CASS
Corpus Approaches
to Social Science

Practical workshop

Using corpora to teach sociolinguistics

TaLC, Cambridge, 18 July
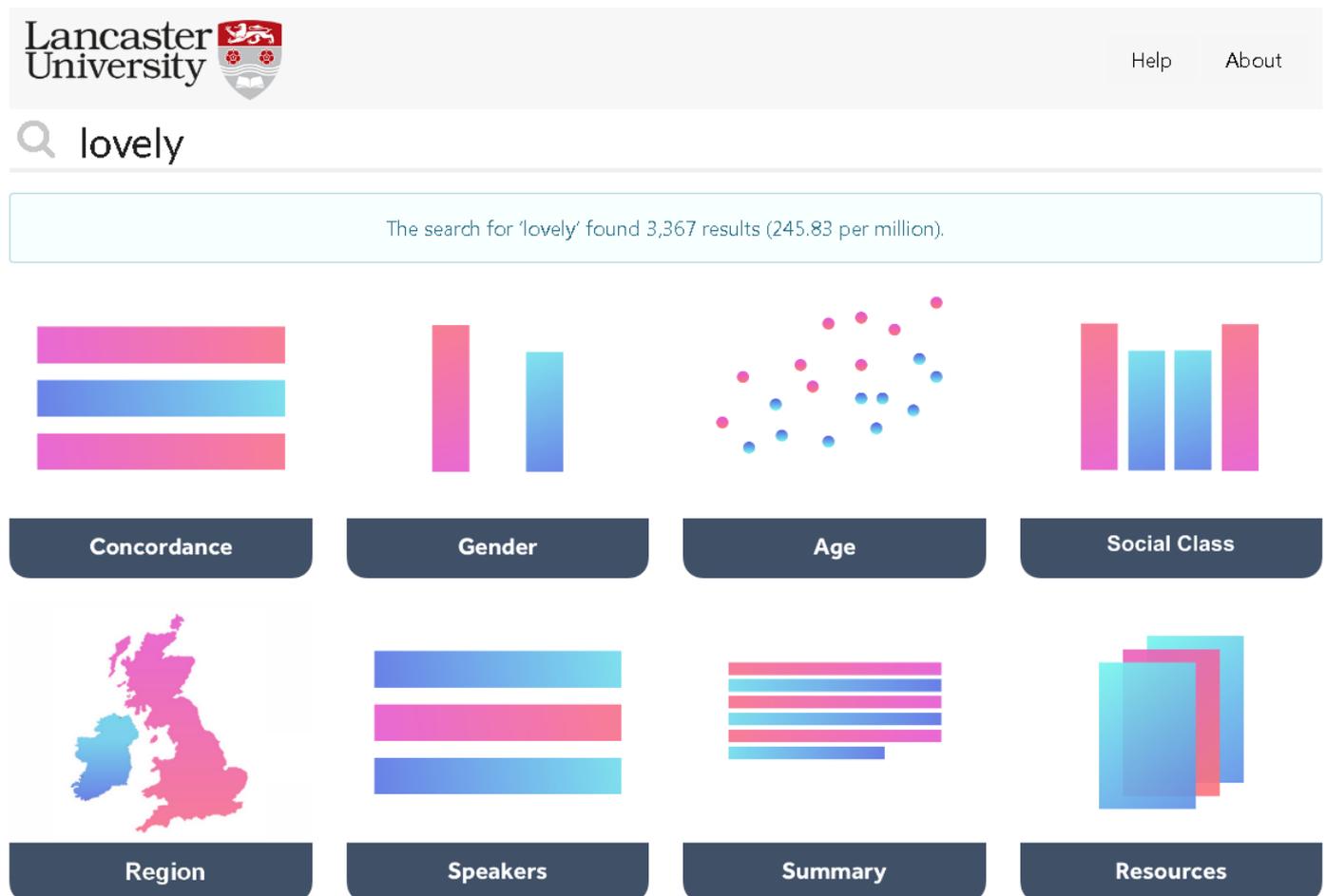
Lancaster University

Help   About

Q lovely

The search for 'lovely' found 3,367 results (245.83 per million).

Concordance

Gender

Age

Social Class

Region

Speakers

Summary

Resources

Dr. Vaclav Brezina & Dr. Dana Gablasova

Beliefs vs Evidence: How do women and men use language?

There are a lot of beliefs about language use and many of these beliefs are related to the gender of the language users. For example, it has been claimed that men speak with more confidence and authority than women, while women speak in a more polite and correct manner. How can we decide whether these beliefs are true or not? In scientific research, we usually look for some evidence that can support of contradict a claim. In linguistic research, one of our key sources of evidence are corpora, large electronic databases of language samples that allow us to investigate how language is really used in communication. Corpora provide empirical evidence about language use – this means that we can directly observe the patterns rather than just draw on our subjective knowledge, experience and intuition. In this worksheet, we will use empirical evidence from corpora to investigate whether some of the claims about how men and women use language are true.

Key terms:　　corpus　　corpora　　language beliefs　　gender　　empirical evidence

Task 1 In this task, we are going to look at the claim that women use emotive words such as 'lovely' and 'cute' more often than men. In BNClab, type in the following words one by one. Go to the *Gender* button and note down whether men or women use the words more often:

- lovely ……………………………………………………………………………………
- cute …………………………………………………………………………………….
- so sweet ……………………………………………………………………………………
- [own word] ……………………………………………………………………………………

What trends related to gender did you notice? …………………………………………………………………………….
………

Task 2 In this task, we are going to look at the claim that men use stronger language than women, in particular, men are believed to use swearwords more often. In BNClab, search for the following words one by one, then go to the *Gender* button and note down whether men or women use the words more often:

- fucking …………………………………………………………………………………
- hell …………………………………………………………………………………
- idiot …………………………………………………………………………………

What trends related to the gender of the speakers did you notice? ……………………………………………………………
………

In some cases, it may seem that gender is the sole factor that can account for the differences in language use. However, when doing scientific research, we should also consider other possible factors and explanations. Let us, therefore, briefly explore the role of social class in the use of swearwords. In BNClab, type in 'fucking' and 'idiot' again, then click on the *Social class* button. Which of the social groups uses it the most in current English?

- fucking ……………………………………………………………………………
- idiot ……………………………………………………………………………

Task 3 In this task, we are going to explore the claim that women use words from certain semantic fields (that is words related by their meaning) more than men. One such semantic area is colour terms. It has been claimed that women use more colour terms than men. In BNClab, type in the word COLOUR in capital letters. This will search for every term related to colour in the corpus.

- How many instances of colour terms are there? ………………………………………………………………………………………
- Go to the *Concordances* button and check which words have been identified as colour terms in the corpus.
- Now go to the *Gender* button and note down whether men or women use colour terms more often………
- You can also have a look at other semantic fields that have been traditionally claimed to be used by men or women such as words related to technology and words related to expressing emotions. To investigate these two semantic fields, type in 'TECHNOLOGY' or 'EMOTION' in BNClab.

Task 4 You have now investigated and observed some trends in language use related to the gender of the speakers. In some cases, men or women tended to use certain features of language more often.
- o Where do you think these preferences can come from?
- o What factors can influence them?

Discuss this with your partner or group.

Space for notes:

Why are people judged by the way they speak?

Attitude refers to people's settled way of thinking and feeling about various matters (e.g. issues, people and events). Language attitudes are beliefs, opinions and prejudices that people have about language. These attitudes can significantly affect people's linguistic behaviour; for example, people can feel strongly about some linguistic features and try to avoid them in their language use. This worksheet looks at language attitudes and how the way people speak can influence how they are perceived and evaluated by others. In particular, it focuses on attitudes to language change. We will discuss where attitudes on language come from, why people have such strong opinions on language and why there is a link between language and our impression of a person.

Key terms   standard language   attitude   language change   ideology

Task 1

In 2017, the new British National Corpus (BNC2014), representing current informal British English, was released jointly by Lancaster University and Cambridge University Press. Following the corpus creation, the linguists behind its compilation publicly discussed some of the changes they had noticed in our spoken language since the 1990s (when the original British National Corpus was released). Among these changes was the higher use of 'like' and 'so' at the start of sentences and an increased use of split infinitives ('to boldly go' instead of 'to go boldly'). The comment below was sent in as a response to an interview on BBC Breakfast from September 2017 in which these language changes were discussed and illustrates the attitudes some people have about language change.

In groups, discuss whether you agree or disagree with the statement.

"As a person who was always brought up to speak properly and grammatically, I was absolutely shocked to read of your pronouncement that starting a sentence with 'so' or 'like' have become part of modern English. According to you, this is the way people speak and is a part of normal everyday speech. Well it is WRONG and people like you should be saying so and encouraging the public to speak properly. I suppose the next thing will be that "Hi there you guys" is quite correct? You set absolutely no example at all. I am sick and tired of hearing terrible English and pronunciation on the radio and television. I also hate the way people pronounce the letter "a" in the northern way, rather than a long "a". Correct and clear speech is so important and sounds lovely.
I await your comments."

Task 2

Look again at the statement above. The writer uses a range of adjective and adverbs to express their attitude on language. Make a list of these adjectives and adverbs.

_____   _____   _____

_____   _____   _____

_____   _____   _____

_____   _____   _____

Now thing about where these attitudes originate. Reflect on the following questions:

• Who decides what the above adjectives and adverbs mean when we talk about language?

• Where do the rules of what is proper and grammatical come from?

• How do these rules relate to how language is spoken?

Let us look at one of the trends in current language use mentioned in the BBC interview, the increase in the use of split infinitive. In BNClab, search for the split infinitive using the following search term:

(to) (ADVERB) (VERB) – This will search for any occurrence of 'to' followed by any adverb and then any verb, findings examples of split infinitives in corpus data.

There are over a thousand occurrences of the patterns in BNClab – we are now going to explore them in more detail.

a. First look at the Concordance tab to see examples of split infinitives. When looking at the examples, what is your own attitude towards split infinitive? Do you find it incorrect? Do you use it often in your own speech or do you try to avoid it? How about your writing? Discuss these questions with a partner or a small group.

b. Go to Gender tab and check the following trends:
• Is the feature really on the increase?
• If yes, who is driving the change, men or women?

c. Go to Social class tab and check how split infinitive is used by different social groups.
• Is there a difference by users from different social classes?
• If yes, how would you explain the difference? What factors may play a role? (e.g. level of education)

Do you agree?

Designed by Freepik

When we first meet a new person, we usually pay attention to that person's characteristics: Are they particularly tall? How are they dressed? Do they like dogs? All of these first impressions help us make up our minds about that person. The way that a person speaks is one of these characteristics, as Miriam Meyerhoff, a prominent sociolinguist, notes:

"We draw very powerful inferences about people from the way they talk. Our attitudes to different varieties of a language colour the way we perceive the individuals that use those varieties. Sometimes that works to people's advantage; sometimes to their disadvantage" (2011:58).

In a study in 1999, American linguists Purnell, Idsardi and Baugh tested these inferences and what effect an accent might have on a person's perceived personality. They had a single speaker emulate three different accents when they called various landlords to arrange appointments: Standard American English, Chicano English, and African American English. Calls in Standard American English resulted in roughly 70% of appointments, while the two other accents only made it to roughly 30%.

Imagine yourself on a bus or train overhearing a conversation from someone behind you who is speaking loudly on the phone with someone else. By the end of the phone call, how much information have you gathered about that person just by the way they speak (age, social background, conversation context, …)? Think of all the various pieces of information that we could learn about a person and discuss this others in your group.

Meyerhoff, M. (2011). Introducing Sociolinguistics. 2nd Edition. London/New York: Routledge.
Purnell, T.; Idsardi, W.; Baugh, J. (1999). Perceptual and Phonetic Experiments on American English Dialect Identification. Journal of Language and Social Psychology. 18/1. pp. 1

Pronouns: The secret life of personal pronouns II

Pronouns are often considered to be a boring word class. Textbooks usually teach us that pronouns are a closed set - this means that, unlike for example nouns and adjectives, the members of the pronoun group do not change and the group does not acquire new members. As a result, personal pronouns are usually learned relatively early and easily in English. However, all of this is not entirely true. When we look at how pronouns are used by English native speakers in their spoken communication today, we can see some interesting patterns and changes over time. In this lesson, we are going to talk about some patterns found in the current informal conversation of English speakers in the UK.

Task 1
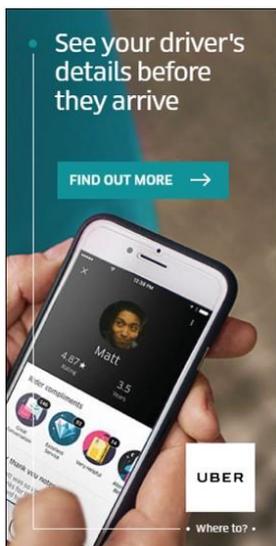
We will consider the changes in use in a well-known pronoun 'they'. This pronoun usually refers to the third person plural such as in this example: *'It was my birthday yesterday and my friends came over for a cake. They didn't leave till well after midnight'*. This is a very straightforward use of 'they'. However, in current English, 'they' is also used for a different function.

Consider the following two examples from current English use in the UK that involve the use of 'they' or 'their' and discuss who these pronouns refer to.

a. Look at the advertisement from Uber and consider what 'they' refers to.

b. Consider also the following extract from an article in *The Guardian* entitled "Why should a primary school separate girls and boys for sports day?" written by Anna Kessel and published on 5 June 2018:
"Back in Inverness one parent reportedly said that their son had been bullied as a result of being beaten by a girl."

Task 2

Read the following concordance lines taken from a corpus of conversations between British English speakers today. Decide in each example whether 'they' refers to the third person singular or third person plural.

1	how are the sprouts ? S: huh ? S: how are the sprouts are	they	acceptable ? S: oh I 'm loving them	
2	accidents you 've seen like bumps and stuff um I have yeah I have yeah I saw someone get their	their	front wheel bent like that was really bad They were alright it was quite a	
3	strong impression based on what you see as to what kind of person	they	are (.) you get huge amount of clues do n't you when you	
4	I like wanted to buy a Minecraft realm with my friends then	they	could go on with me S: well Minecraft costs money	
5	like for a lot of people people do n't like cooking and	they	do n't like the effort S: yeah S: or they just do n't do	
6	hanges it (.) mm I mean like a professional winer person or whatever	they	're called (.) would like (.) professional winer probably not the same	
7	always tell which hand he was writing with the person I sit next to in English broke	their	left hand which was their main hand so they had to write with their right	
8	hint with that (.) I think if you took sent a photo to someone and they	they	stopped replying you would take the hint would n't you	
9	but S: there 's just no point it 's like asking someone if	they	're pregnant if you 're not sure like just do n't like	
10	contacts his lecturer and says oh I 've been talking to this person and	they	're interested in about academic integrity I explained to them how you	
11	hey were travelling abroad and he lives there with her now	they	have a child and erm I was talking to him about it	

Think about the different functions of 'they' that you identified.
• Why do people use 'they' or 'their' to refer to one person?
• Can you identify a pattern when 'they' or 'their' is used to refer to the third person singular?

Task 3

Read the following sentences and change any relevant third person singular pronouns to gender neutral plural pronouns:

Example: The teacher will give you the grade <u>he</u> thinks you deserve.
                  The teacher will give you the grade <u>they</u> think you deserve.

1. You should ask the doctor and she will tell you how much you will need to pay.

2. If a person loves animals, he should protect them.

3. I saw this person on a bus and she waved at me.

4. You should ask your friend what he thinks.

5. You can ask the teacher and he or she will be happy to help.

6. I saw the librarian and she gave me the book.

7. I don't think it's a good idea to ask somebody out when you've just met him.

8. Anybody who litters should pay a fine because he has committed a crime.

Task 4

Pick one of the following topics and debate it in small groups. Make sure to practice using the third person singular form 'they' wherever appropriate.
Or
Pick one of the following topics and prepare a 2 minute speech defending your point of view on the issue. Make sure to practice using the third person singular form 'they' wherever appropriate.

•   Everybody should have access to free healthcare
•   Each child should have at least an hour of sport in school every day Every
•   individual has the right to absolute freedom of speech
•   A doctor should never work more than 12 hours straight without a break

#LancsBox: Analysing corpora in the classroom

Dana Gablasova & Vaclav Brezina, Lancaster University

In this session, we will use #LancsBox, a free multi-platform too for the analysis of language, developed at Lancaster University by a team led by Dr Vaclav Brezina and used in corpus research world-wide. #LancsBox can be used to download students' own corpora (text collections) and for analysing them for lexico-grammatical or meaning-related patterns. In this session, we will look at how DIY corpora can be loaded into #LancsBox and how to search for patterns in these corpora.

Ⓣ **Task 1. Opening and running #LancsBox**. Download or open #LancsBox. If using   #LancsBox on your own computer or laptop, download #LancsBox from <u>http://corpora.lancs.ac.uk/lancsbox</u> and follow the instructions for opening it available on the website.

Ⓣ **Task 2. Downloading the corpora to be analysed in #LancsBox.** The starting point of the tool is the 'Corpora' tab where data is loaded and further processed. There are two options for loading corpora: i) loading **local corpora** from your computer and ii) downloading **corpora provided** by #LancsBox. In this session, we will use corpora available in #LancsBox.

The following corpora are already included in #LancsBox:

- **LOB**: Lancaster-Oslo-Bergen corpus: a 1M words corpus containing 15 genres of written British English.
- **Brown**: a 1M corpus of American English, having the same composition as the LOB corpus.
- **BNC64**: a 1.5M corpus of informal British speech.
- **Shakespeare corpus**: a corpus of all Shakespeare's plays.

We will first work with one corpus (LOB) in order to explore different concordancing functionalities in #LancsBox. In the later tasks, you can work with one of the corpora that have been preloaded for you (available in the 'Corpora' folder):

Downloading a corpus provided by #LancsBox. We will first work with LOB, a one million-word corpus of British English from the sixties in order to explore different #LancsBox functionalities. Follow these steps to download the first corpus:

1. Within the 'Corpora' tab, left-click on 'Corpus' under 'Download'.
2. This will open a list where you can select a corpus. <u>Select LOB</u> and click on 'Download'. Left-click 'Import!' to import LOB into #LancsBox.
3. The corpus will now be imported into the bottom-left section, alongside any other corpora you have already imported.
4. On this screen, you can also see the information about the size and lexical diversity in your corpus. How large exactly is LOB?...

Task 3. Doing concordance analysis in #LancsBox. Concordance lines allow us to identify all occurrences of the target word, phrase or grammatical structure in the corpus and to display them with their context. This helps us identify patterns in language use.

Let us now explore different kinds of searches available in #LancsBox. Go to the **KWIC module** in #LancsBox and do the following searches and note down the frequencies of the search terms:

1. Simple search. This type of query searches for the word as we type it in. Search for the word 'cool' in the corpus and note down the following:

- How many times does your selected word/phrase occur in the corpus? …………………………………………..
- What is the relative frequency in the corpus per 10,000 words?...
- Why do we need relative frequency?
- How can we use this type of search to analyse language use? (Discuss with your neighbour)

2. Searching for lemmas (headwords). 'Lemma' refers to a word with all its inflectional affixes, the same as an entry in a dictionary. For example, if we search for the word 'go' as a lemma, it will find all of its forms such as 'go', 'went', 'going' and 'gone'.

a. First, use the simple search to search for the verb 'be' (i.e. type in 'be'). The frequency of 'be'……

b. Next, search for the verb 'be' as a lemma. Use the arrow in the search window to open advanced search options. In the 'Headword', type 'be'. The frequency of 'be' as a lemma……………………………………………

3. Searching for word classes. You can search for different word classes by typing in the word class category in capital letters, e.g. NOUNS. #LancsBox supports the following word classes:

NOUNS	MODALS
PROPER NOUNS	CONNECTORS
VERBS	PRONOUNS
ADJECTIVES	

Note: When students upload their own corpora into #LancsBox, they get automatically analysed for word classes and can be searched for the.

- How many modal verbs are in the corpus? ……………………………………………………………………….
- How many adjectives are in the corpus? ………………………………………………………………………..
- How could our students use this type of search to analyse their language samples? (Discuss with your neighbour)

4. Searching for linguistic categories. The tool supports smart searches for linguistic categories. You can search for the grammatical categories from the table below by typing in the category in capital letters, e.g. PASSIVES.

PASSIVES	SPLIT INFINITIVE
COMPLEX NOUN PHRASE	PRESENT TENSE
PAST TENSE	PLACE ADVERBIALS
NOMINALIZATIONS	TIME ADVERBIALS

- How many passive constructions are in the corpus? ....
- How can our students use this type of search with their own data? (Discuss with your neighbour).

Task 4. Explore the context in which a word/phrase appears:

Search for the word 'computer'. After the concordance lines are displayed, click on the 'Left' column header and then on the 'Right' column header to sort your results according to the left and then according to the right context.
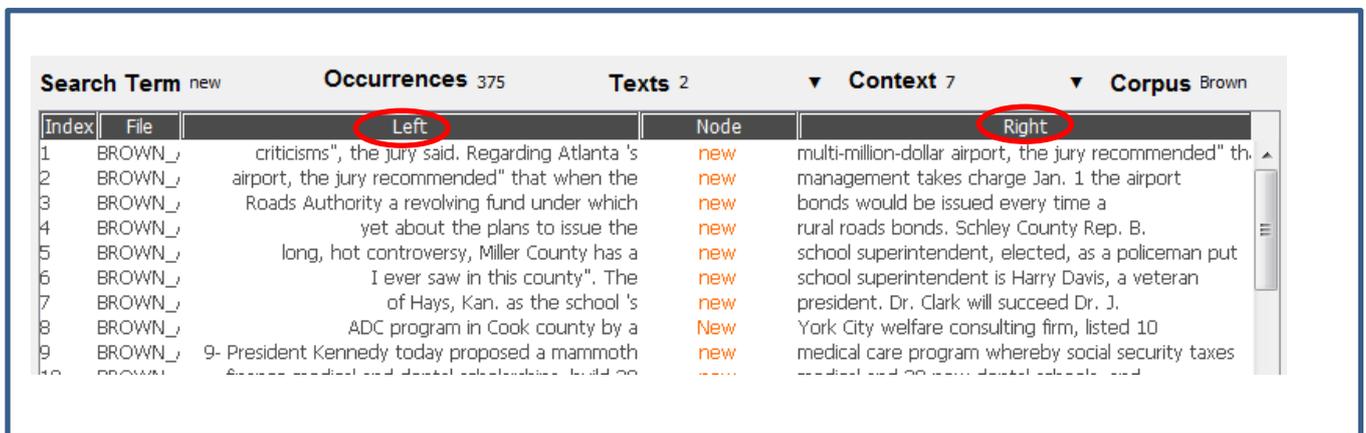
- Can you see any typical contexts in which your target word/phrase occurs? (Tip: You can adjust the number of words appearing in the context by clicking on the arrow next to *Context*).

  ………..

  ………

  ………

 ………

Task 5. Comparing two corpora in #LancsBox: Comparing British and American English.

First, load a second corpus into #LancsBox, following steps from Task 2 (when you loaded LOB). This time, load **the Brown corpus**, a one-million corpus of American English, collected in the 1960s. Brown is a mirror corpus to LOB (this means they have been compiled according to the same principles) and thus we can use the two corpora to directly compare the use of British and American English in the same time period.

1. Upload the second corpus into #LancsBox using the steps from Task 2.
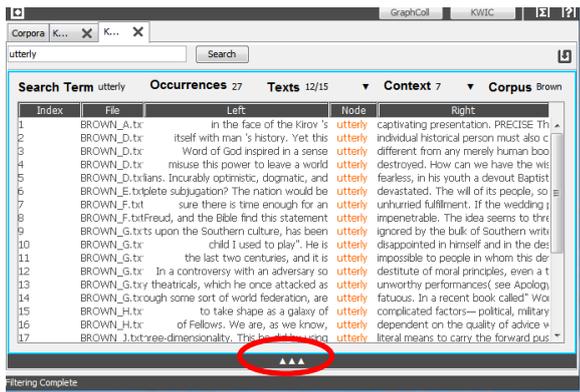
2. Search for the target word/phrase of your own choice using the KWIC function. Then click on the three arrows at the bottom of the window to split the window and open the second panel.
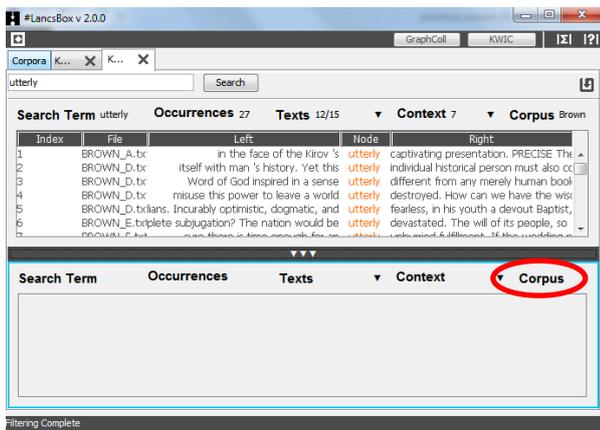
3. In the bottom panel select LOB or Brown as your corpus by clicking on the down arrow (▼) next to 'Corpus'.
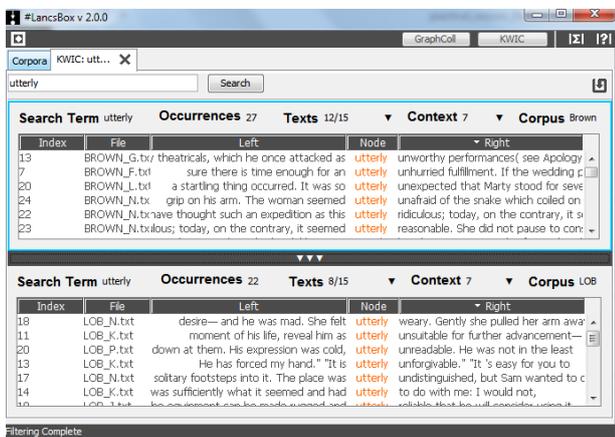
4. Click on the bottom panel to select it; light blue border will indicate the active panel. Type your word into the Search box and search the second corpus. Compare the use of the word/phrase in the two corpora. Use the left/right sort to identify typical contexts of use.

Are there any differences between your corpus and the general corpus?

T **Task 6. Producing graphs.** Start the GraphColl tool in #LancsBox. Use the LOB corpus to produce collocation graphs that show typical association in language. Follow the instructions below.

Build a collocation graph around the word *time,* which is the most frequent noun in English. First, select 03-MI as your statistical measure.

Then change the default settings under 'Threshold' as indicated in the figure below (MI = 5 and above) and search for the word *time*.
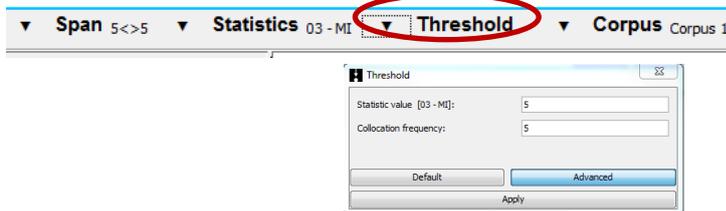

How many collocates does the graph display? Which are the most interesting collocates?

Some of the collocates of *time* such as *t, kungo* might not be completely transparent. Use the right-click function to obtain concordances (KWIC pop-up) and explain these collocates.

t is used as …….

kungo is used as ……

T   **Task 7. Producing network.** Go to the graph you have created in Task 6. It should be similar to the graph displayed in the figure below:
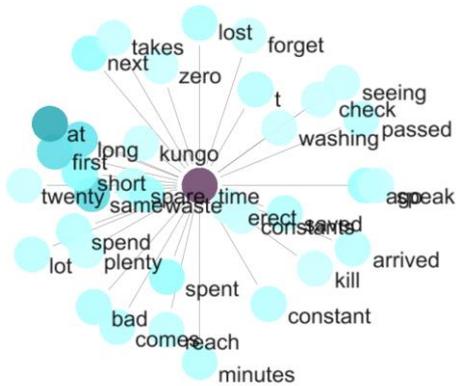
a) Find the collocate *spend* in the graph and double click on it. You should get a collocation network similar to the one displayed below:
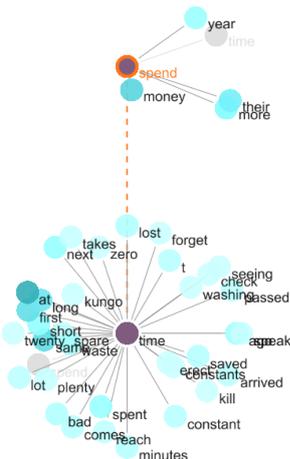
b) Find the second-order collocate *money* in the graph and double click on it. Comment on the connection between *time* and *money* that you can see in the resulting graph that shows collocates around the node *money*.

Tips for using #LancsBox with students

Corpus 1	Corpus 2
Tabloid newspaper	Broadsheet
British newspaper	American newspaper
Coverage of an event from one source	Coverage of an event from another source
Articles from a current newspaper	Articles from a newspaper from the 80s
Blogs from younger users	Blogs from mature users
Academic essays from students from lower years	Academic essays from students in higher years
A current TV series (using fan subtitles)	A TV series from the 90s
…	…

Students can also use #LancsBox to study the language from one source – for example from a novel or a collection of speeches from one person (e.g. a politician, the Queen, a Youtuber or blogger).

Task 8. Do your own search with one of the corpora in #LancsBox. You can use, for example:

- BNC64
- LOB
- Brown
- VU Lancaster Corpus of student writing
- Shakespeare corpus

You can either search for **meaning connections between words** or identify **lexico-grammatical patterns**. You can compare this corpus to other corpora in the 'Corpora' folder.

Some tips for using #LancsBox

1. You can save the results by clicking on the save icon:
2. To find instruction for how to use further #LancsBox concordancing functionalities (e.g. filtering your searches, selecting a sample), go to the #LancsBox website (http://corpora.lancs.ac.uk/lancsbox/help.php) and look at the User Guide.
3. Please feel free to get in touch with us – you can email Dana Gablasova (d.gablasova@lancaster.ac.uk)