

 <p>RÉPUBLIQUE FRANÇAISE Liberté Égalité Fraternité</p> <p>UNC UNIVERSITÉ de la NOUVELLE-CALÉDONIE</p>	VACANCE DE POSTE	
	Un post-doctorant en traitement automatique des langues(TAL/NLP)	
	<u>Date de prise de fonction</u> Vacant à partir du 01/02/2025	<u>Mode de recrutement</u> CDD de 36 mois A temps plein

Présentation de l'établissement

L'Université de la Nouvelle-Calédonie est un établissement pluridisciplinaire qui répond notamment aux besoins de formation et de recherche propres à la Nouvelle-Calédonie. Elle veille à accompagner efficacement les évolutions de la Nouvelle-Calédonie et à répondre à ses besoins spécifiques.

L'UNC, ancrée dans son environnement et sa région, a pour ambition de promouvoir son activité de recherche sur la base de l'excellence et de la reconnaissance nationale et internationale. Cette promotion passe par la mise en valeur de ses enjeux scientifiques, de ses capacités d'innovation et de transfert ainsi que par la qualité des formations qu'elle dispense.

L'UNC mène une politique académique et scientifique dynamique et reconnue.

Ainsi l'UNC est lauréate des appels à projets "Nouveaux cursus à l'université" et "Dispositifs territoriaux pour l'orientation vers les études supérieures" du Programme d'Investissement d'Avenir 3. Sur le plan scientifique, l'université est lauréate d'un appel à projets très sélectif du schéma directeur pour la recherche et l'innovation "Horizon 2020" de la commission européenne.

L'UNC est depuis août 2023 lauréate de la troisième vague du très sélectif appel à projets ExcellencES, avec son projet "DiversitéS", bâti sur sa singularité institutionnelle, son ancrage en Nouvelle-Calédonie et embrassant les diversités, biologiques, culturelles et linguistiques de Nouvelle-Calédonie.

L'UNC en chiffres, c'est 250 personnels, 3700 étudiants, 3 départements de formation (Droit, Economie, Gestion ; Lettres, Langues, Sciences Humaines ; Sciences et Techniques), 1 IAE, 1 IUT, 1 INSPE, 3 équipes de recherche, 2 UMR, 1 école doctorale.

L'UNC, c'est également deux campus dynamiques et chaleureux (Nouvelle en province Sud et Baco en province Nord), des infrastructures modernes (installations dédiées à la recherche et aux pédagogies innovantes, plateaux techniques, studio audiovisuel, Fablab, ...) des installations sportives de qualité, un accès privilégié à la vie culturelle et artistique, et un environnement et une qualité de travail uniques.

Contexte de travail :

Le post-doctorant retenu intégrera l'équipe du projet Jeune Chercheur *DiKaLa* (*Digital tools for Kanak Languages*) financé par l'Agence Nationale de la Recherche. L'équipe est constituée d'une linguiste spécialiste des langues kanak de Nouvelle-Calédonie (coordinatrice scientifique du projet), d'un ingénieur en développement web et applications multilingues et du post-doctorant en TAL/NLP.

Cette équipe bénéficiera du soutien technique et matériel de la Direction au Numérique et aux Services Informatiques de l'UNC

Atouts du poste :

- Participer au développement d'outils numériques pour des langues peu dotées ;
- Contribuer à renforcer la préservation des langues autochtones ;
- Participer au développement d'un nouveau domaine de recherche appliquée à l'UNC ;
- Collaborer avec des experts de divers domaines ;
- Travailler dans une université dynamique avec un environnement de travail à échelle humaine, flexible et orientée vers la recherche et l'innovation.

Missions et caractéristiques particulières du poste :

Le projet ANR DiKaLa a pour objectif la création et la mise en ligne de 3 outils numériques pour les langues kanak : une base de données lexicales multilingue ; un atlas linguistique multimodal ; une plateforme d'archivage de corpus textuels et audio. Ces trois outils s'inscrivent dans une approche intégrative et participative permettant l'ajout de données de la part des utilisateurs et seront rendus librement accessibles en ligne. Ces outils seront développés en accord avec une démarche de connaissance ouverte (open source, open data) afin de garantir la pérennité du code et des données, et conformément aux valeurs de transparence, d'expertise et de transversalité Sciences-Société du projet DiKaLa.

Le post-doctorant retenu aura pour principales missions (non exhaustives) de :

- Effectuer une veille scientifique et technique sur le champ du TAL ;
- Mettre au point des stratégies technologiques d'extraction et de conversion de données textuelles (principalement à partir de lexiques et dictionnaires océrés) pour des langues peu dotées ;
- Evaluer et envisager les meilleures solutions technologiques, ontologiques ou de modèles de machine learning pour atteindre les objectifs visés ;
- Paramétrer des solutions de tokenisation et d'étiquetage morpho-sémantique pour l'annotation de corpus de textes (monolingues, bilingues, audio transcrits et semi-transcrits) ;
- Structurer l'architecture d'une plateforme d'archivage de ressources langagières et de corpus multimédias, en veillant au bon catalogage des métadonnées ;
- Développer ou sélectionner les stratégies technologiques adéquates de Data mining (extraction de données pour traitement lexicographique) ;
- Formater et vérifier la qualité des données linguistiques retenues (textes en langues kanak nativement numériques ou numérisées, documents historiques, données orales transcrites ou non), corriger les erreurs d'OCR, supprimer les bruits et données inutiles, normaliser les données (homogénéisation des formats) pour alimenter les modèles de TAL ;
- Maîtriser ou se former aux outils sélectionnés : Scraping pour la collecte web, bases de données linguistiques ou bibliothèques de documents existants, intégrant l'implication d'experts en linguistique des langues kanak pour assurer la pertinence culturelle et contextuelle des données ;
- Analyser les besoins utilisateurs : collaborer avec les utilisateurs potentiels, notamment des linguistes et locuteurs natifs, pour comprendre les usages concrets et attentes spécifiques ;
- Privilégier une méthodologie agile, prototypage rapide, et utilisation de bibliothèques TAL (par exemple SpaCy, Hugging Face, ou Transformers) pour implémenter et tester les solutions ;
- Adapter des modèles existants (par ex. des modèles de génération de texte ou de classification) aux spécificités linguistiques des langues kanak (peut inclure la fine-tuning des modèles pré-entraînés ou la création de modèles sur mesure) ;
- Déployer les outils dans l'environnement technique défini et former les utilisateurs pour garantir une adoption efficace de l'outil et pour assurer le cycle de vie du logiciel sur sa partie MCO/MCS ;
- Rédiger des notices et produire des tutoriels, en collaboration avec l'équipe du projet.

Compétences et qualifications attendues :

- Être titulaire d'une thèse de doctorat en Linguistique computationnelle, en traitement automatique du langage (TAL), ou en informatique (option en Intelligence artificielle) ;
- Excellent profil au niveau de la recherche (publications, communications, collaborations, etc.) ;
- Très bonne connaissance des principaux outils et algorithmes du TAL (analyse et génération) ;
- Maîtrise des technologies de reconnaissance optique de caractères (tapuscrits et manuscrits) ;
- Excellentes compétences en programmation Python, PyTorch, JSON, Javascript ou d'autres frameworks pertinents, expérience utilisation et idéalement adaptation/modification d'architectures neuronales existantes ;
- Bonnes connaissances des concepts et des techniques avancées en intelligence artificielle ;
- Sens de la communication et de l'écoute ;
- Sens de l'organisation, flexibilité et capacité à s'adapter à des environnements de travail changeants et dynamiques ;
- Grande aisance dans la collaboration et au travail en équipe en milieu multiculturel ;
- Aptitude à identifier des problèmes complexes et à développer des solutions innovantes ;
- Bonne capacité à travailler de manière autonome, à prendre des initiatives, à être créatif et avoir une curiosité d'esprit ;
- Capacité de médiation et de vulgarisation pour expliquer des concepts techniques à des non-spécialistes ;
- Compétences en français exigées ; de bonnes compétences en anglais sont attendues ;
- Un intérêt pour les langues en général, les langues régionales, les langues océaniques, la diversité linguistique et les humanités numériques est vivement souhaité.

Contacts utiles :

Anne-Laure DOTTE, coordinatrice scientifique du projet DiKaLa : anne-laure.dotte@unc.nc
Camille VERBRUGGHE, pôle enseignants et enseignants-chercheurs : recrutement@unc.nc

Les dossiers de candidature (une lettre de motivation en français, un curriculum vitae détaillé en français ou en anglais, une ou deux lettres de référence en français ou en anglais, en particulier de votre directeur de thèse) sont à envoyer **en format pdf** par voie électronique à la direction des ressources humaines de l'Université de la Nouvelle-Calédonie : recrutement@unc.nc

au plus tard le 25/11/2024