#LancsBox X

KWIC

Exercises: Using KWIC

In this worksheet, you will explore concordances and frequency information using the KWIC tool (key word in context) in #LancsBox X. The KWIC tool generates a list of all instances of a search term in a corpus in the form of a concordance. It can be used, for example, to:

■ Find the frequency of a word or phrase in a corpus.
■ Find frequencies of different word classes such as nouns, verbs, adjectives.
■ Find complex linguistic structures such as the passives, split infinitives etc. using 'smart searches'.
■ Define complex searches at different levels of annotation using CQL (Corpus Query Language).
■ Filter concordance lines.

> KWIC — frequency — concordances

Task 1 — Searching corpora

In this task, you will practice different search options in the KWIC tool. Go to the KWIC tool in #LancsBox X, select the BNC2014 corpus, and search for the expressions in the table below. Note down their frequencies and distributions in texts.

Tip: Hover your mouse over the results of a search (e.g. number of occurrences) and click on the "Copy to Clipboard" symbol: 🗐. Then, paste the data in the table below using the keyboard shortcut CTRL + V on Windows, Command ⌘ + V on macOS.

Hits: 571 (0.29) 🗐 Texts: 267/50,210
Copy

Search term	Occurrences (per 1M)	Texts
research		
research*		
research and *		
*ment		
ADJECTIVE research		
FOOD		
SWEARWORDS		
[hw="walk"]		
[hw="walk" pos="V.*"]		

Developed at
Lancaster
University
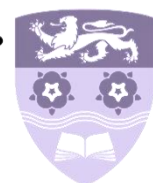
Searching for patterns of language use

This task demonstrates how corpus data can be used to search for patterns of language use and test our intuitions about language. First, answer three questions drawing on your experience with the English language. Then, use corpus evidence by carrying out searches in the BNC2014 to check whether your intuitions were correct.

2a
What particles (e.g. *up, out*) are typically used with the verb *carry*? Make a list:

_____

_____

2b
Which particle is most frequently used with the verb *carry*?

_____

2c
What do we typically *carry out*? Make a list of the nouns that you think are frequently used with the verb *carry out*.

_____

_____

Now let's check your answers using corpus data:

2d
First, search for [hw="carry" pos="V.*"] PARTICLE and sort the concordance lines according to the node. What particles are used with the verb *carry*? List them below and see if they match your list in Task 2a:

_____

_____

2e
Search for the different combinations of the verb *carry* followed by a particle (e.g. *carry on, carry out*) and note down their absolute (and relative) frequency in the table below. Which is the most frequent one? Was your guess in Task 2b correct?

_____

_____

[hw="carry" pos="V.*"] + on, out…	Absolute frequency (texts)

2f
Search for [hw="carry" pos="V.*"]out NOUN and sort the concordance lines according to the node. List the nouns used with the verb *carry out* below and see if they match your list in Task 2c:

_____

_____

Task 3   Filtering concordances

In this task, you will practice the use of filters in the KWIC tool. Go to the KWIC tool in #LancsBox X, search for the following expressions in the BNC2014 and apply filters. Note down their frequencies and distributions in texts.

Search term	Filter	Occurrences (per 1M)	Texts
new	york [anywhere RIGHT]		
old	f* [in R1 position]		
time	this [in L1 position] was [in R2 position]		

Task 4   Advanced CQL searches

In this task, you will practice advanced searches using corpus query language (CQL) in the BNC2014. Answer questions 1- 5 below: *Please note: CQL conventions may differ slightly between corpus tools, including different version of LancsBox. Please refer to the manual of each tool for details.*

Tip: CQL queries use square brackets [ ] to specify individual tokens (word) and their attributes such as the pos-tag (pos), headword (hw), semantic tag (sem) etc. The expression inside quotes " " is interpreted as a regular expression. For instance, N.* stands for all strings starting with N followed by any mumber of characters including no characters at all.

[word="walked" word="walk" pos="N.*" sem="M1"]

1. Write a query to find all instances of "English" as a noun (nouns have pos beginning with N).

2. Write a query to find all instances of "English" as an adjective (adjectives have pos beginning with J).

3. Write a query to find all nouns starting with the letter z.

4. Find all instances of "English" as an adjective followed by any common noun in plural (NN2 in CLAWS7 tagset).

5. Find instances of "English" as a noun that occur immediately after a full stop.

CQL Query Help

[hw="new.*"]	Headword starting with **new**
[word="walk" pos="N.*"]	**walk** as a noun
[word=="Cat"]	**Cat** case-sensitive search
[pos="VB.*"] [pos="R.*"] {0,3} [pos="V.N"]	**to be** followed by a past participle with 0-3 adverbs in between
/punc="\?|\.|,|;"/	To search for **punctuation** use forward slashes and the attribute 'punc'. Special characters such as the question mark or the full stop need to be escaped by using the backlash symbol \