# Lecture 31: Notes

**Using corpus linguistics in research**

**Overview**

This lecture gives the essential information about how corpora have been constructed. It gives a rationale behind corpus linguistics and discusses pros and cons of this methodology. It introduces some central concepts in corpus linguistics: frequency, representativeness etc.

**Review points:**

1. **What is a corpus?**

a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis.

 ...the language stored in a corpus is naturally-occurring,i.e. gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology. […]

1. **What is corpus linguistics?**

 Corpus linguistics is not directly about the study of any particular aspect of language. Rather, it is an area which focuses upon a set of procedures, or methods, for studying language.

1. **Criteria for making a corpus**
* Authenticity
* Size
* Sampling
* Representativeness
* Balance

Balance, representativeness and comparability are ideals which corpus builders strive for but rarely, if ever, attain

1. **Why to use corpora or corpus research?**
* Inform the syllabus
* Create materials/activities
* Create specialized corpora
* Use online resources
1. **How to use corpus research in the classroom?**
2. Teachers can use corpora to create a variety of materials.
* Word frequency lists from readings
* Models of language use and role play actual dialogues
* Practice activities and testing
1. How to access some corpora resources?
2. **Uses of specialized corpora**
* Identify unfamiliar words
* Identify high frequency words
* Use KWICS to generate class activities
* Identify word senses
* Practice inferencing strategies
1. **How can teachers use corpora to create materials and activities**
* Word frequency lists from readings
* Models of language use and role play actual dialogues
* Practice activities and testing
* Specialized corpora
* Use online corpora
1. **How can learners interact with corpora?**
* Learn vocabulary
* Explore extended collocations
* Compare against model texts
* Discover patterns of use
1. **What are challenges in using corpora?**
* Availability of corpora
* Studies that explore the effectiveness of corpus-based materials for teaching
* Time and resources
1. It has been argued that a significant advantage of using corpora is that corpora allow analysts to approach the study of language within the context of the scientific method.
2. If you approach a corpus with a specific theory in mind, it can be easy to unintentionally focus on and pull out only the examples from the corpus that supports the theory (this is technically called a *confirmation bias*). But the theory can never be shown to be false by such an approach, even in principle.