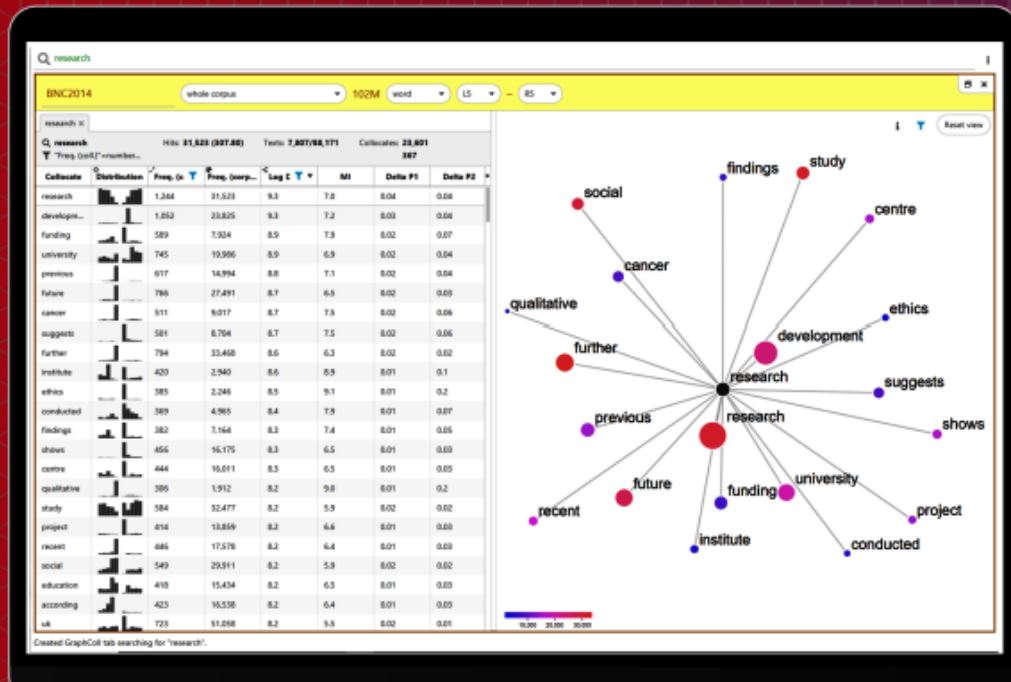# #LancsBox
Innovation in Corpus Linguistics

#LancsBox X is a powerful tool for the analysis of language: millions and billions of words.

Download it for free.

Brezina, V., Platt, W. (2023). #LancsBox X 3.0.0 [software package], lancsbox.lancaster.ac.uk

Words which you need in professional contexts

Test your vocabulary knowledge. Check a box next to a word if you know the word. If you don't know the word, leave the box blank. Then circle the words that will be especially useful in your professional context.

☐ back	☐ lens	☐ stuff
☐ moderate	☐ neen	☐ head
☐ dependent	☐ footballer	☐ desponation
☐ at firse	☐ creature	☐ memory
☐ stone	☐ to register	☐ to wrose
☐ efficiency	☐ to sit	☐ housing
☐ measurement	☐ classical	

Words checked __/20
Words circled __/20

Now, using the BNC2014 v. 2 find the words you didn't know/check. Do they exist and if so, what is their meaning?

Task 2

Words across genres/registers

In this task, you will look at the distribution of words across different registers and genres. This can be done both in #LancsBox on your computer and also in LancsLex, which is an interactive vocabulary website for general use developed at Lancaster University.

Search for the words you have circled in Task 1. In which genre/register do they occur most frequently?
In which genre/register do they occur with the smallest relative frequency?

Word	Genre/register with highest relative frequency	Genre/register with lowest relative frequency

#LancsLex: https://lancslex.lancs.ac.uk

2

Find out more about English vocabulary across genres/registers + improve your lexical knowledge with interactive exercises
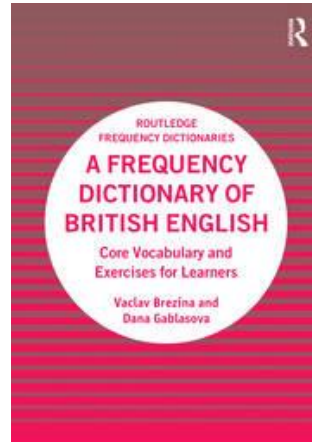

Coming out in November 2023

Task 3 — Language and marketing

In this task, you will be using the GraphColl tool in #LancsBox X to explore associations in British English that are connected with the following words: 'classic', 'comfort', 'cushion' and 'soft'.

Tip: Go to the GraphColl tool, select the BNC 2014 v. 2. and search for the words one by one.

#LancsBox X 3.0.0

🔍 soft

BNC2014 2.0 CLAWS7   whole corpus
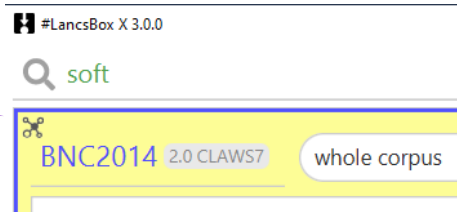
Explore the associations that are unique to these words as well as associations that are shared among/between them. Which of these words would you choose for a marketing campaign for toilet tissues? Why?

Task 4 — Selling your ideas

Select a set of three terms, which are relevant in your professional context:

My context is: _____

My three tems are: _____

Based on my intuition, the most POSITIVE/ ENGAGING of these terms is: _____

Now check the associations of these terms in GraphColl. You can use the whole BNC2014 v. 2 corpus or any relevant subcorpus. Was your intuition correct? Is there anything new you have learned about the use of these terms?

_____

_____

_____

_____

Look at the collocation graphs below. Each graph represents collocates around a node, which has been hidden. Choose an appropriate node from the box below.
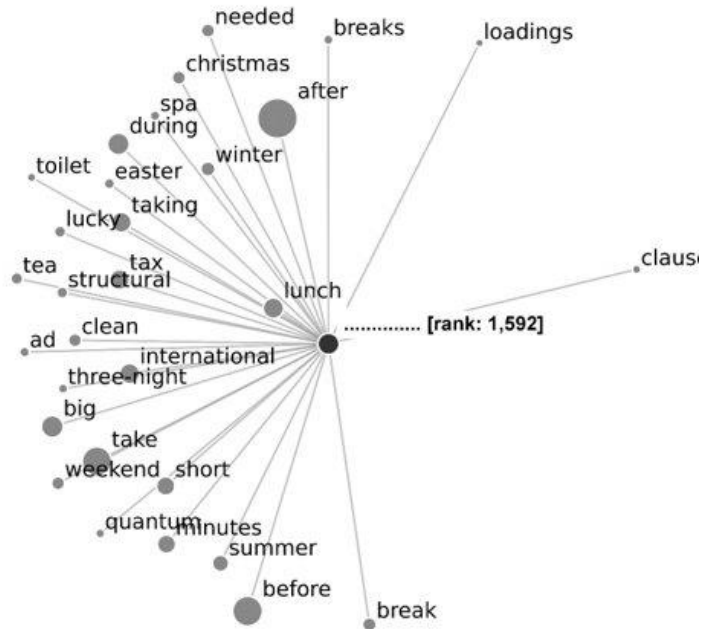
break (n), eat (v), good (adj), play (v), rude (adj), social (adj)

Check your answers in the BNC2014 v.2 using GraphColl and note down examples of use (concordance lines) for the strongest collocations.

1. *Media* as the strongest collocate
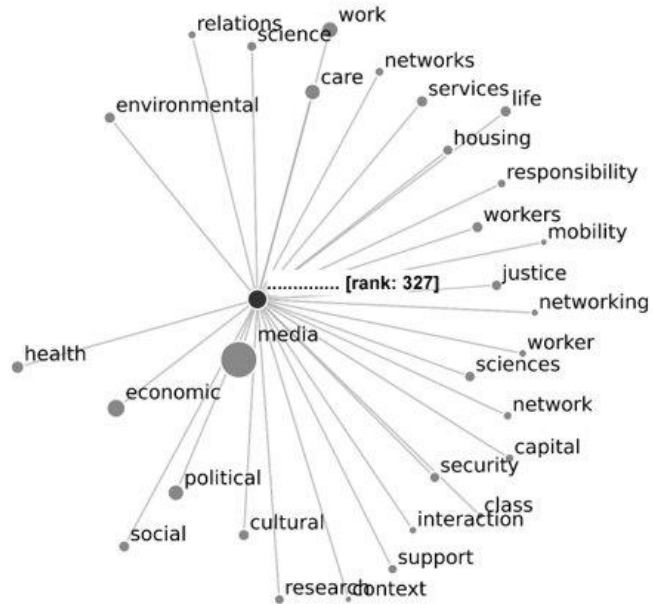
2. *Lunch* as the strongest collocate

Creating your own corpus from the web

Creating your own corpus automatically in #LancsBox X

#LancsBox X offers a feature, which allows automatic download of texts from the web. In this task, you will explore this feature. Imagine that you want to download a corpus made of Wikipedia articles related to your subject area.

a) Locate your initial (seed) Wikipedia website: _____

b) From the corpus drop down menu choose 'add corpora' and select 'Web'.

Add corpora

| Corpus hub | My data | Web |

Corpus full name*   Wiki_linguistics   **1. Name your corpus**

Short display name

Language   English

Initial URLs* ⓘ   https://en.wikipedia.org/wiki/Linguistics

3. Click on "Create corpus"

2. Paste the Wikipedia URL

corpus   Close

c) [OPTIONAL] Choose a different corpus name e.g. 'Wiki_linguistics_tagged'. Repeat the process with Grammatical and Semantic tagging on.

Tagging   ☑ Grammatical ⓘ   ☑ Semantic ⓘ

Select the appropriate tagging. If you are using this feature for the first time, the appropriate taggers and language models will need to be downloaded; this is done automatically.

Understanding your data

Exploring your Wikipedia corpus in the Text tool in #LancsBox X

a) Explore your corpus. Fill in the table below with the descriptive statistics of your Wikipedia corpus:

Corpus size – space tokens (strings separated by space)	
Corpus size – grammar tokens (identified by the tagger)	
Texts	

b) How many times do different words related to your topic occur in the corpus? Search for these in the Text tool.

Create a corpus of job adverts from the Guardian jobs website. Can you see a pattern showing gender preferences in how the adverts are framed?

a) Identify jobs that have been typically male dominated, e.g. engineering and jobs that have been typically female dominated e.g. nursing.

   Male dominated jobs:_____

   Female dominated jobs:_____

b) Go to https://jobs.theguardian.com and search for the key jobs you have identified in a) and copy the urls in turn into #LancsBox X.

   e.g.   https://jobs.theguardian.com/searchjobs/?Keywords=engineer&Page=1
           https://jobs.theguardian.com/searchjobs/?Keywords=engineer&Page=2

c) Create two corpora for these two groups of jobs (see Task 6 for instructions).

d) Compare and contrast these corpora using corpus linguistic tools of your choice e.g. Text, Words (keywords) and GraphColl. What patterns have you noticed? What recommendations can you make to HR departments based on your analysis?

_____

_____

_____

_____

Developed at
Lancaster
University