Chapter 10

Content Analysis

Very often, a political scientist may learn a great deal about individuals, groups, institutions, or even nations through a careful examination of the communication patterns associated with them. Do campaign advertising and election year news reports provide much information about candidates' preferences and abilities? Do the internal memoranda of a large corporation reveal a systematic plan on the part of management to bebe representatives of foreign governments they wish to deal with?

What does the Congressional Record tell us about the relative influence or importance of each U.S. senator? Do diplomatic communiques between the United States and Russia reflect the public perception of a reduction in the level of conflict between these two nations?

These questions and others like them may best be answered by a direct examination of various items of communication. In general, these items fall into one of three classes: those that are internally generated by the individual, organization, or government we are studying and externally directed (such communications as corporate memoranda, which represent or reflect the decision-making process itself), those that are internally generated and externally directed (such publications as the Congressional Record, which are purposefully molded to create a particular image for the source among outsiders and which may reflect or obscure the process and outcome of decision making), and those that are externally generated and internally directed (things like campaign propaganda, which, when viewed from the perspective of the recipient, provide potential source material for decision making). Each class of communication may be different in purpose or effect, as well as in accessibility and usefulness for research, but each provides potential opportunities to further our understanding of political behavior.

In each instance, the most appropriate technique for pursuing these opportunities is content analysis—the systematic counting, assessing, and interpreting of the form and substance of communication. Content analysis provides us with a
method—really a set of methods—by which we may summarize fairly rigorously certain direct physical evidences of the behaviors of, and the relationships between, various types of political actors. In this chapter we shall discuss when it is appropriate to use content analysis, how the technique is applied, and how the results of content analysis should be interpreted, as well as certain limits of content analytic procedures.

PREPARING TO USE CONTENT ANALYSIS

Content analysis may be used to answer research questions whenever there is a physical record of communications by, to, within, or among the political actors we are interested in, as long as the researcher has access to that record. Examples of such a record include books, pamphlets, magazines, newspapers, phonograph records, audiotape and videotape recordings, photographs, transcripts of meetings or proceedings, government documents, memoranda, films, diplomatic communiqués and instructions, political posters and cartoons, political advertising, speeches, and even letters and diaries. Some of these records may be extremely detailed and precise (as is a verbatim transcript of a congressional hearing), while others are much less so (for example, the agenda for the same hearing). Many will have been created independently of the research process (as are newspaper articles by or about the person or group we wish to study), while others must be created by the researchers themselves (as, for example, videotapes of television news programs). But all sources of data for content analysis will have in common one principal characteristic: the existence of a physical record of communication. Whenever such a record exists or can be created, content analysis may serve as an appropriate research method.

The first step in preparing to undertake a content analysis is to define the population of communications we want to study. Here we have a number of options. Which is the best will be determined by our particular research question. For example, if we are interested in studying the development of political themes in twentieth-century American novels, we might define our population as all novels (the type of communication) written by Americans (the type of communication) and published in the United States (the location of communication) between January 1, 1900, and the present (the time period of communication). If we wish to study newspaper coverage of a political campaign, we might define our population as all campaign-related newspaper articles (the type of communication) of two columns or more in length (the size of the communication) published in daily newspapers (the frequency of communication) that are home delivered (the distribution of the communication) in the sixth, seventh, and eighth congressional districts of Ohio (the location of the communication) during the period September 1 to November 5 of the election year (the time period of communication). Or, similarly, if we want to study the level of tension between the leaders of the United States and those of Uganda, we might define our population as all diplomatic messages (the type of communication) passed between the governments of the United States and Uganda (the parties to the communication) during a given time period.
In each instance, we define the population of messages to be studied by establishing sets of criteria to be met by each item. In the examples, these criteria include the type of communication (novels, newspaper articles, or diplomatic notes), the type of communicator, the parties to the communication (the sender or the receiver or both), and the location, frequency, minimum size or length, distribution, and time period of the communication. Although other criteria may be used on occasion, some or all of those listed here will be found in most studies that employ content analysis. The first task in preparing for a content analysis is to choose those criteria that relate most directly to the research question at hand.

Once we have defined our population, we are faced with the problem of deciding which particular cases we shall examine in detail. Because the cases to be analyzed are often limited in number and relatively accessible and because content analysis is generally less expensive per case than other methods (most notably survey research), we are sometimes able to examine every case in a given population—to conduct, in effect, a census. Indeed, the opportunities it offers for the examination of large numbers of cases is one of the major attractions of content analysis as a research technique. More often than not, however, even content analysis must be based on a more limited sample drawn from the larger population. Since documents, newspaper articles, and the like are frequently indexed or otherwise listed in some central location and since such indexes or lists may easily be created by the researcher, the most common sampling procedures used in content analytic studies are the simple random and systematic random techniques. Even when sampling is required, however, the accessibility and relatively low cost of researching messages of various types come into play, and the sample sizes drawn for content analysis may be substantially larger than those employed in other types of research. The result, of course, is a reduction in sampling error and an increased level of confidence in generalizing from our results.

Finally, in preparing to undertake a content analysis, we must decide on our unit of measure, or, as it is more commonly termed, our unit of analysis. The unit of analysis sets of criteria to be met by each item. In the examples, these criteria include the type of communication (novels, newspaper articles, or diplomatic notes), the type of communicator, the parties to the communication (the sender or the receiver or both), and the location, frequency, minimum size or length, distribution, and time period of the communication. Although other criteria may be used on occasion, some or all of those listed here will be found in most studies that employ content analysis. The first task in preparing for a content analysis is to choose those criteria that relate most directly to the research question at hand.

Once we have defined our population, we are faced with the problem of deciding which particular cases we shall examine in detail. Because the cases to be analyzed are often limited in number and relatively accessible and because content analysis is generally less expensive per case than other methods (most notably survey research), we are sometimes able to examine every case in a given population—to conduct, in effect, a census. Indeed, the opportunities it offers for the examination of large numbers of cases is one of the major attractions of content analysis as a research technique. More often than not, however, even content analysis must be based on a more limited sample drawn from the larger population. Since documents, newspaper articles, and the like are frequently indexed or otherwise listed in some central location and since such indexes or lists may easily be created by the researcher, the most common sampling procedures used in content analytic studies are the simple random and systematic random techniques. Even when sampling is required, however, the accessibility and relatively low cost of researching messages of various types come into play, and the sample sizes drawn for content analysis may be substantially larger than those employed in other types of research. The result, of course, is a reduction in sampling error and an increased level of confidence in generalizing from our results.

Finally, in preparing to undertake a content analysis, we must decide on our unit of measure, or, as it is more commonly termed, our unit of analysis. The unit of analysis sets of criteria to be met by each item. In the examples, these criteria include the type of communication (novels, newspaper articles, or diplomatic notes), the type of communicator, the parties to the communication (the sender or the receiver or both), and the location, frequency, minimum size or length, distribution, and time period of the communication. Although other criteria may be used on occasion, some or all of those listed here will be found in most studies that employ content analysis. The first task in preparing for a content analysis is to choose those criteria that relate most directly to the research question at hand.
e, we define the population of messages to be studied by census to be met by each item. In the examples, these criteria include
inclusion of news, newspaper articles, or diplomatic notes, the
or, the parties to the communication (the sender or the receiver),
minute size or length, distribution, and time
communication. Although other criteria may be used on occasion,
listed here will be found in most studies that employ content
ink in preparing for a content analysis is to choose those criteria
relevant to the research question at hand.
ified, our population, we are faced with the problem of decid-
re cases we shall examine in detail. Because the cases to be ana-
ted in number and relatively accessible and because content
less expensive per case than other methods (most notably survey
methods) are sometimes able to examine every case in a given population—to
the census. Indeed, the opportunities it offers for the examination
cases is one of the major attractions of content analysis as a re-
source even if not, however, even content analysis must be
cluded sample drawn from the larger population. Since documents,
and the like are frequently indexed, or otherwise listed in some
study, such indexes or lists may easily be created by the use of
common sampling procedures used in content analytic studies
and systematic random techniques. Even when sampling is re-
easibility and relatively low cost of researching messages of
into play, and the sample sizes drawn for content analysis may be
than those employed in other types of research. The result, of
in sampling error and an increased level of confidence in gen-
erealizing about the content analysis, we must decide on our unit
is more commonly termed, our unit of analysis. The unit of
analysis is simply the particular element or characteristic of a
that we shall examine, count, or assess. The most basic elec-
tion, for example, is the word, and it may be employed in a
manner. As reflected in each one’s public statements, was
Reagan, Jimmy Carter, Gerald Ford, or Richard Nixon most di-
world peace? We might simply sample the statements of each
of the word peace (and, perhaps, some other related
in the text. In speeches before the United Nations during the
, which country was most conciliatory on questions of eliminate
Middle East: Israel, Egypt, Syria, or Saudi Arabia? Again, we might
of all such speeches and count references to such words as
and compromise. In each instance, we identify certain impor-
the frequency with which they appear.
ple procedure, however, we must take care to avoid at least
we must remember that nonstandardized measures can lead to
the years in question, the Israelis have uttered a total of
uding 30 salient references (to the words we wish to count),
ave uttered some 200,000 words, including 100 salient refer-
ences, we might reach either of two conclusions from a study of these speeches
depending on whether or not we have chosen to standardize our indicator. If we
simply count salient references, we will conclude that the Egyptians have been
as concerns as the Israelis with procuring a settlement. If, however, we
standardize our measure to obtain the proportion of all words that are in fact
salient (for example, salient references per 1,000 words), we will conclude that
both sides have shared an equal concern about settling their differences. Which
approach is better? This is a fundamental problem in operationalizing variables,
and the answer is best determined by looking closely at how we have conceptual-
ized the research question initially. The point here is that the use of even such a
sensibly concrete indicator as a number of salient words spoken can entail some
ambiguity. The researcher must recognize and deal with that ambiguity, because
the decisions made (or overlooked) can have a substantial impact upon the conclu-
sions one draws.
A second potential pitfall in the reliance upon raw word counts arises because
the same word may be used in many different ways: “We seek peace, but . . .”; “The
Arab brotherhood can never allow . . .”; “There will be no compromise.” In the ab-
sence of any sort of control, references such as these to peace, brotherhood,
and compromise will be included as positive references and will, at the very least,
inflace the assessment of the interest of the past on the part of one or both sides in reaching an accom-
modation. If such usages are sufficiently common, they may well mislead us altogeth-
er. For this reason, if we choose to count words, we should in most instances choose
to count them in context.
At least two alternatives exist for doing so. First, we may use judges or
or coders—people who are part of the research team or are employed by it—to read
each salient reference in context and to judge that context as positive, neutral, or
negative. Usually more than one coder should read each reference, and a relatively
high level of agreement among coders should be required before a final determina-
tion is reached. (We say more on this point later in this chapter.) This contextual
judgment can then be used to enrich our data by allowing us to count and interate
not only all references to the words we are focusing on but also the proportions of
positive and negative references.
Another possible response to the problem of interpreting individual words in
context, though it is at best only a partial solution, is to move to, or add, a second
unit of analysis—the theme. A theme is a particular combination of words or ideas,
such as a phrase, sentence, or even a paragraph. In effect, when we count
themes, we search for recurring subjects in a text, as, for example, the expressions
cold war, the refugee problem, national health insurance, or the Christian way.
The procedure is similar to that for counting words and represents an improve-
ment to the extent that themes incorporate the modifiers (adverbs, adjectives) and
explanatory text that both accompany usage of a particular word and help to es-
ablish its meaning.
The problem with analysis at the thematic level, however, is that, although it
does not clear the context in which individual words are used, it does so at the
cost of much added complexity. This is true in that the same theme may be refer-
enced in very different ways and by very different sets of words. Sometimes these
references may be very subtle, displaying few or none of the overt characteristics
we are looking for. References to immigration issues, for example, may be veiled in conciliatory words about political asylum, whereas those applied to Christianity may be cloaked in nationalistic rhetoric. Do such words and rhetoric constitute salient references? Is the theme present, or is it not? These questions do not have simple answers. To the contrary, they generally require us to arrive at some clearly stated but potentially limiting definitions and to develop a series of highly formalized decision-making rules (for example, allowing only overt references that contain one or more words or phrases from a given list to be counted), which may make our findings more reliable but at the same time less meaningful.

A third unit of analysis commonly used in content analysis research is the item—the communication itself taken as a whole. What proportion of books published in the United States in 1935 advocated socialism? Which presidential candidate in 1992 was the subject of the greatest number of favorable newspaper editorials? How did letters written by Richard Nixon after his resignation from office differ from those written earlier? In each instance, we treat the item of communication as a unit and examine its overall characteristics. Does it or does it not deal with a particular issue? Does it or does it not reflect a certain set of values or preferences? Such questions lose some of the subtlety of judgment required by lesser units of analysis, and they necessitate the making of summary evaluations, but for precisely these reasons, their analysis is generally more manageable than is that of words or themes, in a sense making fewer demands of the researcher. This is true because variables may be operationalized at a less specific level, one on which events (that is, occurrences of a salient reference) are often more apparent and on which measurement is often more reliable.

Item-based studies of the use of words and themes have become in recent years much easier to perform due to the development of computer data bases like LEXIS/NEXIS, which was discussed in Chapter 3. Suppose, for example, we wanted to know how often George Bush referred to Saddam Hussein as Hitler-like during the months leading up to the 1991 Persian Gulf War. Using NEXIS, we could request both a full-text search to count all of the articles in the New York Times (or any of a large number of other newspapers, magazines, or broadcast transcripts) in which the words "George Bush," "Saddam Hussein" and "Hitler" appeared for each given month of the period under review. Once in the relevant file, and depending on the software one used for access, the instruction might look something like this:

( (George Bush OR (President Bush)) AND ((Saddam Hussein) OR Hitler) )
AND (Date = September 1990)

This string of terms would identify any article published during September 1990 in which Mr. Bush's name appeared and in which Mr. Hussein's name also appeared within 10 words in either direction from the word "Hitler." Similar searches would allow us to test (1) whether Mr. Bush was referred to more often by title rather than by first name as the crisis moved more clearly toward a military resolution (as might be the case, for instance, if the media were subtly enhancing his stature as the nation neared a war); (2) whether Mr. Bush led or lagged behind other political leaders in his use of the Hitler analogy, or (3) numerous other hypotheses related to the framing of public perceptions of the conflict.
ferences to immigration issues, for example, may be veiled about political asylum, whereas those applied to Christianly jingoistic rhetoric. Do such words and rhetoric constitute the theme present, or is it not? These questions do not have a contrary, they generally require us to arrive at some clearly limiting definitions and to develop a series of highly formal rules (for example, allowing only overt references that consist of phrases from a given list to be counted), which may be reliable but at the same time less meaningful.

Analysis commonly used in content analysis research is the in a given social context. If presidential candidates, subject of the greatest number of favorable newspaper editorials by Richard Nixon, after his resignation from office, cited Hitler. In each instance, we treat the item of communication as a substantial characteristic. Does it or does it not deal with a pair does it not reflect a certain set of values or preferences? Such a subtlety of judgment required by lesser units of analysis, making of summary evaluations, but for precisely these reasons generally more manageable than is that of words or themes, in demands of the researcher. This is true because variables may be less specific level, one on which events (that is, occurrences are often more apparent and on which measurement is often of the use of words and themes have become in recent years in due to the development of computer data bases like was discussed in Chapter 3. Suppose, for example, we wanted George Bush referred to Saddam Hussein as Hitler-like during the 1991 Persian Gulf War. Using NEZDI, we could request to count all of the articles in the New York Times (or any of a r newspapers, magazines, or broadcast transcripts) in which or "Saddam Hussein" and "Hitler" appeared for each given period. Once in the relevant file and depending on the access, the instruction might look something like this:

\[(\text{President Bush}) \text{ AND } (\text{Saddam Hussein}) \text{ w/10 Hitler})\] (September 1990)

or identify any article published during September 1990 in which or in which Mr. Hussein's name also appeared Hitler. Similar searches would set off Mr. Bush was referred to more often by title rather than Hitler on the map and would investigate the words used in speeches by Mr. Bush led or lagged behind other political leaders in editorial policy, or (6) numerous other hypotheses related to the frames of the conflict.

The results of item analysis may be at least as meaningful as those of component analysis in many instances. Is it more important that the Egyptians have made, say, seven conciliatory references in a given speech at the United Nations, or simply that they have made a conciliatory speech? Is it more important that the United States has sent to Iraq a note with four overt references to military intervention, three veiled references to the failure to disarm, and two sharply critical references to military expansion, or that the United States has issued a note that can be characterized as contentious in tone? In content analysis, sometimes, though by no means always, we risk losing sight of the forest for the trees or, by analogy, of the overall significance of a communication for its component parts. For this reason, we must be very careful to select a unit of analysis that will allow us the particular perspective on our research question that will prove most advantageous.

**UNDERTAKING A SUBSTANTIVE CONTENT ANALYSIS**

Once we have settled upon a population, a suitable sample, and an appropriate unit of analysis, we are ready to get under way. Substantive content analysis is based on a study of words, themes, and items that focuses on the substantive content of a given communication. Thus, in preparing to analyze these elements, we must anticipate their substance and we must define each possible observation in accordance with our expectations.

What this means, in effect, is that as the first step in undertaking a content analysis of this type, we must create a sort of dictionary in which we define each observable phenomenon we might make according to the particular category it fits. Suppose, for example, we are interested in studying all of the sixth-grade schoolbooks used in Havana, Cuba, last year and in identifying in them all references to Americans and the United States. Before we can proceed with such an analysis, we must define just what constitutes a salient reference. Do we look only for the words "American" and United States? If we do so, we may miss a great many salient references using such derogatory terms as Yankee aggressors, northern imperialists, gringos, invading forces at Guantanamo, and the outlaw regime in Washington. Moreover, some such phrases may have multiple meanings, only one of which refers to the United States. Consider the following problems in arithmetic:

1. If the Cuban people own 1,000 acres of crop lands and the neocolonialist aggressors steal from them 1,000 acres of crop lands, how many acres of crop lands will the Cuban people have left to support themselves?
2. If the African people own 1,000 acres of crop lands and the neocolonialist aggressors steal from them 1,000 acres of crop lands, how many acres of crop lands will the African people have left to support themselves?

If we count all references to "neocolonialist aggressors" as references to the United States, we will regard both problems as anti-American in tone. Anyone familiar with the recent history and ideology of Cuba, however, will regard only the first of these...
references as salient, with the second more likely to have been a pole at the former white government of South Africa.

The point is that we must anticipate not only the references we are likely to encounter but also the contextual elements of their use, and we must devise a thorough and systematic set of decision rules for judging each usage as it occurs. This problem is usually resolved by a combination of prototyping the population of communications to be analyzed (that is, reading through a selection of items to identify the types of salient references most likely to be encountered in a subsequent and more thorough analysis) and developing informed judgments about the contexts and uses of terms. Here, as in the later formal analysis, the observations of several researchers are preferred over those of one.

A more difficult problem arises when we must assign evaluations to salient references—when we must decide whether a particular reference is good or bad, favorable or unfavorable, pro or anti, and so forth—and when a series of such references must be ranked according to their intensity (which is most favorable, which is next most favorable, and so forth). Here we are concerned with developing and applying indicators that are sufficiently refined to tell us not only how the political actor feels but also how strongly the actor feels that way. A situation of this type is illustrated in Figure 10.1. The figure summarizes a number of ways in which a newspaper might endorse a candidate: if our goal is to determine which of several newspapers most strongly supports that candidate, then our immediate task is to decide how to rank these statements according to the intensity of support each reflects.

<table>
<thead>
<tr>
<th>FIGURE 10.1 Sample phrases in newspaper editorials endorsing a candidate (random order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best of the best</td>
</tr>
<tr>
<td>Our first choice</td>
</tr>
<tr>
<td>An outstanding leader</td>
</tr>
<tr>
<td>Lesser of two evils</td>
</tr>
<tr>
<td>Woman (man) of the hour</td>
</tr>
<tr>
<td>Recommend with reservations</td>
</tr>
<tr>
<td>Enthusiastically commend to your attention</td>
</tr>
</tbody>
</table>


the second more likely to have been a poke at the former Fhr. Africa.
must anticipate not only the references we are likely to en-
textual elements of their use, and we must devise a thor-
of decision rules for judging each usage as it occurs. This
ed by a combination of perceiving the population of com-
ed (that is, reading through a selection of items to identify
ences most likely to be encountered in a subsequent and
and developing informed judgments about the contexts and
in the later formal analysis, the observations of several re-
over those of one
s the act of reassigning to salient ref-
decide whether a particular reference is good or bad, fa-
ve or not, and so forth—and when a series of such refer-
cording to their intensity (which is most favorable, which
and so forth). Here we are concerned with developing and
sufficiently refined to tell us not only how the political
strongly the actor feels that way. A situation of this type is
The figure summarizes a number of ways in which a news-
andidate. If our goal is to determine which of several news-
ports that candidate, then our immediate task is to decide-
ents according to the intensity of support each reflects.

FIGURE 10.1 Sample phrases in newspaper editorials endorsing a
candidate (random order)

at of a bad lot
at available
the than the opponent
or best choice
best candidate in a crowded field
anything this policy is most in need of
outstanding leader
smart
part of the nation’s best
two of the wise
renowned
person of the hour
capable
self acceptable
completely acceptable
readily recommend
for our support
enthusiastically commend to your attention
get you to vote for

TABLE 10.1 Distribution of items in the Q-sort method

<table>
<thead>
<tr>
<th>Category (value)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution (percentage)</td>
<td>13</td>
<td>16</td>
<td>18</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Distribution (number of cases)</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Several techniques are available to assist us in making these decisions. Two of
the most prominent of these ranking techniques are the Q-sort method and par-
comparison scaling. Like the Thurstone scaling technique described in Chapter 9,
each relies upon the decision of a group of judges about the meaning or intensity of
a term, though here the judges may be drawn from the issuers of the communica-
tion, the receivers of the communication, a group of scholars familiar with the gen-
eral subject area under study, the general population, or the researchers themselves.
The techniques differ from one another and from the Thurstone method, however,
in the tasks they assign to these judges.

In this regard, the Q-sort is closest to the Thurstone procedure. The Q-sort
uses a forced distribution scale of 9 points in which 1 represents the lowest degree
of intensity of the attribute we wish to measure (for example, the least favorable)
and 9 the highest degree of intensity (for example, the most favorable). In con-
trast to the way it is done in the Thurstone procedure, no provision is made for neutral-
ity or arbitrary judgments. The purpose here is merely to rank in one evaluative
direction. Moreover, judges are far more constrained in their use of the scale than
in the Thurstone procedure. Each judge is given what amounts to a quota for each
of the categories in the scale—an expected number of words or phrases to be as-
signed thereto—and is instructed to distribute a given set of terms so as to meet
those quotas. The quotas are based on the assumption (not necessarily valid) that
the intensity of words and phrases will follow a normal distribution (with cases
clustered toward the middle of the scale and symmetrically decreasing in number
toward the extremes). The judges are thus forced to make relative judgments
about specific words and phrases and to decide where on the scale each one fits. The
procedure is illustrated in Table 10.1.

The table consists of three lines. The first represents the value assigned to each
category of the scale and ranges from 1 to 9. The second represents the percentage
distribution of all cases throughout the nine categories. These numbers are the quo-
tas assigned to each judge. Thus each is required to assign 5 percent of all cases to
category 1, 8 percent of all cases to category 2, 12 percent of all cases to category 3,
and so forth. The third line of the table represents the specific number of cases
these percentages determine for a given research problem. In the table, we have as-
sumed each judge will be asked to rank 50 words or phrases. The table gives us a line
that represents the percentages from line 2 applied to a total n of 50, and they
tell each judge how many of the statements must be assigned to each category.2 In
undertaking a Q-sort, lines 1 and 2 in the table will always be the same, and line 3
will be adjusted according to the number of statements or items to be ranked.

1 For a brief discussion of Q-sort methods, see Bruce McKeown and Dan Thomas, Q Methodology
(NEWbury Park, Calif: Sage, 1988).

2 Since this would result in either fractional entries or an anomaly in the distribution, cases have been
slightly redistributed toward the center of the table in order to preserve normality.
Once this ranking has been completed by a number of judges, we calculate the mean (average) category score for each statement and rank each accordingly. (The rationale for doing so is similar to that underlying the use of interval level statistics for analyzing data derived from Thurstone scales.) We then use these intensity rankings in assigning codes to the texts we are analyzing as the words or themes we have scored appear. In this way we substitute the collective wisdom of a number of judges for the arbitrary judgment of one researcher in deciding the meaning of communication content.

The goal of pair-comparison scaling is the same, but the procedure itself is rather different. Each item to be evaluated by the judges is paired with every other item, in a series of comparisons, and each judge is asked to decide which word or phrase in each pair is the stronger or more intense. Thus if we have five statements for comparison, each judge compares item 1 against items 2, 3, 4, and 5; item 2 against items 3, 4, and 5; and so forth—at each instance designating one or the other as more intense. By counting the number of times each statement is so designated by each judge, by totaling these numbers for each item for all judges, and by dividing by the number of judges (that is, by calculating the average score the judges as a group have assigned to a particular statement), we are able to arrive at a quantitative ranking of the intensity of each item. The higher its mean score, the stronger the judges consider a statement.

At least two problems are associated with both the Q-sort and pair-comparison procedures. First, both techniques rely entirely on the decisions of judges whose criteria for judgment may or may not be appropriate or consistent. The standards for expertise in such undertakings are not always clear, or at least are not always clearly stated, and as a consequence, the judgments themselves are open to question. Indeed, it is not uncommon for a single judge to assign different scores to the same statement in a series of identical tests. Because we are sampling content and not humans here, there is neither a clear reference population, as there is in Thurstone scaling, nor a set of underlying parameters to be approximated. The selection of judges, in other words, is extremely arbitrary. Thus the reliability of results derived by depending upon such judges may be minimal. In addition, judgmental methods can become tedious and cumbersome. A Q-sort of, say, 100 or 200 items that requires repeated determinations of minute shades of difference or a pair comparison of, say, 50 items that requires an examination of 1,225 different pairs \((n(n - 1)/2\), where \(n\) is the number of items\) can try the patience of even the best judge. As a result, such procedures must be approached with a good deal of caution.

UNDERTAKING A STRUCTURAL CONTENT ANALYSIS

In addition to, or in lieu of, words, themes, or other elements that denote the substantive content of a communication, several units of analysis are available that allow structural content analysis. Here we are less concerned with what is said than with how it is said, and while we must retain a concern with the subject matter, we measure something else.
king has been completed by a number of judges, we calculate a category score for each statement and rank each accordingly. Doing so is similar to that underlying the use of interval level scaling data derived from Thurstone scales X. We then use these intensity signs to the texts we are analyzing as the words or phrases appear. In this way we substitute the collective wisdom of a number of judges for the arbitrary judgment of one researcher in deciding the unification content.

An absolute comparison scaling is the same, but the procedure itself is rather different. Each item is evaluated by the judges' paired comparison method, and each judge is asked to decide which word or phrase is the stronger or more intense. Thus if we have five statements, a single judge compares item 1 against items 2, 3, 4, and 5; item 2 against items 3, 4, and 5; and so forth—i.e., in each instance designating one or the other of the five items as the stronger item. Then, by counting the number of times each statement is so designated by dividing by the number of judges, we are able to arrive at a quantitative ranking of each item. The higher its mean score, the stronger the judgment.

Problems are associated with both the Q-sort and pair-comparison methods. Both techniques rely entirely on the decisions of judges; judgment may or may not be appropriate or consistent. The erratic in such undertakings are not always clear, or at least are not stated, and as a consequence, the judgments themselves are indeed it is not uncommon for a single judge to assign different scores to a statement in a series of identical texts. Because we are not always humans here, there is neither a clear reference population, absolute scaling, nor a set of underlying parameters to be approximated. In other words, is extremely arbitrary. Thus the data derived from judges that determine such judgments may be minimal. In statistical methods can be tedious to use, the Q-sort, of course, is that requires repeated determinations of minute shades of air comparison of, say, 10 items that requires an examination of each item (n/2 - 1/2, where n is the number of items) can try the pa
dest judge. As a result, such procedures must be approached with caution.

URAL CONTENT ANALYSIS

In lieu of, words, themes, or other elements that denote the subject of communication, several units of analysis are available that allow for natural analysis. Here we are concerned with what is said than how it is said, and while we must retain a concern with the subject matter, we do not.

<table>
<thead>
<tr>
<th>ID Number</th>
<th>Article Type</th>
<th>Publication Date</th>
<th>Candidate Name</th>
<th>Newspaper</th>
<th>Frequency</th>
<th>Prominence</th>
<th>Graphics</th>
<th>Treatment</th>
<th>Column Inches</th>
<th>Candidate Column Inches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 10.2: Typical coding sheet for structural content analysis**

We may be concerned, for example, with the amount of space or time devoted to a given subject in a particular source. How many words or column inches of newspaper coverage have been accorded each candidate in a particular election campaign? How many articles or pages in political science journals published in the United States are devoted each year to an analysis of governments and politics in Africa? Has the number changed, or has it remained constant over the past three decades?

Alternatively, we might be concerned with other, and perhaps more subtle, aspects of the communication format. Is a particular news item accompanied by a photograph or illustration of some sort? Those that are have been found to attract more attention from readers than those that are not. How large a headline accompanies a news item? Does coverage of a particular subject receive front-page prominence, or is it buried among the trust ads? In answering questions like these, we are less concerned with subtleties of meaning than with styles of presentation. We watch for the presence or absence of prominence, and the extent of treatment of general themes rather than for substantive nuance. The result in many cases is an analysis whose measurements are much more reliable than those employed in a more substance-oriented study (since there is less ambiguity built into the indicators), but one whose lessons may, as a direct consequence, be less rich.

Figure 10.2 illustrates a typical coding sheet for recording data from a structural analysis of content. It is drawn from a study of newspaper coverage of congressional elections. The unit of analysis for this particular study was the candidate insertion, which was defined as any newspaper item that mentioned by name or implication any candidate for Congress in the district in which the newspaper was distributed. Thus each row on the coding sheet summarizes the characteristics of a single candidate insertion.

While we shall wait until Chapter 14 to discuss the assignment of particular numbers to each column on such a sheet, it is worth pointing out here the type of information that is recorded. After each item was assigned a unique identification code,
number, it was classified according to type (news story, feature article, editorial, letter to the editor), the date of publication, the candidate it referred to, the newspaper in which it appeared, the general preferences expressed in the item (if any), its prominence of placement (front page, inside page), the presence or absence of accompanying photographs or drawings, reference to the candidate in the headline of the item, the primary content of the item (news of a campaign event, content of a speech, endorsement), the overall size of the insertion, and the proportion of the insertion actually relating to the candidate in question. The point to note here is that measurements of this kind require only a general concern with the actual substance of each insertion rather than the highly detailed and specific focus necessary in the substantive approach discussed earlier. As a result, structural content analysis is usually easier to design and carry out, and therefore less expensive and often more reliable, than substantive content analysis. And though its results may be less satisfying in that they provide us with what amounts to a sketch of a communication rather than a finished portrait, those results often prove entirely adequate in answering a particular research question.

SPECIAL PROBLEMS IN THE USE OF CONTENT ANALYSIS

Although content analysis is a relatively inexpensive technique that draws on a relatively accessible database and although there are few special ethical dilemmas that we are likely to encounter in undertaking it (unless we are analyzing confidential or classified communications), we must still be careful to avoid several potential difficulties when we use this method.

For one thing, we must be aware that communications are issued, and may be specifically designed, for a purpose, whether it be description, persuasion, exhortation, direction, self-protection, or even obfuscation. In analyzing such communications, therefore, we must attempt to interpret their content in the context of their apparent purpose. For example, it is common to find in the Chinese press statements of the type, "All of the Chinese people believe that the new agricultural policy is a major step forward in the progress toward social revolution." Taken at face value, such statements are demonstrably false, since not every one of many millions of people would be aware of, let alone agree upon the value of, any single policy. From this perspective, we might be inclined to view these statements as the most blatant form of propaganda. We have learned from studying the Chinese press, however, that statements of this type are not printed for purposes of external propaganda at all, but rather are intended to suggest to the Chinese people themselves the beliefs that their government wishes them to hold. In other words, the purpose of such statements of consensus is not descriptive, but directive. Knowing this, we may interpret them as useful indicators of the policy interests of the Chinese leaders rather than as meaningless items of propaganda, and we may employ them to some advantage. The purpose of a communication, then, can provide an important context for understanding its content, and we must attempt, when possible, to ferret out this information. Similarly, the distribution that is accorded a particular item of communication can have significant implications for its meaning. A pamphlet that circulates only
that draws on a rela-
tional dimension that lying could be-
doing is the Chinese press for
the new agricultural motion. Taken at
Every one of many
value of any single
these statements are
the Chinese
purposes of exter-
ternally imposed
 Chinese people
In other words,
wife, but directive,
the policy approa-
ents, and we may
mention, then, can.
As of communication
that circulates only
among Chinese dissidents, a solicitation letter from a candidate or special interest
in a particular mailing list, a document that
reader for a small group of people—each is an example of a commu-
nications with a limited or specialized distribution. Even a newspaper that is gen-
erally available may have a limited or specialized clientele. The New York Times, for
example, has a readership that is generally more affluent and better educated than
that of the New York Daily News, yet both are readily available to all of the city's
newspaper readers. The Wall Street Journal has nationwide distribution, but its
readership does not extend equally to all socioeconomic classes. Very often, if we
are to assess properly the significance of a communication, we must know whom it
reaches. Whether by judgment or evidence (which we have when a document is accom-
panied by a notice of sending and perhaps initiated by all who have read it), we may
rely on an audience survey (such as the kind usually taken by newspapers to document their cir-
culation claims), we must attempt to measure or to estimate how widely a message
has been disseminated and to whom. Having this information enables us to judge
the value or the importance of the material we analyze.

Third, we must try to gauge the degree of our own access to the items at issue.
Have we been provided with free choice over the materials we shall analyze? Are
these materials available in an unbiased manner (that is, do we have access to all of
them), or has some external control been imposed by someone other than the re-
searcher? Or we, for instance, have access only to documents that have been decla-
sified, only to Chinese newspapers that are published for and distributed primarily to
foreigners, only to records of formal meetings of a government commission? The
issue here is one of generalizability, and the question is what the research popula-
tion itself, not to mention the sample, is truly representative. If it is not, the
researcher may, if not exercising care, at the very least be misled and at the worst be
manipulated.
The problem in each of these instances is that the information we require to make
informed judgments may simply be unavailable. We may not know, and may
be unable to ascertain, the purpose of a communication, its distribution, or the de-
gree of access to it that we have been accorded. The slayers here are manifold,
and the context analyst must be sensitive to them. We must not allow appearances
to cloud our judgment but must maintain a healthy skepticism regarding our data
as long as these questions remain unanswered. That is not necessarily to say that
we should not undertake context analysis under conditions of uncertainty, but
merely that we should not lose sight of the uncertainty itself once the analysis is
under way.

Finally, we should say a few words about intercoder reliability. With the excep-
tion of content analysis procedures that have been thoroughly computerized (several
programs embodying concept dictionaries and search or count procedures have
been developed), all content analysis depends on human judgments about commu-
nication content. Judges, after all, do not analyze themselves. They are judged
and graded, counted and classified by Homo sapiens in the form of the researcher.
Therefore individual researchers may differ from one another in their understanding

of the content of a given communication. Indeed, only when some degree of consensus can be reached about that meaning can we have real confidence in our measurements. Inter Ober reliability is the term political scientists use to describe the degree of that consensus. The higher it is, the better. In general, inter coder reliability may be promoted by taking three basic steps:

1. Operationalize all variables carefully and thoroughly. Make sure that all meanings have been clearly stated and as many ambiguities as possible have been eliminated. In effect, this will create common standards of judgment that can be used consistently in classifying and measuring content.

2. Use as many observers (coders) as possible. The larger the number of sub- members to the consensus, the more confidence we can have in it. This may, of course, mean more work and considerable duplication of effort (and, if proper training is not provided, it carries a risk of increased measurement error), but the payoff can be substantial. The limiting factor here is usually cost.

3. Maximize the interaction among the observers. Hold common practice ses- sions and argue out all differences of interpretation so that ultimately the consensus extends not only to the data but also to the real meanings of the operational definitions themselves.

The success of this process can be measured in either of two ways, both of which draw upon statistical concepts that we develop more fully in Chapter 17. One approach, used primarily in substantive content analysis, is to have all observers who are working on a given project analyze and code independently (assign their own numerical values to) the same communication, then to calculate a statistic called a correlation coefficient (Pearson’s r) among the codes recorded by each pair of ob- servers. This coefficient (discussed in detail in Chapter 17) measures the degree of correspondence in the judgments of the researchers on whether and how often a particular word or theme is present. The coefficient ranges from -1 to +1, and read- ings of +.90 or better are usually interpreted as indicating a high degree of inter coder reliability. An alternative measure may be more useful for structural content analysis, in which we are less concerned with the treatment of themes than with their presence or absence and in which duplicated measurement is less necessary. Here we treat the differences between observers as a variable in their own right, and we ask whether that variable is associated with systematic differences in any other variable we have measured. In other words, we are concerned with the possibility that one or more observers have recorded results consistently differently from the others. If it can be assumed that all cases have been distributed to the observers in an unbiased manner (some effort is generally made to distribute them randomly), any systematic differences we observe are more likely to be the result of differences between coders than of underlying differences in the cases that happen to have been assigned.

---

1 More complex variants of this procedure with applications to the measurement of intensity may be found in Robert C. North et al. Concept Analysis: A Handbook with Applications for the Study of International Crime (Evanston, III.: Northwestern University Press, 1965), chaps. 4 and 5.
to the abstrait observer. The coefficients of intercoder reliability here take the form (1 - r) where r is a measure of the variance in each subject variable that is accounted for by differences between coders. By subtracting the *observer error* from 1, we obtain the proportion of error free observations. The coefficient is calculated separately for each variable and should exceed .90 if we are to have confidence in the reliability of our measures.

In sum, content analysis is a widely applicable technique with advantages in cost, sample size, and, often, access to data. Perhaps more than any other technique, however, it demands careful operationalization of all variables and constant monitoring of the process of observation. Its results may be highly informative, but they must be understood in a context that it is often beyond the scope of the content analysis technique itself to describe. For this reason, content analysis is often used to best effect in combination with other datagathering methods (surveys, direct observation) in what are termed multimethod designs.

**SUGGESTIONS FOR FURTHER READING**

The three most useful texts dealing with the techniques of content analysis and their application to political science research are Robert C. North et al., *Content Analysis: A Handbook with Applications for the Study of International Crisis* (Beverly Hills, Calif.: Sage, 1985), which focuses primarily on methods of scaling intensity and measuring reliability; Philip J. Boczar, ed., *The General Inquirer* (Cambridge, Mass.: Massachusetts Institute of Technology Press, 1969), which discusses and gives examples of the application of a widely used computerized program for content analysis based on semantic structures; and Claus Krapp and Irmgard Elster, *Content Analysis as Introduction to the Methodology of Computer-Assisted Coded Data Collection* (Beverly Hills, Calif.: Sage, 1985), which is a broad introduction to content analysis techniques.

A more recent survey can be obtained from Robert E. Smith, *Basic Content Analysis* (Beverly Hills, Calif.: Sage, 1985).


*The formula and computational procedure for the metric, which is a measure of association between one nominal and one interval variable, may be found in Louis C. Parks, *Elements of Applied Statistics for Students of Behavioral Science* (New York: Wiley, 1962), pp. 123-28.*