



TextSTAT Concordancer Exercise: Level I

TextSTAT is a free corpus analysis tool that you can download from the Web site set up by its creator, Matthias Hüning of the Free University of Berlin (<http://www.niederlandistik.fu-berlin.de/textstat/software-en.html>). TextSTAT can help you to build corpora from files stored on your computer or directly from the Internet, and offers a function that can create wordlists as well as a KWIC concordancer.

Doing these exercises will help you learn to:

- [Install TextSTAT](#);
- [Build a corpus from texts stored on your computer](#);
- [Do simple and complex searches using proximity operators](#);
- [Modify the presentation of the search results](#); and
- [Keep a record of your searches by exporting the results as a Word document](#).

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 the compressed folder that contains Supreme Court decisions in English, in .txt format from the CERTT site (**CORPORA AND OTHER RESOURCES > SUPREME COURT OF CANADA DECISIONS > SUPREME COURT DECISIONS (ENGLISH, TXT)**).

NOTE: TextSTAT can process files in Word (.doc), HTML (.htm/.html) or Plain text (.txt) format.

- 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 > EXTRACTING FILES FROM COMPRESSED FOLDERS**).)

Installing TextSTAT

1. Open the Web page that describes TextSTAT (<http://www.niederlandistik.fu-berlin.de/textstat/software-en.html>) and read the introduction to the tool.
2. Download to your U: drive (also called *My Documents*) the compressed folder that contains TextSTAT version 2.8c for Windows by clicking on the appropriate link. Note the name of the file.
3. Open your U: drive (*My Documents*) from the shortcut that appears on the Desktop of your computer, and double-click on the compressed folder you downloaded to automatically extract all the files it contains. Accept the default location for the files by clicking the **OK** button in the dialogue box that appears,



which will automatically create a sub-directory of your U: called *TextSTAT* – 2 in which the files will be stored.

4. To open the program, go to this new sub-directory, find the main program file (which is called *TextSTAT.exe* and indicated by an  icon), and double click on it. The software will open and is ready for use.

NOTE: To use TextSTAT in future, you can simply find this file again and double-click on it to open the software.

If you are working on your own computer, you can also create a shortcut on the Desktop to make accessing TextSTAT even easier. To do this:

- Right-click on the main program file.
- From the contextual menu that appears, choose the Create a Shortcut option. A Windows shortcut appears at the end of the list of files in the sub-directory.
- Select this shortcut by clicking on it, cut it (CTRL + X) and then paste it (CTRL + V) onto the Desktop of your computer.

Choosing files to analyze

1. Create a new corpus (i.e. a list of files):

NOTE: It is important to note that you must start by first creating an empty list, to which you will later add files.

- a. From the **CORPUS** menu, choose the **NEW CORPUS** option.
 - b. In the dialogue box that opens, type a name for the new corpus and choose the location where you want to store it: the sub-directory you created at the beginning of these exercises.
 - c. Click the **SAVE** button to save the new, empty corpus you have created.
2. Select the files to include in the corpus:
 - a. From the **CORPUS** menu, choose the **ADD LOCAL FILE** option.
 - b. In the dialogue box that opens, choose the Supreme Court decision files you extracted from the compressed folder to analyze and click the **OPEN** button to add them to the corpus file list. (You can select several files at once by selecting the first file, holding down the **SHIFT** key on the keyboard and then selecting the last file.)
 3. Save the definition of the corpus (i.e. the list of files):
 - a. From the **CORPUS** menu, select the **SAVE CORPUS** option.

Creating concordances

1. Click the **CONCORDANCE** button.
2. Do a simple search for the string *charge* in the corpus:
 - a. In the search field that appears at the top of the TextSTAT window, type *charge*.
 - b. Click the **SEARCH** button.
 - c. A list of occurrences of the string appears in the left-hand pane of the window. The status bar at the bottom of the TextSTAT window lists the total number of occurrences found. Take a look at the results.

- d. To see a longer context for one of the lines, click on the line to select it, then right-click and from the contextual menu that appears, select the **SHOW MORE CONTEXT** option. A longer segment of text is shown on the *Citation* tab. Once you have read the context and identified the other information provided, click again on the **CONCORDANCE** tab to go back to the results.
3. Adjust the search settings:
 - a. Check the **SEARCH WHOLE WORDS ONLY** check box to search only for the exact string you entered in the search field (as opposed to an undefined character string that contains your search string, at the beginning, middle or end).
 - b. Check the **SEARCH CASE INSENSITIVE** check box to search for the search string without taking case (i.e. the use of capital or lower-case letters) into account (instead of searching only for the string with exactly the same capitalization you used when you entered it).
 - c. Redo the search to see the difference in the results.
 - d. Evaluate this difference. *How could these functions help a translator or other language professional to target pertinent information more easily? In what kinds of situations?*
 - e. *Can you think of any drawbacks of using these search functions?*
 4. In the right-hand pane of the TextSTAT window, adjust the concordancer's display settings:
 - a. **MARK SEARCH STRING** allows you to display the search string in capital letters.
 - b. The context settings (*context left* and *context right*) allow you to specify the number of characters to display before and after the search string.
 - c. The options below the context settings allow you to sort the results by the string found, by the first word form to its left, or the first word form to its right.
 - d. Modify some of these settings, and click the **REFRESH** button to see the new format of the results.
 - e. Evaluate the changes in the display. *How could these functions help a translator or other language professional to identify different types of pertinent information?*
 5. Copy a line from the concordance:
 - a. Select a line from the results by clicking on it.
 - b. Click the **COPY TO CLIPBOARD** () button.
 - c. Paste the line into a word processor document, terminology record, or other document. (You can generally do this using the shortcut **CTRL + V**.)
 6. Do a complex search using a proximity operator (i.e. a search for contexts that contain two strings within a certain distance of one another).
 - a. Click the **QUERY EDITOR** button beside the search field.
 - b. In the first search field in the dialogue box that appears, type *appeal*.
NOTE: You can use wildcards (? or *) to replace one or more characters if you want. As usual, ? replaces 0 or 1 character, and * 0 to many.
 - c. In the second search field, type *hear**.
 - d. From the lists that appear below the fields, choose a minimum number of 0 and maximum number of 5 words to allow between the two strings (to search for occurrences of the words occurring beside one another or separated by 1-5 words).

- e. Click the **START SEARCH** button and then **CLOSE** button to return to the main results window.
7. In the results window, look at how TextSTAT has automatically transformed your search into a regular expression (a formal convention used for example to represent strings that include certain types of variations). **Can you identify how the various elements of the query you entered have been converted into the different parts of the regular expression?**
 - a. Reopen the Query editor, reverse the order of the strings, and then redo the query. **What results do you find? What does this tell you about doing complex searches using this function? What are the advantages and disadvantages of this approach?**
 - b. To find out more about searching using regular expressions, in the help menu (?) click on the **REGULAR EXPRESSIONS – HELP** option, which will take you to a Web page describing available options.
 8. Compare the default settings for the different types of searches:
 - a. Identify a lexical unit that appears in several different forms in the corpus (e.g. singular and plural, with and without capitals). One possibility is *tribunal*.
 - b. Do a simple search for the term in its base form, with the **SEARCH WHOLE WORDS ONLY** and **SEARCH CASE INSENSITIVE** check boxes checked. Analyze the results.
 - c. Redo the search, this time with the two check boxes unchecked. Compare these results to those of the previous search.
 - d. Leave these options unchecked. Click the **QUERY EDITOR** button and enter the same string in the first search field. Run the search (**START SEARCH, CLOSE**).
 - e. Look at the results and the options that are now checked. **What restrictions are imposed on searches using the Query editor? What does this tell you about the possibilities for searching using this function?**

Exporting the results

1. If you wish, export the results of your search:
 - a. From the **EXPORT** menu, select the **CONCORDANCE LIST > MS WORD** option.
 - b. A Word file containing a list of the contexts shown in your results opens automatically. Save this list (**FILE > SAVE AS...**) in the sub-directory that you created in your U at the beginning of this tutorial, giving it the filename of your choice.

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 COMPRESSED FOLDER**).
 - 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. Did you think TextSTAT was easy to use?
2. What did you think about the basic functions it offers?
3. How do you think TextSTAT can help a user to analyze corpora/text collections? What tasks can it help to carry out?
4. What criteria could one use to compare TextSTAT to other monolingual concordancers (e.g. the full text search function in LogiTerm, or WordSmith Tools)? Compare TextSTAT to other concordancers you have used according to these criteria.