

Using and improving #LancsBox software  
on the *Corpus Linguistics: Method,  
Analysis and Interpretation* course

Prof Vaclav Brezina

Lancaster University



Online Courses / Language



## Corpus Linguistics: Method, Analysis, Interpretation

★★★★★ 4.9 (95 reviews)

Get a practical introduction to the methodology of corpus linguistics for researchers in the social sciences and humanities.

Join course

70,092 enrolled on this course



- One of the oldest on the FL platform
- 8 weeks of instructions
- Methodologically oriented: transferable skills
- Consistent interest 2014-2021: 9 runs
- Trained: 70k participants

.innovation in corpus linguistics

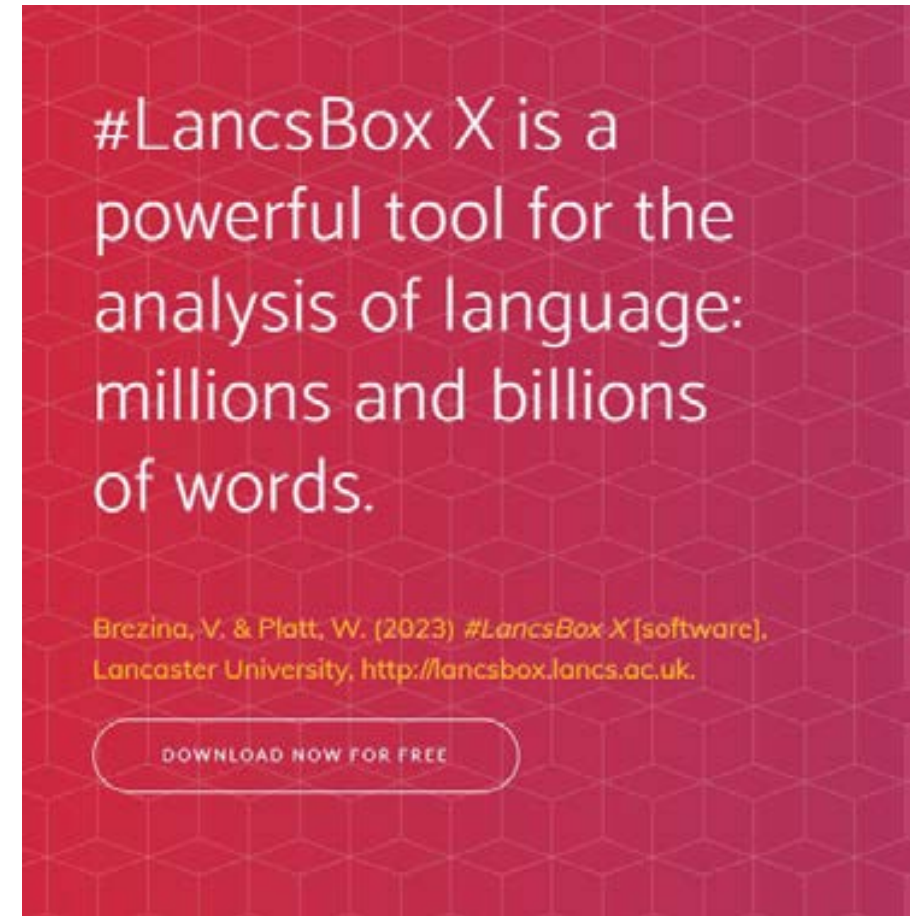
**#LancsBox**

@Lancaster University

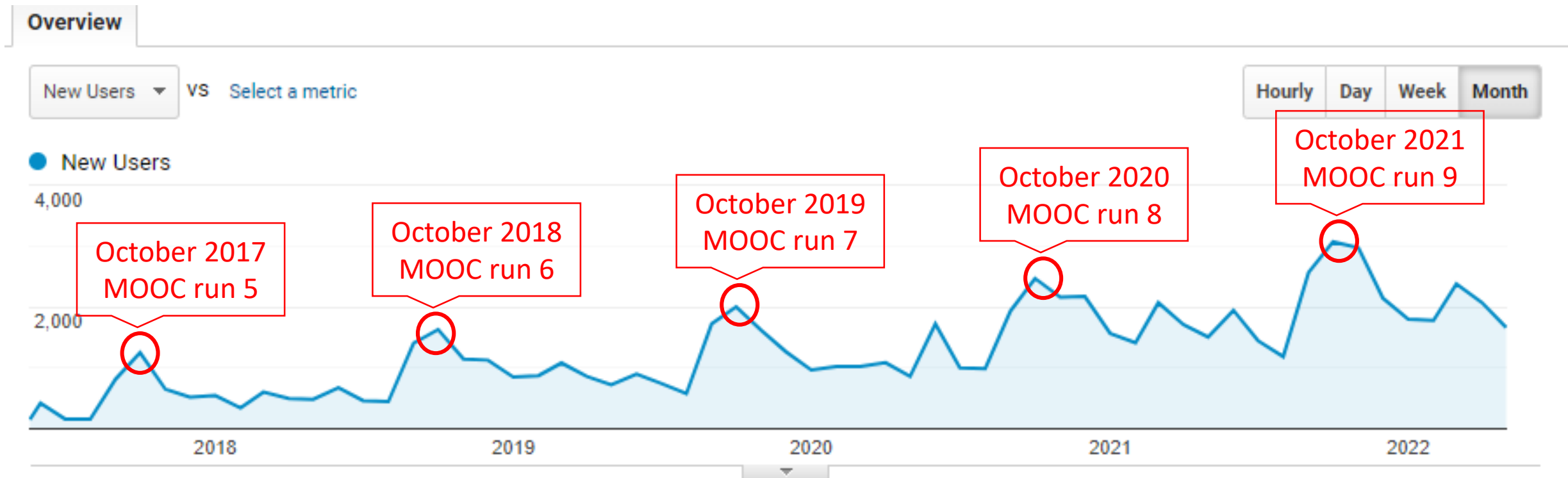




Download #LancsBox v. 6 and X:  
<https://lancsbox.lancs.ac.uk>  
<http://corpora.lancs.ac.uk/lancsbox>



# #LancsBox new users























# Engagement

**Users: 77,901**

Citations: over 500

Training events: over 70

	Country	Users	% Users
1.	 China	16,723	 21.37%
2.	 United Kingdom	9,980	 12.76%
3.	 United States	6,551	 8.37%
4.	 Italy	3,806	 4.86%
5.	 Germany	2,852	 3.65%
6.	 Spain	2,006	 2.56%
7.	 Brazil	1,848	 2.36%
8.	 Indonesia	1,707	 2.18%
9.	 Poland	1,520	 1.94%
10.	 Russia	1,519	 1.94%



# Engagement

Users: 77,901

**Citations: over 500**

Training events: over 70

## Collocations in context

A new perspective on collocation networks\*

Vaclav Brezina, Tony McEnery and Stephen Wattam  
Lancaster University

The idea that text in a particular field of discourse is organized into lexical patterns, which can be visualized as networks of words that collocate with each other, was originally proposed by Phillips (1983). This idea has important theoretical implications for our understanding of the relationship between the lexis and the text and (ultimately) between the text and the discourse community/ the mind of the speaker. Although the approaches to date have offered different possibilities for constructing collocation networks, we argue that they have not yet successfully operationalized some of the desired features of such networks. In this study, we revisit the concept of collocation networks and introduce *GraphColl*, a new tool developed by the authors that builds collocation networks from user-defined corpora. In a case study using data from McEnery's (2006a) study of the Society for the Reformation of Manners Corpus (SRMC), we demonstrate that collocation networks provide important insights into meaning relationships in language.

# Engagement

Users: 77,901

Citations: over 500

**Training events: over 70**

Who will you learn with?



**Tony McEnery**



Has been working for over 20 years to help pioneer new ways to use computers to analyse very large collections of language data.








**Vaclav Brezina**

Corpus linguist, lead developer of #LancsBox.


# Feedback cycle


  **Course team replied** [Follow](#) 20 SEP




Thank you for your wonderful lectures! They are very helpful. I have one question, though. Can we use LancsBox to search concordances from parallel corpora? 1 ST, and 5 TTs, for example.

 Pin  Like 1  Reply  Bookmark  Report

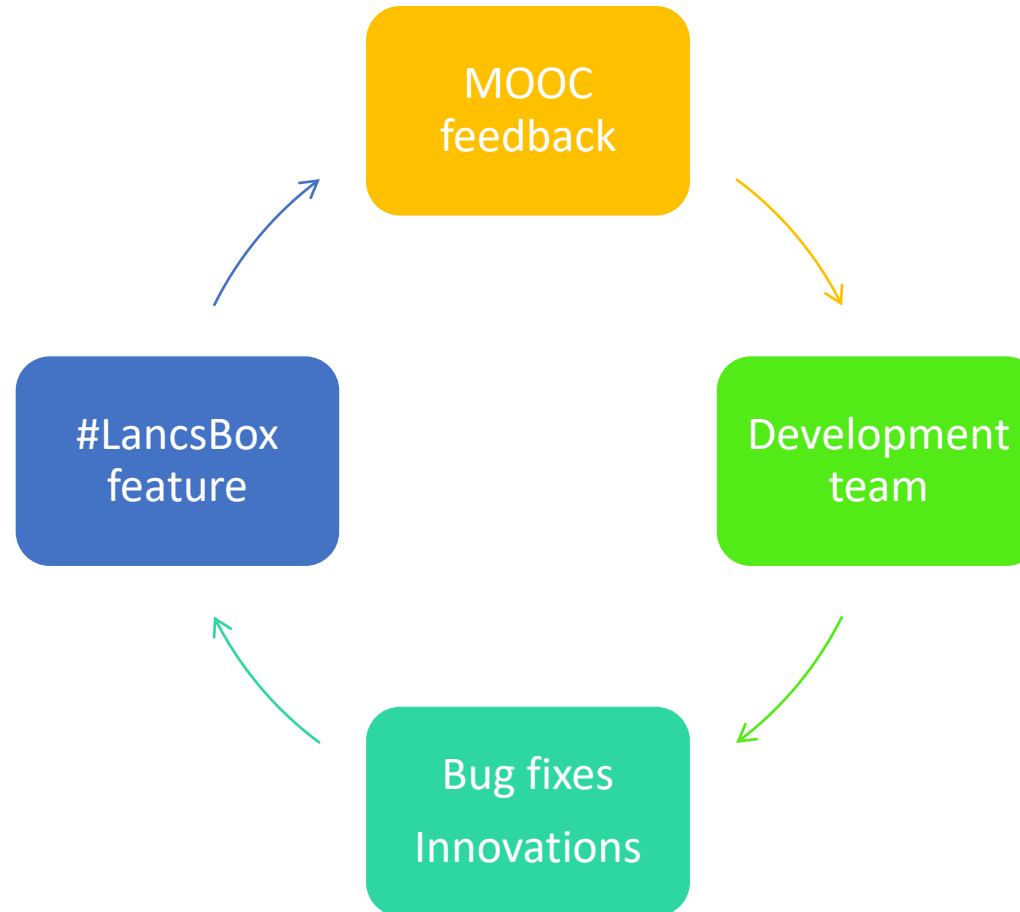
[Mute this conversation](#)

 **Vaclav Brezina** **Educator** 20 SEP

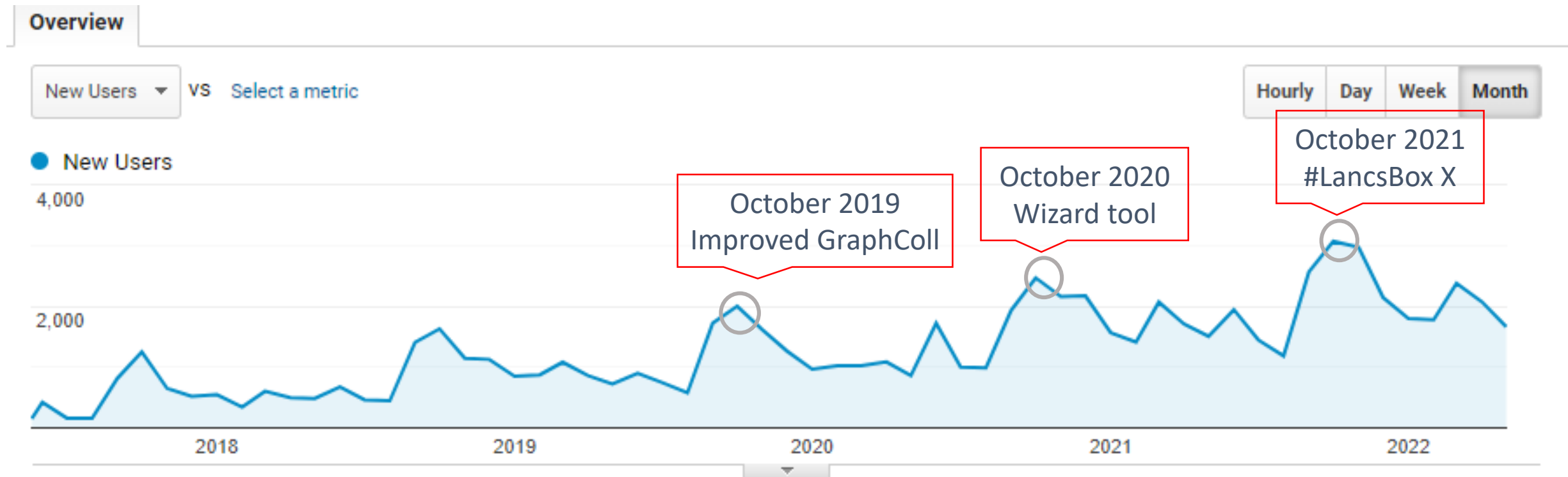
Hi  #LancsBox offers a split-screen view that allows you to compare and contrast concordances for parallel corpora. However, in the current version #LancsBox does not include an alignment feature that would allow extracting translations automatically. I hope this helps.

 Like 1  Reply  Bookmark [Edit](#)

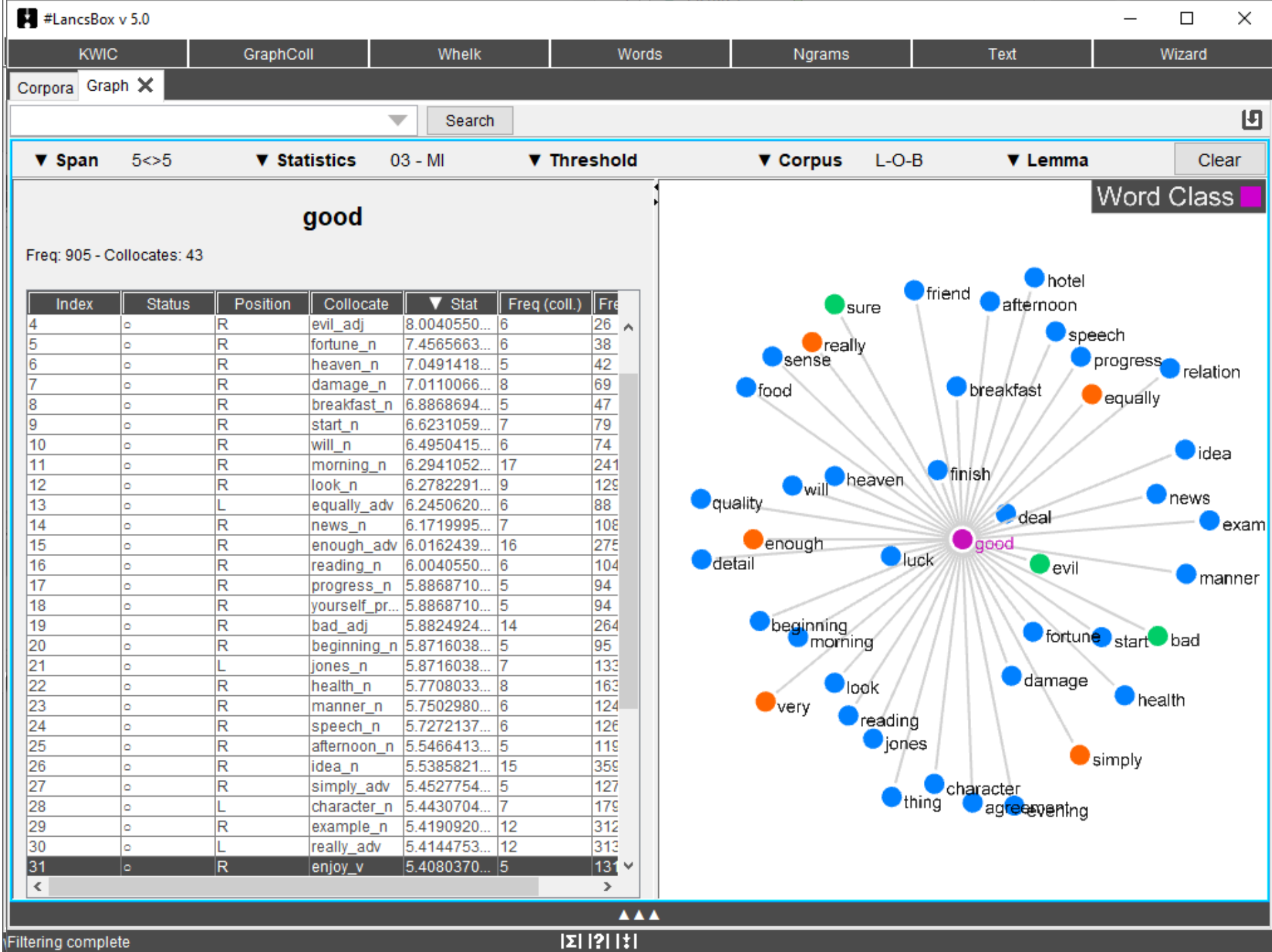
# Feedback cycle



# #LancsBox innovations



# Visualizing MOOC 19





# Visualizing MOOC 19

#LancsBox v 5.0

KWIC GraphCol

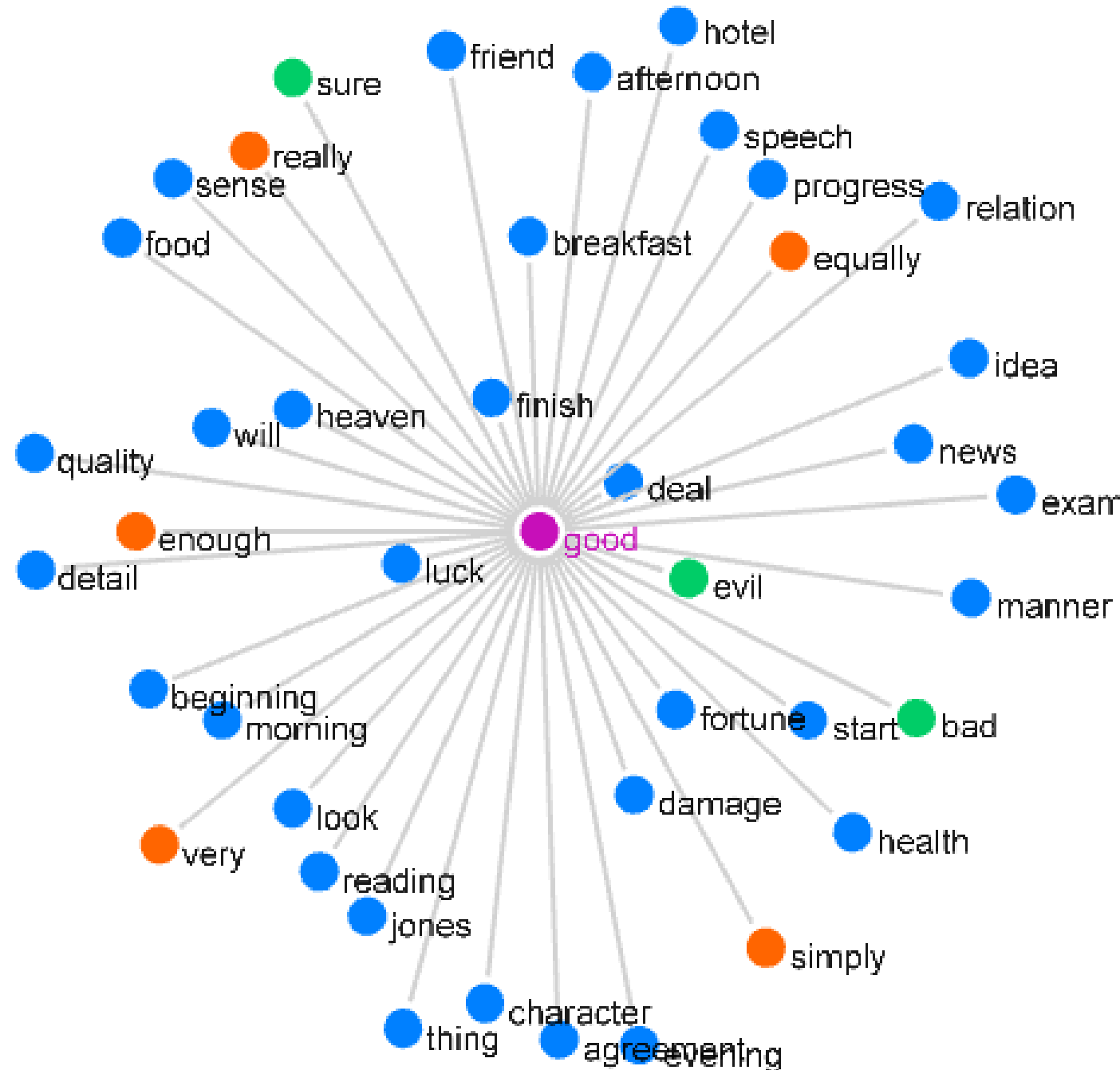
Corpora Graph X

▼ Span 5<>5 ▼ Stat

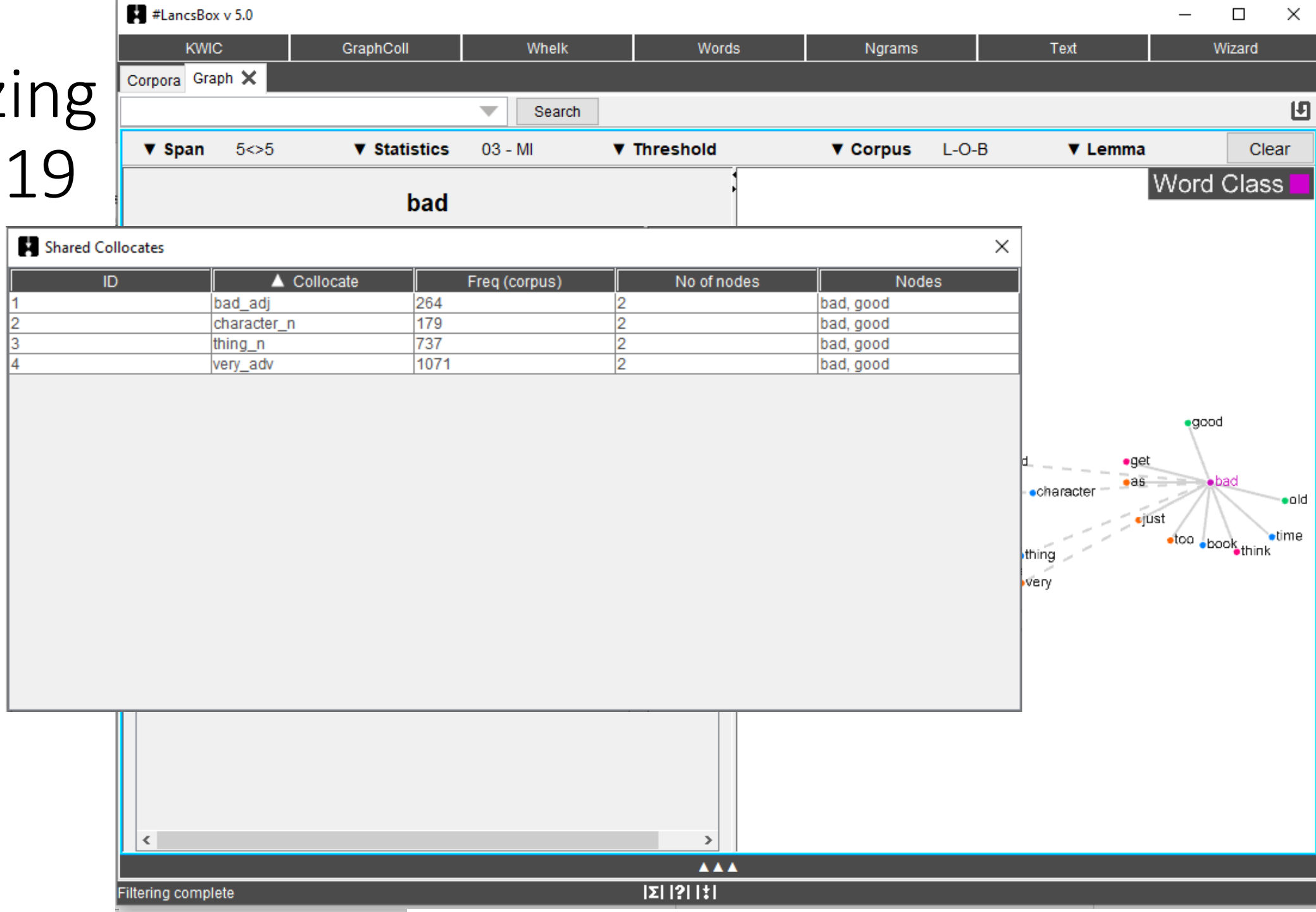
Freq: 905 - Collocates: 43

Index	Status	Position
4	o	R
5	o	R
6	o	R
7	o	R
8	o	R
9	o	R
10	o	R
11	o	R
12	o	R
13	o	L
14	o	R
15	o	R
16	o	R
17	o	R
18	o	R
19	o	R
20	o	R
21	o	L
22	o	R
23	o	R
24	o	R
25	o	R
26	o	R
27	o	R
28	o	L
29	o	R
30	o	L
31	o	R

Filtering complete



# Visualizing MOOC 19



Wizard tool  
MOOC 20

KWIC

GraphColl

Whelk

Words

Ngrams

Text

Wizard

Corpora Wizard X

## Corpora to use

BE06  
Brown  
Climate  
Corpus 1  
Corpus 2

Choose a corpus  
or multiple  
corpora.

All

None

## Tools to use

☒ KWIC  
☒ GraphColl  
☒ Whelk  
☒ Words  
☐ Ngrams  
☒ Text

Settings

Settings

Settings

Settings

Settings

Choose tools and  
modify settings.

All

None

## Search terms

new  
old

Enter search  
term(s), if KWIC,  
GraphColl, Whelk  
or Text are used.

Load search  
terms from a file

File...

## Output

Title report

Folder rs\localadmin\Desktop\report4

Folder...

Modify the name  
of the report and  
its location

Run the tool

Run

# 1 Introduction

## 2 Method

### 2.1 Data

### 2.2 Procedure

## 3 Results

### 3.1 General overview: Frequency

### 3.2 Corpus comparison: Keywor

### 3.3 Specific searches: Concorda

#### 3.3.1 Search term "good" in E

#### 3.3.2 Search term "great" in E

#### 3.3.3 Search term "good" in L

#### 3.3.4 Search term "great" in L

### 3.4 Word associations: Collocati

### 3.5 Statistical analysis

## 4 References

## 5 Appendix

## 3.5 Statistical analysis

So far, descriptive statistical analysis was reported in the sections above. This includes the analysis of frequency and dispersion (sections 3.1 and 3.3), keywords (section 3.2) and collocations (section 3.4). More details about these procedures can be found in Brezina (2018).

The t-test ( $t(28) = -0.82$ ,  $p = 0.421$ ) showed that there is not a statistically significant difference between the corpora with regard to the linguistic variable *good*. This result is visualised below. Cohen's  $d$  ( $-0.3$ , 95% CI  $[-1.05$ ,

### Name



csv



images



tsv



xml

[csv](#).

# Introducing #LancsBox X: MOOC 21/22

#LancsBox X 1.0.0

Q ADJECTIVE [hw="food"]

BNC1994

whole corpus

97M

ADJECTIVE [hw="food"]

Hits: 4,467 (0.46)

Texts: 1,196/7,531

File	Left	Node	Right
Oth...	International plc operates a	fast food	franchise. Specialities refle...
Fict...	Cati's stomach like	undigested food;	Rosalba couldn't possibly
Fict...	...en included beans among	prohibited foods.	Although southern Italians...
Oin...	day. [Scene 4] Food,	good food,	comes to the ward
Fict...	the change from the	hot food	to the offerings of
Oth...	that you live on	dehydrated food,	and since at the
Mag...	not readily come to	hanging food	and birdtables. The Birdfo...
Oth...	are already too many	fast food	operators providing meat i...
Oth...	...ing enormous amounts of	expensive food	and drink if they
Fict...	I would be kept	alive. Food	and drink would be

BNC2014

whole corpus

100M

ADJECTIVE [hw="food"]

Hits: 5,764 (0.58)

Texts: 2,678/88,171

File	Left	Node	Right
Elan...	Having read about the	wonderful food	at Ducksoup on Rosie
Wts...	of students licking the	cheap food	out of their hideous
Wts...	eats a lot of	posh food.	He has no other
New...	peace and dine on	organic food	from the prison gardens.
New...	to afford to buy	good food (	May 5). In my
New...	to spend time preparing	good food.	I raised my children
Fict...	saucers, some tinned and	dried food,	baking trays that have
Mag...	is perfect because the	Whole Foods	business model - organic and
Mag...	...iness model - organic and	expensive food -	was preventing it reaching
Mag...	do so, it launched	Whole Foods	365, effectively offering a

Search completed.



Smarter.

Sleeker.

Faster.

> Smarter.  
Sleeker.  
Faster.

# X IS FOR XML

<u gender="male"/>

<pause/> within <u gender="male"/>

by within PREPOSITIONAL\_PHRASE

<s/> within yes

(<s/> within yes) within <u gender="male"/>

> Smarter.  
Sleeker.  
Faster.

# SUBCORPORA

The screenshot shows a software window titled "BNC2014" with a dropdown menu set to "whole corpus" and a value of "100M". The main dialog box is titled "Define new subcorpus" and has a "Name:" field containing "academic prose: humanities book". A "Clear all" button is in the top right. The dialog is organized into several sections with checkboxes:

- Text** (expanded):
  - mode**:
    - ☐ speech
    - ☐ writing
  - genre**:
    - ☐ academic prose
    - ☐ elanguage
    - ☐ fiction
    - ☐ informal speech
    - ☐ magazines
    - ☐ newspapers
    - ☐ official documents
    - ☐ written-to-be-spoken
  - subgenre**:
    - ☒ academic prose: humanities
    - ☐ academic prose: medicine
    - ☐ academic prose: natural sciences
    - ☐ academic prose: politics/law/education
    - ☐ academic prose: social sciences
    - ☐ academic prose:
  - subsubgenre**:
    - ☐ academic prose: humanities: archaeology
    - ☐ academic prose: humanities:
  - sample**:
    - ☐ beginning
    - ☐ composite
    - ☐ end
  - academic publication**:
    - ☒ book
    - ☐ journal

At the bottom right are "OK" and "Cancel" buttons.

Smarter.  
> Sleeker.  
Faster.

# SIDE-BY-SIDE TOOLS

#LancsBox X 1.0.0

Q [hw="food"]

BNC2014 informal speech 10M

[hw="food"] Hits: 2,863 (2.77) Texts: 705/1,251

File	Left	Node	Right
Sp0m2...	them to pay for my	food	and what else and cos
Sp0m2...	I'm running out of	food	no I mean I have
Sp0m2...	I have run out of	food	I don't have anything
Sp0m2...	you've run out of	food	no that's fine that
Sp0m2...	there'll be like the	food	shop they don't even
Sp2m2...	we still need two single	food	place oh yeah I'm
Sp2m0...	is maybe the lack of	food	and maybe they were just
Sp1m1...	and that there or take	food	with you? yeah no eat
Sp1m1...	there selling all kinds of	food	hot food yeah so you
Sp1m1...	all kinds of food hot	food	yeah so you don't
Sp1m1...	got enough f money for	food	and a couple of drinks
Sp1m1...	I don't each much	food	so that kind of just
Sp1m1...	kind of just covers my	food	no you're quite small
Sp1m1...	I wouldn't eat my	food	of that yeah again thank

Search completed.

Smarter.  
> Sleeker.  
Faster.

# REARRANGE AND RESIZE

#LancsBox X 1.0.0

apple

BNC2014 whole corpus

apple Hits: 4,192 (0.48) Texts: 2,236/88,171

File	Left	N...	Right
...	... Design Museum. Last year,	...	's Jony Ive and I
...	...d place behind Android and	...	s iOS . However, the Nokia
...	the likes of Samsung and	...	If you look at the
...	out? yeah after being the	...	of the eye and everything

BNC2014 informal speech

[hw="f... Hits: 2,863 (2.77) Texts: 705/1,251

File	Left	Node	Right
...	them to pay for my	food	and what else and cos
...	I'm running out of	foo...	no I mean I have
...	I have run out of	food	I don't have anything
...	you've run out of	food	no that's fine that
...	there'll be like the	food	shop they don't even
...	we still need two single	food	place oh yeah I'm
...	is maybe the lack of	food	and maybe they were just
...	and that there or take	food	with you? yeah no eat
...	there selling all kinds of	food	hot food yeah so you
...	all kinds of food hot	food	yeah so you don't
...	got enough f money for	food	and a couple of drinks
...	I don't each much	food	so that kind of just
...	kind of just covers my	food	no you're quite small

BNC2014 whole corpus

Hits: Texts:

File	Left	Node	Right
No content in table			

Search completed.

Smarter.

Sleeker.

> Faster.

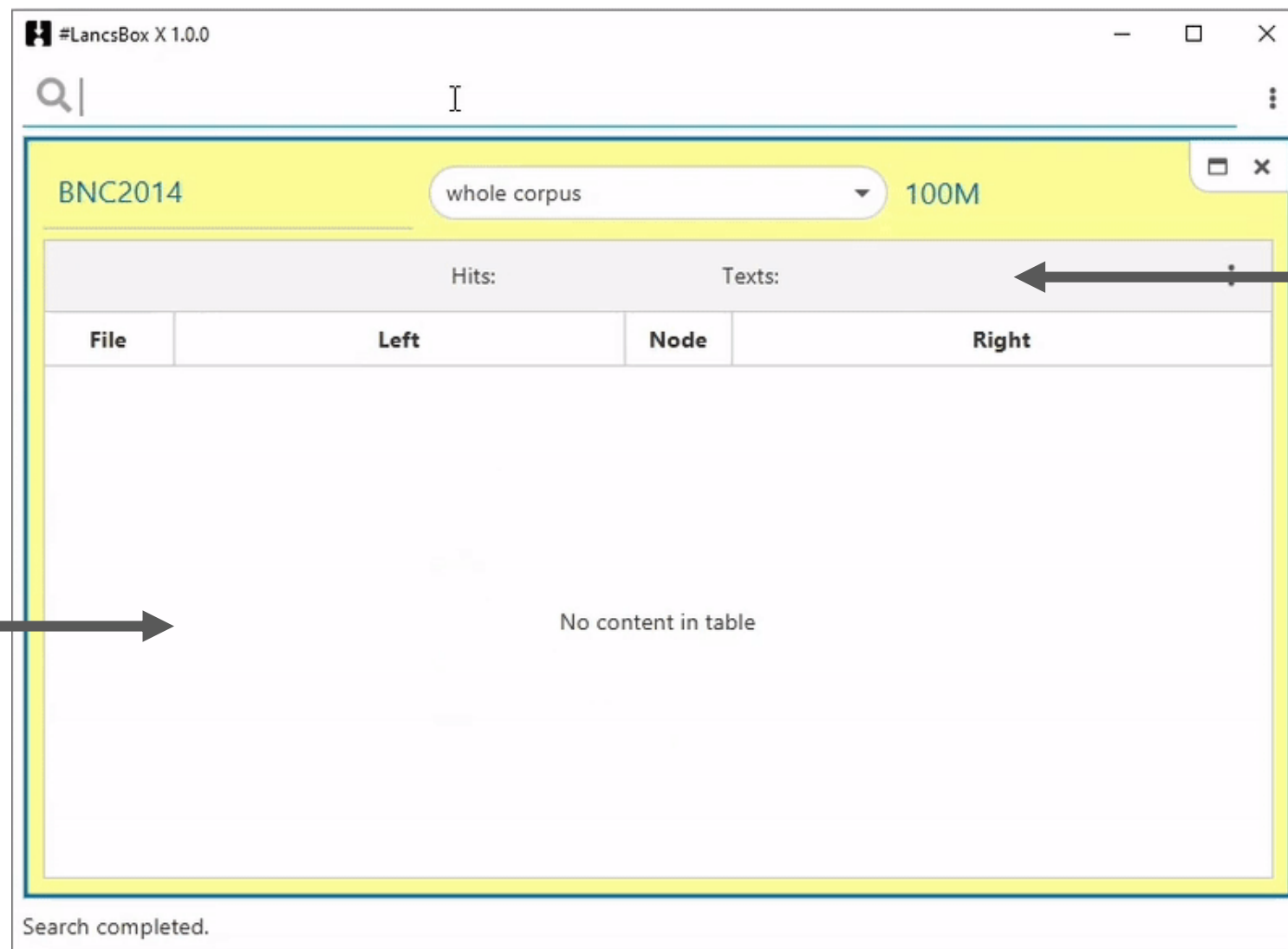


Smarter.  
Sleeker.  
> Faster.

# FASTER RESULTS

No corpus loading

Real-time updates



#LancsBox X 1.0.0

Search | I

BNC2014 whole corpus 100M

Hits:		Texts:	
File	Left	Node	Right
No content in table			

Search completed.

Quick  
stats

