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

Prof Vaclav Brezina

Lancaster University

futurelearn.com/courses/corpus-linguistics

Update

Lancaster University

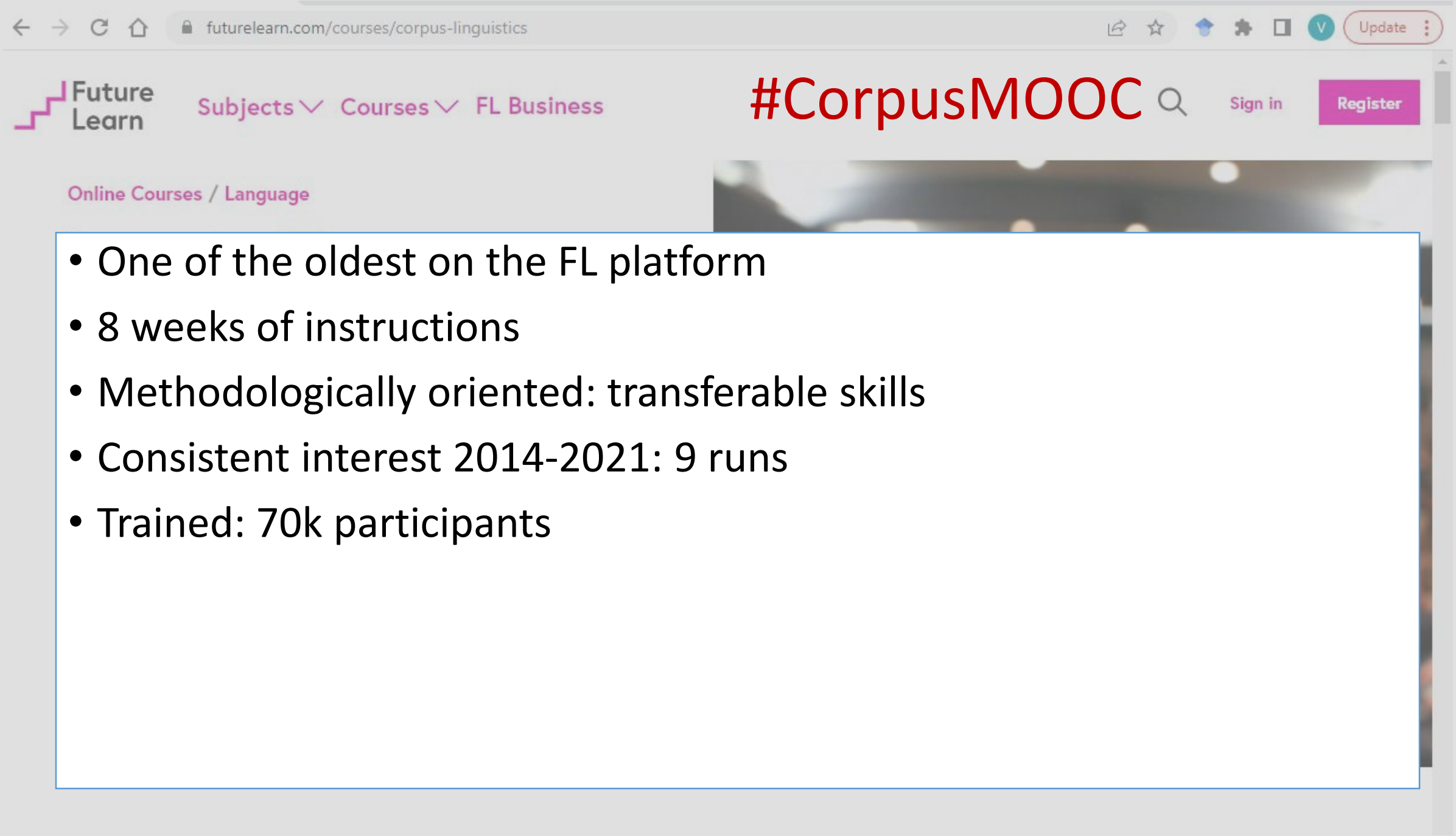
Corpus Linguistics: Method, Analysis, Interpretation

4.9 (95 reviews)

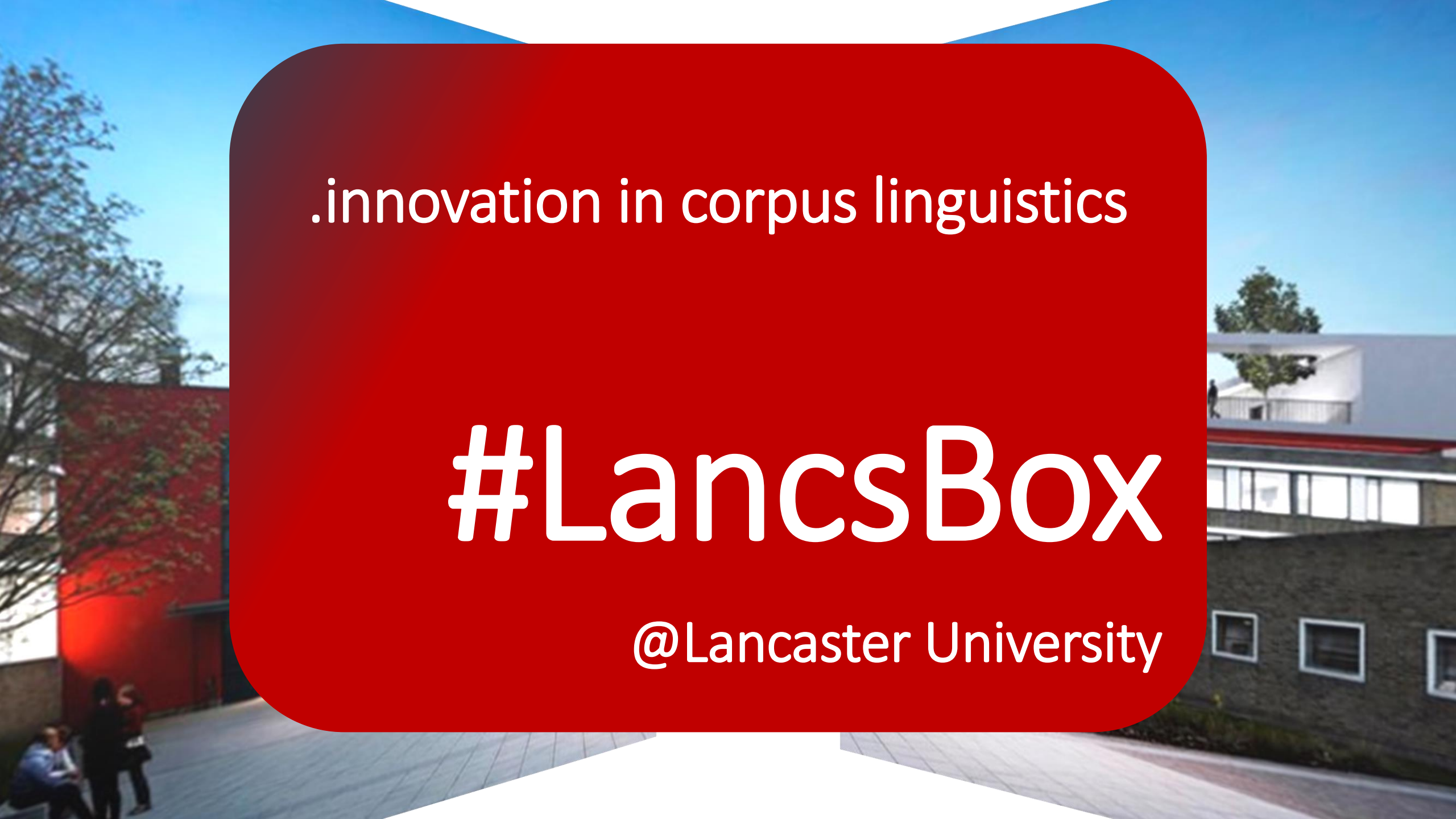
Get a practical introduction to the methodology of corpus linguistics for researchers in the social sciences and humanities.

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

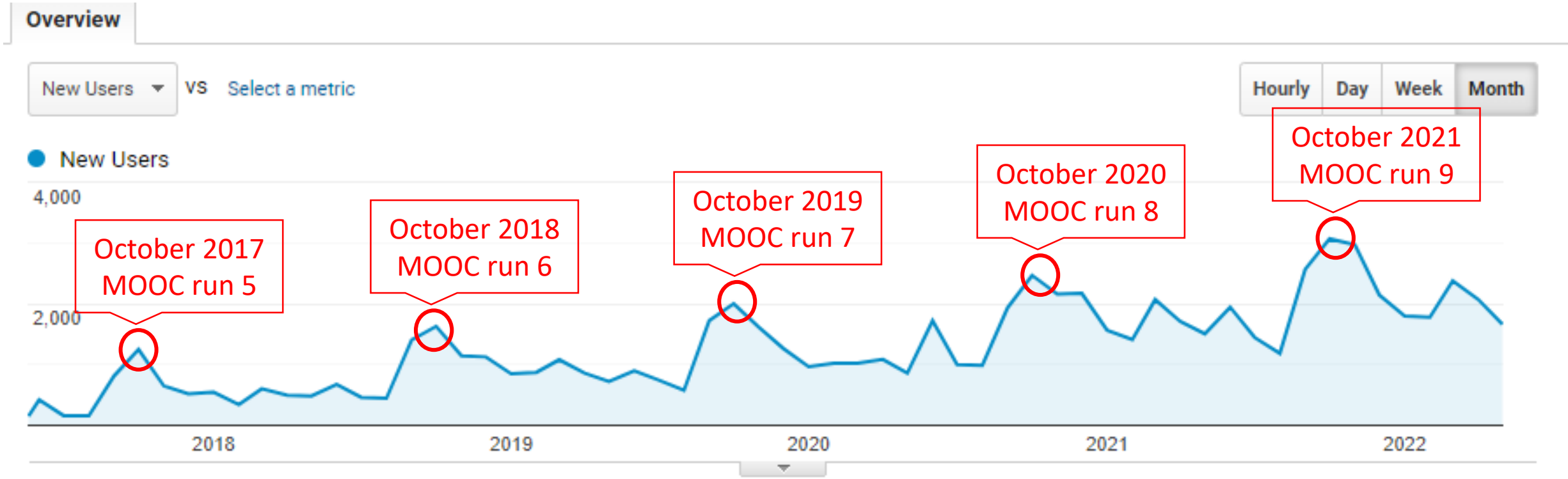
@Lancaster University

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

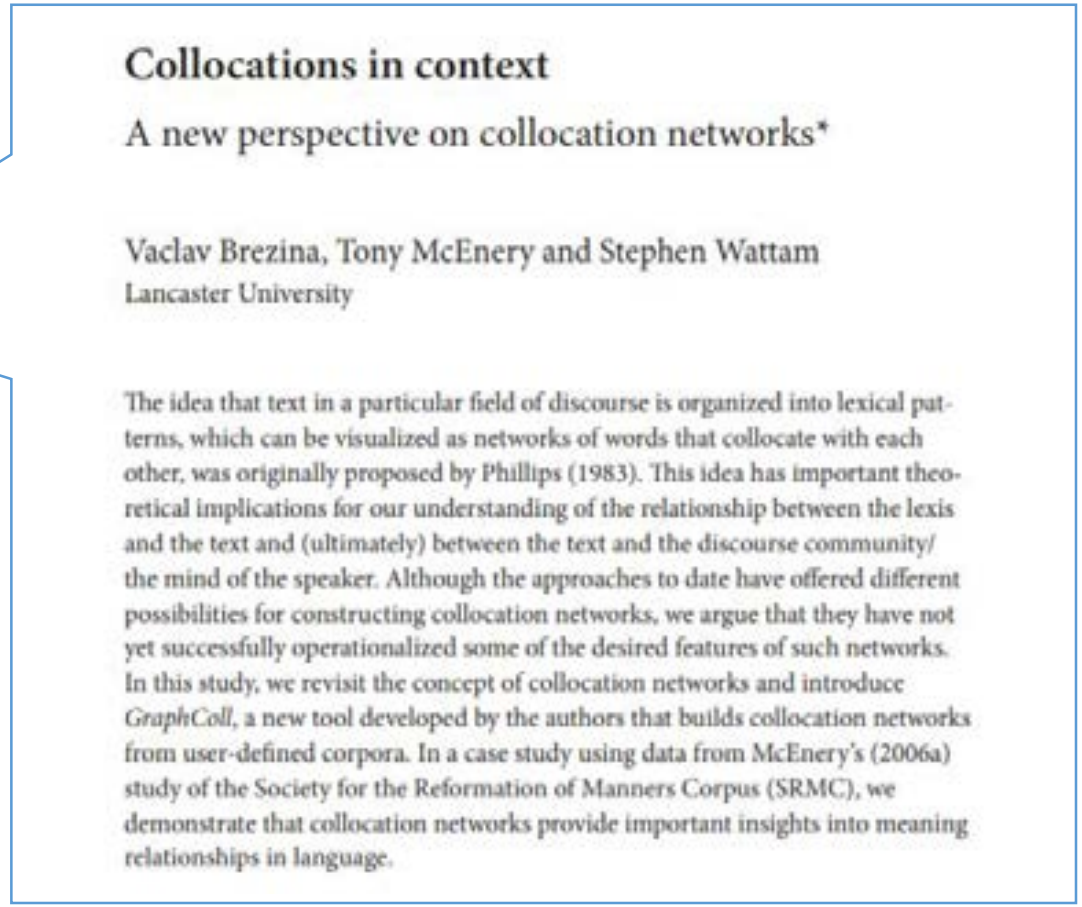
	Country	Users	% Users
1.	China	16,723	21.37%
2.	United Kingdom	9,980	12.76%
3.	United States	6,551	8.37%
4.	Italy	3,806	4.86%
5.	Germany	2,852	3.65%
6.	Spain	2,006	2.56%
7.	Brazil	1,848	2.36%
8.	Indonesia	1,707	2.18%
9.	Poland	1,520	1.94%
10.	Russia	1,519	1.94%

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

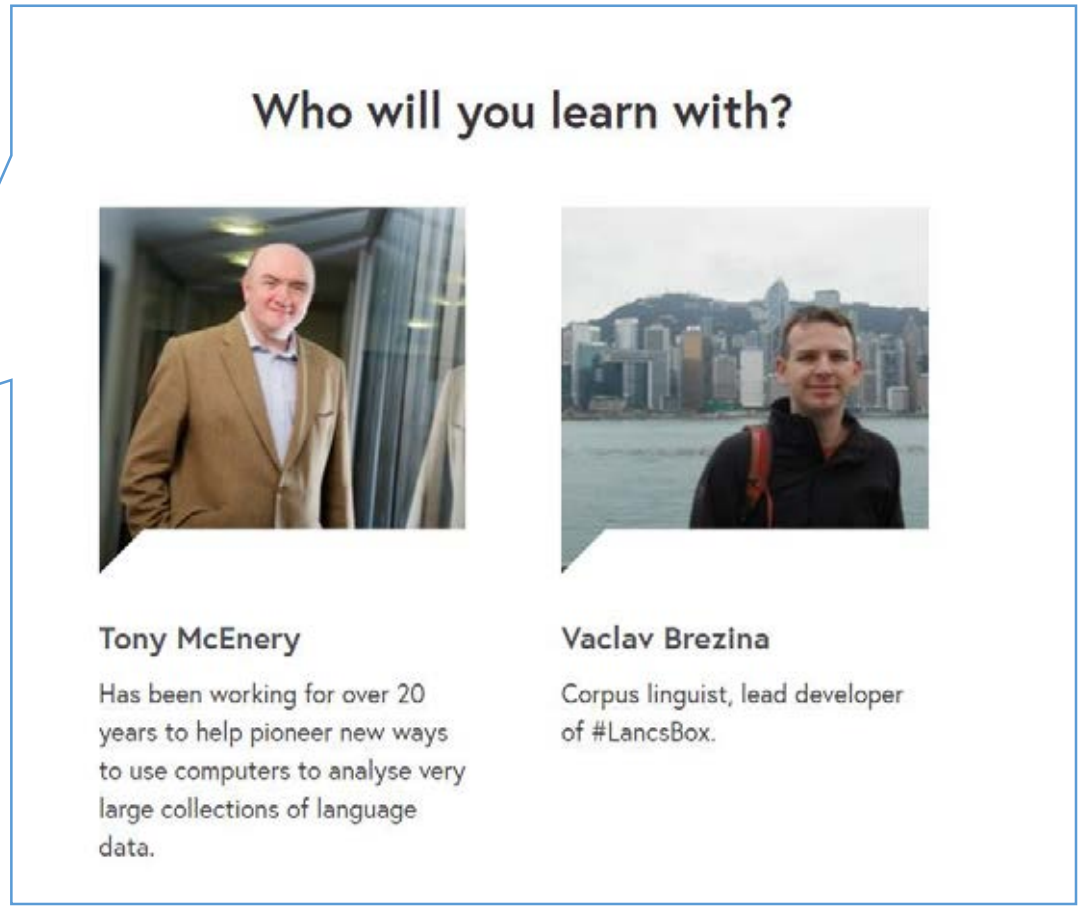
The idea that text in a particular field of discourse is organized into lexical patterns, which can be visualized as networks of words that collocate with each other, was originally proposed by Phillips (1983). This idea has important theoretical implications for our understanding of the relationship 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

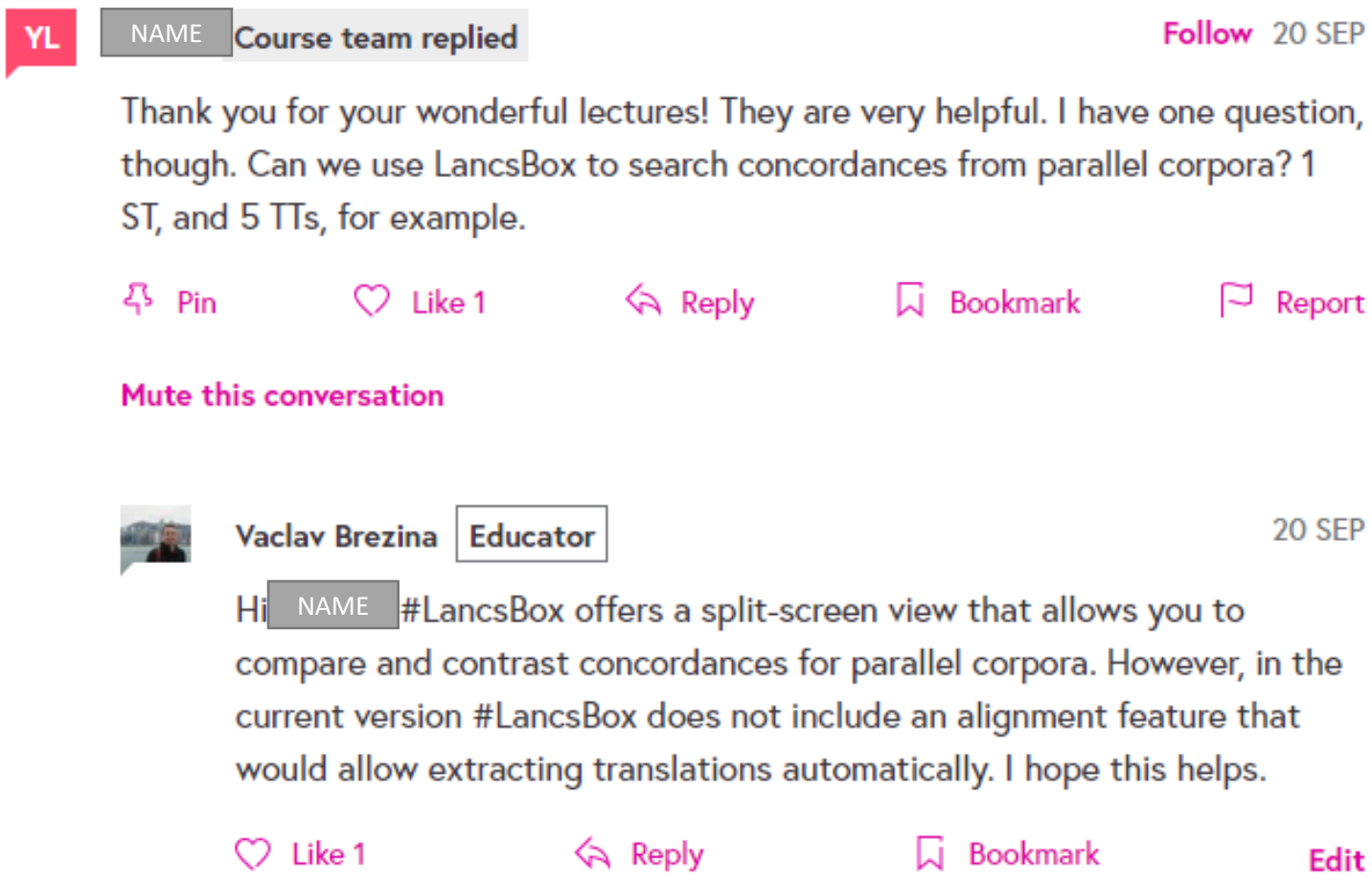
Users: 77,901

Citations: over 500

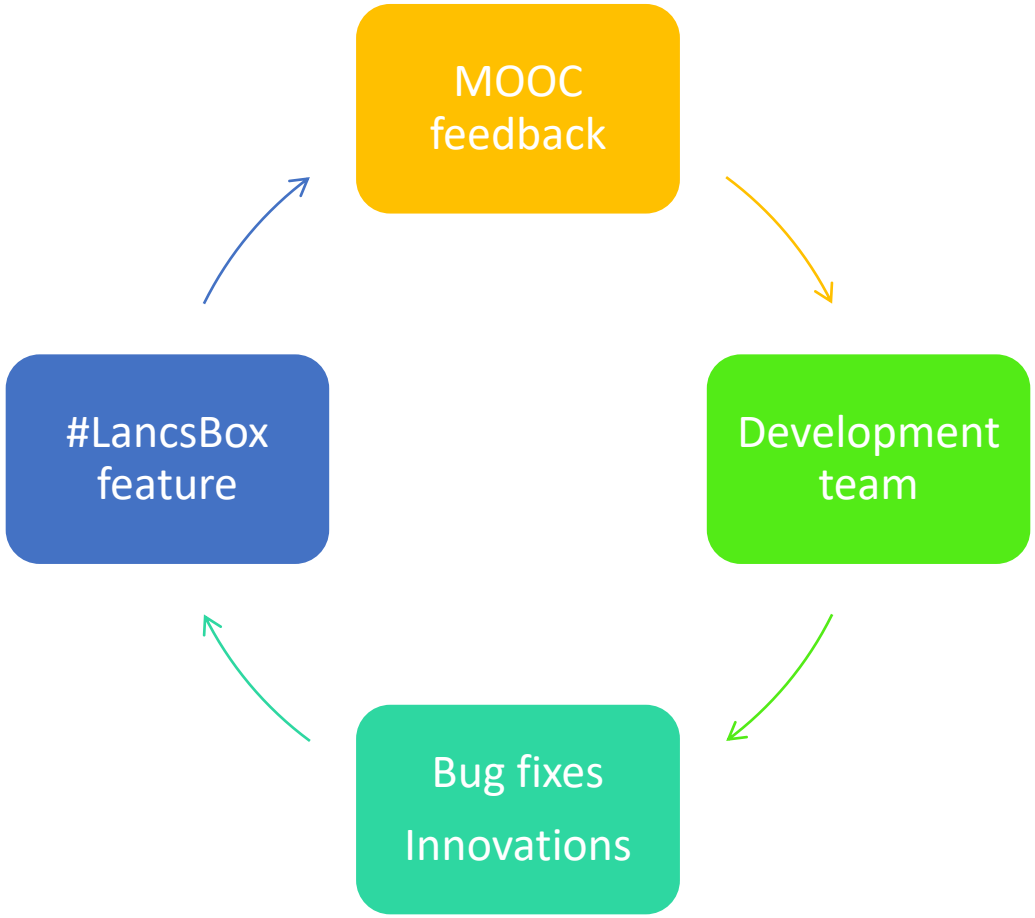
Training events: over 70



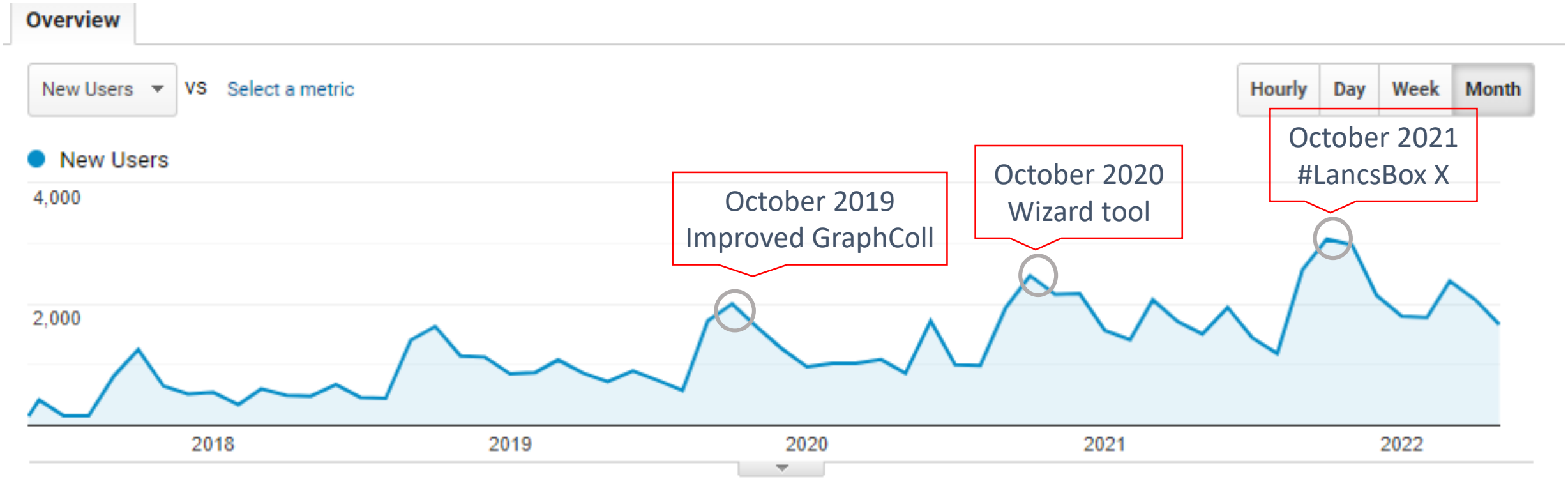
Feedback cycle



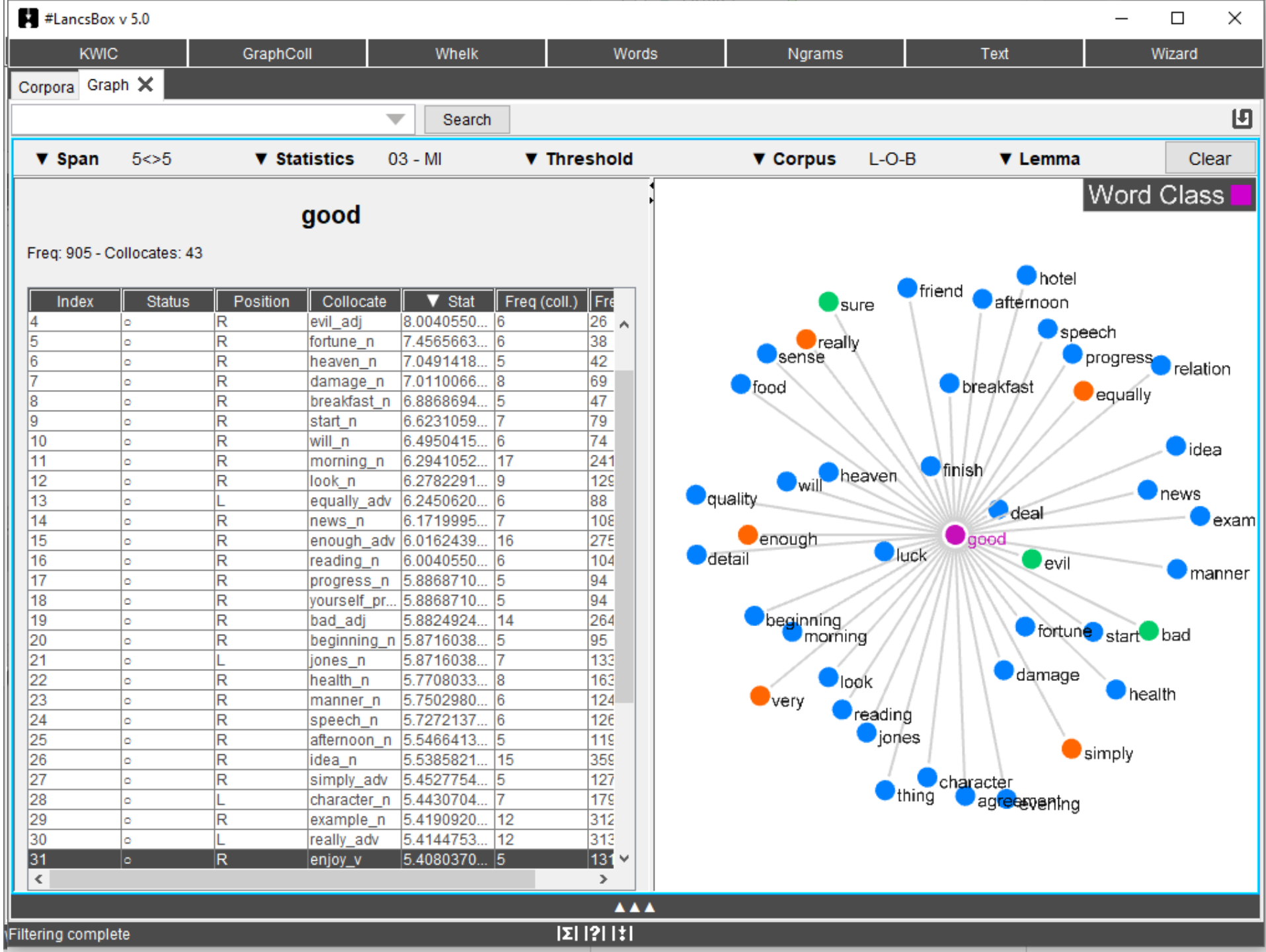
Feedback cycle



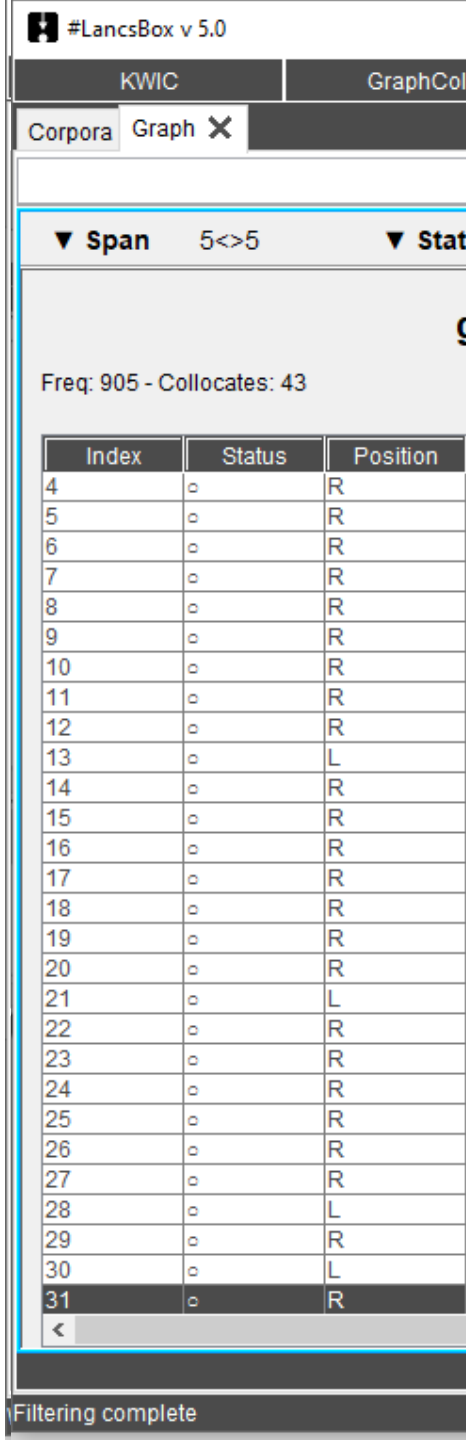
#LancsBox innovations



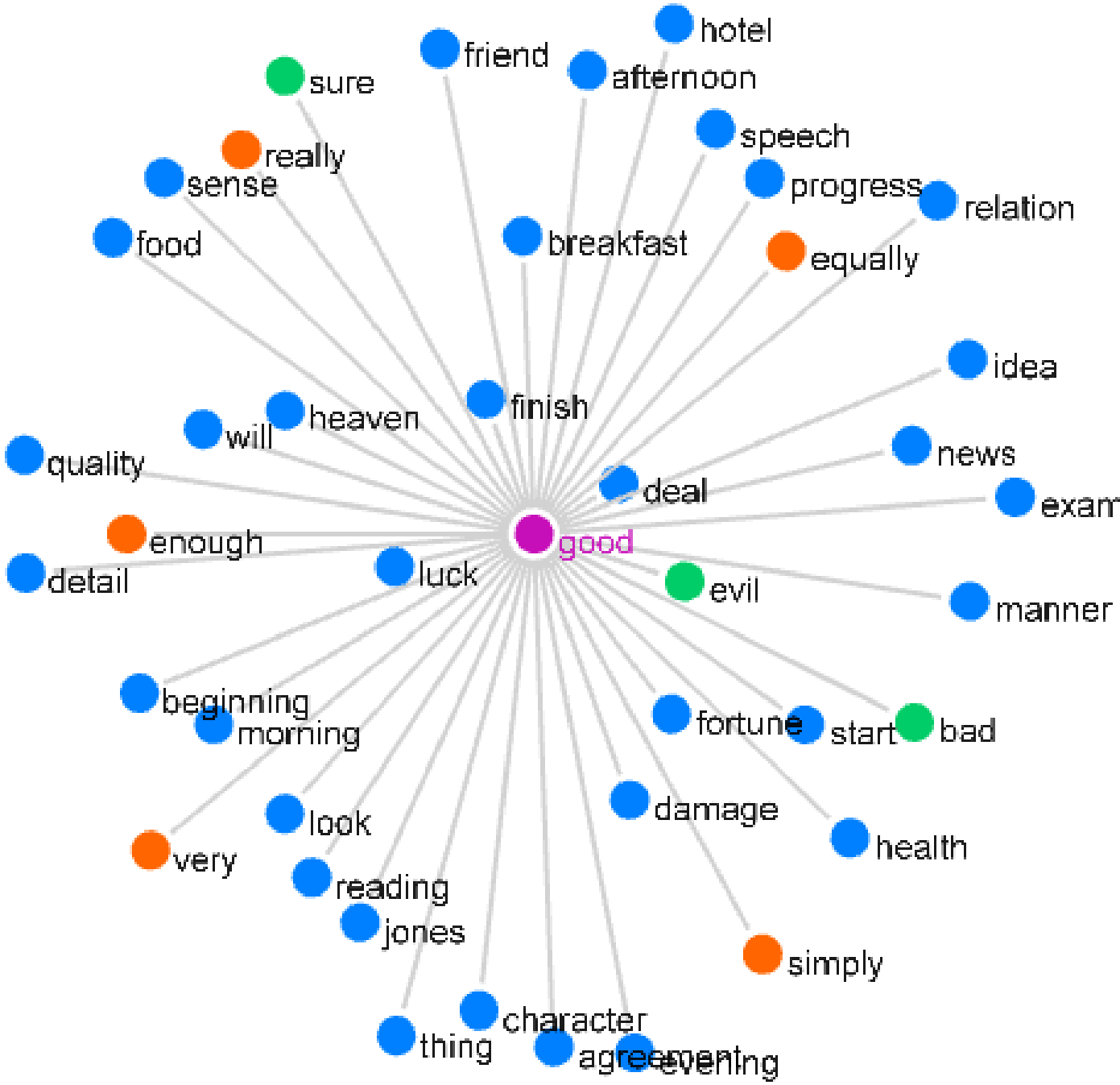
Visualizing MOOC 19



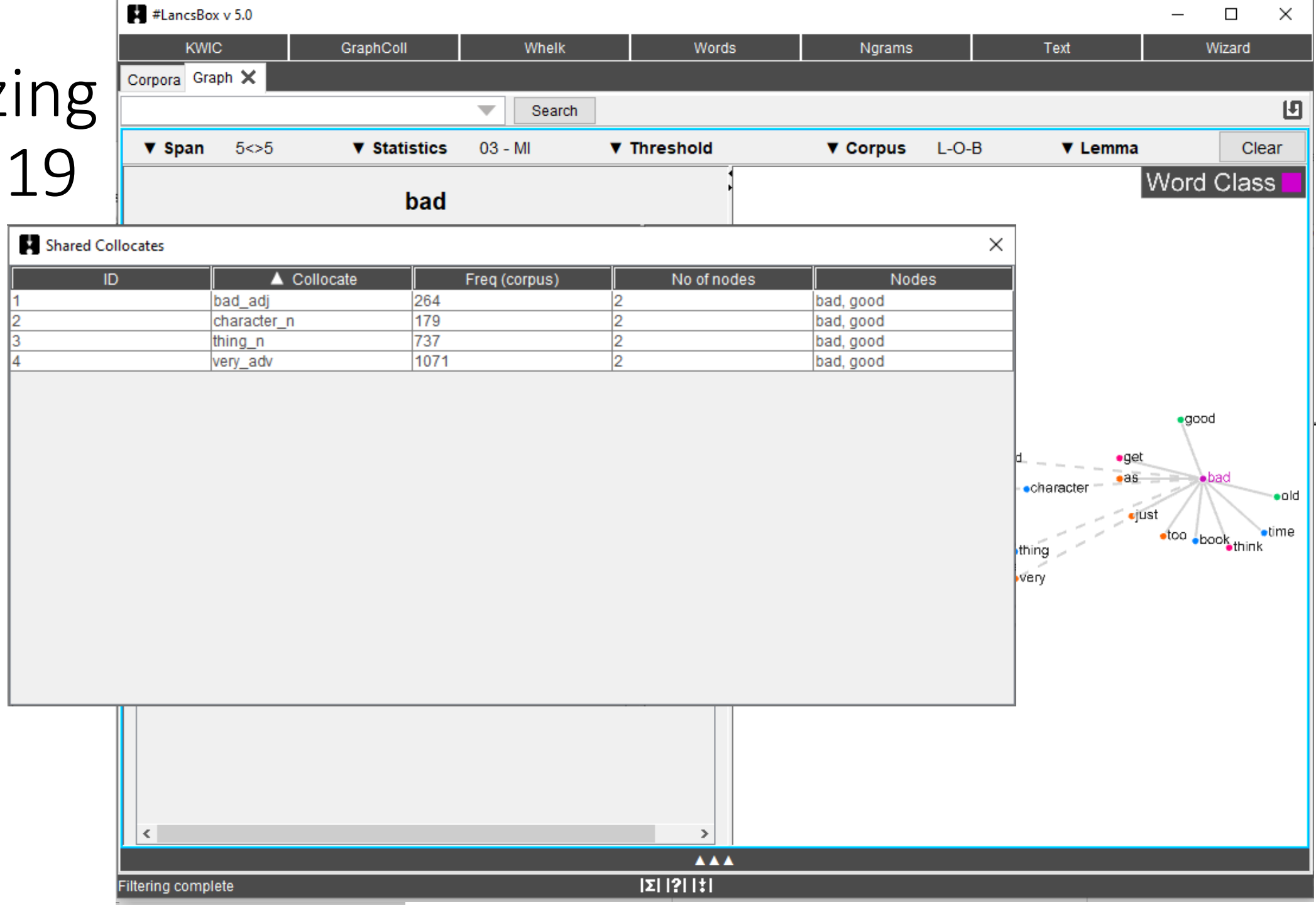
Visualizing MOOC 19



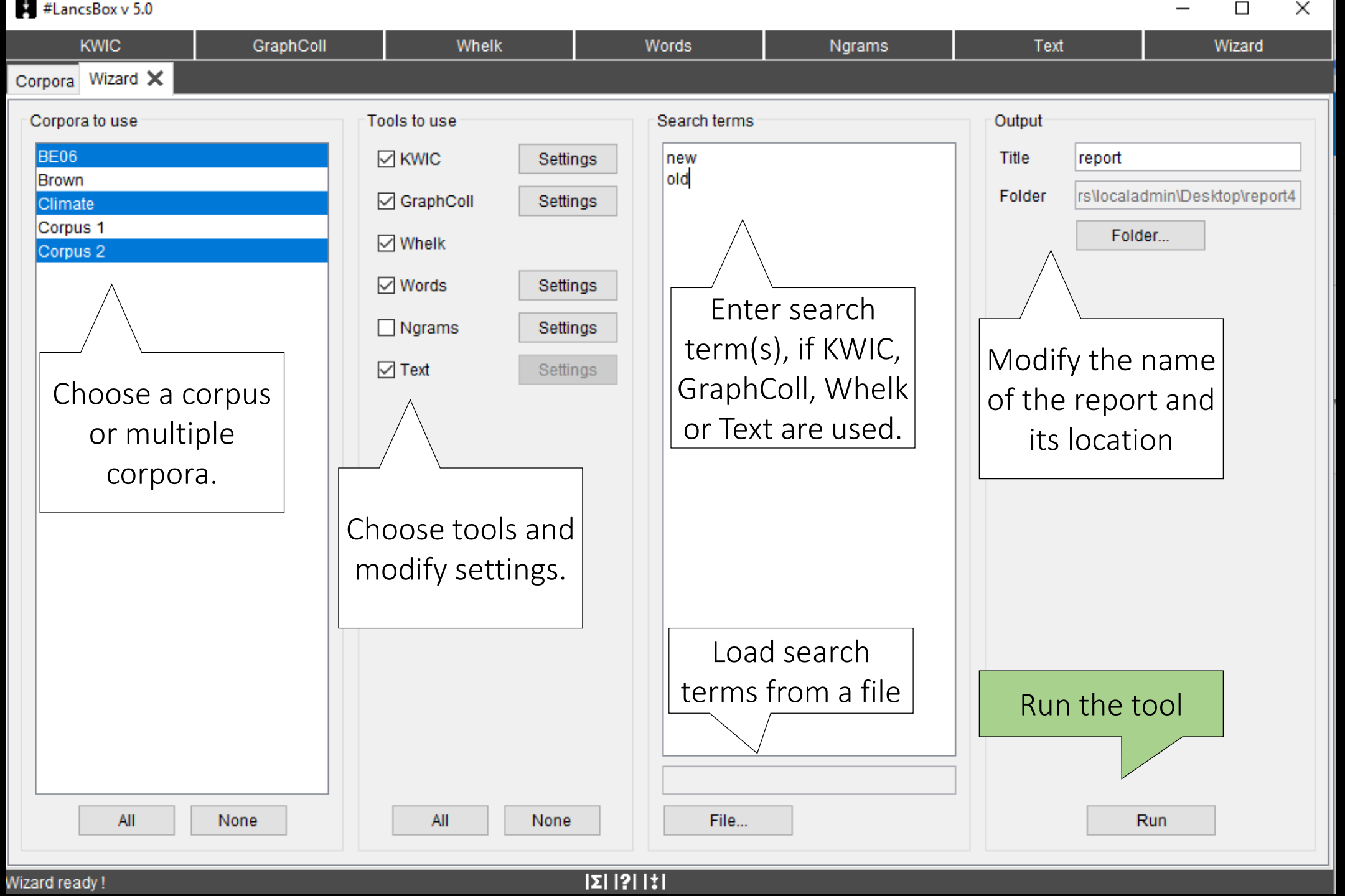
Index	Status	Position
4	○	R
5	○	R
6	○	R
7	○	R
8	○	R
9	○	R
10	○	R
11	○	R
12	○	R
13	○	L
14	○	R
15	○	R
16	○	R
17	○	R
18	○	R
19	○	R
20	○	R
21	○	L
22	○	R
23	○	R
24	○	R
25	○	R
26	○	R
27	○	R
28	○	L
29	○	R
30	○	L
31	○	R



Visualizing MOOC 19



Wizard tool
MOOC 20

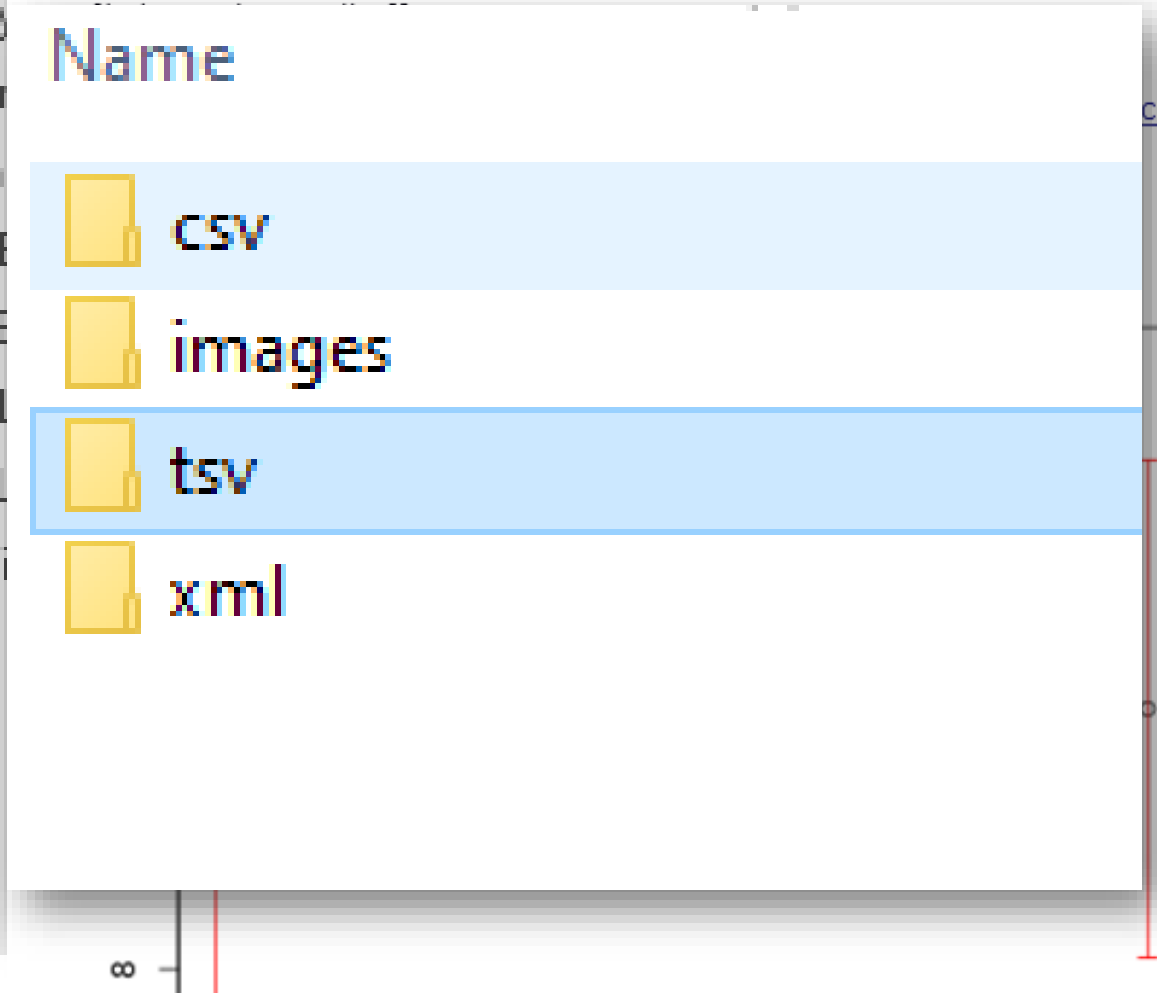

#LancsBox v 5.0
KWIC
GraphColl
Whelk
Words
Ngrams
Text
Wizard
Corpora
Wizard
Corpora to use
BE06
Brown
Climate
Corpus 1
Corpus 2
Choose a corpus or multiple corpora.
All
None
Tools to use
KWIC
Settings
GraphColl
Settings
Whelk
Words
Settings
Ngrams
Settings
Text
Settings
Choose tools and modify settings.
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
Wizard ready !

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

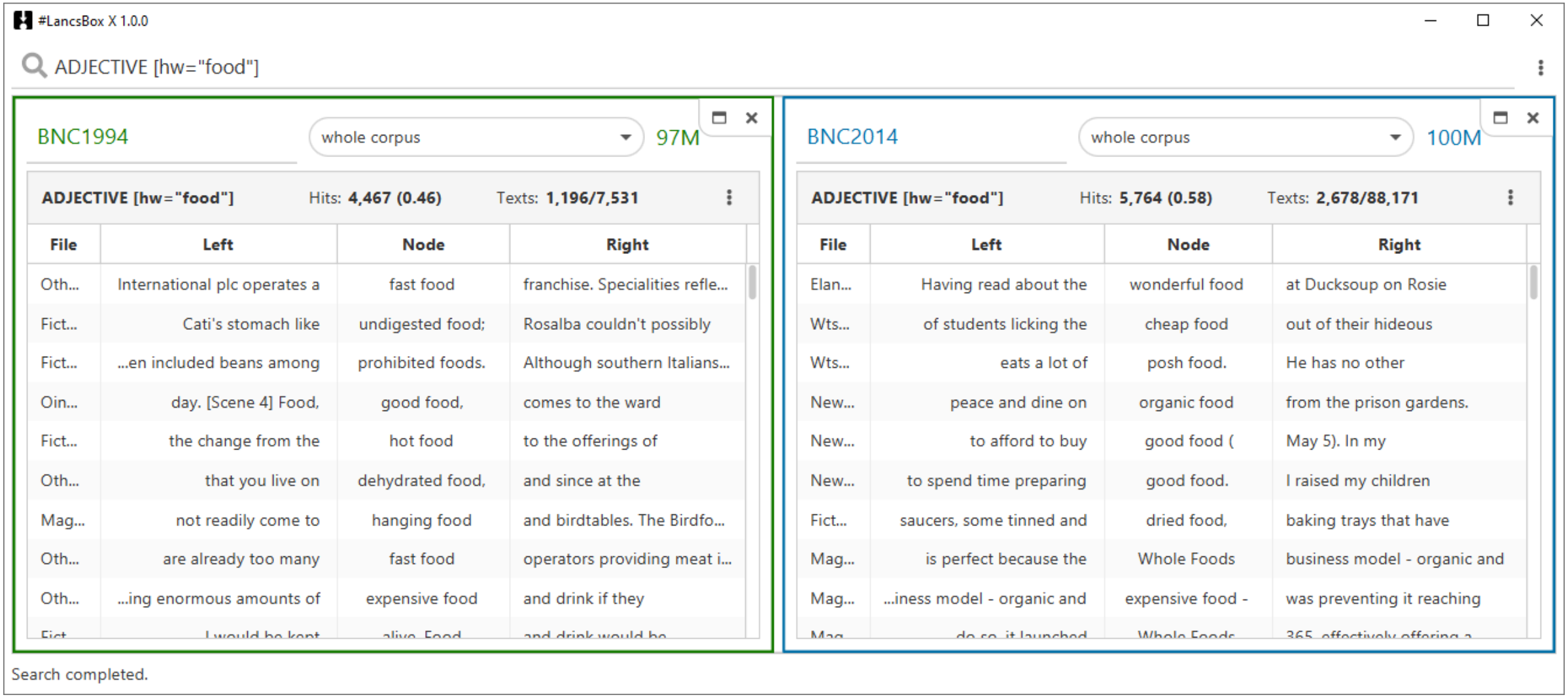
So far, descriptive statistical analysis was reported in the sections above. This includes the analysis of frequency and dispersion (sections 3.1 and 3.3), keywords (section 3.2) and collocations (section 3.4). More details about these procedures can be found in Brezina (2018).

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,



Introducing

#LancsBox X: MOOC 21/22



Smarter.
Sleeker.
Faster.

> Smarter.
Sleeker.
Faster.

X IS FOR XML

<u gender="male"/>

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

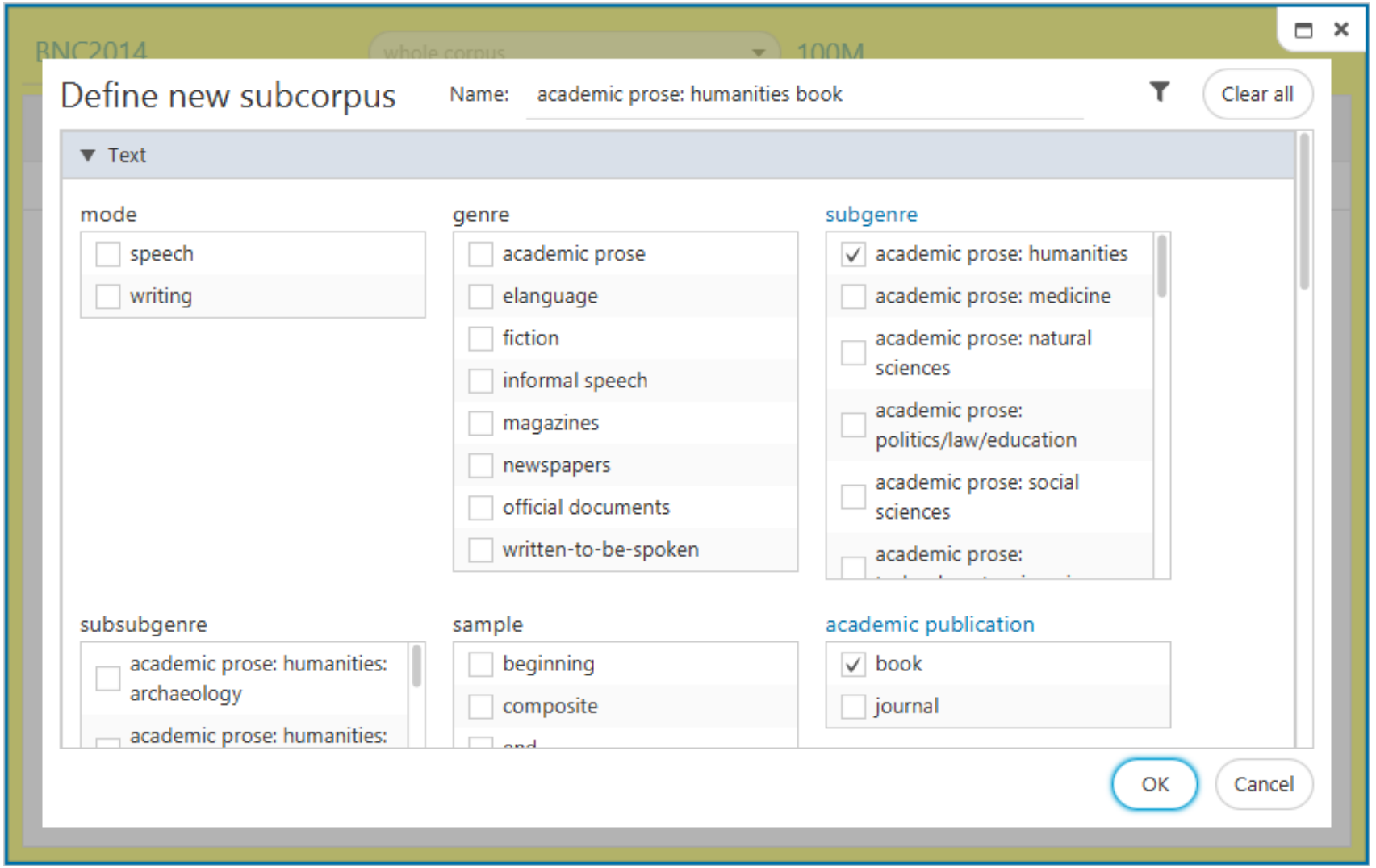
by within PREPOSITIONAL_PHRASE

<s/> within yes

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

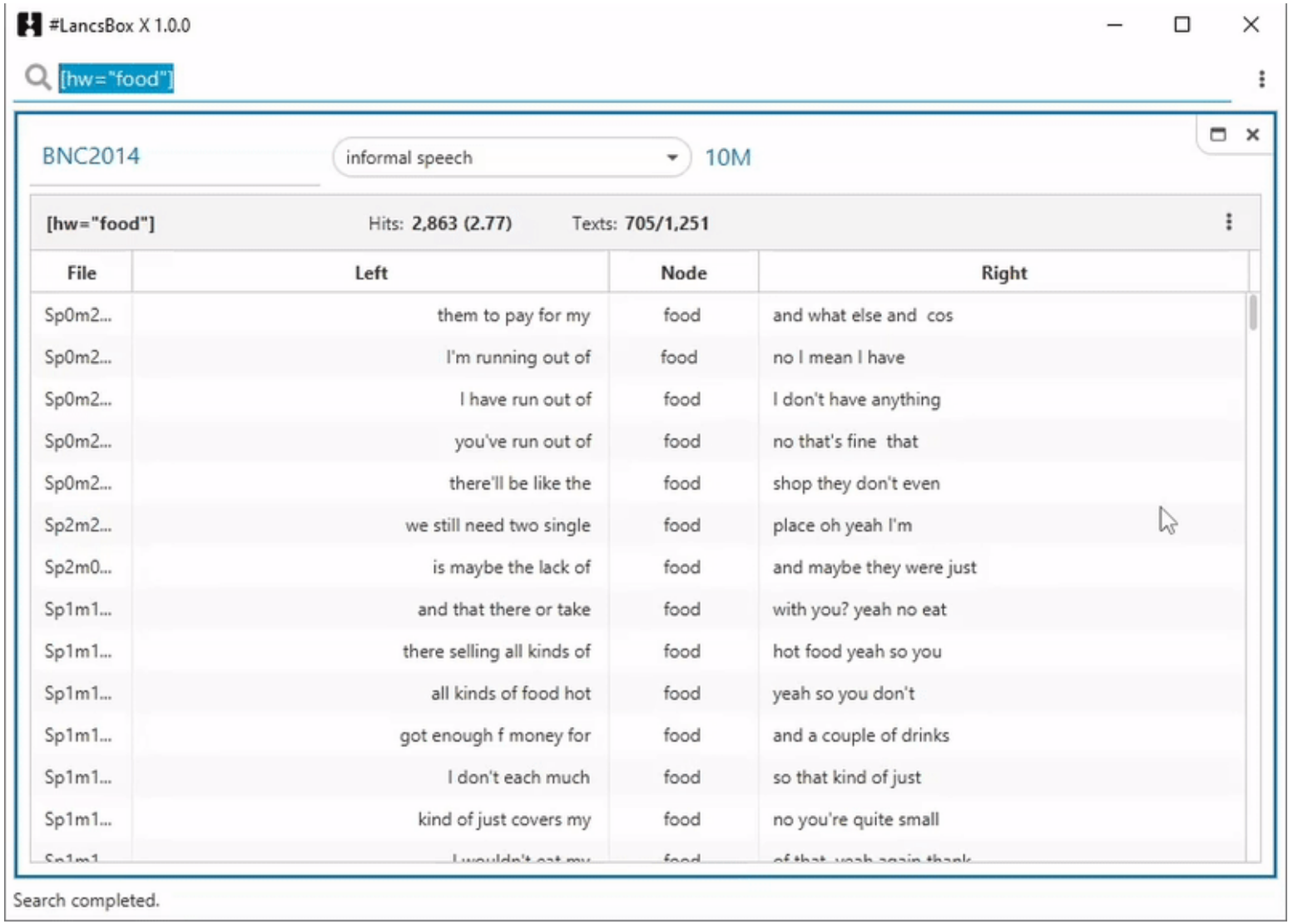
> Smarter.
Sleeker.
Faster.

SUBCORPORA



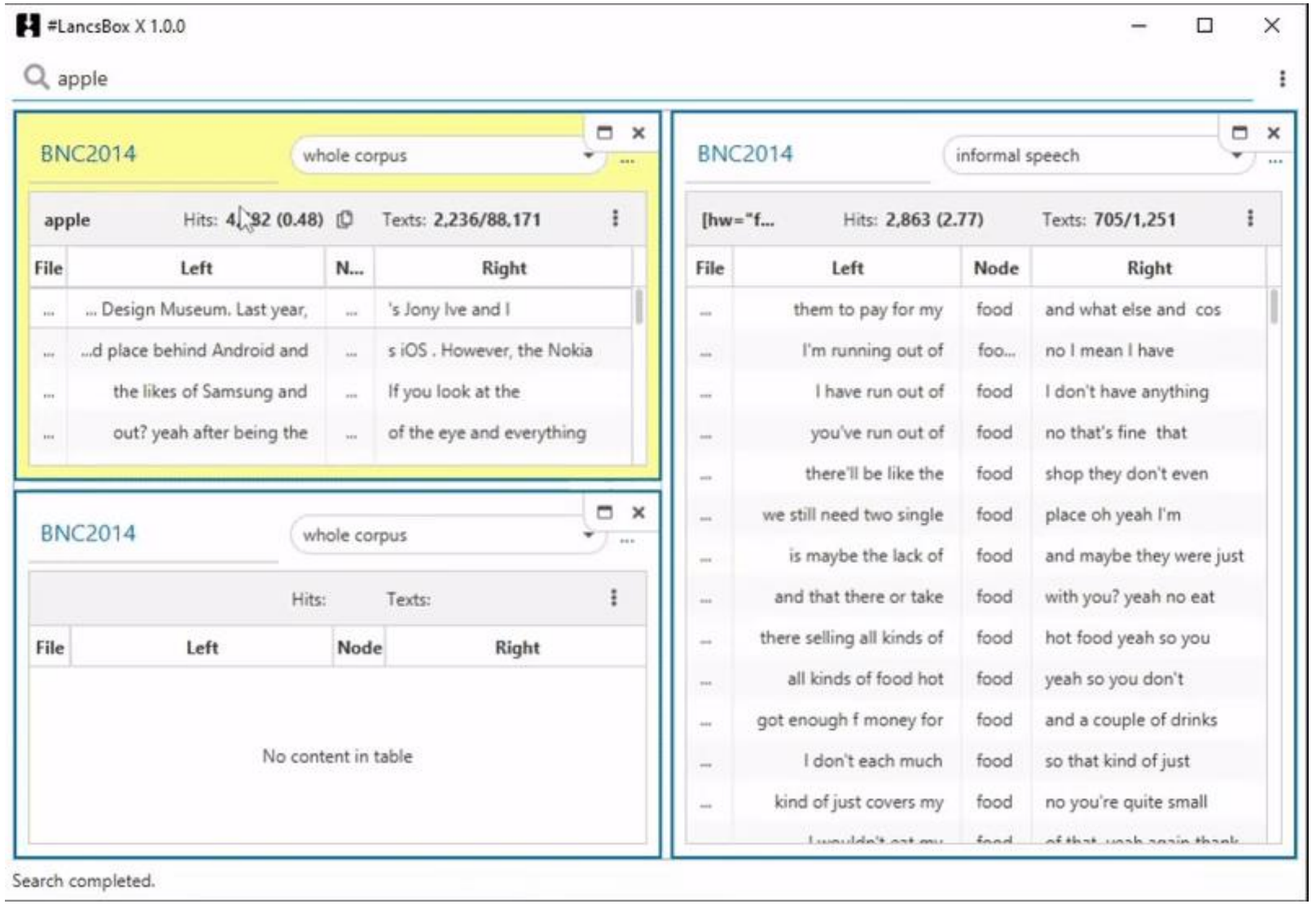
Smarter.
> Sleeker.
Faster.

SIDE-BY-SIDE TOOLS



Smarter.
> Sleeker.
Faster.

REARRANGE AND RESIZE



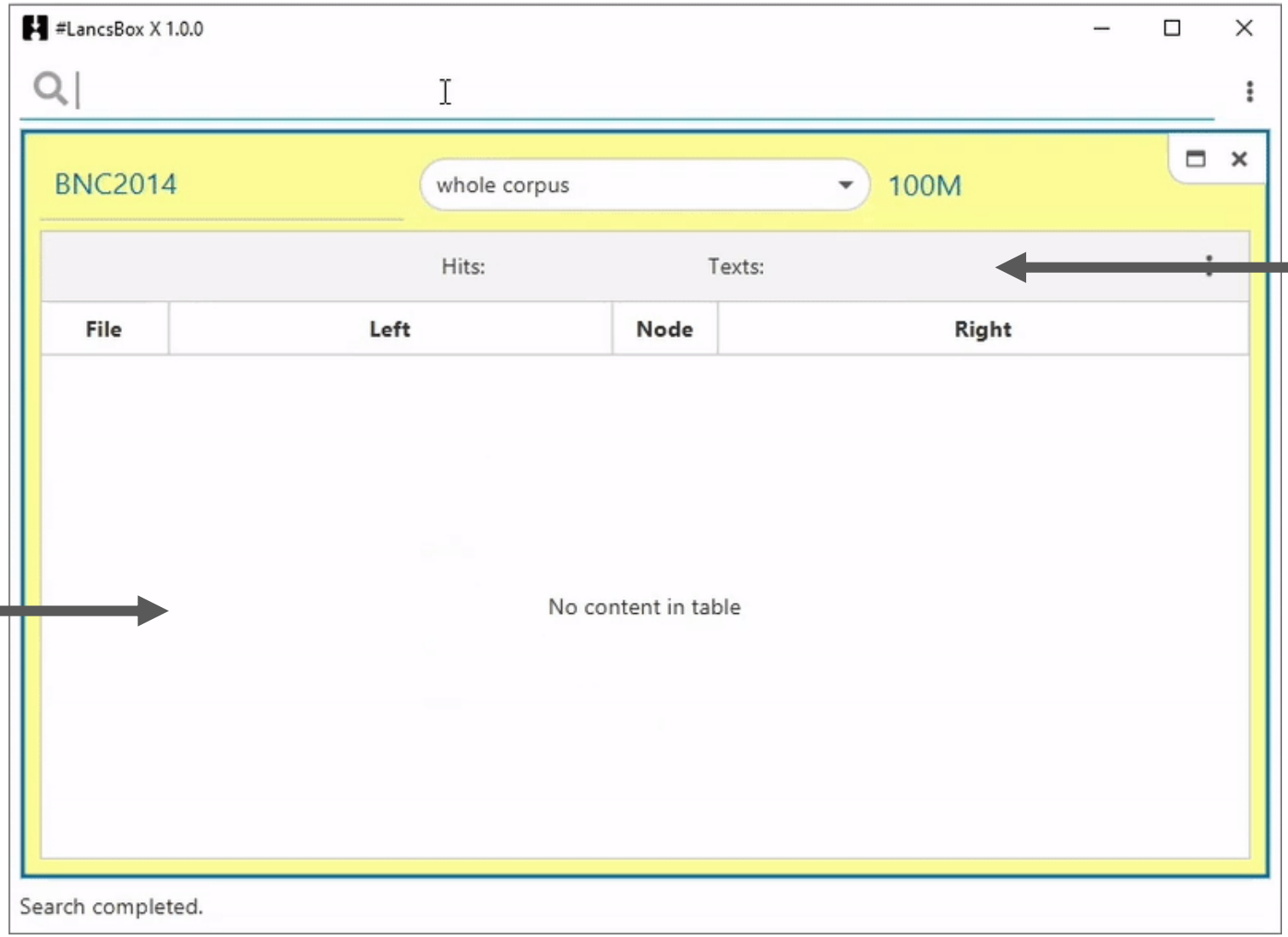
Smarter.
Sleeker.
> Faster.

Smarter.
Sleeker.
> Faster.

FASTER RESULTS

No corpus loading

Real-time updates



Quick stats

for
god
thank
much
goodness
you