Methods of Analysis
Content Analysis
Stephen Petrina (2016/2019)

We have great faith in words. But we don't always say what we mean or mean what we say.
(Galloway, 1971, p. 228)

Content analysis refers to two common practices, excluding concordances of sacred texts. Chronologically, the first is analysis of potential curriculum texts to establish *manifest* content and meaning. The second is analysis of spoken and written texts to establish *latent* content and meaning. Discerning manifest and latent content had long been common to art, criminology, medicine, philology, and theology. Most definitively in the first half of the twentieth century, these distinctions were elaborated in Freud's theory of dreams. "We must make a contrast between the *manifest* and the *latent* content," he (1899/2010) famously reasoned. This "theory is not based on a consideration of the manifest content of dreams but refers to the thoughts which are shown by the work of interpretation to lie behind dreams" (p. 160). Freud's key finding is that manifest expressions signify latent processes, which demand interpretation of latent content.

Content analysis (CA) was profiled in Lasswell's (1935) review of research on the press (pp. 185-206). He concluded that "the study of the space content of the press is useful insofar as it furnishes clues to the distribution of attention" to one issue over another (p. 187). This says nothing about readers' attention and nor does it establish deeper meanings. In 1941, he refocused the method of "content analysis" and demonstrated how counting and coding words can facilitate or index meaning (1941a, pp. 459, 461). At the same time, he turned to Freud's theory. In the first instance, Lasswell (1941b) noted, "we describe manifest content; and in the second, we interpret according to latent meaning" (p. 2). For Lasswell, latent meaning given to manifest content is, in other words, interpretation (Ahuvia, 2001, p. 141). A goal of CA, like semantic analysis, is to saturate data and exhaust meaning, whether through manifest & latent, literal & figurative, denotative & connotative, or explicit & implicit meaning or through text & subtext.

CA includes as its scope of content, what is said (manifest), what is otherwise said, intended, or meant (latent), and what is not said (censored). *Content* in CA extends from information to meanings of any communication or expression. More specifically, *content* refers to samples of text and "content words" or contentives (i.e., nouns, verbs, adjectives, adverbs) along with functives (i.e., pronouns, prepositions, articles, conjunctions, etc.) (Bowman, 1960, p. 47; Tausczik & Pennebaker, 2010, p. 29). CA is applicable to ordinary or natural *and* specialized or technical language. Counting establishes magnitude of content for coding and categorizing.

If CA is a method to "code text into categories and then count the frequencies of occurrences within each category," then what is a category and code (Ahuvia, 2001, p. 139)? *Category* in CA refers to a brief expression or theme for classifying and coding text and subtext. A category captures a pattern and organizes data and codes. *Code* refers on one level to a degree, valence, value, or inflection (e.g., - / +) and on another level to a label or tag "for assigning units of meaning" (e.g., pessimism & optimism) (Miles & Huberman, 1994, p. 56). Analysts distinguish among descriptive codes (descriptors), pattern codes, and analytic codes (concepts) (p. 57).

For example, a sample of text could be a User Agreement or Terms of Use for a social media platform (e.g., Instagram), a category could be Statements on Ownership, and descriptive codes might be License and Share, which could be counted. An analytic code, for establishing latent content, might be Appropriation. Along with a stratification of meaning is the communication or text itself, a union of the manifest and latent. In this example, a Terms of Use Agreement is a particular hybrid of legal and technical text. Four Cs here are common across data analysis while the fifth, counting, is specific to CA:

Conceptualize ←→ Collect ←→ Count ←→ Categorize & Code ←→ Conceptualize

Just as dreamers self-censor reports of dreams, authors often self-censor texts or are somehow censored. Psychoanalysts note that latent content, through various mechanisms, is suppressed and transformed into manifest content. Content analysts rarely make this point but most remain committed to establishing latent content. Although CA is defended as a method of text analysis that facilitates precision and fine-grained description, latent content interpretation necessarily involves imputing meanings, intentions, or motives. Some analysts advocate guarding against imputing "ideas, assumptions, or meanings to ambiguous responses, statements, or behavior" (Elkins & Simeon, 1979, p. 138). Others argue that "no clear line can be drawn between properties 'in' a work and those 'imputed to' it through interpretive procedures" (Lamarque, 2000, p. 96). Author and text remain open to interpretation and of course some interpretations will be more critical or contentious than others.

Latour's (1991/1993) concern with critical interpretation is that it professes "to discern the real motives beneath appearances." This "tradition of the human sciences," he says, "no longer has the privilege of rising above the actor by discerning, beneath his [her or their] unconscious actions, the reality that is to be brought to light" (p. 44). Instead, Latour (2004) asks: "What would critique do if it could be associated with *more*, not with *less*, with *multiplication*, not *subtraction*?... That is, generating more ideas than we have received" (p. 248). This reiterates inclusivism, which was popular in literary theory in the 1950s: "no single interpretation can exhaust the rich system of meaning-potentialities represented by the text" (Hirsch, 1960, p. 471).

Distinctions between what is said and what is implicated are often framed as distinctions between "author meaning" or "speaker meaning" and "sentence meaning." Hirsch (1960) defaults to author meaning as a guiding principle of interpretation: "to verify a text is simply to establish that the author probably meant what we construe his [her or their] text to mean.... It is natural to speak not of what a text says, but of what an author means" (pp. 478-479).

CA reduces to two symmetrical questions: Do people mean what they say? *and* Do people say what they mean? Notably, "orthodox Freudians, occasionally joined by some extreme semanticists, tend to take the position that people *never* mean what they say, and that the manifest replies to a question are inevitably a facade to be pierced rather than anything to be taken seriously" (Riesman & Glazer, 1948, p. 644). This is the problem with which Hirsch (1960) and Latour (1991/1993, 2004) are most concerned. Content analysts nonetheless have good reason to resolve that manifest content implies or implicates latent content— one statement implies or implicates another.

1. Two questions are herein begged: What is content? *and* What is analysis?
   a. What is content?
      i. Chatman (1971, pp. 225-226): What is the nature of literary content? Craig La Driere argues that the "matter out of which the thing (the literary work) is made is language." But here "matter" signifies *materia*, material or substance, in the sense that stone is the matter of a sculpture, or oil pigment and canvas the matter of a painting. Content is surely not that.
         1. In the figurative arts, content is ordinarily taken to be the object represented in the matter as it has been shaped by the form" David" in Michaelangelo's sculpture or "the Mona Lisa" in Leonardo's painting. So the content of a literary work is not the language but what the language stands for, its reference.
         2. Weitz (1971, p. 226): Thus, Professor Chatman ("On Defining 'Form'") argues that every piece of discourse has a clearly identifiable content and form. The content is the message or reference.
      ii. Tausczik & Pennebaker (2010, p. 29): *Content words* are generally nouns, regular verbs, and many adjectives and adverbs. They convey the content of a communication. To go back to the phrase "It was a dark and stormy night" the content words are: "dark," "stormy," and "night." Intertwined through these content words are *style words*, often referred to as function words. Style or function words are made up of pronouns, prepositions, articles, conjunctions, auxiliary verbs, and a few other esoteric categories. In the phrase these words are "it," "was," "a," and "and."
      iii. Ludwig (2013, p. 89): Emerging research on text-based communication suggests that both content and style elements of verbatim reviews are relevant decision inputs that help determine relative diagnosticity and accessibility (Huffaker, Swaab, and Diermeier 2011). This research distinguishes linguistic content and style: At a word level, "*content words* are generally nouns, regular verbs, and many adjectives and adverbs. They convey the content of a communication" (Tausczik and Pennebaker 2010, p. 29).
         1. Yet no content can be communicated without style words. As Tausczik and Pennebaker (2010, p. 29) state, "intertwined through these content words are style words, often referred to as function words. Style or function words are made up of pronouns, prepositions, articles, conjunctions, auxiliary verbs, and a few other esoteric categories." These categories identify not only what people convey (i.e., sentential meaning) but also how they write (sentential style), so both have diagnostic value that affects decisions (Bird, Franklin, and Howard 2002).

2. **What is Content Analysis?**
   a. Content Analysis for Meaning Making
      i. Berelson, *Content Analysis in Communication Research* (1952, p. 18): research technique for the objective, systematic, and quantitative description of the manifest content of communication.

        ii. Osgood (1957, p. 275): attempts to infer the characteristics and intentions of sources from inspection of the messages they produce.

       iii. Stone, Dunphy, Smith, & Ogilvie (1966, p. 5): any technique for making inferences by systematically and objectively identifying specified characteristics within a text.

       iv. Krippendorff (1980, p. 21): a research technique for making replicable and valid inferences from data to their context.

       v. Silverman (2001, p. 12): researchers establish a set of categories and then count the number of instances that fall into each category.

       vi. Ahuvia (2001, p. 139): [a method to] code text into categories and then count the frequencies of occurrences within each category.

       vii. Kotsopoulos (2008, p. 469): Content analysis involves counting units in order to establish magnitude as a means of generating themes (Berg 2004). Units of analysis can be discrete entities or combinations of words, themes, characters, paragraphs, items, concepts, semantics, and so forth.

   b. Content Analysis for Curriculum Making

       i. Waples (1926, p. 1): There are two techniques for analysis of curriculum materials which are easy to contrast, because they appear to work in opposite directions. These two techniques may be distinguished as "functional analysis" and "content analysis."

       ii. Petrina (2007, pp. 223, 224)

          1. There are fundamentally three sources of content: individuals, culture, and nature. Content derived from an individual will be developmental, physical, or psychological. Content derived from nature will tend to be biological or ecological and based on basic needs and survival. Content derived from culture will be institutional, sociological, or spiritual.

          2. We derive content through a number of methods. The content of a discipline is derived from a conceptual analysis of facts, concepts, generalizations, and theories established over time. The content of occupations is derived from a task analysis of work and workers at specific points in time. The content of processes is derived from a systems analysis of processes and methods at specific points in time extended over time. To do a conceptual analysis, one has to make logical inferences from established principles and existing problems. To do a task analysis, one has to make procedural observations of tasks. To do a systems analysis, one has to make systematic observations of problems or processes.

3. Do people mean what they say?

   a. Galloway (1971, p. 228): People of all ages can be heard to say, "You know what I mean." During a conversation, people also repeatedly say "you know," or "I mean." These phrases may be uttered as many as a dozen times during a single conversation. Anytime we talk we scurry for words that capture our meaning. We want to be understood. We have great faith in words. But we don't always say what we mean or mean what we say.