

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
INALCO, 11-15 janvier 2016

NooJ est un environnement de développement utilisé pour formaliser huit niveaux de phénomènes linguistiques : orthographe et typographie, morphologie flexionnelle et dérivationnelle, syntaxe locale et structurelle, grammaire transformationnelle et analyse sémantique. NooJ contient des outils formels adaptés à chaque type de phénomène (grammaires rationnelles, hors contexte, contextuelles et non restreintes) et permet de développer des ressources à large couverture pour de nombreuses langues.

NooJ est aussi utilisé pour analyser des corpus de textes par des linguistes mais aussi par des chercheurs en sciences sociales pour effectuer des analyses historiques, littéraires, sociologiques ou psychologiques de corpus, et aussi par des entreprises pour extraire et annoter des informations scientifiques et techniques, notamment en intelligence économique. NooJ est gratuit et open source et fonctionne sous Windows, Mac OSX, LINUX et Unix, cf. : www.nooj4nlp.net.

Le stage s'adresse particulièrement aux étudiants de Master, doctorants et chercheurs intéressés par la linguistique et la linguistique de corpus ainsi que par l'analyse automatique de textes dans une perspective de sciences humaines et d'application industrielle. Les étudiants en Master qui peuvent et désirent valider le stage auprès de leur université doivent me prévenir dès leur inscription. Le stage est gratuit et dure une semaine : les matins sont dédiés au cours et aux travaux dirigés ; pendant les après-midis, des chercheurs et doctorants présenteront diverses applications de NooJ.

ATTENTION : les places sont limitées : il faut s'inscrire en envoyant son nom, son statut (étudiant / doctorant / chercheur / industriel) et le nom de son institution à max.silberztein@univ-fcomte.fr.

Lundi 11 janvier

9H-12H30 : **Fouille de texte** : textes et corpus, requêtes et concordances, analyses statistiques.

14H-15H : Vincent Bénet (INALCO) : *Ecrire des dictionnaires et des grammaires pour NooJ (exemple du russe)*

15H-16H : Hamid Annouz (INALCO) : *Traitement morphologique de la dérivation verbale du kabyle.*

Mardi 12 janvier

9H-12H30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires, morphologie.

14H-14H40 : Andrea Rodrigo (Univ. Nacional de Rosario, Argentine) : *A proposal for the treatment of clitics in rioplatense Spanish verbs using NooJ*

14H40-15H20 : Zhen Cai (Univ. de Franche-Comté) : *Le dictionnaire électronique du chinois et son utilisation par le logiciel NooJ.*

15H20-16H : Valérie Clerc (Univ. d'Aix-Marseille) : *Analyse lexicale automatique du japonais.*

Mercredi 13 janvier

9H-12H30 : **Syntaxe locale** : requêtes syntaxiques, grammaires locales, annotation automatique de textes, extraction d'entités nommées.

14H-15H : Maximiliano Duran (Univ. de Franche-Comté). *Graphes flexionnels pour la description des verbes composés en quechua.*

15H-16H : Mario Monteleone (Univ. de Salerne, Italie). *Les grammaires locales et la sémantique formelle.*

Jeudi 14 janvier

9H-12H30 : **Syntaxe structurelle** : Grammaires hors contexte, contraintes d'accord et arbres.

14H-15H : Mourad Aouini (Univ. de Franche-Comté) : *Extraction des entités nommées en Moyen Français.*

15H-16H : Héra Fehri (Univ. de Gabès, Tunisie) : *Extraction des plantes médicinales à partir des corpus parallèles Arabe/Français.*

Vendredi 15 janvier

9H-12H30 : **Sémantique** : Analyse transformationnelle et génération de paraphrase automatique, levée d'ambiguïté sémantique, traduction

14H-15H : Slim Mesfar (Univ. de Carthage, Tunisie) : *Annotations sémantiques automatiques : Application à la fouille d'opinion dans le domaine politique.*

15H-16H : Philippe Lambert (Univ. de Lorraine) : *Knowledge Management pour les nanotechnologies.*