

## Offre de stage (5 mois)

### Techniques de TAL et modèles génératifs pour l'analyse d'un corpus littéraire

#### Contexte et présentation du projet

Ce stage s'inscrit dans le cadre de [la chaire "Enjeux écopoétiques contemporains"](#) du laboratoire ALTER à l'Université de Pau et des Pays de l'Adour (UPPA), et en particulier dans l'axe 4 ("Biodiversité") du projet.

L'objectif principal du projet est d'analyser l'évolution du discours sur la biodiversité dans un corpus de 800 romans francophones publiés entre 2001 et 2021. Une chaîne de traitement a été mise en place pour examiner la répartition des entités liées à la biodiversité ainsi que leur utilisation en contexte.

Les analyses s'appuient sur des outils de traitement automatique des langues (TAL) et des modèles génératifs. Le ou la stagiaire aura pour mission d'approfondir ces analyses en enrichissant le corpus (ajout de nouveaux textes et métadonnées, annotation des données), en participant au développement et à l'affinage d'outils adaptés (modèles type BERT, utilisation de grands modèles de langage – LLM) et en évaluant leur performance. Il ou elle pourra également explorer des méthodes générales d'analyse de corpus (modélisation thématique, textométrie, visualisation des résultats, etc.).

#### Missions

Sous la supervision des encadrants, le ou la stagiaire sera amené.e à :

- Participer à l'enrichissement du corpus en intégrant de nouveaux textes et en structurant leurs métadonnées.
- Contribuer à l'annotation des données et à l'amélioration des ressources existantes.
- Affiner les modèles existants (type BERT) et expérimenter l'utilisation des LLM pour l'analyse textuelle.
- Évaluer les performances des outils de TAL mis en œuvre.
- Explorer des méthodes avancées d'analyse de corpus (modélisation thématique, textométrie, visualisation).

#### Profil recherché

- **Formation** : Étudiant.e en M1 ou M2 en humanités numériques, informatique, TAL ou science des données. Les candidatures de L3 seront étudiées en cas d'intérêt marqué pour les thématiques du projet.

- **Compétences requises :**
  - Maîtrise basique d'un langage de programmation, idéalement Python.
  - Familiarité avec la manipulation de données et de métadonnées.
  - Intérêt pour l'analyse de textes et les enjeux écologiques en littérature.

### Équipe et encadrement:

- **Riccardo Barontini**, professeur chaire junior “Enjeux écopoétiques contemporains”, UPPA.
- **Johanna Cordova**, postdoctorante TAL, UPPA.

### Informations pratiques

- **Durée** : 5 mois
- **Début du stage** : dès que possible
- **Lieu** : Le ou la stagiaire sera basé.e à l'UPPA, avec une grande part du travail réalisée en modalité distancielle.
- **Gratification** : Indemnité légale en vigueur (environ 600 €/mois)
- **Date limite de candidature** : 25 mars 2025

### Candidature

Les candidatures (CV et lettre de motivation) sont à envoyer à [riccardo.barontini@univ-pau.fr](mailto:riccardo.barontini@univ-pau.fr) et [johanna.cordova@inalco.fr](mailto:johanna.cordova@inalco.fr) avec pour objet "Candidature Stage TAL – écopoétique".