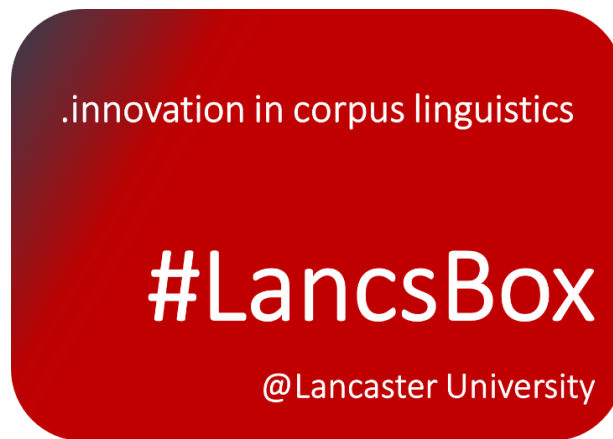


#LancsBox: FAQ

Dr. Vaclav Brezina, Lancaster University

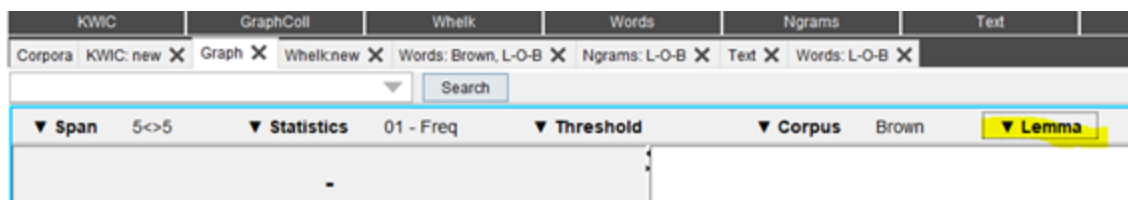


Contents

1. How do I remove function words in GraphColl?.....	1
2. How do I remove <> tags from my corpora?	2
3. How do I save my results?	4
4. My linguistic feature does not appear in the smart searches. What should I do?	4
5. Are my data private/secure on #LancsBox?	4
6. Is the recording of the recent #LancsBox training event (webinar) available?	4
7. I have been using an older version of #LancsBox and want to install an update. Can I keep both the older and the new version?	5
8. What is the maximum size of the corpora that #LancsBox can handle?	5
9. I am a mac user and cannot produce keywords by dragging corpora together in Words?	5
10. Does #LancsBox support right-to-left languages such as Arabic?	5

1. How do I remove function words in GraphColl?

In order to filter graphs according to word classes, you need to change the Unit setting from 'type' to 'lemma' for the tool to have access to the annotation. This setting is available for languages with full support such as English or Chinese.



In the collocate table, apply a filter to the collocate column. In the filter, you can choose a positive or a negative list of word classes e.g. /.*(n|v|adj|adv)/

Please note that everything in forward slashes (/) is interpreted as regex.

2. How do I remove <> tags from my corpora?

By default, #LancsBox disregards tags included in angle brackets <>.

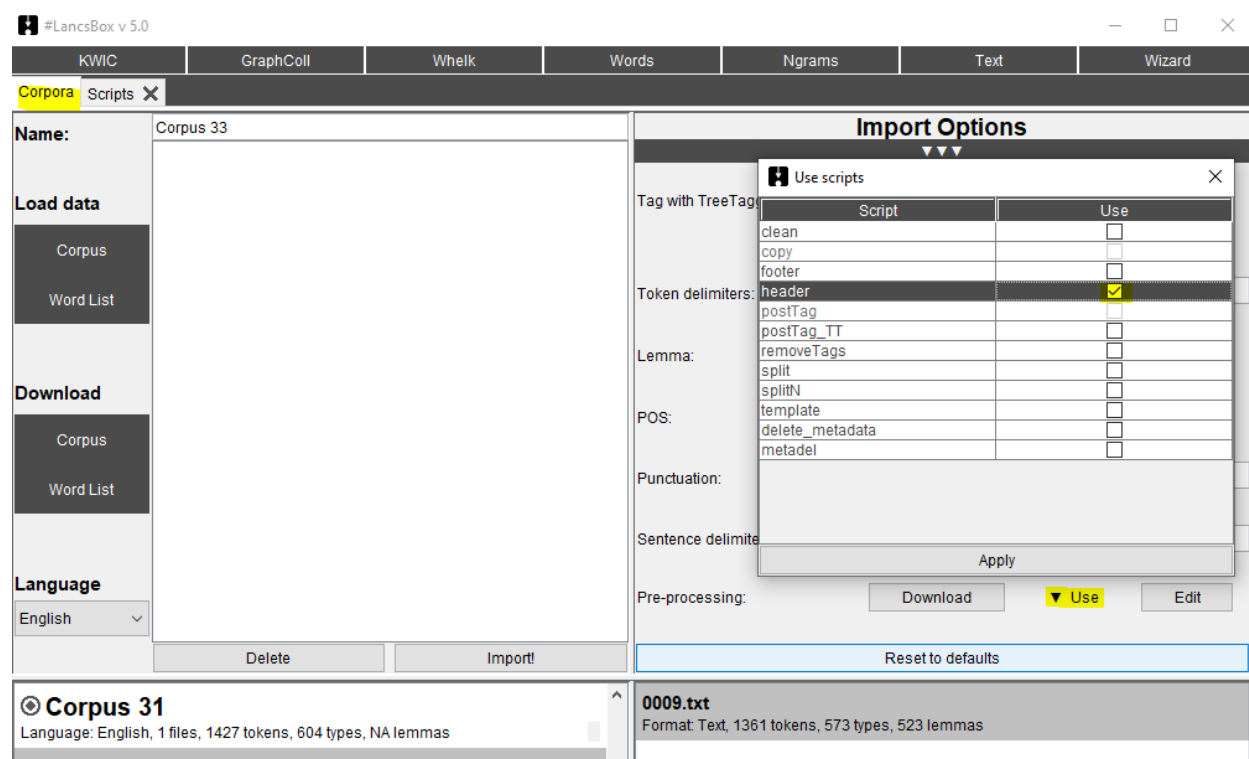
- I. If your files include a header, which you want to delete, you can apply a script inside #LancsBox, which removes the header when loading data.

<header>

some header info

some more info

</header>



- II. If your tags extend over several lines or include some non-standard features, you can download and use a cleaning script called 'metadel'.

<my tag
has multiple lines
and is complicated>

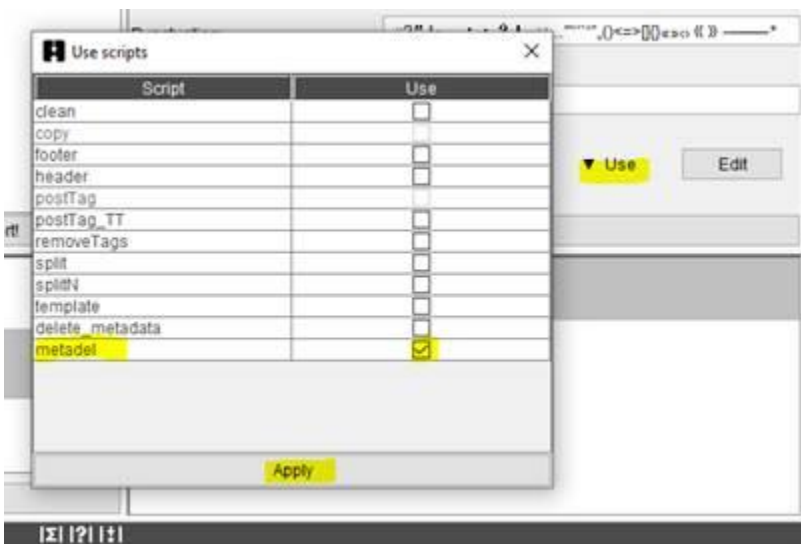
1. Go to 'Import options' on the Corpora tab



2. Under 'pre-processing' click on 'Download'



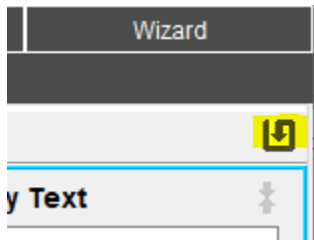
3. Download script that is called 'metadel' (under 'Clean')
4. Under 'Use' select metadel and 'Apply'.



5. Import your corpus.

3. How do I save my results?

That's very easy. Click on the 'Save' icon in the top right corner.



4. My linguistic feature does not appear in the smart searches. What should I do?

Smart searches include only a limited set of common linguistic features. If the feature of your interest does not appear in smart searches, you can still use the functionality of Advanced searches and define the feature for yourself using a combination of token-level, headword and POS searches.



Nice followed by a common noun

5. Are my data private/secure on #LancsBox?

#LancsBox as a desktop tool saves data - texts - only locally (on your machine) so at no point is the data communicated with the server. Nobody but the person or people who has/have access to the local machine can access the data and the data is completely private.

If you use Wizard or the statistics button in KWIC or Whelk, frequencies (a vector of numbers) and names of your variables (search terms) are communicated with the Lancaster server in order to produce statistical analyses. This information is therefore not private.

6. Is the recording of the recent #LancsBox training event (webinar) available?

Yes, this recording can be viewed or downloaded from the #LancsBox website

<http://corpora.lancs.ac.uk/lancsbox/materials.php>.

7. I have been using an older version of #LancsBox and want to install an update. Can I keep both the older and the new version?

Yes. There shouldn't be an interaction with older versions so you can keep these if you like.

8. What is the maximum size of the corpora that #LancsBox can handle?

As a desktop tool, #LancsBox is limited by the resources (disk space, RAM) available on your local machine. On an average laptop, #LancsBox can handle comfortably (with high speed) corpora consisting of several million running words. Above 10M, the speed of processing may be slowed down. The speed of processing data when importing corpora also depends on the number of files in the corpus: the larger the number, the slower the processing.

9. I am a mac user and cannot produce keywords by dragging corpora together in Words?

On some machines with OSX, the interactions in the graphical display are limited (especially in Words and Ngrams). To produce keywords, press the space bar on your keyboard.

10. Does #LancsBox support right-to-left languages such as Arabic?

Yes. If you would like the concordances to be displayed in an appropriate format for right-to-left languages, use the Ctrl+* keyboard combination.