

e-RESEARCH

Methods, Strategies, and Issues

TERRY ANDERSON
(University of Calgary)
HEATHER KANUKA
(Athabasca University)

2020 December 31

b



Contents

Chapter 12

CONTENT ANALYSIS OF ONLINE DOCUMENTS

12.1 THE CODING PROCESS

12.2 DEFINING THE UNIT OF ANALYSIS

Quantitative content analysis usually deals with manifest variables, since they are the only ones to which very high levels of reliability can reasonably be expected among multiple coders. An excellent example of a large quantitative content analysis is Project H, an international research project that involved 107 researchers in some twenty different countries (Allbritton, 1996). Project H analyzed a large number of mostly manifest variables identified in over 100 Usenet and email list discussions. The complicated processes involved in coordinating such a large research team as well as excellent description of methodological and ethical issues involved in this exemplar content analysis project are provided in a 1996 article by the principal investigators of Project H, Sudweeks and Rafaeli. Their article is appropriately entitled "How do you get a hundred strangers to agree: Computer mediated communication and collaboration."

THE CODING PROCESS

Our discussion of quantitative content analysis illustrates the difference between qualitative and quantitative content analysis, yet there are also similarities. Despite our concern with differentiating between different methods that use the same terms, we are cognizant of the research axiom "there can be no quantification without qualification." This expression underlines the necessity of first being able to identify and categorize the variable before we can count, assign a category, or interpret the content. The challenge is twofold: first we must be able to find a commonly understood and easily distinguishable unit of analysis, and, second, we must be able to reliably and consistently classify each of these units.

DEFINING THE UNIT OF ANALYSIS

The following hypothetical example of an email extract among students in an e-learning context provides us with an example of the challenge, that confront the e-researcher engaged in quantitative content analysis.

HI folks,

I really hate the way we have to answer ALL THESE QUESTIONS before getting the data right-how about you? I have very little time this week for this assignment. The last time I tried this I got bogged down with chapter 3 do we really have to know about semiotics, I want to get onto the project analysis first then decide if we need all this theory how about I'll do the first question and you guys do the rest.

Your-confused-comrade-in-arms,

Terry

How many sentences are there in this extract? How many words? How many paragraphs? How many ideas? The difficulty of answering these questions illustrates a number of serious challenges in both qualitative and quantitative content analysis. The challenge of identifying the unit of analysis (or, as it is sometimes termed, unitizing) is critical to reliable content analysis (Rourke, Anderson, Garrison, & Archer, 2001).

The Sentence as Unit of Analysis

In text transcript analysis, a common method is to use a grammatically defined unit of analysis. Some e-researchers have used the sentence as the unit of analysis (Fahy, Crawford, Ally, Cookson, Keller, & Prosser, 1999; Hillman, 1999), however, as the sample email illustrates, online text dialogue often follows its own rules. These rules create a relaxed grammar that falls somewhere between text and voice communication. This often makes the identification of sentences problematic. The sentence as a unit of analysis is also challenging in that the number of sentences can be very large, making for a time-consuming process. In addition, it may also be difficult to identify a relevant variable in each and every sentence.

The Paragraph as Unit of Analysis

Other e-research content analysis (Hara, Bonk, & Angeli, 2000) have chosen the paragraph as the unit of analysis. This unit has an advantage in that it is larger, requiring fewer decisions of the researchers. In addition it should be "a distinct division of written or printed matter that begins on a new, usually indented, line, consists of one or more sentences, and typically deals with a single thought or topic or quotes one speaker's continuous words" (www.dictionary.com). However, our email example shows that often users do not write in clearly defined paragraphs, leaving an unfortunate amount of interpretation to the e-researcher. As the size of the unit expands, so does the likelihood that the unit will encompass multiple variables. Conversely, one variable may span multiple paragraphs. Our experience does not support Hara, Bonk, and Angeli's optimism that "college-level students should be able to break down the messages into paragraphs" (p. 9). Further, once the syntactical criteria are lost, the definition of the unit as a paragraph becomes meaningless, and what the coders are identifying are, in fact, arbitrary blocks of text. Hara, Bonk, and Angeli's ad hoc coding protocol reveals these problems: "when two continuous paragraphs dealt with the same ideas, they were each counted as a separate unit. And when one paragraph contained two ideas, it was counted as two separate units" (p. 9). Thus, the selection of the paragraph presents a very problematic unit of analysis.

The Message as Unit of Analysis

The full email or computer conferencing message has also been the unit of analysis (Ahern, Peck, & Laycock, 1992; Marttunen, 1997). This unit has important advantages. First, it is objectively identifiable. Unlike other units of analysis, multiple raters can agree perfectly on the total number of cases. Second, it produces a manageable set

of cases. Marttunen and Ahern, Peck, and Laycock recorded a total of 545 and 185 messages respectively, a total that would have been considerably larger if the messages had been subdivided. Third, it is a unit whose parameters are clearly determined by the author of the message—they explicitly choose when to end the message. The major disadvantage of the message as the unit of analysis is that often more than a single idea of interest is expressed in a single message. We are familiar with email authors who can never seem to end their message, without “one final point,” while others regularly author extremely sparse messages. Thus, defining the length of a message is as challenging as defining how long a piece of string is. These concerns with grammatically determined units of analysis have promoted some e-researchers to look for units that are defined, not by grammar or syntax, but by meaning.

The Meaning Unit as Unit of Analysis

Henri (1991) rejected the process of a priori and authoritatively fixing the size of the unit based on criteria that are not directly related to the construct under study. Instead, she proposes a thematic or meaning unit. Budd, Thorp, and Donohew (1967) define thematic units as “a single thought unit or idea unit that conveys a single item of information extracted from a segment of content” (p. 34). Quoting from Muchielli, Henri (1991) justifies this approach by arguing that “it is absolutely useless to wonder if it is the word, the proposition, the sentence or the paragraph which is the proper unit of meaning, for the unit of meaning is lodged in meaning” (p. 134). The task of explaining what this enigmatic statement meant to pragmatic researchers was taken up by Howell-Richardson and Mellor (1996). Drawing on speech-act theory, they explained that transcripts should be viewed with the following question in mind: What is the purpose of a particular utterance? A change in purpose sets the parameters for the unit. MacDonald (1998) provides a slightly expanded set of guidelines for identifying a speech segment as the unit of analysis. She observes the following:

These authors also evaded some of the difficulties that Henri’s scheme presents by sticking to manifest content such as the linguistic properties of the posting and the audience to whom it was directed. Coding a complex, latent construct such as “in-depth processing” with a volatile unit such as Henri’s “meaning unit” creates large opportunity for subjective ratings and low reliability.

Our discussion of meaning units thus far illustrates that choosing the unit of analysis for a content analysis of a transcript from an online activity is not an easy task. Our advice is to try coding using a number of units, checking for ease of identification of the unit, ease of classification of the content, and finally, the reliability of both processes by multiple coders. The value of being systematic may also guide the selection of the unit of analysis, since some variables lend themselves to particular units. For example, if the study is looking at the level of argumentation shown by students, the whole message will most logically reveal this complex variable, whereas a study of frequency of postings may make the time stamp on each message the most logical unit of analysis.