Doing corpus linguistics with #LancsBox

In these tasks, we will become familiar with some of the functions of #LancsBox by investigating features of modern-day British English using the BNC2014 Baby+.

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.

Tasks 6 - 7 will introduce how to **create a web corpus automatically** (a brand-new feature for #LancsBox v.6).

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

We will be using the **British National Corpus 2014 Baby+ edition;** this corpus is a four-million-word balanced subset of the BNC2014. Further details can be found here: http://corpora.lancs.ac.uk/lancsbox/docs/pdf/BNC2014Baby.pdf

Task 1. Searches. Go to the KWIC tool in #LancsBox and search for the following expressions in the BNC2014 Baby (provided with #LancsBox). Note down their frequencies and distributions in texts.

Type of search	Search term	Occurrences (per 10k)	Number of texts
Simple	tea		
Simple	weather		
Phrase	mug of tea		
Wildcard	rain*		
Smart Search	DOWNTONER		
Regex	/mate|friend/		
Regex	nick [as headword] V* [as POS]		

Optional task: You are researching how 'hailnames' (informal forms of address) are used in British English. How might you build one query to search for *pal*, *mate* and *buddy* simultaneously? What are some issues you might face?

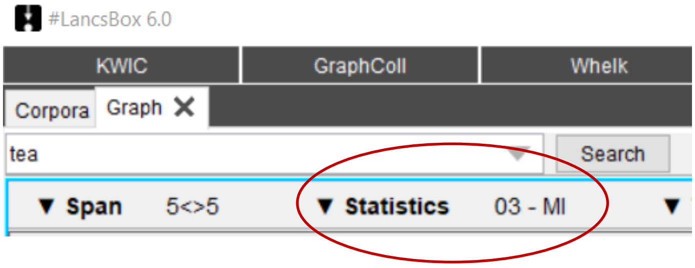
Task 2. Applying filters. Still in the KWIC tool, search for the following expressions and apply filters. Note down their frequencies and distribution in texts.

Search term	Filter	Occurrences (per 10k)	Number of texts
tea	make [anywhere LEFT]		
fish	chips [in R2 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 *tea.* What results did you get?

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



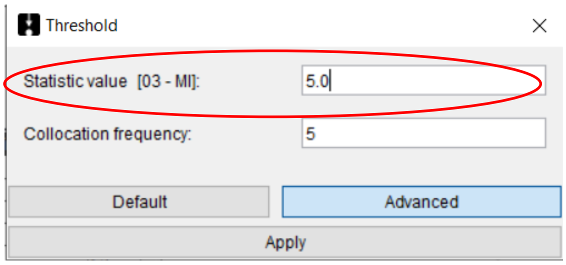
tea

Span 5<>5

Statistics 03 - MI

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 *tea* again.

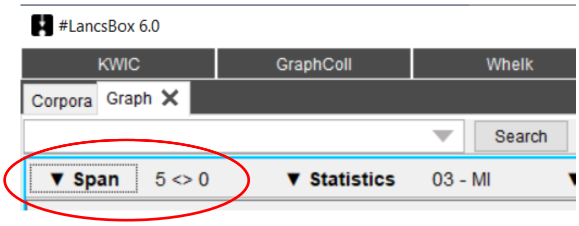


Statistic value [03 - MI]: 5.0

Collocation frequency: 5

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 *tea* again, keeping the same settings from the last steps.



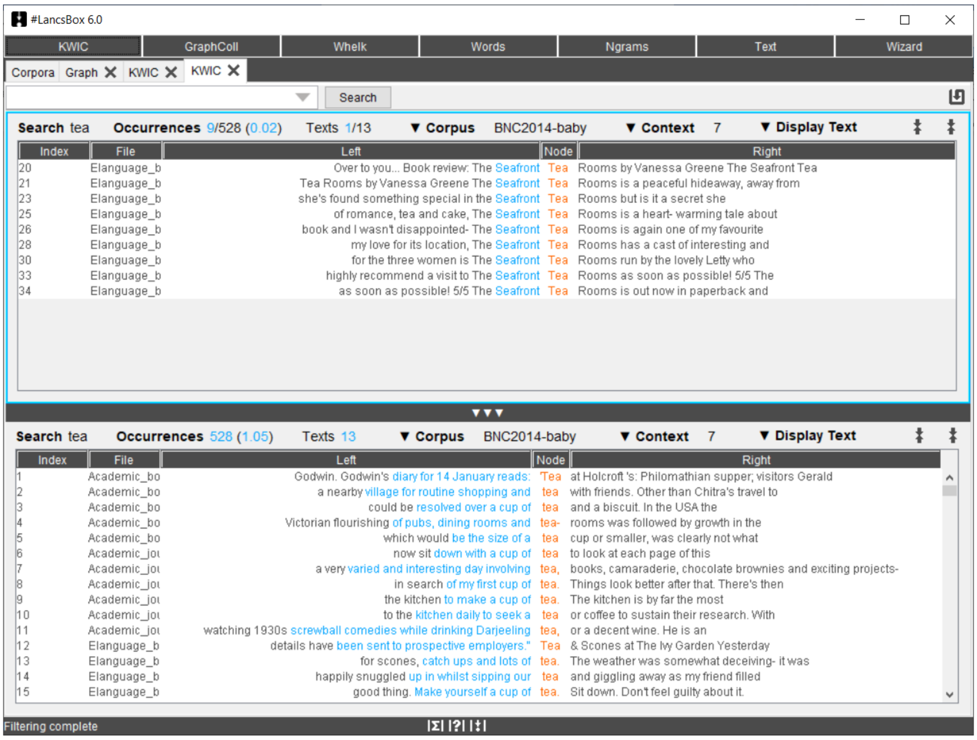
Span 5 <> 0

Statistics 03 - MI

How many collocates are there now?

Optional task: You are researching what collocations British English speakers use when talking about the weather. 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 collocates are displayed.



#LancsBox 6.0

KWIC | GraphColl | Whelk | Words | Ngrams | Text | Wizard

Corpora | Graph ✕ | KWIC ✕ | KWIC ✕

Search

Search tea **Occurrences** 9/528 (0.02) Texts 1/13 ▼ **Corpus** BNC2014-baby ▼ **Context** 7 ▼ **Display Text**

Index	File	Left	Node	Right
20	Elanguage_b	Over to you... Book review: The Seafront	Tea	Rooms by Vanessa Greene The Seafront Tea
21	Elanguage_b	Tea Rooms by Vanessa Greene The Seafront	Tea	Rooms is a peaceful hideaway, away from
23	Elanguage_b	she's found something special in the Seafront	Tea	Rooms but is it a secret she
25	Elanguage_b	of romance, tea and cake, The Seafront	Tea	Rooms is a heart- warming tale about
26	Elanguage_b	book and I wasn't disappointed- The Seafront	Tea	Rooms is again one of my favourite
28	Elanguage_b	my love for its location, The Seafront	Tea	Rooms has a cast of interesting and
30	Elanguage_b	for the three women is The Seafront	Tea	Rooms run by the lovely Letty who
33	Elanguage_b	highly recommend a visit to The Seafront	Tea	Rooms as soon as possible! 5/5 The
34	Elanguage_b	as soon as possible! 5/5 The Seafront	Tea	Rooms is out now in paperback and

Search tea **Occurrences** 528 (1.05) Texts 13 ▼ **Corpus** BNC2014-baby ▼ **Context** 7 ▼ **Display Text**

Index	File	Left	Node	Right
1	Academic_bo	Godwin. Godwin's diary for 14 January reads:	'Tea	at Holcroft 's: Philomathian supper; visitors Gerald
2	Academic_bo	a nearby village for routine shopping and	tea	with friends. Other than Chitra's travel to
3	Academic_bo	could be resolved over a cup of	tea	and a biscuit. In the USA the
4	Academic_bo	Victorian flourishing of pubs, dining rooms and	tea-	rooms was followed by growth in the
5	Academic_bo	which would be the size of a	tea	cup or smaller, was clearly not what
6	Academic_jou	now sit down with a cup of	tea	to look at each page of this
7	Academic_jou	a very varied and interesting day involving	tea,	books, camaraderie, chocolate brownies and exciting projects-
8	Academic_jou	in search of my first cup of	tea.	Things look better after that. There's then
9	Academic_jou	the kitchen to make a cup of	tea.	The kitchen is by far the most
10	Academic_jou	to the kitchen daily to seek a	tea	or coffee to sustain their research. With
11	Academic_jou	watching 1930s screwball comedies while drinking Darjeeling	tea,	or a decent wine. He is an
12	Elanguage_b	details have been sent to prospective employers."	Tea	& Scones at The Ivy Garden Yesterday
13	Elanguage_b	for scones, catch ups and lots of	tea.	The weather was somewhat deceiving- it was
14	Elanguage_b	happily snuggled up in whilst sipping our	tea	and giggling away as my friend filled
15	Elanguage_b	good thing. Make yourself a cup of	tea.	Sit down. Don't feel guilty about it.

Filtering complete

Keeping the same settings, search for *tea* and right click on the collocate *seafront*. Explore the context in which it occurs. Comment on the association between *tea* and *seafront* in this corpus.

____

____

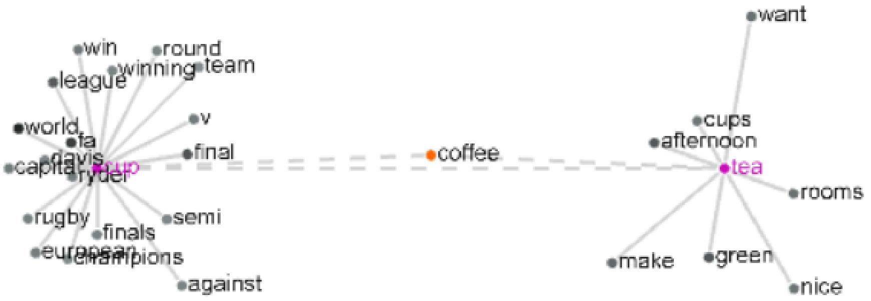
____

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 *tea* to create a new graph (use MI score and settings: span 5L, 5R; Statistic value = 5.0; Collocation frequency = 15).

(b) Find the collocate *cup* 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-->|

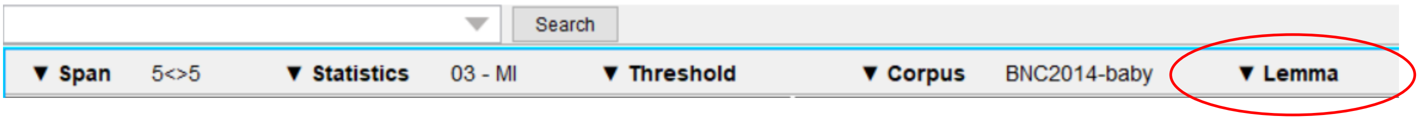


|<------first-order collocation network------>|

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

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

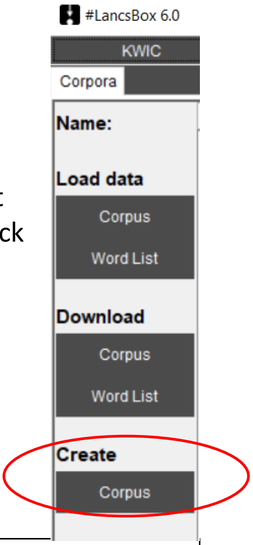
Optional task: You want to investigate both collocates *cup* and *cups*. To do this, change the unit setting from type to lemma.



▼ Span 5<>5 ▼ Statistics 03 - MI ▼ Threshold ▼ Corpus BNC2014-baby ▼ Lemma

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

Task 6. Create a web corpus automatically. #LancsBox v.6 offers a new feature, which allows automatic download of texts from the web.



(a) Locate your initial (seed) website that includes the landing page about your chosen topic.
(b) On 'Corpora' tab, click on 'Corpus' under create (see image)
(c) Paste the URL into the URL box and click on 'Create corpus'
(d) Wait for the process to finish – by default #LancsBox is downloading and converting into text 100 websites at 2 levels of embedding – and explore the log and the folder with the data. Click 'OK'.
(e) Click on 'Import' to import and automatically POS-tag your corpus.

Task 7. Explore and search your corpus using #LancsBox.

(a) Explore the size of your corpus and note it down:

Tokens (running words):	
Types (different words):	
Lemmas (headword + POS category):	

(b) Explore your corpus using relevant search terms:

Search term	**Occurrences (per 10k)**	**Number of texts**

#LancsBox v.6 manual: http://corpora.lancs.ac.uk/lancsbox/docs/pdf/LancsBox_6.0_manual.pdf