

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 a practical introduction to corpus linguistics and an initiation to research as will be carried out in the Master's degree program. Corpus linguistics is an approach to language study based on digital corpora (*i.e.* the plural for 'corpus'). 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 written and oral corpora doing hands-on activities in computer labs, 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 results of a corpus study to the 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 a written vs. an oral 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 corpus? How do you extract data from a written corpus? What units are relevant for the study of different

linguistic phenomena? How do you code and/or observe the recurrence of certain features? What tools are at the disposal of the research community?

To develop their corpus linguistics skills, students will be in charge of a mini-research group project on a specific corpus or linguistic phenomenon of their choosing; they will learn how to work as part of a team, how to build research questions and look for answers grounded in authentic data, how to make methodological decisions along the way and how to synthesize the results of their study in order to make them accessible to the community.

Objectifs

Les objectifs principaux de ce cours sont :

- d'initier les étudiants 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 écrites ou orales 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 descriptive plutôt que prescriptive des phénomènes qu'ils auront découverts tout au long de leur licence afin qu'ils puissent rendre compte, à l'oral comme à l'écrit, de ce qui se passe lorsque des locuteurs natifs utilisent leur langue dans divers contextes.

Évaluation

- * Formule standard session 1 : Contrôle continu : moyenne d'une note de participation aux activités pratiques en classe (10% de la note finale), d'une synthèse orale des enjeux de recherche liés à un corpus oral (20% de la note finale), d'exercices sur la méthodologie de la recherche (20% de la note finale), d'une évaluation de l'extraction de données d'un

corpus écrit (20% de la note finale), d'une synthèse écrite présentant la démarche et les résultats du mini-projet de recherche mené sur un corpus écrit (30% de la note finale).

- * Formule dérogatoire session 1 : Epreuve 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

Avoir suivi les EC fondamentaux de Phonologie et de Grammaire en L1 et L2. Niveau anglais C1-Français B2

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)

> **Sophie Raineri**

Responsable pédagogique
sraineri@parisnanterre.fr