DO IT YOURSELF LEXICOGRAPHY

English speakers find the idea of committees of learned men and women spending hours deciding how their language should be spelt rather cute, but that is what happens with many other languages including Dutch and German.

It is probably just as well that English is allowed to grow naturally. Can you imagine a committee of British, Americans and Australians trying to reach agreement on the spelling of every English word?

When the new spellings came into force in the Netherlands and Germany the QTYP dictionaries became outdated overnight. Our colleagues in the PC world did not have this problem, because there was a market for commercial dictionaries giving the new spelling. But we QL-ers are well used to doing it ourselves, and in practice it is easier to revise our dictionaries than you may think.

Indeed, if you have a reasonable knowledge of a language, it is possible to quickly write a useable QTYP dictionary And if you are prepared to cheat, OCR technology has made the task much easier.

If you obtained your copy of QTYP with a word processor or text editor, you will probably not have a QTYP manual. The Text87 manual gives some helpful information on using QTYP but it is easy to miss this when there are so many other things to learn about the program. To use the spell checker in Text87, you press F6 and are asked tor the name of the QTYP dictionary and those of the word lists. These are the clues about how to make your own dictionary.

Spell checkers are 'off the peg' products and not 'tailor-made'. You cannot expect them to contain your name or those of your family and friends, addresses you use regularly; or specialist words you need for your profession or study. Most spell checkers allow you to make a separate list of these. Every time you come across a word that is not in the dictionary you can save it to a temporary file in memory and later save this file to disk. In Text87 the key presses are F3, Config, Spelling, Save. (On the command line you will see there is also a Load command.) These facilities enable you to adapt the spell checker to your own needs.

One snag is that text lists can take up a lot of memory If you have a big list, it is better to merge it with the main dictionary. This is easy with QTYP. On the Text87 disk or QTYP disk you will find a file called 'QTYP_ded'. You can execute this by a simple line:

EX f1p1_qtyp_ded or equivalent.

When QTYP-ded has loaded, you have a choice of three things:

Create new dictionary
Edit existing dictionary
Rebuild existing dictionary

Let us start by editing an existing dictionary All we do is enter the file name of the dictionary After loading it appears in a window with an active cursor We can now scroll through the dictionary and edit it as required. If we press F3 a small command window opens which allows us to find any word in the dictionary load a secondary word list, save the dictionary or quit the program.

If we choose the second option and load the word list of our specialist words, we see

that it does not load instantly but is loaded in small sections. QTYP is not just loading the word list. It is looking through it word by word, checking for superfluous characters such as punctuation marks, making sure the word is not already in the dictionary and putting it in the correct alphabetical place.

If we press F3 we see the command window now has different options. One of these is to merge the word list into the dictionary Do this and then press F3 again and the original commands are back. Choose save dictionary to file, enter the file name for the saved dictionary and press enter You are now asked whether you wish to compress the dictionary Use 'yes' because the 'no' option does not produce, as you might think, a plain text list. The dictionary is compressed by letter of the alphabet and then saved.

Adding your word list to the existing dictionary is a simple and painless process.

We can now go a stage further and create a completely new dictionary When we choose this option, something rather puzzling happens. You are first asked to load an existing dictionary This is to give the program a few rules about how the dictionary is to be compressed, and any QTYP dictionary will do. This time when you have loaded the dictionary the screen remains blank. Press F3 for the command window and load the word list you have created for the new dictionary. Again this will load in sections. Now press F3 and choose the merge option. Press F3 again, and save the dictionary Once again you can choose between a compressed and an uncompressed dictionary On this occasion if you enter 'no' the dictionary will be saved as a text file, but usually it is better to compress it.

We now have a new dictionary but it will not yet be in its optimal form. The compression techniques were optimised for another dictionary and usually another language. QTYP has a routine to 'tailor-make' the compression for each dictionary and language. This can take some time. The manuals suggest some hours, but do not let this put you off. The manuals pre-date the Gold Card and QL's are now a lot faster

This time we choose the dictionary rebuild option, load our dictionary into it, give QTYP a file name for a temporary log file (e.g. ram1_temp) and then go and make the coffee.

I used this routine on a 52,000 word Dutch dictionary with QTYP working in the background while I was typing this article. The original dictionary was 170,478 bytes long, and QTYP rebuilt this to just 143,885 bytes. On my QXL2 system an English dictionary of 25,000 words took 10 minutes to rebuild and a Dutch dictionary of 70,000 words 40 minutes.

Now the snags. Making the dictionary in QTYP is the easy bit The real work is in the compilation and checking of the word list from which the dictionary is made. This is not too difficult if you are just adding words to an existing dictionary but if you are compiling a completely new dictionary especially if it is not in your native language, you have a big task on your hands. This is where I advise you to cheat.

If you have access to a PC, check your word list using the PC's spell checker When PC word processors were sold on disk, you had to buy language modules separately Now they are sold on CD's and you will usually find several modules on the disk, from which you can install the spell checker, thesaurus and style checker for several languages on your system.

Just Words! has word lists of several languages that could form the basis of a spell checker, although you should check the accuracy of these before use. If you have to build up your own word list, it is a good idea to start by typing in the words from a basic course in the language and make a small dictionary from them. Use this to spellcheck

about 10,000 to 20,000 words of text, and then save and check the extracted words to increase the size of the dictionary Another possibility would be to make a list of words by using a short basic routine to convert all spaces in an ascii text to line feeds (Chr$(10)). If you have access to an OCR reader you could scan texts for new words.

Using these techniques you may get some bogus words, so always control for accuracy preferably using a PC spell checker, before adding to your dictionary Do not worry if your dictionary does not grow quickly Most of us use only about 20,000 words regularly and this number of words will be adequate at first.

This article is a shortened form of an article that appeared in QL Today volume 5 issue 6 page 6