Spotlight Session – Analysing answers to open-ended questions from surveys

Excel format for data preparation:

Notes:

- Column A controls the grouping of the texts in MAXQDA's "Document System" window. Enter the same phrase for all rows if you want all texts to be in a single group, here "Survey Data".

- Column B contains the labels for the Document System in MAXQDA, it should have a unique identifier for each respondent.

- The labels at the top of each column (Row 1) will appear exactly like this in MAXQDA, so use labels that will be meaningful in that different context, and remember that shorter labels are better than long ones.

- If you use Survey Monkey for data collection, there is a special routine in MAXQDA to collect the data from your Survey Monkey account which by-passes this step. Full instructions are available in the MAXQDA Online Manual (click the "(?)" button at the top right-hand corner of the MAXQDA interface).

- When your data is ready, close the Excel workbook, open the analysis project in MAXQDA and use the option **Import > Survey Data > Import Data from Excel Spreadsheet**. You will need to navigate to the folder with the Excel workbook and select it there to start the process.

Dialog during import routine in MAXQDA:

**Import documents from Excel spreadsheet**   ✕

Which column contains the labels for …

… the document group? DOCUMENT GROUP ▾

… the document name? DOCUMENT NAME ▾

Mark the columns to be imported and automatically coded as text.
Mark the columns to be imported as variables.

Column	Code	Variable	
DOCUMENT NAME	☐	☐	
EMP STAT	☐	☑	
FAM CH	☐	☑	
MARITAL	☐	☑	
AGE GRP	☐	☑	
REGION	☐	☑	
Q.1. JOB-SECURE	☐	☑	
Q.2. HOW-AFF	☑	☐	
Q.3. FAULT	☐	☑	
Q.3A COMMENTS FAULT	☑	☐	

Options

☐ Code empty cells

Documents that exist in project before import (same document name in document group):

◉ Import

◯ Ignore for import

◯ Add text to existing documents

ⓘ   [OK]   [Cancel]

Notes:

- This is where you separate variables from qualitative texts. The texts to be analysed are marked in the "Code" column, the short or quantitative data goes into the "Variable" column.

- At the top of this dialog, the program has taken the header labels from columns A & B in the workbook and inserted them in the menu boxes, so you could use different labels for these columns in the workbook but these are the clearest.

- I recommend leaving the "Code empty cells" box blank (as above) because then you get a simple count of all collected responses for each question and non-responses are ignored.

- A similar version of this dialog appears as part of the special routine for importing data from Survey Monkey.

MAXQDA setup for analysing responses to one question:

Notes:

- I have activated the entire Document Group "Survey" in the Document System and activated the Code "Q3A Comments Fault" in the Code System to get this display. I adjusted the window boundaries so that the "Retrieved Segments" window is larger than the "Document Browser" window (I find the horizontal display better for this type of data). All 48 responses to Q3A are now readable in MAXQDA's "Retrieved Segments" window.

- Note that 143 cases in this survey (of 191 people) did not answer this question.

- The Document Browser shows all the responses to the 4 open questions from a single respondent.

- From this display, you could work manually by coding the responses directly from the "Retrieved Segments" window, or you can use a more efficient approach with the option **Analysis > Categorize Survey Data**, however you are also ready for a more automated approach, as outlined below.

- Before starting your coding work, it is good practice to enter the full text of each question into the Code Memo for that question. This means that you can remind yourself of that wording at any stage during the analysis. The wording for the Q3A question in this data was *"Please expand on your understanding of who/what is at fault"*.

Using MAXDictio to analyse the contents of the "Retrieved Segments" window:
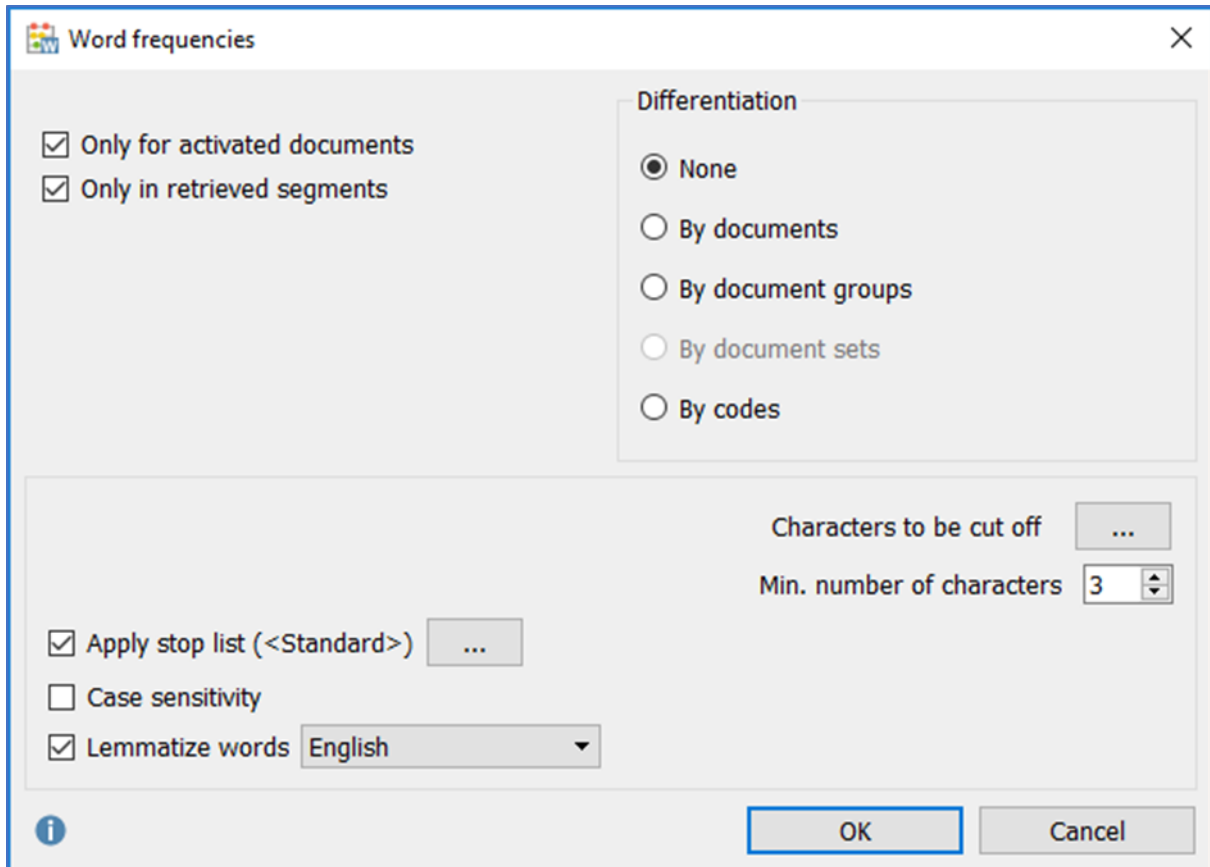
Notes:

- The activated documents and retrieved segments settings limit the routine to the single question that is currently appearing in MAXQDA's "Retrieved Segments" window.

- Differentiation is not helpful in this particular application, so leave the default on "None".

- "Stop lists" can be helpful at avoiding counting trivial words like "of", "and", etc. You can edit them and create your own stop lists for particular purposes, but it may not be necessary for a one-off survey project.

- The "Lemmatize" function will combine words with a common stem, in this example the count for "bank" includes "Banks" and "banking" which is useful in this particular application.

- The "Min. number of characters" setting will exclude words shorter than this length from the count, so "by" will not be counted here but "pay" and "job" will be included. Setting this higher might exclude some important words.

- Click on "OK" to run the frequency count.

Frequency table and search results:

Notes:

- In the "Word frequencies" table, select a word of interest with a click to highlight its row. Use a right-click (Windows) and select "Search results" from the context menu to display the "Search results" table as well as the first. Click on a row in the results table to see its full text highlighted in the Document Browser behind these 2 windows.

- Often the preview display (effectively keyword in context) is enough to identify whether the word has been used in a consistent way in each instance. If an item is inconsistent, exclude it from the next step by highlighting it and clicking on the "no entry" button (shown here for case 163).

- Use the 5th or 6th buttons in the toolbar of the results table to autocode all of the remaining hits in that table.

- In the autocode dialog, think about how much text you want to apply the code to. The option "Only search string" will just code the selected word each time so you may get multiple hits within some individual responses, this will be useful if you

want to distinguish between someone who mentions a concept just once and another person who writes about it a lot. More usually with this kind of data, coding the whole paragraph containing the hit will give you just one coded segment per response - allowing you to simply identify how many respondents mentioned the concept.

- Repeat the search and autocode approach for the most common themes arising for this question (I used "bonuses", "Europe", "Euro", and "Politicians" for this data), then close all of these windows and switch to the new categorization tool in the **Analysis** menu.

Using the Categorize Survey Data tool:
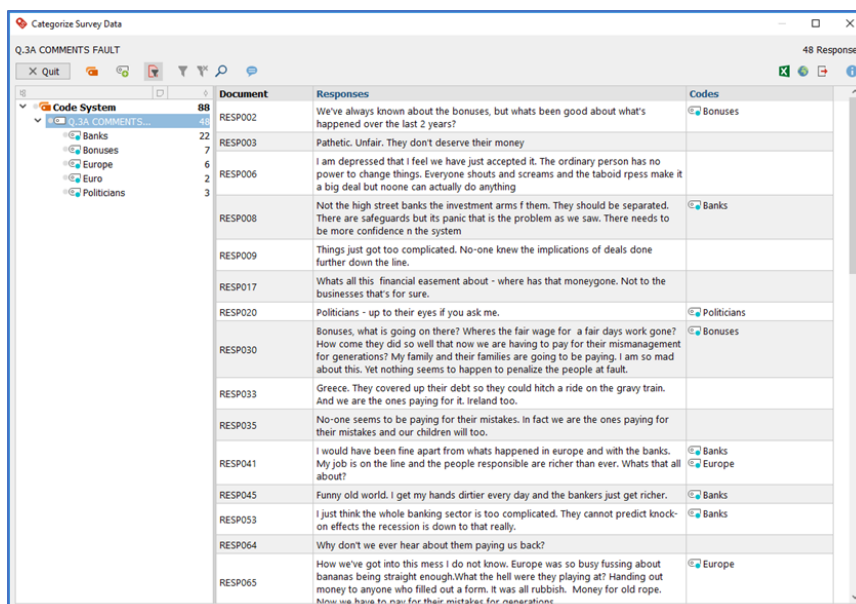
Notes:

- It may be necessary to drag the new codes that you have just created into subcodes of the question in the "Code System" window, because that is the basis on which this tool works. Then select the option **Analysis > Categorize Survey Data**, drag in the code of the question that you are analysing and click "OK" to see the screen shown above.

- You can now see all of the initial coding that you have done with the MAXDictio and autocoding tools, so you can recheck that (if you want), but most likely you will want to focus on the segments that have no code attached as yet.

- Apply a code by dragging from right to left (data segment onto the code).

- As you identify new themes by reading the uncoded segments, you can create codes for them with the 2nd icon to the right of the "Quit" button. If you want to search all of these segments for a particular word, you have a search tool at the 6th icon. If you want to add comments about any segments, you can open the column for those with the 7th icon.

- This display gives you a strong visual check for uncoded segments. It may be useful to create a code for the ones that cannot be coded to a meaningful code (say "No clear answer" or "uncategorizable") and apply it to these.

- By clicking on each thematic code in turn, you can quickly check each subset for consistency, and finally, re-examine the uncategorizable segments.

- Or, you can quit this tool, activate just the "uncategorizable" code for this question, and click on an item in the "Retrieved Segments" window to check the context of the other text responses from that respondent. You can also look at the variable values for those cases to see if they help explain the meaning of the comment (Tip: to see variable values for retrieved segments, go to **Variables > List of Document Variables** and put a tick in the column "Display as tooltip" for the relevant variable).

Reporting the results of your analysis:

Further explanations of who caused the financial crisis in 2008
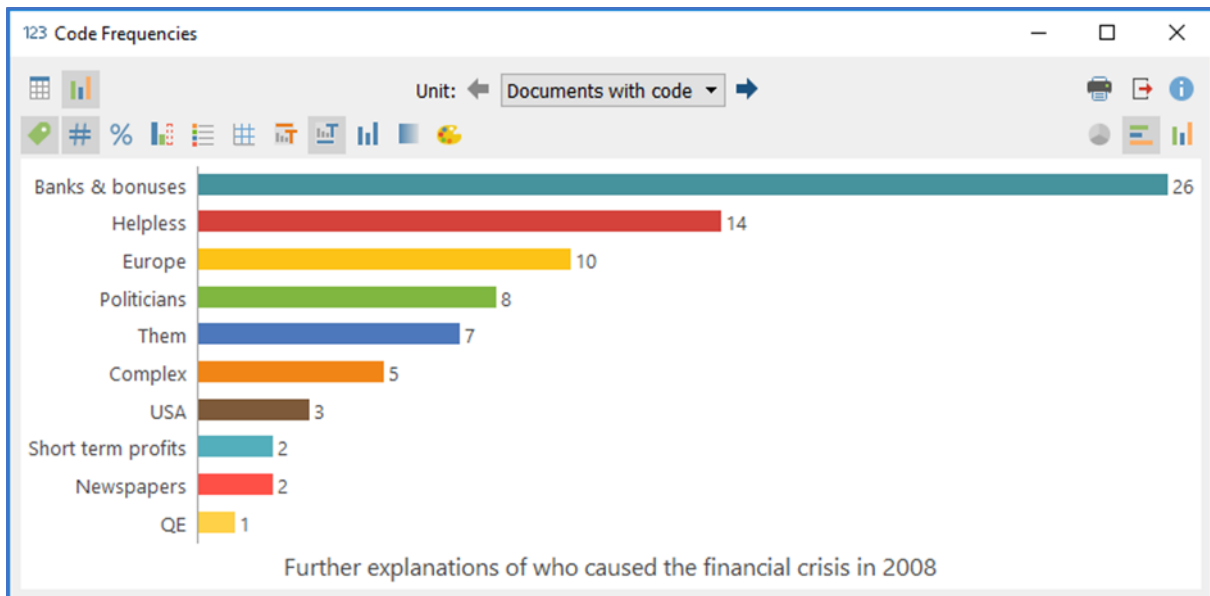
Notes:

- To obtain a display similar to the one above, you need to use the tool **Analysis > Code Frequencies**. It will be best to activate the relevant group of codes in the Code System first so that you can use the option to "Paste activated codes" in the set-up dialog (and you will then need to "Remove" the question code itself, which is the group header, to see the frequencies of only the analysis codes).

- You can explore the table and chart settings, which are all controlled with the toolbars in the display window.

- You can report on the quantitative elements of your survey by using the **Variables** functions. For example, try **Variables > Document Variable Statistics** and select several of the closed question identifiers in the set-up dialog. You can then scroll forwards and backwards through these in table or chart views. Note that you can re-order the tables by frequency or alphabetically and that this controls the equivalent charts.

Combining data from closed and open questions:

You can combine the quantitative and qualitative elements of your survey by using MAXQDA's **Mixed Methods** tools. For example, using this data, we can create a Crosstab of the responses to the question about how people have been personally affected analysed according the age-group of the respondents.

Crosstab	20-29	30-39	40-49	50-59	60-69	70-79	Total
Q.2. HOW-AFF	10.2%	22.4%	21.8%	11.6%	33.3%	0.7%	100.0%
Jobs lost	15.4%	28.2%	38.5%	7.7%	10.3%		100.0%
Family	10.7%	3.6%	21.4%	10.7%	53.6%		100.0%
Pensions		7.7%	11.5%	3.8%	76.9%		100.0%
Insecure	11.1%	50.0%	16.7%	5.6%	16.7%		100.0%
Income down	11.8%	5.9%	29.4%	23.5%	29.4%		100.0%
Not affected	6.3%	25.0%	6.3%	12.5%	50.0%		100.0%
Loans		11.1%	44.4%	33.3%	11.1%		100.0%
Own business struggling		44.4%	33.3%	22.2%			100.0%
Travel	12.5%	12.5%	25.0%		37.5%	12.5%	100.0%
Cut back spending	16.7%	16.7%	33.3%		33.3%		100.0%
Savings					100.0%		100.0%
Morale	20.0%	40.0%	20.0%		20.0%		100.0%
Closed shops		25.0%		25.0%	50.0%		100.0%
Stress		50.0%	50.0%				100.0%
Farming	33.3%		66.7%				100.0%
Holidays		100.0%					100.0%
Σ SUM	9.5%	21.7%	23.4%	10.7%	34.1%	0.6%	100.0%
# N (Documents)	24 (12%)	38 (19%)	51 (26%)	19 (9%)	58 (30%)	1 (0%)	191 (100%)

Notes:

- Here we can see that the comments about effects on pensions were made mainly by people in their 60's.

- This is where the full pay-off for using MAXQDA to analyse your survey data becomes apparent!

Reference:

Fielding, J., Fielding, N., and Hughes, G. (2012): Opening up open-ended survey data using qualitative software: Quality and Quantity.

Graham Hughes (2019)
n.graham.hughes@ntlworld.com