Coding Verbal Data – What to Optimize

Klaus Krippendorff

I have written a second edition of my content analysis book, published in 2004 and a content analysis reader has just come out where you will find much of what I am saying today. There you can also find information on computers for text analysis and a lot of other ideas. Let me approach the problem of analyzing open-ended interviews from the perspective of content analysis.

Content analysts usually start from text that is generated in the world outside and not for the purpose of being analyzed. Text has uses in everyday life, whether it is a political speech, a poem, or a personal diary. Texts have meanings for their writers and for their readers in a context rarely accessible to the analyst.

Content analysis usually has a particular objective, not necessarily shared with ordinary readers of the analyzed texts. So the world of content analysis is far simpler than the world of all of the readers of the analyzed texts. Content analysts assume the right to eliminate meanings that are of no interest to their research project. I think the critical part of content analysis is first of all to start with some kind of research question. What does one want to find out? Second, there has to be some idea about the relationship between the text and the research question. The available theories that connect texts and research questions are not always well established but may well be compelling. Yet, such theories, even if they are mere assumptions, are crucial for the kind of answer one may extract from the text about the research question. Third come methodological issues of how a content analysis is conducted. This pertains to procedures that are more formalized than any unstructured reading of text; usually involve volumes of texts, far larger than can be processed by a single analyst. Relative to the research question under consideration, the texts to be analyzed are typically ambiguous, not that rich in relevant information, convoluted, and finding passages from which the answer to the research question may be inferred involves processes of selection and justified inferences, which can be difficult.

Content analysis is not a mechanical procedure that could be programmed and applied, like factor analysis or analysis of variance. It must be taught and learned and requires a major design effort on the part of analysts. The most important input to a content analysis concerns the relationship between the text and the research question. Ideally, this relationship has to be formulated computably. For coding, this means defining categories and scales of measurement that are analytically processable by what I call an analytical construct. For example, when one looks at or tabulates the pair wise co-occurrence of words, this in itself means nothing unless you have a theory of how co-occurrence relates to the association structure of the authors of the text or reader exposed to that text, etc. So this kind of knowledge of the relationship between the text and the research question has to be formulated so that it can become part of the procedure of a content analysis.

An important part of content analysis is coding raw text into analyzable and inferentially productive categories or scales. In fact, coding is the bread and butter, if you will, of content analysis. When I started writing about content analysis, I remember some researchers equating content analysis with the coding of words and larger verbal expressions into mutually exclusive categories. From my experiences, this is just too primitive a definition. Coding certainly is part of content analysis, but the defining feature of content analysis is the drawing of inferences from coded texts. When you ask how the inferences from texts to research questions are justified, content analysis becomes methodologically interesting.

I mentioned these two books. You can see them later if you would like. Based on the analytical framework I proposed in these books, I want to develop here a similar framework for survey research, particularly with open-ended questions. We start with people who live in their own worlds, some of them are surveyed and some of them are not. They live in a world of explanations, talk, arguments, disagreements, and negotiations, whatever. They are also exposed to the mass media and talk among themselves partly about what they learned from the mass media but always relative to their

everyday lives. This is my starting point. Then we have a survey researcher who lives in their own world, overlapping perhaps but essentially outside the world of the interviewees – both by being scientifically trained and in fact not interacting with the interviewees while analyzing their answers to interviewer questions. The relationship between survey researchers and the sources of the texts they end up analyzing differs quite radically from how content analysts relate to their texts. A content analysis starts from given text and ends with inferences from them. Survey researchers generate their text purposefully, by asking questions of interest to their clients or the hypotheses they explore. Answers are generated by a script, which is standardized and formalized to generate texts in the form of question-answer pairs that are relevant to the current research question. The texts for content analyses usually predates interest in them. The texts that survey researchers analyze are generated purposefully. Survey researchers have the advantage over content analysts in that their data are more closely allied with their research questions, but they also are potentially biased by the questions asked. Surveys are obtrusive; affect the interviewees by the very act of asking questions, content analysis is unobtrusive.

From the perspective of content analysis, most survey researchers do not conceptualize their analytical task according to what I just depicted here. They take for granted that the question/answer pairs can be analyzed as data, tabulated and characterized in terms of frequencies, and that is it. The point is that interviewing intrudes in the lives of interviewees, asking questions of the kind that are not normally asked, questions for which the interviewee has no readymade answers but feels compelled to give one, and questions that interviewees may not want to answer in the public nature of an interview. Interviewing is somewhat artificial. And there need to be analytical criteria for determining what the questions and the answers mean for the respondents. This enters the problematic of coding for what. The statistics that survey researchers report is not of what interviewers think, not what they said, but what they said, as filtered, distorted, and generalized by the categories adopted by the analyst. In content analysis, one expects that the inferences made from text are justified by explicating, as detailed as possible, the analytical constructs relating properties of text to the inferences made from them. In survey research, it seems that the inferences made from answers to open-ended questions are not recognized as inferences but shrunk into the categories used in coding, leaving the necessary inferences to the coders. So, survey researchers are accustomed to statistically analyze the coding categories without much attention paid to the inferences that reading and coding entails. The interview data in the CD we were provided, I have to say, are extremely complex, difficult to summarize and understand, much less codable in terms of simple categories. It turns out that, unlike what I had expected, open-ended interview data are far more difficult to analyze than the texts in content analysis, which usually provide larger contexts for disambiguation. In interview data, the questions provide the context for reading the answers, but this is a very small context and signals considerable difficulties in devising suitable codes.

So, now we have these codes, what is next? Well, I think first of all one ought to think of a theory of public opinion formation as guide to the analysis of survey data. I understand that some of your projects are far more focused on issues of voting. Voting is a very individualistic thing. That means you go into the voting booth and you cast your vote. The processes that lead to a vote certainly are more public in the sense that people form their opinion in talk with others, similar minded people or political opponents, parents, friends, or acquaintances at work, whatever. Add to that the agenda setting capabilities of the mass media. There is much going on prior to an interview and prior to voting. The opinions asserted are not as individualist as a statistic presents them.

I think a theory of public opinion formation ought to inform the inferences that one makes from these codes. Thinking of inferences as inductive, as a matter of counting and aggregating a sample of possible question/answer pairs treats them as context-fee tokens, void of a place in processes of public opinion formation, without the needed inferences. Maybe inductive inferences are satisfactory when surveys are highly structured but entirely inadequate when accounting for open-ended answers to questions that are intended to manifest public opinion. Now when it comes to predictions you can say, well, you can bypass all of this and simply see how it works. The point though to me, even the aggregation of individual opinion ought to be motivated by some or at least be related to some

knowledge of what is going on in the world, what people talk about outside the interviewing situation, and what kind of abductive inferences could possibly be made from the recorded answers.

The next issue is the role of a research question. Ethnographers pride themselves to approach their subjects with a minimum of pre-knowledge, without an agenda other than to understand. I guess survey researchers do start a research design with a research question in mind, whether it is the one that a client posed or derived from a theory that researchers are pursuing. When codes are defined, one needs to be clear about what one wants to find out, know a bit about the linguistic habits of the sources that could provide relevant data, and the inferences required to answer the research question. This is relatively unproblematic. Then there actually is a different kind of question, which is the client's interest in the information that a survey is to provide. Clients pay for the ability of making use of survey findings. Survey research is rarely done for scientific purposes alone, or for the person that hired the researcher. There are whole institutions behind the person that survey researchers are in contact with, a political party that wants to see its candidate elected or a drug company that is interested in learning why their produce lags behind its competitors. These are very different kinds of questions that may not be answered by a frequency account of answers to questions. To be relevant to stakeholder of survey research one needs to be able to translate the answers of a research question into the world of those interested in the findings. The ability to accomplish this translation has very much to do with the codes to be designed to generate findings that are helpful to their stakeholders. If one can afford not to care about the conceptual system of clients, it may well be satisfactory to limit one's concern to the statistical significance of the findings, and to own research question. However, if one has to deal with clients and particularly if one's reputation is at stake visa vie other stakeholders, which I heard being mentioned as a problem, then one has to go beyond scientific research questions and design codes that are relevant to the clients' interest.

Now I want to look at the criteria that are applicable to evaluate the quality of the process of analyzing open-ended survey questions. I'm coming to the logic of the criteria from content analysis, which becomes increasingly part of this kind of analysis. The most obvious criterion we have already talked about is reliability. In the talks we heard so far, reliability seems to be used rather loosely. I prefer a more technical definition as a measure of replicability of a process, particularly that of coding.

A code is always part of instructions to coders that specify the coding process enacted by coders. When we talk about replicability this is not a question of taking two coders, five coders or 100 coders, it is the question of what is it that you want to replicate? Now to replicate something means to have something that you replicate – from my point of view these are the coding instructions. If you have very specific coding instructions you can give it to one coder, you can send it to another agency and ask them to use their coders. If they come to the same results, based solely on the instructions provided, then you can say these instructions are replicable.

Why is replicability so important? If the coding instructions are not replicable, then one cannot claim that research results represent something everyone would see as spelled out in these instructions. Lack of reliability means that one does not know how the units coded – the answers to known questions – end up as described. Under these conditions, one cannot assure anyone else that one knows what one is talking about. Coding instructions that are replicable by different coders working in a variety of circumstances enable researchers to claim that a reality was described and the idiosyncrasies of coders did not interfere with their description. Replicability does not measure whether categorization or scaling was valid, only that coders did it alike, and that those who have the coding instruction can reconstruct in their mind what coders may have seen or read.

By the way, a computer is perfectly reliable but reliability does not mean computer coding would make sense to readers of a coded text, or could be approximated by human coders. I have many years of experiences in reliability assessments and a strong opinion about which coefficients are appropriate and which are misleading, but this might be discussed later.

There is another criterion, the public significance of what is being found. This morning, several examples were given of questions for which interviewees had no knowledge of how to answer them.

Public significance of interview data has to do with whether the question/answer pairs that survey researchers are generating, coding, and analyzing have anything to do with what happened in reality. For example, everyone was laughing hearing from one of the presenters that some women didn't know how many children they have. I was surprised about this response. Could this not have been the problem of an insensitive questionnaire design? If one is Catholic, then any miscarriage is having a baby. If a baby is given up for adoption, or it died young, it makes good sense not to know what is being asked of a mother. If a child is estranged, got lost in another country, or is for any reason not acknowledged as a child, than this is not a child. Questions for which there are no answers signal their insignificance in the interviewees' life. Asking publicly relevant questions is not a problem of the interviewee but of the questionnaire design.

There are issues of so called embarrassing questions, which transgress the distinction between public and private discourse. What is kept private is unlike to have public significance. Asking embarrassing questions are not only yielding evasive answers, they can also affect the relationship between interviewer and interviewee. Public significance is also impaired by the well studied problem of matching the personality type of the interviewer with the questions asked. Women are more likely talking to other women about intimate problems. A male student interviewer asking women such questions is not likely to elicit valid answers. There is also the distortion due to the inequality of power in the act of interviewing. I wrote a paper on how unlike an interview is from the kind of conversations that people engage in when sorting out their opinions with others. A standard interview can only yield data consistent with this rather unusual form of communication. When one asks "are you willing to grant me an interview that takes only 15 minutes of you time?" the interviewer grants the interviewer agency. But after an affirmative answer, the unequal roles are fixed and what transpires is no longer compatible with what would happen in public.

Interferences with the public significance of the answers that one may encounter need to be anticipated when designing the coding instructions or rectified in subsequent analysis. While the best solution to this problem may well be to find situations that are naturally occurring and relatively free of such constraints – content analysis has the advantage of being an unobtrusive technique – open-ended interview data may have the additional burden of structuring the coding process whenever this ideal does not pertain.

There is another criterion that I think we frequently confuse or ignore – semantic validity. A code can be very reliable but wrong. Semantic validity measures the extent to which the results of coding correspond to how ordinary readers, or better still the respondents themselves would categorize what they say. One of the ways we assess the semantic validity in content analysis is to employ experts. Here, one may rely on experts who are familiar with the discourse of the respondents, how they conceptualize what they are asked, and the codes that coders have assigned to what the respondents say. Very often, quantitative research uses categories that are abstract whereas responses are very concrete, practical, and detailed. The questing that semantic validity answers is how far the categorizations of a study corresponds to the categorizations by the authors and readers of a text, in the case of content analysis, and by the interviewees, in the case of coding open-ended answers. In content analysis we have also showed the analytical codes to authors and ask if they represented what they wrote. Very often we found that the code had nothing to do with the writers' conceptions, raising the question of what we were analyzing.

Now, for content analysts it is quite meaningful to say that they are not interested in everything said. Meanings are complex and every text can be read in numerous ways of which we might be interested in only one kind of reading, usually of theoretical or practical importance. For example one might count certain speech disturbances in therapeutic interviews to infer a patient's pathology. The pathology may not have currency in the patient's vocabulary and for therapists all the stories clients tell may be noise, except for the speech disturbances that tells them what they need to hear. However, when the target of analysis is tied to interviewees' conceptions and their public behavior, as I believe all survey research is, coding should at least not violate the conceptions of the source. Now in the case of content analysis it is rarely possible to go back to the author of a text, but in survey research going back to a subsample of interviewees and ask them what they meant may be a little

easier. Semantic validity is particularly important for evaluating computer text analysis and this applies also to the use of computers for processing open-ended interview data. Computers have no reliability problem, but most computer analysis of text are semantically shallow if not invalid by criteria derived from ordinary readers of text or when compared to human coders, who are infinitely more sophisticated in picking up nuances.

Then comes the criterion of construct validity. It is quite possible that the theory of public opinion formation that a researcher has adopted for making inferences from interview data is just meaningless. For example, aggregating interview data by counting and reporting the frequencies of mentions may have nothing to do with how people prioritize their issues. The analytical constructs that justify inferring the answers to a research question from the raw data need to model the reality outside an analysis in order to be valid. Here content analysis provides survey research an important criterion for whenever data are not merely counted but codes are used to allow inferences to be made that have something to do with how people speak, read, conceptualize their world and this includes voting, liking and disliking political candidates, or are concerned with certain issues. In content analysis, there are theories available relating particular textual attributes to questions of bias, attention, authorship, and authenticity whose analytical use can be tested in terms of construct validity.

Another criterion that I think is very important to consider is relevance. Research questions are, as the word suggests, what a researcher hopes to answer. Scientific hypothesis are tested to satisfy truth conditions, which may not be relevant to those eager to obtain information to justify their actions, whether to back a particular candidate for political office, design an advertising campaign, or find a niche for a new drug. We have been discussing your interview question “what are the most important problems facing the United States?” One has to ask who would be interested in its answers and what do frequencies in different categories of responses mean and to whom? Surely, politician might be interested to learn what they should do to appeal to large audiences. However, that politician might be interested only in what he or she can actually affect, what distinguishes his or her campaign from the competitor. A rare and statistically insignificant mention may give rise to a good idea. The point is that the users of survey research judge survey research findings not entirely in terms of statistical significance but in terms of what is relevant to them. To be of service, survey research has to be relevant to their clients and relevance is not necessarily measured in statistical terms.

In sum, all of these are the criteria for coding. I think all coding instructions should be designed with semantic validity in mind, that is, the meanings of the original interview responses should be recognizable in the coding categories. But preserving the meanings of the original is not the only task for the designers of coding instructions.

Public significance refers to the degree to which what is coded has the same currency as in public life. Semantic validity and public significance do not enter reliability considerations.

Reliability test are simple to perform and there are a couple of reliability measures in terms of which comparable standards have been developed. Calculating the agreement among any two coders, often taken from the same research team, is not enough. Their choice must demonstrate that the coding instructions are replicable in a variety of situations for if they are not so replicable, they should not be applied as the data they generate can then not be trusted.

If research goes beyond merely reporting frequencies of observations, occurrences or mentions, which are generally rather uninformative, and infers the answers to given research questions, as content analysts are obligated to do, the validity of the constructs that justify these inferences are important. While the use of structured question did not have to bother with analytical constructs, but the analysis of answers to open-ended question brings this criterion to the fore. Construct validity is not to be confused with the validity of the research results, which is very difficult to establish.

Instead, I suggested relevance as the criterion of the usefulness of the research results to their users. Relevance pays the bills, but this is not to say that the other criteria are less important. Without them,

research results may be deceptive and misleading. Obviously these criteria form a system whose focus is on what precedes and succeeds coding processes.

I would say the coding criteria interact in very particular ways. We know that human coders are reliable when the task is very simple. For example, if one codes television characters, identification of the good and bad characters are extremely reliable. But as soon as researchers ask more complicated questions unreliability becomes a major problem. Traditionally, content analysts have overcome this unreliability by increasing the specificity of the coding instructions, but this adds time to the coding effort. Generally as reliability improves, semantic validity goes down, and preventing the latter from happening increases the costs of coding. This is an almost iron clad interdependency. Optimizing reliability, semantic validity, and efficiency always is a problem of finding a delicate balance between competing criteria that has to be weight by the criterion of relevance.

Computers are perfectly reliable, of course. But without attempting to offend anybody who will later present their computer application, I am getting impatient with advocates of computer text analyses who justify their approach in terms of being perfectly reliable. True, the unreliability of human coders in content analysis and of the answers to open-ended questions is a major problem, especially when the coding task is semantically complex, is done over some period of time, involves large volumes of text, and a large commitment of resources, which are at stake when reliability is insufficient. Now, computer science has offered ways to process great amounts of text in very short time without appreciating that humans read text quite differently from what computers can be programmed to do. For computer text analysis the bottle neck criterion is semantic validity. Unfortunately, most software developers do not care to validate their often wild claims of being able to reveal meanings, extract the underlying concepts, mine information, analyze contents of texts, or reveal hidden associations. Such claims, full of colorful metaphors, must be treated with great suspicion, mainly because evidence of the semantic validity of these claims are rarely ever provided and the simplicity of what they do with text is hidden in elaborate statistics that are difficult to decipher.

This morning, transparency was mentioned as an important criterion for coding and I agree when this means making the coding instructions available with the survey findings. In principle, the computer code of text analysis software makes that software perfectly transparent. In practice, however, it is close to impossible to read that code and understand what it does with text. The appropriate evaluative criterion for computer text analysis is semantic validity. Computer text analysis software processes massive amounts of text with very shallow semantics or it takes an awful amount of time, such as by parsing sentences and producing all possible interpretations of a sentence without producing answers to the kind of research questions that analysts of media content and answers to open-ended questions tend to pursue. Thus, there is another optimizing issue between the high semantic validity of small amount of text that human readers/coders process naturally, and the low semantic validity of large volumes of text that computer content analysis software provides. In my experience, the optimal balance is to combine computer aided content analysis with human coders. Some clerical work on text is best performed by computers; the reading part is best performed by human coders. This is accomplished by qualitative text analysis software such as Atlas-ti and NVivo. If you are serious about including multiple answers to open-ended question, this is the way to develop computer aids.

Before I received your CD with examples of responses to open-ended questions, I was convinced of the benefits of analyzing answers to open-ended questions over the coding of mass mediated material. First of all content analysis is an unobtrusive technique and its texts tend to contain much irrelevant matter that coders need to read and eliminate. Open-ended interviewing is an obtrusive technique that intervenes in the life of interviewees and has the potential to influence the answers in unintended ways, but generate data more pointedly. Survey researchers have more control over the texts they generate, which means that their data naturally are more relevant to a research question and may have higher public significance. This can be a great advantage.

Second, in content analysis, units of analysis – words, expressions, attitudes – occur in contexts and to judge the meaning of these units, contexts usually are large. The textual units in content analysis

may also appear on several inclusive levels – from words to sentences to paragraphs to books to whatever. Answers to open-ended questions are units whose meanings depend almost exclusively on the questions asked, which constitute well defined contexts for coding the answers and no embedded hierarchies.

Third, the research question in the content analysis is often unique. Two content analyses are rarely alike, with a few exceptions. In the kind of survey research you are pursuing, the questions asked seem to be recurring; for example, you may be asking the same kinds of question over many years to see trends. The kind of information that candidates for political office are interested in, for example, "why would you vote for X?" and "why would you not vote for X?" can be asked for different candidates and over and over again. The analytical constructs in content analysis are often designed ad-hoc and to deal with very specific theoretical problems. In coding answers to open-ended question you have the opportunity to develop solid analytical constructs over time as the relationships between questions and answers are simple ones and the categories that you want to analyze and the inferences you want to make are recurrent. While in content analysis there are few standardized categories, you have the chance of developing and testing standardized coding categories and techniques for drawing valid inferences from them.

Fourth and this may be more of a challenge. In content analysis, coding tends to be simple: one unit is coded into exactly one category per variable of interest. When interviewees are asked to give multiple answers to single survey questions, such as "what are your reasons for voting (or not voting) for a candidate for political office, the coding variables consist not of mutually exclusive categories, but of overlapping sets of categories for these answers. Whereas content analysts are often criticized for forcing texts with multiple readings into simple variables, as analysts of multiple answers to questions, you cannot ignore criticism and will have to code vectors. A vector has several dimensions and represents the co-occurrence of distinct entities. The analysis of vectors of unequal dimensionality is not easy, but there are text analyses that have solved this problem, for example, Michael Best's of MIT. I have developed a way to calculate reliabilities for what I call multi-valued data. Here, survey researchers will have to develop multi-valued codes and computational techniques that may in turn contribute to content analysis.

After I examined the sample interview data on the CD you provided us, I was somewhat less enthusiastic about coding these interview data. I want to make a few comments on these and this is all the time I have left.

Multiple answers to one question are not so problematic when they are logical independent and mutually exclusive. Then one can treat them as vectors, as I suggested. But I found the recorded answers to be of rather different kinds. Some are single concepts, some are expressed as noun phrases, some are long arguments, some make comparisons between candidates, some offer predictions about the future of the candidate, and others bypass the question, for example, by declaring they wouldn't vote for the candidate anyhow. These answers are of various logical types and on different logical levels, which makes coding them extraordinarily difficult. The solution probably lies in developing better questions.

(After a clarification, I learned that what I had thought to be a coder's notation was in fact the answer to a follow-up question about which of the named reasons would be the most important one.) Forcing multiple answers into a single code would not only violate the semantic validity of the process, but it would most likely be totally unreliable.

Although I initially thought that the questions would provide an adequate context for disambiguating the answers they elicit; now this seems doubtful. It became evident when eliciting the likes and dislikes of a candidate. For example, someone responded simply "everything" others responded ambiguously, for example, when they asserted they liked the candidate's position on abortion. The latter is a common phrase regarding an issue, but it does not reveal whether the interviewee is pro-choice or pro-life. A human coder might decide on this from how the interviewee answered previous questions. In this case, the interviewee talked about Bill Clinton, but to code the answer as pro-choice

or pro-life, the coder needs to know what Clinton had asserted as his position on this controversy or a previous answer that dealt with how the interviewee felt about abortion. But if coders have some difficulties coding such answers, how could computers be programmed to disambiguate such responses? Some the answers to open-ended questions I read were attitudes, beliefs, predictions, but also quotations.

Quoting or rephrasing what someone else said is really not a respondents' opinion. We discussed this morning that people quote others for good reasons. True. But to know why respondents quote or rephrase others, one has to glean much from the context of what they say, for example whether this is used as reason for liking a candidate or not voting for him or her. Before I examined your CD things seemed far easier. Now the contexts of answers seem to become more and more important. This fact renders the processing of answers to open-ended questions more like the more difficult aspects of content analysis. Survey researchers would benefit from studying and adopting relevant parts of its methodology. Certainly the hope of using computers, especially software that is not context sensitive becomes increasingly remote.

Besides quotes and rephrases, I found expressions of wishes, hopes and fears as well as agreements with candidates, or positions candidates take, which are altogether rather unlike phenomena that cannot easily be accommodated by simple coding categories.

An issue that I have not found discussed adequately, and was surprised to come upon it in the CD, is that questions concerning individual perceptions, assessments, and preference are often answered not just be offering individual opinions, but also by assessing public sentiments, as well as reproducing sound bites recognizably coming from media coverage – all mixed up.

In my opinion, the reason for using open-ended questions stems of course from the researchers' uncertainty about the categories and dimensions in which respondents are thinking and acting. Open-endedness solves the problem of interviewing under conditions of ignorance as to what is out there. The gain in semantic validity is counteracted by making the analysis of these responses extraordinarily difficult, which is why we are here. You asked us what questions you should be asking. To make suggestions of this kind, I would need to know what information would be valuable to you, what should open-ended question elicit? This goes back to the importance of the research question, discussed earlier.

Let me suggest, if you want to obtain information about how respondents vote, voting being an individual act, individual opinions are perhaps more important than public ones. But if you are seeking information about public opinion, what people say to each other in public, for example, in response to a speech or political event, or concerning a local or national problem, or about ethnic prejudices, or who influences whom, individual opinions are less important than what interviewees hear being talked about and how they position themselves in these conversations. In other words, if one wants to know what is going on in public life, I would recommend asking questions of how other people feel or who influenced them, for example, who and how many people in your neighborhood are concerned with the issue you just mentioned, or who among your friends thinks as you do on that issue, what do the democrats you know say about this, or the republicans you have talked with. Opinions are public when talked about with others, when individuals see themselves as in the minority or majority, have hopes to have an impact or are discouraged, willing to engage in political actions or merely hearing others speak. Answers of this kind can be obtained by questions that do not call for individual opinions, opinions that the interviewee may not even have, but that interviewee conceptualizes and reports the attitudes of others, friends, neighbors, coworkers, and various classes of people, the rich and the poor, conservatives and liberals, women and men, etc.

The criterion of public significance, mentioned earlier, applies to questions as well. I think many of the questions that survey researchers are asking are stereotypically individualist, yet most surveys are conducted to inform the public, for example through the mass media, or to inform political actors and institutions that have an interest in public discourse.

The kind of responses that seem to merely reproduce sound bites and stories heard in the mass media, suggest questions concerning interviewees' media habits, preferred news sources, newspapers read, radio stations and television networks attended to, newscasters they like, program they watch – and with whom they watch or discuss what they hear. When combining interview data on media preferences with data from content analyses of diverse media, much may be gained about public opinion. As David Fan demonstrated content analysis of media can substitute for interviews.

In conclusion, I think answers to well crafted open-ended survey questions are likely to contain more information relevant to public opinion research than structured ones, but as I have tried to show, their analysis faces major challenges. I think content analysis can provide much methodological assistance for coding such data. I would be weary of unfounded claims that computer text analyses solve these problems and warn against prioritizing one of the criteria for coding data at the expense of others, for example reliability over semantic validity and relevance to users. Optimizing conflicting quality criteria for coding verbal records may be difficult, but when one recognizes these difficulties, one is on a path to solving these methodological problems. The very fact of this conference suggests being on this path.

This is all the time I had to express my hopes and alert you to what needs to be done.