

- Faculté des lettres et sciences humaines
- www.unine.ch/lettres

Approches structurales : Comment utiliser les corpus (2LN1135)

| Filières concernées | Nombre d'heures | Validation | Crédits ECTS |
|--|------------------------|-----------------|--------------|
| Pilier B A - littérature française | Séminaire: 2 ph | Voir ci-dessous | 6 |
| Pilier B A - sciences du langage | Séminaire: 2 ph | Voir ci-dessous | 6 |
| Pilier principal B A - langue et littérature françaises | Séminaire: 2 ph | Voir ci-dessous | 6 |
| Pilier principal B A - sciences du langage | Séminaire: 2 ph | Voir ci-dessous | 6 |
| Pilier secondaire B A - sciences du langage | Séminaire: 2 ph | Voir ci-dessous | 6 |

ph=période hebdomadaire, pg=période globale, j=jour, dj=demi-jour, h=heure, min=minute

Période d'enseignement:

- Semestre Automne

Equipe enseignante

Cyrielle Montrichard
Bureau : Pierre-à-Mazel 7 (bureau 301)
Téléphone professionnel : 032 718 17 95
Heures de disponibilité : sur rendez-vous
Adresse électronique pour le courriel : cyrielle.montrichard@unine.ch

Contenu

L'objectif principal de ce séminaire est de sensibiliser les étudiant-e-s aux outils et méthodes d'analyse de corpus de données textuelles assistée par ordinateur. Il s'agit ainsi de découvrir une méthodologie pour analyser des textes (littéraire, médiatique, discours oraux) au moyen de logiciels permettant, entre autres, de compter, mesurer l'usage d'une forme (d'un mot, d'un groupe de mots, ...) selon des variables préalablement définies (l'auteur, l'année de publication, le genre, etc.)

Ce séminaire propose dans un premier temps de découvrir la linguistique de corpus et son évolution. Il s'agira de définir et décrire des concepts clés de la discipline (corpus, corpus outillé, statistique de données textuelles, lexicométrie, textométrie) en distinguant les méthodes et outils disponibles à travers la lecture et l'étude d'analyses produites par des chercheurs mobilisant la linguistique de corpus.

Dans un second temps, les étudiant-e-s sont amené-e-s à découvrir un logiciel d'analyse statistique de données textuelles (TXM). Une exploration guidée du logiciel permettra aux étudiant-e-s de mettre en pratique les notions abordées et de découvrir les différents angles de lecture des textes offerts par le logiciel.

Il s'agira également d'explorer une à une les étapes clés d'une analyse en linguistique de corpus : construire une question de recherche, sélectionner les textes, traiter les données, explorer le corpus, analyser les résultats et les interpréter.

NB :

* Il est nécessaire d'apporter un ordinateur portable à chaque séance afin de profiter au mieux des exercices.

* Le séminaire est donné en français.

Forme de l'évaluation

La participation à l'ensemble des séminaires est obligatoire. Il sera demandé la production d'un dossier fondé sur l'exploration d'une hypothèse de travail (mettant en œuvre les notions abordées durant le séminaire) sur un des corpus proposés par l'enseignant. La construction de ce dossier sera guidée : consignes précises sur le format et la structure du dossier ainsi que la possibilité pour l'étudiant-e d'envoyer une première version du travail pour un retour non noté mais commenté de l'enseignante. Délai pour la remise du dossier : 15 janvier 2023.

Remarque :

- la non-remise de ce travail dans les temps impartis, sans motif valable, entraînera un échec à ce travail.
- en cas de bascule à distance la forme de l'évaluation reste inchangée.

Modalités de rattrapage

- Faculté des lettres et sciences humaines
- www.unine.ch/lettres

Approches structurales : Comment utiliser les corpus (2LN1135)

En cas d'échec, le rattrapage consistera en un remaniement du dossier en s'appuyant sur les retours et commentaires de l'enseignante. La remise de ce dossier sera attendue pour le 11 juin 2023

Documentation

Un dossier de lectures (obligatoires et complémentaires) sera proposé en ligne selon les notions abordées. Néanmoins nous proposons trois publications scientifiques qui peuvent être consultées en amont du séminaire :

- Mayaffre, D. (2005). « Rôle et place des corpus en linguistique: réflexions introductives. ». inTexte(X, n°4).[en ligne] URL : http://www.revue-texto.net/1996-2007/Corpus/Publications/Mayaffre_Corpus.html
- Née, E. (2017). Méthodes et outils informatiques pour l'analyse des discours. Rennes : Presse universitaire de Rennes
- Pincemin, B., Heiden, S. (2008). «Qu'est-ce que la textométrie? Présentation ». Site du projet Textométrie. URL:<http://textometrie.ens-lyon.fr/spip.php?rubrique80>

Pré-requis

Aucun.

Aucune connaissance spécifique n'est attendue en informatique ou statistique. Il s'agit d'un cours d'initiation à ces méthodes.

Forme de l'enseignement

Séminaire

Objectifs d'apprentissage

Au terme de la formation l'étudiant-e doit être capable de :

- Interpréter les principaux calculs statistiques utilisés en textométrie
- Analyser un corpus à l'aide d'un logiciel d'analyse de données textuelles
- Produire une recherche sur corpus outillé
- Définir la linguistique de corpus et la textométrie
- Concevoir des graphiques pertinents pour traduire de manière efficace des données chiffrées
- Distinguer les notions de corpus, corpus textuel, corpus lexicographique, base de données et archive
- Sélectionner une plateforme de travail selon les objectifs de l'étude
- Examiner des corpus textuels à l'aide d'outils

Compétences transférables

- Présenter une analyse critique fondée
- Etablir des hypothèses
- Gérer un projet scientifique de l'élaboration des hypothèses à l'analyse et l'interprétation des résultats
- Communiquer les résultats d'une analyse à l'écrit
- Appliquer ses connaissances théoriques à des situations concrètes
- Mettre en oeuvre ses compétences afin des les appliquer dans le cadre d'une étude scientifique
- Synthétiser des résultats