**BNCweb - a custom-made interface to the BNC**

BNCweb - an advanced and flexible interface to the BNC. Free access (registration required) to Lancaster service. See http://www.bnc.info and http://bncweb.lancs.ac.uk/bncwebSignup/.

0. Register for an account on BNCweb (http://bncweb.lancs.ac.uk/bncwebSignup/)

<table>
<thead>
<tr>
<th>Lancaster BNCweb Server Usernames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welcome to the BNCweb username administration system.</td>
</tr>
<tr>
<td>First time users: Register for an Account</td>
</tr>
<tr>
<td>Make changes to an existing account: Username: Enter Here Password: Enter Here Login</td>
</tr>
<tr>
<td>No access? Respond Password Resend Activation Code</td>
</tr>
<tr>
<td>Registered users: Login to BNCweb</td>
</tr>
</tbody>
</table>

- Or try the experimental (beta) service in Oxford at http://ota.oerc.ox.ac.uk/bncweb-cgi/BNCweb.pl. Access is via your Oxford single sign-on (WebAuth) credentials.

**Task 1. A bloody search**

**Who uses the word ‘bloody’ and in what kind of constructions and contexts?**

What does your intuition say? Compare to search results from the BNC.

1. Log in to BNCweb
2. In the Standard Query box, type in the word ‘bloody’.
3. From the ‘Restriction’ drop-down list, choose ‘Spoken Texts’
4. Click the ‘Start Query’ button to run the search.
5. Explore the concordance lines
Can you get a feel for how the word is used? (‘hover’ over the Filename to see meta-data about text)
Note that [unclear] is marked. Any other annotation visible?

Further explorations: distribution.
1. In the ‘New query’ list, select ‘Distribution’ and click on ‘Go!’.
   Explore the list – what does it tell you?

   Look at ‘Domain’ – in which category is the work most frequent? (Why?)

3. At the top of the page, select ‘Speaker: Age’ in the first box. Click ‘Show distribution’.
   Would you say ‘bloody’ is an age-related word?

4. What about gender (choose ‘Speaker: Sex’ in the top box)?
5. 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?

When you want to make another search, click ‘Go’ next to ‘New Query’

**Define spoken restrictions**

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 (pauses, shift in voice quality, etc).

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’

Task 2. Wicked or what?
What does wicked mean, who uses it, and what does it collocate with?

Explore the interface by looking at wicked (see below for help to get started)

Make a search

Making a search in BNCweb is simple. There are many advanced search features that you can explore by and by if/when you need them.

1. Type wicked in the search box and press [Start Query]. A list of hits appears on the screen.

Explore results

What can you find out from the result display and how can you control what you see?

2. Look at the result screen and try to answer (Hint: look at the very top of the page):
   How many ‘wicked’ are there in the corpus?
   What is the relative frequency (hits per million words)?
   What other information do you get at first glance?

3. Place the mouse cursor over one of the codes in the Filename column (or click on it)
   What do you see?

4. Place your mouse cursor over the search work on one concordance line.
What information do you get? (JavaScript enabled browsers only)

5. Click on [Show KWIC view]
   What does this do?

Question: When would you use the [Show in random order] button?

**Sorting**

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

6. 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!]

7. 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’)

**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)

8. 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] (Explore options at your leisure). *The collocations page is displayed.*

9. 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?

10. Set the [Statistics] to [Rank by frequency] and click [Go!]
    In what way are your results different now?

11. In the list of collocates below the parameters box, click on one collocate ([Word] column) – what does the display show?
12. 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 ‘wicked’?
- What is the most frequent noun following immediately after ‘wicked’?
- What is the strongest collocate (log-likelihood) following immediately after ‘wicked’?

Looking at the distribution

Is wicked a written or spoken word? Who uses it? Does it always mean the same? What does your intuition say?

Start from the display of the wicked concordances (if you do not have the concordance display on the screen, use the back button to go back OR make a new search).

13. In the navigation bar above the concordance lines, click on [New Query] and select [Distribution]. Click [Go!]  
Explore the display.
- Can you tell whether wicked is a spoken or written word by looking at the statistics?

(Exploring distribution further Who uses ‘wicked’?)

Go to the Distribution display (see 13 above)

- In [Categories], select ‘Speaker sex’
- In [Categories (for crosstabs only)] select ‘Speaker Age’.
- Click ‘Show distribution’.
- Explore the display. What information do you see? What does it mean?

NB – Please pay attention to the ‘total’ rows. You are looking at hits for which we know the speaker. Are these all instances? Does that matter?

The ‘File-frequency extremes’ function will identify the file which has the highest proportion of a search term. Where is ‘wicked’ used most?

(‘Speaker-frequency extremes’ identifies the speaker who uses a word/phrase most)

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)

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

All about BNCweb: