#LancsBox X

GraphColl

Finding collocates

In this task, you will practice finding collocates and interpreting interpreting collocation statistics. Go to the GraphColl tool in #LancsBox X, select the BNC2014 corpus, and search for the expressions in the table below. Note down top collocates according to different association measures and the colloction frequency.
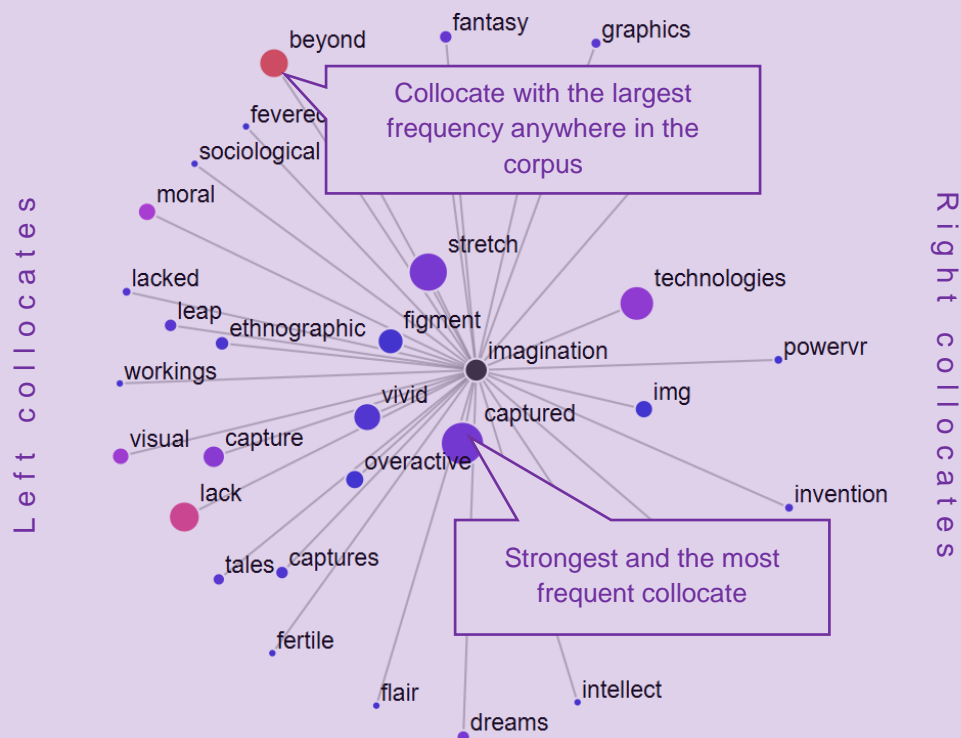
Tip: By default, collocates are sorted and filtered by log Dice. Remove filter from the log Dice column and sort collocates according to [MI ▼] by clicking on this column. A triangle indicates that this column is sorted. For identifying the most frequent collocates use the following column [Freq. (collocation) ▼]

[Freq. (collo...▼] [Freq. (corp...] [Log Di ▼] [MI]

Search term	Top log Dice collocate	Top MI collocate	Most frequent collocate
alcohol	consumption	polyvinyl	consumption
sex	allegations	tantric	with
[hw="drug" pos="N.*"]	alcohol	non-steroidal	or
FOOD	and	100g	the

Collocation graph

A collocation graph shows the relationship between a node, which is in the middle of the graph, and its collocates, which are displayed around the node. The closer the collocate is to the node, the stronger the association. The position of the collocates indicates the position in the text, before or after the node, while the size of the collocate reflects the frequency of co-occurrence. Finally, the colour indicates the frequency of the word anywhere in the corpus on the scale from blue (small) to red (large).

Left collocates — Right collocates

fantasy   graphics
beyond
Collocate with the largest frequency anywhere in the corpus
fevered
sociological
moral
stretch   technologies
lacked
leap   figment
ethnographic   imagination   powervr
workings
vivid   img
visual   capture   captured
overactive
lack   invention
Strongest and the most frequent collocate
tales   captures
fertile
flair   intellect
dreams

2

Task 2   **Interpreting collocation graphs**

This task demonstrates how corpus data can be used to identify and visualise important associations in language and discourse. You will be using BNC2014.

2a   What associations can you think of when you see the word 'alcohol'? Write down at least five.

_____

_____

Now check your answers using corpus data:

2b   In GraphColl, search for `alcohol` and create a collocation graph using log Dice as an association measure (default setting). How many of the associations listed under 2a appear in the graph? Which ones are these?
==(e.g. consumption, smoking, tobacco, alcohol, drugs, drinking, intake, drug, abuse, misuse etc.)==

_____

2c   Now change the unit setting from 'word (lowercase)' to 'lemma'. How did this change the graph?

Labels are added to show the POS for each lemma. Terms such as 'drink', 'abuse' and 'ban'
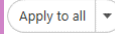_____
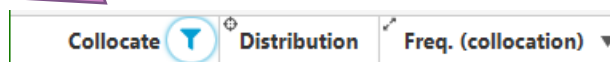now appear as separate items in the graph and table.
_____

Tip: To select 'lemma' as the unit, go to the top bar and in the drop-down find 'lemma'. Other options, such as pos, sem etc. are also possible.

BNC2014 2.0 CLAWS7	whole corpus ▼	102M	lemma ▼

2d   Apply filter `_N` to the collocate column to modify graph that you have created in 2c. This means that only lemmas, which are nouns will be shown. Sort the table according to 'Freq (collocation)'. What are the five most frequent nouns (concepts) associated with alcohol? How often do they occur with 'alcohol'?

==drugs (274) , consumption (198) , alcohol (110) , use (103) tobacco (96)==

Tip: Hover your mouse over the 'Colocate' and click on the filter icon ▼. Type in '_N' and click on the checked box icon ☑ . Finally, click on the Apply to all ▼ button.

Collocate ⊕	Distribution ↗	Freq. (collocation) ▼

Tip: Click on 'Freq (collocation)' to sort the table according to the most frequent collocate. If you wish the graph to show most frequent collocates close to the node, right-click on 'Freq (collocation)' and assign it ◄ Edge length .

Task 3	Creating and interpreting collocation networks

In this task, you will create collocation networks that show association and cross associations related to substance use and abuse. In GraphColl, use the BNC2014 and search first for `alcohol` and then for `[hw="drug" pos="N.*"]`.

3a Fill in the table below.

Search term	Collocate candidates	Collocates matching statistical criteria
`alcohol`	5,861	91
`[hw="drug" pos="N.*"]`	12,727	144

Tip: GraphColl provides detailed information about collocates. The first line indicates possible collocate candidates based on co-ocurrences (15,221). The second line indicates the number of collocates matching given statistical criteria (94)

Q **word**	Hits: **14,408 (140.83)**	Texts: **5,611/88,171**	Collocates: **15,221**
▼ "Log Dice"=numberR...			94

3b Examine shared collocates first in the graph and then in an overview table (click at the ⓘ icon to see the overview table.) How many shared collocates there are? In what contexts do they occur?

19 shared collocates. Terms related to consumption (drink, smoke),

legality/distribution (illicit, sell), specific drugs (alcohol, cannabis, booze,

cocaine), addiction (abuse, use, addiction, misuse, problem, reduce).

3c Interpret the findings from the perspective of corpus-based discourse analysis. What discourses related to alcohol and drug use and abuse have you identified in the BNC2014? What does this indicate about the British society and culture?

The terms appear to focus on drug and alcohol abuse. This indicates is a core issue

discussed in this context. A range of drugs collocate with both terms frequently including alcohol,

cannabis, and cocaine, suggesting that issues with drug abuse is not limited to one drug type.

The fact that these terms occur in British discourse may show that these issues are

acknowledged and discussed to some degree.