Corpus Linguistics

This series of six lunch-time sessions offers an introduction to the use of the corpus in linguistic research. You will learn about tools, resources and methods by examining practical examples from the corpus linguistic literature. This document and all presentations and handouts are available from WebLearn at http://tinyurl.com/669o4zt

**Taught by** Martin Wynne (martin.wynne@oucs.ox.ac.uk) and Ylva Berglund Prytz (ylva.berglund@oucs.ox.ac.uk), IT Services and Faculty of Linguistics.

**Intended audience:** Anybody interested in evidence-based language study.

**Prior knowledge:** Knowledge of basic concepts and research methods in linguistics or a related discipline.

**Resources:** Brief notes are supplied at no charge

**Software:** Various, focus on web-based resources where available.

**Format:** Presentation with practical exercises

**Where next?** Other modules in the series

**Session outline**

1. **Corpus Linguistics: Introduction**  
   In this first session we present some key concepts and techniques and you’ll be able to explore some freely available online resources.  
   **Key topics:**  
   corpus linguistics  
   corpus  
   concordance

2. **Corpus Linguistics: The Tools of the Trade**  
   This week we explore various software applications for displaying, analysing, comparing and otherwise extracting linguistic information from corpora.  
   **Key topics:**  
   concordances  
   online tools  
   downloadable tools

3. **Corpus Linguistics: Collocation**  
   Corpus linguists claim to have identified an important principle is responsible for the creation of much of the meaning of texts – collocation (co-occurrences). What is it, and are the claims true?  
   **Key topics:**  
   collocation  
   colligation  
   meaning
4. **Corpus Linguistics: Annotation**  
   In this session we will be looking at key methods and problems of applying linguistic annotation to a corpus, and how to use the annotated corpus.  
   **Key topics:**  
   annotation  
   tagging/tags/tag-set  
   word class

5. **Corpus Linguistics: Creating a Corpus**  
   In this session we take a look at corpus creation and discuss of core concepts such as representativity, sampling, annotation, and text encoding.  
   **Key topics:**  
   representativity  
   sampling  
   annotation  
   text encoding

6. **Corpus Linguistics: Using the corpus in linguistic research**  
   In this session we take a more in-depth investigation of a specific linguistic research topic, with a critical look at corpus linguistic resources and methods used in a published study.  
   **Key topics:**  
   research  
   corpus choice  
   methods  
   annotation