# Day 5: Classical quantitative content analysis

Kenneth Benoit

Essex Summer School 2012

July 13, 2012

# Thematic and "Classic" content analysis

- Key feature: use of "human" coders to identify themes in a text, possibly through the application of a coding scheme

- Key implementation element occurs *prior to* the construction of a quantitative matrix of information, since the human application of codes or themes determines the dimensions of the matrix

- Human decision-making is the central feature of coding decisions, not a computer or other mechanized tool

- Following the construction of the matrix, all standard quantitative techniques may be applied

# What are "themes"?

- Goal: to identify a limited number of themes which adequately reflect the textual data being analyzed
- "Themes" can mean:
    - specific topics
    - types or modes of argument
    - expressions of sentiment
    - *etc – this is very open-ended*
- "A theme captures something important about the data in relation to the research question and represents some level of patterned response or meaning within the data set."

# Data types

- Texts you've "created" yourself
    - Interview transcripts
    - Focus group transcripts
    - Open-ended survey questions
- "Natural" texts
    - speeches
    - documents
    - essays
    - literature
- Conversations

# Differences from CCA

- Definition: TA is qualitative analytic method for: *Identifying, analysing and reporting patterns (themes) within data. It minimally organises and describes your data set in (rich) detail. However, frequently it goes further than this, and interprets various aspects of the research topic.* (Braun and Clarke, 2006, p.79)

- Thematic analysis focuses on *meaning* – a more discursive interpretation

- Thematic analysis is more exploratory and iterative in the process of developing the procedure at the same time that it is carried out

- Analyst may alter and modify the thematic analysis as new ideas develop, even to the extent of adjusting earlier codings

# Differences from CCA

- Content analysis is far more methodologically pre-defined: it involves the pre-definition of mutually exclusive categories to text units, and is often accompanied by quantitative analysis
- Generally considered important to be very familiar with the texts — possibly these are "texts" that the researcher has generated her/himself
- Much less emphasis on *unitization* in thematic analysis – this is typically open-ended

# Steps

1. Prepare the data for analysis – may involve transcription
2. *Read* the texts:
   2.1 Initial reading, to survey possible themes, getting a sense
   2.2 Re-read the text, using annotation
3. Assign initial themes
4. Examine initial themes and attempt to define them concretely
5. Re-examine the text and code according to the themes
6. Construct the final themes
7. Report the themes – and this does not preclude constructing a quantitative summary

# Examples of quantitative methods for thematic analysis

- Descriptive summaries
- Associations, cross-tabulations
- Over-time measures of an index, possibly correlated with another measure
- Structural equation models
- Dimensional methods such as factor analysis or MDS
- Network analysis
- Cluster analysis
- Other forms of scaling

# Components of thematic/manual coding approaches

Unitizing
: The systematic distinguishing of segments of text that are of interest to the analysis.

Sampling
: Choice (and justification of the choice) of text units to sample, from population of possible text units.

Identifying Themes/Coding
: Classifying each coded unit of text from the sample according to the pre-defined category scheme.

Summarizing
: Reducing the coded data to summary quantities of interest.

Inference and reporting
: The final steps wherein the analyzed results are used to generalize about social world, and communicating these results to others.

# Sampling Texts

- (Mainly we have already covered this on Days 3–4)

- In hand-coded schemes, sampling may be more deliberate

- For the Comparative Manifesto Project, the case study for this topic, "sampling" consists of selecting all texts of a particular class

# Coding and Marking up Texts

- The key step in transforming raw texts into representations that can be analyzed

- Involves reducing and quantifying the data into discrete categories

- Requires a pre-defined scheme with rules for how these should be applied

- This process may be exploratory/iterative/inductive (in the case of thematic analysis)

- Question in designing the scheme is to maximize on the precision-accuracy-reliability frontier

- There should be some sort of settled scheme at the end, however, that could also be subject to a reliability check

# Summarizing Themes

- Involves characterizing the content and (relative) frequency of themes
- Examples

  Theme/category frequencies Coded category frequency measures, such as the proportion of times "economy" is mentioned in a speech, or the proportion of mentions of the environment

  Type/token measures Frequency tabulations of token types and their frequencies within coded themes — the words used in particular themes

  Range/variance Here we might be interested in the total number or the spread or variance of themes used in particular documents or by particular speakers

- May also involve scales or indexes constructed from summary information

# Summarizing: Example

Table 4
Coding scheme identifying meaningful categories of keywords

| Keyword category | Examples of keywords |
| --- | --- |
| Greetings | Regards, thanks, hello, welcome, [all the] best, regs ( = regards), |
| Support | Support, love, care, XXX, hugs |
| Feelings | Feel, scared, coping, hate, bloody, cry, hoping, trying, worrying, nightmare, grateful, fun, upset, tough |
| Health care staff | Nurse, doctor, oncologist, urologist, consultant, specialist, Dr, Mr |
| Health care institutions and procedures | Clinic, NHS, appointment, appt |
| Treatment | Tamoxifen, chemo, radiotherapy, brachytherapy, conformal, Zoladex, Casadex, nerve [sparing surgery] |
| Disease/disease progression | Cancer, lump, mets, invasive, dying, death, score, advanced, spread, doubling, enlarged, slow, cure |
| Symptoms and side effects | Hair, sick, scar, pain, flushes, nausea, incontinence, leaks, dry, pee, erections |
| Body parts | Breast, arm, chest, head, brain, bone, skin, prostate, bladder, gland, urethra, |
| Clothing and appearance | Nightie, bra, wear, clothes, wearing |
| Tests and diagnosis | PSA, mammogram, ultrasound, MRI, Gleason, biopsy, samples, screening, tests, results |
| Internet and web forum | www, website, forums, [message] board, scroll |
| People | Her, she, I, I've, my, wife, partner, daughter, women, yourself, hubby, boys, mine, men, dad, he |
| Knowledge and communication | Question, information, chat, talk, finding, choice, decision, guessing, wondering |
| Research | Study, data, trial, funding, research |
| Lifestyle | Organic, chocolate, wine, golf, exercise, fitness, cranberry [juice] |
| Superlatives | Lovely, amazing, definitely, brilliant, huge, wonderful |

[ ]—square brackets are used to give commonly associated word showing a word's predominant meaning.
( = ) — rounded brackets and = sign used to explain a term's meaning.

(source: Seale et al 2005)

# Inference and Reporting

- This involves drawing conclusions from the research, and these conclusions will depend on the *validity* established by the research design

- Reporting means communicating the results in a clear and relevant fashion. (This can be challenging – see for instance the Schonhardt-Bailey article.)

- No iron-clad rules here – use your discretion as applied to a particular case