

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
INALCO, 65 rue des Grands Moulins, Paris 75013

Salle 5.21 (5^e étage)

6-10 janvier 2020

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

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

14H-14H40 : Jacqueline Léon (CNRS UMR 7597) : Histoire des débuts du TAL en France.

14H40-15H20 : Masako Watabe (Univ. de Franche-Comté) : Dictionnaire NooJ du rromani.

15H20-16H : Hamid Annouz (INALCO) : Traitement formel des valeurs de l'aoriste simple du kabyle.

Mardi

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

14H-14H40 : Andrea Rodrigo (Univ. de Rosario, Argentine) : Automatic processing of discourse connectors in Spanish.

14H40-15H20 : Mourad Aouini (CNRS UPS2259) : Extraction des relations entre entités dans un corpus journalistique.

15H20-16H : Michel Delarche (CRL, Univ. Diderot) : un dictionnaire de l'anglais pour NooJ.

Mercredi

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

14H-14H40 : Alena Saint-Joanis (Univ. de Franche-Comté) : Module NooJ pour l'ukrainien.

14H40-15H20 : Tong Yang (Univ. Paris 3) : Extraction automatique de phrasèmes *N-Adj* du discours culinaire.

15H20-16H : Mario Monteleone (Univ. de Salerne, Italie) : NooJ et l'intelligence artificielle : approche anthropique.

Jeudi

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

14H-14H40 : Magali Bigey (Univ. de Franche-Comté) : NooJ et humanités numériques : analyse des discours de réception.

14H40-15H20 : Zhen Cai (Univ. de Franche-Comté) : Segmentation automatique du chinois avec NooJ.

15H20-16H : Maximiliano Duran (Univ. de Franche-Comté) : Transformations morphosyntaxiques du quechua.

16H-16H40 : Slim Mesfar (Univ. de Carthage, Tunisie) : Web sémantique de l'arabe.

Vendredi

9H-12H30 : **Sémantique** : Analyse et génération de paraphrases automatique, WEB sémantique.

14H-14H40 : Denis Le Pesant (Modyco) : Nouvelles ressources linguistiques pour NooJ.

14H40-15H20 : Héla Fehri (Univ. de Sfax, Tunisie) : Construction de jeux éducatifs avec NooJ.

15H20-16H : Philippe Lambert (Univ. de Lorraine) : Veille informationnelle pour les partenariats académiques.

16H-16H40 : Virginie Léthier (Univ. de Franche-Comté) : Quelles rencontres entre NooJ et la tradition d'analyse du discours outillée ?