

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
INALCO

65 rue des Grands Moulins, 75013 Paris, salle s5.28
7-11 janvier 2019

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

NooJ est utilisé comme outil de linguistique de corpus pour analyser des corpus de textes par des linguistes, et dans les humanités numériques pour effectuer des analyses historiques, littéraires, psychologiques ou sociologiques de corpus. NooJ a aussi été intégré à de nombreuses applications de TAL : reconnaissance d'entités nommées, génération automatique de textes, informatique décisionnelle, interface en langue naturelle, moteur de recherche conceptuel, production automatique de paraphrases, web sémantique, traduction automatique.

NooJ est soutenu par le programme Metashare de la Communauté européenne, est gratuit et open source et fonctionne sous Windows, Mac OSX, LINUX et Unix. Il existe des modules de ressources linguistiques au format NooJ pour plus de 20 langues, cf. : www.nooj-association.org.

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 : le stage est gratuit, mais les inscriptions sont obligatoires et 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 7 janvier

9H-12H30 : **Fouille de texte** : Limite des approches stochastiques. Corpus, requêtes et concordances.

14H-14H40 : Jacqueline Léon (CNRS UMR 7597) : Réception de la TA et débuts du TAL en France.

14H40-15H20 : Hamid Annouz (INALCO) : Traitement de l'homographie du kabyle.

15H20-16H : Zhen Cai (Univ. de Franche-Comté) : Morphologie du chinois.

Mardi 8 janvier

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

14H-14H40 : Denis Le Pesant (Modyco) : Phrases figées du type « Est-ce que je sais ? Ben voyons ! Tu plaisantes ! ».

14H40-15H20 : Maximiliano Duran (Univ. de Franche-Comté) : Analyse transformationnelle des phrases transitives quechua.

15H20-16H : Mourad Aouini (CNRS UPS2259) : Analyse lexicale et morpho-syntaxique du moyen-français.

Mercredi 9 janvier

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

14H-14H40 : Masako Watabe (Univ. Paris 4) : Morphologie productive pour les postpositions en romani.

14H40-15H20 : Alena Saint-Joanis (Univ. de Franche-Comté) : Traitement des verbes ukrainiens avec NooJ.

Jeudi 10 janvier

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

14H-14H40 : Mario Monteleone (Univ. de Salerne, Italie). Ecrire une trame narrative avec NooJ.

14H40-15H20 : Slim Mesfar (Univ. de Carthage, Tunisie). Web sémantique de l'arabe.

Vendredi 11 janvier

9H-12H30 : **Sémantique** : Analyse transformationnelle et génération de paraphrases automatique, levée d'ambiguïté sémantique, le WEB sémantique.

14H-14H40 : Héla Fehri (Univ. de Sfax, Tunisie) : Traduction automatique des verbes de mouvement et de parole arabes.

14H40-15H20 : Tong Yang (Univ. Paris Diderot) : Extraction automatique de phrasèmes nominaux et verbaux.