

Postdoctoral position: Scientific corpus analysis and semantic annotation

Location: Marie and Louis Pasteur University, CRIT Laboratory, Besançon, France

Duration: 24 months

Preferred starting date: between June and October 2026 (to be specified)

Context

As part of an application to the ATTRACT programme (Accueil de Talents de la Recherche et Appui à la Coopération sur le Territoire), the CRIT laboratory is seeking to recruit a high-level early-career researcher for a 24-month postdoctoral position in Natural Language Processing. The objective is to strengthen the laboratory's work on large-scale scientific discourse analysis, in continuity with projects on the modelling of scientific uncertainty across disciplines (ANR InSciM project) and, more generally, on annotation and information extraction from scientific corpora. The research will be conducted under the supervision of Prof. Iana Atanassova.

Scientific focus

The successful candidate will work on large-scale scientific corpora, with a focus on:

- Analysis of scientific corpora (articles, reports, preprints, etc.) using NLP methods, including linguistically informed, machine learning and deep learning approaches.
- Analysis of scientific discourse across disciplines, with applications to epistemology.
- Development or adaptation of methods for text mining, information extraction and/or semantic modelling (e.g. contextual embeddings, language models, classification methods).
- Design of reusable resources and tools (annotation schemes, annotated corpora, models, scripts) in connection with ongoing projects in the laboratory.

The project may be adapted to the profile and research interests of the successful candidate, within the scope of the laboratory's thematic priorities.

Profile

The candidate should:

- Hold a PhD in Natural Language Processing, Computational Linguistics, Computer Science, Data Science, or a closely related field.
- Have defended their PhD less than 10 years ago.
- Have a strong scientific track record (international publications, conference presentations, participation in research projects).
- Demonstrate expertise in several of the following areas:
 - NLP applied to scientific or specialised texts.

- Manual and/or semi-automatic corpus annotation (annotation schemes, guidelines, coordination of annotators).
- Rule-based and linguistically informed methods and algorithms.
- Neural models, embeddings, LLMs, etc.
- Information extraction and discourse analysis.
- Have solid scientific programming skills (Python, NLP/ML ecosystem such as spaCy, HuggingFace, scikit-learn, etc.).
- Currently be affiliated with an institution located outside the Bourgogne–Franche-Comté region and be willing to join UMLP (Besançon) for the duration of the contract.
- Have an excellent command of scientific English and French. Both languages will be used in project meetings.

Experience in epistemology will be particularly appreciated.

The position is part of an application currently being prepared for the ATRACT programme of the Bourgogne–Franche-Comté Region. The successful candidate will benefit from a multidisciplinary scientific environment within the CRIT laboratory and from regional, national and international collaborations.

Application

Interested candidates are invited to send the following documents by e-mail before 28 December 2025:

- A detailed CV (including a full list of publications).
- A cover letter describing:
 - Main research contributions.
 - Possible connections with the research topics described above and with the proposed project.
- The contact details of two referees (name, affiliation, e-mail) who may be contacted for reference letters during the week of 5–9 January.

Applications should be sent to: iana.atanassova@umlp.fr

Suggested e-mail subject line: Application ATRACT Postdoctoral Position - [Last name First name]

Applications will be reviewed on a rolling basis. Online interviews may be organised in early January, in view of submitting the ATRACT proposal before 10 January.

Offre de post-doctorat : Analyse de corpus scientifiques et annotation sémantique

Lieu : Université Marie et Louis Pasteur, Laboratoire CRIT, Besançon, France

Durée : 24 mois

Prise de fonction souhaitée : entre juin et octobre 2026 (date à préciser)

Contexte

Dans le cadre d'une candidature au dispositif ATRACT (Accueil de Talents de la Recherche et Appui à la Coopération sur le Territoire), le laboratoire CRIT souhaite recruter une ou un jeune chercheur de haut niveau pour un contrat post-doctoral de 24 mois en Traitement Automatique des Langues. L'objectif est de renforcer les travaux du laboratoire sur l'analyse de discours scientifique à grande échelle, dans la continuité des projets sur la modélisation de l'incertitude scientifique dans les différentes disciplines (projet ANR InSciM) et plus généralement l'annotation et l'extraction d'information à partir de corpus scientifiques. Les travaux de recherche seront menés sous la direction de Pr. Iana Atanassova.

Thématique scientifique

La personne recrutée travaillera sur des corpus scientifiques à grande échelle, avec un focus sur :

- L'analyse de corpus scientifiques (articles, rapports, préprints...) au moyen de méthodes de TAL, y compris méthodes à base de connaissances linguistiques, apprentissage automatique et profond.
- L'analyse de discours scientifique dans les différentes disciplines, avec des applications en épistémologie.
- Le développement ou l'adaptation de méthodes de fouille de textes, d'extraction d'information et/ou de modèles sémantiques (par ex. embeddings contextuels, modèles de langage, méthodes de classification).
- La mise en place de ressources et d'outils réutilisables (schémas d'annotation, corpus annotés, modèles, scripts) en lien avec les projets du laboratoire.

Le projet pourra être ajusté en fonction du profil et des intérêts de la personne recrutée, dans le cadre des thématiques du laboratoire.

Profil recherché

Le ou la candidate devra :

- Être titulaire d'un doctorat en traitement automatique des langues, linguistique computationnelle, informatique, science des données ou discipline proche.
- Avoir soutenu sa thèse depuis moins de 10 ans.
- Disposer d'un solide dossier scientifique (nombreuses publications internationales, communications, participations à des projets de recherche).
- Présenter une expertise dans plusieurs des domaines suivants :
 - Le TAL appliqué aux textes scientifiques ou spécialisés.
 - Annotation manuelle et/ou semi-automatique de corpus (schémas d'annotation, guidelines, coordination d'annotateurs).

- Traitements et algorithmes à base de règles et connaissances linguistiques.
- Modèles neuronaux, embeddings, LLM, etc.
- Extraction d'information et analyse de discours.
- Maîtriser la programmation scientifique (Python, écosystème TAL/ML type spaCy, HuggingFace, scikit-learn, etc.).
- Être actuellement rattaché,e à un établissement situé hors de la Bourgogne–Franche-Comté et prêt,e à rejoindre l'UMLP (Besançon) pour la durée du contrat.
- Avoir un très bon niveau d'anglais scientifique et du français. Les deux langues seront utilisées lors des réunions de travail.
- Une expérience dans le domaine de l'épistémologie sera particulièrement appréciée.

Le poste s'inscrit dans un dossier en cours de montage pour le dispositif ATRACT de la Région Bourgogne–Franche-Comté. La personne recrutée bénéficiera d'un environnement scientifique pluridisciplinaire au sein du laboratoire CRIT et des collaborations régionales, nationales et internationales.

Candidature

Les candidat·es intéressé·es sont invité·es à envoyer par courriel, avant le 28 décembre 2025, les éléments suivants :

- Un CV détaillé (incluant la liste de publications).
- Une lettre de motivation explicitant :
 - Les principales contributions de recherche.
 - Les liens possibles avec les thématiques décrites ci-dessus et le projet proposé.
- Les coordonnées de deux référents (nom, affiliation, e-mail) susceptibles d'être contactés pour des recommandations dans la semaine du 5 au 9 janvier.

Les candidatures sont à envoyer à : iana.atanassova@umlp.fr

Objet du mail suggéré : Candidature post-doc ATRACT – [Nom Prénom]

Les candidatures seront examinées au fil de l'eau. Des entretiens en visioconférence pourront être organisés début janvier, en vue du dépôt du dossier ATRACT avant le 10 janvier.