

NB This resource assumes you have some knowledge of how to use the corpus tool Sketch Engine. If you do not, you should complete an introductory resource first, or at least watch some online tutorials.

1. Log on to Sketch Engine at www.sketchengine.co.uk and create a trial account if you don't already have one. Once you have an account, you can explore your own collection of texts (corpus) or you can use any of the existing corpora.
2. Once you have logged in, scroll down to the English language corpora to the corpus called 'Brexit corpus'.

Before opening the corpus up, click on the information symbol 'i', which gives you information about the data used to make this corpus.

Question: Why do you think the maker of this corpus made this selection of texts, a mix of webpages, newspapers and tweets, to investigate Brexit?

3. Now click on the 'Brexit corpus' and make a word sketch of the noun 'expat' by selecting the word sketch function in the task menu on the left of the screen. Type in 'expat' in the lemma search box and click on 'show word sketch'. In the word sketch screen, look at the words that modify the noun 'expat'. The modifiers box lists the frequencies for each modifier.

Question: Which proper noun is the most commonly used to pre modify 'expat'?

4. Now make a word sketch for the noun 'migrant'. In this case, a different proper noun is used most frequently to pre modify 'migrant'. Comparing the most frequent proper nouns used to pre modify 'expat' and 'migrant' respectively, the data in this corpus (all from British sources) tends to suggest that 'expats' are from one particular country while 'migrants' come from a larger geographical area.

Question: What is the most commonly suggested origin of 'migrant' in the corpus?

5. The nouns 'expat' and 'migrant' are effectively synonymous - carrying the same denotation. However, the two words differ considerably in terms of their connotations.

Question: What connotations do you think of when you hear 'expat' and 'migrant'?

6. Using the word sketch function, you can test whether your connotations of 'expat' or 'migrant' are supported by the data in the corpus. Look up the frequencies (number per million words) for 'expat' and 'migrant'. While an 'expat', as you saw earlier, is considered to be a British citizen living outside the UK, the corpus mentions them rarely.

Question: What does it suggest about the debates around Brexit that 'migrants' (from the EU primarily, according to the corpus) are more frequently discussed than British 'expats'?

Exploring attitudes and ideologies in a Brexit corpus using Sketch Engine

7. In the word sketches of 'expat' and 'migrant' you can also see how these nouns are used as subject or object in the corpus with particular verb phrases. Look at 'expat' and the number one verb that has 'expat' as its subject. Now look at the number one verb for 'migrant' as subject.

'Expats' - number one verb =

'Migrants' - number one verb =

Question: What does this comparison suggest about attitudes towards expats and migrants?

8. In the activities above, you were given the words 'expat' and 'migrant' to investigate. Often, you will have a gut feeling as to what to look at in your data. If you're not sure, **Keywords** are a good start for getting a sense of the main themes of a text or corpus. Keywords are words that will occur more frequently in a focus corpus than an everyday English reference corpus.

- First make sure the 'Brexit corpus' is selected in Sketch Engine.
- In the blue menu on the left, click on the 'keywords' icon (or simply click on the white 'keywords' button, which opens a new menu. Sketch Engine will select the English Web 2015 as reference corpus to calculate the keywords in the Brexit corpus.
- Once you've done this you will get a screen which compares the number of hits for 'single-words' and 'multi-words' in the Brexit corpus compared to the English Web 2013 as reference corpus. (Multi-words are phrases rather than single words.)
 - a. What are the five most common single- and multi-words in the Brexit corpus?
 - b. What is the most common word class of the single-word keywords?
 - c. Why do you think this is? Think of the nature of the texts that make up this corpus and the topic of Brexit.
 - d. What are the most common phrase types for multi-words?
 - e. Are there any neologisms amongst the noun phrases? What are the contexts for these new phrases?



Teacher notes

2. A mix of different sources might be more representative of the debates people were having about Brexit than just a corpus made up of newspaper articles, which represents the particular media organisations' take on the topic.
3. 'British'
4. EU. Expats are 'British' or 'from the UK', while migrants are mostly 'EU migrants' and some are 'non EU migrants'. While it makes sense for British newspapers to refer to British citizens living abroad as an 'expat' (there is an underlying assumption that the readership will think of these Britons as still British, still belonging to the same nationality, despite being ex patriate), there is also the difference in connotations, which affects how the terms 'migrant' and 'immigrant' are perceived (more negatively than expat).
5. The popular stereotypical perception of expats tend to be white professionals who are living abroad (perhaps for a few years), while migrants/immigrants tend to be non white. The terms 'migrant' and 'immigrant' are often used as synonyms for refugee and asylum seeker. So white Americans or Australians living in the UK are often labelled as expats, while Indians or Chinese living in the UK are labelled as migrant/immigrant.

Since the debates around Brexit at the start of 2016, the terms 'migrant' and 'immigrant' are more commonly used for EU citizens living in the UK. Funnily enough, Facebook adverts on the pages of EU citizens living in the UK often invite EU citizens to join an 'Expatriate Group', such as 'The Dutch Expatriate Group' for Dutch people living in the UK. The term 'migrant' or 'immigrant' is avoided in most of these.

6. Frequency: 'expat' is 3.82 per million and 'migrant' is 228.76 per million.

The word 'migrant' features almost 60 times more frequently than 'expat'. This suggests that in the Brexit debates people were more concerned with EU citizens coming into the UK than with what might happen to UK citizens living in EU countries (estimated to be about one million). While the noun 'expat' suggests a sense of shared identity, the relative lack of discussion of British expats highlights that many were not thinking about their fellow country people's fates following the vote to leave the EU.

7. 'Expats' - number one verb = 'stop'
'Migrants' - number one verb = 'live'

There is a sense that migrants must be stopped, they are perceived as damaging or having a bad influence. While 'expats' are represented as merely 'living'.

8. There is a huge number of proper nouns: 'Corbyn', 'Remain' etc. Most of these are the names of the actual campaign groups/organisations involved and the names of the politicians involved. The fact that so many names are keywords suggests that to a large extent political debates are very much influenced by/are about the people and organisations involved rather than the issues. If debates were more about issues, we might expect to see more abstract nouns in the keyword list such as 'employment' or 'immigration'.

Exploring attitudes and ideologies in a Brexit corpus using Sketch Engine

9. a. and b. Remain appeared to be winning. The keywords associated with Remain appear higher up the list, and also have significantly higher 'keyness' scores than the Vote Leave campaign keywords. When looking just at the keyword list, it appears that there's not much between 'Remain' and 'Vote Leave'. Remain appears in fourth place and Vote Leave in fifth place. However, their respective 'keyness' scores tell a different story:

- Remain: 700.2
- Vote Leave: 587.6

This is a significant gap. 'Remain' has a significantly higher frequency in the corpus than 'Vote Leave'. So it is important to look at the numbers as well as at the ranking in the list!

- c. If the corpus data suggests more is said about Remain, it suggests more support for Remain than for Vote Leave.
- d. As the corpus is made up of newspaper articles alongside online texts as well as tweets on the topic of Brexit, the electronic data might have been produced by younger people, who are more likely to be using social media platforms such as Twitter to discuss political issues. In other words, the data in the corpus might not represent the UK population who were actually voting in the referendum and as a result the data missed the strong support for Vote Leave.

It is useful to remind students of the need to be careful about selecting texts to put in a corpus, because by including some texts and not others, there is a risk of skewing the data and therefore the results.