

# LEXICOMÉTRIE

---

 ECTS  
4 crédits

 Composante  
Ecole  
d'histoire de  
la Sorbonne  
(EHS)

 Volume  
horaire  
24h

 Période de  
l'année  
Automne

## plugin.odf:CONTENT\_PROGRAM\_TAB01\_TITLE

### Description

Cette formation vise à doter les participants-es des connaissances et des savoir-faire nécessaires à l'analyse d'un corpus de sources textuelles à l'aide des méthodes de l'analyse des données textuelles.

Nous aborderons les techniques d'acquisition et de construction de corpus : reconnaissance optique de caractères manuscrits (HTR) et imprimés (OCR), les méthodes de structuration de corpus, de lemmatisation et d'étiquetage morpho-syntaxique du vocabulaire.

Seront également enseignées Les différentes méthodes de mesures lexicales et d'interrogation de corpus étiquetés couramment utilisées en textométrie, les analyses multivariées et les méthodes de classifications ainsi que les méthodes cooccurrence permettant l'étude des relations entre les mots. Une attention sera enfin portée aux techniques venues du domaine du traitement automatique des langues et du machine learning.

Pour ce faire différents logiciels pourront être mobilisés : Lexico, TXM, Iramuteq, Hyperbase, des paquets R ou bien Python.