BNCweb - a custom-made interface to the BNC
Ylva Berglund Prytz & Martin Wynne
Version: January 2014

BNCweb is an advanced and flexible online interface to the British National Corpus (BNC). The BNC is a snapshot of the English language in Britain in the late 20th century, with examples of speech and writing of a wide variety of types. More information about the BNC can be found at [http://www.natcorp.ox.ac.uk/](http://www.natcorp.ox.ac.uk/). This document contains some exercises to help you get started to use BNCweb to explore the usage of words in English.

**Task 1. Who uses the word ‘borrow’ and in what kind of constructions and contexts?**

**1.1 Getting started**

What does your intuition say about the use of 'borrow'? We're going to use the corpus evidence to explore how the word is used. But before logging in, think about the following questions, and write down some of your expectations. We can then compare your intuition to what we find in the corpus.

**What is borrowed and by whom?**

What words do you expect to find together with *borrow*?

Can these words be grouped in some way, for example based on their word class, function, or meaning?

Where would you expect these words (e.g. before or after *borrow*? Immediately adjacent or not?)

Who do you think uses the work *borrow*? In what context or type of language would you find *borrow*?

Are there any words that are NOT used with *borrow*?

Once you have thought about these questions, and maybe discussed them with others, start to follow the tasks below.

1. In a web browser, go to [http://ota.oerc.ox.ac.uk/](http://ota.oerc.ox.ac.uk/) and follow the link to the BNCweb service
2. Log in using your University of Oxford single-sign on credentials
3. In the Standard Query box, type in the word ‘borrow’.
4. Click the ‘Start Query’ button to run the search (see the image below)

5. Explore the concordance lines which are now displayed on your screen. Here are some tips on things that you can try.
   - Take a look at the summary information at the top of the browser page, and try to answer the following questions:
     - How many occurrences of ‘borrow’ are there in the corpus?
     - What is the relative frequency (hits per million words)?
     - What other information do you get at first glance?
   - Can you get a feel for how ‘borrow’ is used? What words do you find before or after ‘borrow’? (Click on the high-lighted ‘borrow’ to see more context)
   - Is the word used in particular types of texts or contexts? (Hover over the 'Filename' to see information about the text).
   - Click on the 'Show KWIC view' button to get an alternative visualization of the fragments of text. This might help to spot patterns of words around 'borrow'.
   - Try the 'Show in random order' button as well. Does this help?
   - Place your mouse cursor over the search work on one concordance line. What information do you get? (This works on JavaScript enabled browsers only.)

So far it is difficult to get an overview of the results, since you have only looked at one page at a time. Let's take a look at some ways of re-ordering the results to it easier to spot patterns of words, and some ways of extracting some information about frequencies and distributions.

### 1.2 Sorting

The BNCweb interface allows sorting on one position, but has some other sorting options which you do not see in all tools.

Above the concordance lines is a box which says [New Query]. Click in it and select [Sort] from the list that pops up. Click [Go!]

- Explore the sort options.
- What does the ‘Position’ option do?
• Why would you use ‘Starting with letter’?
• How many of your examples are immediately preceded by a verb? (*Hint: Select ‘1 Left’ and Tag restrictions ‘any verb’)*

1.3 Looking at the distribution of words

Is *borrow* a word more typically used in writing or speech? What sort of people use it? Does it always mean the same? Think first about what does your intuitions say, then try the following:

1. Start from the display of the *borrow* concordances (if you do not have the concordance display on the screen, use the back button to go back OR make a new search).
2. In the navigation bar above the concordance lines, click on [New Query] and select [Distribution]. Click [Go!]
3. Explore the display. What does it tell you about the use of ‘borrow’?

For example:
• Look at ‘Spoken or Written’. What are the most important and significant figures here which would allow you to make a statement about the distribution of ‘borrow’ in speech and writing?
• Look at ‘Derived text type’ which should give a better breakdown of types of speech and writing events – in which category is the work most frequent?
• At the top of the page, select ‘Speaker: Age’ in the first box. Click ‘Show distribution’.

Note that this category only applies to the spoken texts, as you might expect. Would you say that the usage of ‘borrow’ is strongly age-related?

• What about differences between men and women (choose ‘Speaker: Sex’ in the top box)?
• You can also cross-tabulate two categories, for example age and sex. Explore the results. Does any group use the work more than others? What happens if you look at social class?

Further hints and tips on distributions

• When you want to make another search, click ‘Go’ next to ‘New Query’.
• The ‘File-frequency extremes’ function will identify the file which has the highest proportion of a search term. Where is ‘borrow’ used most? ‘Speaker-frequency extremes’ identifies the speaker who uses a word/phrase most
1.4 Collocations
The BNCweb collocation feature allows you to explore co-occurrence patterns with words or word classes.

- Make sure your display has all your concordances (if not, go back OR sort on [node] with [no restrictions] OR run the search again)
- In the navigation bar above the concordance lines, click on [New Query] and select [Collocations]. Click [Go!]
  Accept the default settings by simply clicking [Submit] (You can explore the further options at your leisure). The collocations page is displayed.

1. Look at the collocation parameters panel (see above).
   Set [Collocation window span] to 4 Left - 4 Right
   Click [Go!]
   What are the top five results? What can you say from that?
2. Set the [Statistics] to [Rank by frequency] and click [Go!]
   In what way are your results different now?
3. In the list of collocates below the parameters box, click on one collocate ([Word] column) – what does the display show?
4. Click on a number in [Observed collocate frequency] column. What does that show?

Exploring collocations further
Try to find:
- What is the most frequent word following immediately after ‘borrow’?
- What is the most frequent noun following immediately after ‘borrow’?
- What is the strongest collocate (log-likelihood) following immediately after ‘borrow’?

1.5 More search features
Using the search box as above, you can also search for lemma, pos and use wild-cards and boolean operators.

POS
Find words with a certain tag: WORD_TAG (plant_NN1’ finds all instances of plant tagged as singular noun
Find a tag: _TAG ( _NN1 finds all singular nouns)
Use wild-cards to replace part of a tag ( _N* finds all nouns)

Lemma
{be} finds all forms of the lemma BE: am, is, are, being etc
{will/N} finds nominal instances of the lemma WILL: will, wills

Wildcards
super* finds all words beginning with super-
*ing_AJ0 finds all words ending in –ing which are tagged as adjectives
Search in part of corpus
Use the [Written restriction] or [Spoken restrictions] options in the left toolbar on the start page to search only part of the corpus (such as texts from newspapers or material spoken by women)

Always look at your results in light of what you have searched - how much of the corpus? How has this sub-set been created?

You can restrict your search to material of certain type. Click on ‘Spoken restriction’ on the start page, left-hand menu and explore the options.

Ticking more than one option in a column (for example Interaction type: monologue AND dialogue) expands your search and give you more hits. Ticking something in more than one column (for example ‘monologue’ AND ‘Region: Midlands’) restricts it – you will get fewer hits.

Include annotation in search
You can use the very powerful CQP syntax to make more complex searches, including ones that takes annotation into account. This would enable you to make use of things like pauses, shift in voice quality, etc., which have been annotated in the corpus.

Type the following into the search box and set Query Mode to ‘CQP syntax’:

"to" [pos = "V.*I" & flags_before contains "pause"]

Can you see what the search does?

For more CQP syntax tips, click on ‘Simple Query Syntax help’
2. Further Information

More about BNCweb:


Another way to access BNCweb

Free access is also available to the service at Lancaster University, with limited functionality -maximum 5,000 hits, no browsing of full texts. Users need to register. See http://www.bnc.info and http://bncweb.lancs.ac.uk/bncwebSignup/.

### Lancaster BNCweb Server Usernames

Welcome to the BNCweb username administration system.

<table>
<thead>
<tr>
<th>First time users:</th>
<th>Register for an Account</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Make changes to an existing account:</th>
<th>Username: [ ]</th>
<th>Password: [ ]</th>
<th>Login</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>No access?</th>
<th>Resend Password</th>
<th>Resend Activation Code</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Registered users:</th>
<th>Login to BNCweb</th>
</tr>
</thead>
</table>

Powered by Deadlock User Management System