19.7
7	Dancing	30	20.5	Movies	43	18.4
8	Outgoing	27	18.5	Camping	43	18.4
9	Reading	26	17.8	Sports	42	17.9
10	Music	23	15.8	Out	37	15.8

Word Frequency

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

Mots présents significativement

Formes	Phi	Khi2	Effectif
democrac	0,28	74	57
referenda	0,27	69	53
unit	0,23	53	32
example	0,22	48	24
represent	0,22	45	32
work	0,21	42	19
europe	0,20	38	64
national	0,19	35	29
scotland	0,18	32	14
kingdom	0,17	29	19
wales	0,17	28	10
scottish	0,17	28	12
recent	0,17	27	15
hitler	0,16	25	13
system	0,16	25	28
century	0,16	25	10
germany	0,16	25	10
society	0,16	25	9
dictator	0,16	25	12
union	0,16	24	14
small	0,15	23	11
number	0,15	23	11
countries	0,15	22	15
refer	0,14	21	9
ireland	0,14	21	10
citizen	0,14	20	10
welsh	0,14	19	8
impose	0,14	19	8
our	0,13	18	64
use	0,13	17	21
past	0,13	17	13
modern	0,13	17	6
wonder	0,13	17	6

Word Frequency

Word Frequency Query

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.

Relational Analysis

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


```
 Contextes d'utilisation

dition of parliamentary      democracy    that none of the other
hink that the mother of      democracy    and the country that pr
iament or parliamentary      democracy  . indeed, at the moment
    of those 40 genuine       democracies  at least half use refe
ablished parliamentary        democracies  is small$
f british parliamentary       democracy    and, secondly, that the
e said, in effect, that       referenda    are used only by dictat
y american states where       referenda    are resorted to without
a dictatorship, just as       referenda    were misused by hitler.
e is apparently keen on       referenda    and, on the example of
 and, on the example of       referenda    as used by germany in t
hould bear in mind that       referenda    are now illegal under t
tional referenda in the       united    kingdom would represent a
ould remain part of the       united    kingdom, on devolution in
st of the states in the       united    states of america frequen
 it be in scotland, the       united    kingdom or europe. in the
are we to ensure within       united    kingdom laws that that wi
             his sole         examples   were from nazi germany a
I do not think that the       example   of germany, any more tha
n referenda and, on the       example   of referenda as used by
uth has been sending          representatives   there for 700 year
 elections and public         representation   is almost the only
ibilities by elected          representatives   and with no ill ef
r of civil servants who       work   in the welsh office, or th
 whatever capacity they       work , it is a matter of burning
erest and then to start       working   out how we should alter
 that none of the other       european   countries can claim$
used. people in eastern       europe   claim that they have a de
who have led the van of       european   civilisation$
that does not allow its       nationals   who are resident$
```

Relational Analysis

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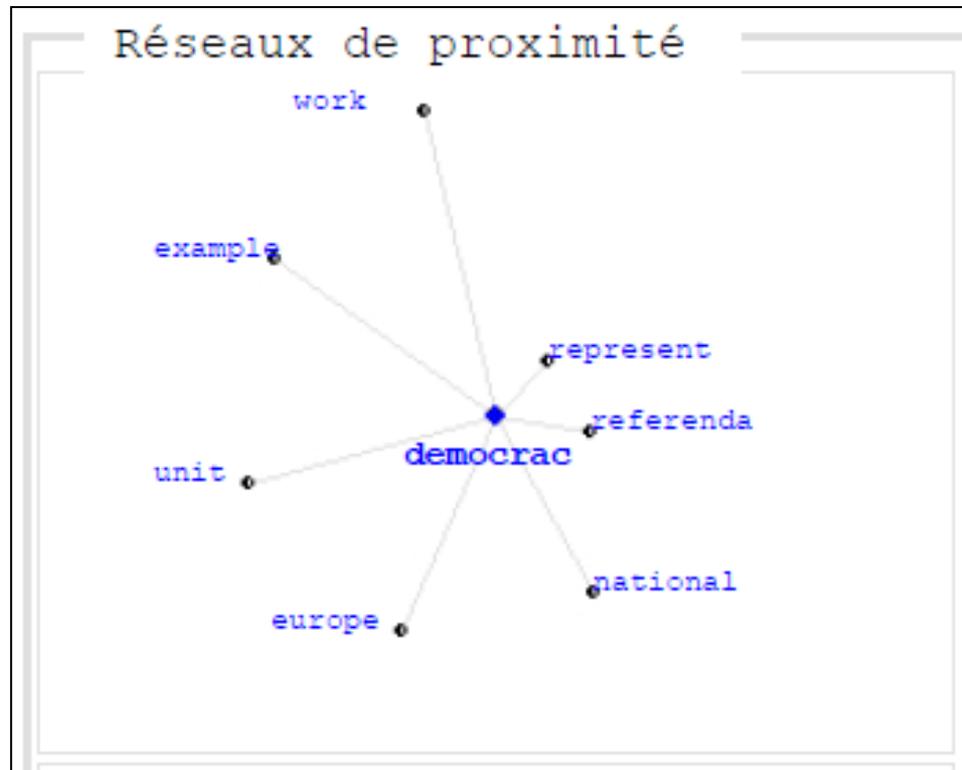
Descriptive Content Analysis

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
On-going research, tracking sexual content in U.S. television programming by Kunkel et al., 2001.

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)

Predictive Content Analysis

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).

Types of Content Analysis

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)

Computer Programs

➢ 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

- Grbich, C. (2013) *Qualitative Data Analysis* (chapter 16) London: Sage.
- Franzosi, R. (2008) *Content Analysis*. London: Sage. A four-volume set in the Sage Benchmarks in Social Methods series. These provide a comprehensive and detailed history and discussion of techniques linking to a range of discipline.
- Kracauer, S. (1952). 'The challenge of quantitative content analysis', Public Opinion Quarterly, 16(4): 631-642.
- Krippendorff, K. (2004). 'Conceptual Foundation', in: Content Analysis: An Introduction to Its Methodology, 2nd edit., Thousand Oaks, CA: Sage, pp. 18-43.
- Kaplan, A. (1943). 'Content analysis and the theory of signs', Philosophy of science, 10(4): 230-247.
- Markoff, J., Shapiro, G. and Weitman, S.R. (1975). 'Towards the integration of content analysis and general methodology, in: Heise, D.R. (ed.). Sociological Methodology, San Francisco, CA: Jossey-Bass, pp. 1-58.
- Mayring, P. (2000). 'Qualitative Content Analysis', Forum: Qualitative Social Research, 1(2), Art. 20.