Using concordances in the classroom: a hands-on workshop

Jenny Kemp & Glynis Scaramuzza
In this session...

- What we do with advanced-level students in a concordance workshop
- Opportunity for practice
Learner Training

Ss already introduced to printed concordance output in classes.

Awareness of concordancers as a self-study tool

Observe - Highlight – Extract - Use Methodology

Patterns of use: must be 2 or more
Your worksheet

- **Student questions**
- **Guided error correction practice**
- **Preparing classes**
- **Exploring other web tools**

**P A T T E R N S  o f  u s e:** must be 2 or more

Link Extractor As discussed here
BYU-BNC: BRITISH NATIONAL CORPUS

100 MILLION WORDS, 1980s-1993

DISPLAY
- LIST  ○ CHART  ○ KWIC  ○ COMPARE

SEARCH STRING
- WORD(S)
- COLLOCATES
- POS LIST
- RANDOM  SEARCH  RESET

SECTIONS  SHOW

INTRODUCTION

This website allows you to quickly and easily search the 100 million word British National Corpus (1970s-1993). The BNC was originally created by Oxford University Press in the 1980s - early 1990s, and now exists in various versions on the web. Note that our version of the BNC was recently updated, and it now uses the CLAWS 7 tagset.

If you find this version of the BNC useful, you may also be interested in other corpora that have been created by Mark Davies of Brigham Young University, including the 450 million word Corpus of Contemporary American English (1990-2010) and the 400+ million word Corpus of Historical American English (1810-2009).

As with other BNC interfaces, you can search for words and phrases by exact word or phrase, wildcard or part of speech, or combinations of these. You can also search for surrounding words (collocates) within a ten-word window (e.g. all nouns somewhere near paper, all adjectives near woman, or all nouns near spin).

With this architecture and interface, you can also easily find the frequency of words and phrases in any combination of registers that you define (spoken, academic, poetry, medical, etc). In addition, you can compare between registers -- for example, verbs that are more common in legal or medical texts, or nouns near break that are more common in fiction than in academic writing.

Download 440 million words of full-text data for COCA (190,000 texts), or 1.8 billion words for GloWbE (1,800,000 texts). With this data, you will have the texts from the corpora on your own computer, rather than having to use the web interface. The data comes in three formats: tables for relational databases, word/lemma/Pos (vertical format), or text (linear format).

Word and Phrase (analyze texts) Enter entire texts and see detailed frequency information on the words in the text, and create word lists based on your text. Click through the words to see detailed information on any word. Highlight phrases in your text and have it search for related phrases in COCA.

Word and Phrase (frequency lists) Search and browse the most complete frequency dictionary of English. See detailed information (all on one page) -- definition, frequency by genre, collocates (nearby words), concordance lines, synonyms, and WordNet-related words, all with useful links from one resource to another.

Word Frequency You can also download lists showing the frequency of the top 60,000 lemmas by genre (and sub-genre), as well as the top 200-300 collocates (nearby words) for these lemmas (4,800,000 node/collocate pairs). There is also a free list of the top 5,000 lemmas in COCA. And now you can download the 100,000 integrated word list from COCA, COHA, BNC, and SOAP -- the largest, corrected frequency list of English.