

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
Auditorium Dumézil
INALCO, 2 rue de Lille, 75007 Paris
3-7 janvier 2022

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 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 aussi utilisé comme outil de linguistique de corpus et dans les humanités numériques pour effectuer des analyses statistiques sur des corpus historiques, littéraires, psychologiques ou sociologiques de corpus, cf. <http://www.nooj4nlp.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 30 langues, cf. : www.nooj4nlp.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 / tutoriels de NooJ.

ATTENTION : le stage est gratuit mais les places sont limitées. 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.
14H40-16H20 : Hamid Annouz (INALCO) : Intégration du kabyle dans NooJ.
16H20-17H : Masako Watabe (INALCO) : Morphologie verbale du rromani.

Mardi

9H-12H30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires, morphologie.
14H-14H40 : Odile Piton (Univ. Paris 1) : Morphologie de l'albanais
14H40-15H20 : Maximiliano Duran (Univ. de Franche-Comté) : Conjugaison et traduction des verbes pronominaux quechua.
15H20-16H : Ilham Blanchete (Univ. de Kenitra, Maroc) : Using NooJ to build an Arabic acquisition software application.

Mercredi

9H-12H30 : **Syntaxe locale** : Requêtes syntaxiques, grammaires locales, annotation de textes, entités nommées.
14H-14H40 : Nicolas Boffo (CIMI) : Détection, classification et analyse d'annonces suspectes sur le net.
14H40-16H20 : Olena Saint-Joanis (Univ. de Franche-Comté) : Grammaires de désambiguïsation pour l'ukrainien.
16H20-17H : Peter Machonis (Univ. Internationale de Floride, Etats-Unis) : Verbes à particule chez les auteurs anglais et américains du XIXe siècle.

Jeudi

9H-12H30 : **Syntaxe structurelle** : Grammaires hors contexte, contraintes d'accord et arbres.
14H-15H : Magali Bigey (Univ. de Franche-Comté) : Réseaux sociaux et infox par l'image : analyse de réception.
15H-16H : Michel Delarche (Univ. Paris Diderot) : Enrichissement des ressources lexicales anglaises de NooJ.

Vendredi

9H-12H30 : **Sémantique** : Analyse et génération de paraphrases automatique, WEB sémantique.
14H-15H : Mario Monteleone (Univ. de Salerne, Italie) : *La Transfer Grammar* de Z. S. Harris avec NooJ
15H-16H : Hela Fehri (Univ. de Sfax, Tunisie) : Construction d'un jeu éducatif multilingue arabe-français-anglais.