CONTENT ANALYSIS

By:
A. Warm Body,
Distinguish Professor

Coding the content of documents (like behavior coding)

- any technique for making inferences by identifying special characteristics of messages (written or oral)
  - artifacts of social communications
- information is condensed (classified) and made systematically comparable by applying a coding scheme
any kind of written document
• field notes from participant observation, letters, novels, transcripts of recorded communications (such as T.V shows, interviews, etc.)

The steps in content analysis:
• fully describe the phenomenon to be studies (e.g. portrayal of the elderly in the media)
• select the media that will be used for data
• derive coding categories
  – choose categories, e.g. status of character, physical attractiveness, context, etc.
  – count presence or absence of a category
  – place each piece into one of many categories (forced choice)
• decide on a sampling strategy --you can’t count it all
• train the coders/raters (reliability is important)
• analyze the data (%’s, compare means and variances?)
example: Laswell, et al (1949);

- Symbol analysis employed during World War II
  - Newspaper content was studied to count the appearance of certain symbols, such as Jews, Stalin, democracy, Russia
  - An affective dimension was added—indulgent, supportive, neutral, etc.
    » Sometimes further divided into those stressing strength, goodness, morality, weakness, immorality, etc.
  - Nowadays this can be done with a computer

- Sampling strategy
  - Newspapers only
  - Sampled every 10th one from an alphabetical list
  - Control geographical location, political and economic orientation, ethnic group

In general, we have to decide—

- What gets counted (words, pictures?)
- What levels of analysis (categories, amounts?),
- What coding frames (every 10th page, every other sentence?)
Types of COMMUNICATION COMPONENTS:

- message,
- sender,
- audience

message components
- explicit themes,
- emphasis on various topics,
- amount of space or time devoted to topics,
- etc.
Putting numbers on prose: counting vs. rating

- counting--little judgment, less discovery
- rating--much judgment, more discovery
- usually have a combination

**count** angry words in text or **rate** anger in sentences or paragraphs

- “**count**” angry words, what **is** an angry word???
- end up using some judgment unless it’s totally spelled out
- and the more it’s spelled out, the more you lose discovery
e.g. Anger

• **count** these words:
  – fed up
  – irritated
  – disgusted
  – etc.

• or **rate** these sentences:
  – I find the idea to be distasteful.

very warm  very angry
0 1 2 3 4 5

examples of elements to count:

• items,
• words
• sentences
• paragraphs,
• characters,
• semantics
• concepts,
• themes,
Items

- represents the whole unit of the sender's message
- may be an entire book, a letter, speech, diary, newspaper, or even an in-depth interview.

Words.

- smallest element
- least judgment
- generally results in frequency distributions
Sentences

• definitely more judgment than words but less than paragraphs, etc.

Paragraphs

• difficulties result in attempting to code and classify
• poor consistency among writer on how to write a paragraph
Semantics

- meanings
  - of overall sentence, paragraph, etc.
- requires a lot of judgment

Characters (persons)

- count the number of times a specific person is mentioned
Concepts

- involve words grouped together into conceptual clusters (ideas)
- e.g. a conceptual cluster may form around the idea of deviance.
  - Words such as crime, delinquency, money laundering, and fraud might cluster around the conceptual idea of deviance
- leads toward more latent than manifest content, more rating, judgment, etc. although word clusters could be simply counted

Themes

- broader than a concept (almost like a mood)
- can be made up of many concepts
- must further specify the unit -- theme of each sentence, each paragraph, the whole book???
Combinations of Elements

- e.g. Berg's (1983) subjective definitions for Jewish affiliational categories

used combination of word, sentence, and paragraph elements

- lifted definition components from interview transcripts cutting across elements to formulate interviewee’s definition

- each definition annotated with the transcript number
example--Interview #60:

• ORTHODOX
  – Well, I guess, Orthodox keep kosher in [the] home and away from home. Observe the Sabbath, and, you know..., actually if somebody did [those] and considered themselves an Orthodox Jew, to me that would be enough. I would say that they were Orthodox.

• CONSERVATIVE
  – Conservative, I guess, is the fellow who doesn't want to say he's Reform, because it's objectionable to him. But he's a long way from being Orthodox.

• REFORM
  – Reform is just somebody that, say they are Jewish because they don't want to lose their identity. But actually want to be considered a Reform, 'cause I say I'm Jewish, but I wouldn't want to be associated as a Jew if I didn't actually observe any of the laws.

• NONPRACTICING
  – Well, a Nonpracticing is the guy who would have no temple affiliation, no affiliation with being Jewish at all, except that he considers himself a Jew. I guess he practices in no way, except to himself.

Three major approaches to categorizing in a coding system:

• common classes,
• special classes, and
• theoretical classes
1. Common classes.

- used by virtually anyone in society
  - (for example, age, gender, mother, father, teacher, boss, lover, etc.)
- essential in assessing whether certain demographic characteristics are related to patterns that arise from other coding

2. Special Classes.

- colloquial categories
- includes jargon of various professions, e.g. petty larceny vs. that other category
3. Theoretical Classes.

- those that emerge in the course of analyzing the data
- category labels generally borrowed from special classes (e.g. psychojargon)
- their substance is grounded in the data
- not immediately knowable until observers spend considerable time with the content

"in vivo codes" vs."sociological constructs"

- in vivo codes are literal
  - actual words
- sociological constructs are formulated by the analyst
  - e.g. "professional attitude," "family oriented," "obsessive workaholic," "educationally minded," might represent examples of sociological constructs
  - may add breadth and depth
In General,

• you can apply an established code but
  – you lose some of the individuality in the data set

• or you can develop your own code but
  – people may accuse you of post hoc or circular reasoning in this case

how to identify categories in the data and apply them--at the same time without being “too circular” in your reasoning

• OPEN CODING
• here’s a quote--don’t copy it
Inexperienced researchers, although they may intellectually understand the process described so far, usually become lost at about this point in the actual process of coding. Some of the major obstacles which cause anguish include the so-called true or intended meaning of the sentence, and a desire to know the real motivation behind a clearly identifiable lie uttered by a subject. If the researchers can get beyond such concerns, the coding can continue. For the most part, such concerns are actually irrelevant to the coding process, particularly with regard to open coding, the central purpose of which is to open inquiry widely. Although interpretations, questions, and even possible answers may seem to emerge as researchers code, it is important to hold these as tentative at best. Contradictions to such early conclusions may emerge during the coding of the very next document. The more thorough analysis of the various concepts and categories will best be accomplished after all of the material has been coded” (Berg, 1990).
• “Inexperienced researchers, although they may intellectually understand the process described so far, usually become lost at about this point in the actual process of coding. Some of the major obstacles which cause anguish include the so-called true or intended meaning of the sentence, and a desire to know the real motivation behind a clearly identifiable lie uttered by a subject. If the researchers can get beyond such concerns, the coding can continue. For the most part, such concerns are actually irrelevant to the coding process, particularly with regard to open coding, the central purpose of which is to open inquiry widely. Although interpretations, questions, and even possible answers may seem to emerge as researchers code, it is important to hold these as tentative at best. Contradictions to such early conclusions may emerge during the coding of the very next document. The more thorough analysis of the various concepts and categories will best be accomplished after all of the material has been coded” (Berg, 1990).
IF YOU DEVELOP YOUR OWN CODE

• YOU LET THE DATA SPEAK FOR THEMSELVES

GROUNDLED THEORY IS

• an interplay of experience, induction, and deduction
  – 1. initial systematic discovery of the theory from the data
  – 2. application of the theory to the data
Remember this--

NOW APPLY IT TO CONTENT ANALYSIS

FACT

THEORY

READ TEXT

READ AGAIN

(induction

deduction

revision

expect something)
Researcher examines newspaper orientations toward changes in seatbelt law!

- First, read each article and ask which ones favor and which against
- Was the decision to label an article pro or con based on the use of certain terms, on presentation of specific study findings, or because of statements offered by particular characters (who?)?
- Were the article's positions more clearly indicated by their manifest content or by some undertone?

Answers to these questions lead to inductive categories in which to slot various units of content

And deductive application of the categories leads to more inductive categories
Quotes (don’t copy)

- “If investigators have begun with specific empirical observations, they should attempt to develop explanations grounded in the data (grounded theory) and apply these theories to other empirical observations (deductive reasoning).”
- “If they are attempting to test theory derived from previous research and previous inductive reasoning, their theoretical orientation should suggest empirical indicators of concepts (deductive reasoning).”

Remember the hermeneutic circle or the hypothetico deductive approach---

specification of the content characteristics

0

induction

deduction

revision

application of rules for counting those characteristics
One way to avoid the accusation that you’re engaging in post hoc or circular reasoning is to:

- divide the data set in half
- develop the code on one half
- apply it to the other half

Another way

- totally replicate the study
- “you think I made it up or imagined it, watch I’ll do it again!”
SO GROUNDED THEORY IS AN:

- interaction of two processes:
  - specification of content characteristics
  - application of explicit rules for counting those characteristics

As you apply the coding rules, you often find more characteristics to code—OR LESS (SEE NEXT EXAMPLE)

specification of the content characteristics

0

induction

deduction

revision

DISCOVER MORE CHARACTERISTICS

application of rules for counting those characteristics
Recent study evaluates effectiveness of Florida delinquency program!

- identified several themes on intake sheets
- set up a tally sheet,
- start to categorize criminal offenses declared by arresting officers
- discovered two distinct classes of crime

1. shoplifting, petty theft, and retail theft all referred to stealing of some type of store merchandise, usually not exceeding $3.50 in value

2. the similar term petty larceny was used to describe the taking of cash whether from a retail establishment, a domicile, or an auto

- The intake sheets at first suggested 4 categories (shoplifting, petty theft, retail theft, petty larceny)

---

Strauss (1987) says—"believe everything and believe nothing" while undertaking open coding

- he lists the following 4 guidelines:
1. Ask the data a specific and consistent set of questions.

- most general
  - What study are these data pertinent to?
  - What was the original objective of the research study?
  - Original purpose of a study may not be accomplished, and an alternative or unanticipated goal identified in the data.

For example,

- Pearson (1987) -- evaluation of a New Jersey intensive probation supervision program -- original aim = demonstrate cost effectiveness
- Objective indicators failed but indirect indicators suggested success
  - Reports from relatives of probationers about changes in attitudes
  - Spouse reported that her husband now sent child-support payments
  - Parent reported that child now showed personal responsibility by doing household chores
Thus, Pearson (1987) points to an unanticipated benefit from the program.

- keep the original study aim in mind and remain open to multiple or unanticipated results that emerge

Strauss (1987) also says--
2. Analyze the data minutely.

- in the beginning, more is better
- like a funnel
- the wide end = many categories, incidents, interactions, etc.
- coded minutely during open coding
- let patterns emerge or reveal themselves
Later,

- more systematic coding can be accomplished, building from the numerous elements that emerge
- The question is when to stop open coding and move on to the speedier, more mechanical coding phase?
- Typically, as researchers minutely code, they eventually saturate the document with repetitious codes.
- as this allows researchers to move more rapidly through the documents, it is usually time to move on.

3. Strauss also says frequently interrupt the coding to write a theoretical note.

- called ‘grounded theory’
- a comment in the document triggers ideas--jot down a note
- otherwise it'll be forgotten
- keep a record of where in each document similar content triggered the idea
For example, study on adolescents' involvements with alcohol, crime, and drugs revealed youths speaking about drugs and criminal activities as though they were partitioned categories (Carpenter et al, 1988). Notes scribbled during coding later led to theories regarding youth’s drug-crime culture.

4. Finally Strauss says that in open coding never assume the analytic relevance of any traditional variable such as age, sex, social class, and so on, until the data shows it to be relevant.
   • even these more mundane variables must "earn their way into grounded theory"
   • What are the study data pertinent to?
   • What are you trying to find out, trying to say??
If researchers are interested in gender differences, they begin by assuming that gender might be analytically relevant, but if the data fail to support this assumption, the researchers must accept this result.

All of our discussion of COMMUNICATION COMPONENTS

- message
- sender,
- audience

has centered on “message”--let’s turn to sender
MESSAGE VS. SENDER

• **What** is said?
  versus
• **How** is it said?

sender

• linkages between message and attributes of the sender are often slight **but**
  – can gain some impression from numerous examples,
    » e.g. recordings, transcripts including literal representations of pauses, mispronounced words, grammatical errors, slang, etc.
now the last --COMMUNICATION COMPONENT

- message,
- sender,
- audience

audience

- analyze content from the audience as well as from the sender of the message
  - e.g. Pornography and Television Violence Commissions
    » write a description or essay about the program you just watched or
    » interview and content analyze the interview transcription
- Also try to get a feel for who is consuming the message
  - when I see ads for feminine napkins on a program I assume the network has calculated that a lot of women will be watching the program--similarly its aimed at men if shaving products are emphasized
Sampling Content

In general, content analysis samples:

• 1. sources (newspapers? which ones? or magazines? which ones?)
• 2. dates of sources (time periods)
• 3. units or chunks of content within the source (headlines? editorials? cartoons?)--or you can simply divide the newspaper into inches, lines, words, quadrants, and sample from within these units
criteria of selection of sources and components:

- should reflect all relevant aspects of the messages and retain, as much as possible, the exact wording used
  - quote: “not merely the arbitrary or superficial application of irrelevant categories.”
  - a priori operational definition of content to be included and excluded
    - try to eliminate analysis in which only content supporting investigator's hypotheses is examined
  - categories also emerge in the course of developing these criteria

SAMPLING STRATEGIES (applicable to all forms of research)

- 1. simple random sampling,
- 2. systematic sampling,
- 3. purposive sampling
1. Simple Random Sampling

- rely on chance to generate a representative sample
- draws subjects from an identified population in such a manner that every unit has precisely the same chance (probability) of being included in the sample
Generalizing

The Population

This is used to make statements about this

The Sample

If you pick only the prettiest flowers from your garden,
If you pick only the prettiest flowers from your garden, is that a random sample?

Would the flower you picked represent your garden?
2. Systematic Sampling

- get a printed list of population members
- sample every Nth name
- sampling interval = divide number of people desired into the full population size
  - e.g. sample of 80 from a population of 2560 means you sample every 32nd person (2560/80 = 32)
  - be sure to start at a random point in the list

One type of systematic sampling = Stratified Sampling

- ensure that a certain segment of the population is correctly represented
- population is divided into subgroups (strata) and the % of the sample taken from each stratum equals its representation in the population
  - e.g. if SIU has 55% men and 45% women, that’s the proportion we put into our sample
3. Purposive Sampling

- researchers use their special knowledge or expertise to select subjects who represent the population
- e.g. use the Eysenck Personality Inventory to select neurotic subjects for your sample
  - must be careful generalizing

Example of sampling in content analysis: Graber (1971)

- She developed a complex sampling scheme for examining newspapers
  - country divided into regions
  - cities in each region divided into population size groupings: over a million, a million to half a million, and fewer than half a million
  - 3/4 of the sample drawn from the most populous states in each region to reflect “voting power”
  - half were selected from states in which Democratic party was dominant and half from those where Republican party was dominant
  - representative sampling of monopoly vs. competitive newspapers
- then all campaign stories were coded
Quantitative vs. Qualitative Analysis

hot debate over whether analysis should be quantitative or qualitative

- Quantitative = tally sheets, specific frequencies of relevant categories
- Qualitative = ratings, examine ideological mind sets, themes, topics, symbols
  while “grounding” such examinations in the data
REMEMBER LEVELS AND UNITS OF ANALYSIS (this sometimes affects the extent to which you can be quantitative)

• levels of prose:
  – words, phrases, sentences, paragraphs, sections, chapters, books, writers, ideological stance, subject topic, etc.

• television programs:
  – segments between commercials, entire program, entire prime time periods, etc.

one side says analysis should be "objective, systematic, and quantitative."

• but quantitative analysis emphasizes "the procedure of analysis," rather than the "character of the data available"
also “quantitative content analysis results in a somewhat arbitrary limitation in the field by excluding all accounts of communications not in the form of numbers, or those which may lose meaning if reduced to a numeric form (definitions, symbols, detailed explanations, and so forth)”

Other proponents, e.g. Smith (1975), suggest that some blend of both quantitative and qualitative analysis should be used.
Smith (1975: p. 218) explains that he has taken this position "because qualitative analysis deals with forms and antecedent-consequent patterns of form, while quantitative analysis deals with duration and frequency of form."

Another hot debate--

- **Manifest versus Latent Content Analysis**
- **manifest content** (those elements that are physically present and countable)
- **latent content** (interpretive reading of the symbolism underlying the physically presented data)
  - e.g. examine an entire speech to rate how "radical" it was, or a novel could be considered in terms of how "violent" it was
manifest content = "surface structure,"

latent content = "deep structure" (meaning)

Holsti (1969) has tried to resolve this debate:

• “It is true that only the manifest attributes of text may be coded, but this limitation is already implied by the requirement of objectivity. Inferences about latent meanings of messages are therefore permitted but . . . they require corroboration by independent evidence”...
also “should offer detailed excerpts from relevant statements that serve to document the researchers' interpretations” ...
“A safe rule of thumb to follow is the inclusion of at least three independent examples for each interpretation”

Blending Manifest and Latent Content Analysis

- Take note of apparent presence of a concept and
- report the frequency at which an indicator appears to suggest or reflect magnitude but
- be cautious about generalizing to actual magnitude
- e.g., If indicators of "positive attitudes toward shoplifting," appear fifty times in one, and twenty-five times in another, do not claim that the first is "twice as likely to shoplift."
- stay close to the “facts”
Content Analysis is continued with some examples and discussion of rating techniques in the next section of the course

: )