



**Sujet :**

**Développement de ressources pour l'analyse linguistique de l'Anglais**

**Niveau** : Bac + 5 - Master

**Durée du stage** : 4 à 6 mois

**Contexte**

Le Laboratoire LIC2M a développé un logiciel d'analyse linguistique multilingue nommé LIMA. Il inclut pour l'anglais un étiqueteur, un analyseur syntaxique, un moteur de résolution de coréférences, etc. Plusieurs de ces modules utilisent un moteur de reconnaissance d'expressions régulières. Certains de ces modules nécessiteraient d'être mis à jour ou remplacés pour tenir compte des travaux récents dans le domaine.

LIMA utilise des ressources pour effectuer l'analyse: dictionnaires, règles de reconnaissances, listes d'entités prédéfinies, etc. Ces ressources sont construites manuellement pour la plupart.

**Objectifs du stage**

Le travail du stagiaire consistera à effectuer une analyse bibliographique approfondie, à effectuer des propositions d'évolution de certains modules de l'analyseur et à enrichir les ressources de l'anglais utilisées par LIMA de manière à améliorer ses performances.

Cette dernière partie consistera à exploiter les résultats de l'analyseur pour repérer des erreurs et à compléter les ressources pour supprimer ces erreurs.

Ceci se fera en utilisant les outils graphiques développés l'an passé dans le cadre de l'évaluation de l'analyse du français et qui pourront nécessiter un travail de généralisation pour leur application à l'anglais.

**Domaines de spécialité requis :**

Linguistique, informatique, morphologie, syntaxe

**Moyens informatiques mis en œuvre :**

Langages : langages de scripts (Bash, Perl, ...)

Logiciels : LIMA, GNU/Linux, éditeurs de ressources

**Autres moyens mis en œuvre (expériences, méthodes d'analyses, autres...)**

**Centre** : Fontenay-aux-Roses

**Pôle ou Direction** : DRT - LIST

**Dépt/Service/Labo** : DTSI/SRCI/LIC2M :

Laboratoire d'Ingénierie de la Connaissance Multimédia Multilingue

**Web** :

<http://www-list.cea.fr/>



## Encadrement

Nom de l'ingénieur responsable : Gaël De Chalendar	Tél. : +33 1 46 54 80 18
E-mail : <a href="mailto:gael.de-chalendar@cea.fr">gael.de-chalendar@cea.fr</a>	Fax : +33 1 46 54 75 80
Secrétariat : DTSI/SRCI	Tél. : +33 1 46 54 91 17
Nom du chef de laboratoire : Olivier Mesnard	Tél. : +33 1 46 54 73 38