

# Atelier Corpus Large : outils et interprétations des données

## Infos pratiques

---

- > ECTS : 3,0
- > Nombre d'heures : 18,0
- > Langue(s) d'enseignement : Anglais
- > Période de l'année : Enseignement septième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Travaux dirigés
- > Ouvert aux étudiants en échange : Oui
- > Composante : Langues et cultures étrangères
- > Code ELP : 4V7AATDP

## Présentation

---

This practical workshop course aims to equip students with the necessary knowledge, tools and methodology to carry out linguistic research building and using a large written corpus, i.e. an electronic collection of texts designed to be representative of language structure and use. After a quick theoretical introduction to key concepts in corpus linguistics, students will develop skills in three areas:

1. Corpus construction: they will learn how to select and format relevant texts to be included in an English-language corpus. They will discover how to enhance the corpus through partially automated and manual annotation (i.e. systematic identification of grammatical, semantic, pragmatic, etc. features in corpus texts)
2. Corpus use: in order to investigate a specific question pertaining to English syntax, semantics and/or pragmatics, they will learn how to use the functionalities of a software tool to retrieve relevant corpus data and obtain useful statistical measures (e.g. statistical significance, word and pattern association measures, keywordness measures).
3. Interpretation of corpus data: they will learn how to interpret the data by combining large-scale

quantitative results with small-scale qualitative analysis in order to answer their research question.

## Objectifs

---

- \* Acquérir les concepts-clés de la linguistique de corpus
- \* Développer les savoirs-faire méthodologiques nécessaires à la construction et l'exploitation d'un grand corpus écrit
- \* Apprendre à mener à bien une recherche linguistique sur corpus écrit

## Évaluation

---

### Formule standard session 1 - Contrôle continu :

Travail collaboratif d'enrichissement/annotation de corpus et compte-rendu oral (30%)

Travail préparatoire écrit conduisant à la rédaction de l'étude (30%)

Etude de corpus écrite portant sur un sujet de recherche choisi par l'étudiant.e (40%)

### Formule dérogatoire session 1 :

Épreuve orale (proposition d'une problématique de recherche et d'une méthodologie adaptée à partir d'un extrait de corpus) d'une durée de 30 min (20 min de préparation + 10 min de passage)

### Session 2:

Epreuve orale sans temps de préparation

## Pré-requis nécessaires

---

Maîtrise des concepts et de la terminologie essentiels en grammaire anglaise, en sémantique et en pragmatique. Compétence basique en informatique. Note : Aucune compétence en programmation informatique requise.

## Bibliographie

---

Lindquist, H. & Levin, M. (2018). *Corpus Linguistics and the Description of English (2nd edition)*. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R. & Tono, Y. (2006) *Corpus-Based Language Studies*. London/New York: Routledge.

Wynne, M. (ed). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. Available online at: <http://ahds.ac.uk/linguistic-corpora/>

## Contact(s)

### > Sophie Raineri

Responsable pédagogique  
sraineri@parisnanterre.fr