

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
INALCO, 8-12 janvier 2018

NooJ est un environnement de **développement linguistique** utilisé pour formaliser dix niveaux de phénomènes : orthographe et typographie, morphologie flexionnelle, dérivationnelle et agglutinative, syntaxe locale, structurelle (en constituants et de dépendance), grammaire transformationnelle et analyse sémantique. NooJ contient des outils formels adaptés à chaque type de phénomène (grammaires rationnelles, hors contexte, contextuelles et non restreintes) ainsi que des outils de développement pour créer, tester, corriger et partager des ressources linguistiques à large couverture, pour de nombreuses langues.

NooJ est aussi utilisé comme outil de **linguistique de corpus** pour analyser des corpus de textes par des linguistes mais aussi pour effectuer des analyses techniques, historiques, littéraires, sociologiques ou psychologiques de corpus. Les moteurs linguistiques de NooJ ont été intégrés dans de nombreuses **applications de Traitement Automatique des Langues** (reconnaissance d'entités nommées, génération automatique de textes, intelligence économique, moteur de recherche linguistique, production automatique de paraphrases, web sémantique, traduction automatique).

NooJ est gratuit et open source et fonctionne sous Windows, Mac OSX, LINUX et Unix, cf. : www.nooj4nlp.net.

Ce stage s'adresse particulièrement aux étudiants de Master, doctorants et universitaires intéressés par la formalisation linguistique, les outils d'exploration de corpus et les applications du TAL. Les étudiants en Master qui peuvent et désirent valider le stage auprès de leur université doivent me prévenir dès leur inscription. Le stage dure une semaine : les matins sont dédiés au cours et aux travaux dirigés ; pendant les après-midis, des chercheurs présenteront diverses applications de NooJ.

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

Lundi 8 janvier

9H-12H30 : **Fouille de texte** : Echecs des approches stochastiques. Textes et corpus, requêtes et concordances.
14H-14H40 : Maximilian Duran (Univ. de Franche-Comté) : Morphologie des mots composés du quechua.
14H40-15H20 : Alena Saint Joanis (Univ. de Franche-Comté) : Module ukrainien pour NooJ : les verbes.

Mardi 9 janvier

9H-12H30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires, morphologie.
14H-14H40 : Mathieu Roy (INALCO) : le module NooJ de kiswahili.
14H40-15H20 : Imen Enasri (Univ. de Franche-Comté) : Analyse automatique de textes juridiques.
15H20-16H00 : Zhen Cai (Univ. de Franche-Comté) : Dictionnaire électronique du chinois.

Mercredi 10 janvier

9H-12H30 : **Syntaxe locale** : requêtes syntaxiques, grammaires locales, annotation de textes, entités nommées.
14H-14H40 : Masako Watabe (Univ. Paris 4) : Locutions post-positionnelles du romani.
14H40-15H20 : Hamid Annouz (INALCO) : intégration du kabyle dans la plateforme NooJ.
15H20-16H : Mourad Aouini (Univ. de Franche-Comté) : Etiquetage morpho-syntaxique du moyen-français.

Jeudi 11 janvier

9H-12H30 : **Syntaxe structurelle** : Grammaires hors contexte, contraintes d'accord et arbres.
14H-14H40 : Denis Le Pesant (Modyco). Application des dictionnaires LVF et DEM.
14H40-15H20 : Mario Monteleone (Univ. de Salerne). Grammaires NooJ et algorithmes éthiques : bloquer les discours de haine en ligne.
15H20-16H00 : Essia Bessaies (Univ. de Carthage, Tunisie). Génération des modèles de réponse pour les questions factuelles en arabe.

Vendredi 12 janvier

9H-12H30 : **Sémantique** : Analyse transformationnelle et génération de paraphrase automatique, levée d'ambiguïté sémantique, le WEB sémantique.
14H-14H40 : Andrea Rodrigo (Univ. Nacional de Rosario, Argentine) : Automatic Treatment of Adverbs: a pedagogical application for teaching Spanish as a foreign language.
14H40-15H20 : Héla Fehri (Univ. de Gabès, Tunisie) : un conjugueur automatique trilingue avec NooJ.