

CHAPTER 12

Content Analysis

Understanding Text and Image in Numbers

*Content analysis means
Counting what can be seen.
Use counting and categories
For basic analyses
And insights on speech, text, or screen.*

Chapter Overview

Media for content analysis can include any recorded communication from papyrus to podcast, published or unpublished. Approaches to studying media content are almost as diverse as the approaches to studying human communication behavior. In a general sense, any study of media content is a content analysis, but in communication research the term usually implies a quantitative approach and a series of specific steps aimed at ensuring systematic sampling, coding, and counting of media content. This chapter provides an overview of content analysis procedures. Other approaches, primarily qualitative, are discussed in Chapter 13.

Starter Questions

- Why would I want to content-analyze text?
- How does analyzing texts differ from analyzing human behavior?
- How valid and reliable are content analyses?
- Why would I prefer analyzing texts to analyzing human behavior?
- Can content analysis be free of bias?

Introduction: Advantages and Disadvantages of Content Analysis

In the context of communication research, content analysis is a quantitative, systematic, and objective technique for describing the manifest content of communications (Berelson, 1952).

- Quantitative means we must count occurrences of whatever we are interested in.
- Systematic means that we must count all relevant aspects of the sample. We cannot arbitrarily pick what aspects get analyzed.
- Objective means that we select units for analysis and categorize them using clearly defined criteria.
- **Manifest** means that we count what is tangible and observable. For example, we cannot count patriotism in consumer advertising because patriotism is ultimately an abstract or **latent** (hidden) notion. What we can count is the frequency with which the word *patriotism* occurs, the frequency with which a national flag appears, or perhaps the number of minutes music defined as patriotic is played.

Some researchers argue that latent content can be studied. The difference between the perspectives of latent content and manifest content is essentially the difference between trying to measure a concept (such as patriotism) and measuring the variables that operationalize that concept (such as number of occurrences of the national flag). You can revisit this discussion in Chapter 5.

By systematically sampling, coding, and counting media content, it is possible to make discoveries related to almost any form of content. What adjectives distinguish left-wing politicians' descriptions of global warming from those of right-wing politicians? How do rock 'n roll lyrics from the 1950s differ from millennium rock in their treatment of love? Are minorities depicted in advertising over- or underrepresented with respect to their numbers in the population? Which political candidates do the bumper stickers on campus vehicles mostly endorse? How does the language differ when people in an intimate relationship communicate by e-mail rather than phone?

Content analysis has been used to examine environmental themes in newspaper advertising (Özkoçak & Tuna, 2011), antisocial behavior in a reality TV show (Wilson, Robinson, & Callister, 2012), portrayals of obesity on YouTube (Yoo & Kim, 2012), risk taking in action movies (Beullens, Roe, & Van den Bulck, 2011), and, more generally, media representations of health, parenthood, food, religion, social groups, politicians, and occupations.

It is possible to analyze the content of almost any recorded medium—press, radio, or web—and also billboards, T-shirts, license plates, lawn signs, photographs, love letters, or tweets.

A strength of content analysis is its emphasis on the systematic coding, counting, and analysis of content. The procedures of any content analysis study should be explicit, precise, and replicable so that other researchers can verify the results of the research. Usually, the raw material of content analysis is readily accessible, be it presidential speeches, advertisements for liquid soap or lingerie, BBC foreign news coverage, comic strips, or *New York Times* editorials. Because content analysis is the study of media content, human participants are not involved, and human research approvals are typically not required. Of course, permission to access content and to publish content analysis studies may be required where the content is unpublished or personal, for example diaries or e-mail correspondence. Because content analyses can involve judgments about the nature of content, you may need to recruit human panelists to provide an independent judgment on how an item of content should be categorized.

Content analysis addresses only questions of content. For example, “Have representations of the military on television changed since 9/11?” is an appropriate question for content analysis.

“Are media representations of the military causing more people to enlist?” is not a question content analysis can answer (though it may contribute to an answer).

A limitation of the method is that it really only has application if used for comparisons. A “one-shot” survey indicating that 72% of voters would vote for Candidate X if the election were held tomorrow produces a finding that is useful to a campaign manager. However, a content analysis indicating that Candidate X used the term *patriotism* 17 times in the course of a campaign is not inherently useful information unless it has some point of reference. What we are interested in as content analysts are questions such as “How does the frequency of the term *patriotism* in Candidate X’s speeches compare with that in Candidate Y’s?” or “How does the use of the term *patriotism* compare with the candidate’s use of the term *environment*?” or “Has the frequency of use of the term *patriotism* increased or decreased over the course of the campaign?”

A further issue, if not a weakness, is interpretation. Traditional perspectives on content analysis insist on counting manifest or observable content. As in the example above, we might count the frequency with which the word *patriotism* appears in a politician’s speeches over time, and detect an overall increase. However, that does not necessarily entitle us to assume that the politician has become more patriotic over time; it entitles us only to assume that his use of the term has become more frequent.

Measuring the frequency of the word *patriotism* by itself tells us little about the speaker’s views of patriotism. An impassioned speech against patriotism may reveal the same frequency of the word as an equally impassioned speech in favor of it. In this case, it is obvious that the word needs to be examined in context. One way to do this is to define our units of analysis as sentences or paragraphs rather than words and then code each sentence or paragraph as positive, negative, or neutral with respect to patriotism.

The above example shows that validity can be a problem with content analysis; on the other hand, reliability can be 100% with computer analysis of text. Once programmed, content analysis software will code all occurrences of a unit the same way. Human coders may differ on how a unit should be coded, but by having each coder code the same content, we can compute each one’s level of intercoder reliability. If the level is unsatisfactory, then we may need to (re)train coders until their level of agreement reaches an acceptable level or revisit our coding scheme.

A Basic Content Analysis

In this example, we follow a student—Rafael—who develops an interest in political biases on campus. He is aware of news media coverage of research suggesting that, overall, higher education faculty are more likely to be liberal than conservative. (See, for example, Gravois, 2007, and Gross, 2012.) If true, a concern stemming from this finding could be that a faculty political bias is manifest in the classroom, making students subject to subtle, or perhaps not so subtle, political influence.

Rafael decides that this topic has both theoretical and practical significance. However, he reasons that interviewing faculty on this topic may not be productive as many will not want to be interviewed or to go public with their political views if they are interviewed. He decides that an indirect measure of faculty’s political views and of their willingness to express them can be obtained indirectly using content analysis.

He reasons that faculty vehicles often carry stickers supporting political candidates. A basic count of these stickers would be an indirect measure of faculty political bias and also a measure of faculty’s willingness to express their views in public.

Generally, because content analysts seek to make generalizations from communication content, they take a scientific approach. Systematic sampling, clear definition of units, and unambiguous categories are important aspects of any content analysis study.

A content analysis study typically has seven parts:

- Develop a hypothesis or research question about communication content.
- Define the content to be analyzed.
- Sample the **universe** of content. “Universe” has the same meaning for media content as “population” does for people.
- Select **units** for coding.
- Develop a **coding scheme**.
- Assign each occurrence of a unit in the sample to a code in the coding scheme.
- Count occurrences of the coded units and report their frequencies.

Based on this procedure, Rafael’s basic content analysis of bumper stickers would take the following form.

Research Questions

RQ₁: Do faculty demonstrate a willingness to publicize their political views?

RQ₂: Do faculty vehicle stickers demonstrate a political bias?

Content to Be Analyzed

Rafael defines the content he is interested in as vehicle stickers supporting political candidates. These stickers may not identify a specific political party, but by definition they promote a specific candidate whose party affiliation Rafael will be able to identify if necessary from the records of a state electoral commission or similar office. He decides to exclude from his study “cause” stickers such as those advocating environmental or social change, or support for the military, peace movements, or education. While many of these cause stickers no doubt reflect a basic political viewpoint, Rafael decides that stickers advocating support of the military or health care or preserving endangered species could come from members of any political party and therefore do not capture political leanings as accurately as political campaign stickers.

Sampling

The universe of content will be faculty vehicles, which can easily be identified by the faculty parking permits on them. Rafael decides to sample faculty vehicles in employee parking lots late morning through early afternoon, midweek. He reasons that this timing will maximize the number of faculty on campus. He decides on systematic probability sampling. That is, he will use a random number to identify his starting vehicle and will then select every *n*th vehicle until he has completed his sample of all faculty parking lots.

Select Units for Coding

The units for coding will be candidate names as these are the most obvious and consistent component of a campaign sticker. He will treat two candidate names appearing on a sticker, for example a presidential and a vice presidential candidate, as one name because the ultimate objective is to code each sticker by party affiliation, not by candidate names.

Develop a Coding Scheme

Developing a coding scheme means developing a classification system or categories into which each sampled unit can be placed. Content analysis requires that all sampled units be placed in a category. Vehicles, for example, could be categorized by name of manufacturer, by color, or by state of registration. Rafael's study design is such that he cannot develop a coding scheme in advance of his study because he cannot predict all the candidate names that he is going to find. This means that he will start recording candidate names at the beginning of his study and may well be adding new names to his list at the end of his study. His coding scheme will be the list of candidate names he builds from his sampling of faculty vehicles.

Assign Each Occurrence of a Unit to a Code in the Coding Scheme

The number of times each candidate sticker occurs will be counted. To simplify data recording, Rafael decides on a simple numbered list. The first name he records will be coded 1, the second name 2, and so on. The numbering system will stop at the point where there are no new names to be added. Each repeat name that he finds will get coded with the number assigned to the first occurrence of that name.

Count Occurrences of the Coded Units and Report Their Frequencies

Rafael will end up with a raw data sheet similar to that shown in Exhibit 12.1. For this study he will have a further step to make, and that is to recode each candidate name, where necessary, into a party affiliation. He has two options as to how to do this. As he is interested in overall political affiliations, he may set up a simple coding scheme in advance—"Democrat, Republican, Other"—and assign each candidate name to one of these categories. If he is planning to look at patterns of affiliation in more detail, another option is to develop categories of political parties that will emerge as he translates candidate names into party affiliation. This means he will end up with a coding scheme such as "Democrat, Republican, Green, Communist, Libertarian," and

EXHIBIT 12.1 Data Coding Sheet for Content Analysis of Political Campaign Stickers

	Candidate Name(s)	Tally	Count	Political Affiliation
1	Williamson	//////	6	Other
2	Chee	///	3	Other
3	Jones	////////	9	Other
4	McLeod	////////////////	18	Republican
5	Szydlo	////////	10	Democrat
6	Menendez	////////	8	Democrat
etc.	etc.			
No Stickers		//////	6	–

so on, depending on the candidate affiliations he finds in his sample. Note that because this coding scheme will cover every party he identifies in his sample, the “Other” category will not be necessary here.

In both cases, however, an additional category is necessary—“No Stickers”—because Rafael is also interested in the percentage of faculty vehicles with no candidate stickers. In other words, as his systematic sampling takes him across the faculty parking lots, he is also recording the sampled vehicles that have no candidate stickers and coding those vehicles as “No Stickers.”

Exhibit 12.2 shows the summary data Rafael is now in a position to interpret with respect to his original research questions. Ninety percent of the vehicles he samples have a candidate sticker on them. The answer to his first research question, then, is that the majority of faculty on campus appear willing to publicly express a political preference—at least on their vehicles. The second research question was whether there was an overall faculty affiliation with one political party more than others. Suppose his data for the 90% of vehicles with political stickers show an even split across all three categories—“Republican,” “Democrat,” and “Other.” Rafael could reasonably conclude that his content analysis of candidate stickers indicates no predominant political affiliation by faculty. His overall discussion of his findings would then center on the notion that while a majority of faculty appear to be willing to “go public” with their political affiliations, it is unlikely that there is an overall faculty mind-set that campus critics should be concerned about.

Rafael’s study has a number of assumptions and problems, some of which are raised in the “Application Exercises” at the end of this chapter. The most obvious is whether willingness to go public with a candidate sticker on one’s vehicle translates in any way to a willingness to go public in the classroom. Indeed, a reverse argument might be made that because faculty feel obligated to restrain their political views in the classroom, they are more likely to express their opinions outside of the classroom in the form of vehicle stickers. It also possible that an analysis of the data by discipline would show a different pattern. That is, if Rafael could identify the owners of the vehicles and the disciplines they teach in—for example, business, science, and humanities—he might discover that one discipline is more likely than others to have a political bias and/or to express it in public.

EXHIBIT 12.2 Summary Data From Exhibit 12.1

Percentage of Vehicles With Campaign Stickers	54/60 = 90%
Campaign Stickers by Party Affiliation	
Democrat	18/54 = 33%
Other	18/54 = 33%
Republican	18/54 = 33%

An Expanded Content Analysis

The above example demonstrates the principles of content analysis. In the following example of an expanded content analysis, we work through the thinking behind a more typical scholarly project that involves more categories, units that do not always fit into an obvious category and the question of how to deal with that problem, and the possibility of working with both text and visual content.

For our example, we will come full circle back to pharmaceutical advertising, as discussed in Chapter 1. The content analysis studies summarized in Chapter 1 under “Three Views of Pharmaceutical Advertising” compared the advertising for different types of pharmaceutical products, examined the relative frequency of emotional and rational appeals, and looked for the presence or absence of specific details related to the performance and side effects of pharmaceuticals.

Many related questions can be asked about the promotion of pharmaceuticals in advertisements, news, and entertainment media. For example, to what extent do newspaper medical columns report the failures as well as the successes of new pharmaceuticals? What pharmaceutical products feature most prominently in daytime television advertising? Which prescription pharmaceuticals appear most frequently in the news coverage of major metropolitan newspapers?

Let’s go down a relatively untraveled road—the representations of pharmaceuticals and their use in contemporary comic strips. Our rationale for this study is that comic strips may promote an image of pharmaceutical use to comic strip readers, just as much as pharmaceutical advertising promotes this product category to consumers.

From a critical perspective, we might propose that any presentation of pharmaceuticals in the everyday life of cartoon characters could legitimize their use to vulnerable reader populations not attuned to comics as a subtle form of advertising. From a public health or criminal justice perspective, we might want to know whether the appearance of a new substance in popular media reflects what is happening “on the street” or is about to be “the trigger” for a new drug culture. Educators may be interested in the potential of comic strips to educate young people about drugs. Of course, not all of these interests are ones that content analysis can address on its own, but let’s assume that a basic starting point is to determine the extent to which pharmaceuticals are actually a part of the comic strip world.

An overview of pharmaceutical use in the comic strip world would reasonably include insight into the nature of the users and the conditions of use. For this study, we will assume a theoretical interest in whether users are male or female, and code accordingly, noting that we could equally well have categorized them by age or occupation. Again somewhat arbitrarily we will categorize the conditions of use as either home or work, recognizing that other such categories are possible (campus, “the street,” and vacation settings, for example).

As noted, content analysis data are most useful when they are used as a basis for comparison rather than standing on their own. The above design already has the potential for some interesting analyses, for example the most and least frequently occurring pharmaceuticals, how men and women differ in their portrayed use of pharmaceuticals, and whether pharmaceutical use is shown as primarily at home or at work.

To get comic strip characters’ use of pharmaceutical substances in context, another comparison seems appropriate, and that is a look at how pharmaceutical use compares with the use of other substances in the form of alcohol and tobacco. We will expand the study to include alcohol and tobacco because our overall interest is in how comic strips portray substance use and because we need a point of reference against which to assess our findings about pharmaceuticals. We start, then, with a research question about the portrayal of all these substances in comic strips and a hypothesis that pharmaceuticals will differ in frequency of occurrence from tobacco and alcohol as follows.

RQ_i: What is the relative frequency of use of tobacco, alcohol, and pharmaceuticals as depicted in comic strips?

H_i: Alcohol and tobacco products will differ from pharmaceuticals in frequency of occurrence in comic strips.

Note that this is a two-tailed hypothesis because we have no evidence that one type of substance is more likely to appear than another.

Define the Universe of Content to Be Analyzed

The content to be analyzed is comic strips. We decide that our interest is in current content, and so we opt for the most recent comic strips. We also opt to capture a full year of seasons and events because we suspect that drinking, smoking, and pharmaceutical consumption may have seasonal patterns. Spring break behaviors, for example, may not be representative of student behaviors during the rest of the year. Commencement may see a peak in consumption of champagne and cigars, and Super Bowl weekend may see a peak in beer consumption in some zip code areas. If comic strips are in fact capturing the ups and downs of life and the associated use of chemical substances, a random sample of just one week of content may under- or overrepresent patterns of behavior overall.

We could sample all comics sold over the counter in book form; we could go to a site such as www.comics.com, which boasts of having over 90 comic strips; or we could look at newspaper editorial cartoons. We decide on a universe of syndicated comic strip pages from the Sunday edition of our major local newspaper because the Sunday comics reach a large national audience and are therefore important because of their potential impact on a wide variety of readers. Because so many of the strips in this paper are nationally syndicated, we make the assumption that our major local Sunday paper will represent most other major papers nationally in terms of comic strip content. This assumption could of course easily be tested by comparing the comic strips printed in a random selection of Sunday newspapers from across the country.

We narrow the universe to be sampled by deciding that we will sample only comic strips that regularly show human characters. We know from casual observation that cartoon animals might smoke, drink, and pop pills, but we decide that our theoretical interest is in humans. That gets our universe down to, say, 20 weekly strips or a total of 1,040 over the course of a year.

Sample the Universe of Content

We now have a sampling frame—the list from which specific samples will be drawn—of 1,040 comic strips.

We decide to systematically sample the first Sunday of every month. If we randomly sampled the year, we might sample out some important seasons, and we have decided that all seasons need to be included. We want to capture all the seasons, so systematic monthly sampling is appropriate. That gives us a manageable sample of $12 \times 20 = 240$ comic strips. Note that sampling at the level of months may eliminate some week-specific comic strip activity that might be relevant to our research, for example Guinness-drinking leprechauns around St. Patrick's Day or keg parties around Super Bowl weekend. For a preliminary study, though, we opt for monthly sampling, recognizing that weekly samples may be needed to provide a more precise picture.

Select Units for Coding

Having selected a universe from which to sample, and developed a sampling frame, we are now faced with the question of what units we will sample. Rafael's content analysis of campaign stickers, above, was based on the reasonable assumption that words—in the form of candidate names—would be appropriate units to sample. There are, however, several other possible units. For example Krippendorff (1980) identifies five possible types of units:

- **Physical units** occupy an observable space in print media or time in audiovisual media. An example in our case would be an entire comic strip or the individual panels within each strip.
- **Syntactical units** are units of language, such as words or sentences, for example the names of political candidates in Rafael's study of vehicle campaign stickers.

- **Referential units** refer to a person or an event, for example a specific cartoon character or an event such as drinking coffee.
- **Propositional units** are structures such as stories or dramas, for example a cartoon character rejecting an offer of alcohol.
- **Thematic units** are broad topics within a structure, such as relationships with the boss or peers, for example the continuing efforts of a son to have his father stop smoking, shown again and again over a series of comic strips.

As we progress from physical units to thematic units, we face an increasing difficulty making decisions about them. Physical units such as photographs are apparent; either they are there, or they are not. Themes such as “environment” or concepts such as “physical fitness” may appear only in the eye of the researcher and typically would be validated by a group of judges looking at the content being analyzed.

In this example, we are looking at substance use in the broad setting of people’s lives, so we decide to examine entire comic strips rather than each panel within a strip for instances of substance use. For example, a comic strip with five frames, each of which shows the same character smoking a cigarette, will be coded as one occurrence of smoking, not five.

Because we are analyzing comic strips, we have a further decision to make. Should we look at visual content only, words only, or both? Each has its advantages and disadvantages. For example, words may show specific brand names and the specific strength of an individual’s relationship to a product. (“I like ___” versus “I love ___,” for example). On the other hand, the visual content is likely to better show the conditions of use, for example whether an individual drinks alone or in company.

We decide that portrayal of substance use implies visual portrayal and that therefore we will analyze the visual content as a priority. However, the presence of text in the comic strips gives us an opportunity to see if text gives us the same results as images. We will therefore also analyze the strips’ text as a way of validating, or not, the results we get from our analysis of visuals.

Our units of analysis will be syntactical for text; that is, we will be looking to count the number of times the names of a specific substance is mentioned. For visual content, our units will be referential; that is, we will count the number of times a substance is shown being used by a character.

Develop a Coding Scheme

Because our interest focuses on the frequency with which substances are referenced in comic strips, we develop a coding scheme with categories as shown in the “Substance” column of Exhibit 12.3. The coding sheet will record the number of times we find a mention or depiction of the substances we are interested in.

A simple study might compare the relative frequency of appearance of different drug categories (the “Overall” column). A more complex study might look at the relative frequency by, for example, setting (“Work” or “Home”) or type of person (“Male” or “Female”) using the substance.

Note that for every category shown a breakdown into more specific categories is possible. For example, “Tobacco” could have subcategories of “Cigarettes,” “Cigars,” “Pipe,” and “Smokeless.” The “Alcohol” category could have subcategories of “Beer,” “Wine,” and “Spirits.” Conversely, it would be possible to coalesce the categories of “Prescription Drugs,” “Over-the-Counter Drugs,” and “Illegal Drugs” into one category called “Drugs” and simply to code every appearance of a substance as either “Tobacco,” “Alcohol,” or “Drugs.” The level of categorization depends on the research question being addressed. A preliminary study might require little more than a “Tobacco, Alcohol, Drugs” set of categories. On the other hand, if comic strips are argued to influence social trends, then the “Cigarettes, Cigars, Pipe, Smokeless” categories for tobacco would be needed to track whether, for example, smokeless tobacco was becoming more popular at the expense of cigarettes or cigars.

EXHIBIT 12.3 Sample Coding Sheet for Content Analysis of Comic Strips

Substance	Frequencies of Appearance				
	User			Setting	
	Overall	Male	Female	Work	Home
<i>Tobacco</i>					
Visuals					
Words					
<i>Alcohol</i>					
Visuals					
Words					
<i>Prescription Drugs</i>					
Visuals					
Words					
<i>Over-the-Counter Drugs</i>					
Visuals					
Words					
<i>Illegal Drugs</i>					
Visuals					
Words					
<i>Generic Reference to Drugs</i>					
Visuals					
Words					

Similar decisions arise with respect to drug categories. There are thousands of prescription drugs available. Listing them by name would obviously provide far more categories than needed for a content analysis study of popular media, so the question arises of how to develop a few workable categories. Consider how pharmaceuticals might be categorized. You could categorize them by nongeneric name, generic name if a generic version exists, active ingredient, common names used by consumers, purpose (e.g., antacid or antibiotic), or method of administration (e.g., oral, injection, ointment, or inhalant).

Intuitively, the thousands of names used by manufacturers for their products provide far more categories than any content analyst would ever use. (A possible exception would be a content analysis of medical or pharmaceutical journals.) For our analysis of comic strips, it is most likely that what we would find in use would be a combination of lay terminology and purpose of the product—for example, sunscreen, contraceptive, antibiotic, pain medicine, and so on.

The possibility of categorizing pharmaceuticals by “Tablet, Liquid, Injection, Ointment, Inhalant” flags one of the differences between content-analyzing text and visuals. It is likely that in our study of comic strips we will see whether characters are popping pills and drinking liquids, but we probably would not get that information from an analysis of the text only. On the other hand, text may be more likely than the cartoon illustrations to provide the names of specific products.

None of the above generalizations can be made with 100% accuracy, of course. For example, the name of a specific substance may come from a graphic of a large pill bottle with the word *aspirin* emblazoned on it.

Ground rules for content analysis are that categories must not overlap, and that no unit can be coded twice. “Substance X,” if shown, or the word *aspirin*, if found, must fit one and only one category. Anticipating that some characters in the Sunday supplement cartoons may be doing illegal drugs, we need a coding scheme that allows for that possibility. Another issue arises. How do we handle occurrences of generic words such as *drug(s)*, *pharmaceutical(s)*, or *medicine(s)*? We can’t ignore a reference to a substance simply because it is generic, so we need to add another category—“Generic”—to our coding scheme.

This discussion of coding should indicate the type and level of decisions that are required in setting up a content analysis study. Two things will help you establish a defensible coding scheme. The first is a clear theoretical background and research question. For example, if your theoretical interest is in the advertising of prescription drugs, you can eliminate nonprescription drugs from your coding scheme. The second, and essential, item is a test run or pilot study that pretests your categories and sampling before you run your analysis. Almost inevitably you will identify categories that can be eliminated and categories that need to be added. You may also detect changes that need to be made to your sampling. For instance, in the above example we may find that monthly sampling is inadequate and that we need to sample comic strips weekly.

Assign Each Occurrence of a Unit to a Code in the Coding Scheme

With coding categories set up, we now proceed to code the units of analysis we find in our comic strips.

Identifying word references should be easy; we are looking for specific names such as *beer*, *cigar*, *latte*, or *aspirin* that clearly identify a substance. With visual occurrences, we may have more of a judgment call. What *is* that character ingesting? Where more than one interpretation is possible, it is usual to form a panel of judges who will vote on how to code an item. This is likely to be necessary in our study.

For any analysis involving multiple coders, it is usual to ask them to code a number of units as a trial run. Formulas are available to help calculate the degree to which the coders are in agreement. If they are in agreement, we can have some confidence that our criteria for coding are clear; if not, it is time to revisit our criteria and redefine them so they are clear.

Count Occurrences of the Coded Units and Report Their Frequencies

The results of a study like this would be reported as the number of occurrences recorded for each of our substance categories.

We would report and discuss the relative frequency of the substances shown in our coding sheet. This gives us the basic comparison that content analysis does well and helps us decide whether our initial hypothesis is supported or not.

Our study thus far has done little more than help us decide the relative occurrence of substances by category even though we have data for both text and visual content.

It would be much more informative if we could assess the conditions under which these substances appear and perhaps the types of people using them. A further level of analysis then would be to categorize substance use according to type of user, for example male or female, blue collar or white collar; or conditions of use, for example home or work. Our coding sheet would then include the columns shown on the right-hand side of Exhibit 12.3.

Content Analysis of Human Interaction

The principal use of content analysis is in analyzing media content such as news, entertainment, or advertising (in part because of the ready availability of media content), but it can also be applied in such areas as organizational communication and interpersonal communication. For example, transcripts of interactions between people constitute texts and are open to content analysis just as much as television advertising, newspaper editorials, or State of the Union Addresses.

For example, **interaction analysis**, pioneered by group dynamics scholar Robert Bales (1950), seeks to capture and understand interactions among members of a group and the different roles that group members play. Three broad categories of group behavior are task-oriented, group-oriented, and self-centered. Task-oriented individuals focus on the group's work, for example asking if the group has all the information it needs or assigning specific tasks to members of the group. Group-oriented individuals work to ensure that the group remains cohesive, for example by making jokes to relieve tension or by showing supportiveness for other members' ideas. Self-centered individuals may refuse to participate or at the other extreme may dominate discussions.

EXHIBIT 12.4 Sample Coding Sheet for Group Behaviors Over Time

Behavior	Frequencies of Appearance		
	Time 1	Time 2	Time 3
<i>Task-Oriented</i>			
Reminds group members of deadlines			
Proposes solutions to problems			
<i>Group-Oriented</i>			
Makes statements supporting a group idea			
Makes statements approving of group's progress			
<i>Self-Centered</i>			
Refuses assignments			
Makes antagonistic statements			

Suppose we are looking at the transcripts or a video recording of a corporate team trying to develop a marketing strategy for a new product. The ebb and flow of discussion over the weeks leading up to the group's final decision can be coded using Bales's (1950) categories or subcategories as shown in Exhibit 12.4. Here, the group is the unit of analysis, and we are coding observed behaviors for three different meetings at Time 1, Time 2, and Time 3.

We could use the results of such an analysis to test a hypothesis that group behaviors change over time, for example that group-oriented behaviors occur more frequently in the preliminary stages as group members clarify their roles and relationships before moving on to task-oriented behaviors. We could compare the results of such analyses for two different groups, one known to produce excellent decisions, the other known to produce mediocre decisions. This might then allow us to identify the characteristics of high-performing groups and provide a basis for training low-performing groups to do better. We could also use exactly the same criteria to analyze individual behaviors. For example, if we were to replace Time 1, Time 2, and Time 3 in the chart with people (Mick, Keith, and Charlie), we could record, analyze, and compare the behaviors of individuals.

We could, of course, combine both approaches and record individual behavior over time. Note that if we did record the behavior of each individual at different times we would have a within-subjects design (Chapter 10). Such data are not statistically independent, and we would need a more sophisticated statistical test to assess any changes over time. By contrast, the data from the study of bumper stickers are statistically independent, and a simple chi-square test will serve to assess whether the distribution of bumper stickers shows significant variation. The statistics used for content analysis assume that there are large counts of items. The chi-square test typically requires a count of at least five in each category.

Purposes and methods for analyzing media content often overlap. For example, content analysis may be used to obtain a critical reading of texts.

Content Analysis Software

Content analysis of language can be automated and expedited with the many content analysis software packages available. For example, Diction software uses word lists to search a text for the qualities of certainty, activity, optimism, realism, and commonality.

IBM's SPSS Text Analytics for Surveys can identify major themes in survey responses, distinguish between positive and negative comments, and extract key concepts and opinions.

Most commercial software packages have the ability to automatically classify large amounts of data and to identify themes, but all such software needs "training" if the results are to be valid. For example, the software needs to know whether the word *right* is being used in the sense of "not wrong," "not left," or the verb form "to correct." The power of such software to process large amounts of data rapidly does not mean the researcher can go on "autopilot" and avoid the need to develop a defensible coding scheme and data analysis procedures.

If word frequency is your only interest, and your documents for analysis are short, you have a crude content analysis device in the form of your word processing software. Simply use this software's "find" command to find how many occurrences of a word a document has. Similarly, you can request readability statistics, which will give you no insight on content as such, but will give you measures of average word and sentence length and a readability score for documents.

Qualitative analyses of text are discussed in Chapter 13.

Ethics Panel: Could Analyzing Media Content Result in Harm?

A literature search using the terms *content analysis* and *ethics* will show you any number of studies focused on the content analysis of codes of ethics. This panel addresses a different issue—the ethical implications of content analysis, and textual analyses more generally. Such analyses focus on text and do not involve human participants directly. Why, therefore, would any of these methods have any impact on individuals that would give rise to ethical concerns?

Consider the applications, overt or covert, of content analysis in the world of electronic communication. “Pop-up” advertisements on your web browser are based on an analysis of the websites you visit.

Many software packages offer e-mail administrators a way of monitoring incoming e-mail for “spam,” viruses, and attempts at “phishing.” Such software can also monitor outgoing e-mail for compliance with company policy. The software administrators can set up e-mail monitoring to analyze content and prevent sensitive content leaving an organization or to identify potential policy violations and thus protect intellectual property, company reputation, and business relationships. Content analysis of employee e-mail is routine in many corporations for all the reasons noted above.

While we can all be thankful for e-mail spam filters that themselves are based on analysis of incoming e-mail message content, we might see such software as an invasion of privacy if applied to our own e-mail.

Questions

- In what ways could the knowledge that one’s business or personal communications are being monitored and analyzed cause psychological distress?
- Should the institutional review boards responsible for human subjects safety be required to review content analysis studies of e-mail traffic even if no human participants are studied directly? Why or why not? Revisit Chapter 3 for more detail on these review boards (IRBs).
- With respect to obtaining permissions, should content analysts treat personal or business e-mails, Internet discussion content, and correspondence differently from television commercials or newspaper pages? Why or why not?

RESEARCH IN PRACTICE

Reaching Out to Hispanics Living With Paralysis

In late 2006, Vanguard Communications began work with the Christopher and Dana Reeve Foundation’s Paralysis Resource Center (PRC) to educate Hispanics about the wealth of information, resources, and services available for people with mobility-related disabilities, as Hispanics were generally underrepresented in the PRC’s call volume. After conducting a series of communications research activities including focus groups and a media analysis, Vanguard and the PRC collaborated to develop the *Vivir Sin Límites* (meaning “limit-free living”) Hispanic outreach campaign.

As the team began its work to design a campaign that would reach Hispanics, it was clear that a few key research questions needed answers. How are people living with paralysis portrayed in the U.S. Hispanic media? Where do Hispanics living in the United States get their health information? What resonates with these audiences, and what would motivate them to call the PRC? To answer these questions, the team conducted a two-pronged research effort consisting of a comprehensive media content analysis and focus group testing.

The media analysis helped establish a baseline from which to compare future media coverage and determine effective strategies for positioning the PRC as the premier source of information about paralysis. The analysis focused on the way people living with paralysis are portrayed in the nation's top English- and Spanish-language print publications including newspapers and Hispanic-focused magazines.

A major finding of the media analysis was that paralysis was rarely discussed in depth. Coverage usually focused on paralysis as a result of a violent crime or an accident. The reality of what it's like to live with paralysis was never discussed, and portrayals of those with paralysis were rarely inspiring or hopeful. Based on these articles, it was clear that readers were provided with little information about the realities of paralysis, including that people with paralysis live healthy, productive lives. To educate Hispanics about the many ways people with paralysis can be independent, productive, and active, the images and messages selected for the campaign needed to showcase diverse personal stories of those who are living independent, productive, and active lives with paralysis.

To ensure that this message and image concept would resonate with Hispanics, draft themes, messages, slogans, and images were tested in a series of focus groups with Hispanics living with paralysis and with Hispanic women who are close friends or family of a person living with paralysis. Focus group participants had highly positive reactions to messages and images that showcased a life that is not limited by a person's physical disability. In particular, visual images such as a person with paralysis engaged in educational pursuits or sports best supported the *Vivir Sin Límites* message. The focus groups also indicated that Hispanics seek their health information in Spanish-language magazines and newspapers, on health websites, and in radio and television programming.

Based on the outcomes of this research, the team conceptualized a positive *Vivir Sin Límites* campaign, incorporating inspiring accounts of people with paralysis living highly productive lives. The campaign was kicked off at a national media event in Washington, DC, featuring Hispanic individuals with paralysis who shared personal stories of how the PRC helped them to live independent and productive lives. The launch highlighted an array of materials available to Hispanics in English and Spanish, including a comprehensive website and a toll-free, bilingual help line. Other campaign materials included a brochure, fact sheets, a bookmark, and a bilingual resource guide about living with paralysis.

To promote the campaign launch and the PRC as the premier source of information for Hispanics on paralysis, outreach was conducted to Hispanic print publications, television and radio outlets, and online resources. Media outreach efforts produced more than 3 million audience impressions, including an interview on the top national Spanish-language morning television program, Univision's *Despierta América*. The PRC also experienced increased call volume, achieving one of the primary objectives of the outreach campaign.

Contributed by Vanguard Communications

*Vanguard Communications is a public relations
and social marketing firm based in
Washington, DC.*

CHAPTER SUMMARY

- Content analysts measure communication content quantitatively.
- Content analysis is a process of systematically sampling, coding, and analyzing media content.
- Content analysis techniques can be applied to both text and visual content.
- Interaction analysis examines groups for the behaviors that hinder or facilitate group processes.

KEY TERMS

Coding scheme

Interaction analysis

Latent

Manifest

Physical units

Propositional units

Referential units

Syntactical units

Thematic units

Units

Universe

APPLICATION EXERCISES

Exercise 1: Sampling

Review the content analysis study of vehicle campaign stickers outlined at the beginning of this chapter. What problems can you identify with the sampling decisions made here? Consider, for example, how time of day and day of week might influence the sample; whether vehicles with out-of-state license plates should be included in the sample; whether national, state, and local election stickers should all be counted; and whether campaign stickers demonstrably more than, say, four years old should be included.

Exercise 2: News Media Bias

Both ends of the political spectrum complain about biased reporting by the news media. Set out the basic elements of a content analysis project that would answer the question of whether a particular news medium you identify is biased toward or against a political figure, government policy, or program. Identify the universe of content you would sample from, the units of analysis, and the coding scheme you would use.

Hint: Visit the Center for Media and Public Affairs website at http://www.cmpa.com/about_methods.htm.

Exercise 3: Stereotyping in Entertainment Media and Advertising

A criticism of advertising and entertainment media such as movies and television is the stereotyping of people by, for example, gender, ethnicity, occupation, or age. Pick one of these types, for example occupational stereotyping, and outline a content analysis study that would test for the presence or absence of stereotyping. Identify the universe of

content you would sample from, the units of analysis, and the coding scheme you would use. Note that for this exercise you will need some operational definition of stereotyping so that you can identify it and code it when you see it.

RECOMMENDED READING

- Bailey, A. A. (2006). A year in the life of the African-American male in advertising. *Journal of Advertising*, 35(1), 83–104. A contemporary example of content analysis of advertising.
- Berelson, B. (1952). *Content analysis in communication research*. New York: Free Press. A pioneering text on content analysis.
- Hirokawa, A. Y. (1988). Group communication research: Considerations for the use of interaction analysis. In C. H. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes* (pp. 229–246). Norwood, NJ: Ablex. Shows ways of coding behaviors in groups.
- Holsti, P. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley. An introductory text on content analysis.
- Krippendorff, K. H. (2012). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks, CA: Sage. The concepts and practice of content analysis. An updated edition that covers electronic texts, social media, and Internet sampling.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage. Has step-by-step instructions and examples.

RECOMMENDED WEB RESOURCES

- Annenberg–Robert Wood Johnson Coding of Health and Media Project: www.youthmediarisk.org. An ongoing project analyzing media content for health-related behaviors. Look at the codebook and sampling procedures in particular.
- Center for Media and Public Affairs: http://www.cmpa.com/about_methods.htm. An explanation of content analysis in practice.
- Matthias Romppel's content analysis resources: <http://www.content-analysis.de>. Links to bibliographies, research, software, and text markup and encoding systems.
- Text Analysis Overview: <http://www.textanalysis.info>. Links and news on text analysis software.

REFERENCES

- Bales, R. (1950). *Interaction process analysis: A method for the study of small groups*. Reading, MA: Addison-Wesley.
- Berelson, B. (1952). *Content analysis in communication research*. New York: Hafner.
- Beullens, K., Roe, K., & Van den Bulck, J. (2011). The portrayal of risk-taking in traffic: A content analysis of popular action movies. *Journal of Communications Research*, 2(1), 21–27.
- Gravois, J. (2007). The battle over studies of faculty bias. *The Chronicle of Higher Education*, 53(21), A8.
- Gross, N. (2012, March 3). The indoctrination myth. *The New York Times*. Retrieved from <http://www.nytimes.com>
- Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.
- Özkoçak, L., & Tuna, Y. (2011). A content analysis: Environment themes and tools in newspapers advertisements. *Online Journal of Communication & Media Technologies*, 1(3), 1–13.
- Wilson, C., Robinson, T., & Callister, M. (2012). Surviving *Survivor*: A content analysis of antisocial behavior and its context in a popular reality television show. *Mass Communication & Society*, 15(2), 261–283. doi:10.1080/15205436.2011.567346
- Yoo, J. H., & Kim, J. (2012). Obesity in the new media: A content analysis of obesity videos on YouTube. *Health Communication*, 27(1), 86–97. doi:10.1080/10410236.2011.569003

STUDENT STUDY SITE

Visit the student study site at www.sagepub.com/treadwell2e to access the following materials:

- Web Resources
- Recommended Readings
- eFlashcards
- Study Questions