

Course: Corpus approaches to discourse analysis

Professor: Alison Duguid (alison.duguid@unisi.it;

Student hours: Monday, Tuesday, Wednesday. Palazzo S. Niccolò

Start date: Monday 5th October

Language and Mind: Linguistics and Cognitive Studies

Curriculum Linguistics and Cognition

Second Year

SYLLABUS Corpus approaches to discourse analysis: Text and corpus analysis in English Studies

Course rationale

Corpus literacy (the knowledge of making use of corpora and software for corpus analysis) **is a key literacy in English studies** the use of corpora and corpus-analytic techniques helps to strengthen students' language awareness, in particular their discourse awareness enabling them to distinguish between different types of texts or discourses. This course aims to develop students' corpus literacy and to explore the effects of register, domain and text type differences through corpus study and to develop awareness and insight into the use and manipulation of language in society in a host of areas.

Objectives

- The aims of this course are to give students the awareness, knowledge, experience and skills of analysis of naturally occurring language texts through a corpus assisted discourse studies approach.

More specifically by the end of the course you will have gained:

Knowledge and understanding: of corpus linguistics and the different kinds of corpora, of corpus approaches to discourse studies, criteria for corpus compilation

The nature of frequency data, the concept of keyness, collocation and collocates,

Application of knowledge and understanding

You will be able to apply your knowledge to: the applications of corpus data to research in linguistics and the social sciences, devising appropriate research questions, compiling a corpus, storing and labelling of texts, and analysing with the help of software programmes, analysis of wordlists, keyword analysis, reading concordance lines, analysis of collocation data

Critical analysis: will be developed through the analysis exercises in class, the examples of previous research and the corpus project

Communication skills: your competence in English will be improved by exposure through the readings, the corpus compilation, the lectures and use in participation in class and the presentation

of the corpus project and by the attention paid to language analysis and the use of reference materials

NB The awareness and skills are built up gradually through the experiences introduced by interactive activities and group work throughout the course. It is possible to do the exam without having attended but those who do might have to spend much more time becoming familiar with the techniques and dealing with the pitfalls of applying knowledge and comprehension acquired through reading only.

There is a very good **free on-line course** which guides you through procedures and provides feedback and a forum as well as many many resources. Those who are unable to attend the course will find it a useful substitute and even attenders will find the resources a great help. It starts on the 28th of September but you can enrol online from now.

Corpus Linguistics, Method, Analysis, Interpretation

<https://www.futurelearn.com/courses/corpus-linguistics>

Programme

- This course takes a functional view of language whereby language is seen as having a social, interactive function (establishing human relationships and negotiating communicative goals) as well as an informational/communicative function.

Topics covered

Core concepts, constructs, techniques in Corpus Linguistics

Text and Corpus

definitions of corpora, purposes and applications Corpora in English studies: Types of corpora:

parameters Corpus specification

background to corpus linguistics Why use corpora? Intuition and introspection vs. corpus evidence

What corpus linguistics can do: applications of corpora in research

Texts and language system: Communicative competence and Generic competence

Keyness and keyword analysis

frequency wordlists and keyword lists

the interpretation of frequency data

Collocations: the systematic co-occurrence of words in use.

Collocations and meaning: Semantic Preference

Collocations and attitude: Semantic/evaluative Prosody

Concordances and concordancing

reading concordances

patterns: frequencies co-occurrence

words and word groups grammatical units meanings attitudes

context and co-text

Using corpus linguistics in language research

Principle of accountability

Text vs. Texts: Cumulative patterns The risks of confirmation bias

Qualitative / Quantitative

Corpus Linguistics and Discourse Analysis

Compiling a corpus

Corpus building

Research questions

CADS, CL and CDA

Examples of research

Final evaluation is carried out by means of an **individual corpus project**. Students decide on a language related research question then choose or compile a corpus (at least 100,000 words) to analyse in order to answer the question. Using appropriate software (e.g. WordSmith or AntConc) and making use of wordlists, keywords and keyness data, concordances and collocation data to obtain information about the discourse type exemplified by the data.

There is a very good **free on-line course** which guides you through procedures and provides feedback and a forum as well as many many resources. Those who are unable to attend the course will find it a useful substitute and even attenders will find the resources a great help. It starts on the 28th of September but you can enrol online from now.

Corpus Linguistics, Method, Analysis, Interpretation

<https://www.futurelearn.com/courses/corpus-linguistics>

Some resources

For some hands on practice you can find here a tutorial on the use of AntConc

<http://programminghistorian.org/lessons/corpus-analysis-with-antconc>

A useful bibliography

www.gabrielatos.com/CLDA-Biblio.htm

CADS: Useful websites which will illustrate the approach used

<http://www3.lingue.unibo.it/blog/clb/>

http://en.wikipedia.org/wiki/Corpus-assisted_discourse_studies

see also **Corpus linguistics on Facebook**

<https://www.facebook.com/groups/6642113001/?fref=nf>

To get an idea of the range of topics studied by Corpus linguistics you can look at the book of abstracts from the last Corpus Linguistics conference this summer, provided here. It shows the variety of interests covered in current research and is an excellent source of bibliographical references on those topics

For some hands on practice: A tutorial on the use of AntConc

<http://programminghistorian.org/lessons/corpus-analysis-with-antconc>

Corpus work is essentially comparative. Useful sources of comparison of word frequency can be found by looking at Word frequencies in Spoken and Written English (<http://ucrel.lancs.ac.uk/bncfreq/>) based on the British national corpus which was compiled in 1991.

Another source for American English is

<http://corpus.byu.edu/coca/>

You can also get frequency data from the **Longman Grammar of Spoken and Written English**. (Biber et al)

Google N-grams: <https://books.google.com/ngrams>

Google trends <https://www.google.com/trends/explore#cmpt=q&tz=>

Where you can see, for example, the rise of the term *selfie*.

Macmillan English Dictionary for Advanced Learners (MEDAL) has a system of stars where 3 stars indicate a very frequent word.

An interesting website

<http://www.macmillandictionary.com/learn/realgrammar.html>