**Exploring key technical vocab with the help of corpus query software**

1. **Create a text corpus**

A text corpus is a (preferably) large collection of textsthat can used for linguistic analysis.

* select documents pertaining to your specialised field
* save them in one word file for easy access

1. **Use the corpus query software**

* download **AntConc** (<https://laurenceanthony.net/software.html> )
* upload your file(s) by opening AntConc and clicking on **File** and then **Open File**
* select **Word** (list) file tab on the top right
* click on **Start** button at the bottom of the screen
* search the word list for field-specific, take a note of 8-10 frequent words

The word list function shows you the total number of occurrences of each word. The most frequent words are generally function words but the content words are of interest as they reflect your specialized field.

- select **KWIC** tab at the top left, type one of your words in **Search Query** at the bottom left

- (by clicking on the blue words in the center of each line you can view the fill co-text > go back)

- go to **Sort Options** at the bottom and see what happens when you sort to the left (1L-2L-3L) and to the right (1R, 2R, 3R)

By analysing which words co-occur with your word, you can learn about its lexico-grammatical behaviour. That means you can find out which words occur more frequently together with your word. These words are called collocates.

Language comes in chunks and this is also how we produce and process language. Learning and using chunks of language, called collocations, makes you a proficient user of English.

This is how you can keep a note of your analysis:

|  |  |  |  |
| --- | --- | --- | --- |
| **word** | **meaning, definition or translation** | **lexico-grammatical behaviour** | |
| **data** |  | real-time **data**  health **data**  healthcare **data**  sensor-generated **data** | **data** analysis  **data** analytics  **data transmission**  (ensure) data privacy |
| **data** for (i.e. treatment decisions)  transmit **data** to (i.e. healthcare teams)  collect/gather **data** on (i.e. aspects of human health  **data**-driven | |