Doing corpus linguistics with #LancsBox

In these tasks, we will become familiar with some of the functions of #LancsBox by investigating features of academic prose in written L2 English.

Tasks 1 - 2 will focus on practicing different types of **searches** using the KWIC tool.

Tasks 3 - 5 will introduce how to create and change the settings of **collocation graphs** using the GraphColl tool.

There are also **optional tasks** you can try during or after the webinar.

We will be using the VU Lancaster corpus of student academic writing (VULC): a corpus of L2 English student essays from VU Amsterdam.

Task 1. Searches. Go to the KWIC tool in #LancsBox and search for the following expressions in the VULC corpus (provided with #LancsBox). Note down their frequencies and distributions in texts.

Type of search	Search term	Occurrences (per 10k)	Number of texts
Simple	however		
Simple	but		
Phrase	according to		
Wildcard	influence*		
Smart Search	NOMINALIZATIONS		
Regex	/however|but/		
Regex	state [as headword] V* [as POS]		

Optional task: You are researching how reporting verbs are used in L2 English writing. How might you build one query to search for *argue*, *claim* and *state* simultaneously? What are some issues you might face?

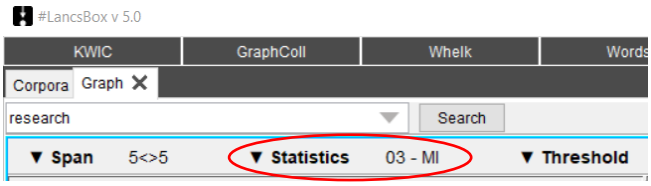
Task 2. Applying filters. Still in the KWIC tool, search for the following expressions and apply filters. Note down their frequencies and distribution in texts.

Search term	Filter	Occurrences (per 10k)	Number of texts
VERBS	however [anywhere LEFT]		
should	be [in R1 position]		

Task 3. Create a collocation graph and change settings. Go to the GraphColl tool, follow the directions and note down the frequencies.

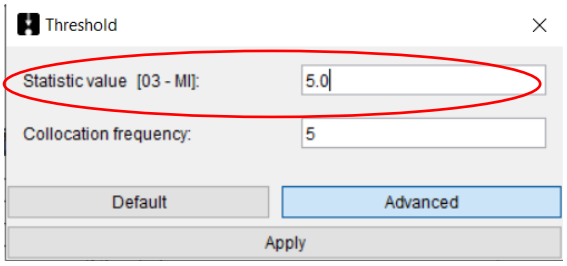
(a) Build a collocation graph by conducting a simple search for *research.* What results did you get?

(b) Change the statistical association measure to MI score. This will delete the current graph. Search for *research* again to create a new graph.



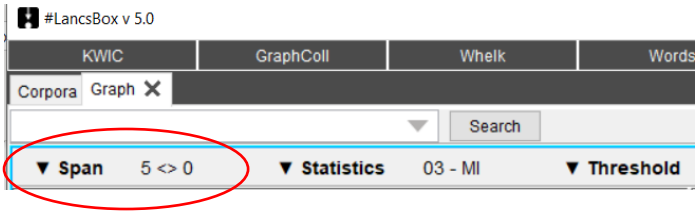
How has the graph changed?

(c) Open the threshold settings and change the statistic value to 5.0 (3 and higher is considered strong for the MI score equation). Search for *research* again.



How has the graph changed now?

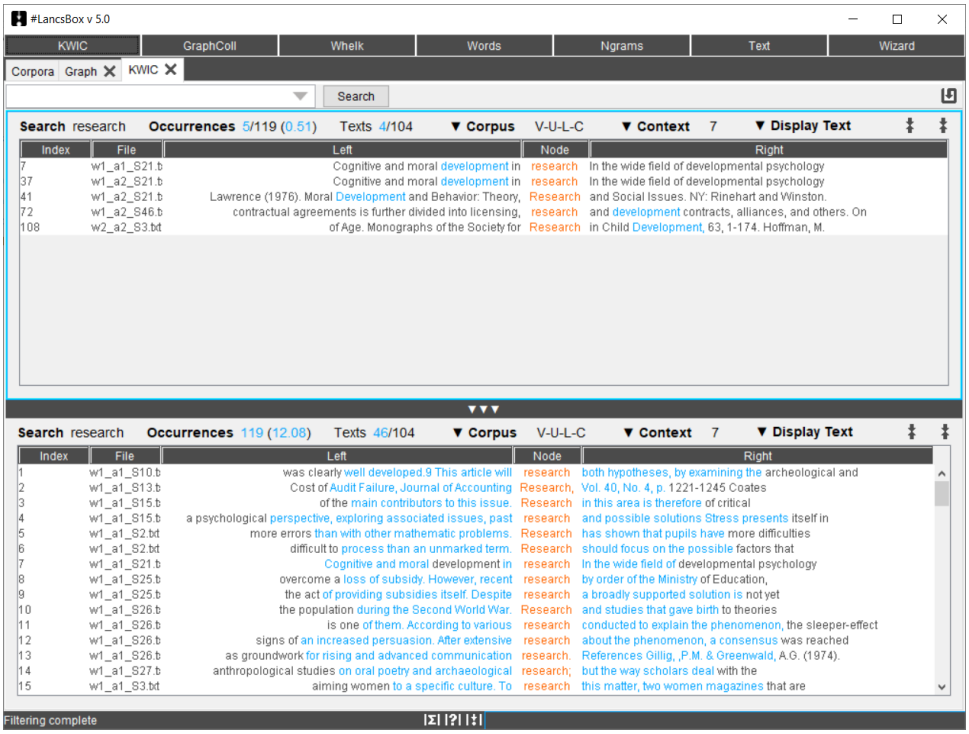
(d) Change the window span to search for five words to the left and zero words to the right (5L, 0R). Search for *research* again, keeping the same settings from the last steps.



How many collocates are there now?

Optional task: You are researching what collocations L2 English speakers use when writing about language. How might you start this study using GraphColl? What settings would you consider changing, and why? Why might you use a narrower or wider collocation window?

Task 4. Combining GraphColl and KWIC view to explore collocation context. To help interpret graphs, GraphColl offers a concordance function, which displays examples of collocate use (KWIC). To display the concordance lines of a given collocation, **right click** on the collocate in the table or in the graph. These concordance lines can be expanded into the full-screen KWIC view by clicking on three arrows (▲▲▲) at the top right. In the top panel, the full-screen view displays examples of the selected collocate co-occurring with the node; in the bottom panel, all other competing candidates for collocates are displayed.



Search for *research* again and right click on the collocate *stress.* Explore the context in which it occurs. Comment on the association between *stress* and *research* in this corpus. (N. B. remember to change your settings back to a Span of 5L, 5R.)

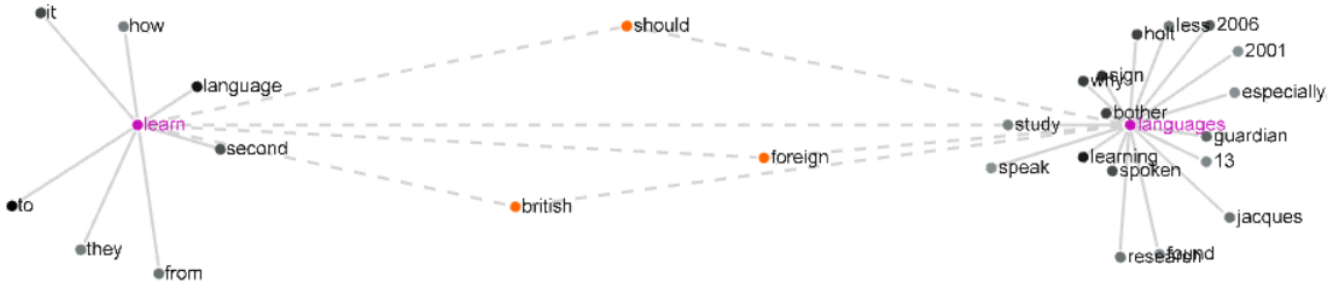
Task 5. Build collocation networks. GraphColl also offers the function to explore second-order collocates through building collocation networks. These are visualisations that help you explore language patterns.

(a) Search for *learn* to create a new graph (use MI score and settings: span 5L, 5R; Statistic value = 5.0; Collocation frequency = 5).

(b) Find the collocate *languages* in the graph (or table) and double click on it. This should create a collocation network similar to the one below. (N. B. you can zoom into a graph using your mouse wheel or change the text size using Ctrl +/-)

|<----------second-order collocation network----------|

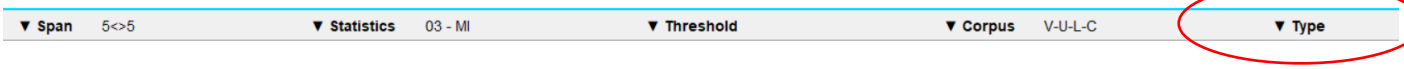
|<----first-order collocation network---->|



|<----------shared collocates---------->|

(c) Comment on the connection between the collocates. Feel free to explore the network further by clicking on any of the collocates.

Optional task: You want to investigate collocates of both *language* and *languages*. To do this, change the unit setting from type to lemma.



How does the graph change? Explore further by changing the span, statistics or threshold settings as you like.

#LancsBox v.5 manual: http://corpora.lancs.ac.uk/lancsbox/docs/pdf/LancsBox_5.0_manual.pdf