WordSmith Tools Concord Tutorial and Exercise: Level I

WordSmith Tools is a corpus analysis tool suite that was developed by Mike Scott. It includes a monolingual concordancer, known as Concord, whose function is to display the occurrences of a user-specified search string (e.g., term, phrase, character string) in its immediate context in a display format known as KWIC (*key word in context*). Note that the WordSmith Tools suite offers a number of other text analysis tools which can be explored through the WordList and KeyWords tutorials and exercises.

You can learn more about WordSmith Tools by consulting www.lexically.net. Be sure to read the Help file for Concord (located in the Writing Centre in the directory V:\WordSmith4 Help files, and also available online). From the *WordSmith Tools* page on the www.lexically.net site, you can also download a demo version of WordSmith Tools which offers the same options as the commercial version, but which limits the number of occurrences that are displayed to 25.

Concord is compatible with the following file formats: .html, .xml, .txt. It offers a range of search functions, a variety of modes for organizing and displaying results, and statistical features for identifying collocations.

Doing these exercises will help you learn to:

- Choose texts for processing,

- Perform simple searches,

- Process the search results (e.g. viewing, sorting, classifying).

Getting ready

1. Save the files you will need for the exercises:

   a. Create a sub-directory of the U: drive (also called **MY DOCUMENTS**). (For instructions, see the file *Creating a sub-directory* on the CERTT site (**ACCESS BY TOOL NAME > WINDOWSXP > CREATING A SUB-DIRECTORY**).)

   b. Download English Wind Power manual from the CERTT site (**CORPORA AND OTHER RESOURCES > TRILINGUAL RESOURCES [EN – FR – SP] > TECHNICAL TEXTS [EN – FR – SP] > WIND POWER [EN – FR – SP] .TXT > WIND POWER MANUAL_EN**).

   c. Extract the files to the sub-directory you created. (For instructions see the file *Extracting files from compressed folders* on the CERTT site (**ACCESS BY TOOL NAME > WINDOWSXP > CREATING A SUB-DIRECTORY**).)

2. Open WordSmith Tools (**START > PROGRAMS > TRADUCTION > WORDSMITH4 > WORDSMITH.EXE**).

3. When WordSmith asks you if you'd like to enable basic functions only, choose **NO**. By choosing **NO**, you will have access to all functions.

4. The program will open a main window from which the various functions can be accessed. The three principal functions — **CONCORD (C), KEYWORDS (K)** and

WORDLIST (W) — are available via buttons. Other functions can be accessed from the **UTILITIES** menu.

Choosing texts

1. Choose the texts for processing :

   a. In the main WordSmith window, go to the **FILE** menu and select **CHOOSE TEXTS**.

   b. In the *Choose texts* dialogue box, the files that have been selected for processing are displayed in the right-hand pane, and the directory structure appears in the left-hand pane.

   c. A demo file that comes with the software (a chapter from *A Tale of Two Cities* by Dickens) appears in the list on the right. Remove this file by selecting it and then pressing the **DELETE** key. If you do not remove this file, your results will include occurrences coming from this text.

   d. From the drop-down menu in the upper left-hand corner of the *Choose texts* window, select the drive where you have stored the documents that you wish to analyze (U: or *My documents*).

   e. You can also choose to see only files of a certain format (e.g. plain text or .txt files) by choosing the corresponding option from the drop-down menu located immediately to the right of the drive menu.

   NOTE: The default selection for this menu is the *.* option, which means that all the files located in the directory will be displayed, regardless of their name or extension. If you choose *.txt, for example, only the files with a .txt extension will be displayed, though the asterisk means that these files may have any name.

   f. In the left-hand pane, browse through the directories on the U: drive to find the files that you wish to process. (Double-click on the yellow folder icons (📁) to open the directories).

   g. Select all the files. (To select all the files at once, select the first one in the list, hold down the **SHIFT** key, and click on the final file in the list).

   h. Click on the long vertical button with two arrows that appears between the left-hand pane (directory structure) and the right-hand pane (list of files). The files that you have selected should now appear in the right-hand pane.

   NOTE: If you think that you will want to consult this group of files regularly, you can save the list by clicking on the *Save favourites* button (💾). You can even add comments about these files in the lower pane of the window. (Be sure to save the resulting list of files in the sub-directory of your U:, where it will be available to you in the future).

   The files can then be added as a block the next time that you use Concord. This can be done by opening the list using the *Get favourites* button (📂). Loading the files can take a few minutes, so be patient!
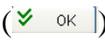
2. Once you have selected the files you want to use, exit the *Choose texts* window by clicking on the green check-mark button (✔).

Generating a concordance

1. Open the **CONCORD** tool by clicking on the *Concord* button in the main WordSmith window.

2. From the **FILE** menu, choose **NEW**.

3. In the *Getting Started* dialogue box that pops up, click on the **SEARCH WORD** tab to ensure that it is active. On this tab, you will see a search field at the top, and, in the lower portion of the tab, some examples of different search options that are available.

NOTE: WordSmith also allows you to conduct a search using a previously prepared file containing a list of search words. However, this option will not be explored in this tutorial. For more information on this option, please consult the WordSmith Help files.

4. Enter the search string *terrain* in the search field and then click on **OK** (⌄ OK) to begin the search.

5. The results will appear in the *Concord* window. Each occurrence of *terrain* appears on a separate line with this search string centred and in colour in the middle of the line. The source of each occurrence is displayed to the right in the *File* column.

6. Take a look at the statistics provided. WordSmith calculates the position of each occurrence in the sentence, paragraph, section and text, and expresses this information as both a raw value and a percentage. (For example, an occurrence that appears at the very end of a text should be in the position expressed as 100% in the sentence, paragraph, section and text, while an occurrence that came at the very beginning would be in position 0%. However, the raw values will vary according to the number of words in a sentence, paragraph, section or text).

   a. Can you think of any useful information that could be gleaned from these statistics?

Displaying the results

1. To adjust the width of the contexts, position the cursor on the dividing line between the columns *Concordance* and *Set* (the cursor will change shape to look like a vertical bar with an arrow on either side) and pull to the right.

2. To adjust the height of a line, place the cursor on the dividing line between two lines (i.e., between the line numbers) and pull down.

3. To adjust the height of all the lines at once, open the **VIEW** menu and choose the options **GROW** (■) or **SHRINK** (▬). (Alternatively, you can press the **F8** key to increase the line height or **CTRL+F8** to reduce it).

4. If you prefer to see only the complete sentence in which the search string appears, choose the **SENTENCE ONLY** option from the **VIEW** menu.

5. To see the complete text from which the occurrence has been taken, double-click on the line in question to select it. The text is displayed on the **SOURCE TEXT** tab with the occurrence of the search string underlined. Once you have finished consulting the text, click on the **CONCORDANCE** tab (at the bottom of the window) to return to the results display.

6. You can also classify your results according to categories that you can define. For example, you could mark those lines that show a potentially interesting collocation for the search string *terrain*.

NOTE: It could be very interesting, for example, to indicate the parts of speech for a character string that could have multiple possible parts of speech (e.g. *power* (n) and *power* (v)), or to indicate different senses for words that have more than one meaning (e.g. *pole* (of the Earth) , *pole* (of a battery) and *pole* (long, narrow cylindrical object)), or even simply to distinguish occurrences that are interesting or problematic from those that are not.

    a. In the *Concord* window, position the cursor on a line in the *Set* column.

    b. Type a letter to represent the category you wish to assign to a given occurrence. (For example, you could type *N* for cases where the collocate of an occurrence is a noun and *A* for cases where it is an adjective, etc.). Once you have classified a given line, the cursor will automatically move on to the next line.

    c. Continue classifying the occurrences by assigning a letter to each line. (In step 7, you will learn how to sort the data according to these classes).

7. To sort the results, go to the **EDIT** menu and choose **RESORT** (⚙). (Alternatively, you can press the **F6** key.) The occurrences will be automatically sorted in alphabetical order according to the information recorded in the *Set* column. (Those lines for which no information has been recorded in the *Set* column will appear first). Click on **RESORT** again to see the occurrences in reverse alphabetical order.

    a. Could this type of sorting be helpful for studying the usage of a search string? How?

8. To save the search results in the form of a concordance that can be opened in WordSmith Tools:

    a. From the **FILE** menu, choose **SAVE** (💾).

    b. In the dialogue box that pops up, choose the drive where you want to save the results and then navigate to the desired folder.

    c. Enter a filename in the field.

    d. Click on the **SAVE** button to save the file.

9. To save the search results in the form of a concordance in plain text format (.txt), which can be viewed and edited in a word processor or imported into a database:

    a. From the **FILE** menu, choose **SAVE AS…> PLAIN TEXT** (📄).

    b. Click on the yellow folder icon (📁) to choose the drive where you want to save the results and then navigate to the folder that you created for your exercises.

    c. Enter a filename and click on the **OPEN** button.

    d. When you are returned to the *Save as Plain text* dialogue box, click on **OK** to save the file.

NOTE: If you save the file in .txt format, you can specify the length of the context to be extracted by going to the Concord window (SETTINGS > SPECIFIC TO CONCORD > CHARACTERS TO SAVE (PER ENTRY) > OK).

10. Sorting results according to other criteria.

NOTE: As noted above, when information has been added to the *Set* column, WordSmith uses this information to sort the results. In order to sort using different criteria, it is necessary to delete the information in the *Set* column. (Note, however, that since you have already saved a copy of the annotated concordance, you can always refer to it if you want to).

a. From the **EDIT** menu, choose **CLEAR SET COLUMN**. Information previously entered in this column should now disappear.

b. From the **EDIT** menu, choose **RESORT**. In the *Concordance Sort* dialogue box that pops up, you can specify up to three different sort criteria to be taken into account (a main or primary sort criterion, a secondary sort criterion and a tertiary sort criterion) by clicking on each of the three tabs. Within each tab, some of the sort criteria that you can choose include:

    i. **FILE**, which sorts the occurrences according to the source file in which they appear.

    ii. **R1**, which sorts the occurrences in alphabetical order according to the word which appears one position to the right of the search string (i.e., immediately after the search string). **R2** sorts by the word two places to the right of the search string, etc.

    iii. **L1**, which sorts the occurrences in alphabetical order according to the word which appears one position to the left of the search string (i.e., immediately before the search string). **L2** sorts by the word two places to the left of the search string, etc.

c. Checking the **CASE SENSITIVE** box will allow case to be taken into account during a sort if desired.

11. Sort the results to highlight different combinations with and collocations of *terrain*:

NOTE: Don't forget to check the ACTIVATED box for criteria in the secondary and tertiary sorts if you want to use these additional criteria for sorting the concordance.

a. Sort the occurrences according to the word that comes immediately before *terrain* by choosing **L1** and clicking on **OK**.

    i. What collocate(s) can you identify in the contexts?

    ii. To what part(s) of speech categories do they belong?

    iii. Do these correspond to the tags that you assigned in the *Set* column?

    iv. In what ways could these collocations be useful to translators or terminologists?

b. Sort the results according to the word appearing 3 places to the left of *terrain*.

    i. What option do you need to choose to do this sort?

    ii. What collocate(s) can you identify in these contexts?

    iii. What is/are the part(s) of speech?

    iv. Do these observations correspond to the tags that you assigned in the *Set* column?

    v. In what ways could these observations be useful to translators or terminologists?

c. Did you identify other relevant collocates of *terrain*? Can you figure out a way to sort the results in order to highlight these collocates?

NOTE: Take note of the collocates that you have identified for this search term. In the follow-up WordSmith Tools Concord Tutorial and Exercises: Level II one the CERTT site (on the CERTT site (ACCESS BY TOOL NAME > WORDSMITH TOOLS > WORDSMITH TOOLS CONCORD, LEVEL II), you will learn another way to identify collocates – this time, automatically.

12. To remove any irrelevant results from the list:

a. Click on the line in question in order to select it.

b. Press the **DELETE** key to "de-activate" the line. (Note that if you change your mind, simply select the line again and press the **INSERT** key to "re-activate" it).

c. Repeat as necessary to "de-activate" any additional irrelevant results.

d. From the **EDIT** menu, choose **ZAP** (🗲) to delete the "de-activated" lines.

NOTE: You can also access the ZAP function using the keyboard shortcut CTRL + Z. You will probably recognize this shortcut which is used by default in Microsoft applications to "undo" a previous action. It is important not to use CTRL + Z in WordSmith unless you are zapping unwanted lines.

13. To close the concordance without saving it, go to the **FILE** menu and choose **CLOSE**, then click on the **NO** button when the system asks you whether you want to save the file.

Wrapping up

1. To make a copy of your files as a backup or to transfer them to another computer outside the Writing Centre:

a. In **MY COMPUTER** or from the **START** menu, find the sub-directory you created to store the files for this exercise.

b. Make a compressed folder that contains this sub-directory. (For instructions, see the file *Create a compressed folder* on the CERTT site (**ACCESS BY TOOL NAME > WINDOWSXP > CREATING A SUB-DIRECTORY**).)

c. Copy this compressed folder to a USB key, or if it is less than 2 MB, send a copy as an attachment to your e-mail.

Questions for reflection

1. As you did these exercises, what did you notice about how Concord works? What do you think of the interface? The search options? The display options? The processing options?

2. What could Concord help you to do? In what kind of situation?

3. What are some of the advantages and disadvantages of using Concord to search a corpus or text collection? Compared to a manual approach? Compared to using another tool?

4. What are some of the main challenges associated with using a concordancer? Do these challenges impact the way in which you perform searches or analyze results?

5. What criteria can be used to evaluate Concord?

6. How does Concord compare to other monolingual concordancers that you have used?