

L'ENSEIGNEMENT SUPERIEUR ET LA RECHERCHE DANS LE CONTEXTE DES GRANDS-LACS ET AFRIQUE DE L'EST: LES ENJEUX LINGUISTIQUES POUR L'INTEGRATION REGIONALE

UNIVERSITE NATIONALE DU RWANDA
NATIONAL UNIVERSITY OF RWANDA

INITIATION A LA LINGUISTIQUE DE CORPUS ET A LA PLATE-FORME *NOOJ*

JOURNEE DE FORMATION
RESEAU LEXICOLOGIE, TERMINOLOGIE, TRADUCTION (LTT)

Prof. Xavier Blanco Escoda
(Universitat Autònoma de Barcelona)

VENDREDI 23 MARS 2012

Présentation de *Nooj*

Nooj est un environnement de développement linguistique capable de formaliser neuf niveaux de phénomènes linguistiques, du niveau orthographique au niveau sémantique (www.nooj4nlp.net).

Les ressources linguistiques peuvent être appliquées à des textes et corpus, et ainsi confirmer ou infirmer ces formalisations (dans une optique de linguistique de corpus par exemple), ou alors pour annoter les textes (pour des applications d'analyse de discours).

Des ressources linguistiques au format *Nooj* sont déjà disponibles en téléchargement libre pour une dizaine de langues ; des équipes travaillent sur une vingtaine d'autres langues.

Nooj est utilisé comme analyseur de discours (analyse de textes littéraires ou d'interviews psychologiques, analyse de textes techniques ou politiques) et comme outil d'enseignement des langues secondes (les "Labs" et "Projects" de *Nooj* sont des supports de cours).

Les objectifs de la formation

Le but de cette formation est de donner aux participants les compétences nécessaires pour qu'ils puissent développer leur propre projet, de façon autonome. À l'issue de la formation, les participants pourront constituer eux-mêmes des corpus informatisés, les interroger et les analyser. Ils pourront développer eux-mêmes des dictionnaires et des grammaires électroniques, les tester et les maintenir. Ils pourront enfin construire des analyseurs capables d'annoter automatiquement leurs corpus pour des applications d'analyse de discours et d'enseignement des langues.

Programme

Nous fournissons le logiciel *Nooj*, des ressources linguistiques, son manuel, des corpus de travail et des références (tutoriels, articles) sur un cédérom. Les participants qui le désirent peuvent travailler sur leur propre ordinateur portable. Nous travaillerons sur quatre projets :

(1) traiter des textes et des corpus ; importation (formats, XML) ; interroger les textes (requêtes, expressions rationnelles) ; concordances (index, coloriage de textes) ; analyses statistiques.

Les participants sont invités à amener leurs propres corpus pour les analyser.

(2) créer, tester et appliquer des dictionnaires ; décrire la morphologie (flexionnelle, dérivationnelle)

Les participants qui désirent travailler sur une autre langue que le français sont invités à nous contacter.

(3) créer, tester et appliquer des grammaires locales ; les réutiliser et les accumuler ; extraction d'entités nommées et analyse de discours

(4) annoter des textes : désambiguïsation et analyse structurelle ; construire des paraphrases automatiquement ; traduction automatique de termes et d'expressions figées.

Formateur

Xavier Blanco est directeur du laboratoire FLEXSEM (*Fonètica, Lexicologia i Semàntica*) de l'Université autonome de Barcelone. Il est spécialisé en lexicographie bilingue et traduction. Il enseigne au master en traitement de l'information et de la communication multilingue (UAB) Il est l'auteur de plusieurs dictionnaires électroniques à large couverture.

Inscription

Les places étant limitées, les personnes intéressées par cette formation doivent manifester leur intérêt dès que possible à l'adresse conferenceunr@bi.refer.org en y joignant un CV.
