

## **APPEL A PROJETS IREEL**

### **PROGRAMME D'INITIATION A LA RECHERCHE EN LICENCE**

- > Clôture de l'appel : le 26 avril 2019
- > Soumission des dossiers à : AO-IdEx-UnivParis@uspc.fr

#### **1. INFORMATIONS GENERALES**

1.1 Titre du projet :  
RELIA - Recherche En Licence Informatique et études Anglophones

1.2 Parcours de formation concernés (plusieurs choix possibles pour un projet interdisciplinaire) :  
Licence Informatique et Licence d'Études Anglophones

1.3 Calendrier du programme d'initiation à la recherche :

☐ Semestre 2 (année universitaire 2018/2019)

X Semestre 1 (année universitaire 2019/2020)

1.4 Catégorie du programme d'initiation à la recherche (plusieurs choix possibles) :

PRATIQUE DE LA RECHERCHE :

☐ Gratification de stages en laboratoires

X Prise en charge de la participation à des colloques, congrès ou séminaires

METHODOLOGIE :

X Modules de formation sur la méthodologie de la recherche (dont potentiellement initiation à la rédaction de rapports en anglais)

☐ Séminaires (sujets scientifiques, éthique et déontologie...)

AUTRES (si oui, prière de détailler) :

---

1.4 Nombre d'étudiants ciblés par le projet :  
Pour initier le programme, 5 groupes de deux étudiants (un par discipline)

---

1.5 Existait-il déjà des programmes d'initiation à la recherche ouverts aux étudiants de votre parcours de formation :  
En informatique nous avons déjà un module mais qui ne concernent que les étudiants de la licence pluridisciplinaire Mathématique et Informatique.

#### **2. PORTEUR/EUSE(S) DU PROJET**

|   |   |
|---|---|
| Nom : BALLIER                                     | Prénom : Nicolas  |
| Fonction : Professeur                             | Établissement et composante : Univ. Paris Diderot, UFR d'Études Anglophones |
| Email : nballier@free.fr                          | Numéro de téléphone : 0157275874  |
| Nom : YUNÈS                                       | Prénom : Jean-Baptiste  |
| Fonction : Maître de conférences                  | Établissement et composante : Univ. Paris Diderot, UFR Informatique         |
| Email : Jean-Baptiste.Yunes@univ-paris-diderot.fr | Numéro de téléphone : 0157279262  |

#### **Contact de soumission des questions**

Azwaw Djebara, chargé de mission formation USPC  
[azwaw.diebara@uspc.fr](mailto:azwaw.diebara@uspc.fr) – 01 43 45 45 05

### 3. DESCRIPTION GENERALE DU PROJET ET DES RESULTATS ATTENDUS (5 PAGES MAX.)

#### Objectifs du cours :

Ce cours vise à la transmission des compétences initiales pour la soumission d'abstract à des conférences dans le domaine des humanités numériques. On cherchera à décrire les problématique linguistiques et les jeux de données envisagés pour les analyser, décrire ces données et établir les traitements automatiques nécessaires.

#### Contexte scientifique :

Insuffisance de la formation à la programmation en Études Anglophones.

Insuffisance de l'ouverture aux autres disciplines (application) en Informatique.

Nécessité de l'émergence d'une dynamique en traduction automatique, traitement automatique de documents en langue naturelle, reconnaissance vocale et compréhension. Le laboratoire CLILLAC-ARP a acquis une certaine expérience pour projeter ses étudiants de Master en conférence et ce dès le M1 pour les plus mûrs d'entre eux.

#### Identification des publics et objectifs poursuivis :

Pour l'UFR d'Informatique, il s'agit de contribuer à diversifier les objets de recherche, à encourager les étudiants à se tourner vers une recherche ouverte aux autres disciplines, à développer leur culture en traitement automatique du langage.

Pour l'UFR d'Études Anglophones, il est vital de faire redescendre un pan important de la recherche en linguistique plus tôt dans les cursus et en particulier les enseignements d'initiation à la programmation. Il s'agit de transmettre une culture de la programmation dans un cursus de LSHS et une initiation au paradigme théorique de «learnability» (*machine learning*) ainsi qu'aux pratiques en sciences des données.

Pour les étudiants en informatique, l'objectif est de s'ouvrir à d'autres disciplines de découvrir les pratiques de celles-ci et d'acquérir une plus large culture scientifique.

Il s'agit d'une démarche à titre expérimental, qui gagnerait à s'inscrire dans la durée, en particulier, une articulation serait possible avec le projet d'École d'été d'initiation à la recherche en Licence prévu dans le document d'accompagnement de l>IDEX transmis aux laboratoires.

#### Contenus:

Intervention de linguistes, d'informaticiens et de spécialistes de l'apprentissage : MINDS (Machine Intelligence and Data Science), Machine Translation, Information Retrieval, Natural Language Processing, Document Processing, Speech Recognition and Understanding.

L'enseignement se concentrera sur les fondamentaux du TAL, la caractérisation des unités linguistiques, la détection automatique dans les textes ; le tout dans une approche centrée sur les activités (learning by doing, active learning) :

Problèmes de représentations de la prononciation (phonologie)

Problèmes de l'analyse automatique des morphèmes (morphologie)

Problèmes de la catégorisation grammaticale (PoS-tagging)

Problèmes de l'analyse en dépendances.

Problème de l'étiquetage de l'anglais du XXe à l'aide d'un dictionnaire phonologique.

En termes d'analyse de données, nous transmettrons les fondamentaux utilisés tant en TAL qu'en apprentissage machine.

#### Fonctionnement

Les étudiants auront un cours toutes les semaines pendant une heure et demie à Paris Diderot au premier semestre. Les étudiants travailleront en binôme : un étudiant(e) angliciste de Diderot en L3 (dans le cadre du projet étudiant de L3) et un étudiant en informatique de Paris Diderot ou de Paris Descartes. Les séances alterneront entre informatique et linguistique toutes les semaines. Après deux séances introductives, le suivi se fera au plus près des

projets conduits par chaque binôme. L'enseignement serait effectué en français mais les productions seraient en langue anglaise (abstract de conférence, blog, poster et article collectif final).

### **Déroulé type des apprentissages :**

Voici les grands axes du séminaire de licence

L'art d'écrire un «abstract», chercher des références, et utiliser Zotero/Bibtex

Architecture d'un article scientifique, utiliser LaTeX pour rédiger

Case d'utilisation n°1 – Les bases de la morphophonologie, le dictionnaire phonétique de CMU

Nettoyer les données, structurer les données (regex), publier sur Github

Bibliothèque python, création d'un Jupyter Notebook

Soumettre un abstract

Approches computationnelles de la linguistique (treetagging et PoS-tagging)

Bibliothèque python : manda/panda

Le corpus Brown et la bibliothèque Linguistica

Autres analyseurs morphologiques : Chipmunk, ParaMA

Parser des données : la bibliothèque spaCy

L'apprentissage par machine en python : Scikit-learn

### **Complémentarité de l'équipe pédagogique :**

L'UFR EA a une expérience pour projeter la recherche de ses étudiants, a instauré une journée de présentation des travaux en M2 depuis 2010. Une collection HAL a été créée en 2018 et qui recense les travaux de Master du CLILLAC-ARP.

L'UFR d'informatique de Paris Diderot a un Master recherche reconnu internationalement comme de très haut niveau et qui s'appuie sur la FSMP (Fondation des Sciences Mathématiques de Paris). Toutefois, peu nombreux sont les étudiants issus des filières classiques qui intègrent un tel Master. Ce projet aiderait à «dédramatiser» ou «décomplexer» certains d'entre eux à persévérer pour continuer dans la recherche et à tout le moins viser un possible Doctorat.

### **Valorisation étudiante :**

La participation à cet enseignement sera valorisée à hauteur de 3 ECTS, sera pour les étudiants de linguistique proposé sous la forme d'un enseignement «projet étudiant» et pour les étudiants d'informatique sera valorisé comme enseignement de programmation.

Impacts attendus :

Féminisation des parcours

Coloration TAL de masters Data Science

### **Livrables (attendus) :**

Abstract de conférence à soumettre

Visibilité des projets conduits par les étudiants via un outil de partage Github (ou équivalent)

Articles de blog scientifique

Posters présentés mi janvier aux étudiants du master 2 à Paris Descartes "Machine Learning for Data Science", sous la responsabilité de Mohamed Natif.

Et, dans l'hypothèse la plus favorable, un article soumis en collectif au semestre 2 (nécessité alors d'un filage sur l'année). Les productions scientifiques se rattacheront à la conférence RECITAL, conférence jeunes

chercheur.euse.s rattachée à la conférence TALN (Traitement Automatique des Langues Naturelles)  
<https://www.irit.fr/pfia2019/taln-recital/taln-recital-soumission/#cfp-recital>

### **indicateurs**

Valorisation des productions par une collection HAL des publications IReEL réalisées avec le soutien de l'IDEX  
Outre les indicateurs suggérés par l'appel (taux de poursuite en master), le projet RELIA propose de prendre en compte :

- Le nombre d'étudiantes associées (rééquilibrage en parité autant que faire se pourra dans le master informatique)

### **Un exemple d'étude de cas :**

Analyse automatique de la morphologie à partir de la segmentation automatique non supervisée pour les dictionnaires du XVIIIe siècle : le logiciel ParaMA (PARAdigm-based Morphological Analyzer) a fait l'objet d'une publication référencées et est disponible sur Github. Les informaticiens analyseraient le code et les opérations importantes des algorithmes d'analyse morphologique ; ils fourniraient aussi une aide au débogage et à la compréhension du fonctionnement. Les anglicistes aideraient à l'analyse morphologique des candidats produits pour les décompositions en composés ou en suffixés.

## 4. BUDGET

Nous avons respecté le principe du financement forfaitaire de 10 000 ou de 15 000 euros envisagé par l'appel, même si nous pensons qu'un prolongement de l'expérimentation pourrait se faire à moindre coût.

Nous avons budgété des pauses cafés régulières afin de motiver les étudiants, de souder les binômes ainsi qu'un repas de cohésion pour lancer le projet.

Nous avons prévu un budget de 1 100 euros par étudiant pour assister à une conférence, permettant de couvrir par exemple trois jours de la conférence TALN (Traitement Automatique des Langues Naturelles) et son édition satellite de sa conférence jeunes chercheur·euse·s RECITAL

<https://www.irit.fr/pfia2019/taln-recital/taln-recital-soumission/#cfp-recital>

<https://www.irit.fr/pfia2019/inscription/>

| Catégorie                                       | Montant       | Nature   |
|---|---------------|--|
| Gratification des stages                        |               |  |
| Prise en charge de colloques/congrès/séminaires | 11 000€       | 10 étudiants, hébergement/inscription/transports   |
| Dépenses RH (autres que stages)                 | 2000€         | Compensation heures UFR<br>12 h UFR EA (516 €)<br>12h UFR informatique Diderot (516 €)<br>3h UFR Mathématiques-Informatique Paris Descartes (129 €)<br>(taux horaire de 43 euros pour les titulaires)<br><br>invitations conférenciers |
| Dépenses fonctionnement                         | 2000€         | Frais pauses café et repas<br><br>Impression posters<br>120 euros  |
| Autre   |               |  |
| <b>Total</b>                                    | <b>15000€</b> |  |

## **5. CURRICULUM VITAE (JOINDRE 2 PAGES MAX.)**

### **M. Nicolas Ballier (UFR Etudes anglophones, CLILLAC-ARP)**

Ne sont présentés que très succinctement les éléments du curriculum en rapport avec le projet.

Détails en [http://www.clillac-arp.univ-paris-diderot.fr/user/nicolas\\_ballier](http://www.clillac-arp.univ-paris-diderot.fr/user/nicolas_ballier)

1997 : Docteur Université Paris X Nanterre  
MdC Université de Rouen, département d'anglais

2005 : Habilité à Diriger des Recherches  
Professeur de linguistique anglaise à Paris 13

2009 Professeur de linguistique anglaise à Paris Diderot

#### **Actions et productions scientifiques :**

>50 productions : articles, conférences, chapitres de livres, édition d'actes, etc.

Membre de comités scientifiques de nombreux colloques nationaux et internationaux

Rapporteur/Membre de jury de plus de 40 thèses et habilitations

Gestion de plusieurs projets, dont deux projets européens pour des cotutelles de thèse

#### **Actions et productions pédagogiques :**

création de plusieurs modules sous moodle

prise en charge de nombreux cours au cours des années, service varié,

cours d'initiation à la recherche en master 1, cours sur les outils informatiques en master1

cycles de formation à R (masterants et doctorants)

Création de la journée des M2, présentation des travaux des étudiants en linguistique anglaise

2010, prolongé depuis en une journée des travaux des étudiants en linguistique des masters d'UFR études anglophones, de linguistique et d'EILA.

#### **Administration**

Ancien responsable de la sélection des étudiants Campusfrance pour le master d'études anglophones

Membre de nombreuses commissions et comités au fil des ans, comités de sélection : Grenoble, Limoges, Nice, Clermont-Ferrand, Paris 8

#### **TICE**

Montage d'un intranet à Rouen pour la licence de science du langage bilingue

Lauréat d'un projet (réalisation avec Franck Zumstein d'un moodle pour la formation des professeurs de langue à l'informatique, le C2I2E)

Supervision de l'installation des logiciels informatiques pour les anglicistes à Paris Diderot

#### **Expérience de la pluridisciplinarité**

Coordinateur d'une licence bilingue science du langage à Rouen (nouvelles technologies, traduction automatique)

Ancien porteur d'axe dans des unités de recherches pluridisciplinaires (à Rouen)

## **M. Jean-Baptiste Yunès (UFR informatique, IRIF)**

Pour les détails prière de consulter les éléments disponibles en ligne (<http://www.irif.fr/~yunes/>)  
Ne sont présentés que très succinctement les éléments du curriculum en relation avec le projet.

1993 : Docteur Université Paris 7  
MdC Université Paris 7 à l'UFR d'Informatique  
2007 : Habilité à Diriger des Recherches

### **Actions et productions scientifiques :**

>50 productions : articles, conférences, chapitres de livres, édition d'actes, etc  
Membre du comité éditorial de la revue JCA  
Rapporteur/Membre jury de thèse et habilitation  
Membre de plusieurs projets FUI, ANR

### **Actions et productions pédagogiques :**

création de nombreux cours au cours des années  
prise en charge de nombreux cours au cours des années, service varié  
production de +100h de cours et tutoriaux filmés  
initiateur de la création d'une salle orientée pédagogie innovante  
deux livres publiés (Dunod, Ellipses) :  
Jean-Marie Rifflet et Jean-Baptiste Yunès (2014) *Fondements de la Programmation : Concepts et Techniques*, Ellipses  
Jean-Marie Rifflet et Jean-Baptiste Yunès (2003) *UNIX : programmation et communication*  
Dunod, 2003

### **Administration**

Membre de nombreuses commissions et comités au fil des ans, comités de sélection : Nice, Caen, CNAM

### **TICE**

Anciennement chargé de mission TICE pour l'Université Paris Diderot

### **Expérience de la pluridisciplinarité**

Actuellement responsable du Master 2 parcours PISE (Master pluridisciplinaire Science et SHS)

Paris, le 26 avril 2019

L'UFR d'études anglophones soutient pleinement le Projet IREEL-RELIA  
– recherche en licence informatique et études anglophones

Porté par Nicolas Ballier (UFR études anglophones) et Jean-Baptiste Yunès  
(UFR d'informatique).

Le directeur-adjoint de l'UFR d'études anglophones de l'université Paris Diderot –  
Paris 7



Charles-Edouard Levillain



Pr. H. Fauconnier  
*Directeur de l'UFR informatique*

L'UFR d'informatique soutient pleinement le projet IREEL :  
RELIA - Recherche En Licence Informatique et études  
Anglophones »

porté par N. Ballier (UFR études anglophones) et Jean-  
Baptiste Yunès (UFR d'informatique)

Paris le 25/4/2019

  
Le Directeur de l'UFR d'Informatique  
de l'Université Paris Diderot - Paris 7

Hugues FAUCONNIER