

# Fiche de poste

## Ingénieur d'étude à mi-temps en linguistique de corpus / documentation et communication

**Durée:** 1 an (renouvelable)

**Emploi-type:** Ingénieur-e d'études en linguistique de corpus / documentation et communication

**Quotité:** mi-temps

**Lieu de travail:** à déterminer sachant qu'une partie du travail sera effectuée à distance (travail de réseau).

### 1. Contexte

Le consortium Huma-Num CORLI (Corpus, Langues, Interactions) est né du rapprochement des deux consortiums précédents de linguistique de la [TGIR Huma-Num](#) : Corpus Écrits (IRCE) et Corpus Oraux et Multimodaux (IRCOM).

Leurs membres ont souhaité poursuivre leurs activités dans le cadre d'un nouveau programme scientifique autour des notions de Langues, Corpus et Interactions.

Le consortium CORLI, coordonné par Christophe Parisse et Céline Poudat et géré par la [MESHS de Lille](#), réunit des chercheurs et enseignants-chercheurs en linguistique, et se donne pour objectif de fédérer les équipes et laboratoires, les chercheurs, enseignants chercheurs, ou ingénieurs engagés dans la production et le traitement de corpus numériques écrits et oraux, quels que soient la langue et/ou le système d'écriture considérés.

### 2. Missions

Les missions de l'ingénieur-e seront d'accompagner les projets du consortium CORLI en assistant les responsables du consortium et les responsables des différents projets et livrables en assurant les tâches de gestion, documentation, mise à jour du site et communication auprès des participants et des partenaires. Suivant ses compétences, il/elle participera également aux projets scientifiques développés par le consortium, et à l'*Open French Corpus* en particulier.

### Projet annotation

La transcription et l'annotation de corpus sont des opérations qui sont au cœur des humanités numériques et la question de l'annotation collaborative de corpus est au cœur du nouveau projet CORLI.

Trois axes sont actuellement privilégiés:

- la mise à disposition d'une plateforme de transcription et d'annotation simple des données langagières - livrable réalisé dans le cadre du projet Palamède, issu d'une collaboration initiée en 2020-2021 entre la MSH Lorraine, CORLI, Huma-Num, l'Atilf et Lit&art et différents concepteurs d'outils de transcription, à commencer par TACT;
- la mise à disposition d'une plateforme d'annotation de haut niveau avec des fonctionnalités d'active learning, en collaboration avec l'équipe d'Inception (TU Darmstadt) avec laquelle nous avons déjà commencé de travailler en 2020;
- une ressource collaborative d'annotation en classe sur le modèle de GUM (<https://corpling.uis.georgetown.edu/gum/>), à laquelle participent les laboratoires CLLE, Loria, Lidilem et BCL.

## Projet citation

Une fois que les corpus sont dans des formats FAIR, ils ont pour but d'être utilisés et réutilisés pour la recherche ouverte. Le projet CITATION a pour but de créer des outils utilisateurs permettant de créer et d'utiliser des citations de corpus ou d'extraits de corpus. Les citations elles-mêmes suivront les standards existants ou proposés par un institut comme RDA (Research Data Alliance). Les outils créés permettront de:

- sélectionner dans des corpus déposés ou dans l'Open French Corpus des emplacements constituant un extrait de corpus
- créer des pages web pérennes permettant de visualiser ou présenter un corpus ou un extrait de corpus (soit sélectionné précédemment, soit manuellement inséré)
- générer des références bibliographiques pointant sur les pages web pérennes et insérées dans le corps et dans la partie bibliographie d'un texte scientifique (aux formats RIS, BibTex, etc. et pouvant donc être utilisées dans un outil comme Zotero).

Ce travail s'intègre dans la philosophie FAIR, de la science ouverte et de l'exploitation des *data papers*. Il permettra d'améliorer l'accès et la visibilité des travaux de création et de dépôt de corpus.

## Projet Open French Corpus

Ce projet a pour objectif de centraliser les corpus de Français déjà existants et issus de divers projets, tous validés et normalisés par la communauté et de les mettre à disposition dans un espace commun avec un outillage approprié pour les utiliser. Ce projet comporte trois phases pouvant se dérouler en parallèle:

- identifier, rassembler les corpus existants ainsi que les méthodes, techniques et formats utilisés pour les constituer
- déterminer un noyau minimum de format, de qualité et de préparation des corpus devant être rendus disponible, présentation d'une chaîne de traitement permettant de normaliser de nouveaux corpus ou de mettre à niveau d'anciens corpus
- rendre disponible les corpus, en téléchargement, en recherche plein texte, en recherche outillée

Tous les projets CORLI suivent une même politique qui est d'utiliser des outils ou des données existantes et d'éviter au plus le développement de technologies à partir de zéro. Au contraire, CORLI s'attache à mieux faire connaître les technologies existantes et à créer des ponts (au besoin en acceptant de faire des développements nécessaires) ou de la documentation pour les rendre disponibles ou mieux les exploiter.

## 3. Activités

- Participer à la gestion du consortium CORLI
  - Suivi des activités des groupes projets
  - Diffusion des comptes rendus et informations diverses
  - Suivi de la gestion financière réalisée par la MSH
- Mettre à jour le site Web
  - Agenda et événements
  - Documentation et informations diverses
- Rédaction de la documentation scientifique avec la participation des membres de CORLI
- Participer scientifiquement ou techniquement à un des projets du consortium (au moins): 1. *Open French Corpus* ; 2. Annotation; 3. Citation.

## 4. Compétences et savoir-faire

- Compétences en gestion de projet. Une expérience antérieure de gestion de projet serait appréciée.
- Compétence en communication (genres de la communication, *i.e.* posters, annonces, réseaux sociaux). Bonnes compétences rédactionnelles. Des compétences en graphisme seraient un plus.
- Connaissance des CMS et de WordPress en particulier.
- Anglais requis - capacité à communiquer en anglais avec nos partenaires (*e.g.* Clarin, TU Darmstadt, réseau CMC-corpora) et à traduire le site Web en anglais.
- Familiarité / expérience avec le domaine de la linguistique de corpus et la construction, constitution de corpus textuels (échantillonnage, structuration, métadonnées, formats)
- Bonne connaissance des méthodes et des outils de la linguistique de corpus

## 5. Savoir-être

- Capacité à travailler en réseau
- Autonomie et prise d'initiative
- Capacité d'écoute et aisance dans les interactions

---

Contact: envoyer CV et lettre de motivation à Christophe Parisse [cparisse@parisnante.fr](mailto:cparisse@parisnante.fr) et Céline Poudat [celine.poudat@univ-cotedazur.fr](mailto:celine.poudat@univ-cotedazur.fr) avant le 31 mars.

Début du contrat possible à partir du 1er mai.