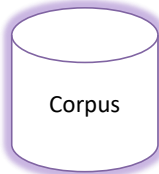
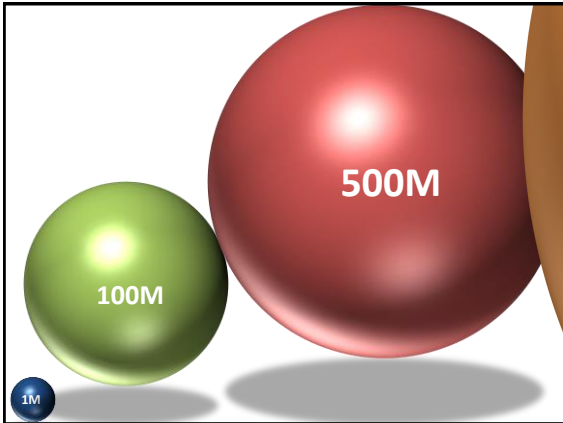





BRITISH NATIONAL CORPUS 2014

Design and compilation
Dr. Vaclav Brezina

Corpus as a sample

BNC 1994: 100M
~3,000 publications



Current British English?

- + We would like to hold on to the **VHS** copies for a further 3 weeks.
- + the Internet could run out of **available addresses by 1995**. [BNC2014-spoken]
- + Tonight's gonna be excellent! I can't wait! **Wicked!**

1994 ————— 2014



BRITISH NATIONAL CORPUS




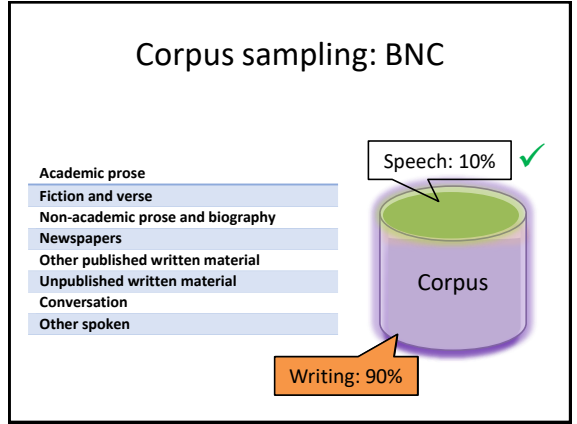
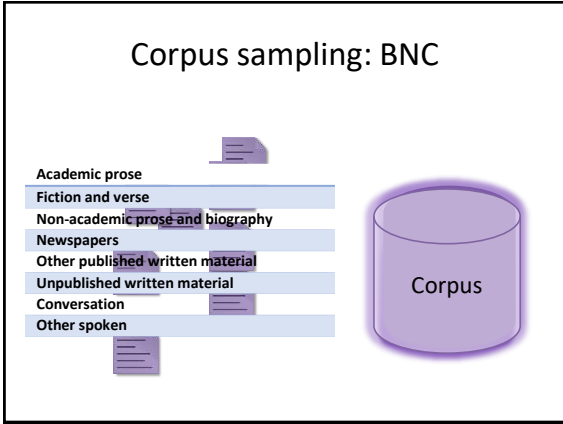
BRITISH NATIONAL CORPUS 2014

CASS

Corpus as a sample

Representative?
Unbiased?





CHALLENGES	OPORTUNITIES
<ul style="list-style-type: none"> To capture regional and social variation. To capture register variation. To maintain balance. To represent 'British English'. To capture the change in language and society. To deal with copyright. 	<ul style="list-style-type: none"> To update the BNC. To understand current British English. To rethink sampling issues. To offer statistical sophistication. To provide a free resource for researchers.

Ginny 5573 Mm...
Keith 5574 and even an English A...
Ginny 5575 She should do... what David does?
Keith 5576 Mm...

Would you like to get involved?

The Written BNC 2014

- Web crawling
- Data from publishers
- Data from authors
- PPSR: public participation in scientific research

cass.lancs.ac.uk

CASS
 ESRC Centre for Corpus Approaches to Social Science (CASS)

Would you like to contribute to the Written BNC2014?

- [Book collection](#) [Already submitted] [British authors]
- [Student essay collection](#) (school-level or university-level essays)
- [Email collection](#) (anonymised)
- [SMS collection](#) (anonymised)
- [Tweet collection](#) (anonymised)

**BNC 2014 Book Collection:
Submission Form**

Dear contributor,

Please do one form per book to upload your scanned files.
Thank you very much for your contribution.

BNC2014 team

The name and photo associated with your Google account will be recorded when you upload files and submit this form. You will be contacted if you need to provide more information. will@bnc2014@gmail.com [Google account](#)

***Required:**

BNC
**BRITISH
NATIONAL
CORPUS
2014**

BNC 2014 Book Collection

Title of the book: *

Your answer:

Corpus and copyright

- Corpus data is collected for research (non-commercial) purposes.
- Only samples of data are taken.
- Following changes to the law in June 2014 (CDPA s. 29A), it is now permissible to copy works to which one has legal access (e.g. via the Library) to enable data and text mining for non-commercial research. This covers creating a corpus of language materials.