

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
INALCO, stage organisé en distanciel
11-15 janvier 2021

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 et dans les humanités numériques pour effectuer des analyses historiques, littéraires, psychologiques ou sociologiques de corpus, cf. <http://www.nooj-association.org/atishs>.

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 (.NET), Mac OSX, LINUX et Unix (JAVA). 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 et sera organisé via la plateforme Zoom. Les inscriptions sont obligatoires : il faut s'inscrire en envoyant son nom, son statut (étudiant / doctorant / universitaire / autre) et le nom de son institution à max.silberztein@univ-fcomte.fr.

Lundi

9H-12H30 : **Fouille de texte** : Approches stochastiques et linguistiques. Corpus, requêtes et concordances.
14H-14H40 : Jacqueline Léon (CNRS UMR 7597) : Histoire des débuts du TAL en France.
14H40-15H20 : Michel Delarache (CRL, Univ. Diderot) : Révision d'un dictionnaire de l'anglais pour NooJ.

Mardi

9H-12H30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires, morphologie.
14H-14H40 : Hamid Annouz (INALCO) : Traitement de quelques valeurs verbales du kabyle avec NooJ.
14H40-15H20 : Masako Watabe (INALCO) : Morphologie verbale du rromani avec NooJ.
15H20-16H : Alena Saint-Joanis (Univ. de Franche-Comté) : Module NooJ pour la langue ukrainienne.

Mercredi

9H-12H30 : **Syntaxe locale** : requêtes syntaxiques, grammaires locales, annotation de textes, entités nommées.
14H-14H40 : Mario Monteleone (Univ. de Salerne, Italie) : Morphosyntaxe et sémantique dans le dictionnaire italien.
14H40-15H20 : Maximiliano Duran (Univ. de Franche-Comté) : Génération de paraphrases en quechua avec NooJ.
15H20-16H : Andrea Rodrigo (Univ. de Rosario, Argentine) : Spanish Gerund: Automatic Treatment using NooJ.

Jeudi

9H-12H30 : **Syntaxe structurale** : Grammaires hors contexte, contraintes d'accord et arbres.
14H-14H40 : Magali Bigey (Univ. de Franche-Comté) : NooJ, un outil pour l'analyse des discours de réception.
14H40-15H20 : Héla Fehri (Univ. de Sfax, Tunisie) : Réponse automatique aux questions complexes en Arabe.
15H20-16H : Essia Bessaies (Univ. de la Manouba, Tunisie) : Annotation des questions cause-effet en arabe standard.

Vendredi

9H-12H30 : **Sémantique** : Analyse et génération de paraphrases automatique, WEB sémantique.
14H-14H40 : Nicolas Boffo (CIMI) : Détection, classification et analyse d'annonces suspectes sur le net.
15H20-16H : Tong Yang (Univ. d'énergie électrique de Chine du Nord) : Extraction automatique de phrasèmes du discours culinaire.