



JOURNEES DE FORMATION

RESEAU DE CHERCHEURS « LEXICOLOGIE, TERMINOLOGIE ET TRADUCTION »

LANGUE FRANÇAISE, DIVERSITE CULTURELLE ET LINGUISTIQUE

AGENCE UNIVERSITAIRE DE LA FRANCOPHONIE

LISBONNE 13 ET 14 OCTOBRE 2009

Conformément à une tradition de plus de deux décennies, les [Huitièmes Journées scientifiques du réseau Lexicologie, terminologie, traduction](#) (Lisbonne, 15-17 octobre) seront précédées de deux journées de formation. Au vu de l'immense succès des journées de formation de Bruxelles (septembre 2005), le comité organisateur a choisi de persévérer dans la voie de l'exploitation des corpus textuels appréhendés dans une perspective multilingue. Soucieux de prendre en compte une diversité de besoins et une hétérogénéité de compétences, nous avons choisi de proposer deux programmes distincts qui s'étaleront chacun sur deux jours. Ils se dérouleront dans les locaux de l'Université nouvelle de Lisbonne les 13 et 14 octobre sous la direction de formateurs chevronnés.

- Le premier programme, « **Linguistique de corpus : constitution et traitement des corpus de spécialité** », permettra de s'initier à la linguistique de corpus appliquée aux langues spécialisées en contexte mono ou multilingue. Il permettra d'évoluer de la problématique de constitution du corpus à son exploitation à l'aide de concordanciers et d'aligneurs pour exploiter au mieux l'information lexicale ou sémantique. Cette formation sera assurée par Manuel Célio Conceição, Rute Costa et Raquel Silva, membres de l'Unité de recherche en « Lexicologie, lexicographie et terminologie », du Centre de linguistique de l'Université nouvelle de Lisbonne ([CLUNL-LLT](#)). [\[Voir le programme détaillé\]](#)

- Le second programme, « **Initiation à la plate-forme Nooj** », s'adressera à celles et ceux qui, déjà initiés à la problématique des corpus, souhaitent apprendre à maîtriser la plate-forme *Nooj*. Véritable outil de développement de recherches linguistiques, *Nooj* – évolution d'*Intex* – est utilisé par une large communauté de chercheurs travaillant sur une dizaine de langues différentes. Il permet de développer ses propres dictionnaires et grammaires électroniques en vue de décrire et de formaliser des connaissances sur les langues naturelles dans un grand nombre de perspectives (syntaxique, lexicale, sémantique, ontologique, didactique...). Cette seconde formation sera animée par le concepteur de *Nooj*, Max Silberstein ([LASELDI](#), Université de Franche-Comté, et [CLI](#), INALCO) et par Xavier Blanco ([FLEXSEM](#), Université autonome de Barcelone). [\[Voir le programme détaillé\]](#)

Les journées de formation offrent une occasion unique de se former à des méthodologies et des instruments au contact d'autres praticiens. Elles constituent un espace de rencontre et de dialogue pour la communauté des membres du réseau LTT avant les journées scientifiques proprement dites.

[Vers le formulaire d'inscription](#)
(à retourner avant le 31 juillet 2009)

JOURNEES DE FORMATION LTT

PROGRAMME 1

LINGUISTIQUE DE CORPUS

CONSTITUTION ET TRAITEMENT DES CORPUS DE SPECIALITE

Présentation de la formation

La formation préparée par l'Unité de recherche en « Lexicologie, lexicographie et terminologie », du Centre de linguistique de l'Université nouvelle de Lisbonne ([CLUNL-LLT](#)) se fonde sur les compétences et l'expérience actuelle du Centre. Son activité de recherche est centrée sur la construction de terminologies mono- et multilingues à partir de corpus de spécialité, en ayant recours à des outils de traitement semi-automatique de la langue.

Plusieurs projets actuellement menés au CLUNL-LLT sont directement en lien avec le thème de la formation et serviront de support à la formation. En particulier, le Centre travaille à l'élaboration de terminologies, des réseaux sémantiques et lexicaux, des bases de données textuelles et de connaissances terminologiques en collaboration avec des entités telles que l'Assemblée de la République portugaise ou la Faculté de médecine de l'Université nouvelle de Lisbonne, entre autres. Ces expériences prennent également en compte les besoins des institutions lusophones, avec lesquelles le CLUNL-LLT collabore. Dans le cadre de la description terminologique, sont menées, par exemple, des études multilingues du vocabulaire du droit et de la médecine.

Les objectifs de la formation

La pertinence et l'importance de l'observation de l'analyse des corpus de spécialités pour l'organisation et la structuration des connaissances justifient la proposition par le réseau LTT d'une formation qui permette de présenter des méthodologies de travail basées sur des présupposés théoriques solides en terminologie, l'objectif étant de diversifier les regards sur la constitution et le traitement de corpus. Le CLUNL-LLT a une longue expérience de constitution de corpus de spécialités et de conception de bases de données textuelles. Les corpus sont constitués d'après des critères spécifiques qui permettent qu'ils soient utilisés à des fins d'organisation des connaissances, ainsi qu'à des descriptions fines de(s) discours de spécialité(s).

Programme

1. Principes théoriques et méthodologiques en terminologie
2. Critères pour la constitution de corpus de spécialité monolingues, corpus parallèles
3. Utilisation de concordanciers, aligneurs de textes, etc.
4. Collocations, reformulations

5. Marqueurs linguistiques et connecteurs pour l'extraction de l'information à partir de corpus de spécialité
6. Systématisation de l'information obtenue à partir du traitement de corpus
7. Réseaux sémantiques et lexicaux
8. Applications à la gestion des ressources terminologiques

Formateurs

Les formateurs, Manuel Célio Conceição, Rute Costa, Raquel Silva, sont membres de l'Unité de recherche en « Lexicologie, lexicographie et terminologie », du Centre de linguistique de l'Université nouvelle de Lisbonne. Ils sont tous les trois fortement impliqués dans les travaux de recherche menés par le CLUNL-LLT.

JOURNEES DE FORMATION LTT
PROGRAMME 2
INITIATION A LA PLATE-FORME NOOJ

Présentation de Nooj

Nooj est un environnement de développement linguistique capable de formaliser neuf niveaux de phénomènes linguistiques, du niveau orthographique au niveau sémantique (<http://www.nooj4nlp.net/pages/nooj.html>).

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.

Formateurs

Max Silberztein est professeur à l'Université de Franche-Comté et à l'Institut national des langues et civilisations orientales (INALCO). Il est membre du Laboratoire de sémiotique linguistique didactique et informatique ([LASELDI](#)), et du Cercle de linguistique ([CLI](#)) de l'INALCO. Le LASELDI est spécialisé dans l'analyse de discours littéraires et politiques. Le CLI est spécialisé dans la formalisation des langues enseignées à l'INALCO.

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.
