

Data and corpora



Niveau d'étude
Master 1



ECTS
3 crédits



Volume horaire
18h



Période de
l'année
Semestre 1

Présentation

DESCRIPTION

Corpus Linguistics studies language using systematically organized collections of spoken and written texts, known as corpora. This course provides foundational knowledge of corpus linguistics, focusing on oral and written corpora. Key topics include the definition and evolution of corpus linguistics, types of corpora (written and oral), and principles of corpus design such as sampling, balance, and ethical considerations. Students will explore analytical tools for quantitative and qualitative studies of linguistic patterns (using large-scale corpora in well-documented languages – English, in particular) and learn about applications in linguistic description / data collection, corpus constitution & annotation, in field situation and in the lab. The course highlights differences between oral and written corpora, as well as challenges in analyzing spoken data, and data maintenance.

OBJECTIFS

On successful completion of this course, students should be able to:

- Constitute oral and written corpora
- Annotate, analyze and maintain them
- Use existing large-scale corpora tools to conduct quantitative studies of specific phenomena

HEURES D'ENSEIGNEMENT

Data and corpora	Cours Magistral	9h
Data and corpora	Travaux Dirigés	9h

PRÉ-REQUIS OBLIGATOIRES

None

Pour en savoir plus, rendez-vous sur > u-paris.fr/choisir-sa-formation