

# Linguistique de corpus

## Infos pratiques

---

- > ECTS : 3.0
- > Nombre d'heures : 24.0
- > Langue(s) d'enseignement : Anglais
- > Période de l'année : Enseignement cinquième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Travaux dirigés
- > Composante : Langues et cultures étrangères
- > Code ELP : 4V5ALNAP

## Présentation

---

This course is an initiation to research in linguistics, as will be carried out in the Master's degree program, and more specifically to corpus linguistics, an approach to language study based on authentic data and digital corpora (i.e. the plural for "corpus"), such as YouGlish for example. Corpora may contain only written material or audio/audiovisual resources but they are all designed to be representative of language use in diverse situations of communication. As students get acquainted with oral and multimodal corpora of French and English, doing hands-on activities in a computer lab, they will discover what corpora are used for in linguistics and what tools are necessary to explore the data they comprise. Through specific case studies, students will also learn how to go from the quantitative and rather descriptive results of a corpus study to the more qualitative interpretations and analyses of those results. This course will notably focus on the following points:

- What is a corpus? How are corpora collected? Following which protocols and with what goals and ambitions?
- What are the specificities of an oral or multimodal corpus, compared to a written corpus? What aspects of language does each allow to explore? How do those different aspects participate in the overall construction of meaning?
- At a more practical level: how do you annotate an oral/multimodal corpus? What units are relevant for the study

of different linguistic phenomena? How do you code and/or observe the recurrence of certain features of pronunciation/orality?

To develop their corpus linguistics skills, students will be in charge of a mini-research project on a specific oral phenomenon of their choosing: they will learn how to come up with hypotheses on the basis of a given corpus, how to build their own corpus and their own research questions, how to look for answers grounded in authentic data, how to make methodological decisions along the way, how to survey what has already been said on their research object, and how to produce reflections that are truly their own, by being able to correctly quote and use outside sources such as bibliographical references, chat GPT, and other AI apps.

## Objectifs

---

Les objectifs principaux de ce cours sont :

- d'initier les étudiant·es à ce qu'implique la recherche en linguistique, que ce soit en termes de formulation de questions de recherche ou de maîtrise d'outils d'annotation de la langue, lorsqu'elle se fonde sur des données orales ou multimodales authentiques compilées dans des corpus ;
- de leur faire découvrir la littérature scientifique dans divers domaines linguistiques afin qu'ils apprennent progressivement à formuler, à l'écrit comme à l'oral, une synthèse de leurs questionnements, de leur démarche, de leurs résultats, et de leurs interprétations de ces résultats ;
- et de leur faire adopter une approche éthique de la recherche, afin qu'ils soient capables de formuler leurs propres réflexions sur la langue anglaise, sans plagier les ressources bibliographiques qu'ils consultent et en ayant le recul nécessaire sur le "savoir" que produisent les applications reposant sur l'intelligence artificielle, telle que chat GPT.

## Évaluation

---

Session 1 :

- Contrôle continu : moyenne d'une note de participation aux activités pratiques en classe incluant la présentation

orale de ses travaux (20% de la note finale), d'une tâche de transcription et d'annotation d'un extrait de corpus multimodal (30% de la note finale), d'une tâche de constitution d'un mini-corpus oral ou multimodal de recherche (20% de la note finale), et d'une synthèse écrite présentant la démarche et les enjeux du mini-projet de recherche mené sur un corpus oral ou multimodal (30% de la note finale).

- Contrôle dérogatoire : Épreuve orale (proposition d'une problématique de recherche et d'une méthodologie adaptée à partir d'un extrait de corpus oral ou multimodal) d'une durée de 30 min (20 min de préparation + 10 min de passage)

#### Session 2 :

Épreuve orale sans temps de préparation (entretien avec l'étudiant·e permettant de lui poser une série de questions portant sur l'ensemble des éléments couverts dans le cours).

## Pré-requis nécessaires

---

anglais C1

+ avoir validé les EC fondamentaux de grammaire et de phonologie en L1 et L2.

## Bibliographie

---

Durand, J., Gut, U., & Kristoffersen, G. (2014). *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press.

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

## Contact(s)

### > Cecile Viollain

Responsable pédagogique  
cviollain@parisnanterre.fr