# Corpus-Assisted Discourse Studies (CADS)

## Worksheet for video 2

Full resource, see: <https://www.ncrm.ac.uk/resources/online/all/?id=20855>

In Video 2, you’ll look at some examples of the ways you can analyse a dataset using corpus linguistics. The dataset in this example is a corpus of student writing. This is divided into writing by students with Chinese as a first language (Mandarin, Cantonese or other) and students with English as a first language. It can also be divided by discipline of study.

**Aims:** To uncover similarities and differences in student writing across disciplinary areas and by students with different first languages

**RQs**

1. How does student writing differ in Engineering and Business Studies?
2. How does student writing differ across students with first language Chinese or first language English?

TASK 1: How might you go about answering these research questions? If you’re familiar with corpus linguistics, you could consider techniques within this methodology. If you’re more familiar with discourse analysis, think about how this might help you.

TASK 2:

In Video 2, you’re asked to respond to the following question. Consider your answer now, and make a list of the words you think of (hint: think about short, functional words).

What are the top 10 words in written English, taken from web sources? (eg blog postings, internet sites).

National Centre for Research Methods (NCRM)  
Social Sciences  
Murray Building (Bldg 58)  
University of Southampton  
Southampton SO17 1BJ  
United Kingdom

**Web** www.ncrm.ac.uk   
**Email** info@ncrm.ac.uk  
**Tel** +44 23 8059 4539  
**Twitter** @NCRMUK