

Practical Session: GraphColl

i You will be using **GraphColl**, a free multi-platform tool for the analysis of collocation networks.

How to cite GraphColl? GraphColl is licensed under BY-NC-ND Creative commons license and is free to use for non-commercial purposes. When you use GraphColl for presentations or publications, please cite:

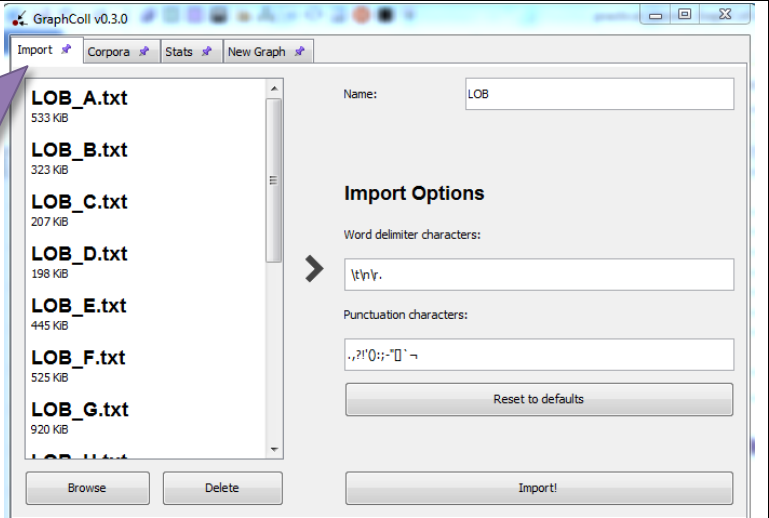
Brezina, V., McEnery, T. & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173.

T **Task 1.** Upload and explore corpora.

- a) Upload the LOB corpus to GraphColl. The corpus is available on the memory stick in the LOB folder.
- b) Check and note down the token counts for each 'corpus' under the 'Corpora' tab.

GraphColl tip:

1. Start GraphColl and go to the 'Import' tab.
2. Click on the 'Browse' button to select files.
3. Type in the corpus name and click on the 'Import!' button.



T

Task 2. Create graphs. Work with the whole LOB corpus.

- a) Build the first-order collocation network around the word *time* using MI score and the default settings.

```
return (
  s  >= 3  && // MI >= 3
  o11 >= 5 // Over 5 within-window
);
```

- b) How many collocates does the graph display? Are all of them useful?
- c) Go back to 'New Graph', select MI as the statistic, change the default settings as indicated in the figure below (MI = 5 and above) and search for the node *time* again.

```
return (
  s >= 5 && // MI >= 3
  o11 >= 5 // Over 5 within-window
);
```

How many results did you get this time?

GraphColl tip:

1. Go to the 'New Graph' tab.
2. From the drop-down menu select the required corpus.
3. From the drop-down menu select the required statistic.
4. Review and/or change the threshold settings.

s minimum value of the statistics
o11 minimum collocation frequency

Import Corpora Stats **New Graph**
Frequency Threshold:

Name: Graph 1
Open... Save as... Load Default

Corpus: **LOB**
15 files, 1023944 tokens, 46134 types
2

Window: 5 <> 5

Statistic: **01 - Freq**
3

Use adj
 Enable d

 01 - Freq
 02 - Mu
03 - MI
 04 - MI2
 05 - MI3
 06 - LogLik
 07 - Z
 08 - Dice

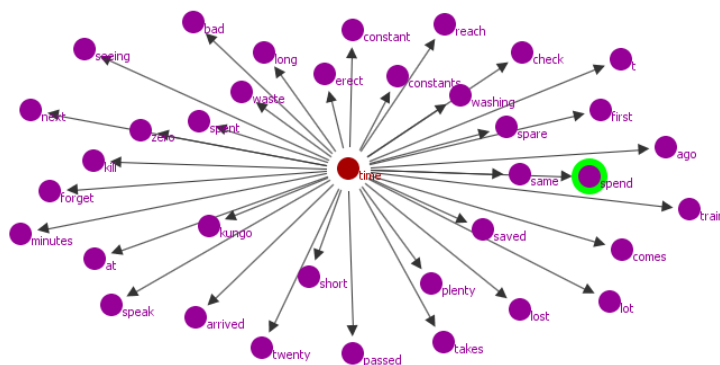
4

```
1 threshold = {
2
3
4 // Variables
5 // -----
6 // s : The value of the stati
7 // o11..o22 : Frequencies (see docs)
8 // e11..e22 : Expected values (see d
9 // r1/c1/n : Row, column, grand tot
10
11
12
13 return (
14   s >= 3 && // MI >= 3
15   o11 >= 5 // Over 5 within
16 );
17
18
```

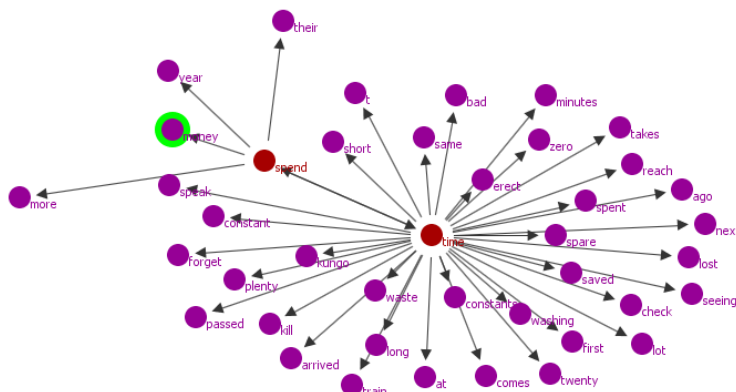
New Graph

T Task 3. Build collocation networks and explore graphs.

- a) Go to the graph you have created in Task 2 c). It should be similar to the graph displayed in the figure below:




- b) Find the collocate *spend* in the graph and double click on it. You should get a collocation network similar to the one displayed below:



- c) Find the second-order collocate *money* in the graph and double click on it. Comment on the connection between *time* and *money* that you can see in the resulting graph that shows collocates around the node *money*.

GraphColl tip:

Double clicking on a collocate creates a new branch of the collocation network treating the original collocate as a new node.



T **Task 4.** Interpret graphs.

To help interpret graphs, GraphColl offers a concordance function which displays nodes in context (KWIC).

- a) In the graph built in Task 3, create concordances for *time*, *money* and *spend* to help you interpret the collocation network.
- b) One of the strong collocates of *time* is *kungo*. What are the contexts in which this collocate occurs? Does this collocate occur across different contexts?
- c) Interpret the following collocates of *time*: *erect*, *washing*, *zero* and *t*

T **Task 5.** Compare graphs.

Compare the use of the word *time* in the academic writing, fiction and news subcorpora of the LOB corpus. Use the MI score with the following threshold settings:

```
return (
  s >= 5    && // MI >= 3
  o11 >= 5  // Over 5 within-window
);
```

T Task 6. Explore statistics.

Use different statistical measures to build collocational networks of *time* and other words of interest. Note how the collocation networks differ with different settings:

- Statistics
- Collocation window
- Thresholds

T Task 7. Compare British and American English.

- Upload both LOB and Brown to GraphColl.
- Search for *frequent* words that you think can show differences between British and American culture, e.g. *government, love, god, gun* etc.
- Compare graphs produced by GraphColl for these words in LOB and Brown. Is there an evidence for meaning differences in the discourse?



Reporting statistics

For the sake of replicability of results, all major parameters that can affect collocate identification should be reported. For this purpose, Brezina et al. (2015, forthcoming) introduce Collocation parameters notation (CPN) that captures all important parameters for collocate identification. CPN has the following format:

Statistic ID	Statistic name	Statistic cut-off value	L and R span	Minimum collocate freq. (C)	Minimum collocation freq. (NC)	Filter
4b	MI2	3	L5-R5	5	1	function words removed
4b-MI2(3), L5-R5, C5-NC1; function words removed						

} Example