

- Faculté des lettres et sciences humaines
- [www.unine.ch/lettres](http://www.unine.ch/lettres)

## Linguistique de corpus : introduction (2PR1466)

| Filières concernées                               | Nombre d'heures    | Validation      | Crédits ECTS |
|---------------------------------------------------|--------------------|-----------------|--------------|
| <b>Pilier B A - Humanités numériques</b>          | <b>Cours: 2 ph</b> | Voir ci-dessous | 3            |
| <b>Pilier B A - linguistique et dialectologie</b> | <b>Cours: 2 ph</b> | Voir ci-dessous | 3            |
| <b>Pilier B A - sciences du langage</b>           | <b>Cours: 2 ph</b> | Voir ci-dessous | 3            |

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

### Période d'enseignement:

- Semestre Automne

### Equipe enseignante

Linda Sanvido  
linda.sanvido@unine.ch  
Bureau 310, Pierre-à-Mazel 7  
Heures de disponibilité: sur rendez-vous

### Contenu

L'objectif principal de ce séminaire est de sensibiliser les étudiant-es aux outils et aux méthodes d'analyse de corpus de données textuelles assistée par ordinateur. Il s'agit ainsi de découvrir une méthodologie pour analyser des textes (littéraires, médiatiques, discours oraux) au moyen de logiciels permettant, entre autres, de compter et mesurer l'usage d'une forme (un mot ou 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 en distinguant les méthodes et outils disponibles à travers la lecture et l'étude d'analyses produites par des chercheur-euses mobilisant la linguistique de corpus.

Dans un second temps, les étudiant-es seront amenés à découvrir un logiciel d'analyse statistique de données textuelles (TXM). Une exploration guidée du logiciel permettra aux étudiant-es 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.

### Forme de l'évaluation

La participation à l'ensemble des séances est obligatoire. Il sera demandé la production d'un dossier fondé sur l'exploration d'une hypothèse de travail (mettant en oeuvre les notions abordées durant le séminaire) sur un des corpus proposés par l'enseignante ou sur un corpus créé par l'étudiant-e.

Délai pour la remise du dossier : 15 janvier 2027.

La non-remise de ce travail dans les temps impartis, sans motif valable, entraînera un échec.

### Modalités de rattrapage

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 5 avril 2027.

### Documentation

Un dossier de lectures (obligatoires et complémentaires) sera proposé sur Moodle selon les notions abordées.

### Pré-requis

Aucun - ce séminaire vise à fournir les bases nécessaires pour l'utilisation des logiciels d'analyse des données textuelles.

### Forme de l'enseignement

Séminaire. Il est nécessaire d'apporter avec soi un ordinateur portable pour effectuer les applications pratiques sur le logiciel TXM. Notez bien qu'il vous sera impossible d'effectuer les exercices sur une tablette.

|      |                                                                                                                                        |
|------|----------------------------------------------------------------------------------------------------------------------------------------|
| URLs | 1) <a href="https://txm.gitpages.huma-num.fr/textometrie/Presentation/">https://txm.gitpages.huma-num.fr/textometrie/Presentation/</a> |
|------|----------------------------------------------------------------------------------------------------------------------------------------|

- Faculté des lettres et sciences humaines
- [www.unine.ch/lettres](http://www.unine.ch/lettres)

## **Linguistique de corpus : introduction (2PR1466)**

### **Objectifs d'apprentissage**

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

- Décrire la linguistique de corpus
- Interpréter les principaux calculs statistiques utilisés en linguistique de corpus
- Produire une recherche sur corpus
- Concevoir des graphiques pertinents pour traduire de manière efficace des données chiffrées
- Analyser un corpus à l'aide d'un logiciel d'analyse de données textuelles

### **Compétences transférables**

- Synthétiser des résultats
- Etablir des hypothèses de recherche
- Appliquer ses connaissances théoriques à des situations concrètes
- Présenter une analyse critique fondée