

# Programmation de modèles linguistiques (II)

L6SOPROG L3 SDL

---

Alice Millour

STIH EA 4509, Sorbonne Université Sorbonne Université

# Plan de la présentation

1. Généralités
2. Retour sur le premier semestre
3. La séance d'aujourd'hui

# Généralités

---

## Organisation du semestre

- CM le vendredi 14h
- TD à 16h
- **Évaluation** : mixte (TDs à rendre, sera précisé la semaine prochaine par V. Lully)

## Ressources

- TAL et Linguistique Informatique 1, ISTE Ed. (Mohamed Z. Kurdi)
- TAL et Linguistique Informatique 2, ISTE Ed. (Mohamed Z. Kurdi)  
*Sémantique, discours et applications*
- Speech and Language Processing (Dan Jurafsky),  
<https://web.stanford.edu/~jurafsky/slp3/>

# Enseignant(s)

3 enseignants : A. Millour, V. Lully, A. Boublenza

5 séances avec moi :

- 31 janvier
- ~~7 février~~
- 14 février **décaler le cours à 13h-17h ?**
- 28 février
- 06 mars
- 13 mars

Me contacter : [alice.millour@etu.paris-sorbonne.fr](mailto:alice.millour@etu.paris-sorbonne.fr) (si je ne réponds pas dans les 48h, me relancer)

Mon bureau : D206 (sur rendez-vous)

Casier : D207 (demander Alice Millour à la personne sur place)

# Quelques règles

- vous arrivez à l'heure (et moi aussi)
- prévenir dans les plus brefs délais en cas d'absence
- faites-nous remonter les problèmes éventuels
- pas de téléphone, pas de mail, pas de réseau social, pas de nourriture (on va faire une pause)

# Retour sur le premier semestre

---

## Programmation et modèles linguistiques 1

Vos impressions ?  
prenez quelques minutes...  
ce que vous avez aimé ?  
pas aimé ?  
pas compris ?



## Programmation et modèles linguistiques 1

Vos impressions ?

prenez quelques minutes...

ce que vous avez aimé ?

pas aimé ?

pas compris ?

Des connaissances en **TAL** / des compétences en **Programmation**

compétences informatiques acquises ?

- ouvrir un fichier texte
- le tokeniser de différentes manières
- calculer l'effectif de chaque mot du texte
- en déduire les mots les plus fréquents
- exploiter des fichiers JSON en lecture et en écriture (index par exemple)
- structurer un code en factorisant via des fonctions Python simples

utilisation d'hypothèses linguistiques **simples** pour réaliser des tâches **complexes**

- identification de la langue
- utilisation des modèles "pauvres" de la sémantique
  - suppression des mots-vides
  - pondération tf-idf
  - lemmatisation

# La séance d'aujourd'hui

---

cours “pivot” : on a le temps

- faire/refaire le devoir maison
- poser toutes vos questions sur le cours du premier semestre
- réfléchir/se renseigner sur les limites de l'approche par mots-clés
- discuter des projets pour le deuxième semestre (système de recommandation)

