Content Analysis

Suphat Sukamolson

CULI

12/16/2010 สุพัฒน์ สุกมลสันต์ 1

An Introduction to Content Analysis

- Content analysis is a research tool used to determine the presence of certain words or concepts within texts or sets of texts.

- Researchers quantify and analyze the presence, meanings and relationships of such words and concepts, then make inferences about the messages within the texts, the writer(s), the audience, and even the culture and time of which these are a part.

Uses of Content Analysis

1. Detect the existence of ideas, concepts and "truth" hidden in the texts.

2. Identify the intentions, focus or communication trends of an individual, group or institution.

3. Describe attitudinal and behavioral responses to communications of the informants.

4. Determine psychological or emotional state of persons or groups.

5. Reveal the differences in communication content of persons or groups.

Two Main Types of Content Analysis

1. **Conceptual Analysis or Thematic Analysis**

   – A concept is chosen for examination, and the analysis involves quantifying and tallying its presence. The focus here is on looking at the occurrence of selected terms within a text.

2. **Relational Analysis or Semantic Analysis**

   – It begins with the act of identifying concepts present in a given text or set of texts, and seeks to go beyond presence by exploring the relationships between the concepts identified.

Six Questions before Doing Content Analysis

According to Krippendorff (1980), six questions must be addressed in every content analysis:

1. Which data are analyzed?
2. How are they defined?
3. What is the population from which they are drawn?
4. What is the context relative to which the data are analyzed?
5. What are the boundaries of the analysis?
6. What is the target of the inferencces?

สุพัฒน์ สุกมลสันต์

8 Steps
for Conducting Conceptual Analysis

สุพัฒน์ สุกมลสันต์

Step 1: Decide the Level of Analysis

- First, the researcher must decide upon the level of analysis. For example, *to investigate factors leading to be high achievers in English*, the researcher must decide whether to code for a single word, such as "love/like," or for sets of words or phrases, such as "family supports."

สุพัฒน์ สุกมลสันต์

Step 2: Decide How Many Concepts to Code For

- Decide how many different concepts to code for. This involves developing a pre-defined or interactive set of concepts and categories. The researcher must decide whether or not to code for every single positive or negative word that appears, or only certain ones that the researcher determines are most relevant to the research objective.

Step 3: Decide Whether to Code for Existence or Frequency of a Concept

- This is important, because it changes the coding process.

- When coding for existence, "love/like" would only be counted once, no matter how many times it appeared.

- This would be a very basic coding process and would give the researcher a very limited perspective of the text. However, the number of times "love/like" appears in a text might be more indicative of importance.

Step 4: Decide on How You Will Distinguish Among Concepts

- Decide on the level of generalization, i.e. whether concepts are to be coded exactly as they appear, or if they can be recorded as the same even when they appear in different forms.

- For example, "love" might also appear as "like" or "crazy." The research needs to determine if the 3 words mean radically different things to him/her, or if they are similar enough that they can be coded as being the same thing, i.e. "positive attitude."

Step 5: Develop Rules for Coding Your Texts

- Create translation rules that will allow the researcher to streamline and organize the coding process so that he/she is coding for exactly what he/she wants to code for. Developing a set of rules helps the researcher insure that he/she is coding things consistently throughout the text, in the same way every time.

Step 6: Decide What To Do with "Irrelevant" Information

- Decide whether irrelevant information should be ignored or used to reexamine and/or alter the coding scheme. In the case of this example, words *like* in "like my friend" and "love my mother," as they appear in such contexts, would be ignored.

- The researcher may either deleting or skipping over unwanted material, or viewing all information as relevant and important and using it to reexamine, reassess and perhaps even alter his/her coding scheme.

Step 7: Code the Texts

- Code the text. This is done either by hand, i.e. reading through the text and manually writing down concept occurrences, or through the use of various computer programs.

- Coding with a computer is one of contemporary conceptual analysis' greatest assets. By inputting one's categories, content analysis programs can easily automate the coding process and examine huge amounts of data.

- However, automation is very dependent on the researcher's preparation and category construction. When coding is done manually, a researcher can recognize errors far more easily.

Step 8: Interpret Your Results

- Once the coding is done, the researcher examines the results and attempts *to draw whatever possible conclusions and generalizations*.

- Given that the conceptual analyst is dealing only with quantitative data, the levels of interpretation and generalizability are very limited. The researcher can only extrapolate as far as the data will allow.

- However, it is possible to see *trends*, for example, that are indicative of much larger ideas. Using the example from step three, if the concept "love/like" appears 50 times, compared to 15 appearances of "family supports," then the researcher can pretty safely extrapolate that "positive attitude" is a key factor to be high achievers.

8 Steps
for Conducting Relational Analysis

สุพัฒน์ สุกมลสันต์

Step 1: Identify the Question

- The question is important because it indicates where you are headed and why. Without a focused question, the concept types and options open to interpretation are limitless and therefore the analysis difficult to complete.

- For example, for the "Factors making students be high achievers in English" might be:

  – What make the informants have high achievement in learning English?

สุพัฒน์ สุกมลสันต์

Step 2: Choose a Sample or Samples for Analysis

- Select the texts that the informants have already provided.

- For relational content analysis, the primary consideration is how much information to preserve for analysis.

- One must be careful not to limit the results by doing so, but the researcher must also take special care not to take on so much that the coding process becomes too heavy and extensive to supply worthwhile results.

สุพัฒน์ สุกมลสันต์

Step 3: Determine the Type of Relationships to Examine

- Determine what type or types of relationships you would like to examine, e.g.

1. **Affect extraction:** This approach provides an emotional evaluation of concepts explicit in a text. It is problematic because emotion may vary across time and populations.

2. **Proximity analysis**: This approach is concerned with the co-occurrence of explicit concepts in the text. For example, if a researcher believes that "motivation" is a key factor to be a high achiever, the whole text is scanned for this word.

3. **Cognitive mapping:** This approach analyzes of the results from the two previous approaches by representing the relationships of ideas/concepts visually for comparison. Whereas the first 2 approaches function primarily within the preserved order of the text, cognitive mapping attempts to create a model of the overall meaning of the text.

Step 4: Reduce the Text to Categories and Code for Words or Patterns

- At the simplest level, a researcher can code merely for existence. This is not to say that simplicity of procedure leads to simplistic results.

- Many studies have successfully employed this strategy.

- For example, to study the changes of concept in a classroom, a researcher may look at the change in the presence of concepts over the course of the semester, and compare a map analysis from the beginning of the semester to one constructed at the end.

Step 5: Explore the Relationships Between Concepts

Points to consider:

1. **Strength of Relationship**: Refers to the degree to which two or more concepts are related. These relationships are easiest to analyze, compare, and graph when all relationships between concepts are considered to be equal.

2. **Sign of a Relationship**: Refers to whether or not the concepts are positively or negatively related.

3. **Direction of the Relationship**: Refers to the type of relationship categories exhibit. Coding for this sort of information can be useful in establishing, for example, the impact of X on Y.

สุพัฒน์ สุกมลสันต์

Step 6: Code the Relationships

- One of the main differences between conceptual analysis and relational analysis is that the statements or relationships between concepts are coded.

- At this point, assign codes to the relationships in an effort to determine whether the ambiguous words used in the texts are just fillers, or hold information for the study.

Step 7: Perform Statistical Analyses

- Conduct statistical analyses of the data coded during the relational analysis. This may involve exploring for differences or looking for relationships among the key terms that have been identified in the study.

- For more information about conducting statistical analysis, learn more about
  – Descriptive and inferential statistics
  – Parametric and non-parametric tests
  – Trend analysis
  – Path Analysis using LISREL and AMOS

Step 8: Map the Representations

- In addition to statistical analysis, relational analysis often leads to viewing the representations of the concepts and their associations in a text (or across texts) in a graphical form.

- Relational analysis is also informed by a variety of different theoretical approaches, e.g.
  - **Linguistic content analysis**
  - **Decision Mapping**
  - **Mental Model**

Some Theoretical Influences on Relational Analysis

1. **Linguistic Approach**

    It focuses to analyze texts on the level of a linguistic unit, typically single clause units, e.g. Discourse Analysis and Unit-Thought Analysis.

    –   One example of this type of research is Gottschalk (1975), who developed an automated procedure which analyzes each clause in a text and assigns it a numerical score based on several emotional/psychological scales.

    –   Another technique is to code a text grammatically into clauses and parts of speech to establish a matrix representation (Carley, 1990).

Some Theoretical Influences on Relational Analysis

2. Cognitive Science Approach

- It includes the creation of decision maps and mental models to represent the relationships between ideas, beliefs, attitudes, and information available to an author when making a decision within a text.

- These relationships can be represented as logical, inferential, causal, sequential, and mathematical relationships.

- Typically, two of these links are compared in a single study, and are analyzed as networks. For example, Heise (1987) used logical and sequential links to examine symbolic interaction.

Some Theoretical Influences on Relational Analysis

3. Mental Model Approach

It is a group or networks of interrelated concepts that are thought to reflect conscious or subconscious perceptions of reality.

- According to cognitive scientists, internal mental structures are created as people draw inferences and gather information about the world. Mental models are a more specific approach to mapping because they can be numerically and graphically analyzed.

Example 1
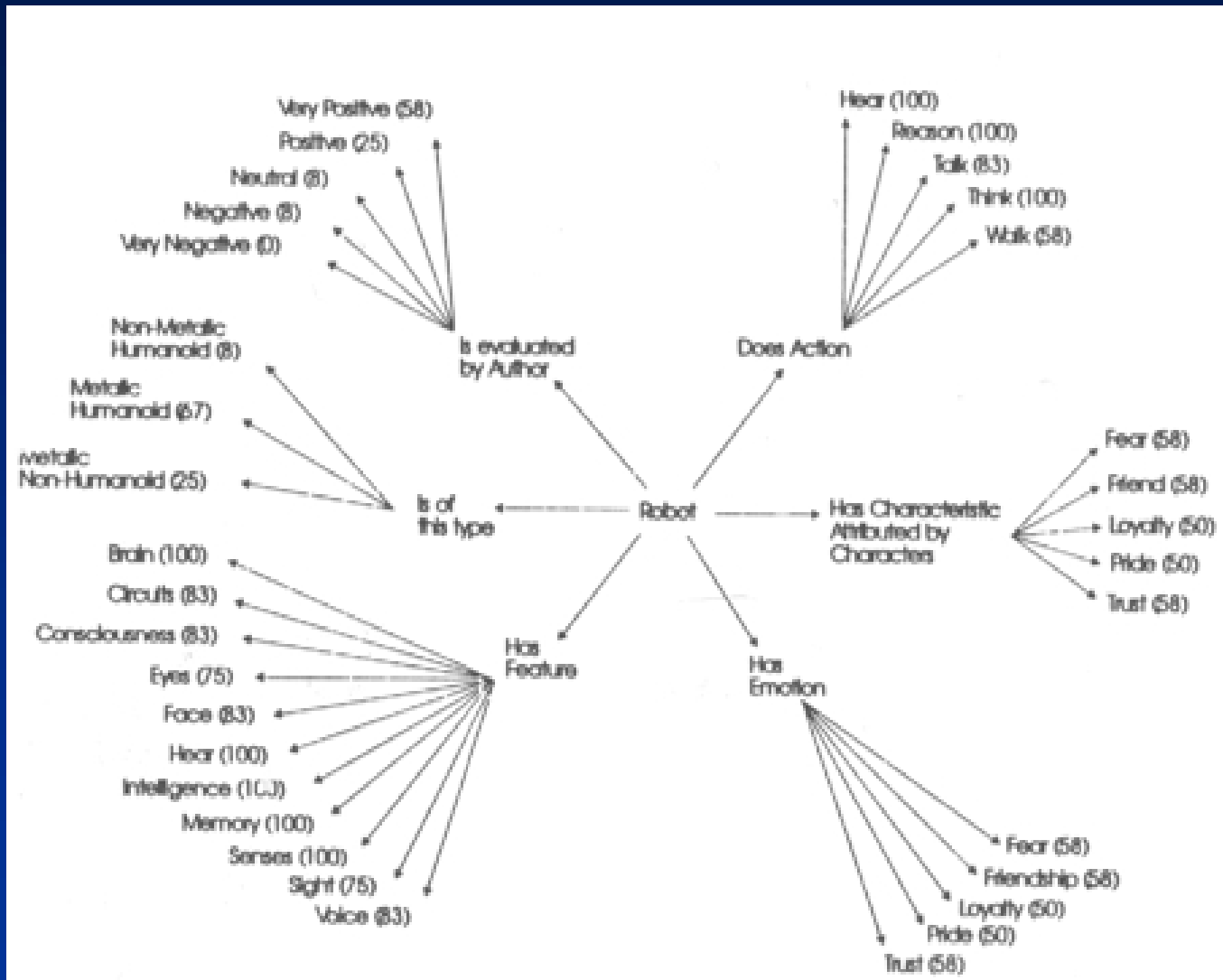
Example 2
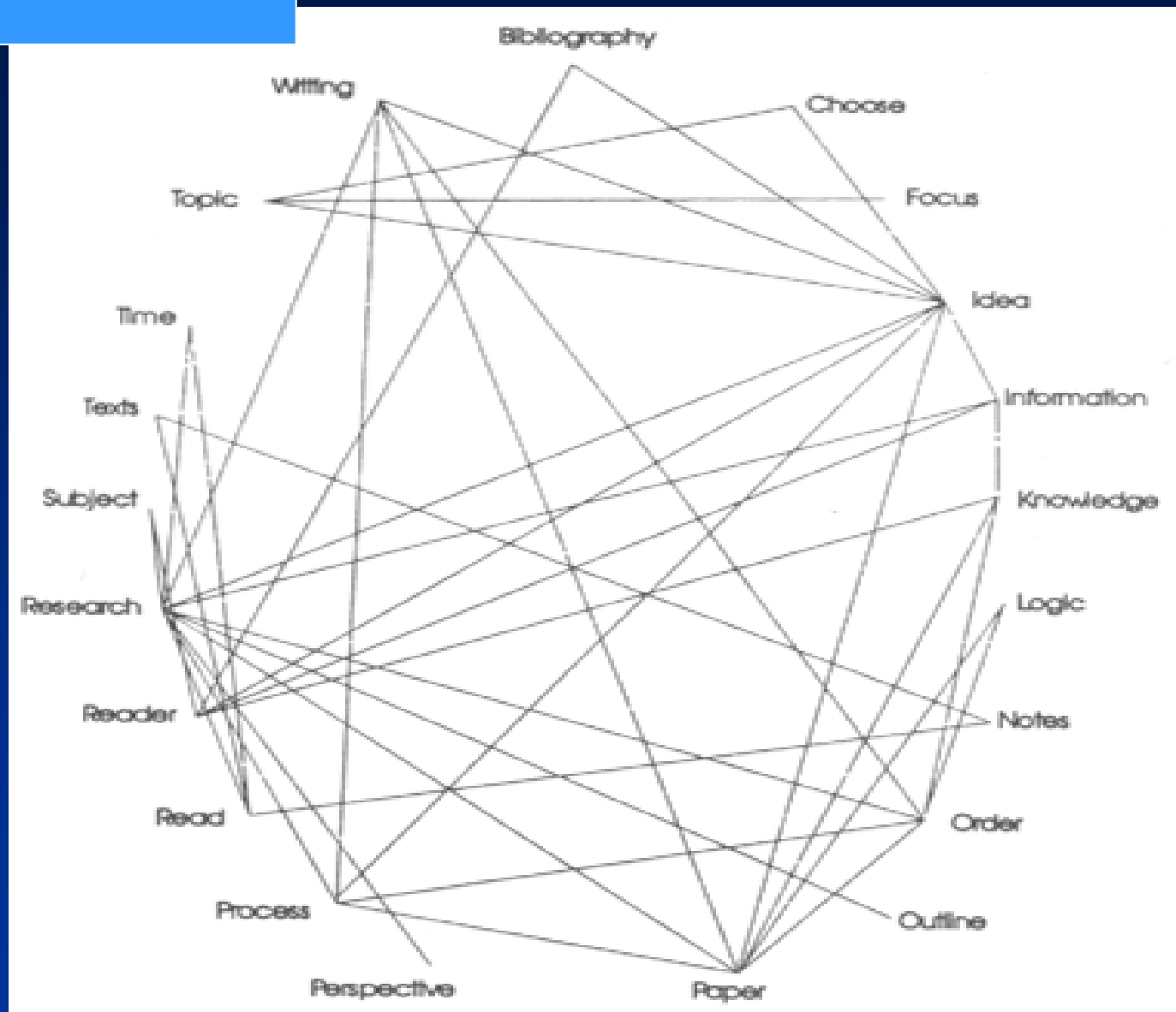
How to assign codes?

There are two approaches to coding data: Emergent and *a priori* codings.

A. E*mergent Coding*

1. At least 2 coders/researchers independently review a sample of text and come up with a set of codes that form a checklist.

2. They compare notes and reconcile any differences that show up on their initial checklists.

3. They use a consolidated checklist to independently apply coding.

4. The researchers check the reliability of the coding (a 95% agreement is suggested or 0.80 for Cohen's kappa).

If the level of reliability is not acceptable, then they repeat the previous steps. Once the reliability has been established, the coding is applied on a large-scale basis. The final stage is a periodic quality control check.

สุพัฒน์ สุกมลสันต์

How to assign codes?

B. Priori Coding

- When dealing with a priori coding, the categories are established prior to the analysis based upon some theory.

- Professional colleagues agree on the categories, and the coding is applied to the data.

- Revisions are made as necessary, and the categories are tightened up to the point that maximizes mutual exclusivity and exhaustiveness.

Coding Units

There are several different ways of defining coding units, e.g.

1. Physical Unit

Use their natural or intuitive borders of the text to be analyzed. For instance, newspaper articles, letters, or poems all have natural boundaries.

2. Syntactic Unit

Use syntax created by the author, such as words, a phrase or a sentences.

3. Referential Unit

Use a word and a phrase to represent other words or phrases. For example a paper might refer to Former Prime Minister Taksin as "Prime Minister Taksin," "the 23rd Prime Minister," or "Squared Face Person."

4. Propositional Unit

Break down the text in order to examine underlying meanings. For example, in a sentence that would read, "I love to talk to friends in class while the teaching is teaching English," we would break it down to: The student does not pay attention to the teacher/ The English lesson is boring.

Computer Programs for Content Analysis

- Now many computer programs can be used for Content Analysis, e.g.
  - Hyper Transcribe
  - Hyper Research
  - Nud*ist
- They are used to assign codes, to find frequency counts of texts, audio files and graphic files, and to perform some descriptive statistics.

Advantages of Content Analysis

- Look directly at communication via texts or transcripts, and hence gets at the central aspect of social interaction.

- Allow for both quantitative and qualitative operations

- Provide valuable historical/cultural insights over time through analysis of texts.

- Allow a closeness to text which can alternate between specific categories and relationships and also statistically analyzes the coded form of the text.

- Can be used to interpret texts for purposes such as the students' learning strategies, learning habits, etc.

- Is an unobtrusive means of analyzing interactions.

- Provide insight into complex models of human thought and language use.

Disadvantages of Content Analysis

- Can be extremely time consuming

- Is subject to increased error, particularly when relational analysis is used to attain a higher level of interpretation

- Is often devoid of theoretical base, or attempts too liberally to draw meaningful inferences about the relationships in a study.

- Is inherently reductive, particularly when dealing with complex texts.

- Tends too often to simply consist of word counts.

- Often disregards the context that produced the text, as well as the state of things after the text is produced.

- Can be difficult to automate or computerize.

Issues of Reliability

The reliability of a content analysis study refers to its:

1. *Stability (intra-rater reliability)*, or the tendency for coders to re-code the same data in the same way over a period of time consistently;

2. *Reproducibility (inter-rater reliability)*, or the tendency for a group of coders to classify categories membership in the same way; and

Reliability problems usually grow out of the ambiguity of word meanings, category definitions, or other coding rules.

So, the researcher should have a list of codes and use them consistently.

สุพัฒน์ สุกมลสันต์

How to find agreement between coders/raters

Use Cohen's Kappa, which approaches 1 as coding is perfectly reliable and goes to 0 when there is no agreement other than what would be expected by chance.

Kappa is computed as:

$$k = \frac{p_a - p_c}{1 - p_c}$$

where:

p_a = proportion of units on which the raters agree

p_c = the proportion of units for which agreement is expected by chance.

		Rater 1			Marginal Totals
		Academic	Emotional	Physical	
Rater 2	Academic	.42 (.29)*	.10 (.21)	.05 (.07)	.57
	Emotional	.07 (.18)	.25 (.18)	.03 (.05)	.35
	Physical	.01 (.04)	.02 (.03)	.05 (.01)	.08
		.50	.37	.13	1.00

$$p_a = 0.42 + 0.25 + 0.05 = 0.72$$

$$p_c = 0.29 + 0.18 + 0.01 = 0.48$$

$$k = \frac{0.72 - 0.48}{1 - 0.48} = 0.462$$

Landis & Koch (1977, p.165) have suggested the following benchmarks for interpreting kappa:

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00– 0.20	Slight
0.21– 0.40	Fair
0.41– 0.60	Moderate
0.61– 0.80	Substantial
0.81– 1.00	Almost Perfect

More about reliability of coding

- If there are more than 2 coders (researchers) assigning codes to the same text, many methods can be used to find reliability coefficients, e.g.

  – Guilford's Method, Hoyt's ANOVA Method, Intra-Class Correlation Method and Kendall's Concordance Method

- If one coder (researcher) analyzes a text more than 2 times repeatedly, Pearson's Chi-squared Test can be used instead.

Points to consider

1. To make valid inferences from the text, it is important that the classification procedure be reliable in the sense of being consistent. Different people should code the same text in the same way.

2. Reliability problems usually grow out of the ambiguity of word meanings, category definitions, or other coding rules.

Issues of Validity

- The validity of a content analysis study refers to the correspondence of the *categories* to the *conclusions*, and the *generalizability* of results to a theory.

The researcher should have a list of multiple classifiers or codes and use them consistently.

- For example, a content analysis study might measure the occurrence of the concept category "communist" in a text. Using multiple classifiers, the concept category can be broadened to include synonyms such as "red," "Soviet threat," "pinkos," "godless infidels" and "Marxist sympathizers." "Communist" is held to be the explicit variable, while "red," etc. are the implicit variables

สุพัฒน์ สุกมลสันต์

Issues of Validity

- The *generalizability* of one's conclusions is very dependent on how one determines concept categories, as well as on how reliable those categories are.

The researcher should consider how much he/she can interpret, infer or extrapolate the findings and try best not to be over- or under-generalize the results.

If at all possible, the researcher should try to have some sort of validation study built into the design. In qualitative research, validation takes the form of triangulation. Triangulation lends credibility to the findings by incorporating multiple sources of data, methods, investigators, or theories.

References

Colorado State University. (1993-2007). **Content Analysis**.
Retrieved on Jan. 7, 2008 from
http://writing.colostate.edu/index.cfm

Stemler, Steve (2001). **An overview of content analysis**.
Practical Assessment, Research & Evaluation, 7(17). Retrieved
on November 26, 2007 from
http://PAREonline.net/getvn.asp?v=7&n=17.