

Proposition de Contrat Postdoctoral

Le laboratoire ATILF propose un CDD sur travaux scientifiques de 12 mois en linguistique informatique – dans le cadre de l’obtention d’un financement ANR de trois ans (MSH Lorraine –Inalco– CNRS ATILF).

Le chercheur travaillera avec l’équipe constituée autour du projet ALIENTO (Analyse Linguistique, Interculturelle d’Énoncés sapientiels et Transmission Orient/Occident-Occident/Orient)

Début du contrat : 4 janvier 2016.

DESCRIPTION DU PROJET DANS LEQUEL LE POSTE S’INSCRIT ET MISSION DU POST-DOCTORANT

ALIENTO est un projet interdisciplinaire innovant qui s’inscrit dans le domaine nouveau des humanités numériques. Notre démarche interdisciplinaire croise les pratiques scientifiques entre spécialistes des textes, linguistes, parémiologues et chercheurs en informatique. Le dialogue des disciplines à propos d’un objet commun (la circulation des énoncés sapientiels brefs médiévaux) ouvre de nouvelles perspectives de recherche tant méthodologiques que conceptuelles. En ce sens, ALIENTO occupe une place centrale dans le traitement scientifique et méthodologique des textes sapientiels, tant par la nature de ses corpus et la manière de les traiter, sa gestion du multilinguisme, sa réflexion en matière d’étiquetage, que par l’originalité de son approche qualitative de l’exploitation de corpus (interopérabilité homme-machine).

➤ **Objectifs du projet :**

Le projet vise à l’élaboration d’un outil informatique (un logiciel transférable à d’autres textes semblables) à partir d’un important corpus de référence constitué de 8 textes apparentés ayant circulé dans la Péninsule ibérique au Moyen Âge (en latin, arabe, hébreu, espagnol et catalan soit 582 pages de textes pour un nombre d’énoncés sapientiels brefs évalué à 9570 unités permettant d’extraire les connaissances des corpus annotés sur les sources, la transmission, la circulation et la postérité des énoncés sapientiels de la Péninsule Ibérique (IX^e-XV^e siècle) entre les Trois Cultures).

L’objectif est d’apporter des réponses actuellement hors de portée faute d’outils pertinents pour traiter des corpus apparentés mais multilingues et en grande partie non alignables ; de donner, enfin, grâce à la création collective d’un outil informatique approprié d’automatisation de l’appariement des données sapientiellles multilingues (dont l’arabe et l’hébreu), la possibilité aux médiévistes d’accéder aux savoirs les uns des autres, au-delà de la barrière des langues et des disciplines.

Les étapes de travail sont :

– l’annotation des textes à l’aide des balises XML/TEI compatibles, portant sur le sens, la forme, la structure, la leçon de chaque énoncé sapientiel bref repéré, de façon à pouvoir croiser les énoncés sapientiels et développer par expérimentation collaborative avec les chercheurs informaticiens un protocole de balisage performant pour ce type d’unités, transférable à d’autres corpus (formalisation ; normalisation ; création d’une ontologie commune à ces types d’énoncés) ;

- L'élaboration d'un outil capable d'effectuer les appariements et constituer une interface trilingue permettant de consulter ces résultats, c'est-à-dire, les liens établis entre les énoncés sapientiels brefs ;
- L'analyse des résultats des appariements
- La diffusion des résultats scientifiques dans le domaine des humanités permettant d'étudier la circulation des énoncés sapientiels brefs, leurs transformations et leurs degrés d'acculturation en fonction de leur intertextualité propre.

Le présent projet d'ANR ALIENTO se situe dans le prolongement de 4 années de recherche.

Pour en savoir plus : www.aliento.eu

➤ **Mission du post-doctorant :**

Le/La post-doctorant(e) aura une formation en linguistique informatique. Sa mission sera de proposer et développer une méthodologie utilisant toutes les annotations (sens, forme, structure, traduction multilingue) ajoutées par les chercheurs du projet pour calculer l'appariement des énoncés sapientiels. L'objectif est donc de développer un outil permettant de détecter automatiquement les similarités entre énoncés pour les relier afin d'en faciliter la consultation ultérieure. Le travail sur les aspects multilingues sera une véritable plus-value sur les résultats attendus.

Il/Elle travaillera à Nancy (Laboratoire ATILF, CNRS/Université de Lorraine) en étroite collaboration avec l'équipe ALIENTO (Paris-Nancy).

Compétence :

- Doctorat en linguistique informatique
- Bonne connaissance des langages de balisages (XML)
- Bonne connaissance des approches probabilistes en linguistique

➤ **Contrat :**

Durée : 12 mois

Aucune condition de nationalité n'est exigée. Pour les candidats non francophones, la langue de communication peut être l'anglais

Début du contrat : 4 janvier 2016

Date limite de candidature : 20 novembre 2015

Salaire brut mensuel : 2525€ à 2913€ /mois selon expérience

Envoyer un CV de trois pages maximum, le résumé de la thèse et une lettre de motivation à :

Étienne Petitjean : etienne.petitjean@atilf.fr

Jean-Marie Pierrel : jean-marie.pierrel@atilf.fr