Task 1: Finding collocates

In this task, you will practice finding collocates and 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 collocation 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.

<table>
<thead>
<tr>
<th>Search term</th>
<th>Top log Dice collocate</th>
<th>Top MI collocate</th>
<th>Most frequent collocate</th>
</tr>
</thead>
<tbody>
<tr>
<td>alcohol</td>
<td>consumption</td>
<td>polyvinyl</td>
<td>consumption</td>
</tr>
<tr>
<td>sex</td>
<td>allegations</td>
<td>tantric</td>
<td>with</td>
</tr>
<tr>
<td>[hw=&quot;drug&quot; pos=&quot;N.*&quot;]</td>
<td>alcohol</td>
<td>non-steroidal</td>
<td>or</td>
</tr>
<tr>
<td>FOOD</td>
<td>and</td>
<td>100g</td>
<td>the</td>
</tr>
</tbody>
</table>

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).
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.

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 ‘Collocate’ and click on the filter icon . Type in ‘_N’ and click on the checked box icon . Finally, click on the button.

  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.
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.*"].

### Task 3: Creating and interpreting collocation networks

#### 3a

Fill in the table below.

<table>
<thead>
<tr>
<th>Search term</th>
<th>Collocate candidates</th>
<th>Collocates matching statistical criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>alcohol</td>
<td>5,861</td>
<td>91</td>
</tr>
<tr>
<td>[hw=&quot;drug&quot; pos=&quot;N.*&quot;]</td>
<td>12,727</td>
<td>144</td>
</tr>
</tbody>
</table>

#### 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 are there? 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.