

Chapter 3

How to get texts ready for analysis

Although during later working sessions with AQUAD it is sufficient to click on an icon on the computer screen (to load AQUAD) in order to immediately afterwards continue working on your analysis, at the very outset you must do a number of things that make your data files available for AQUAD. In particular, you must

- format your text documents;
- convert the data you entered into your word processor into a format called "RTF";
- make a list of all the data files you want AQUAD to work with for the research project at hand;
- have AQUAD number the lines of your texts.

In case you work with multimedia files in your project, please, refer to chapter 4, where all the necessary details are explained how to prepare graphic files (for instance, scanned drawings or text pages), sound recordings or video recordings for analysis in AQUAD.

3.1 Format your text documents

You may use whatever word processor you prefer for entering your data. Before they can be worked with in AQUAD, however, you need to prepare each data file in a certain way. Since AQUAD allows you to attach codes to segments of text directly on the screen, it needs some space in which to place the code names (and the number of the line where the coded segment ends). Therefore, if you want to use this option (rather than doing your coding exclusively on a paper print-out and entering the coding information into the computer without seeing the text), you are asked to make a wide right margin when you enter your data into a word processor. In fact, each of your lines should be no longer than 60 characters, and they could be even shorter if your texts contain a lot of information and therefore require heavy coding.

If you prefer the paper-option for coding (two-step coding), then the length of the lines does not matter. However, long lines may carry more than one unit of meaning which has to be coded. Since you may attach up to 40 codes to each line of your text, and since units of meaning may overlap, there will be no problem for handling your codes in AQUAD. But for you coding is much clearer if you reduce the probability of multiple meanings in the same line of text by shortening the lines. At the same time a broad right margin gives you more space to mark text segments and to enter code names.

This "formatting" is done while your data are entered into your text processor. In most word processors you determine the maximum length of the lines by setting the margins accordingly (no margin on the left, a margin on the right wide enough to make sure not more than 60 characters/spaces will fit on the line before the next word is wrapped to the following line). You should not use a proportional font, where each character gets exactly as much space as it needs, that is, a "W" consumes more space on a line than an "I": WI... – and consequently the number of characters on lines of defined width varies according to the text's content. Instead you should apply a non-proportional font like Courier, which reserves the same amount of space for each character, as you can see here: WI . . .

What can you do if your texts are already written and formatted with a word processor? This is no problem at all. All word processors, even simple text editors offer options to re-format texts as described above. In most word processors you can define the layout of pages ("style" definition) and load this definition together with your text files, which then become adapted to the new format automatically. Please, refer to the manual of your word processor for details.

3.2 Provide the proper file names

When you save your data files, AQUAD requires that you use only eight characters for naming your files. The rule for file names in AQUAD is:

- File names are composed of a name part and an extension part.
- You define only the name part, consisting of maximally 20 characters.
- Extensions are added automatically by your text program (for instance: *.rtf; see below) and by AQUAD.
- Suggestion: All text files in a project get the same name (maximally 17 characters), but are distinguished by a different three-digit number.
- Never use "work" as a file name, because this word is reserved for internal use by AQUAD.

You will probably want to use the characters and numbers to indicate what the content of the particular file is, so you can easily recognize it. For example, if your data consist of 24 interview transcriptions which you have numbered, you might name the original files that are the basis of your analysis

inter_001, inter_002, ... , inter_024.

When AQUAD works with these files, it creates additional parallel files, as for instance files that hold the code information for a particular file. The program constructs the new file names from the elements of the original names, so that the connection to the original file can be recognized. In our example, files with the names

inter_001.aco, inter_002.aco, ... , inter_024.aco

would be set up by AQUAD. (After coding you can see these new names in the subdirectory which you have declared as location for storing your code files.

If your data files are already saved with names that deviate from this convention, you may rename them. You can do this either within Microsoft's WINDOWS "File Manager" or "Windows Explorer" – or you may switch into the DOS window (explained later in this manual), make sure the DOS prompt indicates your subdirectory for text files, and use the command RENAME in the following way:

RENAME oldname newname

Let us assume, our interview files were named originally according to the interviewees' first names (which we would not recommend to secure anonymity). Let us further assume the first two of these original files are *john.txt* and *mary.txt*. We rename these files as follows:

RENAME john.txt inter-001.txt
RENAME mary.txt inter-002.txt

and we do so in the same way for the remaining 22 files of our project.

3.3 Convert the data files into RTF (Rich Text Format)

Each data document that you want to use in AQUAD *must* be converted into a plain ANSI text file or into RTF. The first abbreviation stands for "American National Standards Institute", the second for "Rich Text Format". When you create text, your word processor fills your documents with formatting and other information that is invisible to you. These commands instruct the program, for instance, when to break up a line and move on to the next line, where to set margins, tabs and indents, they specify line spacing, page breaks etc., tell the printer what font to use and many other things. Unfortunately the commands are unique to each word processor. For instance, the last sentence would be stored in WordPerfect, the text processor we used for this manual, as

Unfortunately the commands are unique to each word processor.

These program-specific characters need to be stripped away, if you want to exchange your texts via e-mail or to import your data into another program; they must be "standardized." In technical language, you must convert your documents into ANSI documents, the "standard code" or into RTF.

Most word processors offer a facility to convert text documents into ANSI or into RTF files when saving them within the function "Save as...". Please, read the manual of your word processor for further information. Key-words in the manual's index like "ANSI", "DOS text", "Import/Export", "text-in/text-out", "Rich Text Format", etc., may lead you to the information needed. If there is no choice between formats when saving your texts, then there should be at least a small additional program for text conversion. In this case, you should look in the directory, where your word processor is stored, for *.exe (program) files, whose names contain sequences of letters like "cnv", "conv", "textcnv", etc. (cnv or conv usually signals "conversion"). Usually, an utility program of this kind has to be started separately from your word processor to convert your data before you can import them into AQUAD.

With older versions of some word processors it may happen that the conversion facility contains bugs. The converted files then contain codes that are invisible in the text. The result in our own experience once was that AQUAD got caught up during loading. In this case the only thing you can do is to use the mode of your text program that reveals the codes so you can see them, or to use a special text editor. The annoying characters (mostly only one!) must be deleted manually. With newer versions of word processors such problems, to our knowledge, do not occur.

If your word processor is ancient enough not to include the option to save text in ANSI format or in RTF, find a small utility program that includes this procedure and use it on your data before you enter them into AQUAD.

The next section can be skipped if you are reading this chapter for first orientation. It is important only once you have been working a while with AQUAD and discover belatedly that you still need to correct something in one or the other of your data files.

Editing data files after they have been converted into ASCII format or into RTF

If you notice errors in your original texts (or the line-numbered or coded data files, see chapter 6), you may wish to correct them or add something to them. However, there is a basic rule in qualitative analysis that text documents should not be changed during the analysis. Many people found that versions 3.x of AQUAD made it too easy to violate this rule and they asked for some changes. What you can do now since version 4 if you happen to detect for instance an irritating typing error is only to go back to your word processor and edit your original document, convert it again into ANSI format or RTF, and save it in the path determined for data texts. There is one caution, though! Whenever you make changes in a text file

- (1) after you have started to code this text and
- (2) this change includes adding or deleting complete text lines
- (3) and you have the ANSI/RTF file of this text line-numbered again

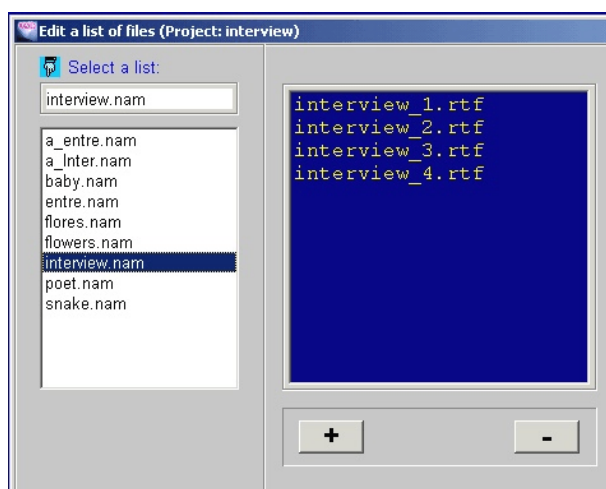
then the code entries in your code files still refer to the line numbers before editing this text! In this case – and only in this case – the line numbers stored together with your codes *will not be valid for the newly*

formatted data. This means that all coding you might have done will be invalid, as well, since it refers to the old line numbers. Therefore, we recommend that you keep editing to a minimum, that is, you keep it to changes like correcting typing errors, which will not lead to adding new or deleting old lines. Or you give the newly formatted files different names that identify them as different versions. This latter option is also suggested in cases where two people want to work independently on the same data.

3.4 Prepare a list of all data files (one or more file "catalogs")

When you determined the settings for your new project (see chap. 2), you listed already all cases available for analysis in your project (text files with your transcriptions) in a file catalog. However, especially in the beginning of a project, you may want other file catalogs than one major catalog listing *all* files of your project. It is a good idea to begin your work with one or more small samples of data files. Therefore, you may wish to edit your comprehensive file catalog and create one or more smaller catalogs.

You can "*edit a list of files*" by activating this option in the sub-menu which appears after you click on "*Files*" in the main menu.



The white window on the left shows, which lists of file names are available at the moment. Before you start your own project, only lists of files of the example projects will shown, which were copied to your hard disk during the installation. Among these lists you see the list "*interview.nam*". After selecting this list by clicking on its name, the blue window to the right is filled with the names of those data files, which belong to this project, here:

interview_1.rtf,
interview_2.rtf,
interview_3.rtf, and
interview_4.rtf.

Supposed we want to work only with the first and the third file, we click on the "*minus*" button in the box below the list of file names and activate the function to delete file names; then we click on the second and the fourth name and thus delete them from the list.

Of course, we can add these files again or other files to our project by clicking on the "*plus*" button. This will open additional boxes for selecting the directory of files to be added and a box containing the names of data files in this directory. Clicking on the names of files (blue box), which we want to add, will copy them immediately into the new list of files. Don't be afraid to experiment with this option. Nothing will happen to your original file catalog (see chap. 5), because the editing is done with a temporary copy of this catalog.

After your new file catalog is finished, click on the "*OK*" button. Your entries will be saved instead of your original file catalog -- unless you decide to use a different name for your new catalog. AQUAD will warn you whenever some file would be overwritten by an edited version, and you get a chance to enter a different name for saving the edited file -- here your new file catalog -- separately.

At this point, you are probably eager to begin coding your data. Before you start, however, you should know that all data texts, which were not imported automatically into AQUAD when you determined the settings of a "*New project*", have to be converted manually by applying the function "*Import of texts (RTF)*."

3.5 Import RTF files

At the moment, we have plain ANSI files or RTF text files stored in the directory for data texts. When you activate either the option "*Import of texts (TXT)*" or "*Import of texts (RTF)*" in the sub-menu "*Files*," the data texts listed in the file catalog, which you are going to select now, will be copied to another path and transformed into AQUAD's internal data bank format.

As you can see in the screen shot on the following page, you have four options to select ANSI formatted or RTF text files for importing:

- You may import *all files* belonging to the actual project, that is all the files you listed in the original file catalog, when you determined the settings of this project initially (see chap. 2/2.2). We *strongly* recommend to apply this option.
- You may import all files listed in the actual file catalog (here: "interview.nam"). If you did not edit your initial list of files (see chap. 2/2.3) and did not activate one of the newly created catalogs, this option will cause the same effect as option #1: since all your data texts are listed in the initial list of files (if you followed the recommendation), they will all become imported now.
- You may import all files from a particular file catalog. In this case, AQUAD opens a selection box where all your file catalogs are displayed. You pick one of them (the content is shown in the box to the right) – and the files listed there will be imported.
- Finally, you may decide to import just one (or several) files from a particular file catalog; in this case you mark the last of the four options (see figure below).

