

**Post-doctorat à l'Université Grenoble Alpes (LIDILEM) en linguistique et traitement de corpus**

**Extraction et modélisation de phrases préfabriquées à partir de corpus d'interactions**

(22 mois à partir du 1/10/23, date de début négociable)

**-- Champs de recherche --**

Linguistique, Traitement automatique des langues

**-- Mots-clés --**

Phraséologie – Corpus oraux – Analyse syntaxique – Pragmatique – Interactions

**-- Contexte et objectifs --**

Le projet ANR PREFAB (<https://prefab.hypotheses.org>) réunit des linguistes et des spécialistes du traitement automatique de quatre laboratoires français, dont le LIDILEM, qui pilote le projet. Il a pour objectif d'aborder une thématique émergente, l'étude des phrases préfabriquées des interactions comme *ça marche ! tu plaisantes ! comment dirais-je ? tu peux le dire. c'est OK*. Ce thème n'a pas encore fait, pour le français, l'objet d'études sur de grands corpus, malgré son importance pour la didactique des langues étrangères ou le traitement automatique des langues. L'objectif du projet est double : il vise, d'une part, à extraire ces phrases à partir de grands corpus (corpus oraux transcrits, corpus écrits d'interactions et corpus romanesques) et, d'autre part, à modéliser les constructions les plus saillantes correspondant à ces phrases.

La personne recrutée aura pour fonction de réfléchir en collaboration avec les équipes du projet à une chaîne de traitement permettant l'extraction des phrases préfabriquées à partir de corpus d'interactions et la modélisation de ces expressions exploitant les grammaires de construction.

**-- Activités principales -- :**

Plusieurs tâches sont prévues dans le cadre du projet :

- participer avec l'équipe à la réflexion sur la modélisation des phrases préfabriquées, à partir d'observations sur un sous-ensemble d'expressions dans une étude pilote portant sur des données extraites des différents corpus.
- prétraiter les corpus du projet comportant de l'oral transcrit, des interactions écrites et des romans, en utilisant des analyseurs syntaxiques récents tenant compte des modèles en usage (UD, SUD).
- extraire les candidats phrases préfabriquées à l'aide de méthodes exploitant les structures syntaxiques et l'analyse sémantique automatique.
- modéliser les principales constructions associées aux phrases préfabriquées les plus saillantes, en utilisant un modèle inspiré des « constructicons ».

**-- Groupe de recherche --**

Le postdoctorat sera intégré au Laboratoire LIDILEM EA 607 (Linguistique et Didactique des Langues Etrangères et Maternelles, <https://lidilem.univ-grenoble-alpes.fr/>) qui regroupe actuellement une soixantaine de membres permanents et environ 70 doctorants dans les domaines suivants : descriptions linguistiques, sociolinguistique, acquisition, constitution et exploitation de corpus, didactique des langues, traitement automatique des langues, étude des formes nouvelles d'interaction suscitées par les usages numériques.

Le postdoctorat s'effectuera en collaboration étroite avec trois autres équipes de recherche (laboratoires ICAR, ATILF et BCL)

**-- Profil du candidat --**

La personne recrutée doit avoir de solides compétences en description sémantique et pragmatique et en linguistique de corpus. Des connaissances en phraséologie et syntaxe de l'oral seront appréciées, ainsi qu'une aisance avec les chaînes de traitement en TAL (notamment via la ligne de commande) et les formats de données annotées (XML, CONLL, TSV). Une compétence native (ou quasi-native) en français est requise.

La personne recrutée devra faire preuve d'autonomie et d'initiative. Elle devra être en mesure de travailler en équipe et de communiquer ses résultats à l'oral et à l'écrit.

**-- Salaire --**

À partir de 2480 euros bruts par mois selon expérience.

**-- Processus de candidature --**

Envoyer un CV et une lettre de motivation détaillée à :

Agnès Tutin (LIDILEM) : [agnes.tutin@univ-grenoble-alpes.fr](mailto:agnes.tutin@univ-grenoble-alpes.fr) et Olivier Kraif (LIDILEM) : [olivier.kraif@univ-grenoble-alpes.fr](mailto:olivier.kraif@univ-grenoble-alpes.fr)

**Date limite** : 30 juin 2023