

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/283785303>

# Analyzing Open-ended Questions by Means of Text Analysis Procedures

Article in *Bulletin de méthodologie sociologique: BMS* · October 2015

DOI: 10.1177/0759106315597389

---

CITATIONS

46

READS

6,050

1 author:



**Roel Popping**

University of Groningen

69 PUBLICATIONS 1,020 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



New developments in quantitative text analysis [View project](#)



Nominal scale agreement [View project](#)

## **Analyzing open-ended questions by means of text analysis procedures**

Roel Popping

University of Groningen

Department of Sociology

NL 9712 TG 31

Email: r.popping@rug.nl

Appeared in *Bulletin de Méthodologie Sociologique*, 2015, 128: 23-39, doi:  
10.1177/0759106315597389.

### **Abstract**

Assume one has open-ended questions in a survey and seriously wants to analyze the answers to these questions. Now text analysis might be applied. This text discusses a number of choices to be made when a thematic text analysis is to be applied. It starts with requirements to be posed to the open-ended questions themselves and sketches choices in the development of a code system. Here the coding comes immediately into view. Coding can be performed from an instrumental or a representational perspective. In the first the coding is performed from the point of view of the investigator, it can be performed in a run of a computer program. In the second the point of view of the respondent is acknowledged. Now the computer can be used as a management tool, but the coding itself must be performed by a human coder. The choice for one of these methods depends on what the investigator is looking for and has consequences for the way how to proceed. When the representational way of coding is applied also questions about intercoder reliability must be posed. The codes to be used are decided upon before or during the coding process.

**Keywords:** open-ended questions, text analysis, content analysis

**Running title:** open-ended questions

## **1. Introduction**

An open-ended question (sometimes called open question) in a survey or public opinion poll is a question in which possible answers are not suggested, and the respondent answers in his or her own words. These questions allow a spontaneous response. Originally these questions were to be used in a pilot study. Closed questions, characterized by assisted response, constructed in an a priori way possibly will fail to provide an appropriate set of alternatives meaningful in substance or wording to respondents. In a pilot study the investigator has the opportunity to fine-tune the question wording and also there will be clarity regarding the answering possibilities. The investigator has the possibility to overcome the argument that it is impossible to construct codes in advance for a question. Next these could be entered in a survey as a closed question (Lazarsfeld, 1944). Nevertheless the open-ended question has become part of many surveys. There are several arguments why this is good, but also why this is not good at all. The debate has been well summarized by Converse: “[t]he open / closed debate was shaped in good part by institutional needs and capacities, and by ideologies about research, remaining largely untouched by research” (Converse, 1984: 279). The distinction between the type of response in an open-ended question) and a closed question is a way to measure the issues saliency (Geer, 1991). It is also a way to analyze the normative burden on some responses: normative constraints are better revealed with assisted techniques than with spontaneous techniques. Today the debate is still going on. Fribourg and Rosenvinge (2013) show that open-ended questions show more in-depth information than closed questions in the field of mental health, but there are also more missing data. Hruschka et al (2004) made a similar comparison with respect to data on HIV prevention. In this study the codes for the open-ended question were obtained by the coders. These codes overlapped only moderately with the codes from the closed question. Scholz and Züll (2012) looked at who is not answering the open-ended question. But also new issues came up. As an example: Smyth, Dillman, Christian and McBride (2009) discuss how the open-ended question should be presented on the screen in web-surveys. We do not go into these debates.

When the open-ended questions are entered in a general population survey questionnaire the answers should be analyzed. So far this aspect did get little

attention.<sup>1</sup> This contribution focuses on the analysis of the answers to open-ended questions by using computer-assisted text analysis. Text analysis is part of a broader class of methodologies called “content analysis”— a generic term for statistical analyses of qualitative data (e.g., words, gestures, art forms, etc.).

The text actually starts with distinguishing a number of types of open-ended questions as are used in surveys and a discussion on pros and cons. Next the opening to computer-assisted text analysis is made. A number of possibilities is shown, these all contain some restrictions. The codes to be used must be known before coding starts or must be developed during the coding process. The text analysis can be based on human judgments. Now intercoder agreement comes up, as one needs a reliability check. This all should give insight in how to use open-ended questions in a research project so that the data obtained are really used. It also gives some insight in whether open-ended questions should be used in surveys.

## **2. Open-ended questions**

Open-ended questions are usually seen as opposed to closed questions. The closed question might be a knowledge question which requests basic, limited, factual information, having a correct answer that is incorporated in a list of alternatives that is presented to the respondent. It might also refer to an attitude question for which there is no correct answer. The open-ended question is supposed to catch information that is not seized by a closed question. The answers to open-ended questions are statements. These are linguistic interactions, often framed by a sequence of questions (open and closed). This point has a significant impact on the shaping of data, analysis and software use strategies. The open-ended questions must be closed in the analysis process, this is the informational paradigm. To do this some action is required.

Three types of open-ended questions are distinguished:

1. The technically open-ended question;
2. The apparent open-ended question;
3. The really open-ended question.

The *technically open* open-ended question is found when the user has to fill in some digits or letters having a specific meaning. This type is among others found when the year of birth is asked for. Here four or the last two digits have to be filled in.

In the situation of an *apparent open* open-ended question the respondent is asked to write down the answer that comes from a list, a list that is too long to be included in

the questionnaire. An example is: “Which newspaper do you read at home?” Now the respondent is supposed to fill in one (or more) newspapers that are available to the investigator. This type of question is used for all questions that have answers that fit in a list: religious affiliation, ethnic background, language spoken at home, magazines read, and so on. It is a self-completion format.

When investigators talk about *really open-ended* questions they usually have questions in mind that put forward answers by the respondents. Usually a specification or an argumentation is asked for. Examples of such questions from election studies are (Van der Eijk et al, 1988): “Why did you vote for this party” (p. 260), “Are there things you appreciate very much in this party or for which you have sympathy” (p. 162) or “Can you tell how you understand ‘left’ in politics” (p. 167). Often the main question is split into two parts. First the closed question: “Do you like ...?” having the codes “yes,” “no,” “don’t know,” followed by the open question “Why?” A great diversity in answers is possible and these might provide alternative explanations to those that closed questions can capture. It is possible that an answer can only be understood in combination with the answer to the accompanying closed question. More general, if no context information is available (coder) understanding of the answer might be problematic.

Another possible problem, that might impede understanding, is that the answer is a transcription by the interviewer with gaps in terms of function words: particles, pronouns, adverbs disappear. The coder also has to keep this possibility in mind. If it was the respondent him or herself who wrote the answer in for example an on-line questionnaire he or she might have skipped several words.

Each time the respondent might answer with a long and detailed comment in which his or her situation or argument is explained. On the other hand it is also allowed that only one short answer is given. The answer to the “why vote for this party” question might have been: “I always do so, it is a tradition.” This is a quite common answer and does not contain any information regarding political or governmental preferences. In the Dutch elections study of 2006 (NKO2006) this was mentioned in 167 out of the 2173 valid answers. Sometimes the question consists of several sub questions or has the form of an incomplete utterance to be completed by the respondent. Usually an open-ended question asks for who, what, when, where, why.

The answers to open-ended questions will usually be descriptive. Respondents will present listings, will demonstrate knowledge, or will present an explanation or a

motivation. These will be based on facts, but can also be based on attitudes or evaluations. Sometimes the arguments will be mentioned explicitly; sometimes one has to read them between the lines, which can be quite subjective. Note, these answers will be different from answers in depth interviews; such answers usually contain more details and may describe processes.

Technically and apparent open-ended questions are alternatives for closed questions, they seem better tools to get an answer to a specific question. Really open-ended questions might be used for theory construction. Open-ended questions can be helpful when new theory is to be developed. The questions are used in a pilot and serve to indicate how the categories for a closed question should be formulated. In case the respondents have to give longer answers one might be able to look for co-occurrences of themes that are foreseen, i.e., there might be unexpected relations between variables.

Open-ended questions must be specific to provide meaningful, interpretable data; therefore the formulation of the question is relevant. The formulation is to be neutral, but also inviting for an answer. It should be as short as possible and should contain the correct questioning word. This is especially true when there is no interviewer who can help in understanding the question. In a self-completion interview it is not possible to ask for more detail or to guide the respondent into another direction. The question can at most include a direction (which is not a manipulation). Online surveys even can integrate some kind of probe questions which are able to "react" depending on the answer entered. These questions cannot substitute a human interviewer. The answer must be that good that the investigator can use it. This might be different for specific types of research. A non-response study might use different answers than a substantive study. Compare the differences in the following examples. The first question leaves the interpretation up to the respondents, which could differ from the question's intended meaning: "What are important problems?" The second question provides specific information on what is being gathered: "Mention the most important political problems in your country today?" (Note both questions are actually apparent open-ended questions.) The answer written down by the respondent might be formulated in perfect sentences, but it is also possible it consists of catchwords or cryptic words and contains a lot of (grammatical) spelling mistakes. In the words of Kammeyer and Roth (1971: 61): "Responses to open-ended questions are usually less

than completely clear; they often contain ambiguous words and phrases; and they are frequently ungrammatical and poorly worded.” Respondents in a survey are confronted with a cognitive process that includes three steps. They must interpret the question and deduce its intent. Next, they must retrieve relevant information and integrate that information into a single judgment. Finally, they must translate the judgment into a response by selecting one of the alternatives offered (Krosnick, 1999: 546-547). This will also work for open-ended questions. With respect to such questions the last step will be a difficult one. Geer (1988: 370) however reported that most people are able to communicate an adequate answer.

During the analysis problems that rise due to the wording of the answer here must be taken into account. If there is a human interviewer who is supposed to write down the answers, this person can prevent some problems. This interviewer must among others be trained in asking for a clear answer. Afterwards only very few problems in understanding the answer can be solved.

From the open-ended question the investigator might have additional information that he or she could not get by using closed questions, but this information still needs to be classified in some way. For this text analysis is a proper tool.

A special case of the really open-ended question is the following. A set of codes belonging to a question should be exhaustive. In order to realize this often the code “other” is used as the last one. As this code is not informative the respondent gets the opportunity to specify what is in the actual situation meant by “other”. The respondent does not have the bad feeling that his or her opinion cannot be expressed. The investigator might get an answer that is after all very valuable. In practice however, the investigator hopes that the respondents will not use this code.

It is expressed that open-ended questions may generate richer and more spontaneous information in questionnaires administered by interviewers than in self-completion questionnaires. In such questionnaires responses are typically less detailed because the burden of recording the response falls on the respondent (McColl et al, 2001: 22). It is also possible one expects to get a more accurate answer than when a closed question is used.

### **3. Opening to text analysis**

Schuman and Presser (1981) already indicated that open-ended questions are not used that much because the coding analysis of the answers is so difficult compared to

closed questions. The issue has not received that much attention in the literature, while in the meantime there have been a lot of developments.

The input for the text analysis consists of as many pieces of texts as there are respondents in the study. It is possible that a piece of text is empty when the respondent did not give an answer. Attached to the piece of text is some identification mark. The output consists of a matrix having in the row the texts (which might even be split up) and in the columns the themes. The cells inform whether a theme is found in the textual data or not. One exception will be mentioned later on. The identification mark allows merging with the data found for the closed questions. When the themes are exclusive, and the investigator cares only about the occurrence, the theme variables can be reduced to one variable having the themes as separate codes.

A computer-program can count the occurrence of themes in the texts, here in the answers by the respondents.<sup>2</sup> In general these texts are indicated as units of analysis or sampling units. In text analysis studies a distinction is made between units of analysis and recording units. Recording units are “the specific segment of content that is characterized by placing it in a given category” (Holsti, 1969: 116). A unit of analysis can consist of several recording units. Therefore it might be necessary to bring the codes at the very end to the level of the unit of analysis. How this is done depends on the actual research situation. In text analysis studies often weighting is used. As an example, say there are two recording units within the unit of analysis, now each recording unit is weighted for 0.5. Together they sum to 1. When the text consists of the answer to an open-ended question the respondent is the unit of analysis.

There are several differences between a text analysis as usually intended and a text analysis on the answers to open-ended questions. Usually the data are unobtrusive; the investigator will collect a lot of irrelevant and unstructured text. Besides the context and the nature of the units vary widely. In the situation of the open-ended question, the data can be obtrusive; the texts are more relevant, and pre-structured by questions. Here, the questions are the context for the answers. The units are convenient to the question-answer pairs.

#### **4. What information does one get?**

The final coding of the answer by the respondent must reflect as good as possible what was said and intended by this respondent when the open question was answered. Each open-ended question has its own set of codes. Per question the answer is written



down by the respondent, this answer tells what the respondent has in mind. But the answer might also be written down by the interviewer, who even can give an own interpretation. Later on the answer as written down is coded by a coder (or computer program). This brings us to the semantic validity of the codings. A meaning category or code has semantic validity if there is consensus among persons familiar with the language and texts under study that the words or phrases judged into this category do reflect the category's meaning (Krippendorff, 2012: 323). Assurance of semantic validity requires more than a face validity check for similarities among words' meanings when taken out of context.

Construct validity seeks agreement between a theoretical concept and a specific measuring device or procedure. It refers to the degree to which inferences can legitimately be made from the operationalizations in the study to the theoretical constructs on which those operationalizations were based. It “is not only concerned with validating the measure, but also the theory underlying the measure” (Holsti, 1969: 148). So, the validity is on the results of an investigation and the theory in which the problems behind the investigation are represented.

## **5. Text analysis**

Before three types of computer assisted text analysis have been distinguished (Roberts & Popping, 1993). These are thematic text analysis (in which one looks for the occurrence or co-occurrence of themes), the semantic text analysis (where one uses the subject – verb – object [SVO] relations as found in clauses, the sentences or parts-of-sentences that - explicitly or implicitly - contain an inflected verb, an optional subject and/or object, plus all modifiers related to this verb, subject, and object), and the network text analysis (where the SVO relations are even combined into networks). Semantic and network text analysis are especially interesting when long answers are available, consisting of at least some (complete) sentences. These approaches are not discussed here.

When open-ended questions are used (especially in sociology, political science or communication studies), the investigator usually wants to look for the occurrence of themes within the answer, not for the co-occurrence of themes. In text-analysis studies this co-occurrence is problematic, because one has to question whether this really indicates a relation between the themes other than occurring in the same part of text. In

the situation of open-ended questions the investigator generally is not confronted with this problem.

An important question in text analysis is from whose point of view the data are to be analyzed. In the instrumental approach texts are interpreted according to the researcher's theory. The approach ignores the meanings that the texts' authors may have intended. When the representational perspective is applied, texts are used as a means to understand the author's meaning (Shapiro, 1997).

Today many investigators still follow the instrumental approach, as is shown by Shapiro. In these investigators' studies a computer program usually takes care of the coding. A "fixed dictionary" of thematic codes having a one-to-one correspondence with words and phrases in texts is used. Heinrich (1996) asserts that the fundamental advantage of computer-assisted text analysis is the intersubjectivity of the result. Personal evaluations by the coder have no effect on the results. He was referring to this instrumental approach, and it is questionable whether this is the best type.

Researchers using the representational approach must develop dictionaries that contain themes reflecting the perspectives of the texts' authors. Coders must use sympathetic "understanding" (or "Verstehen") to encode the texts according to the meanings their sources intended. At issue is no longer "how" to encode text (instrumental approach), but "whether" one chooses to apply one's own theory or one's sources' theories to the texts under analysis. Sometimes one also needs human coders when the instrumental approach is followed. This occurs in such situations where the size of a text is to be determined in terms of amount of space needed or where the place on the page in the newspaper is to be indicated.

The computer program generally takes care of the coding when the instrumental approach is followed. But when the representational view is followed, the computer is especially used as a management tool; the coding is actually performed manually. Textual analysis software also allows to study the language components of the responses and further to provide more consistent analysis strategies with an interpretivist paradigm.

### *5.1. Instrumental thematic text analysis applied on survey data*

In a computerized instrumental text analysis the investigator is looking for the occurrence of themes. Such themes are found in the texts by using so called search entries. The occurrence of a search entry denotes the occurrence of the corresponding

theme, no matter the context in which the search entry was used. It is possible that in a text or open answer several different search entries (and so themes) are used. In survey data when made available for investigators these are found back as “first argument mentioned”, “second argument mentioned”, and so on. The question for the most important national problem in the Dutch elections study 2006 even resulted in seven different arguments given by respondents.

Once a dictionary is available the coding by computer goes very fast, often ambiguity in the texts is ignored. In the situation of open-ended questions this concerns especially idiomatic ambiguity in the identification of themes. When computers were first enlisted in the encoding process, it was soon discovered not only that they afforded perfect intercoder (actually, inter-computer) agreement, but that they did so at a cost to the quality of the encoded data. In short, computers are dumb. Upon encountering a sentence such as, “You'll eat your words!”, they obediently classify the word, “eat,” as an occurrence of the theme “consume,” instead of as “regret” (as in the sentence, “You'll regret your words.”). Until the coder (or computer) has been trained to recognize such idiomatic uses of words, ambiguity will remain in the process of classifying these words under one or another theme. Ambiguity starts already with the fact that a word can be a homonym, can have different meanings. The word “spring” can refer to a season in the year, but also to the beginning of a river.

More problems might be found in texts. They may contain negations, pronouns or value judgments. There are no standard ways to solve these problems. The sentence “I am going” differs from “I am not going.” The one performing the coding (a human or a computer) must recognize this, but does this happen? Sometimes it is possible to replace the negation, now the sentence above is “I am staying.” In some cases however it might be that as a consequence of this rephrasing the meaning of the sentence changes.<sup>3</sup>

Pronouns like “he,” “she,” “his,” “hers” are not useful. They refer to a person and often even to a role performed by a person. Again, is it possible to replace the pronouns and here this doesn't affect what is said in the sentence. In general, for grasping judgments (“this is a much better method”, “he is a bad boy”) there is at this moment no other alternative than manual coding.

## *5.2. Representational thematic text analysis applied on survey data*

If an investigator wants to follow the representational way of coding usually human coders are used who select a text fragment and assign a theme to this fragment. This way of working allows interpretation by the coder; the coder can capture the latent meaning of a text. The method is time consuming, as human coders have to perform the coding task. The coder has to read between the lines, not only the manifest content might be looked for, but also the latent meaning might be considered. This meaning can only be understood when the point of view of the sender of the message is taken into account. Here is the main difference compared to the instrumental approach to coding. Here is also the argument for why a computer program will have difficulty in performing the coding task.

In the situation where – short - answers to open ended questions are available it is often difficult to distinguish between the two approaches. Usually manual coding is performed. Here interpretation of the answer by the coder is relevant. Does he make a correct interpretation and will he not fall into the trap of ambiguity. For such a good job a coder training is necessary to optimize the coding task. Besides at least a part of the coding task must be performed by several coders. This allows computing intercoder reliability: are the codings consistent?

## **6. Coding**

Coding is “the process whereby raw data are systematically transformed and aggregated into units which permit precise description of relevant content characteristics” (Holsti, 1969: 94). That is, it is the process in which recording units are identified and linked to the conceptual categories. The rules by which this is accomplished serve as the operational bond between the investigator's data and the theory and hypotheses. If coding is performed by humans, it is necessary that the coder is able to accurately identify the recording units (or example clauses). The coder also has to apply the concept categories correctly. For the coder, but also if computer coding is applied “[t]he task of coding is, after all, one of extracting from the subject’s words the essential meaning of his expressed feelings, beliefs, or knowledge about some social object” (Kammeyer & Roth, 1971: 82).

### *6.1. Coding system*

The codes or themes that are coded usually refer to the “what” (substance) or the “how” (form) (Holsti, 1969: 28). The way the set of codes is found is a process that is

generally underexposed. Both deductive and inductive reasonings are followed. First it is necessary to ask why the open-ended question is included in the questionnaire. The investigator knows why the question is relevant and the content of the codes becomes more or less clear based on this relevance. By asking “why did you vote for party X” the investigator might want to learn whether the respondent has ideological arguments in mind or whether qualities of the party leader are relevant. A problem is how to ask this in a closed question. It might also be that the question is relevant but the investigator is not able to indicate yet what should be in the codes. The same question “why did you vote for party X” gives rise to lots of alternatives for sets of codes. There is more than the two arguments just mentioned. Later on we will discuss one such alternative. There are also situations in which the question in itself is not relevant. This is when it functions as an exhaust valve for the respondent (“is anything left that you want to tell us?”) or to complete a set of codes.

For the development of concept categories three ‘ideal’ types are distinguished:

- A set of concept categories is developed *a priori*, based on theory underpinning the research project. These categories are an operationalization of theoretical notions;
- The concept categories are ‘data-driven’, i.e., they are constructed *a posteriori*, based on words or phrases in the texts that are analyzed. Here it is possible that each individual coder develops themes.
- A combination of these two approaches. The first type will probably not hold for open-ended questions with more complex answers, and the second type might end up in an unsystematic and unstructured list of categories. In the combination the development of a structured categorical system starts with some theoretically based categories. The fine-tuning and extension that follow is text based.

Criteria for the process of developing codes are among others the following. The codes should fit to the question, should have relevance and semantic *validity* (they must represent what respondents had in mind saying). Each category is formulated in a neutral way, in common language, must be specific and concrete and the formulation is as short as possible. The categories should be mutually exclusive, the whole set should be collectively exhaustive and contain a logical ordering. One might also want to consider the costs and time required to perform the coding task.<sup>4</sup>

In most studies the author does not explain how the codes that are used have been found. An exception is Kurasaki (2000). She describes how she developed a priori her codebook using a grounded theory approach. The steps followed were: annotate five transcripts together with respect to the interview's contents, sort the annotations into similar codes and sub-codes and label the thematic categories. General openings to the finding of codes are found in Contas (1992) and Bogdan and Biklen (2002: 167 ff).

### *6.2. Complexity*

Coding tasks might be ordered on complexity. The more complex the task the more difficult it will be to get a high amount of agreement in the assignments by independently operating coders, or in other words, the higher the measurement error. Crittenden and Hill (1971: 1078) distinguish three types of coding tasks: A, B1, and B2. These types have been summarized: "Type A coding tasks require a coder to find a specific answer to an explicit question at a given place on an instrument. Type B1 coding tasks involve locating relevant information within a larger context ..., type B2 coding tasks are those where the coder has not only to locate relevant information, but also to evaluate the relative importance of two or more possible responses to arrive at a single code" (Montgomery & Crittenden, 1977: 236). This distinction will come back later.

Crittenden and Hill (1971) were looking at human coders, who also had to develop a code system. The types of tasks they distinguished can also be used when the codes are already available, and the manifest or latent content of texts is to be coded by humans or by a computer. The tasks are becoming more complex. This will have consequences for the results of the coding task. As this task becomes more complex reliability will usually decrease. Such has been investigated: "The more complex, wide range, and inclusive a category is, the more likely that manual coding will find more instances of applicability for this code. However, for codes that are easy to operationalize using a limited number of specific words and expressions, computer content analysis rarely fails in finding these cases" (Linderman, 2001: 107).

Type A coding tasks require a coder to find a specific answer to an explicit question at a given place on an instrument. Crittenden and Hill (1971) indicate that variables like age, department affiliation and salary are part of this type. These are the technically and apparent open-ended questions as distinguished before. This type of coding is applied very often in case open-ended questions are used. The formulation

of the question is relevant, not: “which languages are spoken at your home?” but “which language is spoken most often at your home?” Only one answer is permitted. Looking at open-ended questions in NKO2006 the coding of the answers to knowledge questions like “Who is the leader of party X” or “What is the function person Y has” are of this type. In general this coding task can be performed very well by using a computer program in which the instrumental view is applied.

Type B1 coding tasks involve locating relevant information within a larger context. This type is not found in NKO2006. A variation on a question used before however might serve as an example: “Give one reason for voting for that party.” Type B2 coding tasks are those where the coder has not only to locate relevant information, but also to evaluate the relative importance of two or more possible responses to arrive at a single code. In NKO2006 respondents often used several arguments when they explained why they voted for a specific party. For example one respondent has answered that the decision to vote for a specific party was a strategic choice. The respondent however also agreed most with the party’s social policy. It is possible that such answers are coded into different variables, i.e., groups of concept categories of different types. This type of coding tasks can only be performed correctly when the representational view on coding is followed. One possible reason for low agreement in type B1 and type B2 coding tasks might be the not meeting of the requirement of taking into account by the coder the frame of reference of the respondent (Montgomery & Crittenden, 1977: 236).

These why-vote-for-a-specific-party-arguments are coded into a measure of conceptualization of ideology (Popping, 2012)<sup>5</sup>. The task, a B2 coding task, has been performed using a computer program for instrumental text analysis, using a dictionary, and twice by human coders using the representational view on coding. Many answers are not coded by the computer program. Three important reasons for this omission are in general: The dictionary can be (1) incomplete – it was constructed before coding has started and not alternative texts have been checked - , (2) inaccurate – some search entries are incorrectly linked to a certain concept category- and (3) unable to give an interpretation – see the column at the right in Table 1. The formulation of the arguments is such that based on the dictionary no classification is possible. The column at the left contains examples of formulations that allowed a classification.

\*\*\* Table 1 about here \*\*\*

In case the human coders disagreed it was most of all with respect to assigning to the categories Nature of the times and No issue content. An example where this occurred is with respect to the following answer: “the broom needs to go through; attention should be given to other people, not only to the guys with the big money.” The answer could not be assigned based on information the dictionary. The same is found with respect to the answer: “bombshell”. I would code the first answer as Nature of the times as the respondent asks attention for the common people. The second answer indicates that the present coalition parties should not return, it contains no issue.

The assignments by the coders resembled each other very much ( $\pi = 0.86$ ;  $N = 2173$ ). For the greater part this result is possible because explicit rules have been followed during the coding process, as is requested by Popping and Roberts (2009). There were no systematic differences in the assignments by the coders. More explicit rules than followed however are possible. The comparison of the assignments by the coders to the computer coding does not show that good results ( $\pi = 0.59$ , resp.  $0.57$ ;  $N = 1481$ ). The difference in sample size is due to the fact that a number of answers could not be coded by the program as they were not in the dictionary. This can for the greater part be ascribed to the fact that the coders could understand what the respondents had in mind. Another consequence of using a computer program is that the program is not able to choose the most relevant one in case more than one argument is given (type B2 coding). The study shows that the way of coding can cause enormous differences in the final classification. This already becomes visible when one looks at the frequency distributions (Table 2).

\*\*\* Table 2 about here \*\*\*

The computer program often did choose for “the nature of times” (good or bad times are linked to the party in control) where the coder decided “no issue content” (no reason – often the party leader is mentioned) is going on. This shows that the decision with regard to a specific search entry used for the instrumental coding works out differently when a human codes the answers. Looking at the answers not coded by the computer program one sees that relatively more answers are assigned to the categories “no issue content” and “group interests” (opinion is connected to the group one



belongs to - labor union, interest group, class, and race) and less to the category “the nature of times”.

Coding really open-ended questions is hard. Therefore it is necessary the investigator wonders whether the question is really needed and if so, he or she should before data are collected have concrete ideas with regard to analyzing this question. If such ideas are not there it turns out that very often the question is not analyzed at all. Therefore the respondent has made an unnecessary effort represented.

## **7. Discussion**

The main goal of this contribution is to make visible a number of steps and decisions to be made when open-ended answers in a survey are to be coded. In doing this implicitly two main questions have been addressed. The first deals with the way the coding is to be performed. The second is on whether open-ended questions should be used at all in surveys.

The way text analysis works has been explained. Two ways of coding have been discussed. Looking at the way most coding is accomplished, the instrumental way is to be followed. This is supported by artificial intelligence where the position is taken that the computer can learn how to do the coding based on previous assignments. Looking at the level of abstraction of the codes used, at decisions to be taken, at latent meanings, at ambiguity in the answer, the representational view is preferred.

The way to be followed is up to the investigator, both have benefits and back draws.

Is it skillful to use open-ended questions in a survey? Yes, as far as technically and apparent open-ended questions are concerned. These are clear questions, resulting in clear answers. It is different when really open-ended questions are used. Often it seems doubtful whether the answers contribute information that is not captured in closed codes. Sometimes the amount of information is so broad, that it becomes almost impossible for an investigator to categorize this.

The example study has been used to give an impression. It shows that in this field a lot of research is still needed.

## **Notes**

1. In Van der Eijk et al (1988) the open-ended questions are coded. In the data set made public the original question has been replaced by one or more questions having fixed answering possibilities.
2. Today many computer programs for text analysis are available. Most of these programs count the number of occurrences of themes. As input the programs need a text file containing the answers to an open-ended question. Sometimes computer programs are used that do not (only) count the number of occurrence of themes, but allow some kind of analysis. Brugidou (2003) describes the analysis of answers to open-ended questions by using a program that identifies homogeneous subsets of verbatims on the basis of their lexical profile.

The coded answers appear in a file; this file can easily be merged with the file containing the coded answers to the closed questions.

3. One way to overcome the problem is to code both search entries “going” and “not going”. In case both corresponding themes appear in a recording unit, the occurrence of the theme which is recognized via the word “going” is to be removed. This is done afterwards in the program that is used for the statistical analyses.
4. Alternatives are available for the looking for codes and the coding to follow. Starting from a matrix containing co-occurrences between significant words in the answers by the respondents a method resembling principal component or cluster analysis can be applied. The first method was already proposed by Iker (1974); it results in factors that can be labeled. The second is used by among others Schonhardt-Bailey (2005). Here clusters are found that can be labeled, but in between also hierarchical category structures can be identified. Jackson and Trochim (2002) used this approach to analyse answers to open-ended questions. At this moment I consider this approach most of all as an alternative way to find codes.
5. For details on the measure of conceptualization of ideology see Converse (1964). Determining under which code one falls is very complex, respondents cannot do this themselves.

## References

- Bogdan, R.C. & Biklen, S.K., 2002. *Qualitative Research for Education. An Introduction to Theory and Methods*. Boston, MA: Allyn & Bacon.

- Brugidou, M., 2003. Argumentation and values: An analysis of ordinary political competence via an open-ended question. *International Journal of Public Opinion Research*, 15 (4): 413-430.
- Contas, M.A., 1992. Qualitative analysis as a public event: The documentation of category development procedures. *American Educational Research Journal*, 29 (2): 253–66.
- Converse, J.M., 1984. Strong arguments and weak evidence: The open / closed questioning controversy in the 1940s. *Public Opinion Quarterly*, 48 (1B): 267-282.
- Converse, P. E., 1964. The nature of belief systems in mass publics. Ideology and discontent. In: Apter, D.E. (ed.), *Ideology and Discontent*. New York: The Free Press: 206-261.
- Crittenden, K.S. & Hill, R.J., 1971. Coding reliability and validity of interview data. *American Sociological Review*, 36 (6): 1073-1080.
- Fribourg, O. & Rosenvinge, J.H., 2013. A comparison of open-ended and closed questions in the prediction of mental health. *Quality and Quantity*, 47 (3): 1397-1411.
- Frisbie, B. & Sudman, S., 1968. The use of computers in coding free responses. *Public Opinion Quarterly*, 32 (2): 216-232.
- Geer, J.G., 1988. What do open-ended questions measure? *Public Opinion Quarterly*, 52 (3): 365-371.
- Geer J.G., 1991. Do open-ended questions measure “salient” issues? *Public Opinion Quarterly*, 55(3): 358-368.
- Heinrich, H.A., 1996. Traditional versus computer aided content analysis. A comparison between codings done by raters as well as by Intext. In: Faulbaum, F. & Bandilla, W. (eds.), *SoftStat'95. Advances in Statistical Software 5*. Stuttgart: Lucius & Lucius: 327-233.
- Holsti, O.R., 1969. *Content Analysis for the Social Sciences and Humanities*. London: Addison Wesley.
- Hruschka, D., Cummings, B., Cobb St. John, D., Moore, J., Khumalo-Sakutukwa, G. & Carey J. W., 2004. Fixed-choice and open-ended response formats: A comparison from HIV prevention research in Zimbabwe. *Field Methods*, 16 (2): 184-202.

- Iker, H.P., 1974. An historical note on the use of word–frequency contingencies in content analysis, *Computers and the Humanities*, 8 (1): 93–98.
- Jackson, K.M. & Trochim, W.M.K., 2002. Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research methods*, 5 (4): 307-336.
- Kammeyer, K.C.W. & Roth, J.A., 1971. Coding responses to open-ended questions. In: Costner, H.L. (ed.), *Sociological Methodology 1971*. San Francisco: Jossey Bass.
- Krippendorff, K., 2012. *Content Analysis. An Introduction to its Methodology*. Thousand Oaks: Sage.
- Krosnick, J.A., 1999. Survey Research. *Annual Review of Psychology*, 50: 537-567.
- Kurasaki, K.S., 2000. Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field Methods*, 12 (3): 179-194.
- Lazarsfeld, P.F., 1944. The controversy over detailed interviews – an offer for negotiation. *Public Opinion Quarterly*, 8 (1): 38-60.
- Linderman, A., 2001. Computer content analysis and manual coding techniques: A comparative analysis. In: West, M.D. (ed.), *Theory, Method, and Practice in Computer Content Analysis*. Westport: Ablex: 97-109.
- McColl, E., Jacoby, A., Thomas, L., Soutter, J., Bamford, C., Steen, N., Thomas, R., Harvey, E., Garratt, A., Bond, J., 2001. Design and use of questionnaires: A review of best practice applicable to surveys of health service staff and patients. *Health Technology Assessment*, 5 (31): 1–256.
- Montgomery, A.C. & Crittenden, K.S., 1977. Improving coding reliability for open-ended questions. *Public Opinion Quarterly*, 41 (2): 235-243.
- Popping, R., 2012. Human or machine coding in text analysis. *Bulletin de Méthodologie Sociologique*, 115: 79-88.
- Popping, R. & Roberts, C.W., 2009. Coding issues in modality analysis. *Field Methods*, 21 (3): 244-264.
- Roberts, C.W. & Popping, R., 1993. Computer supported content analysis. Some recent developments. *Social Science Computer Review*, 11 (3): 283-291.
- Schonhardt-Bailey, C., 2005. Measuring ideas more effectively: An analysis of Bush and Kerry’s National Security speeches. *PS, Political Science and Politics*, 38 (3): 701-711.

- Scholz, E. & Züll, C., 2012. Item non-response in open-ended questions: Who does not answer on the meaning of left and right? *Social Science Research*, 41 (6): 1415-1428.
- Schuman, H. & Presser, S., 1981. *Questions and Answers in Attitude Surveys*. New York: Academic Press.
- Shapiro, G., 1997. The future of coders: Human judgments in a world of sophisticated software. In: Roberts, C.W. (ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*. Mahwah, NJ: Lawrence Erlbaum Associates: 225-238.
- Smyth, J.D., Dillman, D.A., Christian, L.M. & McBride, M., 2009. Open-ended questions in web surveys. *Public Opinion Quarterly*, 73 (2): 325-337.
- Van der Eijk, C., Irwin, G.A. & Niemöller, B., 1988. *Dutch Parliamentary Election Panel-Study 1981-1986*. Amsterdam: Steinmetz Archives, Dutch Interuniversity Election Workgroup.

### **Appendix – Computer programs for text analysis**

Most computer programs for quantitative text analysis have been developed by researchers themselves and usually are written in the context of a specific study. This generally implies the program is not for general use. But if others want to use the program, this is fine. Documentation usually is poor and the program is as it is. This implies there has not always been a complete control for imaginable bugs and errors are not captured. With respect to the data presented in this text the manual coding was performed in SPSS, simply because the data were available in a string variable in that program. A new variable to contain the codes was inserted next to this string variable. For the automatic machine coding a program was used I once wrote for my course on text analysis. The possibilities this program offers are very limited, but error handling is taken care of. The number of investigators using quantitative text analysis (unfortunately) is not that high, for that reason I do not expect that general programs that are well documented, free of errors and available for all kinds analysis do come up.

In the meantime one can look at some requirements for such programs. Questions concerning the program itself from the point of view of the user should at least include the following:

1. Which operating system is to be used (most investigators use computers running under Windows or UNIX)?
2. Can the program be run on languages other than English (also think of languages that use other than Latin characters, like Cyrillic, Chinese, Japanese, and so on – in technical terms, this refers to the UTF - Unicode Transformation Format).
3. How is the input the program requires organized (each text on a separate file or all texts in one file; how to recognize the beginning of a new text and a new paragraph)? Most programs require a plain text file and cannot deal with all kinds of characters as used in text editors like Word, WordPerfect, or OpenOffice.
4. What will be the output and how does it look like (minimal a data matrix in which occurrences of themes are counted and which can serve as input for a program for statistical analysis. There are several programs however that can perform specific analyses, usually because these programs are written in the context of a specific research question)?
5. If used for manual coding, which additional facilities are offered (keyword in context; how to catch manifest and latent meanings)?

Besides, the user should be comfortable with the program. The order in which buttons and edit lines should be used must be logical and if possible in line with the way of thinking of the layman user. No cryptic remarks should appear in the programs windows.

Next there are questions concerning that what is investigated. Actually these should not be linked to the computer program, but there are programs that have a dictionary included. This makes these programs can only be used for specific research questions.

6. Is a standard dictionary available? (Often the answer is no, but if there is a dictionary it is often related to a specific research question. Still these dictionaries might be very complete.
7. If not, how to develop such a dictionary and when is it complete (the dictionary becomes especially complex if search entries are used that are in prefix, suffix, and infix position)?

At least there is one important other question which often comes too late:

8. What will be the recording unit? In the situation where one has open ended questions having short answers, the unit usually is the complete answer by the respondent. But how to decide in case one has longer texts. Especially in the situation where co-occurrences are looked for this question becomes very relevant.

One can make things even more complex. What happens when for example machine learning comes in? Here the most probable concept category is selected based on a comparison to a small dataset that has been really coded. To do so, Bayesian statistics is often used. This however goes too far for our purpose here to consider some points that are relevant when one chooses a program or wants to develop one.

Some general computer programs that are available for quantitative thematic text analysis at this moment are: TCA and Yoshikoder (which are free), WordStat, TextQuest and T-Lab (which have a test version). Wordscores contains a set of Stata programs and ReadMe a set of R programs. TCA is most of all for manual coding (can handle latent themes), the other programs are for machine coding.

**Table 1 – answers assigned by the coder and in dictionary or just not**

<i>Category</i>	<i>assigned by coder and in dictionary</i>	<i>assigned by coder but not in dictionary</i>
Ideologues	Liberal thoughts.	Because their ideology is closest to me.
Near ideologues	This is the only social democratic party.	Because that one is closest to Gods word and that is how one wants to live. I always do and because of religious conviction.
Group interests	I belong to the working class, so I vote for this party.	Works for the higher educated. They are there for the common people.
Nature of the times	Because of the social policy and the attention for the environment. A strategic choice; social policy.	Only party that tries to survive in a healthy and clean Netherlands. Improvement of the road network.
No issue content	Trust this party. Always voted for this party. Best choice.	The most solid political party; least clamorous party. The party where I feel best.



**Table 2 - frequency distribution based on different way of coding**

Category	Coding by computer program in %	Coding by one coder in %	Coding by one coder in % (all data)
Ideologues	.9	1.0	1.0
Near ideologues	.1	2.6	2.5
Group interests	4.2	5.4	6.4
Nature of the times	51.9	44.6	39.7
No issue content	42.9	46.4	50.4
	N = 1481	N = 1481	N = 2173

Categories as defined by Converse (1964)

