

Semaine intensive NooJ

Max Silberztein, ADCOST-ELLIADD, Université de Franche-Comté

INALCO, 20-24 janvier 2014

65 rue des Grands Moulins, 75013 Paris

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 informatiques adaptés à chaque type de phénomène (grammaires rationnelles, hors contexte, contextuelles et non restreintes) et permet de développer des ressources à large couverture dans une perspective de linguistique descriptive.

NooJ est utilisé par des linguistes pour formaliser les langues et appliquer les descriptions linguistiques à des corpus de textes, 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, est disponible en open source JAVA 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 descriptive et la linguistique de corpus ainsi que par l'analyse automatique de textes dans une perspective de sciences humaines et 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 20 janvier

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

14H-15H : Vincent Bénét (INALCO) : Problèmes de conception des ressources dictionnaires pour NooJ : entrées, étiquettes morphologiques et sémantiques, paradigmes (exemple du russe).

15H-16H : Hamid Annouz (INALCO) : les verbes berbères.

Mardi 21 janvier

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

14H-15H : Nicolas Boffo (Univ. Montpellier 3, Institut Polytechnique de Hanoi) : la temporalité en vietnamien.

15H-16H : Xavier Blanco (Univ. Autonome de Barcelone) : Élaboration de dictionnaires électroniques pour la langue espagnole.

Mercredi 22 janvier

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

14H-15H : Denis Le Pesant (Modyco) : Problèmes d'analyse automatique des sentiments.

15H-16H : Valérie Clerc (INALCO) : le module japonais de NooJ.

Jeudi 23 janvier

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

14H-15H : Meng Sun (INALCO) : analyse d'opinion en chinois.

15H-16H : Maximiliano Duran (Univ. de Franche-Comté) : les verbes du quetchua.

Vendredi 24 janvier

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

14H-16H : Philippe Lambert (Institut Jean Lamour CNRS) : NooJ pour le Knowledge Management en Nanosciences : de l'analyse du besoin à l'indexation de ressources documentaires.

15H-16H : Anna Pappa (Univ. Paris 8) : grammaire de désambiguïsation de l'article défini / pronom personnel pour le français/grec.