1.0 Analysing Open-Ended Survey Question Data in CAQDAS Packages: Data Preparation for MAXqda

The following material provides step-by-step guidance for preparing the texts, collected from a large number of respondents answering several open-ended questions in a survey situation, for qualitative analysis using MAXqda software.

There are several stages in this procedure, some of which may not be relevant in your circumstances. Some users may have alternative methods or short-cuts, in which case please use your own judgment as to which elements from below to apply. Our purpose here is to provide a comprehensive guide, which has been tested and proved to work, for the benefit of those who have not achieved this task successfully before.

Unlike some other CAQDAS packages, MAXqda includes specific routines for working with this sort of data and so the steps outlined below do not represent a "workaround" but a mainstream activity.

Outline:
1.1   Create a spreadsheet containing all of the survey texts and respondent identifiers.
1.2   Create another spreadsheet with socio-demographic attribute variables and respondent identifiers.
1.3   Save the spreadsheets as tab-delimited text files.
1.4   Import the survey text file into MAXqda using a designated routine.
1.5   Import the attributes file into MAXqda using a designated routine.
1.6   Verify the accuracy of the data transfers.

Detailed Steps:
1.1   Create a spreadsheet containing all of the survey texts and respondent identifiers.

This step is illustrated here using Microsoft Excel, however it is likely that almost any spreadsheet program could be used for these tasks as no specialised functions are required.

The recommended structure to aim for is one in which you arrange the data in a table, or grid, with a separate row for each respondent (or case). The columns should separately hold both the unique identifier for each respondent and the texts of the answers to the open ended questions.

TIP: Before carrying out any processing of the data, try to examine it in its most original format in order to identify some of the longest individual responses and to look for unusual characters in the texts. The longer responses may get truncated in some conversion processes (for example on being brought into SPSS if the 256 character default was not changed) and it is useful to be able to check that you have the fullest versions of these before you carry out the analysis work. Unusual characters,

---

other than basic alpha-numeric and punctuation ones, sometimes affect conversion processes so it is a good idea to check that these have been copied faithfully before doing analysis work.

MAXqda has specific requirements for the first row and the first two columns of the table. The first row must contain the unique identification names for the separate questions, these will appear in the code section of the program and should be short and unambiguous. The first column must contain the name of the "Textgroup" that will become a container for all of the respondents in MAXqda, and the second column must contain the unique identifiers for the respondents. This is illustrated in Figure 1 below.

Figure 1: Spreadsheet of survey texts prepared for MAXqda

Note the following in Figure 1.

- Cell A1 must contain the word "textgroup", and cell B1 must contain "textname". The strings in C1, D1 etc are the names of the separate questions in the survey.
- All the rest of the cells in column A have the word "Survey", this is the group name under which all of the respondents will appear. You may choose any word here, the important point is that the same word must be used throughout the column.
- Column B contains the unique identifier for each respondent. This has been formatted as the survey serial number prefixed with "R.". There are many acceptable formats for this but it is recommended that a consistent length and layout is used. The letter prefix causes the spreadsheet program to interpret each item as a text string instead of a number. MAXqda accepts purely numerical identifiers, so the "R." prefix is not strictly necessary. It is essential, however, that each respondent has a unique identifier and serial numbers are probably more reliable for this purpose.
- In columns C and D the format has been set to Word Wrap and Auto-Fit Row-Height so that the longer responses can still be seen. Before proceeding any further, you should check that any particularly long responses are included in full.

- It is not necessary for there to be a response text in every cell in the table, for example in this illustration respondent R.26222 did not answer question QBETT2.
- The inadvertent inclusion of a blank row or column in this table does not appear to cause a problem when the data is imported into MAXqda.

After preparing this table, save it in spreadsheet format as a back up and in case further editing may be necessary.

TIP: We have found it difficult at a later stage to add a further set of question texts to an existing MAXqda project, and so we recommend that you take care in this first step to include all of the response materials that you will need to analyse in this project.

1.2 Create another spreadsheet with socio-demographic attribute variables and respondent identifiers.

The second table is similar to the first in some respects but contains the attributes of the respondents which may be relevant to the subsequent analysis.

TIP: It is quite possible to add further attributes for all respondents during the analysis, so the decision over which attributes to include at this point is not an irrevocable one with MAXqda. However it is worth making sure that you use all of the data that you can reasonably expect to consider during the analysis, as it is easier to incorporate it at this stage.

For further comments and guidance on selecting variables to use in such analysis please see the notes [here].

Figure 2: Spreadsheet of attribute data prepared for MAXqda

	textgroup	textname	sex	age	area	work	flowarn	where	tenure
432	Survey	R.26221	F	age65\74yr	Shrews	Retired	fYwY	Outer	Prp10\20yr
433	Survey	R.26222	F	age75\99yr	Shrews	Retired	fYwY	Outer	Prp05\10yr
434	Survey	R.26225	F	age75\99yr	Shrews	Retired	fYwY	Outer	Prp20\99yr
435	Survey	R.26226	F	age55\64yr	Shrews	Wrkfull	fYwY	Outer	Prp05\10yr
436	Survey	R.26227	F	age18\24yr	Shrews	Wrkfull	fYwY	Outer	Prp01\03yr
437	Survey	R.26229	F	age35\44yr	Shrews	Wrkfull	fYwY	Outer	Prp10\20yr
438	Survey	R.26232	F	age25\34yr	Shrews	Wrkfull	fYwY	House	Prp01\03yr
439	Survey	R.26234	M	age35\44yr	Shrews	Wrkfull	fYwY	House	Prp05\10yr
440	Survey	R.26414	F	age25\34yr	Newark	Selfemp	fYwY	House	Prp01\03yr
441	Survey	R.26415	F	age65\74yr	Newark	Retired	fNwN	Notfl	Prp05\10yr
442	Survey	R.26416	M	age35\44yr	Newark	Wrkfull	fNwY	Notfl	Prp01\03yr
443	Survey	R.26420	M	age65\74yr	Newark	Retired	fYwY	Outer	Prp20\99yr
444	Survey	R.26421	M	age65\74yr	Newark	Retired	fYwY	Outer	Prp20\99yr
445	Survey	R.26422	F	age65\74yr	Newark	Retired	fYwY	Outer	Prp20\99yr
446	Survey	R.26424	M	age25\34yr	Newark	Wrkfull	fYwN	Outer	Prp01\03yr
447	Survey	R.26425	F	age65\74yr	Newark	Another	fNwN	Notfl	Prp05\10yr
448	Survey	R.26426	F	age35\44yr	Newark	Wrkpart	fNwY	Notfl	Prp01\03yr
449	Survey	R.26427	M	age35\44yr	Newark	Wrkfull	fYwY	Outer	Prp05\10yr
450	Survey	R.26432	M	age75\99yr	Nottin	Retired	fYwY	Outer	Prp20\99yr
451	Survey	R.26433	F	age45\54yr	Nottin	Wrkpart	fYwY	House	Prp20\99yr
452	Survey	R.26437	M	age75\99yr	Nottin	Retired	fYwY	House	Prp20\99yr
453	Survey	R.26438	M	age75\99yr	Nottin	Retired	fNwY	Notfl	Prp20\99yr

Note the following in Figure 2:
- Columns A and B should be identical to those set up for the texts in Figure 1.

- The first row should contain the names of the attributes.
- There should be data in every cell of this table. When this table is imported into MAXqda any blank rows or columns will be ignored, however an empty cell will remain blank and may not be recognised as missing data in subsequent analysis reports, so it may be best to enter a specific missing data label in such cells.


TIP: It may be advisable to do some work to recode scale variables into ordinal groups, as shown for the age variable above. Some routines within MAXqda will work with a numerically scaled variable but these are more likely to be used for filtering data or cross-tabulations than in regression calculations.

After preparing this table, save it in spreadsheet format as a backup and in case further editing may be necessary.

### 1.3   Save the spreadsheets as tab-delimited text files.

Each of the spreadsheets created in steps 1 and 2 needs to be saved as a separate tab-delimited text file for the import into MAXqda. If you have saved both tables as worksheets within a single Microsoft Excel workbook you will need to copy each to a new workbook and then save those separately. (In Microsoft Excel you can easily copy a worksheet by right-clicking on the name tab at the bottom of the screen, selecting the *Move or Copy* command in the menu that this brings up, clicking a tick into the *Create a copy* box, and selecting "(new book)" from the pull-down list at "To book:" – see Figure 3.)

Figure 3: Copy worksheet to a new file in Microsoft Excel
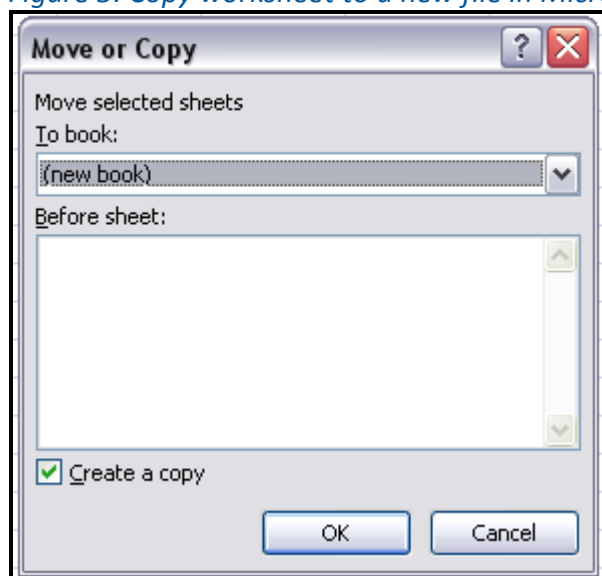
Use the *Save As* command with the field *Save as type* set to "Text (Tab delimited)" and select a file name and storage location which you will be able to access from MAXqda at the import stage (see Figure 4).

Figure 4: Save As command in Microsoft Excel
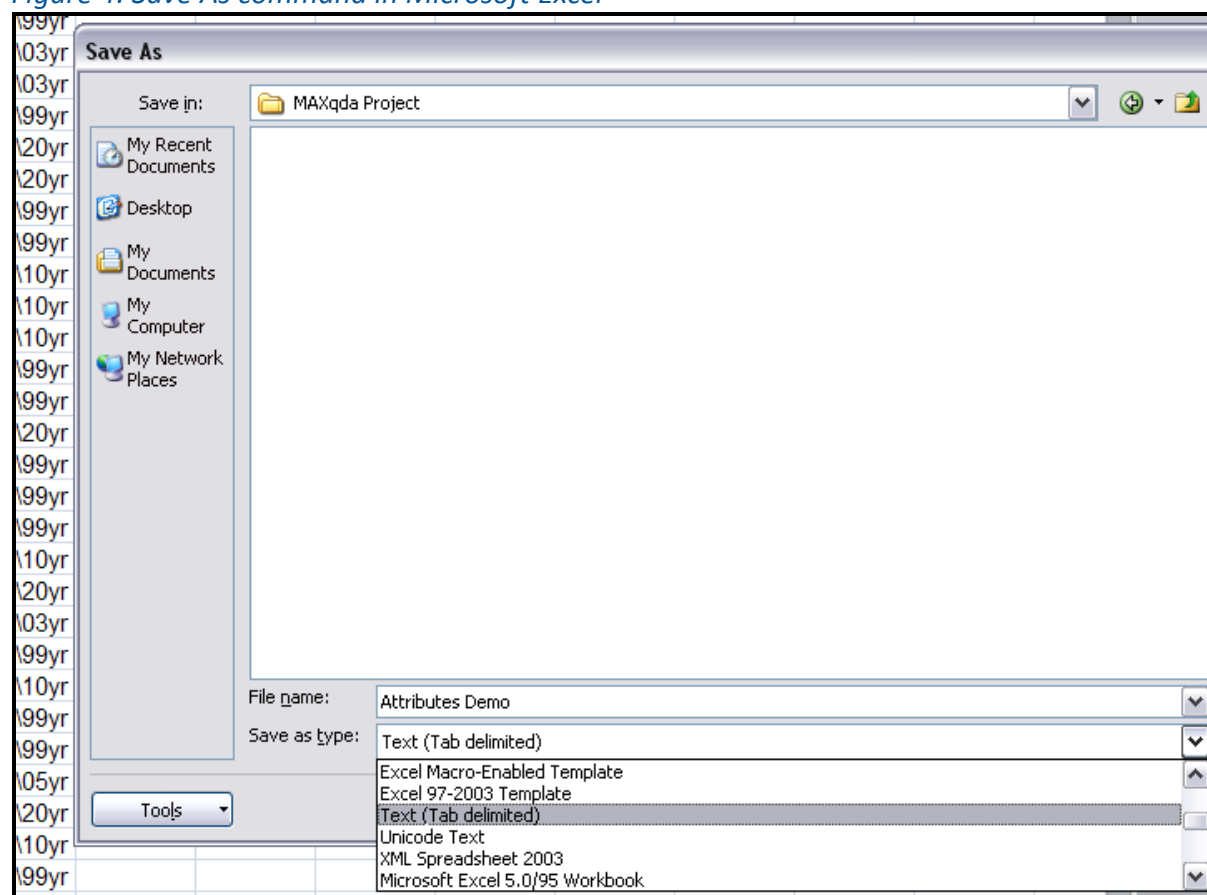
1.4 Import the survey text file into MAXqda using a designated routine.

Open a new project in MAXqda. From the menu in the top bar select *Text / Import Structured Text from Microsoft Excel*. You will be presented with an information screen which tells you how to structure the Microsoft Excel spreadsheet, and the instructions above should prove to be consistent with these, so press OK to continue. The next dialogue prompts you to navigate to the location where you stored the text file created at step 3 above (the file with the response texts is the one you need at this point), and when you have selected the correct file click on "Open".

The next dialogue asks if you want to ignore empty cells, but gives no clue over the significance of this decision, we have obtained satisfactory results by clicking "Yes" at this point.

TIP: Testing our trial data with both answers to this dialogue has revealed the effects. If you answer "Yes" to ignore empty cells the import process will be quicker because a "No" answer means that the program creates a coded segment for every cell in the spreadsheet grid whether it has data in it or not. This will also affect the coding and analysis stages by clogging your screen with empty segments.

The import process may seem to be slower in MAXqda than in other CAQDAS programs, but a lot of work is being done by the computer where the equivalent has to be done manually in ATLAS.ti or NVivo. In our example a data file with 1,257 respondents and 8 questions took around 15 minutes to process on a fairly high specification PC (July 2009). It is apparent how the process slows down as it builds the database because the textnames are displayed in the Document System panel as they are

completed. The number displayed in the top right corner of the Document System panel is the total number of coded segments in the system. In Figure 5 this can be seen to be 1,456, and this represents the total number of separate question responses made in this dataset. The highlighted respondent, R.97117, answered two questions (shown by the "2" in that column) whereas case R.97116 did not answer any of these questions. Each respondent's response to each question represents one segment. The same total appears in the Code System panel below, where the segments are sorted and coded according to the question names from the top row of the spreadsheet – for example in Figure 5 the question QBLO2 was answered by 87 respondents.

Figure 5: MAXqda immediately after importing a set of Survey texts
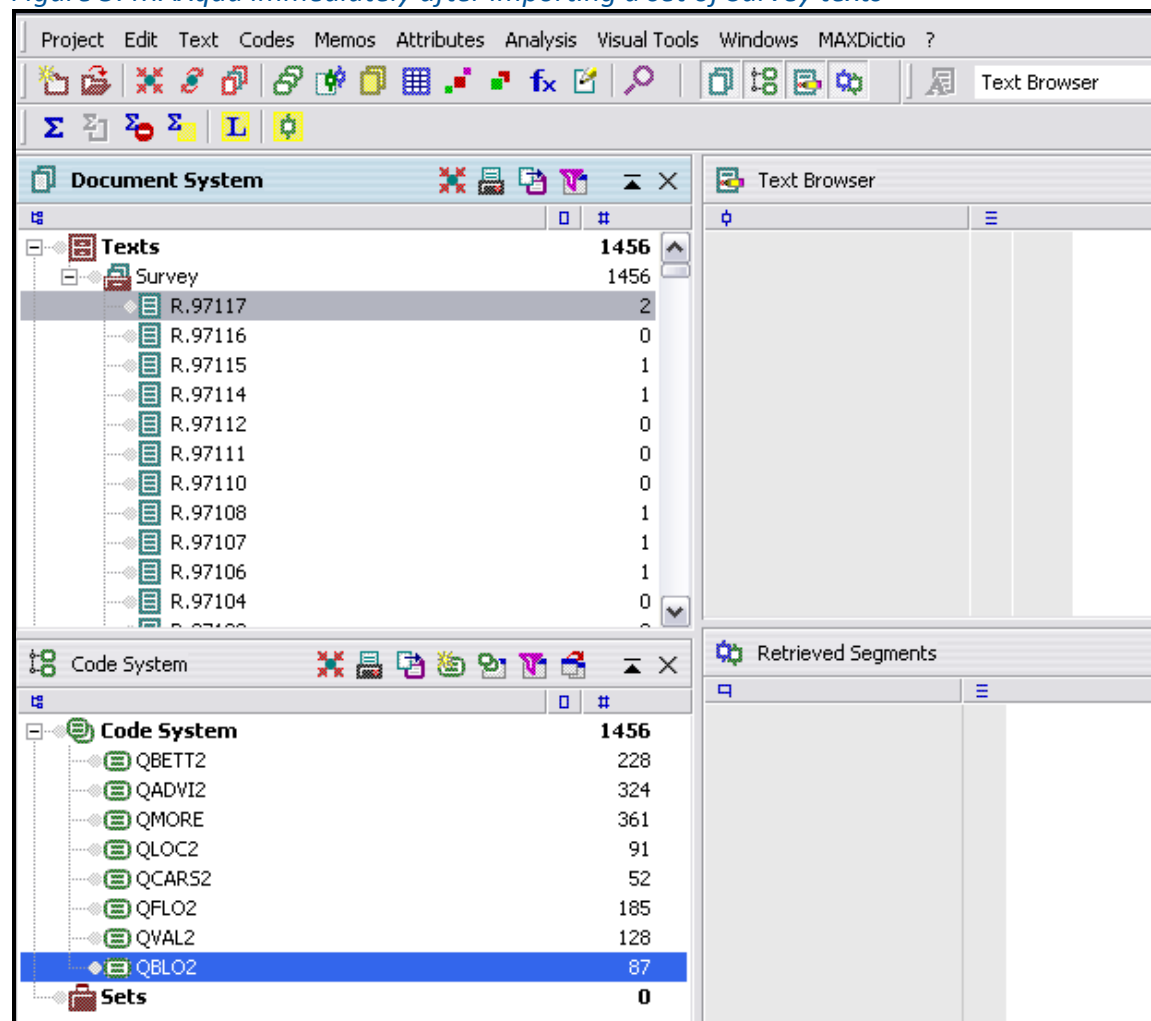
### 1.5 Import the attributes file into MAXqda using a designated routine.

The process to import the attributes is started from the Main Menu by selecting *Attributes / Import*. The dialogue screen requires you to navigate to and select the tab-delimited text file with the attribute variables that you saved in Step 3 above. Having done that, click on "Open" and the process will run.

TIP: This procedure is much faster than the text import, if you watch closely you should observe a counter in the middle of the screen clocking up to the total number of respondents.

To confirm that the data has been imported you should open the attributes table. The command *Attributes / Edit* shows an icon which is also located on the top toolbar (the grid symbol just beneath the attributes command in Figure 5 above), use either the full command or the toolbar button to open the attributes table. The table may need some editing and arranging to obtain a satisfactory display. It is shown in a separate window which can be moved and resized to suit your preferences, and you can choose which variables are to be visible. Locate the button in the attributes window which toggles between Variable View and Data View (the third from the left in the Attributes Table view in Figure 6 below). In Variable View tick just the variables that you want to see displayed and untick the ones that you want hidden, then switch back to Data View to see the effect. You can also adjust the width of any column by dragging its border in the title bar.

Figure 6: MAXqda with attributes table open in Data View
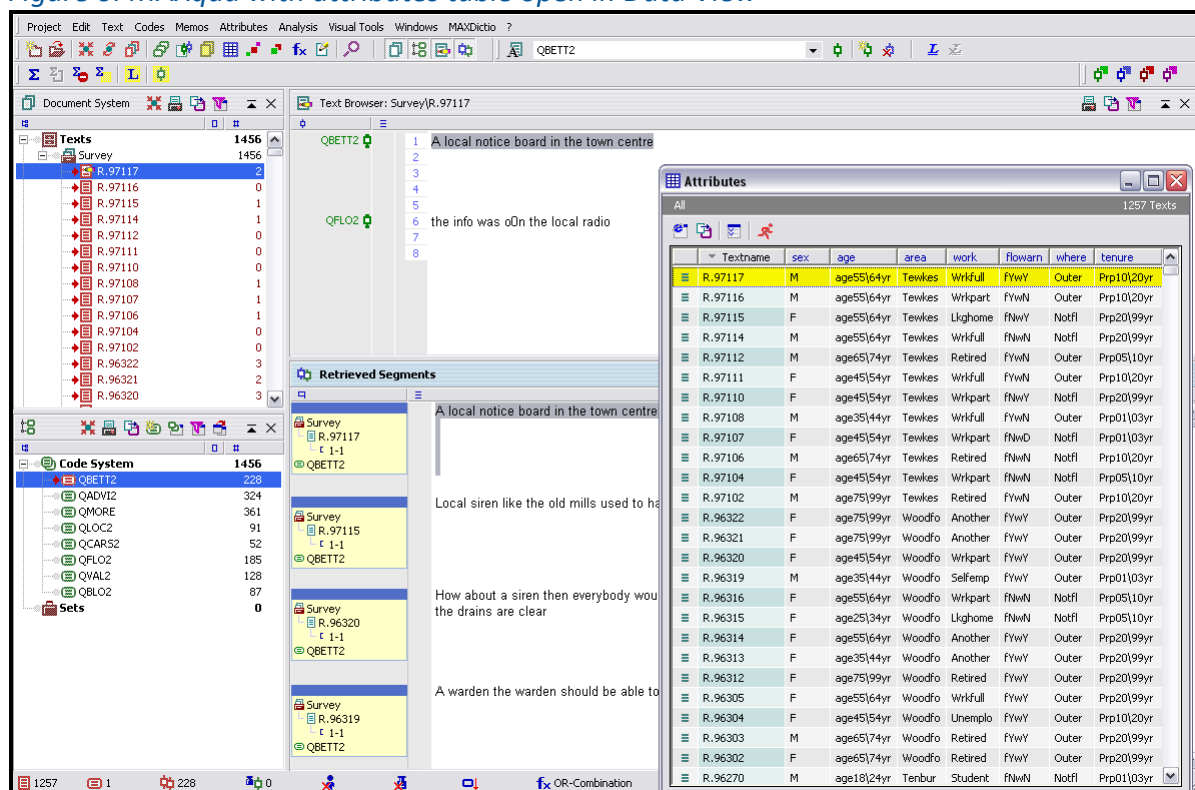
In Figure 6, note that all of the Survey text group documents have been activated. When activated, the texts are linked interactively to the data in the attributes window. Selecting a respondent in the document system (R.97117 in this illustration) causes the attributes for that respondent to be highlighted in yellow in the attributes window.

1.6   Verify the accuracy of the data transfers.

The first checks to confirm that the data has been imported correctly are those that confirm the total number of texts and respondents. In Figure 6 above, note that in the top right corner of the attributes window "1257 Texts" is displayed. This is confirmation of the number of respondents whose attributes have been imported. When you activate all of the texts in the group (by right clicking on the group header, "Survey", and selecting *Activate All Texts*) the number beside the first icon in the

extreme bottom left corner of the screen shows the total number of texts (i.e. respondents) in the group (also 1257 in this example). These are two indications of successful data imports.

Next, you should check that the longest response texts which you identified in the first step above have been copied successfully. Move the attributes window to one side so that you can see the Text Browser and Retrieved Segments panels. Deactivate all of the texts if they are currently activated, and activate all of the question codes in the Code System panel instead. Now, when you activate a single text you will be able to see all of that informant's responses in the Retrieved Segments panel. Use this method to call to screen the longest individual response texts and confirm that they are complete.

Finally, you may check the import of attributes by scrolling through the attributes window quickly, looking for obvious gaps, blanks, or out of pattern cells.

2.0   Analysing Open-Ended Survey Question Data in CAQDAS Packages: Initial Coding Approaches for MAXqda

There are, of course, many different ways to analyse responses to open-ended questions. This page is not a step-by-step guide on how to do analysis, it is rather a series of observations about how the features of MAXqda might interact with a particular type of dataset. This page should be read in the context of the related materials concerning the use of MAXqda for the analysis of open-ended survey question data, accessible from the main *Analysing Survey Data* page.

The tools discussed below are illustrated with examples from the same post flooding event survey that was used to illustrate the data preparation processes. For a summary of the project from which this data derives see here. This data is characterised by a fairly large number of short statements.

Outline:
2.1   Reading the texts – by respondent or by question?
2.2   Developing a coding scheme – manually or by using word frequency tools?
2.3   Text searching and autocoding.
2.4   Coding – data indexing versus data reduction.
2.5   Checking summarising codes – consistency and omissions.
2.6   Looking for similarities or differences?

Detailed Guidance:

2.1   Reading the texts – by respondent or by question?

Out of the four CAQDAS packages reviewed in this section, MAXqda offers the greatest flexibility over how to read the response texts on screen. The choice between reading all of the texts respondent by

respondent or reading all of the responses question by question has to be made at some stage, but in this program it is not determined by any earlier text preparation decisions and remains a free choice throughout the analysis phase. A discussion of the issues relating to this decision may be found here.

To read the texts respondent by respondent, open the first text by double-clicking on its name in the Document System panel (top left in Figure 7) and all of the responses made by that respondent will be displayed in the Text Browser panel (top right in Figure 7). To move on to the next respondent, double click on the next name in the Document System panel – the 'open' name is highlighted and shows a pencil icon superimposed on the text symbol.

To read all of the responses to one question it is necessary to activate all of the texts and also to activate the code for that question, the responses will then be displayed in the Retrieved Segments panel (bottom right in Figure 7). Activating all of the texts can be achieved by right clicking on the group header title in the Document System panel and selecting *Activate all texts* from the context menu. Activating the question code requires a right click on the question name in the Code System panel (bottom left in Figure 7) and selecting *Activate* from that context menu. A left click on the information box beside any response text in the Retrieved Segments panel highlights that text, brings all of the responses by that respondent into the Text Browser panel, and highlights that respondent's attributes if the Attributes window has been opened.

Figure 7: MAXqda display showing the four working panels
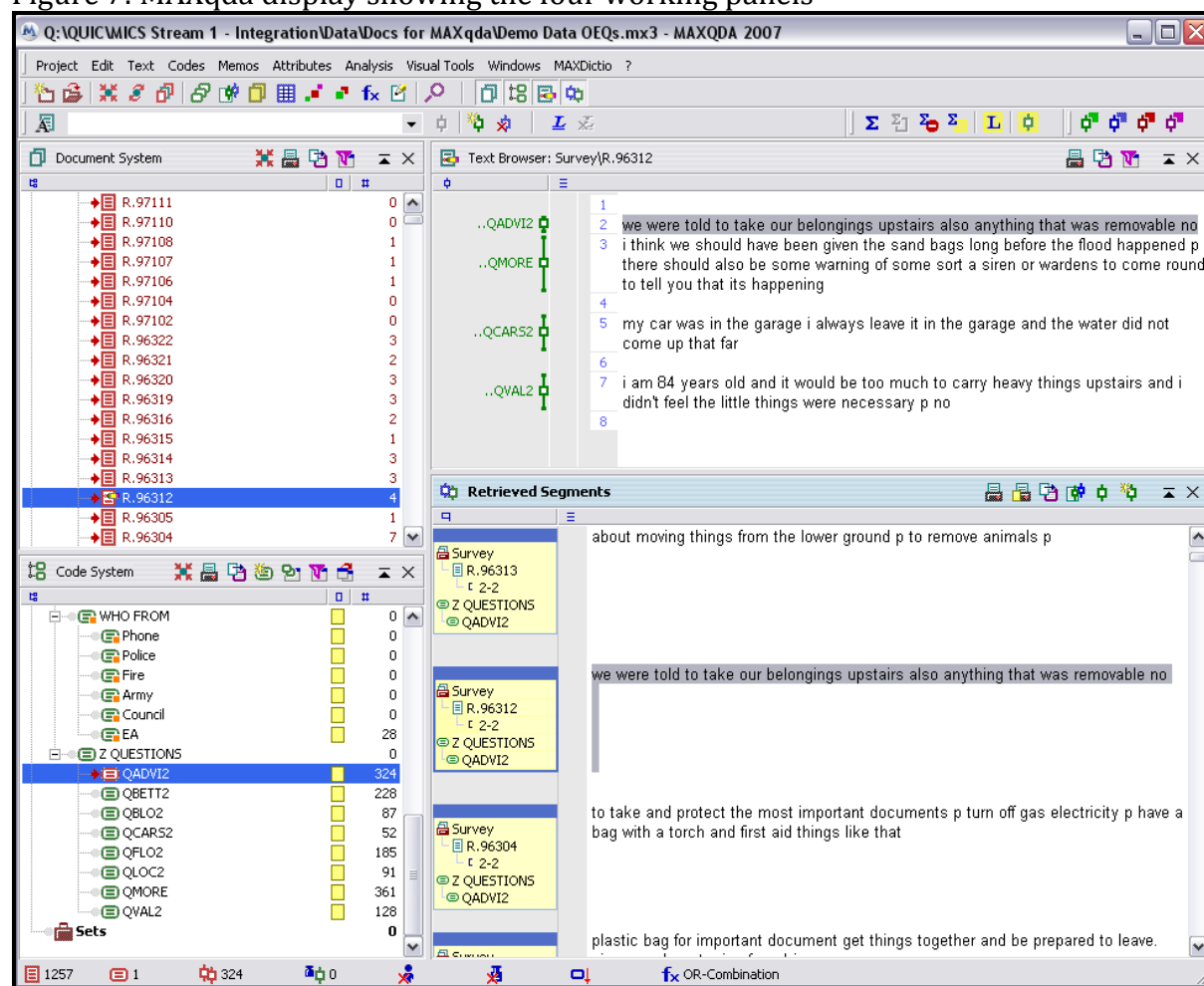
In Figure 7, all 1,257 texts have been activated (indicated by the red icons in the Document System panel), the code for question QADVI2 has been activated (red icon in the Code System panel), some of the 324 responses to QADVI2 are visible in the Retrieved Segments panel (which has scroll buttons to display the others), the response by R.96312 has been highlighted in the Retrieved Segments panel and this has opened all of that person's answers in the Text Browser panel (only four of the eight questions in this dataset were answered by this respondent).

2.2   Developing a coding scheme – manually or by using word frequency tools?

The nature of the analytic strategy will affect whether it is appropriate to develop a coding scheme manually or by using word frequency tools. If working deductively a coding scheme may be derived from, or informed by, existing literature or (theoretical) frameworks; in these situations the following comments will not really be relevant. If, on the other hand, you are working inductively and therefore intend to generate coding categories from the ideas mentioned in the response texts themselves then you have a choice as to how to proceed. You may work 'manually' by reading the texts and choosing categories that seem to be mentioned in those texts or alternatively let the software help by creating a list of the most frequently used words in the texts and allow code development to be informed by this list.

The manual, or maybe that could be termed "human", method will be required at some stage if really accurate coding is needed, because only human readers can detect all of the subtleties of human expression involving multiple ways of phrasing any particular idea. However to get started, particularly in a large dataset, it could be worth trying the word frequency method to get an early idea of the range and salience of words used. The most frequently used words may be expected to provide indications of the most frequently expressed concepts, although multiple possible meanings for some words can complicate this assumption.

For MAXqda the word frequency functions are held within the module "MAXDictio" which is an optional add-on to the program. The word "MAXDictio" should be visible in the top menu bar if this module has been installed with your version of the software. Before selecting the initial menu option it is important to set the correct activations so that the word count is run on the passages relevant to your current work. In the context of a survey it is likely that you will want to develop a separate set of codes for each question's responses (see section 1 above and further comments after Figure 10 below), so activate all of the texts and just one question's code as the first step. Then select the *MAXDictio* menu and the *Word Frequencies* option.

Figure 8: MAXDictio - Word Frequency options
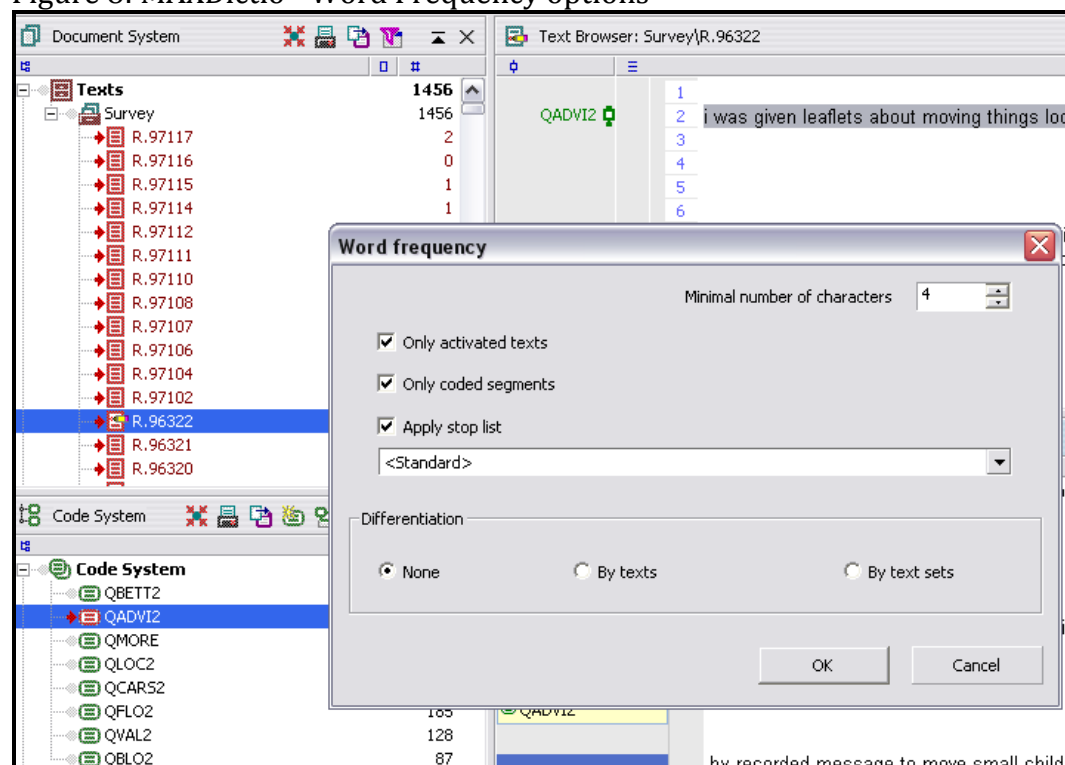
In Figure 8, above, the group of texts under the heading Survey have all been activated, along with the code for QADVI2, and text R.96322 has been opened. In the Word frequency dialogue box, placing ticks in the first two boxes, *Only activated texts* and *Only coded segments*, will restrict the count to the full batch of responses to question QADVI2. The *Minimal number of characters* field has been adjusted to length 4 in this illustration in order to reduce the number of trivial words included in the output list – this is a matter of judgement.

In Figure 9, below, the results of the procedure are illustrated. Confirmation that the words counted have been limited to those for the selected question is apparent in the number of Texts included – the table shows "In 324 texts …" and the highlighted code to the left shows that there are 324 segments in QADVI2.

Not all of the results are meaningful as indications of useful concepts, for example the fourth most frequently used word in this table is "from" and this does not seem to lend itself to a thematic code, so judgement will still be needed to select coding categories, but some useful information can be seen very quickly. This data is about advice on how to cope with the threat of being flooded, and so words like "upstairs, sandbags, valuables, furniture" can be picked out at once.

TIP: It is worth noting that care should be taken over the word length threshold; for this exercise it was set at 4 letters, if it had been set any higher the word "pack" would have been excluded, yet its 44 mentions here may be interesting.

Figure 9: MAXDictio - Word frequency output table
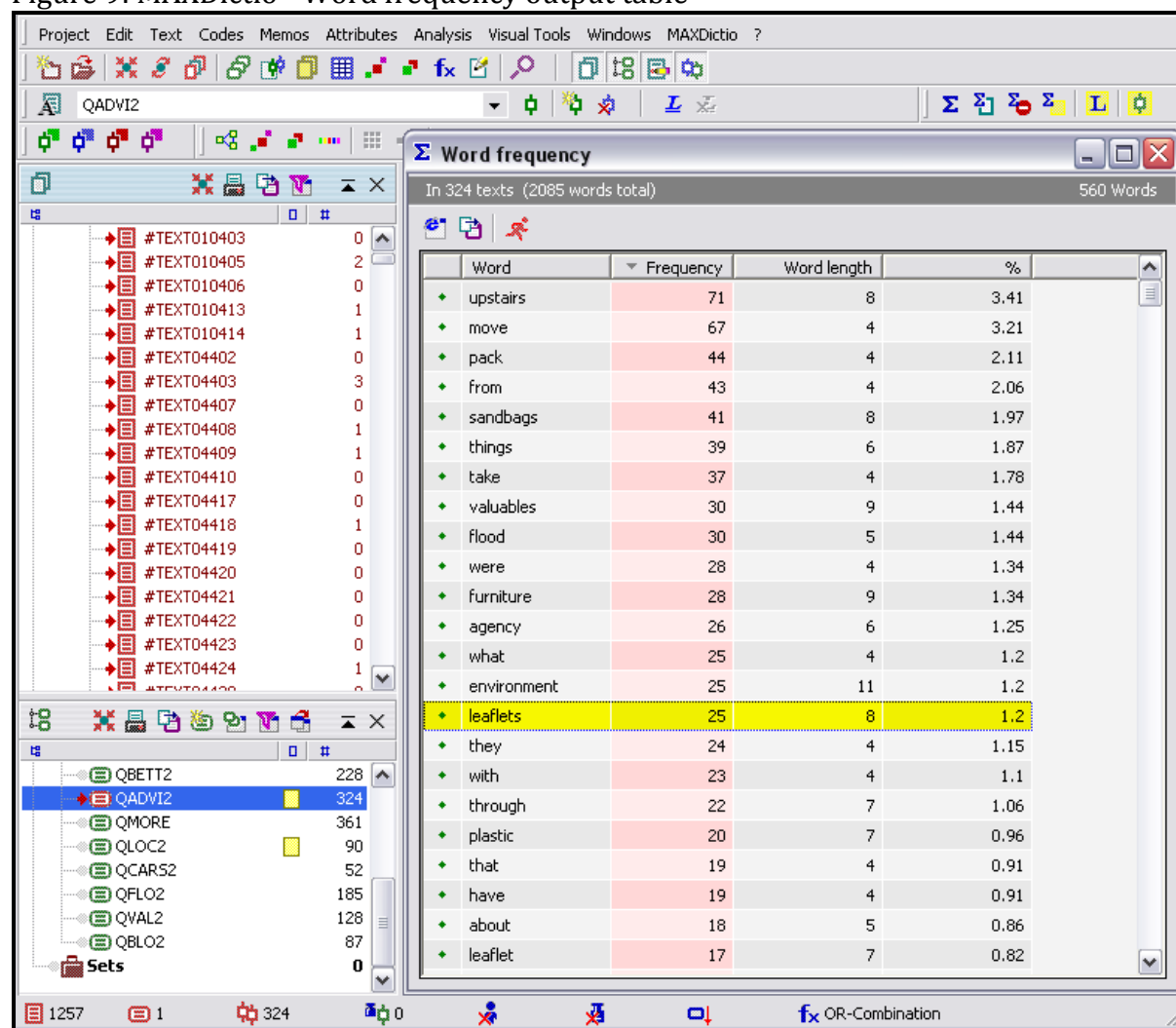
The data can be re-sorted according to any of the columns by clicking on the relevant column header. The example shown in Figure 9 has been sorted in descending order of frequency. It may be interesting to look at the longest words by sorting on the word length column for some complex ideas. An alphabetical sort may also reveal some minor misspellings, splitting a frequently cited word between two or more counts, for example in Figure 9 it can be seen that "leaflets" were mentioned 25 times but when that is combined with the 17 mentions of "leaflet" the total of 42 makes this an even more significant topic.

In some situations it may be useful to examine the contexts in which a frequently used word has been found in order to assess its suitability as a thematic code. A left-click on the required word in the Word Frequency table highlights it (as with "leaflets" in Figure 9 above), then a right-click opens a small context menu from which the option *Create Index* should be selected. This opens a second output window in which all occurrences of the required word are listed, identified by the text number within which they are located. The number of hits in this window may well exceed the frequency number in the other window because this index is not restricted to the single question code as before. However, by sorting the list on the paragraph column data the relevant occurrences will all be grouped together. A click on any row of this "Search Results" window brings that response up in the Text Browser window with the search word highlighted so that it can be read in its full context (see Figure 10).

TIP: There is no need to close this window when you want to check another frequently used word, simply right-click on the new word in the Word Frequency window, reselect Create Index and a new list will be generated in the Search Results window, already sorted by paragraph number (if you had used that previously). In this way it is quite easy to examine the detailed usage of keywords before you decide whether to use them as the basis of thematic codes or not.

Figure 10: Index of uses for a word selected from Word Frequency table, sorted by paragraph
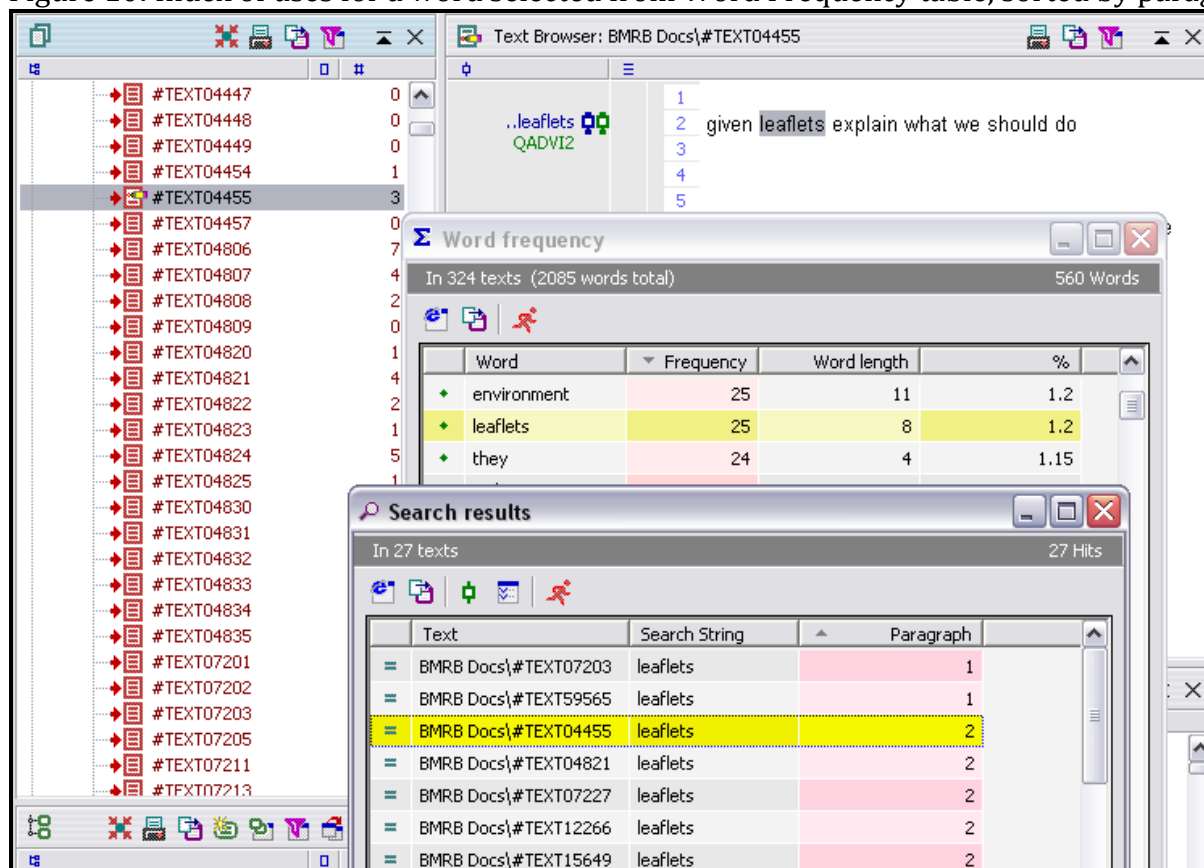
As you develop ideas for your coding scheme you should also consider an important aspect of its structure. Do you want to create a separate set of codes for each question in your data or have a common coding scheme across all questions? There is no clear solution to this problem for open-ended questionnaire data, as it depends on your own preferences and working practices, the nature of your data and your analytic approach. If the survey questions were largely unrelated to each other, perhaps because they were widely spaced apart at different points in the questionnaire, then a separate group of sub-codes for each question may be suitable. However if you expect to be interested in the way themes occur across multiple questions, then having a more unified coding structure may be more suitable.

2.3   Text searching and autocoding.

Although most qualitative analysts will naturally prefer to do all coding work manually it is quite reasonable in some circumstances to use MAXqda's autocoding process. For example quite a lot of responses may be simply "don't know" because the open-question has not triggered a specific

response. Coding such material individually would be tedious work but by running a *Lexical Search* for that phrase (probably first with and then without the apostrophe, or both together using more sophisticated search criteria) and then using the *Autocode* function, these can be allocated to a code quickly and efficiently. With practice more positive common concepts may be identified and also rapidly coded in this way. It should be easier to follow this procedure and then eliminate any incorrect codings, following a consistency check, than to code a large number of very similar statements manually.

In the example illustrated in Figure 11, below, the full set of response texts have been activated and the question code "QADVI2" has been activated, limiting the search to the responses to the single question but including all of the respondents. The menu option *Analysis / Lexical search* has been initiated and a new search string "environment" has been entered in the dialogue box (to locate references to "Environment Agency" in order to allocate these comments to the "EA" code). The checks in the options *Only in active texts* and *Only in retrieved segments* are necessary to limit the search to the responses just defined.

Figure 11: Basic Lexical Search to locate a common word or phrase
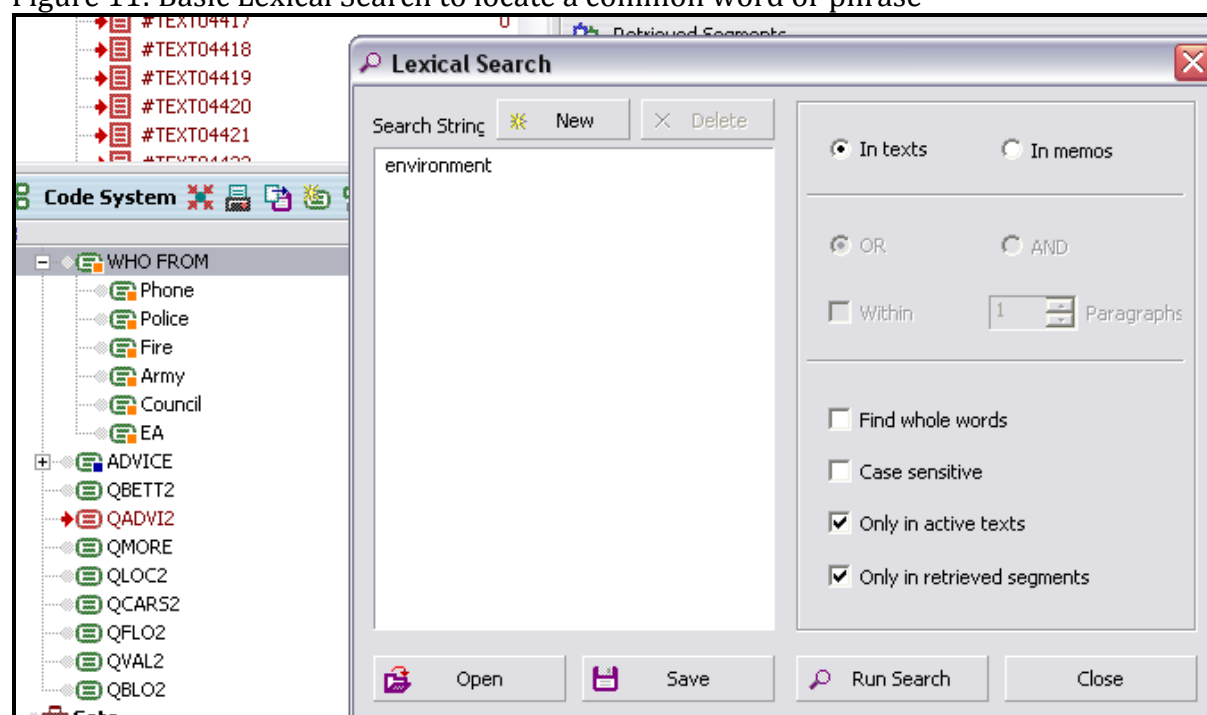

Click on Run Search to generate the output, which is illustrated in Figure 12.

The output shows that 26 texts were found in the activated group which contained the word "environment". By clicking on any line in the search results window, the full text for that respondent is brought up in the Text Browser panel where it can be checked for accuracy or relevance to the expected code. It is not essential to check all of the hits at this stage but it is worth making sure that your search has made sense by looking at a few of them.

Figure 12: Lexical Search output
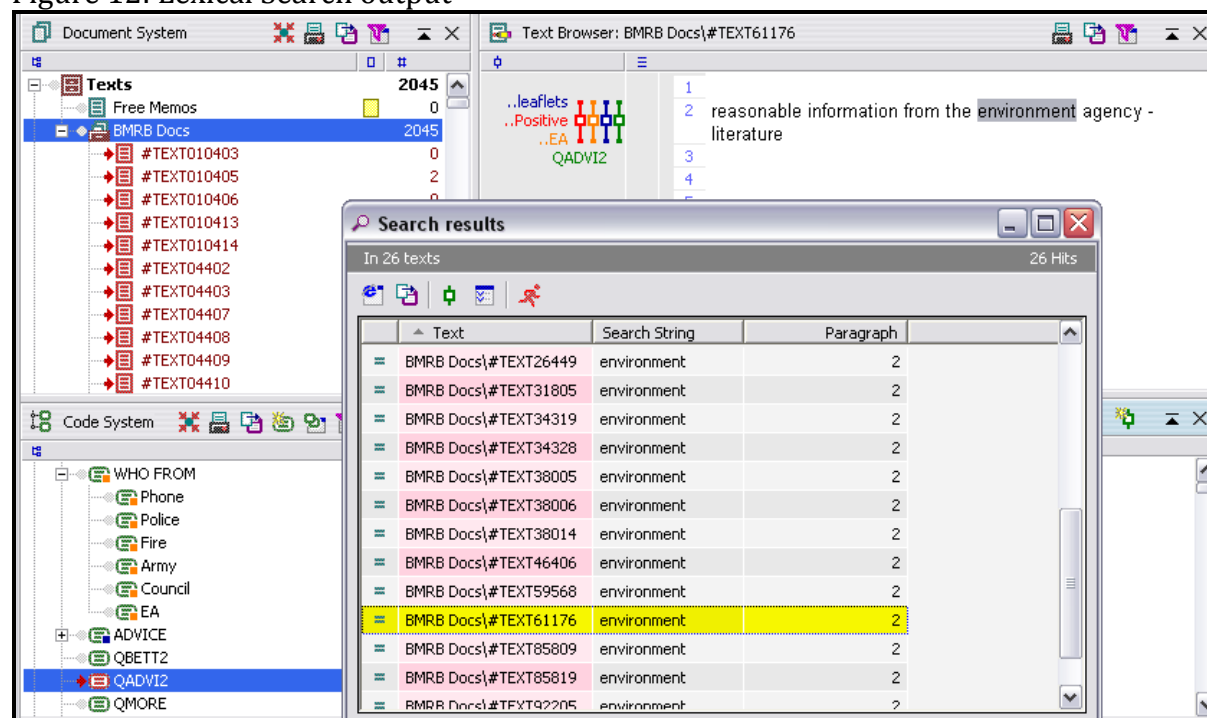
The next step will be to click on the green coding icon in the search results window to pull-up the autocode dialogue shown in Figure 13. However, before doing this it is worth clicking once on the code label (in the code System panel on the left) for the appropriate code (in this example "EA") as this will make it available in the autocode dialogue.

Figure 13: Autocode the results of a Lexical search
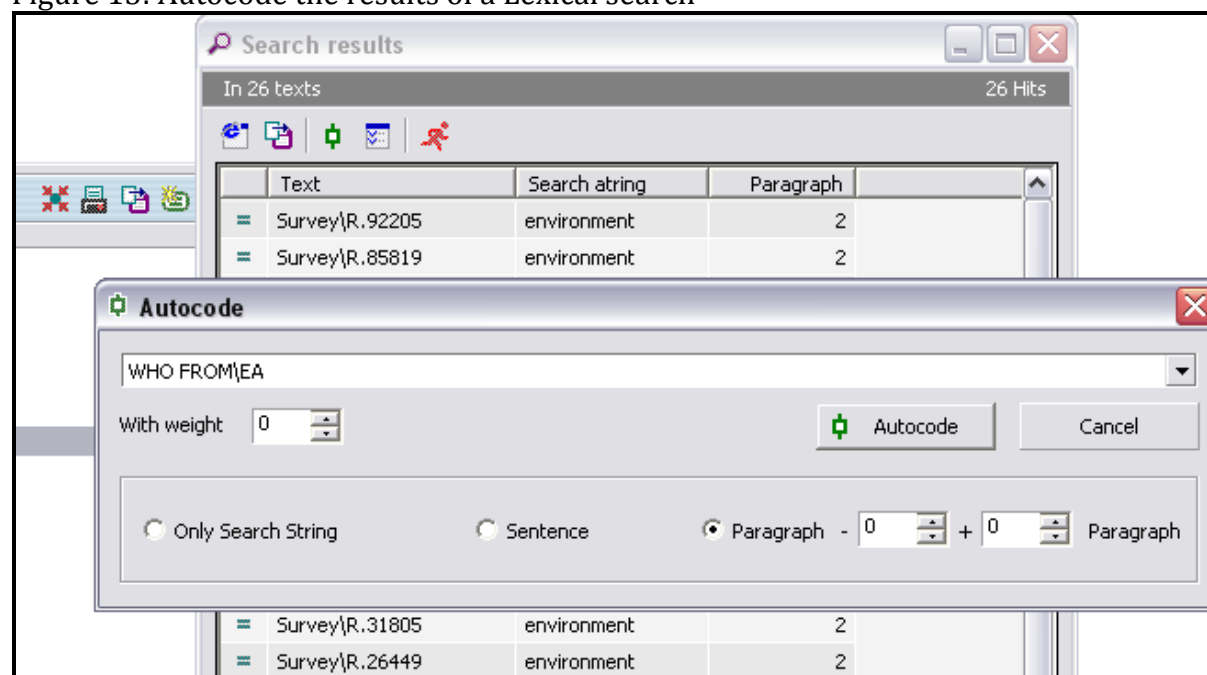
It was the click on the code "EA" before commencing the autocode which has placed that as the expected code for this dialogue. If the code you want to use is not listed in the drop-down menu it will

be necessary to cancel this command, click on the required code in the Code System panel (or create it if necessary) and then restart the autocode process. It is not necessary to close the Search Results window whilst doing this. For this example in the lower part of the dialogue we have chosen to autocode the whole paragraph containing the search word but shorter or longer coding options are also available. A click on the Autocode button now completes the sequence of operations resulting in 26 responses being coded at once.

TIP: The decision as to how much text should be included in the autocode should be considered carefully. In contrast to some other CAQDAS packages there is no penalty in terms of loss of a link to the respondent's ID when short passages are coded, because MAXqda will always include the relevant Text number in outputs and reports. For on-screen enquiries it will always be possible to view the attributes of the respondent who contributed any coded segment under review. But by extending the code to the paragraph, as indicated here, it will be possible to read the full context of the response within which a theme has been coded. On the other hand, if your data is especially rich and you may want to investigate whether one theme is generally mentioned before another, then it will be advisable to restrict the autocoding to more precisely determined segments.

Before moving on you should now check the accuracy of the coding just carried out. Reset the activations in the Code System window (a single icon click achieves this) then activate the code you have just used ("EA" in our example) to bring all of the items allocated to that code into the Retrieved Segments panel. You can now read each of the texts in turn to satisfy yourself that they have been correctly coded, and you have the option of removing the code from any responses where the automatic allocation was incorrect (for example in our illustration if the word "environment" had been used in some other sense than a reference to the Environment Agency).

It is unlikely that a single search will exhaust the potential autocoding for one code, and the process can be repeated with variations on the search theme. To continue this example we searched on the word "agency" and found two more relevant responses which had used that word but not "environment". Clearly, ideas for further lexical searches may be found in the Word frequency tables suggested earlier, and those tables also indicate the variations in spelling which should be included in the searches.

TIP: Some users find it productive to export the Word Frequency table to a spreadsheet (like MS Excel) in order to carry out more powerful sorting and grouping routines to identify related words, and this can be seen as a way of extending the rigour of the analysis.

However, it is not necessary to run a separate search and autocode procedure for each possible word. Figure 14, below, shows how several words or phrases can be combined in a single search.

In Figure 14 four separate strings are set up for the search, and they have been combined with the "OR" function so that a hit will be located if any one of these strings is found. The check box "Find whole words" has been ticked to catch references to the Environment Agency by its acronym while excluding the many irrelevant words in which the letters ea may be found together. However this setting may prevent this search from locating variations to the other search words (for example "environmental"), but this can be accommodated by using the '*' wildcard extension (which works in conjunction with the whole words setting).

When multiple strings have been used in this way the search results window shows which word or phrase has been located in each hit and indicates the number of texts which contain them without

double counting the occasions when two or more terms appear in the same text (such as "environment agency").

Figure 14: Complex Lexical Search to locate multiple words or phrases
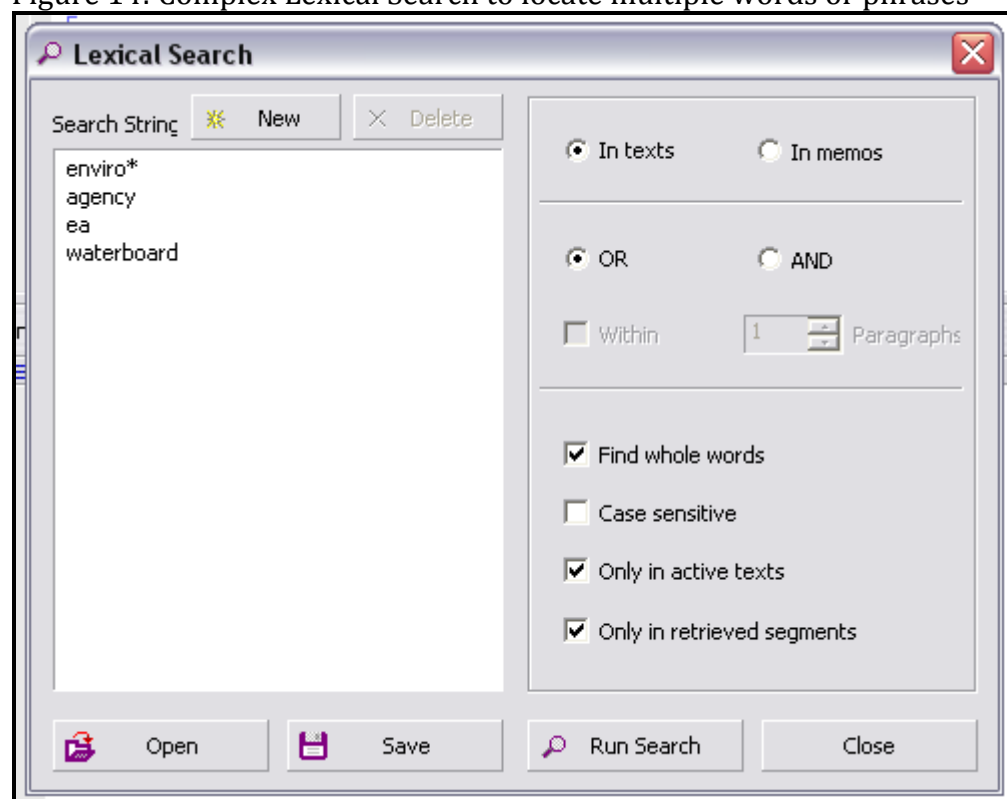
The autocoding process will not complete the task if data reduction to accurate quantities of references is the goal of the analysis (see section 4 below). Our searches above did not locate the responses which made reference to "floodline pack" or "rivers authority" which were eventually coded to the same "EA" code following careful reading of the full texts. So a combination of autocoding and human interpretation is needed to achieve a high level of accuracy. But time can undoubtedly be saved through the use of well-directed search and autocoding routines.

2.4   Coding – data indexing versus data reduction.

The actual techniques of manually applying codes to segments of text are not discussed here. They are common to all applications of the program and are clearly explained in MAXqda's help manual and in other sources. However, the possible uses to which the analysis of responses to open-ended survey questions may be put is a matter worth discussing further.

As a coding scheme is developed and applied to textual data, the analyst will inevitably encounter uncertainty and doubt. Does the text in front of me represent something different from others I have read before which mentioned a particular keyword? A common solution to this is to be generous and inclusive, applying specific codes to a range of comments that initially appear to be connected to those concepts, with the good intention of returning later and checking the work. This activity may be described as "data indexing" as it facilitates the retrieval of various passages that appear to relate to a particular topic.

When open-ended questions have been asked in survey situations it may be appropriate to generate numerical summaries of the data, probably in the form of statements of the type "X% of responses to this question mentioned Y". The obvious source of the numbers for this output is the coding of concept "Y". However the statement will only be valid if the use of that concept in every one of the responses allocated that code is consistent and equivalent, because the code that is used in this way has effectively replaced the words recorded for each respondent. The original textual data has been reduced to the code label.

When put this way it should be apparent that work needs to be done to refine the inclusive indexing codes before they can be safely used as summarising reducing codes. In this example data one respondent answered the question about the advice they had received by saying "... move self and belongings upstairs. Contact floodline. I think there were a few tags to put on things in the kitchen. The pack itself was very well done and well thought out ..." while another dismissed this as "... a flood pack stating obvious, silly stickers ...". Initial index coding may have allocated a code "Given flood pack" to both of these passages but it could be potentially misleading to include both in a percentage of respondents who referred to the flood packs as though these comments are equivalent to each other.

2.5   Checking summarising codes – consistency and omissions.

There are a variety of tools in MAXqda to assist with the refinement of codes when they have to be reduced to summarise what was originally said. Two particular aspects should be considered, firstly confirmation that all of the passages connected to any one code are all sufficiently similar to be treated as equivalent, and secondly confirmation that no other passages that are also equivalent have been omitted from that code.

The first step in confirming consistency or equivalence is to extract all of the passages that have been allocated to a code and read them carefully. In MAXqda a retrieval of this kind is achieved by activating all of the texts and just the specific code that is being checked. If the code has only been used in connection with a single question then this retrieval can be achieved using mouse clicks in the Code System panel. However, if the code has been used with more than one question and the requirement is to confirm its consistent use within a single question's responses then a more complex *Text retrieval* operation is necessary. First activate all of the texts, the specific code to be checked, and the code for the question to which the check is to be limited. Then use the *Analysis / Text Retrieval* option (or the fx button on the main toolbar) and change the function if necessary to *Intersection* by using the drop-down menu. Next click on the *All activated codes* button to bring the required codes into window A of the dialogue box. The Execute command should then put these operations into effect and in the Retrieved Segments panel display just the passages from the selected question that have been coded to the selected code.

When checking all of the passages linked to a single code it should be possible to write a concise definition of that code. MAXqda provides a memo function for each code and this is the ideal location to store that definition. A right click on the code label in the Code System panel brings up a context menu, select the command *Code Memo* from this menu, and type the definition into the large text space in the lower part of the dialogue box. If the code has been applied in part by using autocoding routines then the specific words used in those routines should be listed here. When this has been done a yellow

box is added to the line for that code in the Code System panel and subsequently by moving the mouse pointer over this box you can bring up a temporary display of the code definition.

TIP: If you find it difficult to write a concise definition of a code then it may be inferred that you should not refer to the number of references to that code in any data reducing statements.

It is more difficult to search for code omissions, these are passages which are closely equivalent to those already allocated to a particular code but which have not yet been allocated themselves. One possible method for doing this is to filter-out all of the segments which have been allocated the code and then identify passages in the remaining data which may be equivalent.

TIP: The way this is done probably varies according to the way in which you have coded initially. If you have used autocoding routines to apply the code it would be good practice to read the remaining passages as with manual coding, but if you have initially applied codes manually then it may be appropriate to use text searches using key words connected to that code on the remaining text passages to confirm the accuracy of those procedures.

To filter-out the segments which have been allocated to a particular code it is necessary to use the fx function a little differently. We will use the example of the "EA" code described separately in the autocoding section of this page (section 3); to filter-out the responses already coded to EA, activate all of the texts, the question QADVI2 and the code EA, then select the Text retrieval function (fx) to bring up a dialogue box similar to that in Figure 15. Change the function in the top part of the dialogue to *If Outside*, select *All activated codes* to bring both the question and the thematic code into Window A, and then highlight the thematic code and click on the Remove button to take it out again. In Window B click on the ... button and select the thematic code in that list. The dialogue should now look something like Figure 15 below. Note in this illustration how the Result bar shows "Segments found: 296", and this is the difference between the 324 segments in QADVI2 and the 28 segments coded to EA (both visible in the Code System panel in Figure 15). Click on the *Execute* button to apply the filter.

Figure 15: Filter-out responses to QADVI2 already coded to EA
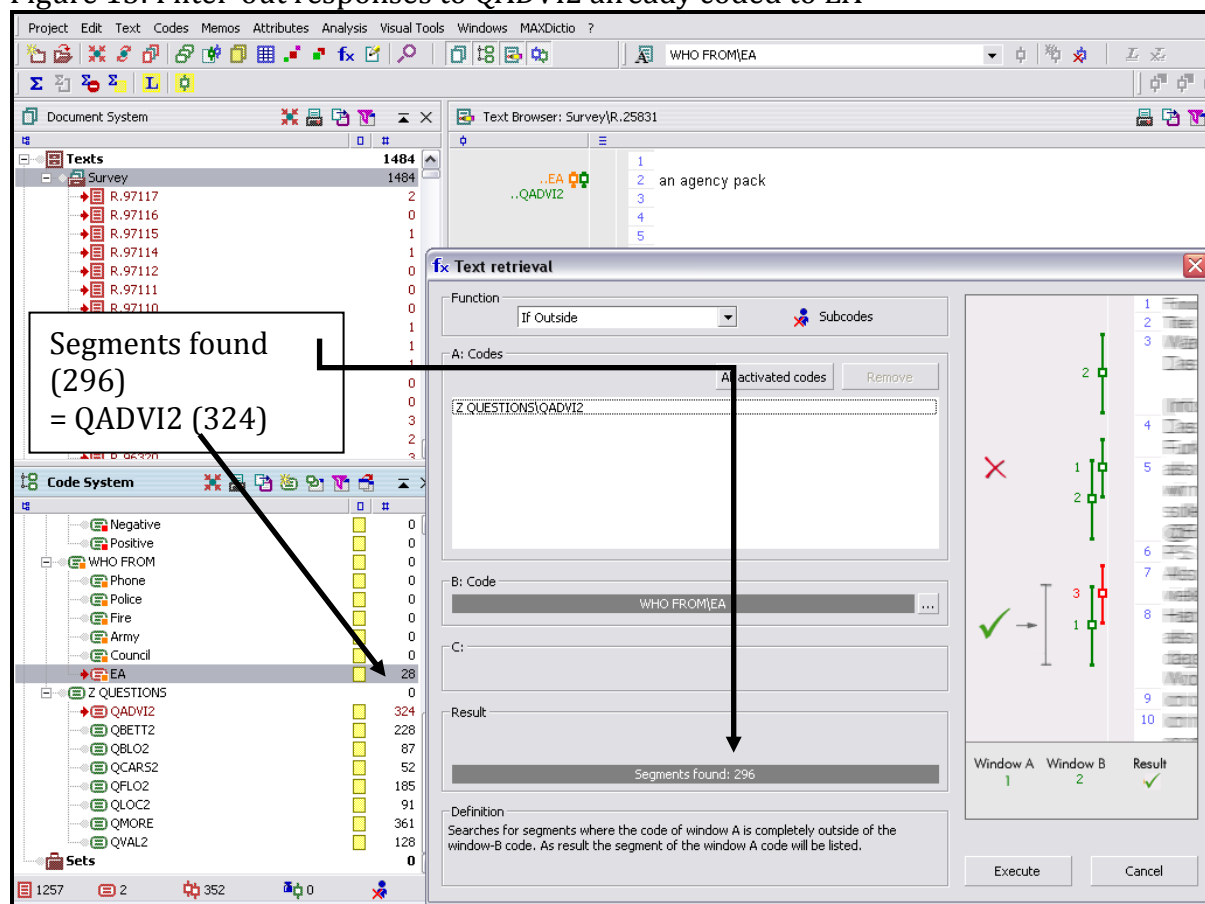
With the coded segments filtered out, it is possible to run repeated Lexical searches on the remaining retrieved segments for key words that may be associated with the thematic code being checked (see instructions in section 3 above for Lexical searching). Zero returns, or no "hits", are satisfactory as they confirm that no texts including the search term have been found that are not already coded to that theme.

TIP: It may be even quicker to run a Word Frequency calculation on the remaining retrieved segments and review that for key words, again the absence of these words in the results would be satisfactory.

2.6   Looking for similarities or differences?

When analysing the responses to open ended survey questions it may well be easy to slip into the expectation that the most frequently used codes, or rather the concepts to which they refer, are the most significant. After all, these are the items that seem to have the most statistical importance. However, it should always be worth looking out for contributions which are different from the common ideas. One-off comments will never feature in the quantitative tables because, by definition, they lack numerical support. But a small number of individuals may well take the opportunity of an open-ended question to add an unexpected thought and these contributions represent a challenge and an opportunity for the analyst.

Perhaps it is worth asking yourself what was the purpose behind the inclusion of an open ended question in the survey. In many situations previous research will have revealed the most likely

answers and these will have been included as response categories in closed questions asked elsewhere in the survey, but then an open question has been included to pick up other ideas. In these situations it is the unusual answers which may be of most interest.

For example in the data used as an example for these instructions the question QADVI2 was used to ask respondents what advice they had been given in order to prepare themselves for an impending flood about which they had been warned. It may be interesting to note that out of 324 responses to this question, just three people mentioned warm clothing or blankets. Now it may be the case that for most people the need for warm clothing as you sit out a flood in an upstairs room is too obvious to be worth mentioning, but this may also be a clue that there was a potentially significant gap in the advice actually given to the flood victims in the incidents under consideration. It seems that the value of a detailed qualitative analysis of the responses to such a question is an opportunity to pick up the unexpected ideas which would be so easily overlooked in a statistical analysis.

In conclusion:

There are several step-by-step instructions in the above materials. These have been included to help those who are not familiar with certain aspects of the way MAXqda works. However this is not intended to imply that these are the correct/only/best ways of analysing the responses to open-ended survey questions in MAXqda. These are merely examples of procedures that do work, particularly with data of the sort shown in the examples. However it will always be the case that different data may require different procedures, but we hope that these examples will help some analysts to get over the problem of using unfamiliar software or of using familiar software in an unfamiliar way.

3.0   Quantitative Analysis Strategies for analysing Open-ended Survey Questions in MAXqda

In common with other pages in this section of the website, this page is a series of observations about how the features of MAXqda might interact with a particular sort of dataset. This page should be read in the context of the related materials concerning the use of MAXqda with Open-ended Survey Questions, in particular the Data Preparation Instructions and the Qualitative Analysis Strategies, since the quantitative strategies outlined below can only be effected after the data have been imported and coded systematically in a MAXqda project.

The tools discussed below are illustrated with examples from the same post flooding event survey that was used to illustrate the data preparation processes. For a summary of the project from which this data derives see here. This data is characterised by a fairly large number of short statements.

Outline:

3.1   Summary of codes applied to a single question
3.2   Activation by Variable
3.3   Crosstabs
3.4   Quote Matrix
3.5   Code Relations Browser
3.6   Document Comparison Chart
3.7   Summary conclusion

Detailed Guidance:

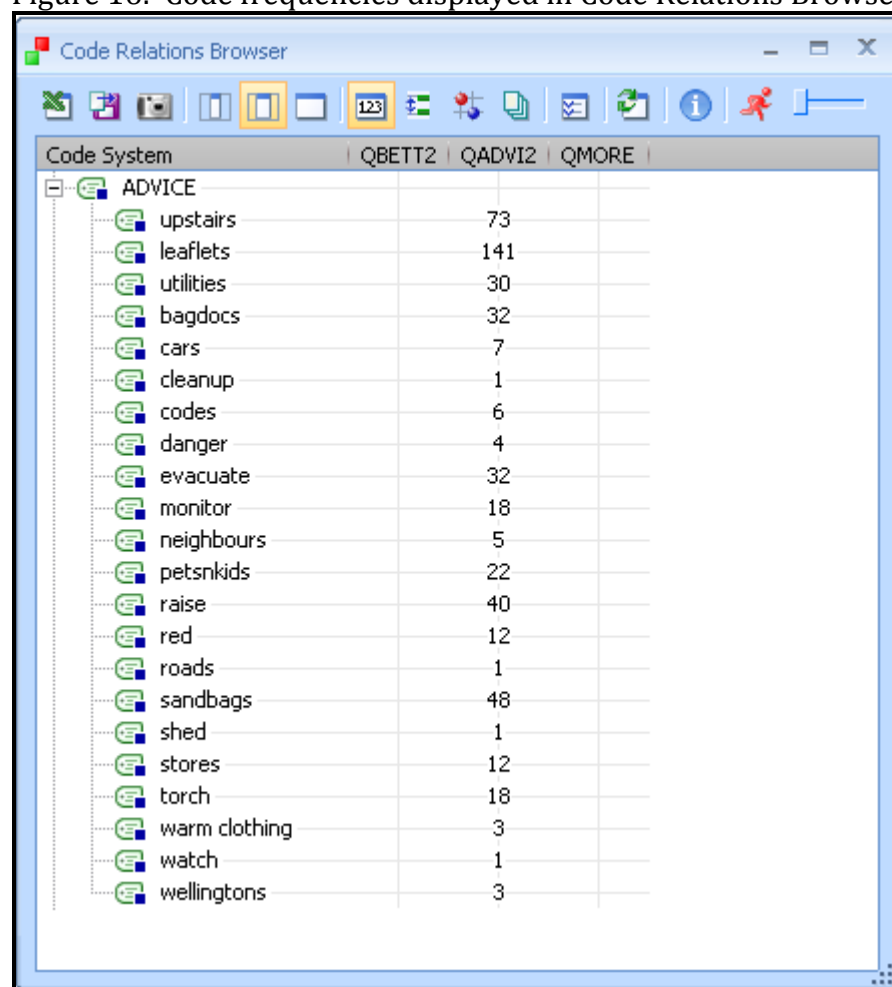### 3.1   Summary of codes applied to a single question

In many situations the most important output required from the analysis of some open-ended survey question responses will be a simple summary of the number of times each thematic code has been applied to the set of responses to a single question. In MAXqda this can be generated with a particular application of the Code Relations Browser (which will be discussed in more detail in section 5 below).

If you have created a separate code group for each question and been careful to avoid using any codes with more than one question, then you could simply read the code frequencies off the Code System panel on the main working screen by scrolling to the relevant group of codes. However, if you would like a printed summary, or if you would prefer to confirm the separation of codes for each question, then the following notes should be helpful.

First activate all of the response documents in the Document System (control + click on the group header does this effectively). Then activate the group(s) of thematic codes whose frequencies you wish to report (again control + click on the code group header). Now select the menu option *Visual tools / Code Relations Browser* (or the appropriate icon on the visual tools toolbar). A small dialog box asks whether this is "Only for activated documents and codes?" and you should click on the "Yes" box. The next dialog box asks you to select the codes on the x-axis in the report, and this is where you select the code(s) for the question(s) that you wish to report on. In this example we selected three question

codes "QBETT2", "QADVI2", & "QMORE". On clicking "OK" the program runs the calculations and generates the output table as shown in Figure 16 below.

Figure 16: Code frequencies displayed in Code Relations Browser

In Figure 16, above, just one group of thematic codes was activated prior to running the Code Relations Browser, the "ADVICE" group of codes and, as can be seen, these codes were only used with the question "QADVI2" as the other question columns are blank. This table can be saved, copied or exported to a spreadsheet, and it provides the basic frequency data for interpreting the questions responses.

3.2   Activation by Variable

A fundamental concept in MAXqda is the idea of "activation". Several important functions in the program work directly with this concept. When a number of documents and one code have been "activated", then the Retrieved Segments panel will display all of the segments that are in both the activated documents and the activated code. When more than one code is activated then the retrieval of segments will also be governed by the logical operator specified in the Text Retrieval dialog box (the default is "OR" meaning that a segment will be displayed if it has been coded to any of the activated codes).

The activation concept is very powerful when it is combined with the variables (MAXqda 2007 called these "attributes") using the function "Activation by variable". This option can be found in the Mixed Methods menu, where a shortcut icon is displayed (the fifth icon on the Standard Toolbar, highlighted in Figure 17 below). In the dialog screen that opens (see Figure 17 below) click on the "Reset" button to remove any previous activation formula that may be showing, and then on the "New" button to commence creating a new activation formula.

There are three elements at least that make up an activation formula, a variable, a comparison function, and a value of the selected variable. When you click on the "new" button another dialog screen opens showing a list of all the variables in the project, both system and user defined, with a check box beside each for marking your selection. In this example the variable "work" was marked, followed by the "OK" button in that dialog screen. Then, in the dialog shown in Figure 17, the comparison function "=" was selected with its radio button (it is the default setting) and the value "wrkfull" was selected using the drop-down menu in the lower right corner of the dialog. The operation is completed when the "Activate" button in this dialog is clicked.

Figure 17 – Activation by variable dialog screen in MAXqda
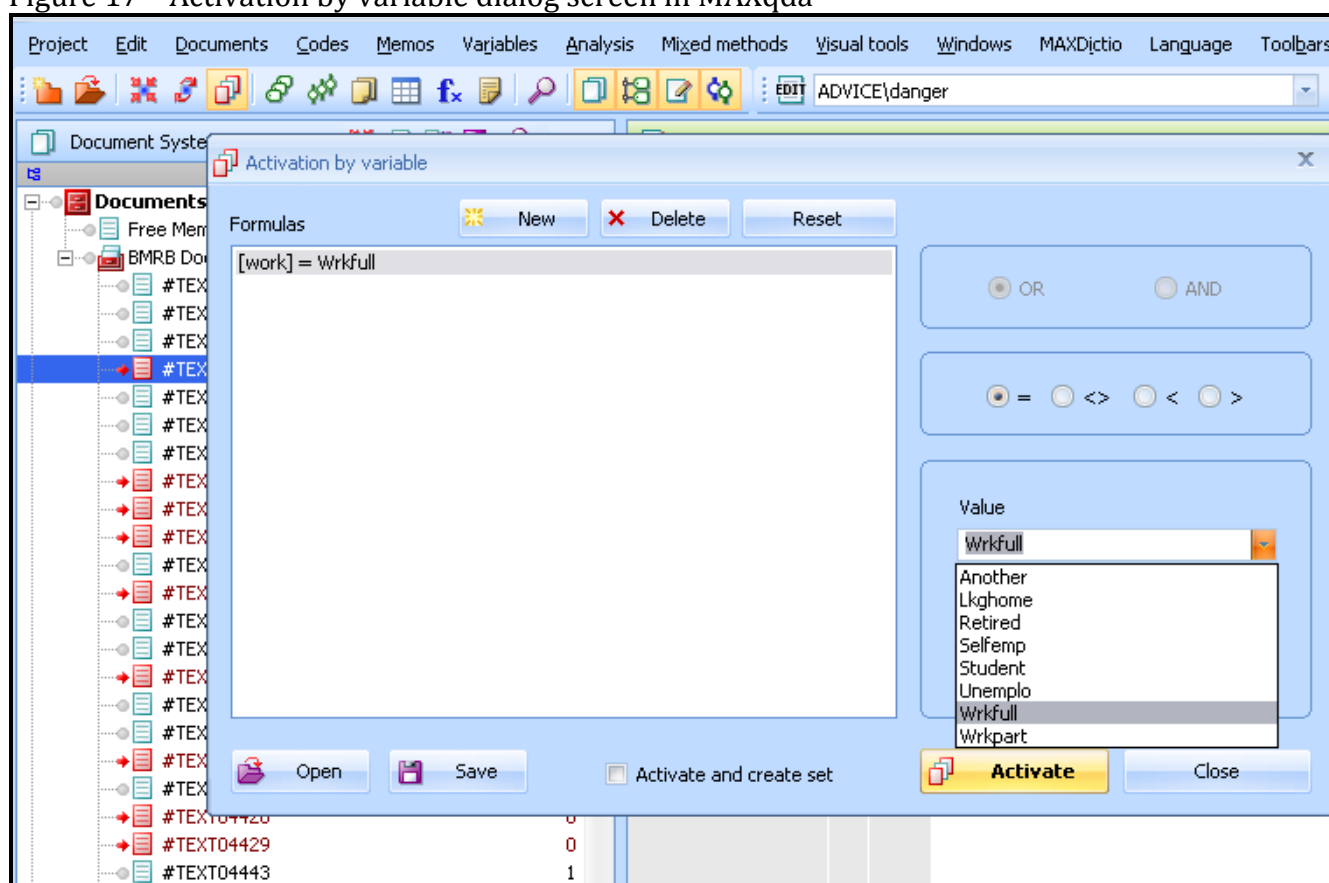
As can be seen in the Document System panel on the left in Figure 17, some of the documents have been activated (showing a red icon with a red arrow) while others have not (showing a blue icon and a grey arrow). Complex combinations of parameters can be set when the Boolean operators ("OR", "AND") are used, the choice between these is greyed-out until a second line of the formula has been created with the "New" button and a variable for that line has been selected.

It will soon be found that expressions such as "Males, aged 50 or over, who are working full time" can be generated quite easily. Combinations involving both the "AND" operator and the "OR" operator may need more care because there is no bracketing function to remove the ambiguity that can arise in those situations. The "Save" and "Open" buttons which can also be seen in Figure 17 make it possible to store and reuse an activation formula, this is only likely to be worthwhile with a complex formula which is likely to be reused, as the basic procedure is quite straightforward.

TIP: Note that MAXqda sometimes has difficulty resetting activations when these are being changed frequently. To be sure of obtaining accurate results it may be best to use the global "Reset activations" button which is the third from the left on the standard toolbar (showing in Figure 17 above) between different activation operations.

Having activated a particular group of documents in this way, successive activation of particular codes can be used to read the responses or, by using the count of retrieved segments shown in the bottom bar of the MAXqda window, to note the frequencies with which those codes have been applied to this set of activated documents. However, this is principally a qualitative activity with numerical data appearing as a by-product.
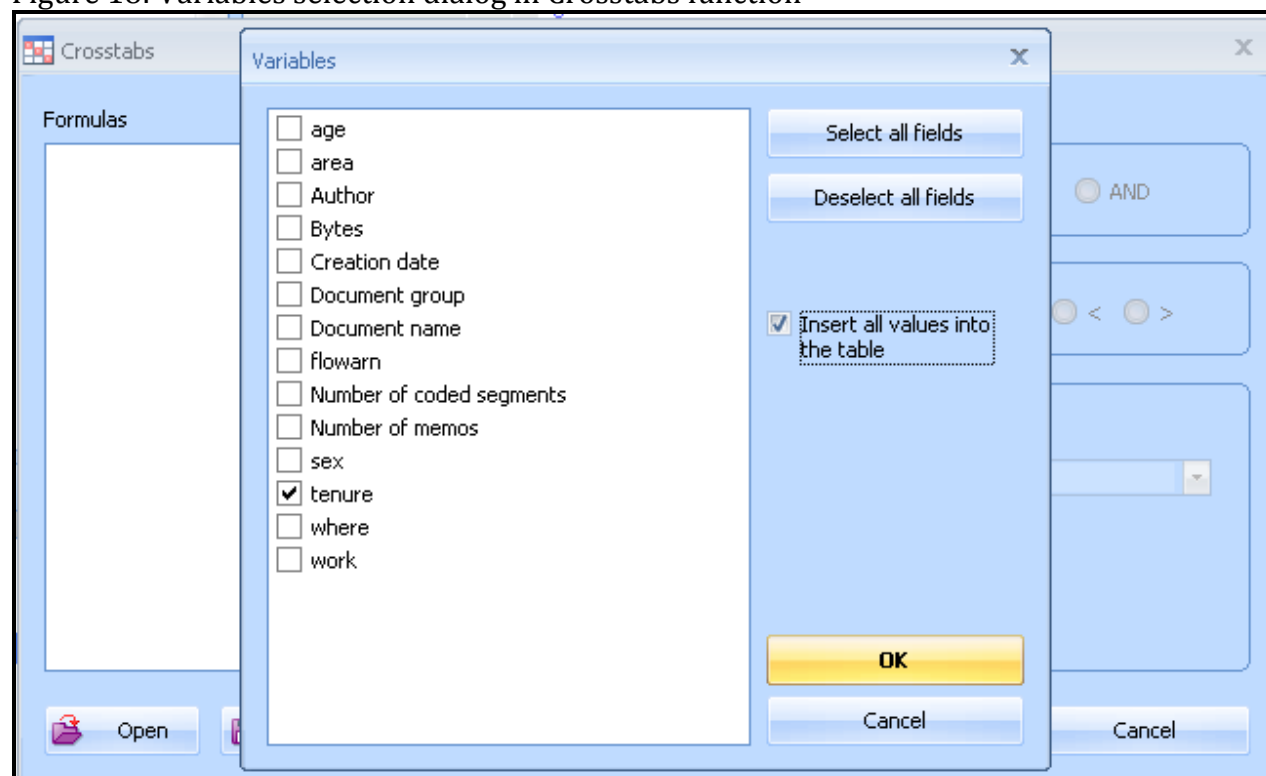
3.3   Crosstabs

A much more quantitative output can be generated using the "Crosstabs" option from either the Mixed methods or the Visual tools menu, or by using its icon from the Visual tools toolbar. None of these options will work (they are greyed-out) until at least some documents and some codes have been activated, so activation is the first step.

In general Crosstabs will make most sense if all of the documents are activated, as the variable values will be used in columns within the output table, and so it is not advisable to use the Activate by variable option at this point. Similarly, the rows of the table will show values for all of the activated codes, and so it will probably be useful to activate whole groups of thematic codes. So you should use the initial activation settings to define the population of documents and codes that will make up the table.

With an appropriate selection of activated documents and codes, then select the Crosstabs option. The dialog screen that opens is similar to the one for Activation by variable shown in Figure 17, but it will be used in a different way. The dialog settings are only concerned with specifying the columns of the table in terms of the variables, their values, and the sequence in which they will appear. The codes are fixed by the activation settings. Start by clicking on the "Reset" button to ensure a fresh set of choices, and then click on the "New" button to bring up the supplementary dialog shown in Figure 18.

Figure 18: Variables selection dialog in Crosstabs function

In Figure 18 the supplementary dialog is called "Variables" and it overlays the "Crosstabs" dialog which currently has no formula showing. Select the variable of interest, here it is "tenure" (the length of time the respondent has lived in the home potentially vulnerable to flooding) which has been ticked. In most instances it will also be useful to put a tick in the box to the right "Insert all values into the table", as shown here, this will ensure that all of the possible values taken by the "tenure" variable are included in the output. Click "OK" to return to the main crosstabs dialog screen.
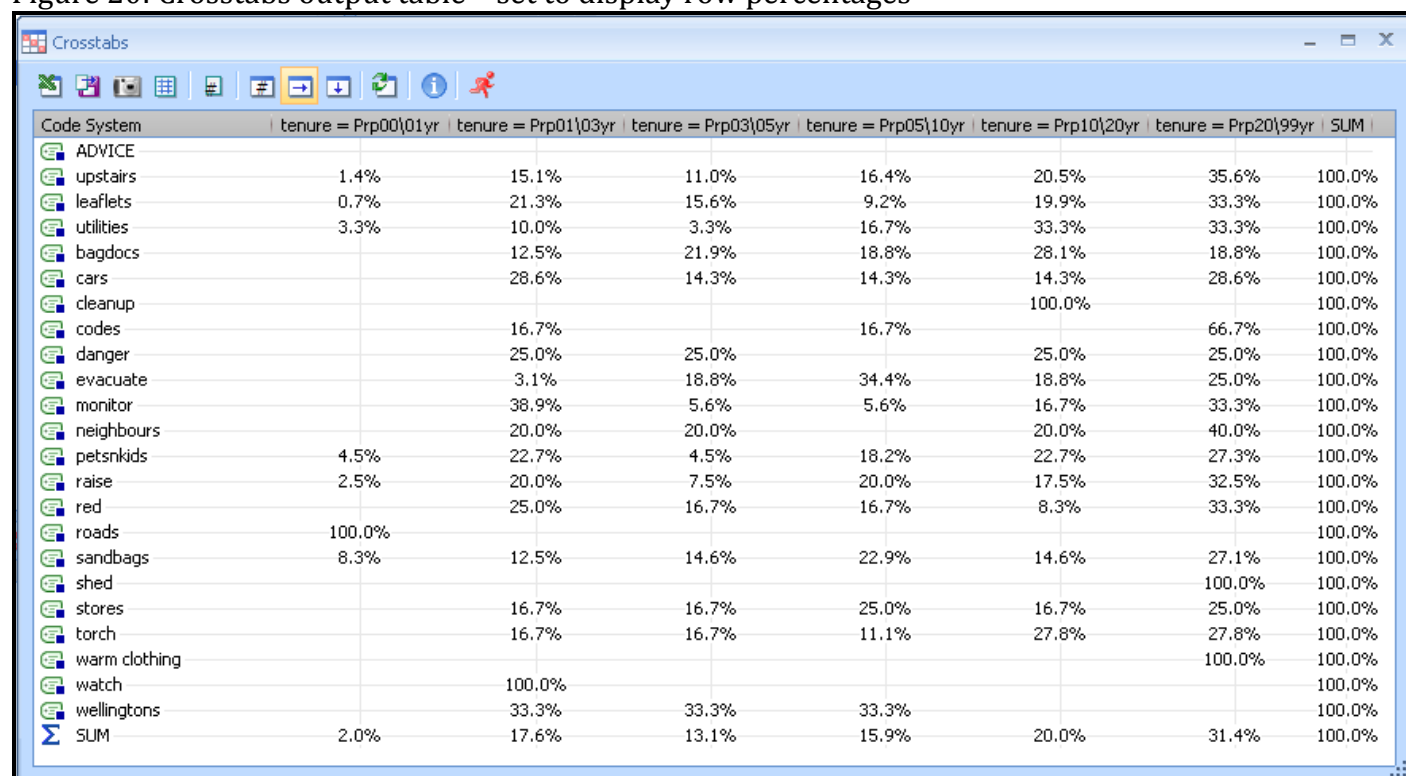
Figure 19: Crosstabs dialog after selection of one variable

Turning to Figure 19 above, it can be seen that the formula now includes the six separate values of the Tenure variable. No special operators are required for this form of output.

TIP: As the six values were originally created in ascending order by length of time, the sequence shown will make sense when it is converted to columns in the table. However, with some variables this may not be so straightforward. If you need to vary the sequence you can use manual commands to create the rows in this formula one at a time, selecting the values from the drop down list on the right instead of using the "Insert all values" option shown in Figure 18.

After clicking on the "OK" button within this dialog you will be asked to confirm that the procedure should be run only on the activated documents. The calculation may take a few moments, depending on the volume of data and the capacity of your computer. The result is displayed in a new window and may look something like Figure 20, below.

Figure 20: Crosstabs output table – set to display row percentages

Code System	tenure = Prp00\01yr	tenure = Prp01\03yr	tenure = Prp03\05yr	tenure = Prp05\10yr	tenure = Prp10\20yr	tenure = Prp20\99yr	SUM
ADVICE							
upstairs	1.4%	15.1%	11.0%	16.4%	20.5%	35.6%	100.0%
leaflets	0.7%	21.3%	15.6%	9.2%	19.9%	33.3%	100.0%
utilities	3.3%	10.0%	3.3%	16.7%	33.3%	33.3%	100.0%
bagdocs		12.5%	21.9%	18.8%	28.1%	18.8%	100.0%
cars		28.6%	14.3%	14.3%	14.3%	28.6%	100.0%
cleanup					100.0%		100.0%
codes		16.7%		16.7%		66.7%	100.0%
danger		25.0%	25.0%		25.0%	25.0%	100.0%
evacuate		3.1%	18.8%	34.4%	18.8%	25.0%	100.0%
monitor		38.9%	5.6%	5.6%	16.7%	33.3%	100.0%
neighbours		20.0%	20.0%		20.0%	40.0%	100.0%
petsnkids	4.5%	22.7%	4.5%	18.2%	22.7%	27.3%	100.0%
raise	2.5%	20.0%	7.5%	20.0%	17.5%	32.5%	100.0%
red		25.0%	16.7%	16.7%	8.3%	33.3%	100.0%
roads	100.0%						100.0%
sandbags	8.3%	12.5%	14.6%	22.9%	14.6%	27.1%	100.0%
shed						100.0%	100.0%
stores		16.7%	16.7%	25.0%	16.7%	25.0%	100.0%
torch		16.7%	16.7%	11.1%	27.8%	27.8%	100.0%
warm clothing						100.0%	100.0%
watch		100.0%					100.0%
wellingtons		33.3%	33.3%	33.3%			100.0%
SUM	2.0%	17.6%	13.1%	15.9%	20.0%	31.4%	100.0%

In Figure 20 the basic layout of the table is apparent with the selected set of codes appearing in rows and the variable values appearing as column headings in the sequence that they were listed in Figure 19. At this stage the precise content of the table cells has some flexibility. A group of three icons on the toolbar within this window provide alternative displays of document counts (the hash icon), row percentages (the right arrow icon, here highlighted), or column percentages (the down arrow icon). Note that the table also includes both row and column totals. Another icon with a hash symbol (5th from the left in Figure 20) works as a toggle, switching the display between counts of documents containing coded segments and counts of the coded segments themselves – this only makes a difference if you have applied some codes more than once in single documents (which is less likely with this sort of data than in other circumstances). All of these changes work with the current display without rebuilding it from scratch, so it is possible to switch between them quickly.

TIP: At the time of writing we observed some unexpected results in terms of the sequence in which the codes are shown in this table. It is possible to change the sequence of codes in the Code System panel by dragging selected codes up or down the list, or to sort them alphabetically with a context menu option applied to a group header. However, when the Crosstabs table is generated, the program appears to use the sequence of codes as it stood when the data was loaded at the start of the current session, and it ignores any re-ordering that may have been done during that session. So, if you want to achieve an output with a specific sequence for the rows of codes, try adjusting the Code System to show that sequence, then close the program and reopen it before running the Crosstabs.

TIP: Having said, above, that Crosstabs makes sense generally when all documents are included, it should be noted that more sophisticated analysis can be achieved by combining the Activation by variable option with Crosstabs. In order to examine whether there is different variation between men and women over how their length of tenure at the home in the flood risk area affects their views it may be necessary to activate all the documents from interviews with men and then run the Crosstabs

as shown above just on those activated documents, before repeating the process whilst the interviews with women have been activated. The two outputs can then be compared.
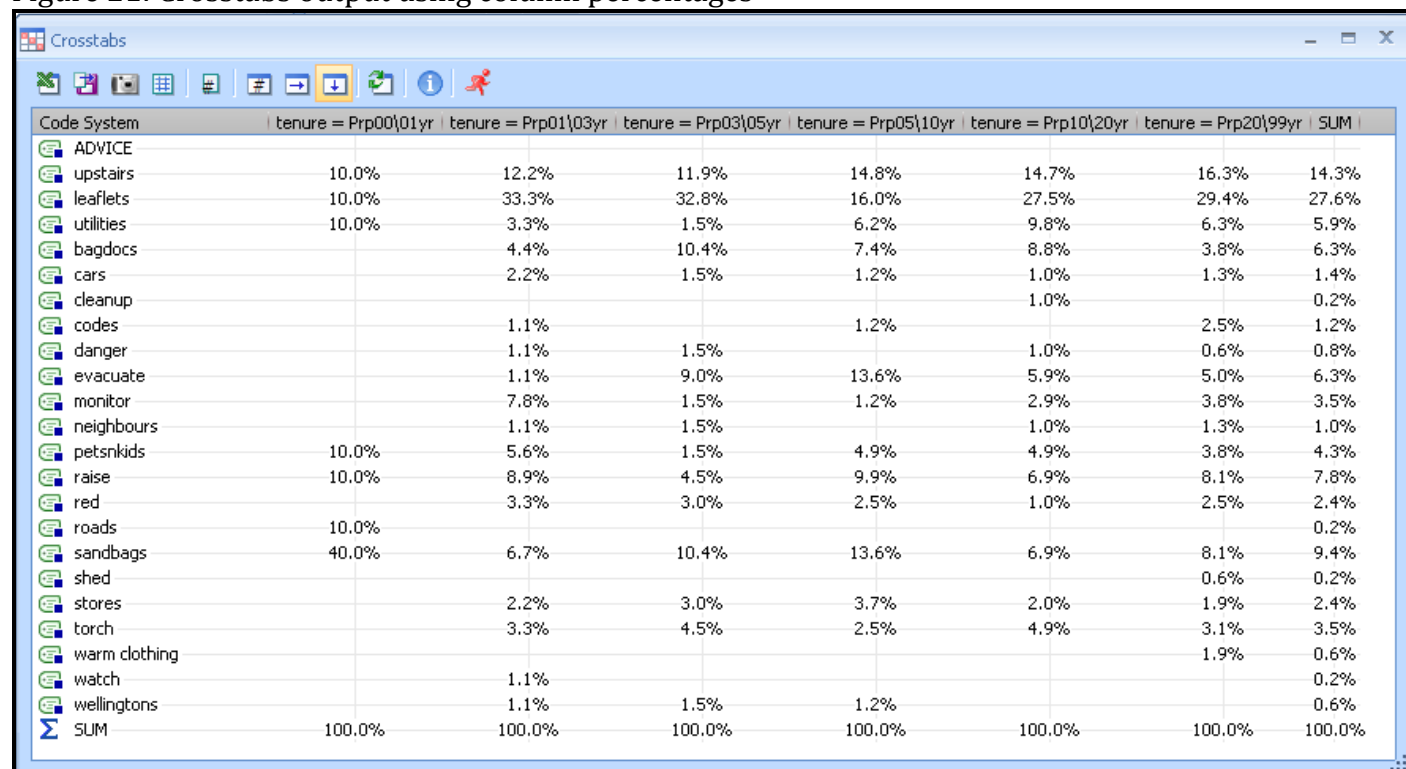
TIP: It is only necessary to activate some codes in order to enable the Crosstabs option in the first place. If no documents have been activated, the program automatically includes all documents in the calculations. However if some, or all, of the documents have been activated the program displays an additional confirmatory dialog "Only for activated documents?" with a Yes/No choice as the last step. This is superfluous if all the documents have already been activated.

Having provided detailed guidance on how to prepare the Crosstabs output in MAXqda, it is perhaps worth briefly considering how such a table might be interpreted. In many forms of qualitative analysis such a table could easily be over-interpreted because of the small numbers of cases involved. However, here in a survey situation where in all probability there is a large number of cases, such a quantitative output may well be justifiable. (The data being used in this example are responses to a question about what advice the respondents remembered receiving prior to the flooding incident).

An important problem in crosstabulations arises when there are large differences between the numbers of cases across the columns, this may be partly overcome by looking at the row percentages as shown in Figure 20. The bottom row of the display shows the percentages of all coded segments for each attribute value, and these can be seen to range from 2.0% to 31.4%. What is interesting here is that references to "upstairs" as shown by the values on the top line of data are in similar proportions along that row to those of the totals, so the upstairs concept shows no significant variation across this attribute variable relative to the total responses. On the other hand the code "evacuate" shows a very different pattern across its row in the table, with above proportion percentages in the middle columns (tenure between 3 and 10 years) and below proportion percentages amongst the newer and longer term residents. Such a variation may be worth investigating by going back to the segments that have that code and reading them in the light of this apparent pattern. A double click on any value cell in the Crosstabs table will bring the segments that match that attribute and code combination into the Retrieved Segments window where they can be read and compared.

An alternative way of thinking is to use the column percentages setting. Figure 21, below, uses the same underlying data as Figure 20 but with this change of display. Now it is possible to consider whether the proportions of the codes applied to the responses of one column differ from those in another.

Figure 21: Crosstabs output using column percentages

Code System	tenure = Prp00\01yr	tenure = Prp01\03yr	tenure = Prp03\05yr	tenure = Prp05\10yr	tenure = Prp10\20yr	tenure = Prp20\99yr	SUM
ADVICE							
upstairs	10.0%	12.2%	11.9%	14.8%	14.7%	16.3%	14.3%
leaflets	10.0%	33.3%	32.8%	16.0%	27.5%	29.4%	27.6%
utilities	10.0%	3.3%	1.5%	6.2%	9.8%	6.3%	5.9%
bagdocs		4.4%	10.4%	7.4%	8.8%	3.8%	6.3%
cars		2.2%	1.5%	1.2%	1.0%	1.3%	1.4%
cleanup					1.0%		0.2%
codes		1.1%		1.2%		2.5%	1.2%
danger		1.1%	1.5%		1.0%	0.6%	0.8%
evacuate		1.1%	9.0%	13.6%	5.9%	5.0%	6.3%
monitor		7.8%	1.5%	1.2%	2.9%	3.8%	3.5%
neighbours		1.1%	1.5%		1.0%	1.3%	1.0%
petsnkids	10.0%	5.6%	1.5%	4.9%	4.9%	3.8%	4.3%
raise	10.0%	8.9%	4.5%	9.9%	6.9%	8.1%	7.8%
red		3.3%	3.0%	2.5%	1.0%	2.5%	2.4%
roads	10.0%						0.2%
sandbags	40.0%	6.7%	10.4%	13.6%	6.9%	8.1%	9.4%
shed						0.6%	0.2%
stores		2.2%	3.0%	3.7%	2.0%	1.9%	2.4%
torch		3.3%	4.5%	2.5%	4.9%	3.1%	3.5%
warm clothing						1.9%	0.6%
watch		1.1%					0.2%
wellingtons		1.1%	1.5%	1.2%			0.6%
SUM	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

The left-most column of data is probably unreliable for this comparison as it has far fewer respondents in that category, those living at the property for less than one year. This is confirmed by toggling to the number of segments display and observing that the sum total of that column is just 10 coded segments compared to an average of 100 for each of the other columns. The remaining columns appear to be remarkably consistent with each other and the proportions of the total column. The most significant difference seems to be in the column of data for those who had lived from 5 to 10 years at the property, who had much less recall of receiving leaflets with advice (16.0% of their codes compared to 27.6% of the codes for the whole sample), and this group also reported advice about evacuation more frequently than the overall average (13.6% compared to 6.3%). There is no obvious explanation for these variations, and the overall conclusion is that length of tenure at the property did not influence recall of the advice received.

3.4 Quote Matrix

It is possible in MAXqda to replace the numerical display of the Crosstabs output with a textual report of all the segments that are referenced by the table of codes. As text obviously takes up more space than numbers, this option should be used with some care otherwise the analyst is likely to generate large volumes of waste paper. The textual version of Crosstabs is called the "Quote Matrix".

The Quote Matrix can be generated in two ways, either directly by selecting the option with that name from the Mixed methods menu, or it can be derived from a Crosstabs table. The two-stage process, via Crosstabs, is probably the safer one to use as the intermediate stage permits a check to be made on the quantity of segments to be extracted before the final step is taken. There is an icon on the toolbar in the Crosstabs output, the fourth from the left in Figure 21 above, which will generate the quote matrix from the Crosstabs table that is displayed.

Figure 22: Example of a Quote matrix output

	where = House	where = Outer	where = Notfl
EFFECTIVE			
None			
Common			
Apathetic	take stuff upstairs and cant remember what else BMRB Docs\#TEXT82510, 2	given leaflet from environment agency but did not take much notice BMRB Docs\#TEXT07202, 2 a leaflet can't remember BMRB Docs\#TEXT25619, 2 package in plastic bag didn't read it! BMRB Docs\#TEXT85532, 2	can't remember BMRB Docs\#TEXT07226, 2 the plastic cover was to put valuables in p cant remember anything else BMRB Docs\#TEXT59504, 2
Negative	" it was pathetic. move furniture etc; switch off electric, given plastic bags for documents. no advice on clean up " BMRB Docs\#TEXT15650, 2 a flood pack stating obvious silly stickers BMRB Docs\#TEXT25622, 2 getting wellington's totally inappropriate info BMRB Docs\#TEXT25629, 2	loads of advice but only my garden was at risk BMRB Docs\#TEXT15648, 2	what to do with livestock and furniture it seemed so farfetched at the time my husband has lived here most of his life and remembered only seeing it the once BMRB Docs\#TEXT34321, 2 told warnings had been changed and were given a small card p nothing elses was said BMRB Docs\#TEXT92219, 2
Positive	the estate came and warned me making sure i was alright p they moved the furniture upstairs p they brought sandbags round back and front i cannot see what else they could have done n BMRB Docs\#TEXT76641, 2	red card through the post very good BMRB Docs\#TEXT17922, 2 they covered everything they gave the definitions of the things for severe and mild and about sandbags .p. 0 BMRB Docs\#TEXT33718, 2	

In Figure 22, above, a quote matrix has been generated to show all the segments from interviews with women that were coded as to how effective they thought the prior advice had been. This output is sent directly to the word-processing program by MAXqda, so this is an image of that document. The yellow highlights were inserted automatically and serve to distinguish the speaker identifiers from the response texts.

As can be seen in this example, this sort of output would quickly become unwieldy if too many columns are used (ie too many variable values are selected) or if there were a large number of segments in any of the cells of the table. It is for this reason that we suggest running the crosstabs calculation first in order to confirm that the number of texts in the output will be manageable.

However, it can also be seen that when the quantities of texts are suitable, it is much easier to compare the language and styles used by the various speakers across this range of views and attributes, and thus interpretation of the numerical pattern may be made more easily than would be possible by pulling each cell's contents into the Retrieved segments panel one at a time within the program.

3.5   Code Relations Browser

The Code Relations Browser is a routine that displays counts of co-occurrences or overlaps between selected codes in the data. It is a visual tool that is designed to aid the identification of patterns of association in the coded data. The output from this routine is a table where both the rows and columns represent codes, so the primary focus of this tool is the relationship between each pair of codes in the selection, however if Activation by variable is used prior to running this tool it is possible to examine such relationships in the context of any particular variable value.

Figure 23 shows an example of the Code Relations Browser using our demonstration data once more. Here a set of codes used to identify the range of topics mentioned in the responses to a question about advice received is related to a set of codes used to identify the respondents' evaluative judgements on that advice. It is quickly apparent that the negative comments outweigh the positive ones, and that they are mainly associated with the "leaflets" code.

Figure 23: Code Relations Browser output
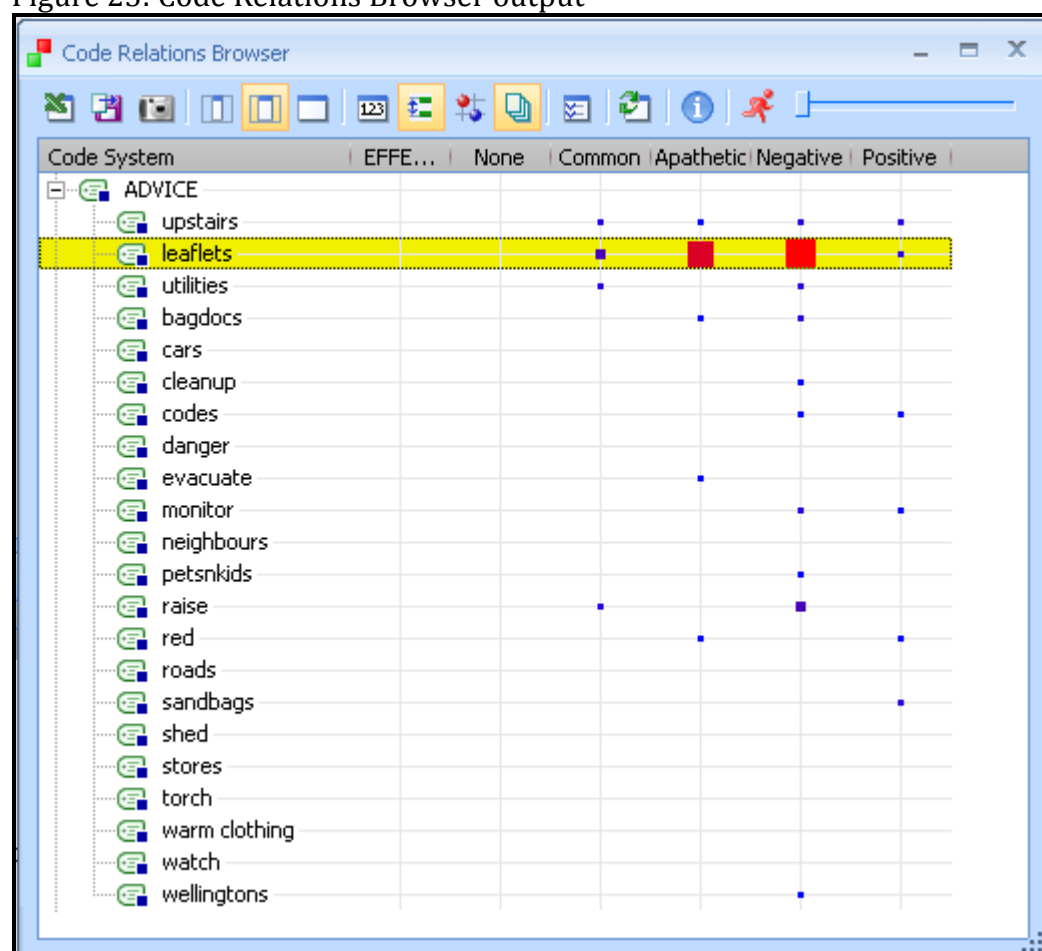
The steps to generate this sort of output are as follows.

Firstly, reset all activations and then activate just the thematic codes that you want to see represented in the rows of the output table. In Figure 23 this involved activating the "Advice" group of codes only. Then select the Code Relations Browser option, either from the Visual tools menu or the appropriate icon on the Visual tools toolbar (if that has been opened). The first dialog box asks if you want to run the analysis "Only for activated codes?" and you should answer "Yes" to this question. The next dialog screen invites you to select the "Codes on x-axis" and this is where you will select the columns for the output table. Here it seems to help if your coding scheme has been organised in hierarchical groups as the dialog offers you all the codes at the highest level in the code system, and selecting one of these causes all of its sub-codes to be shown as separate columns in the table. The "OK" button then generates the output table.

The toolbar of icons within the output table (as shown in Figure 23) includes several ways of modifying the display. In Figure 23 the 7th icon from the left (with numbers "123") is a toggle switching the display between count numbers and coloured symbols. If this is toggled to its "off" position (not highlighted) then the 9th icon from the left becomes relevant and this is a toggle between coloured squares or circles with size proportional to the count numbers, both of these have been turned off in Figure 23 so the display is shown as squares. The icon between the two just mentioned, ie the 8th from the left, is another toggle that switches a "near" setting on or off – this changes the numerical data and so triggers a recalculation every time it is switched – this interacts with a setting in the options button (11th from the left) which allows you to set the number of paragraphs to be included in a near overlap. The 10th icon from the left is yet another toggle, this time switching between a document count (when it is highlighted) and coded segment count.

TIP: The interaction between all of these toggles permits 12 different displays to be shown (without considering further changes with the options icon) so some care is necessary to make sure that the combination of settings used makes sense in terms of the nature of your data and the way in which it has been coded. If you have always applied the thematic codes to the whole of each relevant response, which may be the case if you have used autocoding extensively, then the "document" and "near" toggles may not make any difference to the output. If every response is set out in a single paragraph, which is likely if you have imported the survey data via a spreadsheet, then changing the number of paragraphs in the "options" may not be meaningful.

3.6   Document Comparison Chart

The Document Comparison Chart provides a visual check on the extent of the coding that has been applied. It does not have much direct contribution to make to the analysis of the sort of data used in these examples. Figure 24 below shows an example of part of the output from this routine.

In Figure 24 the display shows the coding status of 20 respondents to one question in this dataset. The coloured bars show where there is coded data, and the colours shown reflect the colour scheme defined for each code in the code system. Each question in the dataset has been allocated a column in this display, which is equivalent to the paragraph number that shows in the Document Browser window. When the data was first imported into the project the response to each question was allocated a code with green colouring, whilst the thematic codes applied so far to paragraph 2 have been allocated blue, orange and red colourings (for groupings labelled "Advice", "Who from" and

"Effective" respectively). White space in the second column of this display indicates that there is no data for that question from that respondent.

Figure 24: Document Comparison Chart display example
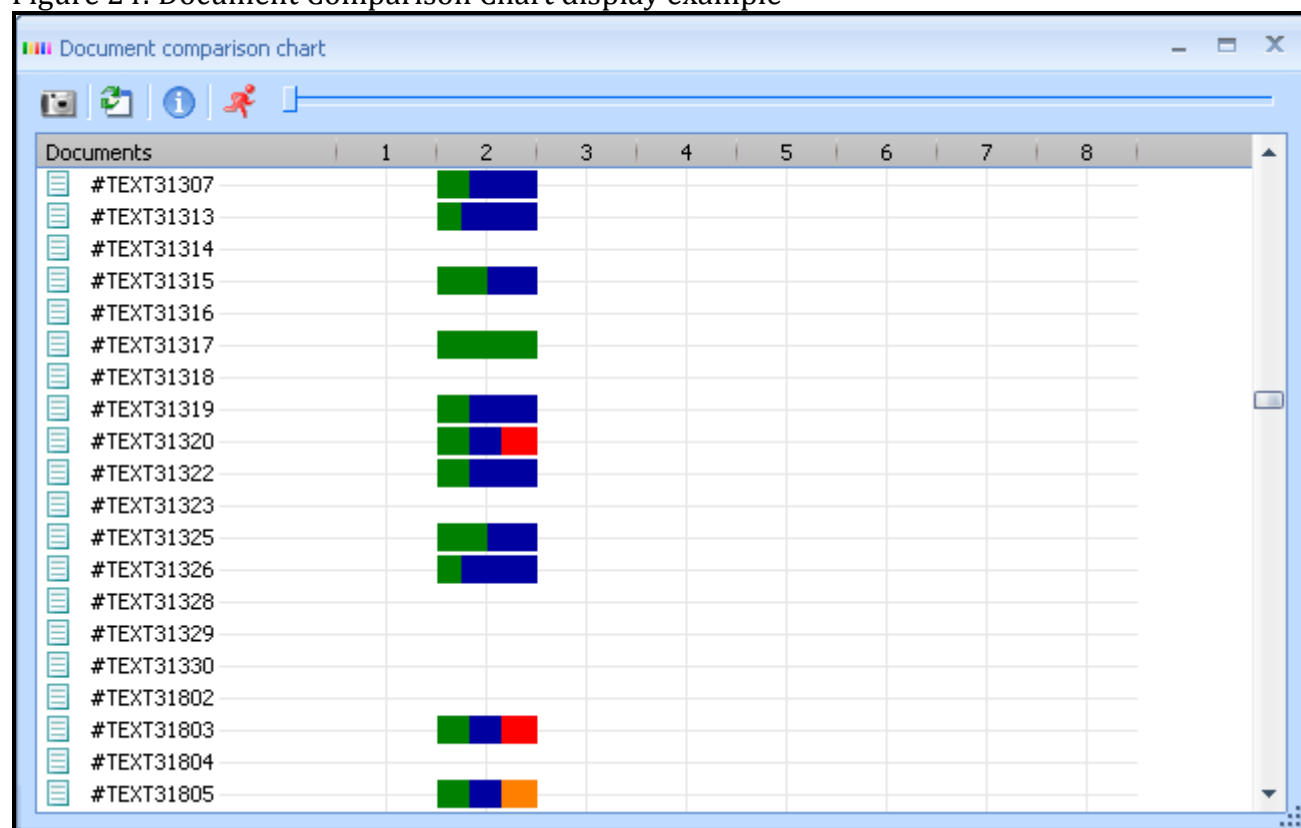
This display was generated by first activating the code for the second question, and the sets of thematic codes relevant to that question, and then running the function on "activated codes only". Those activations make the display simpler by only showing the column for the question responses which are being checked.

In the bottom row of the display, respondent number 31805 has had equal numbers of codes from the green, blue and orange groups (as interpreted from the equal-sized blocks of colour), while number 31803 just above has had equal numbers of green, blue and red codes applied to the second question. What may be significant is that respondent number 31317, near the top of this display, has only green coding for the second question, and this indicates that there is some text for that question but no thematic code has been applied to it – that response should be checked to make sure that there is a good reason for it remaining uncoded.

TIP: This Document Comparison Chart procedure provides a visual method for quickly checking which responses to the activated question have not had any thematic codes applied to them. It cannot be used to identify a specific omitted code, the procedures set out elsewhere on this website should be used for that purpose.

3.7   Summary conclusion

Many qualitative analysts, who would be familiar with these CAQDAS programs, may be reluctant to use quantifying statements about the data that they analyse. This would be for the very good reason that most often the size of the sample that they are working with is too small to justify a numerical conclusion. However, when one is working with survey data then there is more likelihood that the sample size is sufficiently large and sufficiently random to support some quantified statements.

At the same time many quantitative analysts may be wary of applying statistical techniques to data which has been coded separately from the data collection process. However, provided the data collection process was sufficiently robust to capture the data accurately, then the fact that the codes have been applied by research analysts rather than the respondents themselves should not be a reason to water-down the analysis processes. Because the coding process is transparent and replicable, the coded data may be used with more confidence than that derived from the 'amateur' coders of the respondents to closed questions.

Thus, if quantitative conclusions are being drawn from the rest of the survey data, then there should be no reason why quantitative conclusions cannot be drawn from the open-ended question responses as well.

4.0   Export Data to Statistical Packages after analysing Open-ended Survey Questions in MAXqda

In common with other pages in this section of the website, this page is a series of observations about how the features of MAXqda might interact with a particular sort of dataset. This page should be read in the context of the related materials concerning the use of MAXqda with Open-ended Survey Questions, in particular the Data Preparation Instructions and the Qualitative Analysis Strategies, since the export strategies outlined below can only be effected after the data have been imported and coded systematically in a MAXqda project.

The tools discussed below are illustrated with examples from the same post flooding event survey that was used to illustrate the data preparation processes. For a summary of the project from which this data derives see here. This data is characterised by a fairly large number of short statements.

Outline:
   4.1   Use of Code Matrix Browser
   4.2   Editing in MS Excel
   4.3   Checking data transfer

Detailed Guidance:

4.1 Use of Code Matrix Browser

Creating a Microsoft Excel file with the count of code frequencies for each respondent is quite straightforward in MAXqda. It is a special application of the Code Matrix Browser. Before starting the routine you should activate all of the cases (this can be done simply with a Ctrl + click on the group header in the Document System window, or a right click on the group header followed by selecting "Activate all documents" from the context menu that appears). Then you need to activate all of the codes that you want included in the export table – this may not be absolutely every item in your Code System as, for example, there would be no need to include the question identifier codes that the system set-up when you first imported the data into the program – Ctrl + click on each hierarchical group of codes that you want to include will achieve this.

With the activations in place, select *Visual tools / Code Matrix Browser* from the menu bar or the appropriate icon from the visual tools toolbar, and confirm "Yes" to the dialog "Only for activated documents and codes?". The routine will then process the data, which may take a few moments if you have large numbers of cases and codes.

Figure 25, below, shows a portion of the output for our example data. This takes the form of a very wide matrix with 1,257 columns of data (because we had 1,257 respondents) and a row for each code. It is possible to view different parts of the table by using a slider to the right of the toolbar in the report window or by using a Windows slider at the bottom of the report. We applied relevant codes only once to each response so the frequencies in this table are either 1 or blank.

Figure 25: Code Matrix Browser for all cases and selected codes

The procedure for exporting this report is simply to click on the "Excel" icon at the left of the toolbar in the Browser window (as can be seen in Figure 25, above). This opens MS Excel, imports the data, and transposes the data so that the codes appear in columns and the cases in rows. It also inserts a zero in all of the blank cells of the table. So the result in our example is a spreadsheet with 1,257 rows and 37 data columns, as illustrated in Figure 26, below.

Figure 26: Excel table following export from MAXqda

4.2   Editing in MS Excel

If you are intending to move this data into SPSS or another statistical program you may need to do some further editing. The code names in the top row of the table may appear unhelpful with the hierarchical group name and backslash preceding the important part of the name, but these can be manually edited within Excel quite easily to match whatever restrictions the statistical software may impose on variable names. The case identifiers may be rather more difficult to turn into a form that enables matching with those already in the statistical software because the distinguishing number is preceded by unwanted text.

The first step in cleaning up the case identifiers is to insert two new columns between the current column A and the first column of code data, ie between columns A and B in Figure 26. The corrected IDs will be created in these new columns. In our example the required data are the last five digits of

the text in column A, and these can be extracted by typing the logic =RIGHT(A2,5) into the cell at B2 (in the new empty column just previously inserted) and then copying that logic all the way down column B. It will then be necessary to convert these logical statements that display the correct digits into absolute values in the new column C. This final step involves copying all of column B and then pasting it into column C with a "paste special" command to store "values" only. When this has been completed and checked for accuracy, columns A and B can be deleted and the workbook can be saved.

The column headers in row 1 of the spreadsheet are likely to be used by the statistics package as the names of the new variables whose data is in those columns. Here the problem lies with the prefix path labels that MAXqda has inserted in front of the code name. On the assumption that the number of variables exported is not very large it is probably simplest to edit these manually as with any text cell in MS Excel.

When you are satisfied that the row and column identifiers are in the correct format for importing into the statistics program, save and close the spreadsheet file.

4.3   Checking data transfer

Whenever large amounts of data are moved between programs there will be potential sources of error so, before continuing with the analysis in your statistical package, it would be advisable to check a small sample of cases in that package to confirm that the code frequencies have been matched with the correct cases. This is not simply a matter of confirming that the total frequency for each imported code variable is the same in MAXqda and your statistical program but also that individual cases have correct codes.