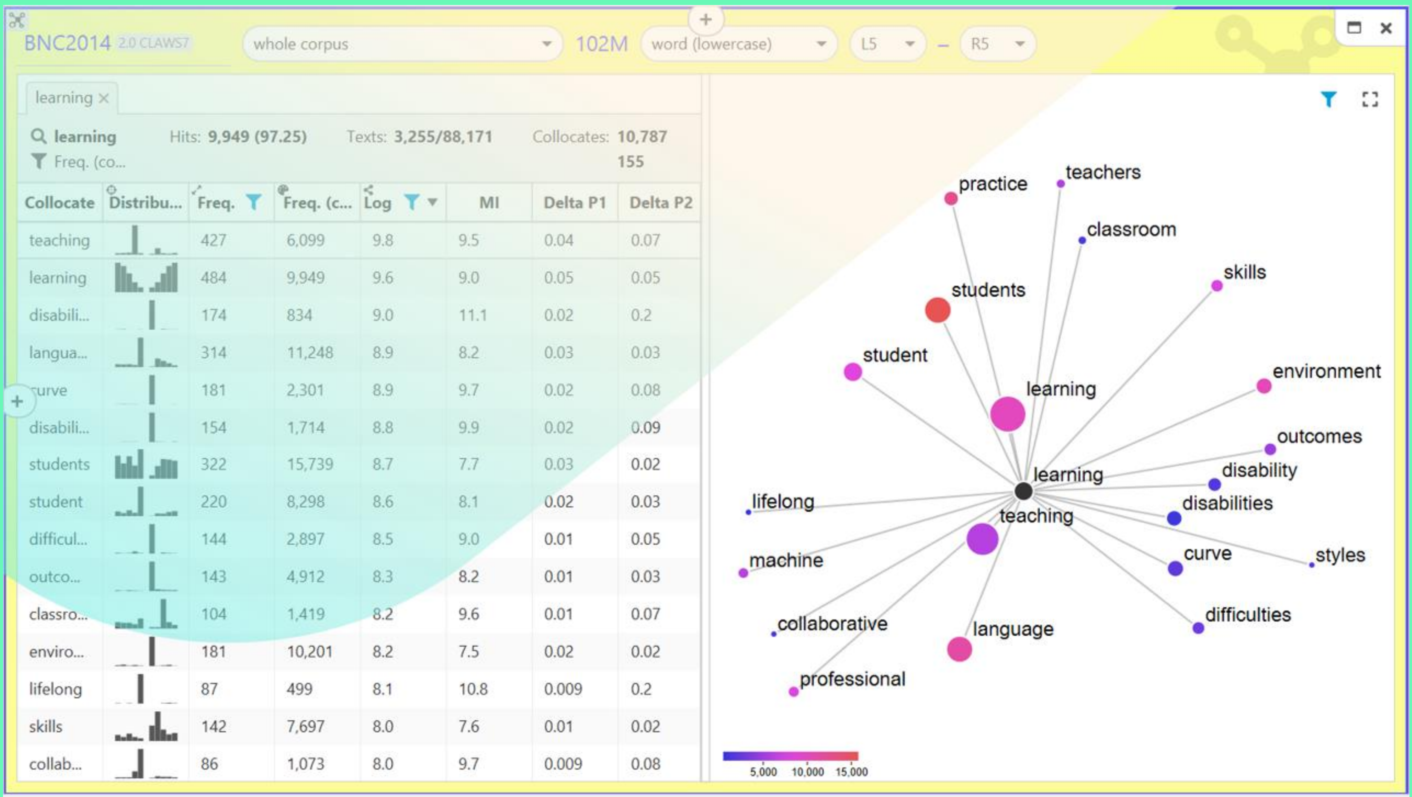
#LancsBox X



BNC2014 2.0 CLAWS7 | whole corpus | 102M | word (lowercase) | L5 – R5

learning ×

learning | Hits: **9,949 (97.25)** | Texts: **3,255/88,171** | Collocates: **10,787**
Freq. (co... **155**

Collocate	Distribu...	Freq.	Freq. (c...	Log	MI	Delta P1	Delta P2
teaching		427	6,099	9.8	9.5	0.04	0.07
learning		484	9,949	9.6	9.0	0.05	0.05
disabili...		174	834	9.0	11.1	0.02	0.2
langua...		314	11,248	8.9	8.2	0.03	0.03
curve		181	2,301	8.9	9.7	0.02	0.08
disabili...		154	1,714	8.8	9.9	0.02	0.09
students		322	15,739	8.7	7.7	0.03	0.02
student		220	8,298	8.6	8.1	0.02	0.03
difficul...		144	2,897	8.5	9.0	0.01	0.05
outco...		143	4,912	8.3	8.2	0.01	0.03
classro...		104	1,419	8.2	9.6	0.01	0.07
enviro...		181	10,201	8.2	7.5	0.02	0.02
lifelong		87	499	8.1	10.8	0.009	0.2
skills		142	7,697	8.0	7.6	0.01	0.02
collab...		86	1,073	8.0	9.7	0.009	0.08

Data Driven Learning

Professor Vaclav Brezina
@ Lancaster University

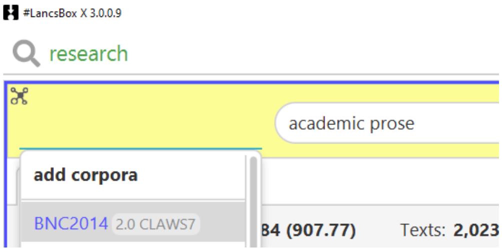
LET'S START

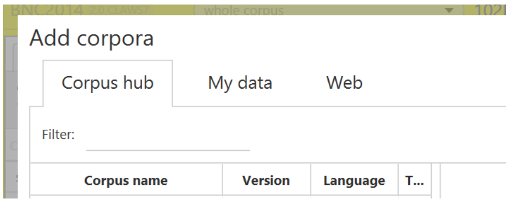
Starting with #LancsBox X

#LancsBox X is a powerful software tool for the analysis of large amounts of language. It can be used with your own data or the data provided.

The tool is very easy to use with an intuitive and flexible UI.

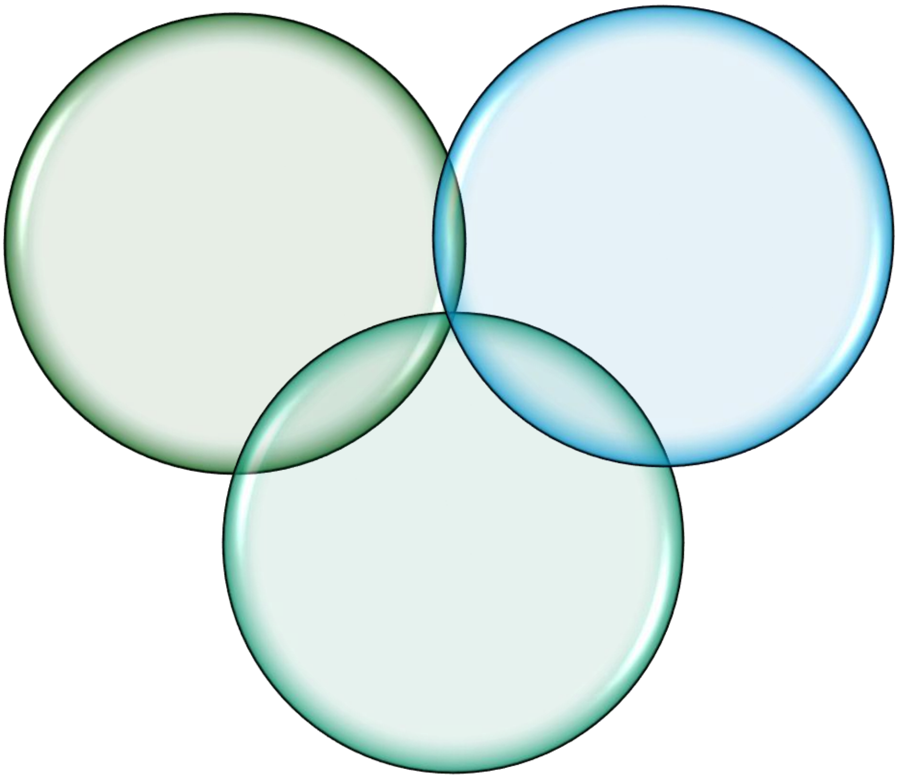
1. Download the most recent version of #LancsBox X from https://lancsbox.lancs.ac.uk
2. Go to 'add corpora'> 'Corpus Hub' and select and download the British National Corpus 2014 version 2.





THE STORY

Why is DDL important?



Lexical knowledge

Check your vocabulary knowledge

In this section, we will focus on the exploration of concordance lines to find meaning of words (their uses in context). We will also discuss frequency information. You will learn

- to use the KWIC tool
- to sort and filter concordance lines
- to use the summary table

KWIC

frequency

context

meaning

Task 1

Vocabulary test

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.

□ back
□ moderate
□ dependent
□ at firse
□ stone
□ efficiency
□ measurement

□ lens
□ neen
□ footballer
□ creature
□ to register
□ to sit
□ classical

□ stuff
□ head
□ desponation
□ memory
□ to wrose
□ housing

Words checked: __________

Task 2

Discovering the meaning of words

Find the meaning of the words from Task 1. Using the KWIC tool, search for the words you were not sure about in the British National Corpus 2014. Note down their frequencies, genres where they typically occur and their meaning.

Word	Frequency	Dominant genre/register?
moderate		

Developed at
Lancaster University

Collocations in context with GraphColl

In this section, you will explore collocation graphs and networks using the GraphColl tool in #LancsBox X. The GraphColl tool identifies collocations and displays them in a table and as a collocation graph or network. It can be used to:

- Find the collocates of a word or phrase.
- Find colligations (co-occurrence of grammatical categories).
- Visualise collocations and colligations.
- Identify shared collocates of words or phrases.

GraphColl collocation graphs collocation networks

Task 3 Finding collocates

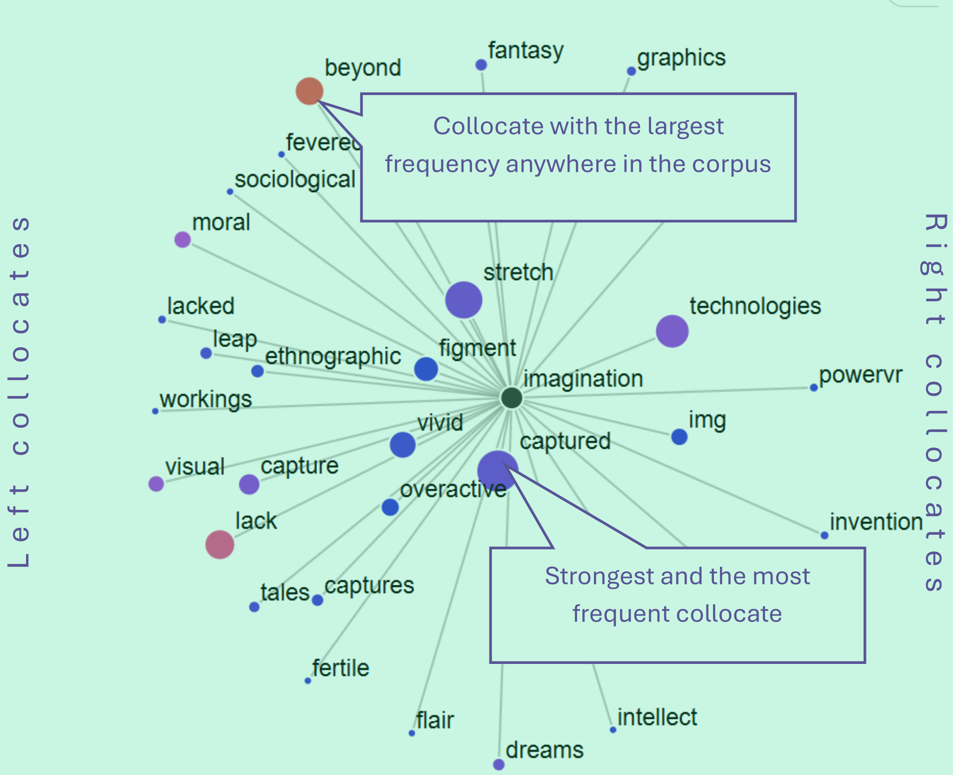
In this task, you will practice finding collocates and interpreting collocation statistics. Go to the GraphColl tool in #LancsBox X, select the BNC2014 and search for the expressions in the table below.

Note down top collocates according to log Dice and the collocation frequency.

Search term	Top 3 log Dice collocates	Most frequent collocate
`vision`		
`avid`		
`unclear`		
`[hw="disagree" pos="V.*"]`		
`[hw="ring" sem="B.*"]`		
`[hw="ring" sem="Q.*"]`		

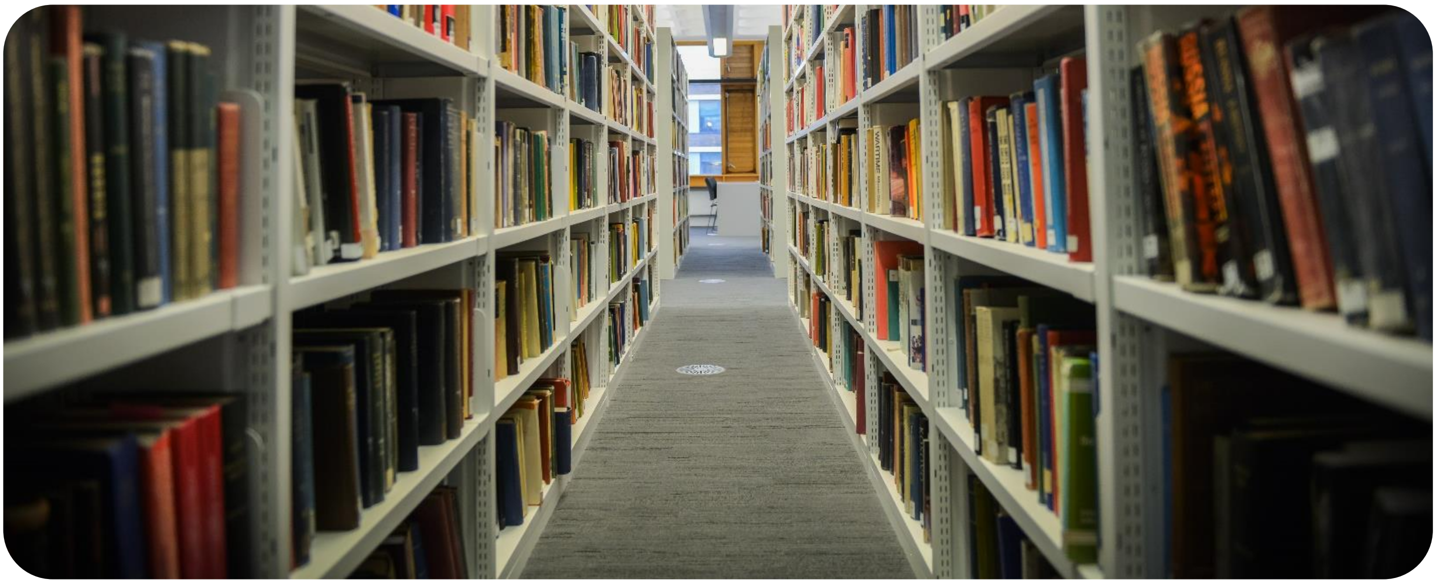
Collocation graph

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).



Collocate with the largest frequency anywhere in the corpus

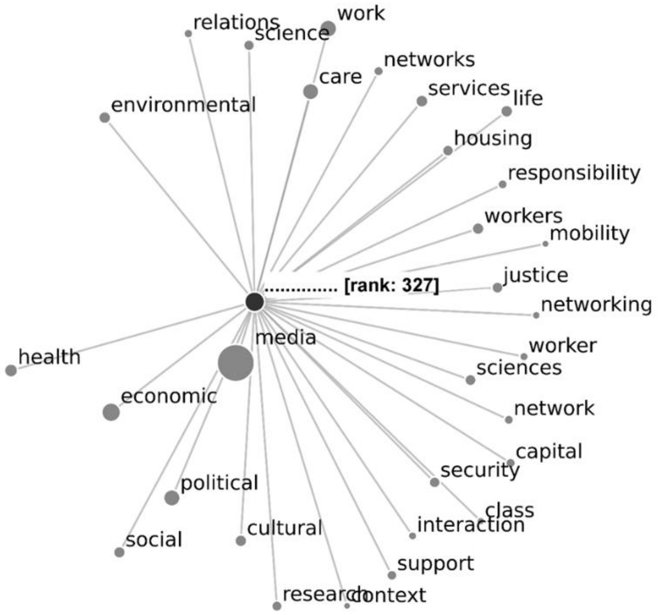
Strongest and the most frequent collocate

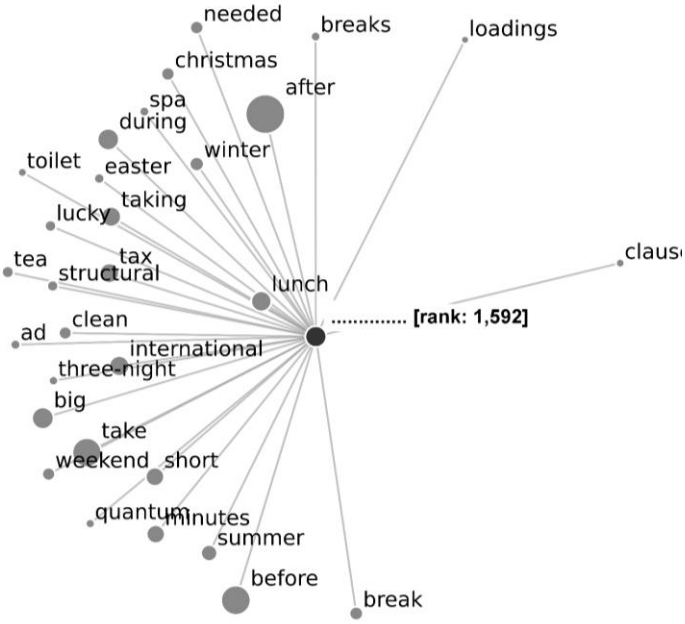


Task 4 Understanding collocation graphs

Look at the collocation graphs below. Each graph represents collocates around a key word, which has been hidden. Choose an appropriate key word from the box below.

break (n), eat (v), good (adj), play (v), rude (adj), social (adj)





Keywords and Key phrases

In this section, you will explore words and phrases important in a particular corpus when compared to a reference corpus. These are called *keywords* and *key phrases*. You will learn how to:

- Create and visualize a frequency list.
- Create a keyword list.
- Understand keyword statistics.

Wordlists keywords key phrases

Task 5 Understanding wordlists and n-gram lists

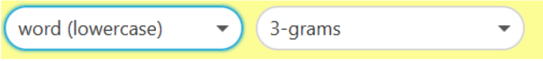
Select the Academic prose subcorpus of the BNC2014. Create and visualize words and n-grams.

1. First create a wordlist based on lemmas. How many lemmas does the list contain? ______________

2. Click on the filter icon and select nouns (type in _N). How many nouns are there? ____________

3. What are the top 5 academic nouns? ____

   ____

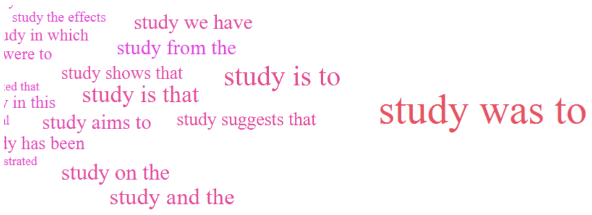
4. Change the settings as follows  word (lowercase) 3-grams and identify all trigrams.

What are the top 3 academic trigrams? ____

5. Click on the filter icon and select trigrams starting with ‘study’, the most frequent academic noun.

   How many trigrams have you identified? ____________

6. Visualize and interpret these results.



Task 6

Understanding keywords and key n-grams

Keywords and key phrases are words that occur with a considerably higher frequency in a given (sub)corpus compared to a reference (sub)corpus. In this task, you will explore key 2-grams in the Academic prose subcorpus compared to the whole BNC2014.

1. First create a 2-gram list based on the Academic prose subcorpus and note down the first three items:

   ____

2. Click on the keyword icon 🔑 and select the whole BNC2014 as a reference corpus.

3. Note down the top 10 key bigrams, i.e. combinations typical of the academic prose subcorpus.

   ____

   ____

4. Write a short description of key features of academic writing based on the bigrams above.

   ____

   ____

   ____

   ____

   ____

   ____

Developed at
Lancaster
University

#LancsBox X in your research

In this section, you will design a small study in the area of your interest. The focus is on:

- Conceptual grounding.
- Research question(s).
- Operationalization and study design.
- Data collection.
- Data analysis.

Research design methodology

Task 9 Designing a corpus study

In this task, you will design a mini-study. First, think about a topic in the area of your interest.

Topic: ____

Then, think of a specific question you would like to ask:

Specific question: ____

Is it a yes/know question? If no, think of an aspect of your question that can be formulated as a yes/no question.

Yes/No RQ: ____

Data: Is there an available corpus that can be used to answer the RQ? Yes/No ____________________

If no, can the data be obtained online? URL____

Operationalization:

#LancsBox X tools: ____

Search terms: ____

Comparisons: ____

Possible challenges: ____