

## UFA tool instructions

This tool can be used to:

- Perform Usage fluctuation analysis on corpus historical data.
- Automatically compare collocation use across time.
- Categorise collocates as consistent, initiating, terminating and transient.
- Visualize usage divergence through a 'peaks and troughs' graph.

Instructions:

- 1) Prepare dataset capturing collocation use across time.** Each text file represents a particular period. The example below is taken from the EEBO corpus analysed using Lancaster CQPweb.

```

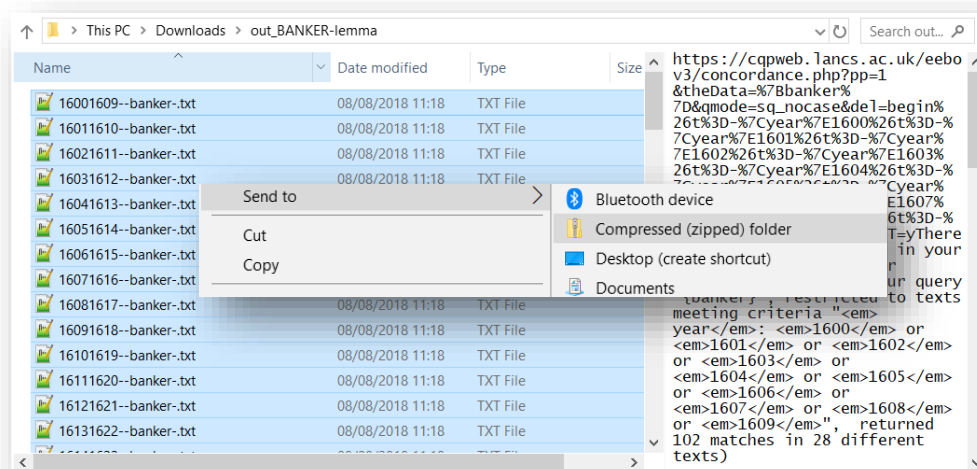
1 https://cqpweb.lancs.ac.uk/eebov3/concordance.php?pp=1&theData=%7Bbanker%7D&qmode=sq\_nocase&del=begin%26t%3D-%7Cyear%7E1600%26t%3D-%7Cyear%7E1601%26t%3D-%7Cyear%7E1602%26t%3D-%7Cyear%7E1603%26t%3D-%7Cyear%7E1604%26t%3D-%7Cyear%7E1605%26t%3D-%7Cyear%7E1606%26t%3D-%7Cyear%7E1607%26t%3D-%7Cyear%7E1608%26t%3D-%7Cyear%7E1609&del=end&T=yThere are 368
different words in your collocation database for "[hw="banker"%c}"].
(Your query "{banker}", restricted to texts meeting criteria
...
4 No. w total no. in whole corpus Expected collocate
frequency Observed collocate frequency In no. of texts Mutual
information value
5
6 1 usurers 3968 0.003 8 6 11.2
7 2 money 188212 0.16 10 5 5.968
8 3 hands 327530 0.278 7 3 4.655
9 4 called 768363 0.652 6 3 3.202
10 5 these 1985969 1.685 13 6 2.948
11 6 ( 3147389 2.67 7 5 1.39
12 7 at 3182175 2.7 7 4 1.374
13 8 who 2315377 1.964 5 5 1.348
14 9 then 2700348 2.291 5 5 1.126
15 10 are 5017099 4.257 9 5 1.08

```

List of collocates (col. 2) and the information about their frequency (col. 5) and association measure value (col 7)

Line 1: header

- 2) Select all text files with collocations (Ctrl+A) and zip them. Make sure that the files are zipped from within the folder.



- 3) Go to the UFA tool, locate the zip file and fill in the rest of the form. The default settings help process standard CQPweb collocation output.

Bootstrapping test  Neighbour clusters  Peaks and troughs  UFA  Videos

1. Indicate historical period. For help click [here](#).  
From:  To:  Window:  info about the time period

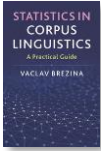
2. Upload a zip file with collocation files.  
 out\_banker.zip zip file location

3. Provide info about data.  
Row no. of header row:   
Regex for identifying collocates:  info about how the data is structured  
Column delimiter:   
Column no. of column with collocates:   
Column no. of column with collocation (node + collocate) frequencies:   
Column no. of column with relevant association measure:

4. Define a collocate.  
Statistic cut-off value:  statistical parameters for classification of collocates  
Latency threshold (LT):  sampling points  
Consistency threshold (CT):  %  sampling points

5. Decide if you want to run the analysis with frequency cut-off point.  
 Yes, absolute cut-off  Yes, relative cut-off  No  
relative cut-off point applied when different amounts of data are available for different periods

6. Provide additional info.  
Frequency cut-off value:   
Regex for identifying node frequency in header (relative cut-off):  additional info if cut-off point is chosen



Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.

- 4) Click 'Analyze'.
- 5) The output

summary and full collocation dataset download.

SUMMARY: The usage fluctuation analysis (UFA) identified 34 unique collocates across 91 periods. [Full table with collocates](#) is also available.

CONSISTENT:

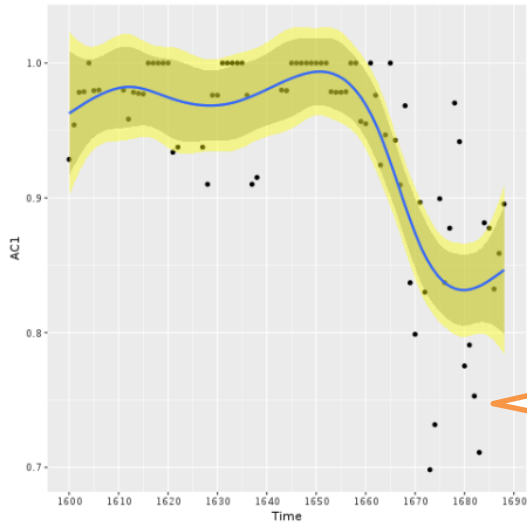
INITIATING: case (1671-),

TERMINATING:


TRANSIENT: bags (1675-1677, 3.3%), banker (1665-1674, 11%), banqui (1675-1677, 2.1%), broke (1682-1683, 2.2%), brokers (1673-1684, 12.1%), called (1675-1677, 2.1%), creditors (1672-1674, 3.3%), do (1630-1638, 2.2%), exchange (1613-1622, 11%), exchangers (1629-1638, 11%), feats (1614-1622, 9.9%), goldsmiths (1676-1690, 16.5%), hands (1600-1690, 30.8%), has (1688-1689, 2.2%), interest (1687-1687, 1.1%), keep (1629-1638, 11%), lombard-street (1682-1683, 2.2%), london (1681-1690, 11%), merchants (1629-1690, 44%), money (1600-1690, 58.2%), mr. (1674-1683, 11%), order (1668-1672, 5.5%), paid (1668-1684, 12.1%), paris (1684-1689, 6.6%), private (1688-1690, 3.3%), scriveners (1661-1670, 11%), these (1601-1601, 1.1%), took (1687-1689, 3.3%), trade (1671-1690, 8.8%), turner (1682-1683, 2.2%), usurers (1600-1670, 15.4%),

classification of collocates

UFA GRAPH:



Graph: points of divergence in collocation use

 R code that performs the analysis can be viewed and copied when going with the mouse pointer to [R code](#)