

## 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.

**T** **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]		

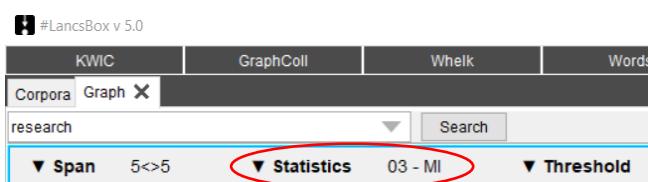
**Optional task:** You are researching how reporting verbs are used in L2 English writing. How might you build one query to search for *argue*, *claim* and *state* simultaneously? What are some issues you might face?

**T** **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]		

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

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

---

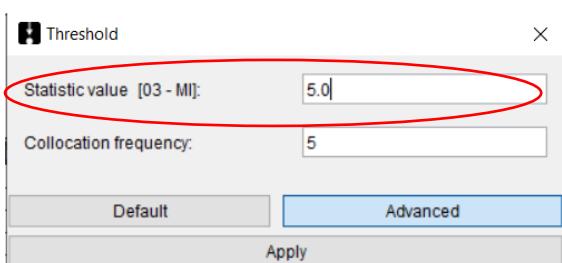


---



---

(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?

---

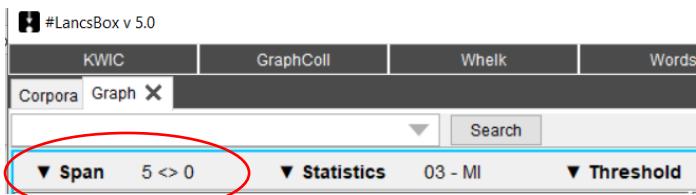


---



---

(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?

---



---

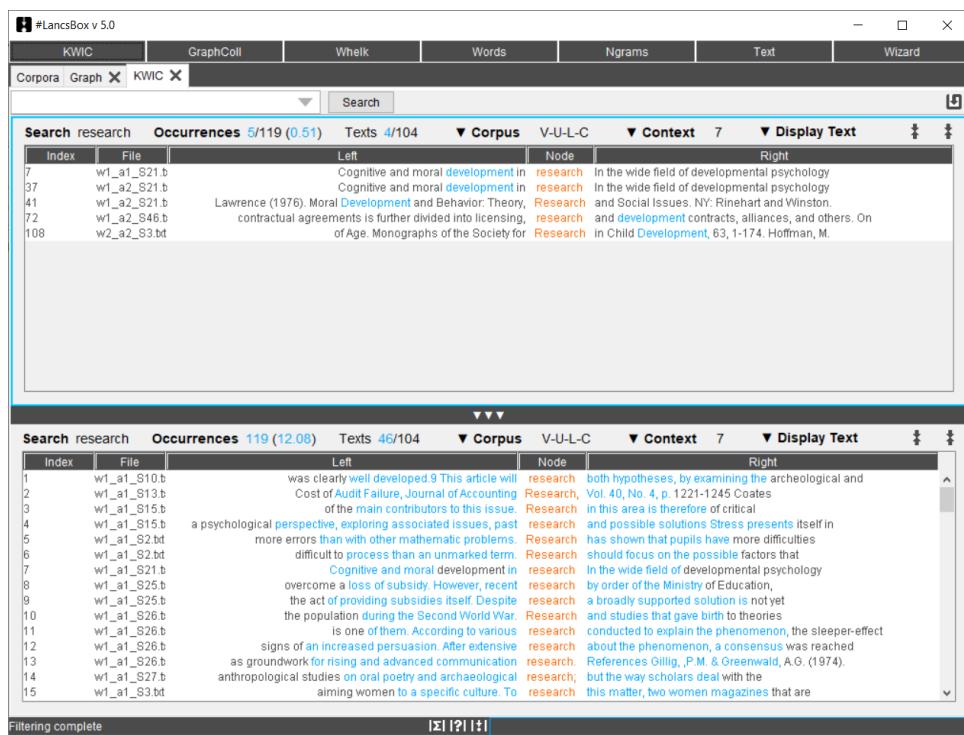


---

**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.



The screenshot shows two search results tables in the #LancsBox v 5.0 software. The top table is for the search term 'research' with 5/119 occurrences (0.51). The bottom table is for the search term 'research' with 119 occurrences (12.08). Both tables include columns for Index, File, Left, Node, and Right, with the Right column showing the context of the word 'research' in the texts.

Index	File	Left	Node	Right
7	w1_a1_S21.b	Cognitive and moral development in	research	In the wide field of developmental psychology
37	w1_a2_S21.b	Cognitive and moral development in	research	In the wide field of developmental psychology
41	w1_a2_S21.b	Lawrence (1976). Moral Development and Behavior: Theory	Research	and Social Issues. NY: Rinehart and Winston.
72	w1_a2_S46.b	contractual agreements is further divided into licensing,	research	and development contracts, alliances, and others. On
108	w2_a2_S3.b	of Age. Monographs of the Society for	Research	in Child Development, 63, 1-174. Hoffman, M.

Index	File	Left	Node	Right
1	w1_a1_S10.b	was clearly well developed. 9 This article will	research	both hypotheses, by examining the archaeological and
2	w1_a1_S13.b	Cost of Audit Failure, Journal of Accounting	Research	Research, Vol. 40, No. 4, p. 1221-1245 Coates
3	w1_a1_S15.b	of the main contributors to this issue.	Research	Research in this area is therefore of critical
4	w1_a1_S15.b	a psychological perspective, exploring associated issues, past	research	and possible solutions Stress presents itself in
5	w1_a1_S2.b	more errors than with other mathematical problems.	Research	research has shown that pupils have more difficulties
6	w1_a1_S2.b	difficult to process than an unmarked term.	Research	should focus on the possible factors that
7	w1_a1_S21.b	Cognitive and moral development in	research	In the wide field of developmental psychology
8	w1_a1_S25.b	overcome a loss of subsidy. However, recent	research	by order of the Ministry of Education,
9	w1_a1_S25.b	the act of providing subsidies itself. Despite	research	a broadly supported solution is not yet
10	w1_a1_S26.b	the population during the Second World War.	Research	and studies that gave birth to theories
11	w1_a1_S26.b	is one of them. According to various	research	conducted to explain the phenomenon, the sleeper-effect
12	w1_a1_S26.b	signs of an increased persuasion. After extensive	research	about the phenomenon, a consensus was reached
13	w1_a1_S26.b	as groundwork for rising and advanced communication	research	References Gillig, P.M. & Greenwald, A.G. (1974).
14	w1_a1_S27.b	anthropological studies on oral poetry and archaeological	research;	but the way scholars deal with the
15	w1_a1_S3.b	aiming women to a specific culture. To	research	this matter, two women magazines that are

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.)

---



---

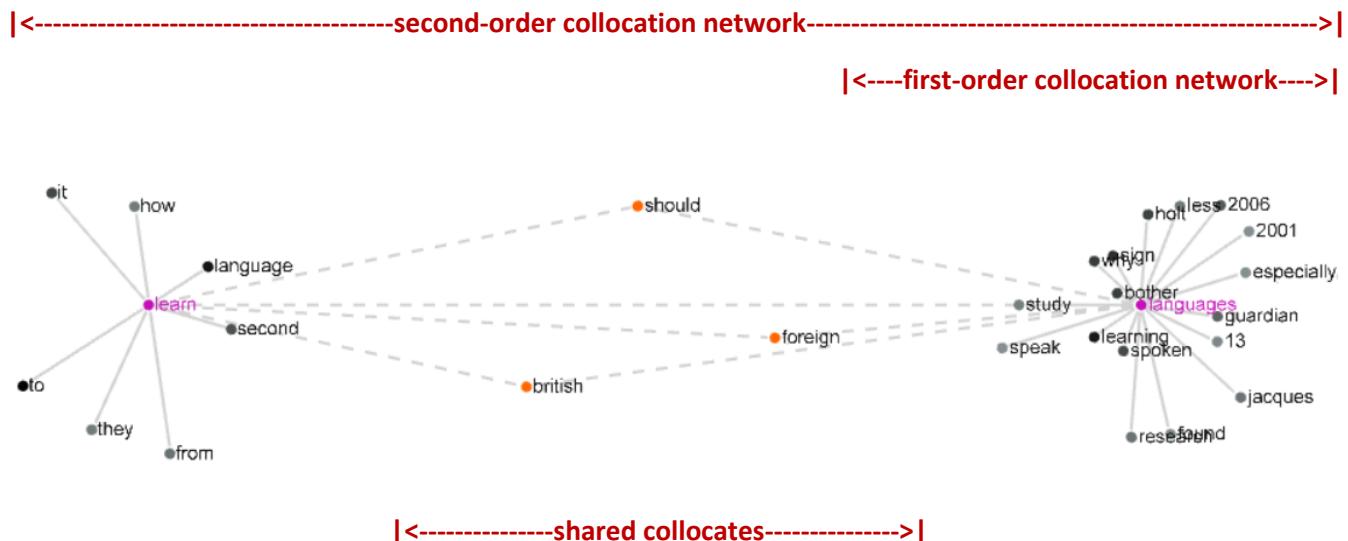


---



**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.

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



(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.

▼ Span 5<>5 ▼ Statistics 03 - MI ▼ Threshold ▼ Corpus V-U-L-C ▼ Type

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](http://corpora.lancs.ac.uk/lancsbox/docs/pdf/LancsBox_5.0_manual.pdf)