We would like to hold on to the VHS copies for a further 3 weeks.

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!
Corpus sampling: BNC

- Academic prose
- Fiction and verse
- Non-academic prose and biography
- Newspapers
- Other published written material
- Unpublished written material
- Conversation
- Other spoken

Speech: 10%

Writing: 90%

CHALLENGES

- 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.

OPPORTUNITIES

- To update the BNC.
- To understand current British English.
- To rethink sampling issues.
- To offer statistical sophistication.
- To provide a free resource for researchers.

Would you like to get involved?

The Written BNC 2014

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

Would you like to contribute to the Written BNC2014?

- Book collection (already submitted [British authors])
- Student essay collection (undergraduate or university-level essays)
- Email collection (anonymised)
- SMS collection (anonymised)
- Tweet collection (anonymised)
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.