

Resources

Freq	Rel [%]
1682	131.4
1301	115.6
881	89.7
684	58.9
Freq	Rel [%]
2301	99.6
2071	100.5

The Centre for Corpus Research at Birmingham has a wide range of corpus resources and tools for research purposes. This page provides information on these resources.

Corpus interfaces

Bank of English

For access, you require a Bank of English user's user name and password. The corpus is accessed by Telnet connection. Users of computers running Unix-based operating systems (e.g. Mac or Linux) can simply open a 'Terminal' window and type in: telnet titania.bham.ac.uk. You will then be prompted to enter your user name and password. For Windows users: if Telnet is installed, go to 'Run' in the 'Start' menu and type in: telnet titania.bham.ac.uk. You will then be prompted to enter your user name and password. If you do not have Telnet, you can use a freeware programme called '**PuTTY** (<http://www.chiark.greenend.org.uk/~sgtatham/putty/>)'; you can create a 'Telnet' connection using this programme.

CQPWeb

We have now placed the Bank of English on CQPWeb, on a Birmingham server. This allows access to the corpus and (in course of time) other corpora through the CQPWeb interface. It requires a bham.ac.uk email address for you to be able to register. Registration is at <http://www.cqpweb.bham.ac.uk/usr/?thisQ=create&uT=y> (<http://www.cqpweb.bham.ac.uk/usr/?thisQ=create&uT=y>).

Thereafter you can use the portal at <http://www.cqpweb.bham.ac.uk/> (<http://www.cqpweb.bham.ac.uk/>)

BNCWeb

A web-based client program for searching and retrieving lexical, grammatical and textual data from the British National Corpus (BNC). It requires a bham.ac.uk email address for you to be able to register. Registration is at <http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php>, and thereafter you can use the portal at <http://bncweb.lancs.ac.uk> (<http://bncweb.lancs.ac.uk>). This gives access to the British National Corpus.

Sketch Engine

Institutional access to the **Sketch Engine** (<http://the.sketchengine.co.uk/>) interface is available on any computer on the University network (but not outside the network). Approximately 160 corpora are included.

Wordbanks Online

Institutional access to the **Wordbanks** (<http://wordbanks.harpercollins.co.uk/auth>) Online service is available on any computer on the University network (but not outside the network). Wordbanks Online is the HarperCollins interface for the 550 million word version of the Bank of English.

Wordsmith Tools

Birmingham University users can use the networked version of the programme from any machine on the university network, provided that the user has logged into the network (there is a **PDF worksheet** (http://paulslals.org.uk/ccr/wst/wsmith_worksheet.pdf), which explains how to access the programme and the basics of using Wordsmith Tools).

CLAWS Part of Speech Tagger

The Centre has a licence for the CLAWS tagger (UCREL, Lancaster). Staff or students who are interested in POS-tagging large quantities of data for research should contact Paul Thompson.

WMatrix

The Centre also has a licence for the **WMatrix** (<http://ucrel.lancs.ac.uk/wmatrix/>), suite of semantic and POS annotation and analysis tools developed by Paul Rayson (UCREL, Lancaster). Staff or students who are interested in using WMatrix for research should contact Paul Thompson.

The Centre has a large number of corpora for use by researchers and students at the university. These include:

- AHRC corpora
- Australian Corpus of English
- British Academic Spoken English corpus
- British Academic Written English corpus
- British National Corpus
- Bank of English
- Brown Corpus
- COCA
- COLT

- DANTE
- Europarl
- Freiburg-LOB (FLOB)
- Freiburg-Brown (Frown)
- German Parole
- Global Web-Based English (GloWbE).
- The Helsinki Corpus of English Texts: Diachronic Part
- The Helsinki Corpus of Older Scots
- ICLE
- Innsbruck Computer-Archive of Machine-Readable English Texts (ICAMET)
- The International Corpus of English - East African component
- Italian Corpus
- Italian Newspaper
- Kolhapur Corpus (India)
- Lampeter
- Lancaster/IBM Spoken English Corpus (SEC)
- LOB Corpus
- London Lund Corpus
- Micase corpus
- Multilingual Plato
- Newdigate Newsletters
- Polytechnic of Wales Corpus
- Wellington New Zealand Spoken
- Wellington New Zealand Written
- Wolverhampton Business English

Access to some of these corpora is restricted; for further information, contact Paul Thompson at ccr@contacts.bham.ac.uk (<mailto:ccr@contacts.bham.ac.uk>)

[Privacy](#) | [Legal](#) | [Cookies and cookie policy](#) | [Accessibility](#) | [Site map](#) | [Website feedback](#) | [Charitable information](#)

© University of Birmingham 2015

