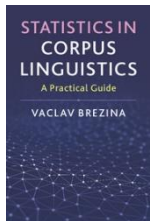


# Exploring Collocations with #LancsBox v. 4

Vaclav Brezina, Dana Gablasova | 7 December 2018

#LancsBox is a free tool available at: <http://corpora.lancs.ac.uk/lancsbox>

You can find detailed information about quantitative analysis of corpora and statistics (including many linguistic examples and case studies) in:



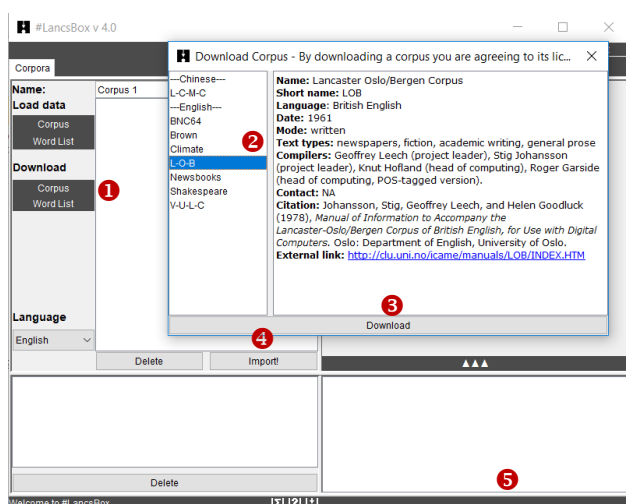
Brezina, V. (2018). [Statistics in Corpus Linguistics: A Practical Guide](#). Cambridge: Cambridge University Press.

## 1 Starting with #LancsBox: Load your own corpus or existing corpora

The starting point of the tool is the 'Corpora' tab where data is loaded and further processed (POS-tagged etc.). There are two options for loading corpora: i) loading local corpora from your computer and ii) downloading corpora provided by #LancsBox.

**T** **Task 1. Import the LOB corpus.** Start #LancsBox. Follow the instructions below; the steps are highlighted in the image.

- 1 On the Corpora tab under 'Download' click on 'Corpus'.
- 2 From the list select 'L-O-B'.
- 3 Click on 'Download'
- 4 and then on 'Import'.
- 5 Wait while the corpus is being processed (notice the progress bar at the bottom of #LancsBox window).



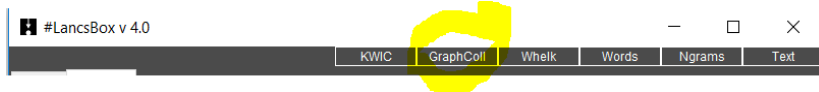
## 2 Collocations in context with #LancsBox: Collocation graphs and networks

The GraphColl tool identifies collocations and displays them in a table and a graph.

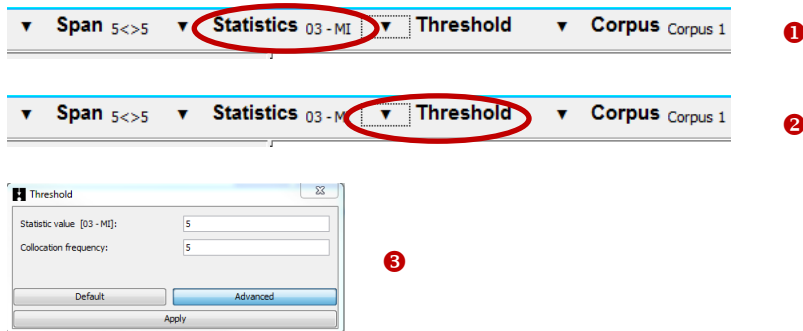
It can be used to:

- Find the collocates of a word, phrase or word class.
- Visualise the collocations of a word or phrase and identify connections between multiple words (collocation network); the graphs illustrate how closely the search term and collocations occur together, the strength of the collocation (as measured by a particular association measure) , and the position of the collocations in text in relation to the search term.

**T** **Task 2. Create graphs.** In the main menu click on 'GraphColl'. Work with the LOB corpus.



- a) Build a collocation graph around the node *good* using MI score. Change the 'Statistic' to 03-MI and 'Threshold' to MI = 5 as indicated below and then search for 'good'.



- b) How many collocates does the graph display? Are all of them useful?
- c) Which of the collocates occur predominantly to the left of the node *good* and which ones to the right? Explore different view options by clicking on the display button ( **Free** ) in the top right corner.

Left: .....

Right: .....

- d) Some of the combinations with *good* such as *supreme*, *bad* might not be obvious. Use the right-click function to obtain concordances (KWIC pop-up) and explain these collocates. Note down examples.

*supreme* .....

*bad* .....

- e) Change the settings to lemma ( **Lemma** ) as the unit in the collocation graph. Search for 'good' again. Click through the display button in the top right corner ( **Word Class** ) until a colour-coded graph appears.

SEE NEXT PAGE

